COLECCIÓN: EL NÚMERO DE ORO
DIRECTOR: *Act. Alberto Landro*

Acerca de la Probabilidad – 2da. edición
Alberto Landro

Álgebra para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Análisis Matemático I para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Análisis Matemático II para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Los matemáticos que hicieron la historia
Alejandro E. García Venturini

Análisis de Series de Tiempo, univariadas y multivariadas
Heriberto Urbisaia – Juana Brufman

Decisión Estadística Bayesiana, a modo de introducción
Emma Fernández Loureiro de Pérez

Estadística no Paramétrica, a modo de introducción
Emma Fernández Loureiro de Pérez

Teoría de los Conjuntos Borrosos, a modo de introducción
Emma Fernández Loureiro de Pérez

Estadística: Herramientas de Inferencia
Gabriela Kurincic

Estadística: Probabilidades y Distribuciones
Gabriela Kurincic

Los Métodos Cuantitativos en las Ciencias Sociales
Alejandro E. García Venturini – Federico Castelli

Aplicaciones del Análisis Matemático a la Economía
Blanca R. Vitale

Modelos para el Análisis de Series de Tiempo
Juan Carlos Abril

Análisis Matemático I para estudiantes de Ingeniería
Alejandro E. García Venturini – Mónica Scardigli

Cálculo Financiero
Juan R. Garnica Hervás - Esteban O. Thomasz - Romina P. Garófalo

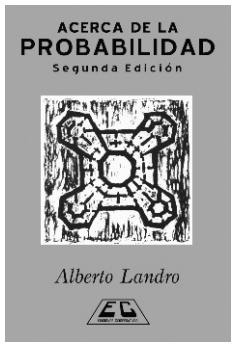
ELEMENTOS DE ECONOMETRÍA DE LOS FENÓMENOS DINÁMICOS

Parte I

LOS PROCESOS ESTOCÁSTICOS LINEALES UNIDIMENSIONALES

De los mismos autores

Acerca de la probabilidad



CENTRO DE INVESTIGACIONES EN ECONOMETRÍA
FACULTAD DE CIENCIAS ECONÓMICAS
UNIVERSIDAD DE BUENOS AIRES

ALBERTO H. LANDRO MIRTA L. GONZÁLEZ

ELEMENTOS DE ECONOMETRÍA DE LOS FENÓMENOS DINÁMICOS

Parte I

LOS PROCESOS ESTOCÁSTICOS LINEALES UNIDIMENSIONALES

EDICIONES COOPERATIVAS

Ninguna parte de esta publicación, incluido el diseño de cubierta puede ser reproducida, almacenada o transmitida en manera alguna ni por ningún medio, ya sea electrónico, mecánico, óptico de grabación o de fotocopia sin permiso previo del Editor. Su infracción está penada por las leyes 11723 y 25446.



Landro, Alberto H.
Elementos de econometría de los fenómenos dinámicos: Parte I : los procesos estocásticos lineales unidimensionales / Alberto Landro y Mirta L. González. - 1a ed. - Buenos Aires: Ediciones Cooperativas, 2009.
600 p. ; 26x19 cm.

ISBN 978-987-652-030-0

1. Econometría. I. González, Mirta L. II. Título
CDD 330

© 2009 Alberto H Landro – Mirta L. González
Derechos exclusivos

© 2009 Ediciones Cooperativas
Tucumán 3227 (1189)
Buenos Aires – Argentina
☎ (54 011) 3528-0466 / (15) 4937 6915
http://www.edicionescoop.org.ar
info@edicionescoop.org.ar

1º edición, Febrero 2009

Colección: *El número de oro*
Director: *Act. Alberto H. Landro*

Imagen de tapa: *El enigma del tiempo* – Rogelio Polesello

Hecho el depósito que establece la ley 11.723

Impreso y encuadernado por:
Imprenta Dorrego. Dorrego 1102, C.A.B.A.
1ª. ed. Tirada: 100 ejemplares. Se terminó de imprimir en Febrero 2009.

Editorial asociada a:
CÁMARA ARGENTINA DEL LIBRO

IMPRESO EN ARGENTINA – PRINTED IN ARGENTINE

¡Viva la Pepa!

Prólogo

"No puede haber sino borradores. El concepto de texto definitivo no corresponde sino a la religión o al cansancio".
(J.L. Borges, Prefacio al Cementerio marino de Valéry)

Comenzamos a escribir este trabajo hace varios años. Afin de perfeccionarlo -afán absurdo que nunca finaliza- lo sometimos a múltiples revisiones y luego, de acuerdo, con la recomendaciones de un sabio maestro, lo dejamos reposar. Pero, este reposo, como todos los reposos, tiene un término y consideramos que ha llegado a su fin pues corremos el riesgo de que todo el material aquí analizado envejezca demasiado y se convierta en una mera referencia histórica. De modo que, con todos los defectos inherentes a un texto provisorio, decidimos finalmente entregarlo a la consideración de ustedes.

A partir de los lineamientos teóricos sobre el estudio de la probabilidad planteados por el Profesor Bruno de Finetti, que insistía en la necesidad primordial de resolver los problemas conceptuales "*...a fin de que el desarrollo de los razonamientos no se reduzca a un mero juego formal sobre expresiones matemáticas o a enunciados pseudo-filosóficos o pretendidamente prácticos, vacíos y simplistas*", nuestra modesta intención fue realizar un tratamiento unitario sobre la teoría de la probabilidad y de los procesos estocásticos, más desde el punto de vista conceptual que matemático, procurando lograr un balance adecuado entre un análisis riguroso y un detalle innecesario, aplicando este modo probabilístico de pensamiento a la explicación y predicción del comportamiento de fenómenos dinámicos de la economía y las finanzas.

Ocurre que, en general, estos fenómenos son no-integrables en el sentido de Poincaré y que, por lo tanto, poseen en forma intrínsecamente objetiva las propiedades de irreversibilidad en el tiempo y de aleatoriedad, por lo que sus estados forman una sucesión de variables aleatorias orientadas en el tiempo y, en consecuencia, la noción de trayectoria asociada a su comportamiento dinámico pierde sentido y su representación se reduce entonces al cálculo de modelos estocásticos definidos por distribuciones de probabilidades condicionadas.

Esta circunstancia nos condujo a una nueva parábola acerca de la aleatoriedad -en la que el protagonista es el tiempo- que trata sobre las dificultades que trae aparejada la estimación de los modelos dinámicos y su selección y evaluación de acuerdo con ciertos criterios de optimización y cuyo resultado final es este compendio sobre la teoría de los procesos estocásticos, dividido en tres partes dedicadas, respectivamente, a las series cronológicas lineales unidimensionales, lineales multidimensionales y no-lineales.

Los temas tratados en este primer volumen incluyen una introducción a la teoría general de los procesos estocásticos, un detallado tratamiento de los modelos para procesos de parámetro discreto, estacionarios y no.-estacionarios (incluyendo aquéllos correspondientes a procesos con orden de integración fraccionario) y un estudio sobre los aspectos particulares de la modelización lineal de los fenómenos del mercado de capitales.

Agotadas las consideraciones preliminares, lo único que resta es darles la bienvenida a este maravilloso universo Borgiano en el que "*...el presente es indefinido, el futuro no tiene realidad sino como esperanza presente, el pasado no tiene realidad sino como recuerdo presente*".

AHL - MLG

Agradecimientos

Este trabajo fue elaborado en el ámbito del Centro de Econometría, dependiente de la Sección de Investigaciones en Matemática, Estadística y Econometría del Instituto de Investigaciones en Administración, Contabilidad y Matemática de la Facultad de Ciencias Económicas (UBA), por lo que queremos agradecer, en primer lugar, a sus directoras, Profesora Juana Z. Brufman y Profesora María Teresa Casparri y a su subdirector, Profesor Heriberto L. Urbisaia, por los atinados consejos con que nos favorecieron durante los últimos 43 años y por su invalorable apoyo al Centro de Econometría.

Vaya también nuestro agradecimiento al Señor Julián Siri por la investigación bibliográfica y por su infinita paciencia puesta al servicio de la detección de errores en el texto y de la consecuente recomendación de las correcciones y a nuestros alumnos por los innumerables aportes originados en sus comentarios cotidianos en el aula.

Queremos, además, expresar nuestro reconocimiento a los profesores Fausto I. Toranzos, José Barral Souto y Elías A. De Césare, quienes ya no están entre nosotros, por sus inolvidables enseñanzas sin las cuales, probablemente, esta modesta contribución nunca hubiera visto la luz.

Finalmente, queremos agradecer al Maestro Jorge Luis Borges por habernos permitido el atrevimiento de tomar prestada su pluma, en forma inconsulta, para aliviar nuestra pobreza de lenguaje y ayudar a transmitir los principios fundamentales que justifican las técnicas desarrolladas en este trabajo.

AHL - MLG

Advertencia

Los autores tienen el agrado de comunicar a los Pepos, por este único medio que, a partir de mañana, ya podemos ir a ver a las focas.

AHL - MLG

INDICE

Parte I: Los procesos estocásticos lineales unidimensionales

Capítulo 1: Acerca de la teoría general de los procesos estocásticos

1.1.- Introducción: Acerca de fenómenos dinámicos y procesos estocásticos 3
1.2.- Definición de proceso estocástico 7
1.3.- Clasificación de los procesos estocásticos 9
1.4.- El concepto de estacionariedad
1.4.1.- En los procesos unidimensionales 16
1.4.2.- En los procesos k-dimensionales 20
1.5.- El concepto de ergodicidad 26
1.6.- Algunas consideraciones preliminares acerca del teorema de descomposición De Wold 34
1.7.- Acerca de la definición estricta de algunos procesos que son de aparición frecuente en las aplicaciones financieras, biométricas y actuariales
1.7.1.- Las cadenas de Markov
1.7.1.1.- Definición 37
1.7.1.2.- Clasificación de los estados 43
1.7.1.3.- Probabilidades de recurrencia 44
1.7.1.4.- La condición de estacionariedad y las probabilidades en el límite 48
1.7.1.5.- Probabilidades de permanencia 49
1.7.1.6.- El "random walk" (RW) 51
1.7.2.- El proceso de Poisson
1.7.2.1.- Homogéneo o con función intensidad de ocurrencia constante 53
1.7.2.2.- Compuesto 64
1.7.2.3.- Ponderado 66
1.7.2.4.- Generalizado (o no-homogéneo) 67
1.7.3.- El proceso de nacimiento-fallecimiento
1.7.3.1.- El proceso de nacimiento-fallecimiento con funciones intensidad dependientes del tamaño de la población 71
1.7.3.2.- El proceso puro de nacimiento con función intensidad dependiente del tamaño de la población 83
1.7.3.3.- El proceso de nacimiento-fallecimiento lineal 85
1.7.3.4.- El proceso de nacimiento-fallecimiento con funciones intensidad dependientes del tiempo y del tamaño de la población 94
1.7.4.- El proceso de nacimiento-migración-fallecimiento
1.7.4.1.- El proceso de nacimiento-inmigración-emigración-fallecimiento lineal unidimensional 105
1.7.4.2.- El proceso de nacimiento-migración fallecimiento lineal bidimensional 108
1.7.5.- El proceso de nacimiento-fallecimiento lineal bidimensional 112
1.7.6.- El proceso de enfermedad-fallecimiento 115

Capítulo 2: Los procesos estacionarios de parámetro discreto

2.1.- Definiciones 125
2.2.- La función de autocovarianzas 126
2.3.- La estimación del valor medio 127
2.4.- La estimación de la función de autocovarianzas 128
2.5.- La función de densidad espectral 131
2.6.- La estimación de la función de densidad espectral 132
2.7.- La función de autocorrelaciones 133
2.8.- La estimación de la función de autocorrelaciones 134
2.9.- Las expresiones MA y AR de orden infinito como dos formas equivalentes de representación lineal de un proceso estocástico
2.9.1.- La representación MA 138
2.9.2.- La representación AR 142
2.10.- Los procesos autorregresivos de orden p
2.10.1.- Las condiciones de estacionariedad e invertibilidad 144
2.10.2.- La función de autocorrelaciones 145
2.10.3.- La estimación de los coeficientes 146
2.10.4.- La función de autocorrelaciones parciales 149
2.10.5.- La selección del orden de autorregresividad 155
2.10.5.1.- Criterios eficientes y criterios consistentes 157
2.10.5.2.- Las medidas de distancia 159
2.10.5.3.- Criterios de selección eficientes basados en la medida De Kullback-Leibler 161
2.10.5.4.- Criterios de selección eficientes basados en la medida L2 172
2.10.5.5.- Criterios de selección consistentes 180
2.10.5.6.- El método de Quenouille 188
2.10.6.- La condición de linealidad 189
2.10.7.- El proceso AR(1) 191
2.10.8.- El proceso AR(2) 197
2.11.- Los procesos de promedios móviles de orden q
2.11.1.- Las condiciones de estacionariedad e invertibilidad 226
2.11.2.- Las funciones de autocorrelaciones y autocorrelaciones parciales 227
2.11.3.- La estimación de los coeficientes 228
2.11.4.- El proceso MA(1) 232
2.11.5.- El proceso MA(2) 233
2.12.- Los proceso autorregresivos de promedios móviles de orden (p,q)
2.12.1.- Las condiciones de estacionariedad e invertibilidad 238
2.12.2.- Las funciones de autocorrelaciones y autocorrelaciones parciales 241
2.12.3.- La estimación de los coeficientes 245
2.12.4.- La estimación de los órdcenes p y q 249
2.12.5.- El proceso ARMA(1,1) 251
2.13.- Las funciones de autocovarianzas y autocorrelaciones inversas 252
2.14.- Análisis de la adecuacidad del modelo
2.1.4.1.- Acerca de la existencia de autocorrelación en el proceso $\{\varepsilon_t\}$ 258
2.1.4.2.- Acerca de la no-Normalidad del proceso $\{\varepsilon_t\}$ 260
2.1.4.3.- Los tests para decidir acerca de la adecuacidad de una distribución 264
2.15.- La teoría de la predicción del Kolmogorov-Wiener para series cronológicas estacionarias 276

Capítulo 3: Los procesos no-estacionarios de parámetro discreto

3.1.- No-estacionariedad en la varianza 291
3.1.1.- Las transformaciones de potencia 291
3.1.2.- La estimación del coeficiente λ 295
3.1.3.- Algunas transformaciones alternativas a las de potencia 300
3.2.- No-estacionariedad en el valor medio
3.2.1.- El método de la diferenciación sucesiva 301
3.2.2.- La representación MA 308
3.2.3.- La representación AR 311
3.3.- Los tests de las raíces unitarias
3.3.1.- El caso de los procesos AR(1) 312
3.3.2.- El caso de los procesos AR(p) 328
3.3.3.- El caso de los procesos ARMA(p,q) 329
3.3.4.- Las raíces unitarias múltiples 332
3.4.- Modificaciones en la estructura temporal 344
3.5.- Las medidas de persistencia 347
3.6.- La teoría de la predicción en procesos no-estacionarios 349
3.7.- Los procesos afectados por factores periódicos
3.7.1.- Las condiciones de estacionariedad e invertibilidad 353
3.7.2.- La función de autocorrelaciones en las representaciones multiplicativas 355
3.7.3.- La función de autocorrelaciones en las representaciones no-multiplicativas 366
3.7.4.- La estimación de la función de autocorrelaciones 368
3.7.5.- La estimación de los coeficientes 368

4.- Capítulo 4: Acerca de la econometría del mercado financiero

4.1.- Los procesos "random walk" 375
4.2.- Los tests para contrastar las hipótesis RW 385
4.3.- La martingala 409
4.4.- La mixingala 417
4.5.- El movimiento Browniano 420
4.6.- El proceso de Itô 431
4.7.- Procesos estocásticos y mercado de capitales
4.7.1.- Las hipótesis clásicas sobre el comportamiento del mercado de capitales 439
4.7.2.- Los modelos clásicos representativos de los precios de un activo 443
4.7.3.- El comportamiento de la rentabilidad como un proceso fractal 448
4.7.4.- El proceso de Uhlenbeck-Ornstein 449
4.7.5.- Acerca de la estructura temporal de la tasa de interés 451
4.7.6.- La estimación de la volatilidad
4.7.6.1.- La volatilidad histórica 454
4.7.6.2.- La volatilidad implícita 459
4,7,7,. Las distribuciones representativas del comportamiento de las rentabilidades
4.7.7.1.- Las distribuciones Normales 462
4.7.7.2.- Las distribuciones no-Normales 465

4.7.7.3.- Los métodos de predicción en procesos autorregresivos con varianza infinita 485
4.7.8.- Las distribuciones Normales compuestas
4.7.8.1.- El modelo de Clark 493
4.7.8.2.- El modelo de Praetz-Blattberg-Gonedes 496
4.7.9.- Los modelos para el estudio de las rentabilidades que utilizan procesos con "salto"
4.7.9.1.- El proceso de difusión con "salto" 497
4.7.9.2.- El modelo de Kou 499
4.7.9.3.- El modelo de Press 501
4.7.9.4.- El modelo de Oldfield-Rogalski-Jarrow 502
4.7.9.5.- El modelo de Merton 503
4.7.10.- Los modelos para el estudio de las volatilidades 505

5.- Capítulo 5: Una introducción a los procesos de parámetro discreto con orden de integración fraccionario

5.1.- Acerca de la memoria de los procesos 513
5.2.- Los rpocesos WN con orden de diferenciación fraccionario
5.2.1.- Definición 514
5.2.2.- las funciones de autocorrelaciones y autocorrelaciones parciales 517
5.2.3.- La función de densidad espectral 520
5.3.- Los procesos ARFIMA(p,d,q)
5.3.1.- Definición 520
5.3.2.- La función de autocovarianzas 521
5.3.3.- La función de densidad espectral 522
5.3.4.- La estimación de los coeficientes 523
5.4.- El coeficiente de Hurst 528
5.5.- Los tests para contrastar la hipótesis ARFIMA 529
5.6.- Los procesos GARMA 545
5.7.- Los procesos ARFIMA y el movimiento Browniano 548

Bibliografía 553

Indice temático 583

Capítulo 1

Acerca de la teoría general de los procesos estocásticos

"La causa verdadera
es la sospecha general y borrosa
del enigma del Tiempo;
es el asombro ante el milagro
de que a despecho de infinitos azares,
de que a despecho de que somos
las gotas del río de Heráclito,
perdure algo en nosotros:
inmóvil".

(Borges, J.L.: "Fervor de Buenos Aires")

1.1.- Introducción: Acerca de fenómenos dinámicos y procesos estocásticos

Sea un fenómeno dinámico $Y(t,\omega)$ que evoluciona en el **dominio del tiempo** $(t \in T)$ y cuya configuración varía en un dominio conocido como **dominio de los estados** (o **dominio de las fases** o **de las variables**) $\left(\omega \in \Omega(Y)\right)$.

Cada estado $\omega(t)\,(t \in T)$ del fenómeno queda definido por la **realización** simultánea, en el momento t, de los infinitos factores (**variables de estado**) que forman su **estructura causal**. La sucesión de los estados $\left\{\omega(t), t \in T\right\}$ define la **trayectoria** del fenómeno.

Para fenómenos que se encuentran en estado de equilibrio dicha trayectoria admite una representación en términos de dinámica clásica. El fundamento de esta representación mecánica del comportamiento de los fenómenos de la naturaleza se basa en una concepción determinística (al menos a nivel macroscópico)[1] y reversible en el dominio del tiempo, en la que la diferencia entre el pasado, el presente y el futuro no posee ningún significado; en la que, conocido su presente, se puede reproducir su pasado y

[1]. En realidad debería decirse representación mecánica o cuántica, dado que la mecánica cuántica orotodoxa señaló el final del determinismo microscópico pero a nivel macroscópico se mantuvo en el paradigma de la reversibilidad según el cual, una vez definida la función de onda, se puede predecir todo el futuro o reproducir todo el pasado del fenómeno.

calcular su futuro; es decir, en la que el presente no es sino un punto que separa el pasado del futuro[2].

Los fenómenos en estado de equilibrio se caracterizan por estar definidos por sistemas en los que las interacciones entre los factores que forman su estructura causal no existen o pueden ser eliminadas (a estos sistemas se los conoce también como **integrables en el sentido de Poincaré**)[3]. Ahora bien, Poincaré demostró que generalmente los fenómenos son no-integrables, en otros términos, que el estado natural de los fenómenos fácticos es de no-equilibrio (en el que las interacciones entre los factores que componen la estructura causal no pueden ser eliminadas)[4], un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, conocida como **estabilidad asintótica**, genera nuevos estados coherentes y nuevas estructuras complejas[5] que sólo son imaginables en el ámbito de la irreversibilidad en el tiempo.

Las características más importantes de estas estructuras, que son asintóticamente independientes de las condiciones iniciales, se resumen en los **atractores**, definidos como conjuntos de puntos en el espacio de los estados distribuidos de forma más o menos densa, al cual tiende la trayectoria del fenómeno cuando $t \to \infty$ [6]. Estos conjuntos: **i)** pueden asimilarse a un punto o a un ciclo límite definido por una curva cerrada, una superficie o un volumen y constituyen los llamados **atractores simples**[7]; **ii)** pueden observar una disposición irregular y poseer dimensión fraccionaria o fractal o **iii)** pueden observar una distribución densa, pero no-compacta y constituyen los atractores denominados **extraños** o **caóticos**[8]. La presencia de un atractor en un sistema dinámico implica la existencia de un flujo de correlaciones seriales significativas, viceversa, la no-existencia de correlaciones seriales significativas indica la ausencia de atractores e implica que el fenómeno puede ser considerado aleatorio puro[9].

De las consideraciones anteriores se puede concluir que todo razonamiento acerca de la

---

[2]. Un punto cuya existencia no tiene demasiada importancia ya que, de acuerdo con este interpretación, ni el ser ni el devenir poseen elementos temporales.

[3]. Es decir, fenómenos en los que las variables de su estructura causal representan factores independientes entre sí (o cuyas interacciones pueden ser incorporadas a su definición). En términos matemáticos estos sistemas integrables se caracterizan porque sus variaciones están definidas por un grupo, lo cual implica suponer que una transformación de un punto t_1 a un punto t_2 es igual a la suma de las transformaciones en los intervalos $[t_1,t)$ y $[t,t_2)$ (para $t_1,t,t_2 \in T$).

[4]. En términos matemáticos estos fenómenos se caracterizan porque sus variaciones están definidas por un semi-grupo.

[5]. Ver Pirgogine, I. (1962)(1980) (en Bocchi, G.; Ceruti, M. (eds.)(1985))(1994)(2001)(2003), Prigogine, I.; Nicolis, G. (1977)(1989), Prigogine, I.; Stengers, I. (1979)(1988)(1997), Kondepudi, D.; Prigogine, I. (1998).

[6]. La dinámica completa de un fenómeno no-integrable está definida por la característica de su conjunto atractor. Debe tenerse en cuenta que, si bien a partir de distintos estados iniciales el proceso puede moverse a atractores diferentes, a cada estado inicial le corresponde un atractor.

[7]. Como se verá con más detalle en la Parte 2, estos atractores pueden ser estables o inestables, según que, cuando se somete a un punto del atractor a una perturbación que lo coloca en forma infinitesimal fuera del atractor, la trayectoria perturbada pueda retornar al atractor o ser expulsada del mismo.

[8].Definidos por una combinación de órbitas periódicas inestables y órbitas aperiódicas. Como se verá en la Parte 3, es posible demostrar que existe una vinculación directa entre la existencia de atractores caóticos, los resultados obtenidos por Chaitin, G. (1987)(1988) (1990) sobre complejidad algorítmica y el teorema de incompletitud de Gödel (ver Feferman, S. (ed.) (1986)).

[9]. A estos atractores se deben agregar los quasiperiódicos, los cuales constituyen un nexo entre el ciclo límite y el comportamiento caótico y están definidos básicamente por una combinación de ciclos límite que "casi" se intersectan. Obsérvese que esta teoría de los atractores permite comprender el comportamiento de un gran número de fenómenos que no pueden ser explicados a partir de una interpretación clásica.

representación de los fenómenos dinámicos está íntimamente relacionado con la discusión sobre la irreversibilidad del tiempo, cuyo origen se puede situar en la famosa controversia entre Parménides y Heráclito. La interpretación de Parménides (que coincide con la representación clásica) consistía en considerar que no hay nada nuevo, que todo estaba, está y estará por siempre[10], en tanto que la de Heráclito, como se sabe, se basaba en el cambio continuo[11]. Es recién a principios del siglo XIX (el siglo del evolucionismo), con las obras de Charles Darwin en biología, Georg Friederich Wilhelm Hegel en filosofía y Jean-Joseph Fourier y Nicolas Leonard Sadi Carnot en física y, posteriormente con la introducción de las leyes de la termodinámica y del concepto de entropía por Rudolph Clausius, que a la concepción clásica se agrega la idea de tiempo irreversible[12], de tiempo con dimensión existencial, que permite independizar el dilema de Parménides-Heráclito de su contexto metafísico y expresarlo en términos formales[13].

La teoría clásica de la irreversibilidad se basa en la teoría de Boltzmann, L. (1866) según la cual el concepto de irreversibilidad-entropía (que incorpora la flecha del tiempo a la naturaleza intrínseca de los fenómenos y, en consecuencia, establece una diferencia fundamental entre el pasado que existe y el futuro que no existe)[14] está ligado al concepto de aleatoriedad y, en consecuencia, de probabilidad.

Debe tenerse en cuenta que, en situaciones alejadas de los estados de equilibrio, la producción de entropía conduce al sistema a comportamientos complejos (no reducibles a trayectorias individuales)[15] tales que, para condiciones de contorno dadas, admiten más de una solución posible y, en consecuencia, a la aleatoriedad. Lo que demuestra que la irreversibilidad y la aleatoriedad no son un producto de la ignorancia

---

[10]. La obra de Parménides de Elea data del 475 a.C. (circa). Su pensamiento filosófico está contenido en un extenso trabajo escrito en verso (de dudosa calidad, según la opinión de sus exégetas), dividido en dos partes: "*El camino de la verdad*" y "*El camino de la opinión*". Su lógica, aunque parezca primitiva. constituye el punto de partida de la dialéctica Platónica, la lógica Aristotélica y todo el razonamiento filosófico de la tradición occidental. Su proposición fundamental es "*Lo que es "es", y es imposible que no sea. Este es el camino de la verdad*". Como su interpretación de *es* es "existe", esta proposición admite una lectura del tipo: el UNO primordial permanece por siempre inalterado e inalterable. Parménides establece, mediante un proceso regular de demostración lógica, que su UNO-realidad "*no nace ni muere, no se mueve ni cambia, que es homogéneo e indivisible, que es continuo e igual en todas sus partes y que tiene la forma de una esfera que ocupa la totalidad del espacio*" (Amstrong, A.H. (1957)). Esta interpretación de la naturaleza a través de la unidad, el reposo y el límite (de origen Pitagórico) lo condujo, como se menciona en el texto, no sólo a una postura diametralmente opuesta a la de Heráclito, sino también a una negación de la posibilidad lógica de la existencia del mundo tal como puede ser percibido por los sentidos.

[11]. La obra de Heráclito de Éfeso data del 500 a.C. (circa). Su pensamiento está contenido en 130 breves fragmentos que quedan de su trabajo en prosa "*De la naturaleza*". Su propuesta fundamental, "*No es posible ingresar dos veces al mismo río*", indica su concepción de que todas las cosas se hallan en un cambio perpetuo, que en ninguna parte será posible hallar "*reposo eterno*" o "*estabilidad inmutable*" (Amstrong, A.H. (1957)), sino conflicto perpetuo ("*La guerra es el origen de todas las cosas*"). Debe tenerse en cuenta, no obstante, que su interpretación de una naturaleza de cambio y conflicto no debe ser asimilable a un supuesto comportamiento caótico, sino que su tesis implicaba la consideración de un mundo regido por un principio inmanente de orden y medida.

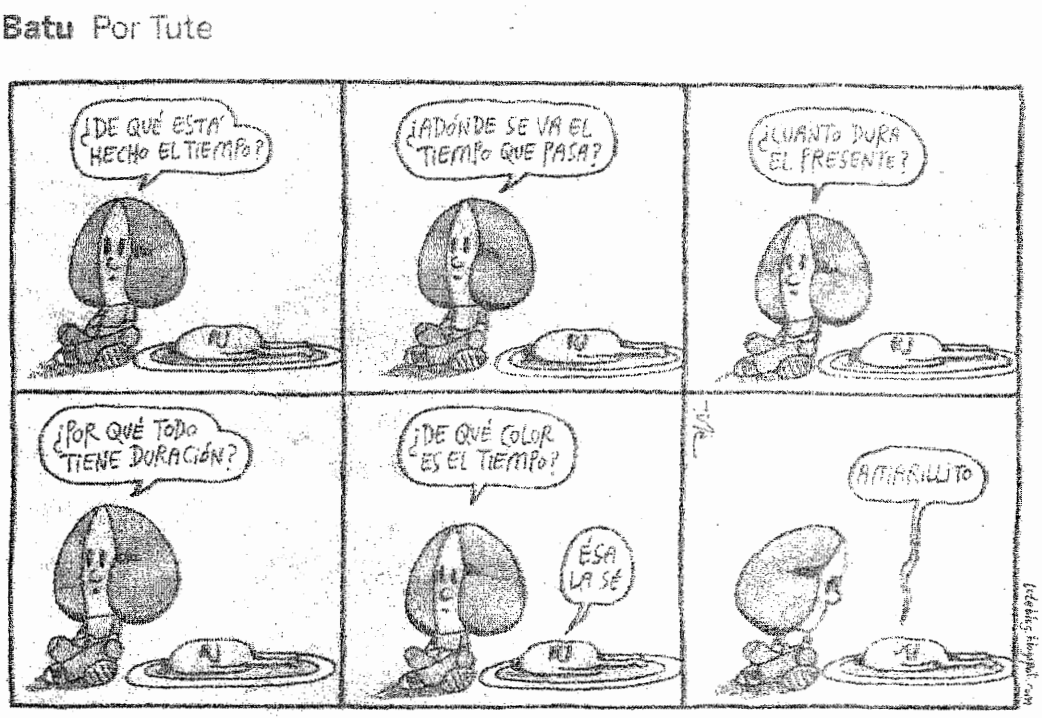
[12]. Asimilando el tiempo interno del fenómeno al tiempo de su degradación (ver Schneider, I. (1974/75), Brusch, S.G. (1976), Landro, A.H. (2002)).

[13]. Se incorpora la idea de tiempo interpretado como el tránsito de la posibilidad a la realidad, es decir, la idea del devenir.

[14]. El planteo de L. Boltzmann estaba dirigido a formular una aproximación evolutiva de la física, tratando de emular la obra de Ch. Darwin en biología, pero fue muy resistido ya que sus contemporáneos lo entendieron como un intento de negación de las leyes de la física Newtoniana, consideradas como el ideal del conocimiento objetivo.

[15]. A partir de las publicaciones de Gibbs, J.W. (1902) y Einstein, A. (1903)(1905)(1907) se plantearon dos formas de interpretar la dinámica: **i)** la individual, en términos de trayectorias de la dinámica clásica o de funciones de onda en la teoría cuántica y **ii)** la estadística, en términos de colectivos representados por una distribución de probabilidades (una función de densidad en el ámbito de la mecánica clásica o un operador densidad en la teoría cuántica) que posee la propiedad de satisfacer la ecuación de Liouville, la cual constituye el punto de partida de la física estadística (ver Krüger, L.; Gigerenzer, G.; Morgan, M. (1987)).

del observador, sino propiedades objetivas intrínsecas a los fenómenos dinámicos no-integrables[16] (que la aleatoriedad que genera la irreversibilidad no es epistemológica sino ontológica)[17]. En esta circunstancia la noción de trayectoria pierde sentido, los estados a asumir por un fenómeno forman una sucesión de variables aleatorias orientadas en el tiempo o, lo que es lo mismo, un **proceso estocástico**[18] y su representación se limita solamente a calcular las probabilidades de que un fenómeno, que se encuentra en un estado cualquiera de su dominio de estados en un momento t_0, pase a otro estado cualquiera de dicho dominio en un momento $t > t_0\ (t_0, t \in T)$ [19].



---

16. Contrariamente al concepto clásico en el que la dinámica y la aleatoriedad pertenecían a dominios diferentes y que, en consecuencia, las fluctuaciones con respecto a las trayectorias determinísticas debían ser resueltas como correcciones a las ecuaciones de la dinámica clásica y de la teoría cuántica, la noción de irreversibilidad demuestra que la aleatoriedad inherente a esta representación es una consecuencia de esta interpretación más general de la dinámica.

17. No se basa en la hipótesis Tomista "el observador nunca puede saber", sino en el principio "ni el observador ni el fenómeno nunca pueden saber" (ver Atlan, H. (1979), Gardner, M. (1979), Stengers, I. (en Bocchi, G.; Ceruti, M. (1985)), Landro, A.H. (2002)).

18. Todas las consideraciones realizadas precedentemente parecen sugerir que la teoría de los procesos estocásticos constituye el aporte más importante realizado por la física a la interpretación de la naturaleza de la aleatoriedad y, en consecuencia, a la teoría de la probabilidad.

19. Obsérvese que esta proposición contiene los siguientes axiomas sobre el tiempo: **i)** T denota un intervalo en el eje real; **ii)** todo $t \in T$ representa un instante; **iii)** la relación $t_0 < t$ define un ordenamiento del conjunto T; **iv)** $Y(t_0)$ representa un estado del fenómeno anterior a $Y(t)$ (ver Bunge, M. (1978)). De acuerdo con Ouspensky, P.D. (1982), "*Por tiempo entendemos la distancia que separa a dos estados en el orden de sucesión que los vincula*".

1.2.- Definición de proceso estocástico

De las consideraciones analizadas en la sección precedente se puede concluir que un proceso estocástico $\{Y(t,\omega), t \in T, \omega \in \Omega(Y)\}$ puede ser interpretado como una **función aleatoria** de dos variables: **i)** para un valor fijo de t es una variable aleatoria, $Y(\omega) = Y(t,\omega)\ (\omega \in \Omega(Y))$ y **ii)** para un valor dado, ω, del dominio de los estados, es una función del tiempo, $Y(t) = Y(t,\omega)\ (t \in T)$.

Sea $\{Y(t_1), Y(t_2), ..., Y(t_n)\}$ un conjunto finito de las variables aleatorias del proceso $\{Y(t)\}$ y sea $F_{Y(t_1),...,Y(t_n)}[y(t_1), y(t_2), ..., y(t_n)]$ la correspondiente función de distribución de probabilidades conjunta[20] (donde el vector $\{y(t_1), y(t_2), ..., y(t_n)\}$ denota la sucesión de realizaciones de las correspondientes variables aleatorias).

Como se sabe, para que una función $F_{Y(t_1),Y(t_2),...,Y(t_n)}[y(t_1), y(t_2), ..., y(t_n)]$ pueda ser considerada una función de distribución de probabilidades de una variable aleatoria n -dimensional $\{Y(t_1), Y(t_2), ..., Y(t_n)\}$ debe cumplir las **condiciones de consistencia de Kolmogorov**[21]: **i) de simetría**: debe ser invariante con respecto a cualquier permutación de las variables $Y(t_j)\ (j = 1,2,...,n)$, simbólicamente:

$$F_{Y(t_1),...,Y(t_i),...,Y(t_j),...,Y(t_n)}[y(t_1),...,y(t_i),...,y(t_j),...,y(t_n)] =$$

$$= F_{Y(t_1),...,Y(t_j),...,Y(t_i),...,Y(t_n)}[y(t_1),...,y(t_j),...,y(t_i),...,y(t_n)]$$

ii) de compatibilidad: se debe verificar que[22]:

$$\lim_{\substack{y(t_{k+1})\to\infty \\ ... \\ y(t_n)\to\infty}} F_{Y(t_1),Y(t_2),...,Y(t_n)}[y(t_1), y(t_2), ..., y(t_n)] = F_{Y(t_1),Y(t_2),...,Y(t_k)}[y(t_1), y(t_2), ..., y(t_k)] \qquad (k < n)$$

A partir de estas condiciones se puede concluir que una función de distribución de orden n genera $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ funciones de distribución de orden $k\ (= 1,2,...,n-1)$. El conjunto de las $\sum_{k=1}^{n}\binom{n}{k} = 2^n - 1$ funciones de distribución generadas define el **sistema de funciones de distribución** de la función de

---

[20].En la nomenclatura de algunos autores a estas funciones se las conoce como las **fidis ("finite-dimensional distributions")** del proceso (ver Mikosch, Th. (2006)).

[21]. Ver Kolmogorov, A.N. (1956).

[22]. Ver Billingsley, P. (1986). Obsérvese que este teorema permite a partir de la función de distribución conjunta $F_{Y(t_1),Y(t_2)}(y(t_1), y(t_2))$, derivar la distribución marginal $\lim_{y(t_2)\to\infty} F_{Y(t_1),Y(t_2)}[y(t_1), y(t_2)] = F_{Y(t_1)}[y(t_1)]$ y, en consecuencia, la distribución de probabilidades condicionada $F(Y(t_2) / y(t_1)) = \int_{-\infty}^{y(t_2)} \frac{f_{Y(t_1)}(y(t_1), u)}{f_{Y(t_1)}(y(t_1))} du$.

distribución de orden n. Ahora bien, de acuerdo con el **teorema de existencia de Kolmogorov**[23], todo sistema de funciones de distribución que cumplan las condiciones de consistencia define, en forma estricta, una función de probabilidades n-dimensional (propiedad que puede ser generalizada para $n \to \infty$). En consecuencia, se puede asegurar que un proceso estocástico queda **definido en forma estricta** o **en el sentido de Slutsky** si es posible definir la función de distribución de probabilidades conjunta para cualquier subconjunto finito de sus variables aleatorias[24]:

$$F_{Y(t_1),Y(t_2),\ldots,Y(t_n)}\left[y(t_1),y(t_2),\ldots,y(t_n)\right] =$$

$$= p\left\{\left[Y(t_1) \le y(t_1)\right] \cap \left[Y(t_2) \le y(t_2)\right] \cap \ldots \cap \left[Y(t_n) \le y(t_n)\right]\right\} =$$

$$= \prod_{j=1}^{n} p\left(\left[Y(t_j) \le y(t_j)\right] / \left\{\left[Y(t_{j-1}) = y(t_{j-1})\right] \cap \left[Y(t_{j-2}) = y(t_{j-2})\right] \cap \ldots \cap \left[Y(t_1) = y(t_1)\right]\right\}\right)$$

(para $t_1, t_2, \ldots, t_n \in T; n = 1,2,\ldots$). Donde, como se mencionó en la página precedente, el vector $\{y(t), t \in T\}$ denota la sucesión de realizaciones de las correspondientes variables aleatorias del proceso que constituye el **flujo de información natural**[25] que permite definir la familia de probabilidades condicionadas[26]:

$$p\left(\left[Y(t_j) \le y(t_j)\right] / \left\{\left[Y(t_{j-1}) = y(t_{j-1})\right] \cap \left[Y(t_{j-2}) = y(t_{j-2})\right] \cap \ldots \cap \left[Y(t_1) = y(t_1)\right]\right\}\right)$$

$\left(t_j > t_{j-1} > t_{j-2} > \ldots > t_1\right)$ que define la **función de probabilidades de transición**[27] del proceso del estado $y(t_{j-1})$ al estado $y(t_j)$ (obsérvese que, en este sentido, el proceso $\{Y(t)\}$ puede ser interpretado como una población hipotética formada por sus infinitos flujos de información natural posibles)[28].

---

[23]. Kolmogorov, A.N. (1956). Conocido también como **teorema de Kolmogorov-Daniels** (ver también Neveu, J. (1965), Loéve, M. (1977), Williams, D. (1979), Billingsley, P. (1986)).

[24]. La **distribución del proceso** queda, entonces, definida por la colección de las fidis. Debe tenerse en cuenta que, como se vio más arriba, la inversa de esta implicación sólo se verifica si las funciones de distribución de probabilidades cumplen las condiciones de consistencia de Kolmogorov.

[25]. Como se verá en la próxima sección, a este flujo de información natural se lo conoce también como la **filtración natural** del proceso.

[26]. El planteo probabilístico riguroso de los procesos estocásticos surgió recién a comienzos de la década de 1920, asociado a la noción de independencia de variables aleatorias, fundamentalmente en el ámbito de la física, con la notable excepción del modelo de Bachelier, L. (1900) para explicar el comportamiento de los precios de los activos en el mercado de valores de París (ver Sec. 4.4). Debe tenerse en cuenta que, si bien la noción de independencia entre eventos, propuesta por G. Cardano y formalizada por A. De Moivre en 1739, surgió en la segunda mitad del siglo XVI , hasta fines del siglo XIX el concepto de independencia fue interpretado simplemente como ausencia de dependencia (ver Landro, A.H. (2002)).

[27]. Esta denominación aparece por primera vez en el trabajo de Einstein, A. (1916-1917) relacionado con las transiciones de los estados de energía experimentados por los átomos (W_{mn} en la nomenclatura de Einstein) y reaparece recién en el trabajo de Heisenberg, G.W. (1925) sobre mecánica cuántica. La definición de probabilidades de transición propuesta por Einstein y Heisenberg constituyó la base del concepto de probabilidad en mecánica cuántica desarrollado por Born, M. (1926)(1926).

[28]. Como alternativa a esta definición de un proceso estocástico a partir de la función de distribución conjunta para un conjunto finito de sus variables aleatorias, conocida como **distribucionista**, existe una **interpretación constructivista**, basada en la definición del proceso como función de otros procesos más simples (a menudo formados por variables aleatorias independientes e idénticamente distribuidas (iid)). Como se verá en la Sec. 1.7 y en el Cap. 4, muchos procesos que se presentan frecuentemente en las aplicaciones

Como se mencionó en la página precedente, cada observación, $y(t)$ denota una realización de una variable aleatoria $Y(t)$ y la sucesión de estas realizaciones, de acuerdo a lo expresado en la Sec. 1.1, constituye una trayectoria del proceso $\{Y(t)\}$. Luego $\{Y(t)\}$ puede ser interpretado como el conjunto de sus infinitas trayectorias posibles. Generalizando esta idea, en lugar de considerar al proceso $\{Y(t)\}$ como un fenómeno dinámico con una trayectoria estocástica, es posible considerarlo -haciendo abstracción del dominio del tiempo- como una variable aleatoria cuyo dominio sea el espacio $\Omega(Y(t))$ que define la trayectoria en su totalidad. A partir de este razonamiento, la variable $Y(t)$ puede ser interpretada como un punto en el espacio $\Omega(Y(t))$.

Alternativamente, en casos más complejos (en los que no se cuenta con la información suficiente para definir la estructura de probabilidades de transición), la definición del proceso se obtiene mediante la definición de su sistema de momentos, en particular, los momentos de primero y segundo orden, que incluyen los valores esperados ($E[Y(t)]$), las varianzas ($\sigma^2[Y(t)]$) y las autocovarianzas:

$$\gamma(Y(t_i),Y(t_j)) = E\left(\{Y(t_i) - E[Y(t_i)]\}\{Y(t_j) - E[Y(t_j)]\}\right) \qquad (t_i,t_j \in T; i \neq j)$$

que miden el grado de influencia que $Y(t_i)$ ejerce sobre el comportamiento de $Y(t_j)$.

Debe tenerse en cuenta que, si la distribución conjunta de las variables $Y(t)$ es Normal, los momentos de primero y segundo orden caracterizan en forma completa las propiedades del proceso estocástico. Si no se puede suponer la condición de Normalidad, pero puede considerarse la hipótesis de linealidad -en el sentido que cada valor esperado del proceso puede ser generado a partir de la combinación lineal de sus realizaciones anteriores y de las realizaciones pasadas y contemporáneas de otros procesos relacionados con él-, entonces, los momentos de primero y segundo orden caracterizan las propiedades más importantes del proceso. En términos más generales se demuestra que, si la sucesión de probabilidades de transición es de cuadrado integrable, entonces las características estructurales del proceso están resumidas en los momentos de primero y segundo orden.

1.3.- Clasificación de los procesos estocásticos

Los procesos estocásticos pueden ser clasificados de acuerdo a:

1) La naturaleza del dominio de las variables: Si $\Omega(Y(t))$ está definido por un conjunto numerable de puntos (finito o infinito) se dice que el proceso es **discreto en el dominio de las variables**[29], en caso contrario, se dice que el proceso es **continuo en el dominio de las variables.**

(como, por ejemplo, los procesos de Markov, los de incrementos independientes, los "random walk" y las martingalas) pueden ser definidos a partir de sumas de variables aleatorias independientes.

[29]. En particular, si $\Omega(Y(t))$ está definido por un conjunto finito, se dice que el proceso $\{Y(t)\}$ está formado por variables **simples**.

Por otra parte, si $\Omega\left(Y(t)\right) \in R_1$, se dice que el proceso es **unidimensional**. Si $\Omega\left(Y(t)\right)$ está definido por el producto cartesiano $\Omega\left(Y(t)\right) = \Omega\left(Y_1(t)\right) \times \Omega\left(Y_2(t)\right) \times ... \times \Omega\left(Y_k(t)\right)$ (donde $\Omega\left(Y_j(t)\right) \in R_1 \quad (j = 1,2,...,k)$) entonces queda definido un **proceso k-dimensional en el dominio de las variables** (también conocido como **proceso-vector de orden k**):

$$\left\{Y(t,\omega)\right\} = \left\{Y_1(t,\omega), Y_2(t,\omega), ..., Y_k(t,\omega); t \in T; \omega \in \Omega\left(Y_j(t)\right); j = 1,2,...,k\right\}$$

formado por la combinación de los **procesos marginales**[30] $\left\{Y_j(t)\right\}$ $\left(j = 1,2,...,k\right)$, con función de distribución de probabilidades conjunta de la forma:

$$F_{Y(t_1),Y(t_2),...,Y(t_k)}\left[y(t_1), y(t_2), ..., y(t_k)\right] = p\left\{\left[Y(t_1) \le y(t_1)\right] \cap \left[Y(t_2) \le y(t_2)\right] \cap ... \cap \left[Y(t_k) \le y(t_k)\right]\right\}$$

Este proceso vector será discreto o continuo, según que los correspondientes dominios, $\Omega\left(Y_j(t)\right)$ $\left(j = 1,2,...,k\right)$, sean discretos o continuos.

Una generalización inmediata de la definición anterior permite obtener un proceso definido por un vector de variables aleatorias de infinitas dimensiones $Y(t) = \left[Y_1(t) \quad Y_2(t) \quad ...\right]$ con un espacio de estados de dimensión infinita $\Omega = \Omega\left(Y_1(t)\right) \times \Omega\left(Y_2(t)\right) \times \Omega\left(Y_3(t)\right) \times ...$.

2) La naturaleza del dominio del tiempo: De la misma forma que en el punto precedente, el proceso será continuo o discreto en el dominio del tiempo según que T sea continuo, ($T = (-\infty, \infty)$), o discreto, ($T = ..., -2, -1, 0, 1, 2, ...$). Que T sea discreto implica que las transiciones del proceso en su espacio de estados se generan a intervalos regulares y, viceversa, que T sea continuo implica que las transiciones se verifican en momentos aleatorios[31]. A los procesos discretos en el dominio del tiempo se los denomina también **de**

---

[30]. También denominados **procesos componentes**.

[31].- El concepto de continuidad en el dominio de las variables fue adoptado en forma natural por la teoría de la probabilidad a partir de la aplicación de la teoría de la medida desarrollada por Borel, E. (1898)(1905) y Lebesgue, H. (1904). En cuanto al concepto de continuidad en el dominio del tiempo, su incorporación a la teoría moderna de la probabilidad requirió un tiempo mucho mayor. Si bien las **"cantidades aleatorias continuas"** habían sido estudiadas desde el siglo XVII, los fenómenos de resultado eventual ordenados en una sucesión continua en el tiempo, introdujeron una estructura conceptual novedosa, la cual tuvo dos líneas de desarrollo: una en el ámbito de la física y otra en el de la matemática. La primera tiene su punto de partida en la formulación de la ecuación de Boltzmann, L. (1872) y alcanzó su punto culminante en los trabajos de Einstein, A. (1905)(1906) sobre la teoría del movimiento Browniano, los de von Smoluchowski, M. (1915)(1916)(1918) sobre la ecuación de difusión de Fokker-Planck y en la ecuación de Chapman, S. (1928). A excepción de los intentos aislados de Bachelier, L. (1900) y Lundberg, O. (1902) en aplicaciones económicas y a la teoría del seguro, las primera teorías matemáticas sobre los procesos estocásticos continuos en el dominio del tiempo surgieron repentinamente con los trabajos de de Finetti, B. (1929)(1929)(1929)(1930), Hostinsky, B. (1929)(1929)(1930), Kolmogorov, A.N. (1928)(1929)(1929)(1929) y con la memoria de Reichenbach, H. (1929), la cual representa un cambio profundo en el concepto de probabilidad. Debe tenerse en cuenta que estos trabajos no observan ninguna relación con el ámbito de la física. Fue el trabajo de Kolmogorov, A.N. (1931) el que, a partir de su análisis general de los procesos estocásticos mediante la utilización de la teoría de la medida a partir de la construcción del espacio producto, unificó los desarrollos físico y matemático . Su publicación de 1934 (conocida como el "Grundbegriffe") se transformó en el símbolo de la teoría moderna de la probabilidad. En su capítulo 5 figuran dos novedades matemáticas esenciales: **i)** la teoría de las probabilidades condicionadas para conjuntos infinitos de eventos elementales y su aplicación a casos (como por ejemplo, el movimiento Browniano) en los que la condición tiene probabilidad nula (las probabilidades condicionadas son definidas como variables aleatorias) y **ii)** la teoría general de los procesos estocásticos. Entre las continuaciones inmediatas del "Grundbegriffe" cabe mencionar los trabajos de Khinchin, A. (1934) sobre una aproximación probabilística a la teoría ergódica y la teoría de los procesos estacionarios, los de Hopf, E. (1934) sobre el estudio de sistemas

parámetro discreto o **series cronológicas**[32]

Asimismo, se dice que un proceso $\{Y(t)\}$ es **continuo en valor medio cuadrático** si se verifica la siguiente convergencia uniforme: $\lim_{h\to 0} E\left\{\left[Y(t) - Y(t-h)\right]^2\right\} = 0$. Como corolario de esta definición, se dice que un proceso $\{Y(t)\}$ es **diferenciable en valor medio cuadrático** si existe un proceso $\{Z(t)\}$ tal que $\lim_{h\to 0} E\left\{\left[\frac{Y(t+h) - Y(t)}{h} - Z(t)\right]^2\right\} = 0$. En este caso, se puede escribir $\frac{d}{dt}Y(t) = Z(t)$ (donde $\frac{d}{dt}$ denota, en este caso, el **operador diferencial en valor medio cuadrático**).

Igual que en el dominio de las variables, se dice que un proceso $\{\Omega(Y(t))\}$ es k-dimensional continuo en el dominio del tiempo si $T \in R_k$ [33].

De acuerdo a lo expresado en estos dos criterios de clasificación y haciendo abstracción de su dimensión, se pueden distinguir, entonces, cuatro clases de procesos[34]:

* **discreto** en el dominio del tiempo-**discreto** en el dominio de las variables,

* **discreto** en el dominio del tiempo-**continuo** en el dominio de las variables,

* **continuo** en el dominio del tiempo-**discreto** en el dominio de las variables,

* **continuo** en el dominio del tiempo-**continuo** en el dominio de las variables.

3) La forma de la relación de dependencia entre las variables $Y(t)$:

i) Como se mencionó en la sección precedente:

$$\Omega(Y(t)) = \{Y_1(t,\omega),...,Y_k(t,\omega); t \in T; \omega \in \Omega(Y_j(t)); j = 1,2,...,k\}$$

define el dominio de las trayectorias del proceso k-dimensional $\{Y(t)\} = \{Y_1(t), Y_2(t),..., Y_k(t)\}$ determinadas al momento t. El conjunto $\{\Omega(Y(t_m)), \Omega(Y(t_\ell)),...\}$ $(t \geq t_m > > t_\ell > ...)$ de estos dominios define el llamado **flujo**

---

dinámicos, los de Doob, J.L. (1936)(1937)(1938) sobre procesos continuos y discretos en el dominio del tiempo, los de Cramér, H. (1937) y los de Lévy, P.P. (1937).

[32]. Debe tenerse en cuenta que, como se verá en el Cap. 2, también se denomina serie cronológica a la sucesión finita y ordenada en el tiempo de las realizaciones de las variables aleatorias que integran el proceso.

[33]. Un ejemplo históricamente paradigmático de este caso es el proceso representativo de la velocidad de una partícula suspendida en un líquido con t representando su posición en un espacio Euclídeo tridimensional $\{Y(t), t \in R_3\}$.

[34]. Debe tenerse en cuenta que este análisis es sumamente esquemático y que, como se verá inmediatamente, ignora otras clasificaciones basadas en la estructura de probabilidades de los procesos, como por ejemplo, en procesos estacionarios y no-estacionarios, ergódicos y no-ergódicos (ver Srinivasan, S.K.; Mehata, K.M. (1988)).

de información (o **filtración**) de $\{Y(t)\}$ [35]. Se dice, entonces, que $\{Y(t)\}$ es un proceso **adaptado** con respecto al flujo de información $\Omega(Y(t))$, cuando su realización al momento t se produce sólo cuando se producen las realizaciones al momento t de sus k procesos marginales $\{Y_j(t); j = 1,2,...,k\}$ [36].

En particular, si $\Omega(Y(t)) = \{Y(t)\}$, como se mencionó en páginas anteriores, el conjunto $\{y(t_m), y(t_\ell),...\}$ de las realizaciones pasadas del proceso $\{Y(t)\}$ constituye su **flujo de información natural** (o **filtración natural**).

ii) Se dice que el proceso unidimensional $\{Y(t); t \in T\}$ es **estrictamente autorregresivo de orden m** ($m = 0,1,2,...$) (**estrictamente AR(m)**), si su estructura de probabilidades de transición asume la forma[37]:

$$p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left\{\left[Y(t_m) = y(t_m)\right] \cap \left[Y(t_\ell) = y(t_\ell)\right] \cap ...\right\}\right) =$$

$$= p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left\{\left[Y(t_m) = y(t_m)\right] \cap \left[Y(t_\ell) = y(t_\ell)\right] \cap ... \cap \left[Y(t_1) = y(t_1)\right]\right\}\right)$$

$(t_1 < t_2 < ... < t_\ell < t_m < t_n)$. Esto significa que el estado a asumir por el proceso en el momento t_n depende exclusivamente de los estados asumidos en los m momentos anteriores, $t_1, t_2, ..., t_m$ ($y(t_1), y(t_2), ..., y(t_m)$) o, mejor dicho, que la información contenida en el infinito pasado del proceso puede ser resumida en las realizaciones de sus m estados inmediatamente anteriores[38].

Se dice que un proceso $\{Y(t)\}$ es **débilmente autorregresivo de orden m** (**débilmente AR(m)**) si se verifica que[39]:

$$E\left[Y^s(t_n) / y(t_m), y(t_\ell), ...\right] = E\left[Y^s(t_n) / y(t_m), ..., y(t_1)\right] \qquad (s = 1,2,...)$$

Se dice que un proceso $\{Y(t)\}$ es **débilmente linealmente autorregresivo de orden m** si el valor esperado de $Y(t_n)$, condicionado por sus realizaciones anteriores, puede ser expresado de la siguiente forma:

---

[35]. A esta filtración, formada por la filtración natural autorregresiva de $\{Y(t)\}$ más la información que proporcionan las realizaciones contemporáneas y anteriores de los procesos que integran su estructura causal, se la denomina **filtración aumentada**.

[36]. Esta caracterización de la relación de dependencia entre las variables corresponde al análisis del proceso en el dominio del tiempo. El análisis alternativo en el dominio de las variables implica la caracterización de sus periodicidades. Debe tenerse en cuenta que existen variaciones originadas en la modificación de las realizaciones condicionantes y una heterogeneidad originada en las variaciones del conjunto de información.

[37]. En particular, como se verá en la Sec. 1.7.1.1, estas probabilidades se denominan **probabilidades de transición simple**.

[38]. El orden de autorregresividad define la llamada **"memoria"** del proceso.

[39]. Debe tenerse en cuenta que este valor esperado condicionado puede ser interpretado como una variable aleatoria definida a partir de un vector $\left[Y(t_m) \quad Y(t_\ell) \quad ... \quad Y(t_1)\right]$, que proporciona la información con que cuenta el observador acerca de la estructura de la variable $Y(t,\omega)$ como una función de $\omega \in \Omega(Y(t))$. Para un análisis más detallado de la esperanza matemática condicionada, ver Williams, S. (1991), Billingsley, P. (1995), Landro, A.H. (2002).

$$E\left[Y(t_n) / y(t_m), y(t_\ell), \ldots, y(t_1)\right] = a_0 + a_1 y(t_m) + \ldots + a_m y(t_1)$$

En particular, se dice que $Y(t)$ es **estrictamente autorregresivo de orden cero** o (**estrictamente AR(0)**) si se verifica que:

$$p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left\{\left[Y(t_m) = y(t_m)\right] \cap \left[Y(t_\ell) = y(t_\ell)\right] \cap \ldots\right\}\right) = p\left(y_a \le Y(t_n) \le y_b\right)$$

$\left(t_1 < t_2 < \ldots < t_\ell < t_m < t_n\right)$ o, lo que es lo mismo que:

$$F_{Y(t_1),Y(t_2),\ldots,Y(t_n)}\left[y(t_1), y(t_2), \ldots, y(t_n)\right] = \prod_{j=1}^{n} F_{Y(t_j)}\left[y(t_j)\right]$$

es decir, si las variables aleatorias $Y(t)$ que integran el proceso son estocásticamente independientes[40].

Si se verifica que:

$$\lim_{h \to \infty} p\left(\left[y_a \le Y(t_{n+h}) \le y_b\right] / \left\{\left[Y(t_n) = y(t_n)\right] \cap \left[Y(t_m) = y(t_m)\right] \cap \ldots\right\}\right) = p\left(y_a \le Y(t_{n+h}) \le y_b\right)$$

se dice que el proceso $\{Y(t)\}$ es **asintóticamente estrictamente AR(0).**

Como una extensión del concepto de proceso con variables idénticamente distribuidas (id) surge el concepto de **intercambiabilidad**. Se dice que un proceso $\{Y(t)\}$ es intercambiable si se verifica que la distribución de probabilidades conjunta para una sucesión finita de sus variables aleatorias ($\{Y(t_1), Y(t_2), \ldots, Y(t_n)\}$) es invariable respecto de cualquier combinación de las variables marginales. Es decir, si se verifica que[41]:

$$f_{Y(t_h)}\left(y(t_h)\right) = f_Y \qquad (h = 1,2,3,\ldots,n)$$

$$f_{Y(t_h),y(t_i)}\left(y(t_h), y(t_i)\right) = f_{Y(t_j),Y(t_k)}\left(y(t_j), y(t_k)\right) \qquad (h,i,j,k = 1,2,3,\ldots,n)$$

.....

$$f_{Y(t_h),Y(t_i),\ldots,Y(t_a)}\left(y(t_h), y(t_i), \ldots, y(t_a)\right) = f_{Y(t_j),Y(t_k),\ldots,Y(t_b)}\left(y(t_j), y(t_k), \ldots, y(t_b)\right) \qquad (h,i,\ldots,a,j,k,\ldots,b = 1,2,3,\ldots,n)$$

Obsérvese que , contrariamente a lo que ocurre con la condición de idénticamente distribuidas que considera exclusivamente a las distribuciones marginales, la intercambiabilidad -dado que no presupone la condición de independencia, sino solamente un cierto tipo de simetría en la distribución- se funda en el comportamiento de

---

[40]. Obsérvese que un $AR(0)$ estricto es un proceso que no depende de su filtración natural, lo cual no implica que el proceso no dependa de una filtración aumentada. Como se verá en el Cap. 6, puede suceder que, si a la filtración natural autorregresiva se le agrega la información proporcionada por las realizaciones contemporáneas y anteriores de un conjunto de procesos relacionados causalmente con $\{Y(t)\}$, se mejore su predicción.

[41]. Ver Chow, Y.S.; Teicher, H. (1978).

las distribuciones de probabilidades conjuntas. Se puede concluir, entonces, que todo proceso iid es intercambiable, pero que la implicación inversa no se verifica necesariamente.

Si se verifica que:

$$p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left\{\left[Y(t_m) = y(t_m)\right], \left[Y(t_\ell) = y(t_\ell)\right], \ldots\right\}\right) = p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left[Y(t_m) = y(t_m)\right]\right)$$

se dice que $\{Y(t)\}$ es un proceso **estrictamente autorregresivo de orden 1** (o **estrictamente AR(1)** o **proceso de Markov estricto**)[42]. En este caso particular se verifica que:

$$f\left[Y(t_n) / y(t_m)\right] = \int_{\Omega(Y(t))} f\left[Y(t_n) / y(t)\right] f\left[y(t) / y(t_m)\right] dy(t) \qquad (t_m \le t \le t_n)$$

expresión conocida como **condición de Kolmogorov-Chapman**[43].

Obviamente, si se verifica que:

$$E\left[Y^s(t_n) / y(t_m), y(t_\ell), \ldots\right] = E\left[Y^s(t_n) / y(t_m)\right] \qquad (s = 1,2,\ldots)$$

se dice que $\{Y(t)\}$ es **débilmente AR(1)** (o **proceso de Markov débil**).

De la misma forma, si se verifica que:

$$p\left(\left[y_a \le Y(t_n) \le y_b\right] / \left\{\left[Y(t_m) = y(t_m)\right] \cap \left[Y(t_\ell) = y(t_\ell)\right]\right\}\right) \qquad (t_n > t_m > t_\ell)$$

se dice que $\{Y(t)\}$ es un **proceso estrictamente AR(2)** o **proceso de Yule**.

Si las variables $Y(t)$, además de ser independientes, son idénticamente distribuidas (iid)[44] con esperanza matemática nula y varianza constante, $\sigma^2\left[Y(t)\right] = \sigma_Y^2 < \infty$, se dice que el proceso $\{Y(t)\}$ es un **ruido blanco estricto (WN-estricto)**[45]. Si en este proceso las variables $Y(t)$ se distribuyen de acuerdo con una función Normal, entonces se dice que $\{Y(t)\}$ es un **ruido blanco Normal (WN-Normal, WN-N(0, σ_Y^2))**.

Si se verifica que $E\left[Y(t_n) / y(t_m), y(t_\ell), \ldots\right] = E\left[Y(t_n)\right] = 0$, es decir, si las autocovarianzas entre las

---

[42]. En particular, todo proceso $\{Y(t)\}: AR(1)$ discreto en el dominio del tiempo y en el dominio de las variables, define una **cadena de Markov** (ver Sec. 1.6).

[43]. Esta expresión se debe a Kolmogorov, A.N. (1935). El aporte de Chapman, S. (1928) está relacionado con su introducción en el contexto de la teoría dinámica de los gases.

[44]. Si las variables $Y(t)$ son iid, se dice que el proceso $\{Y(t)\}$ es **completamente homogéneo**.

[45]. Se puede concluir en forma inmediata que todo proceso WN es $AR(0)$, pero que la implicación inversa no se verifica necesariamente.

variables aleatorias $Y(t)$ son nulas, $\gamma\left[Y(t_i),Y(t_j)\right] = 0\left(i \neq j, t_i, t_j \in T\right)$ y, además, sus varianzas son constantes, $\gamma\left[Y(t_i),Y(t_i)\right] = \sigma^2\left(Y(t_i)\right) = \sigma_Y^2 < \infty \left(t_i \in T\right)$, se dice que $\{Y(t)\}$ es un **ruido blanco débil**[46].

Asimismo, si se verifica que $\sigma^2\left[Y(t_j)\right] = \sigma_Y^2 \left(t_j \in T\right)$ y,

$$\gamma\left[Y(t_i),Y(t_j)\right]\begin{cases} \neq 0 & para\ 0 < \left|t_i - t_j\right| \leq q \qquad \left(i \neq j; t_i, t_j \in T\right) \\ = 0 & para\ \left|t_i - t_j\right| > q \qquad \left(i \neq j; t_i, t_j \in T\right) \end{cases}$$

se dice que el proceso $\{Y(t)\}$ es **incorrelacionado de orden q**.

Si se verifica que $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \gamma\left[Y(t_i),Y(t_j)\right] = 0\left(i \neq j; t_i, t_j \in T\right)$, se dice que el proceso $\{Y(t)\}$ es **asintóticamente incorrelacionado.**

Si se verifica: **i)** que $E\left[Y(t_n) / y(t_m), y(t_\ell), ...\right] = E\left[Y(t_n)\right] = 0$; **ii)** que las autocovarianzas condicionadas son tales que $\gamma\left[Y(t_n),Y(t_m) / y(t_\ell), y(t_k), ...\right] = 0\left(m \neq n; t_n, t_m, t_\ell, t_k, ... \in T\right)$ y **iii)** que las varianzas condicionadas son constantes, $\sigma^2\left[Y(t_n) / y(t_m), y(t_\ell), ...\right] = \sigma_Y^2 < \infty \left(t_n, t_m, t_\ell, ... \in T\right)$, entonces se dice que $\{Y(t)\}$ es un proceso **ruido blanco condicionado (WN-condicionado)**. Si, por el contrario, se verifica que $\sigma^2\left[Y(t_n) / y(t_m), y(t_\ell), ...\right]$ $\left(t_n, t_m, t_\ell, ... \in T\right)$ depende del tiempo, se dice que el proceso $\{Y(t)\}$ es un **WN-con heterocedasticidad condicionada.**

4) La existencia de relación de dependencia entre las variables aleatorias definidas por los incrementos de las variables Y(t): Se dice que el proceso unidimensional $\{Y(t)\}$ $(t \in T)$ es **de incrementos independientes** si se verifica que las variables aleatorias $Y(t_k) - Y(t_j)$ e $Y(t_i) - Y(t_h)$ $\left(t_h < t_i \leq t_j < t_k\right)$ son independientes entre sí. Dado que $f\left(Y(t_j) / y(t_{j-1})\right) = f_j\left(Y(t_j) - Y(t_{j-1})\right) (j = 2,3,...)$, si se verifica que:

$$f_{Y(t_1),...,Y(t_n)}\left(y(t_1),...,y(t_n)\right) = f_{Y(t_1)}\left(y(t_1)\right)\prod_{j=2}^{n} f\left[Y(t_j) - Y(t_{j-1})\right]$$

se puede concluir que, si $\{Y(t)\}$ es un proceso con incrementos independientes, entonces es estrictamente

---

[46]. Algunos procesos determinísticos-caóticos tienen momentos de primero y segundo orden similares a los de un proceso estocástico. Si esta similitud es con respecto a un proceso WN, queda definido un proceso determinístico conocido como caos blanco ("white chaos", WC, que pertenecerá, inevitablemente, a una de dos clases posibles: "for ladies" o "for men"). Por ejemplo, para un proceso discreto en el dominio del tiempo del tipo "tent-map" de la forma $Y_t = a^{-1}Y_{t-1}\left(0 \leq Y_{t-1} \leq a\right)$ y $Y_t = (1-a)^{-1}(1 - Y_{t-1})\left(a \leq Y_{t-1} \leq 1\right)$ se verifica que la función de autocorrelaciones es similar a la de un proceso $AR(1)$ y, cuando $a = \frac{1}{2}$, es similar a la de un proceso $AR(0)$ (ver Sakai, H.; Tokumaru, H. (1980)).

$AR(1)$.

Sea un proceso $\{Z(t)\}$ tal que $Z(t_1) = Y(t_1)$ y $Z(t_j) = Y(t_j) - Y(t_{j-1})$ $(j = 2,3,...)$, de modo que $Y(t_j) = \sum_{i=1}^{j} Z(t_i)$ $(j = 1,2,...,n)$, se puede concluir, entonces, que $\{Y(t)\}$ es linealmente dependiente con respecto a sus incrementos anteriores, $Y(t_j) = Y(t_1) + \sum_{i=2}^{j}[Y(t_i) - Y(t_{i-1})]$. Luego, se puede concluir que la distribución de probabilidades de la variable $(Y(t_j) - Y(t_{j-2}))$ es igual a la distribución de la variable $(Y(t_j) - Y(t_{j-1})) + (Y(t_{j-1}) - Y(t_{j-2}))$. Viceversa, si $\{Z(t)\}$ es un proceso formado por variables aleatorias independientes, entonces el proceso $\{Y(t)\}$, tal que $Y(t_n) - Y(t_1) = \sum_{j=1}^{n} Z(t_j)$ $(n = 1,2,...)$, es de incrementos independientes[47].

Se dice que el proceso unidimensional $\{Y(t), t \in T\}$ es **de incrementos incorrelacionados** si se verifica que:

$$E\{[Y(t_k) - Y(t_j)][Y(t_i) - Y(t_h)]\} = E[Y(t_k) - Y(t_j)]E[Y(t_i) - Y(t_h)] \qquad (t_h < t_i < t_j < t_k)$$

es decir, si las autocovarianzas $\gamma\{[Y(t_k) - Y(t_j)][Y(t_i) - Y(t_h)]\}$ son nulas para todo $t_h < t_i \leq t_j < t_k$ [48].

Debe tenerse en cuenta que, dado que la clase de los procesos de incrementos incorrelacionados es más amplia que la de los procesos de incrementos independientes, todo proceso de **incrementos independientes** es de **incrementos incorrelacionados**, pero que la implicación inversa no se verifica necesariamente.

1.4.- El concepto de estacionariedad

1.4.1.- En los procesos unidimensionales

Se dice que un proceso $\{Y(t); t \in T\}$ es **estacionario en sentido estricto** (o **invariante** u **homogéneo en el tiempo**), si las distribuciones de probabilidades conjuntas para las sucesiones de variables $\{Y(t_1), Y(t_2), ..., Y(t_n)\}$ e $\{Y(t_1 + h), Y(t + h), ..., Y(t_n + h)\}$ son iguales, para todo $h > 0$ y para cualquier subconjunto $(t_1, t_2, ..., t_n) \in T$:

---

[47]. Como se verá en la Sec. 4.1, los procesos con incrementos independientes integran la familia de los procesos **"random walk"** (los cuales, a su vez, forman parte del conjunto de las **martingalas**). En particular, si el proceso formado por los incrementos independientes de $\{Y(t)\}$ tiene varianza constante, entonces se dice que $\{Y(t)\}$ es un **"random walk homocedástico"**.

[48]. Como se verá en la Sec. 4.1, los procesos de incrementos incorrelacionados integran la familia de los procesos **"random walk" débiles** o **"random walk"-3**.

$$F_{Y(t_1),Y(t_2),\ldots,Y(t_n)}\left[y(t_1),y(t_2),\ldots,y(t_n)\right] = F_{Y(t_1+h),Y(t_2+h),\ldots,Y(t_n+h)}\left[y(t_1+h),y(t_2+h),\ldots,y(t_n+h)\right]$$

(para $(t_i + h \in T; i = 1,2,\ldots,n)$). Es decir, si:

$$\left\{Y(t_1),Y(t_2),\ldots,Y(t_n)\right\} \overset{d}{=} \left\{Y(t_1+h),Y(t_2+h),\ldots,Y(t_n+h)\right\} \quad (t_i + h \in T; i = 1,2,\ldots,n)$$

Haciendo $n = 1$, será $F_{Y(t_i)}(y(t_i)) = F_{Y(t_i+h)}(y(t_i+h))\,(t_i, t_i + h \in T; i = 1,2,\ldots,n)$. En particular, para $h = -t_i$, será $F_{Y(t_i)}(y(t_i)) = F_{Y(0)}(y(0))\,(t_i \in T; i = 1,2,\ldots,n)$. Se concluye por lo tanto que, si un proceso es estacionario en sentido estricto, las funciones de distribución de probabilidades $F_{Y(t_i)}(y(t_i))\,(t_i \in T;$ $i = 1,2,\ldots,n)$ son independientes del tiempo, es decir que todas las variables $Y(t)$ tienen la misma distribución de probabilidades, $F_{Y(t_1)}[y(t_1)] = F_{Y(t_2)}[y(t_2)] = \ldots = F_{Y(t_n)}[y(t_n)] = F_Y(y)$ [49] y, por lo tanto, que $dF_{Y(t)}(y(t)) = dF_Y(y)$. Es decir que la estacionariedad estricta implica la igualdad de las funciones de densidad marginales[50].

Para $n = 2$, se verificará que:

$$F_{Y(t_i),Y(t_j)}\left[y(t_i),y(t_j)\right] = F_{Y(t_i+h),Y(t_j+h)}\left[y(t_i+h),y(t_j+h)\right] = F_{t_i-t_j}(y)$$

$\left(para\ t_i,t_j,t_i+h,t_j+h \in T; i,j = 1,2,\ldots; i \neq j\right)$. Es decir que la función de distribución conjunta depende de la distancia que separa a las variables del proceso $(t_i - t_j)$, pero no depende ni de t_i ni de t_j (esta propiedad implica que los parámetros de las distribución conjunta no dependen de las realizaciones $y(t_i)$ e $y(t_j)$ sino de las distancias en el tiempo que las separa)[51].

Con un razonamiento similar, se dice que un proceso $\{Y(t)\}$ es **de incrementos estacionarios** si se verifica que $Y(t_j) - Y(t_i) \overset{d}{=} Y(t_j+h) - Y(t_i+h)\,(t_i,t_j \in T)$. Debe tenerse en cuenta que el hecho que un proceso sea de incrementos estacionarios, no implica que sea estacionario.

En general, dados un proceso $\{Y(t)\}$ estrictamente estacionario y una función continua $g(\bullet)$, se demuestra que el proceso $\{g[Y(t)]\}$ también es estrictamente estacionario.

Dada la correspondencia biunívoca entre funciones de distribución de probabilidades y funciones

---

[49]. Ver Khinchin, A.Y. (1934).

[50]. Obsérvese que, en este sentido, la estacionariedad estricta puede ser interpretada como una extensión de la condición de **homogeneidad absoluta** para procesos autorregresivos.

[51]. Obsérvese que, igual que en la condición de estacionariedad estricta, la intercambiabilidad supone la invariancia con respecto a la combinación de las variables pero, contrariamente a la condición de estacionariedad, no supone que la dependencia de las variables $Y(t)$ sea una función de la distancia entre las variables.

características [52], la definición de estacionariedad estricta puede ser expresada, también, en términos de la función característica asociada a la distribución $F_{Y(t)}(\bullet)$:

$$C_{Y(t_1),Y(t_2),\ldots,Y(t_n)}\left(w_{t_1},w_{t_2},\ldots,w_{t_n}\right) = C_{Y(t_1+h),Y(t_2h),\ldots,Y(t_n+h)}\left(w_{t_1},w_{t_2},\ldots,w_{t_n}\right)$$

(para $t_i, t_i + h \in T; i = 1,2,\ldots,n$).

Dado un proceso $\{Y(t)\}$ estrictamente estacionario, se dice que es **reversible en el dominio del tiempo** si, para todo k entero, las distribuciones de probabilidades conjuntas para las sucesiones de variables $\{Y(t_1),Y(t_2),\ldots,Y(t_k)\}$ e $\{Y(-t_1),Y(-t_2),\ldots,Y(-t_k)\}$ son iguales para todo subconjunto $(t_1,t_2,\ldots,t_k) \in T$:

$$F_{Y(t_1),Y(t_2),\ldots,Y(t_k)}\left[y(t_1),y(t_2),\ldots,y(t_k)\right] = F_{Y(-t_1),Y(-t_2),\ldots,Y(-t_k)}\left[y(-t_1),y(-t_2),\ldots,y(-t_k)\right]$$

Se dice que un proceso unidimensional $\{Y(t); t \in T\}$ es **estacionario de orden s en sentido amplio** o **débilmente estacionario de orden s** ($s = 1,2,\ldots$), si sus momentos de orden s satisfacen la condición de **invariancia** con respecto a la variable tiempo:

$$E\left[Y(t_1)Y(t_2)\ldots Y(t_s)\right] = E\left[Y(t_1+h)Y(t_2+h)\ldots Y(t_s+h)\right] \qquad (t_i, t_i + h \in T; i = 1,2,\ldots,s)$$

En particular, todo proceso estocástico con valor medio constante, $E[Y(t)] = m$, momentos de segundo orden finitos, $E\left\{\left[Y(t)\right]^2\right\} < \infty$ y autocovarianzas, $\gamma\left[Y(t_i),Y(t_j)\right] = E\left\{\left[Y(t_i)-m\right]\left[Y(t_j)-m\right]\right\}$ $(t_i,t_j \in T; i,j = 1,2,\ldots)$, dependientes exclusivamente del valor $|t_i - t_j|$, $\gamma\left[Y(t_i),Y(t_j)\right] = \gamma_{|t_i-t_j|}(Y)$ $(t_i,t_j \in T; i,j = 1,2,\ldots)$, se dice que es **estacionario en las covarianzas**[53].

De las definiciónes anteriores se puede concluir fácilmente que todo proceso estacionario en sentido estricto que posea, además, momentos de segundo orden finitos, es también estacionario en las covarianzas[54]. Pero, dado que los momentos de orden superior a dos pueden no ser homogéneos en el tiempo, la implicación inversa no se verifica necesariamente. Sólo si el proceso estacionario en las covarianzas es Normal, será

---

[52]. Debe tenerse en cuenta que, dada una variable aleatoria $Y(t)$, su función característica está definida como la transformación de Fourier de su función de densidad, $C_Y(w,t) = E\left(e^{iwY(t)}\right) = E\left\{\cos\left[wY(t)\right]\right\} + iE\left\{sen\left[wY(t)\right]\right\}$ $(w \in R)$ y, dado que $E\left|e^{iwY(t)}\right| = 1$ (para todo w y todo $Y(t)$), de acuerdo con la desigualdad triangular, se verifica que $\left|E\left(e^{iwY(t)}\right)\right| \le E\left(\left|e^{iwY(t)}\right|\right) = 1$. Se puede concluir, entonces, que la función característica existe para cualquier variable aleatoria $Y(t)$ y, por lo tanto, que la variable queda completamente identificada por su función característica, aún en el caso que no todos sus momentos existan.

[53]. De esta expresión se puede concluir fácilmente que, si $\{Y(t)\}$ es estacionario de segundo orden, su función de autocovarianzas es par, es decir, $\gamma_{t_i-t_j}(Y) = \gamma_{t_j-t_i}(Y)$ $(t_i,t_j \in T; i,j = 1,2,\ldots)$ (esta propiedad será tratada nuevamente en la Sec. 2.2).

[54]. Como se verá en el Cap. 4 al tratar los procesos vinculados a la econometría del mercado financiero, el proceso representativo de las rentabilidades de un activo es habitualmente estrictamente estacionario pero, como en general su varianza es infinita, puede no ser estacionario de segundo orden.

también estrictamente estacionario[55].

Como un corolario de la definición de estacionariedad de segundo orden, se demuestra que, dado un proceso $\{Y(t)\}$ estacionario en las covarianzas, continuo en valor medio cuadrático y tal que $m(Y) = 0$ se verifica que[56]:

$$E\left\{\left[Y(t_i) - Y(t_j)\right]^2\right\} = E\left\{\left[Y(t_i)\right]^2\right\} + E\left\{\left[Y(t_j)\right]^2\right\} - 2E\left[Y(t_i)Y(t_j)\right] =$$
$$= 2E\left\{\left[Y(t_i)\right]^2\right\} - 2E\left[Y(t_i)Y(t_j)\right] =$$
$$= 2\left[\gamma_0(Y) - \gamma_{|t_i - t_j|}(Y)\right]$$

Sea un proceso $\{Y_t\}$ $(t = 1,2,...)$ unidimensional, discreto en el dominio del tiempo continuo en el dominio de las variables. Como una extensión inmediata de las definiciones anteriores, se dice que el proceso $\{Y_t\}$ es estrictamente estacionario si la distribución de probabilidades conjunta para las sucesiones de variables $\{Y_1, Y_2, ..., Y_n\}$ e $\{Y_{1+h}, Y_{2+h}, ..., Y_{n+h}\}$ son iguales para todo $h > 0$:

$$F_{Y_1,Y_2,...,Y_n}(y_1, y_2, ..., y_n) = F_{Y_{1+h},Y_{2+h},...,Y_{n+h}}(y_{1+h}, y_{2+h}, ..., y_{n+h}) \qquad (n = 1,2,...)$$

y se dice que es estacionario en las covarianzas si se verifica que: **i)** $E(Y_t) = m(Y)$ $(t = 1,2,...)$; **ii)** $E\left(|Y_t|^2\right) < \infty$ $(t = 1,2,...)$ y **iii)** su función de autocovarianzas depende exclusivamente del orden j, $\gamma(Y_t, Y_{t+j}) = \gamma_j(Y)$ $(j = 0,1,2,...)$.

Ejemplo 1.1:

Sea un proceso $\{Y(t)\}$ estrictamente estacionario que admite una representación $AR(p)$ $(p \neq 0)$. Su función de distribución de probabilidades conjunta $F_{Y(t_1),Y(t_2)}(y(t_1), y(t_2))$ se mantendrá invariante en el tiempo siempre que la distancia $|t_1 - t_2|$ se mantenga constante. La naturaleza de la función de distribución conjunta puede ser inferida a partir de la forma que asuman los diagramas de dispersión para los pares de variables $(Y(t), Y(t+j))$ $(j = 1,2,...)$. En el **GRAFICO 1.1** se observa la existencia de correlación entre las variables $Y(t)$ e $Y(t+j)$: positiva para los pares de variables $(Y(t), Y(t+2))$ e $(Y(t), Y(t+4))$ (gráficos **1.1.b** y **1.1.d**) y negativa para los pares de variables $(Y(t), Y(t+1))$ e $(Y(t), Y(t+3))$ (gráficos **1.1.a** y **1.1.c**).

---

[55]. Cuando en el texto se califica a un proceso simplemente como estacionario, debe interpretarse que es estacionario en las covarianzas.

[56]. Como se verá en el Cap. 4, ésta es una propiedad muy útil en las aplicaciones al mercado de capitales.

Ejemplo 1.2:

Sea un proceso $\{Y(t)\}$ débilmente $AR(0)$. Como se observa en el **GRAFICO 1.2**, los diagramas de dispersión indican una distribución al azar de los pares de variables $(Y(t), Y(t+j))$ $(j = 1,2,...)$.

1.4.2.- En los procesos k-dimensionales

Sea $\{\Omega^{(k)}(t)\} = \{\{Y_1(t)\},\{Y_2(t)\},...,\{Y_k(t)\}\}$ $(t \in T)$ un **proceso k-dimensional** formado por la combinación de los procesos marginales $\{Y_j(t)\}$ $(j = 1,2,...,k)$, con función de distribución de probabilidades conjunta de la forma $F_{Y_1(t_1),...,Y_1(t_n),...,Y_k(t_1),...,Y_k(t_n)}(y_1(t_1),...,y_1(t_n),...,y_k(t_1),...,y_k(t_n))$.

Se dice que el proceso $\{\Omega^{(k)}(t)\}$ está formado por procesos **independientes** si, para todo conjunto finito $t_1, t_2, ..., t_n \in T$, su función de probabilidades conjunta puede ser expresada como el producto de las distribuciones de probabilidades de los procesos marginales:

$$F_{Y_1(t_1),...,Y_1(t_n),...,Y_k(t_1),...,Y_k(t_n)}(y_1(t_1),...,y_1(t_n),...,y_k(t_1),...,y_k(t_n)) = \prod_{j=1}^{k} F_{Y_j(t_1),...,Y_j(t_n)}(y_j(t_1),...,y_j(t_n))$$

Generalizando esta definición, se puede asegurar que una sucesión infinita $\{Y_1(t), Y_2(t), Y_3(t),...\}$ es de procesos independientes si cualquier subconjunto finito $\{Y_1(t), Y_2(t),..., Y_k(t)\}$ está formado por procesos independientes.

Se dice que el proceso k-dimensional $\{\Omega^{(k)}(t)\}$ es estrictamente estacionario si se verifica que:

$$F_{Y_1(t_1),...,Y_1(t_n),...,Y_k(t_1),...,Y_k(t_n)}(\bullet) = F_{Y_1(t_1+\tau),...,Y_1(t_n+\tau),...,Y_k(t_1+\tau),...,Y_k(t_n+\tau)}(\bullet) \qquad (t_1, t_2, ..., t_n \in T)$$

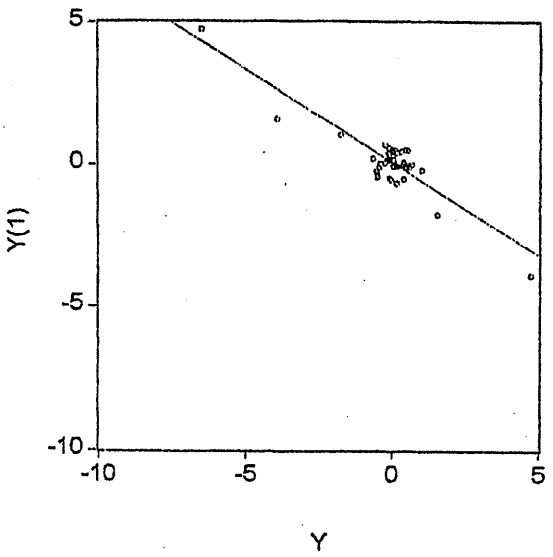
Si los procesos $\{Y_j(t)\}$ $(j = 1,2,...,k)$ son **conjuntamente estacionarios en sentido estricto**, entonces se puede asegurar que también son individualmente estacionarios en sentido estricto. Pero la implicación inversa sólo se verifica si los procesos individualmente estacionarios son, además, independientes[57].

Por otra parte, que los procesos $\{Y_j(t)\}$ $(j = 1,2,...,k)$ sean individualmente estacionarios de segundo orden es condición suficiente (pero no necesaria) para que el proceso k-dimensional $\{\Omega^{(k)}(t)\} = \{Y_1(t), Y_2(t),..., Y_k(t)\}$ sea **conjuntamente estacionario de segundo orden.**

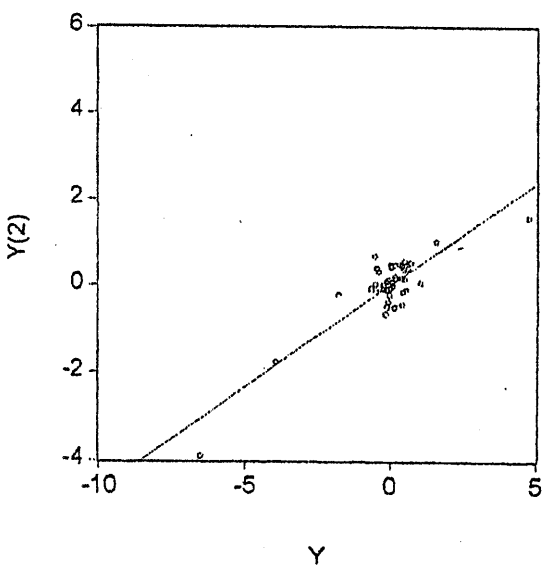
---

[57]. Debe tenerse en cuenta que la estacionariedad conjunta es una propiedad de la distribución de probabilidades conjunta, en tanto que la estacionariedad individual es una propiedad de las distribuciones de probabilidades marginales.

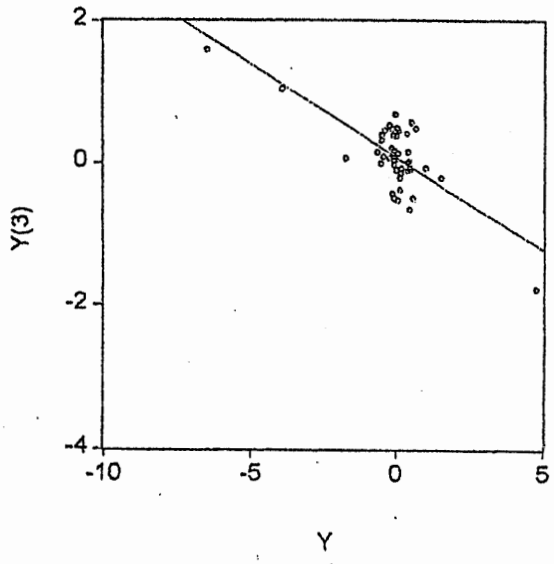
GRAFICO 1.1: Diagramas de dispersión correspondientes a los pares de variables (Y(t),Y(t+j)) (j=1,2,3,4), que definen un proceso {Y(t)}:AR(p)



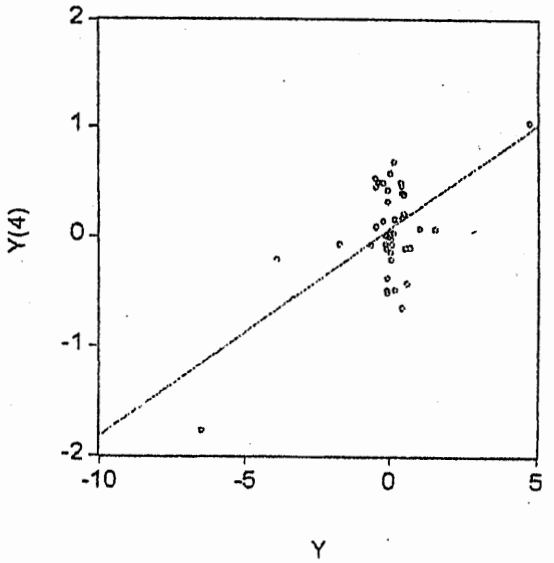
(a)



(b)

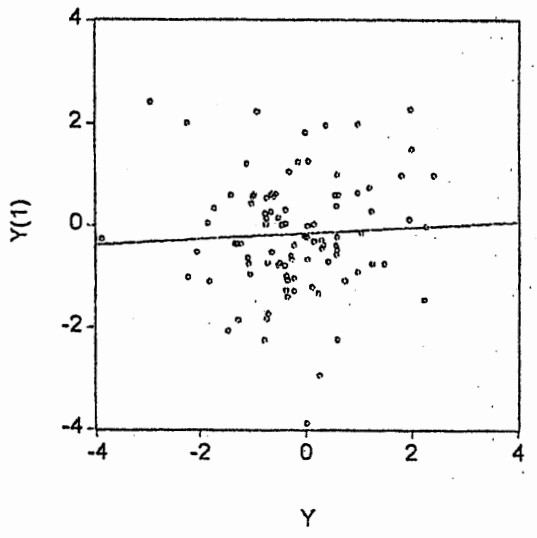


(c)

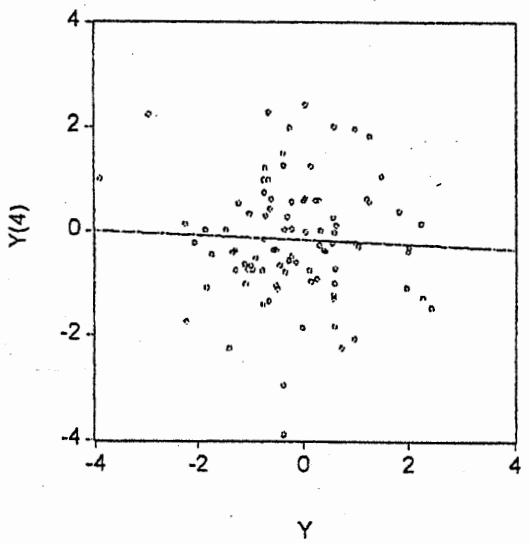


(d)

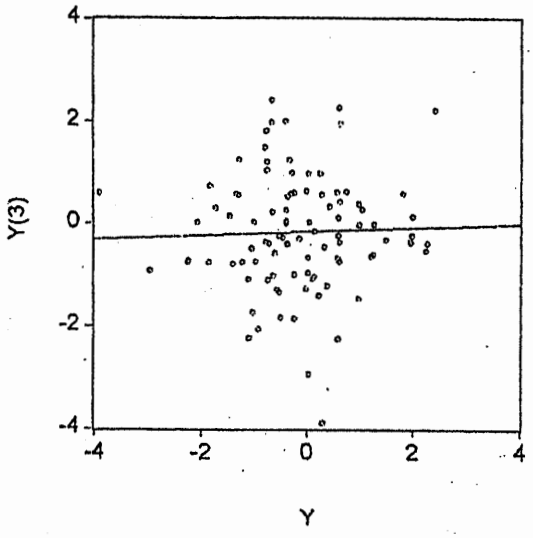
GRAFICO 1.2: Diagramas de dispersión correspondientes a los pares de variables (Y(t),Y(t+j)) (j=1,2,3,4), que definen un proceso {Y(t)}:AR(0)



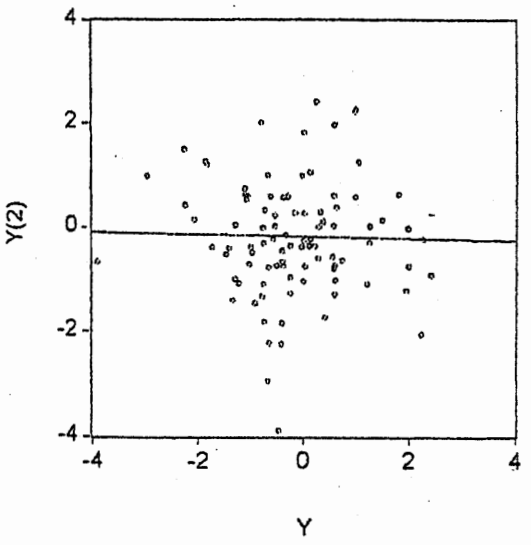
(a)



(b)



(c)



(d)

Sea $\left\{\Omega_t^{(k)}\right\} = \left\{\left\{Y_t^{(1)}\right\},\left\{Y_t^{(2)}\right\},...,\left\{Y_t^{(k)}\right\}\right\}$ un proceso estocástico k-dimensional discreto en el dominio del tiempo, como una extensión de las definiciones anteriores, se dice que $\left\{\Omega_t^{(k)}\right\}$ es estacionario de segundo orden si se verifica que el vector formado por los momentos de primer orden y la matriz de autocovarianzas y covarianzas cruzadas de los procesos marginales satisfacen la condición de invariancia con respecto al tiempo:

$$E\left(\Omega_t^{(k)}\right) = \begin{bmatrix} E\left(Y_t^{(1)}\right) \\ E\left(Y_t^{(2)}\right) \\ ... \\ E\left(Y_t^{(k)}\right) \end{bmatrix} = \begin{bmatrix} m\left(Y^{(1)}\right) \\ m\left(Y^{(2)}\right) \\ ... \\ m\left(Y^{(k)}\right) \end{bmatrix} = m\left(\Omega^{(k)}\right)$$

$$\Gamma_j\left(\Omega_t^{(k)}\right) = E\left\{\left[\Omega_{t-j}^{(k)} - m\left(\Omega^{(k)}\right)\right]\left[\Omega_t^{(k)} - m\left(\Omega^{(k)}\right)\right]^T\right\} =$$

$$= \begin{bmatrix} \gamma\left(Y_{t-j}^{(1)}, Y_t^{(1)}\right) & \gamma\left(Y_{t-j}^{(1)}, Y_t^{(2)}\right) & ... & \gamma\left(Y_{t-j}^{(1)}, Y_t^{(k)}\right) \\ \gamma\left(Y_{t-j}^{(2)}, Y_t^{(1)}\right) & \gamma\left(Y_{t-j}^{(2)}, Y_t^{(2)}\right) & ... & \gamma\left(Y_{t-j}^{(2)}, Y_t^{(k)}\right) \\ ... & ... & ... & ... \\ \gamma\left(Y_{t-j}^{(k)}, Y_t^{(1)}\right) & \gamma\left(Y_{t-j}^{(k)}, Y_t^{(2)}\right) & ... & \gamma\left(Y_{t-j}^{(k)}, Y_t^{(k)}\right) \end{bmatrix} =$$

$$= \begin{bmatrix} \gamma_j\left(Y^{(1)}\right) & \gamma_j\left(Y^{(1)}, Y^{(2)}\right) & ... & \gamma_j\left(Y^{(1)}, Y^{(k)}\right) \\ ... & ... & ... & ... \\ \gamma_j\left(Y^{(k)}, Y^{(1)}\right) & \gamma_j\left(Y^{(k)}, Y^{(2)}\right) & ... & \gamma_j\left(Y^{(k)}\right) \end{bmatrix} =$$

$$= \Gamma_j\left(\Omega^{(k)}\right) \qquad (j = 0,\pm 1,\pm 2,...)$$

Se dice que la sucesión de procesos marginales $\left\{\left\{Y_1(t)\right\},\left\{Y_2(t)\right\},...,\left\{Y_k(t)\right\}\right\}$ $(t \in T)$ **converge en-probabilidad** al proceso $\left\{\Omega(t)\right\}$ si las variables aleatorias que se generan para cada t en particular, convergen en-probabilidad a la variable $\Omega(t)$ [58].

Por otra parte, se demuestra que, si los procesos de la sucesión $\left\{Y_j(t)\right\}$ $(j = 1,2,...,n)$, además de converger en-probabilidad a $\left\{\Omega(t)\right\}$, son individualmente estacionarios en sentido estricto, el proceso-límite $\left\{\Omega(t)\right\}$ también es estrictamente estacionario [59].

---

[58]. Se dice que una sucesión de variables aleatorias $\left\{Y(t_1,\omega),Y(t_2,\omega),...,Y(t_n,\omega)\right\}$ $\left(\omega \in \Omega(Y)\right)$ converge en-probabilidad a una variable $Y(t,\omega)$ $\left(\omega \in \Omega(Y)\right)$ si se verifica que $\lim_{n\to\infty} p\left(\left|Y(t_n,\omega) - Y(t,\omega)\right| < \varepsilon\right) = 1$ $(\varepsilon > 0)$.

[59]. La convergencia en-probabilidad es una de las formas de **convergencia estocástica**, entre las que cabe mencionar: **i)** La **convergencia casi con certeza**: Dada una sucesión de variables $\left\{Y(t_j,\omega)\right\}$ $\left(j = 1,2,...,n;\omega \in \Omega(Y)\right)$, si se verifica que

1.5.- El concepto de ergodicidad

Como se puede concluir de las definiciones dadas en la sección precedente, decir que un proceso es **estacionario** implica suponer que, de alguna forma, los elementos de su estructura estocástica (sus momentos de primero y segundo orden en el caso de la estacionariedad débil, su distribución de probabilidades en el caso de la estacionariedad estricta) son fijos a través del tiempo. Afirmar que, además, es **ergódico** significa suponer que esta estructura generadora del proceso, a pesar de que debe ser considerada en el dominio de las variables, puede ser inferida en el dominio del tiempo, a partir de la observación de la muestra finita que constituye la serie cronológica formada por una de sus realizaciones posibles (teorema de Birkhoff-Khinchin, ver Birkhoff, G.D. (1931), Khinchin, A.I. (1932)). En otros términos, se dice que un proceso es ergódico cuando sus **momentos muestrales finitos**, obtenidos a partir de una serie cronológica finita convergen, en cierta forma, a los correspondientes **momentos muestrales infinitos**, los cuales coinciden con los **momentos poblacionales** del proceso. Es decir, dada una función $h(Y_t)$ tal que $E\left[\left|h(Y_t(w))\right|\right] < \infty$, se dice que el proceso estacionario $\{Y_t\}$ es ergódico si se verifica que:

$$\lim_{n\to\infty} \overline{h_Y^{(n)}} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} h\left[Y_t(w)\right] = E\left[h(Y_t(w))\right] = \int_{w\in\Omega(y)} h\left[Y_t(w)\right] f\left[Y_t(w)\right] dY_t(w)$$

(donde h_Y denota el tiempo promedio de una función de los estados $h\left[Y_t(w)\right]$), suponiendo que el límite en el dominio del tiempo existe.

En particular, dado un proceso $\{Y(t)\}$ estacionario y tal que $E(Y(t)) = m(Y(t)) = 0$, si se verifica que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Y(t)dt = E\left[Y(t)\right] = m(Y(t)) = \int_{w\in\Omega(Y)} Y(t) f(Y(t)) dY(t)$$

$\sum_{j=1}^{n} p\left(\lim_{n\to\infty}\left|Y(t_n,\omega) - Y(t,\omega)\right| = 0\right) = 1\ (\omega \in \Omega(Y))$, se dice que $Y(t_n,\omega)$ converge casi con certeza (con probabilidad igual a uno) a $Y(t,\omega)$ ($Y(t_n,\omega) \xrightarrow{ccc} Y(t,\omega)$). La convergencia casi con certeza es más fuerte que la convergencia en-probabilidad. Es decir, $Y(t_n,\omega) \xrightarrow{ccc} Y(t,\omega) \Rightarrow Y(t_n,\omega) \xrightarrow{p} Y(t,\omega)$, pero la implicación inversa no se verifica necesariamente. **ii)** La **convergencia completa**: Dada una sucesión de variables aleatorias $Y(t_j,\omega)\ (j = 1,2,\ldots,n)$, se dice que $Y(t_n,\omega)$ converge en forma completa a $Y(t,\omega)$ ($Y(t_n,\omega) \xrightarrow{c} Y(t,\omega)$) si se verifica que $\lim_{n\to\infty} \sum_{j=n}^{\infty} p\left(\left|Y(t_j,\omega) - Y(t,\omega)\right| > \varepsilon\right) = 0\ (\varepsilon > 0)$,. De acuerdo con el teorema de Borel-Cantelli, la convergencia completa implica la convergencia casi con certeza: $Y(t_n,\omega) \xrightarrow{c} Y(t,\omega) \Rightarrow Y(t_n,\omega) \xrightarrow{ccc} Y(t,\omega)$. La implicación inversa se verifica sólo en el caso en que las variables $Y(t_j,\omega)$ sean independientes. **iii)** La **convergenciae en-distribución** (o **convergencia débil**): Dada una sucesión de variables aleatorias $Y(t_j,\omega)\ (j = 1,2,\ldots,n)$, si se verifica que $\lim_{n\to\infty} F_{Y(t_n,\omega)}(y) = F_{Y(t,\omega)}(y)$ (donde y es un punto de continuidad de $F_{Y(t,\omega)}(\bullet)$), se dice que $Y(t_n,\omega)$ converge en-distribución a $Y(t,\omega)$ ($Y(t_n,\omega) \xrightarrow{d} Y(t,\omega)$ o $Y(t_n,\omega) \Rightarrow Y(t,\omega)$). Esta es la forma más débil de convergencia estocástica (debe tenerse en cuenta que esta forma de convergencia puede ser definida aún en el caso en que las variables $Y(t_n,\omega)$ posean distribuciones diferentes). Las consideraciones anteriores, permiten concluir, entonces, que $Y(t_n,\omega) \xrightarrow{c} Y(t,\omega) \Rightarrow Y(t_n,\omega) \xrightarrow{ccc} Y(t,\omega) \Rightarrow Y(t_n,\omega) \xrightarrow{p} Y(t,\omega) \Rightarrow Y(t_n,\omega) \xrightarrow{d} Y(t,\omega)$ (ver Tremayne, A. (1993), Landro, A.H. (2002), Rosenthal, J.S. (2006)).

se dice que el proceso $\{Y(t)\}$ es **ergódico de primer orden** o **ergódico en el valor esperado** y, si se verifica que[60]:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}Y(t)Y(t+\tau)dt = E\left[Y(t)Y(t+\tau)\right] = \gamma\left(Y_t,Y_{t+\tau}\right) = \int_{w\in\Omega(Y)}Y(t)Y(t+\tau)f\left[Y(t)\right]dY$$

se dice que el proceso $\{Y(t)\}$ es **ergódico de segundo orden** o **ergódico en las autocovarianzas**[61].

Los llamados **teoremas ergódicos**, vinculados con las leyes débil y fuerte de los grandes números[62], analizan las condiciones de independencia asintótica bajo las cuales se puede garantizar la convergencia de los momentos muestrales[63].

Un **primer teorema ergódico** es la **ley débil de los grandes números** para sucesiones de variables iid[64], la cual demuestra que, dado un proceso unidimensional formado por variables iid, $\{Y_1,Y_2,...\}$, con valor

---

60. La teoría ergódica se inicia con el trabajo de Boltzmann, L. (1868) sobre la justificación de la definición de valores promedio en la teoría cinética. En este ámbito, el **problema ergódico** fundamental consiste, precisamente, en hallar hipótesis que justifiquen la coincidencia de la medida geométrica de un conjunto de estados A (denominada **promedio de fase**) con el límite del tiempo relativo que el proceso transcurre en el conjunto A (la **probabilidad del tiempo promedio**, según la nomenclatura de Boltzmann, L. (1871)).

61. De acuerdo con Boltzmann, L. (1868), el fundamento de la hipótesis ergódica es justificar la introducción de una distribución de probabilidades apropiada para calcular los valores promedio del proceso dinámico. Sea $T\left(Y_t(w)\right)$ la función que representa la ley que rige la evolución del proceso $\{Y_t(w)\}$, entonces el tiempo promedio queda definido de la siguiente forma $h_T = \lim_{t\to\infty}\frac{1}{t}\int_0^t f\left[T\left(Y_t(w)\right)\right]dw$. En particular, si se verifica que $h\left[Y_t(w)\right] = I_A$ (donde I_A es una función indicador para el conjunto de estados A), el tiempo promedio representa el tiempo esperado de permanencia del proceso en el conjunto de estados A. Como se menciona en el texto, la justificación de este principio se obtiene si se demuestra que el tiempo promedio existe y es único, es decir, si $I_A = p(A)$.

62. Debe recordarse que la elegante teoría unificada de los límites de sumas de variables aleatorias independientes está formada por dos familias: **i)** el **teorema central del límite** y **ii)** la **ley de los grandes números**. La primera analiza las condiciones de convergencia en-distribución y la segunda las condiciones de la convergencia en-probabilidad. En la familia de la ley de los grandes números se pueden distinguir dos modalidades: **débil** (asociada con la convergencia en-probabilidad) y **fuerte** (asociada con la convergencia casi con certeza). Dada una sucesión de variables aleatorias $\{Y_t\}$, en el primer caso se demuestra que, bajo ciertas restricciones (ya comentadas en las páginas precedentes), $\lim_{n\to\infty}p\left\{\left|\frac{1}{n}\sum_{t=1}^{n}Y_t - \frac{1}{n}\sum_{t=1}^{n}E\left(Y_t\right)\right| < \varepsilon\right\} = 1\ (\varepsilon > 0)$ y en el segundo caso, se demuestra que $p\left\{\lim_{n\to\infty}\left[\frac{1}{n}\sum_{t=1}^{n}Y_t - \frac{1}{n}\sum_{t=1}^{n}E\left(Y_t\right)\right] = 0\right\} = \quad = 1$.

63. En Boltzmann, L. (1884)(1887) la hipótesis ergódica se asimila al postulado que afirma que la trayectoria de un proceso ergódico pasa a través de todos los puntos del dominio de los estados. Define como **"monode"** al conjunto de una clase especial de sistemas estacionarios con la misma energía. En particular, si la única restricción posible es la ecuación dada por el valor de la energía total, Boltzmann denomina al conjunto como un **"ergode"** (si existen más restricciones, lo denomina como un **"subergode"**) y hace mención de la propiedad **"ergodisch"** de un sistema. Previamente Boltzmann, L. (1875) denomina como una **"ergal"** a la función potencial del sistema.

64. En realidad, el teorema aplicado en este caso es la versión de Chebychev, P.L. (1887), la cual constituye la primera generalización de la ley de los grandes números -con respecto a las demostraciones de Bernoulli, J. (1713), y de Poisson, S.D. (1837)-

esperado $E(Y_t) = m(Y)$ y varianza $\sigma^2(Y_t) = \sigma_Y^2 < \infty$ $(t = 1,2,...)$ [65], la variable aleatoria definida por el promedio de un conjunto finito de sus realizaciones $\overline{Y}_{(n)} = \frac{1}{n}(Y_1 + Y_2 + ... + Y_n)$ tiene esperanza matemática $E(\overline{Y}_{(n)}) = m(Y)$ y varianza:

$$E\left\{\left[\overline{Y}_{(n)} - m(Y)\right]^2\right\} = \frac{1}{n}\sigma^2\left(\sum_{t=1}^{n} Y_t\right) = \frac{1}{n^2}\sum_{t=1}^{n}\sigma^2(Y_t) = \frac{1}{n}\sigma_Y^2$$

Como corolario de la propiedad de consistencia del estimador $\overline{Y}_{(n)}$, resulta que $\lim_{n\to\infty} E\left\{\left[\overline{Y}_{(n)} - m(Y)\right]^2\right\} = 0$, es decir, que $\overline{Y}_{(n)} \xrightarrow{m_2} m(Y)$, lo cual implica que $p\left[\lim_{n\to\infty}\left(\overline{Y}_{(n)} - m(Y)\right)\right] = p\left[\lim_{n\to\infty}\left(\overline{Y}_{(n)} = E(Y)\right)\right] = 1$, es decir, que $\overline{Y}_{(n)} \xrightarrow{p} m(Y)$, de modo que se puede concluir que $\{Y_t\}$ es **ergódico de primer orden**[66].

Un segundo **teorema ergódico** (basado en la versión de Bernstein, S. (1927) de la ley débil de los grandes números para procesos estacionarios de segundo orden) demuestra en primer lugar que, dado un proceso $\{Y_t\}$, estacionario en las covarianzas, con valor medio $m(Y)$ y **asintóticamente incorrelacionado**, es decir, tal que su función de autocovarianzas, $\gamma_j(Y)$ converje a un límite de Cesaro igual a cero cuando $j \to \infty$ (en otros términos, tal que $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\gamma_j(Y) = 0$), se verifica que $\overline{Y}_{(n)} = \frac{1}{n}(Y_1 + Y_2 + ... + Y_n)$ es un estimador insesgado de $m(Y)$.

Además, teniendo en cuenta que:

$$E\left\{\left[\overline{Y}_{(n)} - m(Y)\right]^2\right\} = E\left\{\left[\frac{1}{n}\sum_{t=1}^{n}(Y_t) - m(Y)\right]^2\right\} =$$

$$= \frac{1}{n^2}\left\{\sum_{t=1}^{n}\left(Y_t - m(Y)\right)^2 + 2\sum_{t<\tau}\sum\left[Y_t - m(Y)\right]\left[Y_\tau - m(Y)\right]\right\} =$$

---

que sustitye los supuestos sobre la distribución de probabilidades de las variables Y_t por las de existencia de los momentos de primero y segundo orden.

[65]. Dado que, si existe el momento $m_s(Y)$ $(s > 1)$ entonces todos los momentos de orden menor que s también existen, la condición sobre la varianza implica la condición sobre el valor esperado finito.

[66]. Debe tenerse en cuenta que, de acuerdo con la versión de Khinchin, A.Y. (1924), dada una sucesión de variables aleatorias iid con valor esperado $E(Y_t) = m(Y)$ $(t = 1,2,...)$, se verifica que $\lim_{n\to\infty} p\left(\frac{1}{n}\sum_{t=1}^{n}\left|Y_t - m(y)\right| < \varepsilon\right) = 1$ $(\varepsilon > 0)$. Obsérvese que, de acuerdo con esta versión, los postulados del primer teorema ergódico se verifican aún en el caso en que las variables Y_t iid sean de varianza infinita, es decir, como se mencionó en la página precedente, aún cuando $\{Y_t\}$: *iid* sea estrictamente estacionario pero no-estacionario en las autocovarianzas.

$$= \frac{1}{n^2}\left[n\gamma_0(Y) + 2\sum_{t=1}^{n}\sum_{\tau=1}^{n-1}\gamma_{t-\tau}(Y)\right] =$$

$$= \frac{1}{n^2}\left[2\sum_{t=1}^{n}\sum_{j=0}^{n-1}\gamma_j(Y) - n\gamma_0(Y)\right]$$

y, dado que $\lim_{n\to\infty}\frac{1}{n}\gamma_0(Y) = 0$ y que, para $k < n$, se verifica que:

$$\frac{2}{n^2}\sum_{t=1}^{n}\sum_{j=0}^{t-1}\gamma_j(Y) = \frac{2}{n^2}\left[\sum_{t=1}^{k}\sum_{j=0}^{t-1}\gamma_j(Y) + \sum_{t=k+1}^{n}\sum_{j=0}^{t-1}\gamma_j(Y)\right] \leq$$

$$\leq \frac{2}{n^2}\left|\sum_{t=1}^{k}\sum_{j=0}^{t-1}\gamma_j(Y)\right| + \frac{2}{n^2}\left|\sum_{t=k+1}^{n}\sum_{j=0}^{t-1}\gamma_j(Y)\right| =$$

$$= \frac{2}{n^2}\left|\sum_{t=1}^{k}\sum_{j=0}^{t-1}\gamma_j(Y)\right| + \frac{2}{n^2}\left|\left(\sum_{t=k+1}^{n}t\right)\left(\frac{1}{t}\sum_{j=0}^{t-1}\gamma_j(Y)\right)\right| \leq$$

$$\leq \frac{2}{n^2}\left|\sum_{t=1}^{k}\sum_{j=0}^{t-1}\gamma_j(Y)\right| + \frac{2}{n^2}\sum_{t=k+1}^{n}t\varepsilon \leq$$

$$\leq \frac{2}{n^2}\left|\sum_{t=1}^{k}\sum_{j=0}^{t-1}\gamma_j(Y)\right| + 2\varepsilon$$

se demuestra que $\lim_{n\to\infty}\frac{2}{n^2}\sum_{t=1}^{n}\sum_{j=0}^{t-1}\gamma_j(Y) = 0$ y, por lo tanto, que $\lim_{n\to\infty}E\left\{\left[Y_{(n)} - m(Y)\right]^2\right\} = 0$ [67]. Es decir, que $\overline{Y}_{(n)}$ converge en-valor medio cuadrático a $m(Y)$ y, por lo tanto, que $\overline{Y}_{(n)} \xrightarrow{p} m(Y)$ [68].

Sea, ahora, un proceso $\{Y_t\}$ estacionario de segundo orden con valor esperado nulo y función de autocovarianzas absolutamente sumable. Como corolario del teorema anterior se demuestra, entonces que, dado un proceso $\{Z_{t,j}\}$ definido de la siguiente forma: $Z_{t,j} = Y_tY_{t+j} - E(Y_tY_{t+j})$, se verifica que:

---

[67]. Debe tenerse en cuenta que, bajo el supuesto de incorrelación asintótica del proceso $\{Y_t\}$, se verifica que, para $t > k$, $\left|\frac{1}{n}\sum_{j=0}^{m}\gamma_j(Y)\right| \leq \varepsilon$.

[68]. Obsérvese que la incorrelación asintótica puede ser considerada condición suficiente para la consistencia del estimador $\overline{Y}_{(n)}$ y, por lo tanto, para la ergodicidad de primer orden. De acuerdo con las condiciones "mixing" fuertes (debidas a Rosemblatt, M. (1956) y Cogburn, R. (1960)), esta propiedad, según la cual la dependencia entre dos variables cualesquiera del proceso converge rápidamente a cero cuando la distancia entre las mismas aumenta, implica que dicho proceso posee "memoria corta" (para un mayor detalle acerca de las condiciones necesarias para que un proceso pueda ser considerado de memoria corta ver Sec. 2.5.).

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n-j}Z_{t,j} = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n-j}\left[Y_tY_{t+j} - E\left(Y_tY_{t+j}\right)\right] = \lim_{n\to\infty}E\left\{\left[\hat{\gamma}_j^{(n)}(Y) - \gamma_j(Y)\right]^2\right\} = 0$$

Es decir, que $\hat{\gamma}_j^{(n)}(Y)\xrightarrow{m_2}\gamma_j(Y)$ y, en consecuencia, que $\hat{\gamma}_j^{(n)}(Y)\xrightarrow{p}\gamma_j(Y)$. Generalizando este resultado, se puede concluir, entonces que, dado un proceso $\{Y_t\}$ estacionario de segundo orden y con función de autocorrelaciones absolutamente sumable, el estimador:

$$\hat{\gamma}_j^{(n)}(Y) = \frac{1}{n}\sum_{t=1}^{n-j}\left(Y_t - \overline{Y}_{(n)}\right)\left(Y_{t+j} - \overline{Y}_{(n)}\right)$$

converge en probabilidad al verdadero valor de la autocovarianza de orden j:

$$\hat{\gamma}_j^{(n)}(Y)\xrightarrow{p}\gamma_j(Y)$$

y, en consecuencia, que $\{Y_t\}$ es **ergódico de segundo orden** o **ergódico en las covarianzas**[69].

Debe tenerse en cuenta que las condiciones supuestas sobre las autocovarianzas para la convergencia en-probabilidad involucran a los momentos mixtos de cuarto orden ($E\left(Y_tY_{t+j}Y_{t+k}Y_{t+k+j}\right)$), circunstancia que restringe la aplicabilidad de este resultado y que esta restricción desaparece si se puede asegurar que el proceso estacionario $\{Y_t\}$ es Normal y asintóticamente incorrelacionado.

Sea $\{Y_t\}$ un proceso estrictamente estacionario y tal que $E(Y_t) = m(Y)$. Se demuestra, entonces, que la sucesión $\overline{Y}_{(n)} = \frac{1}{n}\left(Y_1 + Y_2 + ... + Y_n\right)\left(n = 1,2,...\right)$ obedece a los postulados de la **ley fuerte de los grandes números**[70], es decir, se demuestra que:

---

[69]. Para un análisis más detallado sobre las propiedades de los estimadores de la función de autocovarianzas ver Sec. 2.4.

[70]. Para la ley fuerte de los grandes números y sus generalizaciones la evolución histórica de las condiciones de convergencia puede ser resumida de la siguiente forma: **i)** que las variables Y_t sean independientes y con distribución $b(1,\theta)\left(t = 1,2,...\right)$ (Borel, E. (1909)); **ii)** que las variables Y_t sean independientes y tales que: $E(Y_t) < \infty\left(t = 1,2,...\right)$; $\sigma^2(Y_t) < \infty\left(t = 1,2,...\right)$ y $\sum_{t=1}^{\infty}\frac{1}{t^2}\sigma^2(Y_t) < \infty$ (Kolmogorov, A.N. (1930); **iii)** que las variables Y_t sean iid y tales que $E\left(|Y_t|\right) = m(Y) < \infty$ (Kolmogorov, A.N. (1931)). Esta última versión clásica de la ley fuerte de los grandes números constituye una expresión formal de la interpretación de la probabilidad como un límite fuerte de las frecuencias relativas (ver Landro, A.H. (2002)). En los casos en que todos los momentos del proceso $\{Y_t\}$ existan y sean acotados, se agrega la siguiente restricción (conocida como **ley del logaritmo iterado**) referida a la velocidad de la convergencia de las sumas parciales $\left|\frac{1}{n}\sum_{t=1}^{n}Y_t - m(Y)\right| \le \sqrt{n\ln(\ln(n))}$ (ver Khinchin, A.Y. (1924), Kolmogorov, A.N. (1928)(1929), Hartman, P.; Winter, A. (1941), Hardy, G.H.; Littlewood, J.E.; Pólya, G. (1952)). Como se verá en las páginas siguientes, el teorema central del límite proporciona la medida de la diferencia $\frac{1}{n}\sum_{t=1}^{n}Y_t - m(Y)$ (que converge a cero casi con certeza), en la convergencia en-distribución a una función de probabilidades. La ley del logaritmo iterado proporciona, en cierto sentido, el primer término del desarrollo asintótico de $\frac{1}{n}\sum_{t=1}^{n}Y_t - m(Y)$. Obsérvese que, en tanto que, a partir de la ley de los grandes números se obtiene

$$p\left\{\lim_{n\to\infty}\left[\overline{Y}_{(n)} - m(Y) = 0\right]\right\} = 1$$

o, lo que es lo mismo, que $\overline{Y}_{(n)} \xrightarrow{ccc} m(Y)$.

Las consideraciones realizadas hasta aquí se refieren a la aplicación de la teoría ergódica a los procesos estacionarios. En términos más generales, dados: **i)** un espacio L de funciones $f(\bullet)$ para los cuales existe alguna definición de integral, $\int f(\bullet)$ y **ii)** un operador H que transforma una función $f \in L$ en una función $H(f) \in L$, la teoría ergódica consiste en el análisis de las iteraciones H^n de los operadores H que satisfagan las siguientes condiciones: **i)** si f es no-negativa, entonces la transformación $H(f)$ también es no-negativa y **ii)** $\int|H(f)| \leq \int|f|$. De acuerdo a lo considerado hasta aquí, dado un proceso $\{Y_t\}$ estacionario: **i)** el espacio L será el de las funciones $f(\{Y_t\}) = f(Y_1, Y_2, \ldots)$ tales que $E\left[f(Y_1, Y_2, \ldots)\right] < \infty$; **ii)** la definición de integral es el valor esperado $\int f(\{Y_t\}) = E\left[f(Y_1, Y_2, \ldots)\right]$ y **iii)** la transformación H debe ser tal que $H\left[f(Y_1, Y_2, \ldots)\right] = f(Y_1, Y_2, \ldots)$.

Una forma alternativa de fijar las condiciones de independencia asintótica que garanticen la ergodicidad de un proceso es utilizando las **medidas de dependencia temporal** en sucesiones de variables aleatorias, basadas en las llamadas **condiciones "mixing"**.

A partir de estas condiciones y considerando la división del espacio de las trayectorias del proceso en dos subconjuntos formados por las variables Y_t futuras y pasadas, $\Im = \{Y_{t+\tau}, Y_{t+\tau+1}, Y_{t+\tau+2}, \ldots\}$ y $\Re = \{\ldots, Y_1, Y_2, \ldots, Y_t\}$, respectivamente, se pueden definir medidas de dependencia de la forma:

$$\alpha(\tau) = \sup_{\tau} \alpha(\Im, \Re) = \sup_{\tau} |p(A \cap B) - p(A)p(B)| \qquad (A \in \Im, B \in \Re)$$

$$\phi(\tau) = \sup_{\tau} (\Im, \Re) = \sup_{\tau} |p(A / B) - p(A)| \qquad (p(B) > 0, A \in \Im, B \in \Re)$$

$$\psi(\tau) = \sup_{\tau} (\Im, \Re) = \sup_{\tau} \frac{|p(A \cap B) - p(A)p(B)|}{p(A)p(B)} \qquad (p(A) > 0, p(B) > 0, A \in \Im, B \in \Re)$$

---

que $\frac{\sqrt{n}}{c_n}\left(\frac{1}{n}\sum_{t=1}^{n} Y_t - m(Y)\right) \xrightarrow{p} 0$ (para cualquier sucesión de constantes c_n tal que $c_n \uparrow \infty$), a partir de la ley del logaritmo iterado se puede concluir que $\frac{1}{c_n\sqrt{n \ln\ln(n)}}\left(\frac{1}{n}\sum_{t=1}^{n} Y_t - m(Y)\right) \xrightarrow{ccc} 0$. Es decir que la ley del logaritmo iterado proporciona una frontera entre las convergencias en-probabilidad y casi con certeza.

Se demuestra[71] que estas medidas son no-negativas y tales que $\alpha(\tau) \leq \frac{1}{4}$, $\phi(\tau) \leq 1$ y $\psi(\tau) \leq 1$.

Estas condiciones "mixing" son más fuertes que las condiciones de ergodicidad analizadas en las páginas precedentes, es decir que, dado un proceso $\{Y_t\}$ estacionario, el cumplimiento de cualquiera de estas condiciones implica ergodicidad[72].

Se demuestra, además que, cuando $\tau \to \infty$, la condición $\psi(\tau) \to 0$ implica la condición $\phi(\tau) \to 0$ y ésta implica la condición, ya mencionada, que la sucesión $\rho_j(Y)$ sea tal que $\lim_{j\to\infty} \rho_j(Y) = 0$. A su vez ésta implica la condición $\alpha(\tau) \to 0$. Es decir, cuando $\tau \to \infty$ se verifica que:

$$\psi(\tau) \to 0 \Rightarrow \phi(\tau) \to 0 \Rightarrow \alpha(\tau) \to 0 \ .$$

Como un resultado adicional de la aplicación de estas condiciones "mixing" se demuestra que: **i)** todo proceso $\{Y_t\}$ estrictamente estacionario, tal que $E\left(|Y_t|^{2+j}\right) < \infty \ (j = 1,2,...;t = 1,2,...)$ y que cumple la condición α-mixing $\sum_{t=1}^{\infty}\left[\alpha(t)\right]^{j/(2+j)} < \infty$ y **ii)** todo proceso $\{Y_t\}$ estacionario de segundo orden que cumple la condición ρ-mixing:

$$\lim_{n\to\infty}\sum_{j=1}^{n}\rho_{2^j}(Y) < \infty$$

obedecen a los postulados del **teorema central del límite**[73]:

---

[71]. Ver Hall, P.; Heyde, C.C. (1980).

[72]. Sea, por ejemplo, un proceso $\{Y_t\}\ (t = 1,2,...)$ estrictamente estacionario, tal que $E\left(|Y_t|^{2+j}\right) < \infty \ (j = 1,2, \ ...;t = 1,2,...)$ y que cumple con la condición α-mixing $\sum_{t=1}^{\infty}\left[\alpha(t)\right]^{j/2+j} < \infty$, se verifica, entonces, que $\{Y_t\}$ es ergódico.

[73]. Como se mencionó en páginas anteriores, el conjunto de los denominados teoremas en el límite está formado por las familias de la ley de los grandes números (que analiza la probabilidad de que el evento "la suma de una sucesión de variables aleatorias difiera de su valor esperado en menos de un ε real positivo"), de la ley del logaritmo iterado (que analiza la velocidad de la convergencia de las sumas parciales) y del teorema central del límite, cuyo objetivo es determinar las condiciones necesarias y suficientes para poder asegurar la convergencia en distribución a la función Normal de la sucesión $\{Y_t\}\ (t = 1,2,...)$. Las sucesivas generalizaciones de este último teorema fueron simplificando estas condiciones de convergencia de la siguiente forma muy resumida: **i)** las variables $Y_t\ (t = 1,2,...)$ deben ser independientes y con distribución $b(1,\theta)$ (De Moivre, A. (1733)); **ii)** las variables $Y_t\ (t = 1,2,...)$ deben ser iid y tales que $\sigma^2(Y_t) = \sigma^2 < \infty\ (t = 1,2,...)$ (Laplace, P.S. (1810)); **iii)** las variables $Y_t\ (t = 1,2,...)$ deben ser independientes y tales que $E(Y_t) = 0\ (t = 1,2,...)$, $\sigma^2(Y_t) = \sigma_t^2\ (t = 1,2,...)$ y $|Y_t| \leq b > 0$ (Chebychev, P.L. (1870)); **iv)** las variables $Y_t\ (t = 1,2,...)$ deben ser independientes y tales que $E\left(|Y_t|^{2+k}\right) < \infty\ (k > 0;t = 1,2,...)$ y $\lim_{n\to\infty}\frac{1}{\left(\sigma_{(n)}\right)^{2+k}}\sum_{t=1}^{n}\left|Y_t - E(Y_t)\right|^{2+k} = 0$ donde $\sigma_{(n)}^2 = \sigma^2\left(\sum_{t=1}^{n}Y_t\right)$ (Lyapunov, A.M. (1900)(1901)(1901) (1901)); **v)** las variables $Y_t\ (t = 1,2,...)$ deben ser independientes y tales que

$$\lim_{n\to\infty} p\left(\frac{1}{\sigma_{(n)}}\sum_{t=1}^{n}\left(Y_t - m_Y\right) \le u\right) = F_N(u)$$

Por otra parte, dado un proceso $\{Y_t\}$ estacionario de segundo orden y tal que cumple la condición ρ-mixing $\lim_{n\to\infty}\sum_{j=1}^{n}\left[\rho_{2^j}(Y)\right]^s < 0\left(s = \frac{1}{2},1\right)$, se puede asegurar que obedece a los postulados del **teorema central del límite funcional**[74]. Como se verá en la Sec. 4.4 al tratar los procesos "random walk" continuos en el dominio del tiempo, esta convergencia se verifica también en procesos no-estacionarios con raíces unitarias[75].

Sea $\{Y_t; t = 1,2,...\}$ un proceso estacionario y sea la varianza de su suma parcial:

$$\sigma^2\left(\frac{1}{n}\sum_{t=1}^{n}Y_t\right) = \sigma^2_{(n)}(Y) = \frac{1}{n^2}\left[\sum_{t=1}^{n}\sigma^2(Y_t) + 2\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\gamma\left(Y_t, Y_{t+j}\right)\right]$$

Dado que el primer miembro de esta expresión debe converger a cero y que, de acuerdo con la condición de estacionariedad supuesta, el primer sumando del segundo miembro converge a cero rápidamente, se puede concluir que la cuestión de la convergencia está relacionada con el comportamiento asintótico del segundo sumando del segundo miembro, es decir, de las autocovarianzas. En este sentido, teniendo en cuenta que[76]:

$$\gamma\left(Y_t, Y_{t+j}\right) = E\left\{\left[Y_t - E(Y_t)\right]\left[Y_{t+j} - E\left(Y_{t+j}\right)\right]\right\} =$$

$$= E\left(E\left\{\left[Y_t - E(Y_t)\right]\left[Y_{t+j} - E\left(Y_{t+j}\right)\right]\right\} / \left(Y_1, Y_2, ..., Y_t\right)\right) =$$

$E\left(Y_t^2\right) < \infty \left(t = 1,2,...\right)$ y $\lim_{n\to\infty}\left\{\frac{1}{s_{(n)}}\sum_{t=1}^{n}E\left(\left[Y_t - E(Y_t)\right]^2 I_{\left(|Y_t - E(Y_t)| > s_{(n)}\right)}\right)\right\} = 0$ (donde $I_{(\cdot)}$ denota una función indicador) (Lindeberg, J.W. (1922)); **vi)** las variables $Y_t \left(t = 1,2,...\right)$ deben ser independientes y tales que $E\left(Y_t^2\right) < \infty \left(t = 1,2,...\right)$ y $\lim_{n\to\infty}\left\{\max_{1\le t\le n}\left[\frac{\sigma^2(Y_t)}{\sigma^2_{(n)}}\right]\right\} = 0$ (Feller, W. (1935)(1937)).

[74]. Esta extensión del teorema central del límite será analizado en la Sec. 4.4, al tratar la transformación de un proceso "random walk" (discreto en el dominio del tiempo y continuo en el dominio de las variables) en un movimiento Browniano (ver Billingsley, P. (1968), Breiman, L. (1968), Phillips, P.C.B. (1987), Stock, J.H.; Watson, M.W. (1988), Dhrymes, P.J. (1998)).

[75]. Ver Phillips, P.C.B. (1987), Stock, J.H. Watson, M.W. (1988), Banerjee, A.; Dolado, J.; Galbraith, J.W.; Hendry, D.F. (1993), Davidson, J. (1994), Dhrymes, P.J. (1998). El problema de los procesos no-estacionarios por la presencia de raíces unitarias será tratado en el Cap. 3.

[76]. En este desarrollo se aplicaron sucesivamente las siguientes propiedades de la esperanza matemática: **i)** $E\left[E(Y/X)\right] =$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f(y/x)dy\right] f_X(x)dx = E(y)$$ y **ii)** $E\left[g_1(X)g_2(Y)/X\right] = g_1(X)E\left[g_2(Y)/X\right]$ (ver Landro, A.H. (2002)).

$$= E\left\{\left[Y_t - E(Y_t)\right]\left[E\left(Y_{t+j} / (Y_1, Y_2, ..., Y_t)\right) - E\left(Y_{t+j}\right)\right]\right\}$$

y, de acuerdo con la desigualdad de Cauchy-Schwartz[77], se puede escribir:

$$\left|\gamma\left(Y_t, Y_{t+j}\right)\right| \leq \sigma_{(n)}(Y_t)\sqrt{E\left\{\left[E\left(Y_{t+j} / (Y_1, ..., Y_t)\right) - E\left(Y_{t+j}\right)\right]^2\right\}}$$

A partir de esta relación y, teniendo en cuenta que, dado el supuesto de estacionariedad de $\{Y_t\}$, el primer factor del segundo miembro es acotado, McLeish, D.L. (1975)[78] propuso, como alternativa a la ya mencionada de ser absolutamente sumable, la siguiente condición necesaria de convergencia para la función de autocovarianzas:

$$\sqrt{E\left\{\left[E\left(Y_{t+j} / (Y_1, ..., Y_t)\right) - E\left(Y_{t+j}\right)\right]^2\right\}} \leq c_{t+j}\psi_j$$

(donde $c_{t+j} < \infty$ y $\lim_{j\to\infty} \psi_j = 0$) para asegurar la ergodicidad.

Si, dado un proceso $\{Y_t\}$, tal que $E\left(|Y_t|^2\right) < \infty$, es posible hallar dos constantes, c_t y ψ_j, como las mencionadas precedentemente, se dice que $\{Y_t\}$ es una **mixingala**[79]. Esta denominación de **condición de mixingala** se aplica también a la condición de convergencia de las autocovarianzas de McLeish.

1.6.- Algunas consideraciones preliminares acerca del teorema de descomposición de Wold

Sea $\{Y_t\}$ un proceso estocástico unidimensional estacionario de segundo orden con valor esperado nulo y varianza σ_Y^2 y tal que admite una representación $AR(p)$ de la forma:

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} + \varepsilon_t^{(p)} = Y_t^{(p)} + \varepsilon_t^{(p)}$$

Donde[80]:

1) $\left\{Y_t^{(p)}\right\}$ representa la parte de Y_t **"explicada"** por la representación $Y_t^{(p)} = a_1 Y_{t-1} + ... + a_p Y_{t-p}$ (conocida también como la estructura temporal de $\{Y_t\}$).

---

[77]. Ver Landro, A.H. (2002).

[78]. Ver también Andrews, D.W.K. (1988).

[79]. Denominación que se origina en la circunstancia que este proceso combina los conceptos de **dependencia de martingala** y de condiciones "mixing". La **martingala** será tratada en la Sec. 4.3.

[80]. Como se verá en el Cap. 2, este teorema -que fue la tesis doctoral de Wold- constituye un aporte fundamental a la interpretación de muchos aspectos conceptuales del análisis de los procesos estocásticos no-estrictamente periódicos (ver Wold, H. (1938), para una demostración rigurosa, ver también Anderson, T.W. (1964), Parzen, E. (1967), Hannan, E.J. (1970)).

2) $\left\{\varepsilon_t^{(p)}\right\}$:$WN-debil$ denota el denominado **proceso de las perturbaciones** o **de las innovaciones**, que representa la parte de $\{Y_t\}$ **"no-explicada"** por la combinación lineal de las variables $Y_{t-1},Y_{t-2},...,Y_{t-p}$ (es decir que representa la influencia que ejercen sobre el comportamiento de $\{Y_t\}$ los infinitos procesos que forman su estructura causal)[81] y que, como se explicó en la Sec. 1.3, posee las siguientes propiedades: **i)** $E\left(\varepsilon_t^{(p)}\right) = E\left(Y_t - Y_t^{(p)}\right) = E(Y_t) - \sum_{j=1}^{p} a_j Y_{t-j} = 0$;. **ii)** $E\left[Y_{t-j}\varepsilon_t^{(p)}\right] = E\left[Y_{t-j}\left(Y_t - Y_t^{(p)}\right)\right] = E\left(Y_{t-j}Y_t - \sum_{i=1}^{p} a_i Y_{t-i}Y_{t-j}\right) =$

$= \gamma_j(Y) - \sum_{i=1}^{p} a_i \gamma_{|j-i|}(Y) = 0 \ \ (j = 1,2,...,p)$ (obsérvese que se podría haber arribado a este resultado de una forma exclusivamente conceptual, teniendo en cuenta que, si existiera algún tipo de relación entre las variables $\varepsilon_t^{(p)}$ y las variables Y_t anteriores, entonces parte del comportamiento de las innovaciones podría ser "explicado" por dichas variables Y_{t-j} pero, en ese caso, los coeficientes a_j no cumplirían la condición de minimizar el error medio cuadrático $E\left[\left(Y_t - Y_t^{(p)}\right)^2\right]$); **iii)** $E\left(\varepsilon_{t-j}^{(p)}\varepsilon_t^{(p)}\right) = \gamma_j\left(\varepsilon^{(p)}\right) = E\left[\left(Y_{t-j} - Y_{t-j}^{(p)}\right)\varepsilon_t^{(p)}\right] =$

$= E\left(Y_{t-j}\varepsilon_t^{(p)}\right) - E\left(Y_{t-j}^{(p)}\varepsilon_t^{(p)}\right) = E\left(Y_{t-j}\varepsilon_t^{(p)}\right) - E\left[\left(a_1 Y_{t-j-1} + a_2 Y_{t-j-2} + ... + a_p Y_{t-j-p}\right)\varepsilon_t^{(p)}\right] = 0 \ \ (j > 0)$ (debe tenerse en cuenta que suponer lo contrario, es decir, que $\gamma_j\left(\varepsilon^{(p)}\right) \neq 0$ implicaría reconocer que, al menos en parte, el comportamiento de $\varepsilon_t^{(p)}$ podría ser explicado por $\varepsilon_{t-j}^{(p)}$ ($\varepsilon_t^{(p)} = f\left(\varepsilon_{t-j}^{(p)}\right)$) pero, en ese caso, los coeficientes a_j tampoco cumplirían la condición de minimizar $E\left[\left(\varepsilon_t^{(p)}\right)^2\right]$); **iv)** $E\left[\left(\varepsilon_t^{(p)}\right)^2\right] = \sigma^2\left(\varepsilon_t^{(p)}\right) = \sigma_\varepsilon^2$ y **v)** $\varepsilon_t^{(p)}: N\left(0,\sigma_\varepsilon^2\right)$.

3) Los coeficientes a_j se obtienen de aplicar el criterio de optimización de los cuadrados mínimos ordinarios[82], es decir, de minimizar la expresión:

$$E\left[\left(\varepsilon_t^{(p)}\right)^2\right] = E\left[\left(Y_t - Y_t^{(p)}\right)^2\right] = E\left(\sum_{i=1}^{p}\sum_{j=1}^{k} a_i a_j Y_{t-i}Y_{t-j} - 2\sum_{j=1}^{p} a_j Y_t Y_{t-j} + Y_t^2\right)$$

Luego, suponiendo que la solución exista, los coeficientes $a_i \ (i = 1,2,...,p)$ quedan definidos de la siguiente forma $a = \left[\Gamma_k(Y)\right]^{-1}\gamma(Y)$ donde, $a' = \left[a_1 \ \ a_2 \ \ ... \ \ a_p\right]$; $\gamma'(Y) = \left[\gamma_0(Y) \ \ \gamma_1(Y) \ \ ... \ \ \gamma_{p-1}(Y)\right]$; $\gamma_j(Y) = E\left(Y_{t-j}Y_t\right) =$

$= a_1\gamma_{j-1}(y) + a_2\gamma_{|j-2|}(Y) + \ ... + a_p\gamma_{|j-p|}(Y) \ (j = 0,1,2,...,p)$ y :

---

[81]. Cada innovación $\varepsilon_t^{(p)}$ representa la nueva información que actualiza el valor del proceso para cada valor de t y que, por lo tanto, replantea para cada t nuevas condiciones iniciales y determina la diferencia entre un proceso autorregresivo y una ecuación en-diferencias (ver Sec. 2.10.1).

[82]. Este criterio de optimización será tratado en el Cap. 6.

$$\Gamma_p(Y) = \begin{bmatrix} \gamma_0(Y) & \gamma_1(Y) & \dots & \gamma_{p-1}(Y) \\ \gamma_1(Y) & \gamma_0(Y) & \dots & \gamma_{p-2}(Y) \\ \dots & \dots & \dots & \dots \\ \gamma_{p-1}(Y) & \gamma_{p-2}(Y) & \dots & \gamma_0(Y) \end{bmatrix}$$

denota la **matriz de autocovarianzas de Toeplitz**[83]. La cual, dado que se ha supuesto que el proceso $\{Y_t\}$ es estacionario, es simétrica con respecto a ambas diagonales[84].

Con un razonamiento similar al anterior, se puede escribir:

$$Y_{t-1} = b_2 Y_{t-2} + b_3 Y_{t-3} + \dots + b_p Y_{t-p} + \varepsilon_{t-1}^{(p)} = Y_{t-1}^{(p)} + \varepsilon_{t-1}^{(p)}$$

donde los coeficientes b_j son tales que minimizan la expresión $E\left[\left(Y_{t-1} - Y_{t-1}^{(p)}\right)^2\right]$ y se verifica que

$$E\left[\varepsilon_{t-1}^{(p)}\right] = 0$$

$$E\left[Y_{t-j}\varepsilon_{t-1}^{(p)}\right] = 0 \,(j = 2,3,\dots)$$

$$E\left[\varepsilon_{t-1}^{(p)}\varepsilon_t^{(k)}\right] = E\left[\left(Y_{t-1} - Y_{t-1}^{(p)}\right)\varepsilon_t^{(p)}\right] = E\left[Y_{t-1}\varepsilon_t^{(p)}\right] - E\left[Y_{t-1}^{(p)}\varepsilon_t^{(p)}\right] = 0$$

$$\sigma^2\left(\varepsilon_{t-1}^{(p)}\right) = \sigma_\varepsilon^2 \text{ y } \varepsilon_{t-1}^{(p)} : N\left(0, \sigma_\varepsilon^2\right).$$

A partir de estas resultados se puede escribir:

$$Y_t = \varepsilon_t^{(p)} + a_1\left[Y_{t-1}^{(p)} + \varepsilon_{t-1}^{(p)}\right] + a_2 Y_{t-2} + \dots + a_p Y_{t-p} =$$

$$= \varepsilon_t^{(p)} + a_1\left(b_2 Y_{t-2} + b_3 Y_{t-3} + \dots + b_p Y_{t-p}\right) + a_2 Y_{t-2} + \dots + a_p Y_{t-p} =$$

$$= \varepsilon_t^{(p)} + a_1\varepsilon_{t-1}^{(p)} + a_2 Y_{t-2} + a_3 Y_{t-3} + \dots + a_p Y_{t-p} =$$

$$= \varepsilon_t^{(p)} + a_1\varepsilon_{t-1}^{(p)} + Y_{(t-2)}^{(p)} = \eta_{t-1}^{(p)} + Y_{(t-2)}^{(p)}$$

Luego, operando en forma iterativa, se concluye que[85]:

$$Y_t = \varepsilon_t^{(p)} + b_1^*\varepsilon_{t-1}^{(p)} + b_2^*\varepsilon_{t-2}^{(p)} + \dots + b_j^*\varepsilon_{t-j}^{(p)} + Y_{(t-j-1)}^{(p)} = \eta_{t-j}^{(p)} + Y_{(t-j-1)}^{(p)}$$

---

[83]. Ver Grenander, U.; Szego, G. (1958).

[84]. Debe tenerse en cuenta que $Y_t^{(p)}$ y $\varepsilon_t^{(p)}$ están definidos en forma unívoca, pero los coeficientes a, en algunos casos, pueden no cumplir esta condición de unicidad (ver Cramér, H. (1946)).

[85]. Las condiciones de linealidad asociadas al teorema de Wold dependen de los supuestos que se realicen sobre las características del proceso $\left\{\varepsilon_t^{(p)}\right\}$.

donde $Y_{t-j}^{(p)} = a_j Y_{t-j} + a_{j+1} Y_{t-j-1} + ... + a_p Y_{t-p}$; $\gamma\left[\varepsilon_{t-h}^{(p)}, Y_{t-j-1}^{(p)}\right] = 0\ (h \le j)$; $\left\{\varepsilon_t^{(p)}\right\} : N\left(0, \sigma_\varepsilon^2\right)$; $E\left\{\left[Y_{t-i}^{(p)}\right]^2\right\} \ge E\left\{\left[Y_{t-j}^{(p)}\right]^2\right\}$ $(i > j; j = 0,1,2,...)$ y $b_j = \dfrac{E\left[Y_t \varepsilon_{t-j}^{(p)}\right]}{E\left\{\left[\varepsilon_{t-j}^{(p)}\right]^2\right\}}$.

Como corolario de estas condiciones, se demuestra que $E\left(\eta_{t-j}^{(p)} Y_{t-j-1}^{(p)}\right) = 0\ (j = 0,1,2,..., p-1)$.

Suponiendo que los procesos $\left\{\eta_{t-j}^{(p)}\right\}$ y $\left\{Y_{(t-j-1)}^{(p)}\right\}$ convergen en valor medio cuadrático, respectivamente, a los procesos $\{\eta_t\}$ y $\{Y_{(t)}\}$[86], se puede escribir $Y_t = \eta_t + Y_{(t)}$. De modo que el proceso $\{Y_t\}$ queda expresado como la suma de dos procesos estacionarios incorrelacionados de valor esperado nulo: **i)** el proceso $\{\eta_t\}$ ergódico, definido como una combinación lineal de innovaciones puramente aleatorias, y **ii)** el proceso $\{Y_{(t)}\}$, linealmente determinístico, que representa el pasado remoto del proceso $\{Y_t\}$.

Sea $\sigma_\eta^2(p)$ una función no-decreciente de p que define la varianza de la combinación lineal de las innovaciones. Obviamente, se verifica que $\sigma_Y^2 \ge \sigma_\eta^2(p) \ge 0$. La sucesión $\{\sigma_\eta^2(1), \sigma_\eta^2(2), ...\}$ debe tender a un límite σ_η^2 . Si este límite es igual a σ_Y^2, significa que la combinación lineal de las realizaciones del infinito pasado remoto del proceso no posee ninguna capacidad explicativa (predictiva) del comportamiento del proceso $\{Y_t\}$, es decir que, como se vio en la Sec. 1.3, el proceso es **no-autorregresivo**. Si se verifica que el límite de la sucesión de varianzas, $\sigma_\eta^2(p)$, es igual a cero, se puede concluir que: **i)** el comportamiento de $\{Y_t\}$ podría ser explicado en forma determinística, a partir del conocimiento de su infinito pasado y **ii)** cualquier conjunto finito de realizaciones del proceso poseería las mismas cualidades explicativas del comportamiento de $\{Y_t\}$, no importa cuán alejado en el tiempo se encontrara (condición de estacionariedad).

1.7.- Acerca de la definición estricta de algunos procesos que son de aparición frecuente en las aplicaciones financieras, biométricas y actuariales

1.7.1.- Las cadenas de Markov

1.7.1.1.- Definición

Se dice que $\{Y(t)\}\ (t \in R_1)$ es un **proceso de Markov** si, conocido el estado actual del proceso, la expectativa sobre su comportamiento futuro no se ve afectada por el conocimiento de su comportamiento pasado, es decir si, conocido su estado al momento t_m (y_m), su comportamiento para todo $t > t_m$ es independiente de los estados asumidos por el proceso para todo $t < t_m$. De modo que sus **probabilidades de transición simple** (o **en un paso**) quedan definidas de la siguiente forma:

---

[86]. Desde un punto de vista estrictamente riguroso, esta convergencia debe ser demostrada (ver Wold, H. (1938), Anderson, T.W. (1971)).

$$p\Big(\big[y_a \le Y(t_n) \le y_b\big] / \Big\{\big[Y(t_m = y_m)\big] \cap \big[Y(t_\ell) = y_\ell\big] \cap ...\Big\}\Big) =$$

$$= p\Big\{\big[y_a \le Y(t_n) \le y_b\big] / \big[Y(t_m) = y_m\big]\Big\} =$$

$$= p\big[(y_m, t_m), (A, t_n)\big] \qquad \big(t_n > t_m > t_\ell > ...; A = \{y / y_a \le Y \le y_b\}\big)$$

A partir de esta definición, se puede concluir, inmediatamente, que sus **probabilidades de transición en dos pasos** (probabilidades de que un proceso que se encuentra en un **estado** y_ℓ, en el momento t_ℓ, se encuentre, al cabo de dos transiciones, en el estado y_n, en el momento t_n siendo $t_n > t_m > t_\ell$) satisfacen la **condición de Kolmogorov-Chapman**[87]:

$$p\Big\{\big[Y(t_n) = A\big] / \big[Y(t_\ell) = y_\ell\big]\Big\} =$$

$$= \int_{-\infty}^{\infty} p\Big\{\big[Y(t_m) = z\big] / \big[Y(t_\ell) = y_\ell\big]\Big\} p\Big\{\big[Y(t_n) = A\big] / \big[Y(t_m) = z\big]\Big\} dz \qquad \big(A = \{y / y_a \le Y \le y_b\}\big)$$

En particular, todo proceso de Markov discreto en el dominio de las variables (infinito numerable o finito) y discreto en el dominio del tiempo, define una **cadena de Markov**[88]. Luego, las **probabilidades de transición simple** de que un proceso, que se encuentra en el estado i (que asumió el valor $i = 0,1,2,...$) en el momento $t-1$, pase al estado j (asuma el valor j) ($j = 0,1,2,...$) en el momento t, para una cadena de Markov, pueden ser denotadas de la siguiente forma:

$$p\Big\{\big(Y_t = j\big) / \big(Y_{t-1} = i\big) \cap \big(Y_{t-2} = h\big) \cap ...\Big\} = p\Big[\big(Y_t = j\big) / \big(Y_{t-1} = i\big)\Big] =$$

$$= p\big[(i, t-1), (j, t)\big] = p_{ij}(t-1, t) \qquad \big(i, j = 0,1,2,...; t = 1,2,...\big)$$

Estas probabilidades, ordenadas según sus **estados de partida** y **de llegada**, definen una **matriz estocástica** de la forma[89]:

---

87. Ver Sec. 1.3.

88. De acuerdo con la teoría de las fluctuaciones en estadística molecular de von Smoluchowski, M. (1906), la forma más inmediata de obtener un proceso de Markov (utilizada por Markov, A.A. (1908)) es a partir de una interpretación constructivista, mediante la suma de variables aleatorias independientes. El proceso surgido en el ámbito de esta teoría de las fluctuaciones representaba el número de partículas Brownianas contenidas en un espacio dado, obtenido de observaciones realizadas a intervalos regulares y constituye el primer ejemplo de lo que, desde los comienzos de la década de 1930, se denominó cadena de Markov. Este proceso observó dos corrientes de desarrollo: La primera, liderada por von Mises, R. (1920)(1921)(1921)(1928) (1930)(1931)(1933)1934) y Kolmogorov, A.N. (1931), basándose en una hipótesis de ergodicidad formulada dinámicamente, condujo a una formulación puramente probabilística. La segunda proviene de las propuestas de Gibbs, J.W. (1912) sobre la mezcla de líquidos, reformulada por Poincaré, H. (1912) en términos discretos. Los trabajos de Kolmogorov fueron continuados por Feller, W. (1936).

89. Una matriz estocástica se caracteriza porque todos sus elementos son no-negativos y la suma de los elementos de cada fila es igual a la unidad, $\sum_{j=0}^{\infty} p_{ij}(t-1, t) = 1 \; (i = 0,1,2,...)$.

$$P(t-1,t)=\begin{bmatrix} p_{00}(t-1,t) & p_{01}(t-1,t) & p_{02}(t-1,t) & \dots \\ p_{10}(t-1,t) & p_{11}(t-1,t) & p_{12}(t-1,t) & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

La condición de Kolmogorov-Chapman para las probabilidades de transición en dos pasos puede ser expresada, en este caso, de la siguiente forma:

$$p\left\{Y_t = j / \left[(Y_{t-2} = i) \cap (Y_{t-3} = h) \cap \dots\right]\right\} = p\left[(Y_t = j) / (Y_{t-2} = i)\right] = p_{ij}(t-2,t) =$$

$$= \sum_{h=0}^{\infty} p_{ih}(t-2,t-1)p_{hj}(t-1,t)$$

De modo que la **matriz de probabilidades de transición en dos pasos** queda definida como:

$$P(t-2,t) = \left[p_{ij}(t-2,t); i,j = 0,1,2,\dots\right] = \left[\sum_{h=0}^{\infty} p_{ih}(t-2,t-1)p_{hj}(t-1,t)\right] = P(t-2,t-1)P(t-1,t)$$

(en particular, si el dominio de las variables es finito, $\{0,1,2,\dots,k\}$, la matriz P es de orden $(k+1)$).

De acuerdo con la condición de Kolmogorov-Chapman, operando en forma iterativa, las **probabilidades de transición en n pasos** pueden ser expresadas de la siguiente forma:

$$p\left\{(Y_t = j) / (Y_{t-n} = i) \cap (Y_{t-n-1} = h) \cap \dots\right\} = p\left[(Y_t = j) / (Y_{t-n} = i)\right] = p_{ij}(t-n,t) =$$

$$= \sum_{h=0}^{\infty} p_{ih}(t-n,t-1)p_{hj}(t-1,t) \qquad (i,j = 0,1,2,\dots; n = 1,2,\dots; t = n,n+1,\dots)$$

Luego, la matriz de probabilidades de transición en n $(= 1,2,3,\dots)$ pasos, queda definida como:

$$P(t-n,t) = \left[p_{ij}(t-n,t); i,j = 0,1,2,\dots\right] = P(t-n,t-n+1)P(t-n+1,t-n+2)\dots P(t-1,t)$$

Existe una correspondencia biunívoca entre la familia de matrices de probabilidades de transición $\{P(t-n,t), n = 1,2,\dots\}$ y la cadena de Markov que le dio origen, de modo que, dada una cadena de Markov, queda definida una, y sólo una familia de matrices $\{P(t-n,t), n = 1,2,\dots\}$ y, viceversa, dada una familia de matrices de probabilidades de transición, queda definida una, y sólo una cadena de Markov.

Si las probabilidades de transición dependen de n (es decir, del número de pasos), pero son independientes de t, es decir, si se verifica que:

$$p\left[(Y_t = j) / (Y_{t-n} = i)\right] = p\left[(Y_n = j) / (Y_0 = i)\right] = p_{ij}^{(n)} \qquad (i,j = 0,1,2,\dots; t,n = 1,2,\dots)$$

se dice que la cadena es **homogénea** o **con mecanismo de transición estacionario.** En este caso se verifica que la distribución de probabilidades conjunta para cualquier sucesión finita de sus variables aleatorias puede

ser expresada como el producto de la probabilidad inicial por las probabilidades de transición simple[90]:

$$p\left[\left(Y_{t-n} = h\right)\cap\left(Y_{t-n+1} = i\right)\cap\left(Y_{t-n+2} = j\right)\cap\ldots\cap\left(Y_t = \ell\right)\right] = p_h^{(0)} p_{hi} p_{ij} \ldots p_{k\ell} \quad (h,i,j,k,\ell = 0,1,2\ldots)$$

(donde $p_h^{(0)}$ denota la **probabilidad inicial** del proceso, es decir, la probabilidad de que el proceso elija a $h\,(= 0,1,2,\ldots)$ como estado inicial)[91] y, por lo tanto, la matriz de probabilidades de transición en n pasos queda definida por[92]:

$$P^{(n)} = \left[p_{ij}^{(n)}; i,j = 0,1,2,\ldots\right] = \left[\sum_{h=0}^{\infty} p_{ih}^{(n-1)} p_{hj}\right] = P^n$$

Debe tenerse en cuenta que $p_{ij}^{(n)} \leq p_{ij}^{(n-1)} \leq \ldots \leq \sum_{j=0}^{\infty} p_{ij} = 1$, lo que demuestra que la sucesión $p_{ij}^{(n)}$ $(n = 2,3,\ldots)$ es convergente. Como corolario de este resultado se puede concluir que la probabilidad no-condicionada $p_j^{(n)} = p\left(Y_n = j\right)$ puede ser expresada de la siguiente forma:

$$p_j^{(n)} = p\left(Y_t = j\right) = \sum_{i=0}^{\infty} p\left[\left(Y_t = j\right)/\left(Y_{t-n} = i\right)\right] p\left(Y_{t-n} = i\right) = \sum_{i=0}^{\infty} p_i p_{ij}^{(n)}$$

Supóngase, en particular, una cadena de Markov homogénea con un dominio de los estados $\{0,1\}$, con matriz de probabilidades de transición de la forma[93]:

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - p_{01} & p_{01} \\ p_{10} & 1 - p_{10} \end{bmatrix} \qquad \left(0 < p_{01} < 1; 0 < p_{10} < 1\right)$$

Si se considera que inicialmente el proceso se encuentra en el estado 0, es decir que el vector de las probabilidades iniciales es de la forma $P^{(0)} = \left[p_0^{(0)} \quad p_1^{(0)}\right] = \left[1 \quad 0\right]$, se verifica que:

$$p_{00}^{(n)} = p\left[\left(Y_n = 0\right)/\left(Y_0 = 0\right)\right] = p\left(Y_n = 0\right) =$$

---

90. Debe tenerse en cuenta que la propiedad de estacionariedad en el mecanismo de transición no implica la estacionariedad estricta del proceso. Para que una cadena de Markov homogénea sea estrictamente estacionaria debe verificarse que su distribución de probabilidades no-condicionada debe ser estacionaria (ver Sec. 1.4.1). Las condiciones para el cumplimiento de esta propiedad serán tratadas en la Sec. 1.7.1.4.

91. Como se verá más adelante, en general, las probabilidades iniciales no son relevantes en el estudio de las propiedades de las cadenas.

92. Nótese que, en el caso de las cadenas homogéneas, toda la información necesaria para definir el proceso está contenida en la matriz $P = \left[p_{ij}; i,j = 0,1,2,\ldots\right]$. Definidas las probabilidades p_{ij} $\left(i,j = 0,1,2,\ldots\right)$ y la distribución de probabilidades inicial, $p\left(E_0 = i\right) \equiv p_i^{(0)}$ $\left(i = 0,1,2,\ldots\right)$ (probabilidad de que la cadena se encuentre en el estado i antes de que se produzca la primera transición), la cadena de Markov queda definida en forma unívoca.

93. Como se verá en la Sec. 1.7.1.3, esta matriz de probabilidades de transición corresponde a una **cadena no-reducible**.

$$= p\left[\left(Y_n = 0\right) \cap \left(Y_{n-1} = 0\right)\right] + p\left[\left(Y_n = 0\right) \cap \left(Y_{n-1} = 1\right)\right] =$$

$$= p\left[\left(Y_n = 0\right) / \left(Y_{n-1} = 0\right)\right] p\left(\left(Y_{n-1} = 0\right) + p\left[\left(Y_n = 0\right) / \left(Y_{n-1} = 1\right)\right] p\left(Y_{n-1} = 1\right) =$$

$$= p_{00} p\left(Y_{n-1} = 0\right) + p_{10} p\left(Y_{n-1} = 1\right) =$$

$$= \left(1 - p_{01}\right) p\left(Y_{n-1} = 0\right) + p_{10}\left[1 - p\left(Y_{n-1} = 0\right)\right] =$$

$$= \left(1 - p_{01} - p_{10}\right) p\left(Y_{n-1} = 0\right) + p_{10}$$

Operando en forma iterativa, se obtiene que:

$$p_{00}^{(n)} = p_{10} + \left(1 - p_{01} - p_{10}\right)\left[p_{10} + \left(1 - p_{01} - p_{10}\right)\right] p\left(Y_{n-2} = 0\right) =$$

$$= p_{01} + p_{10}\left(1 - p_{01} - p_{10}\right) + \left(1 - p_{01} - p_{10}\right)^2 p\left(Y_{n-2} = 0\right) =$$

$$= p_{10} + p_{10}\left(1 - p_{01} - p_{10}\right) + \left(1 - p_{01} - p_{10}\right)^2 \left[p_{10} + \left(1 - p_{01} - p_{10}\right) p\left(Y_{n-3} = 0\right)\right] =$$

$$= \ldots = p_{10} \sum_{j=0}^{n-1} \left(1 - p_{01} - p_{10}\right)^j + \left(1 - p_{01} - p_{10}\right)^n =$$

$$= \left(1 - p_{01} - p_{10}\right)^n + p_{10} \frac{1 - \left(1 - p_{01} - p_{10}\right)^n}{p_{01} + p_{10}} =$$

$$= \frac{p_{10}}{p_{01} + p_{10}} + \left(1 - p_{01} - p_{10}\right)^n \frac{p_{01}}{p_{01} + p_{10}} \qquad (n = 1,2,\ldots)$$

Teniendo en cuenta que $p_{00}^{(n)} + p_{01}^{(n)} = 1$, se obtiene que:

$$p_{01}^{(n)} = \frac{p_{01}}{p_{01} + p_{10}}\left[1 - \left(1 - p_{01} - p_{10}\right)^n\right] \qquad (n = 1,2,\ldots)$$

De la misma forma, suponiendo que el vector de las probabilidades iniciales sea de la forma $P^{(0)} = \begin{bmatrix} 0 & 1 \end{bmatrix}$, se obtiene que:

$$p_{11}^{(n)} = \frac{p_{01}}{p_{01} + p_{10}} + \left(1 - p_{01} - p_{10}\right)^n \frac{p_{10}}{p_{01} + p_{10}} \qquad (n = 1,2,\ldots)$$

$$p_{10}^{(n)} = 1 - p_{11}^{(n)} = \frac{p_{10}}{p_{01} + p_{10}}\left[1 - \left(1 - p_{01} - p_{10}\right)^n\right] \qquad (n = 1,2,\ldots)$$

Luego, se puede concluir que:

$$P^{(n)} = \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{bmatrix} = \frac{1}{p_{01} + p_{10}} \left\{ \begin{bmatrix} p_{10} & p_{01} \\ p_{10} & p_{01} \end{bmatrix} + \left(1 - p_{01} - p_{10}\right)^n \begin{bmatrix} p_{01} & -p_{01} \\ -p_{10} & p_{10} \end{bmatrix} \right\} \qquad (n = 1,2,\ldots)$$

Teniendo en cuenta que $0 < 1 - p_{01} < 1$ y que $\left|1 - p_{01} - p_{10}\right| < 1$, se demuestra en forma inmediata que, a medida que n aumenta indefinidamente, existe una matriz con columnas constantes como límite de $P^{(n)}$:

$$\lim_{n\to\infty} P^{(n)} = \frac{1}{p_{01}+p_{10}}\begin{bmatrix} p_{10} & p_{01} \\ p_{10} & p_{01} \end{bmatrix}$$

(obsérvese que las probabilidades $p_{ij}^{(n)}\ (i,j = 0,1,2,...)$ convergen en forma geométrica, $\sup_{i,j}\left|p_{ij}^{(n)} - \pi_j\right| \leq$ $\leq c(1-p_{01}-p_{10})^n$, donde $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$).

Como se verá en la Sec. 1.7.1.4, la existencia de esta matriz límite está vinculada con la estacionariedad de la distribución del proceso

Suponiendo que el vector de las probabilidades iniciales sea de la forma $P^{(0)} = \left[p_0^{(0)} \quad p_1^{(0)}\right]$, con un razonamiento similar al desarrollado precedentemente, se obtiene que:

$$p_{00}^{(n)} = \frac{p_{10}}{p_{01}+p_{10}} + (1-p_{01}-p_{10})^n\left(p_0^{(0)} - \frac{p_{10}}{p_{01}+p_{10}}\right) \qquad (n = 1,2,...)$$

$$p_{01}^{(n)} = \frac{p_{01}}{p_{01}+p_{10}} - (1-p_{01}-p_{10})^n\left(p_0^{(0)} - \frac{p_{10}}{p_{01}+p_{10}}\right) \qquad (n = 1,2....)$$

$$p_{11}^{(n)} = \frac{p_{01}}{p_{01}+p_{10}} + (1-p_{01}-p_{10})^n\left(p_1^{(0)} - \frac{p_{01}}{p_{01}+p_{10}}\right) \qquad (n = 1,2,...)$$

$$p_{10}^{(n)} = \frac{p_{10}}{p_{01}+p_{10}} + (1-p_{01}-p_{10})^n\left(p_1^{(0)} - \frac{p_{01}}{p_{01}+p_{10}}\right) \qquad (n = 1,2....)$$

Sea una cadena de Markov homogénea, finita, con matriz de probabilidades de transición de la forma:

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0k} \\ p_{10} & p_{11} & \cdots & p_{1k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k0} & p_{k1} & \cdots & p_{kk} \end{bmatrix}$$

(donde $p_{ij} > 0$ para $i,j = 0,1,2,...,k$) y sea $\{y_0,y_1,...,y_n\}$ una muestra de tamaño $n+1$. La función de verosimilitud de esta muestra será de la forma:

$$L = p_{y_0}^{(0)}\prod_{j=1}^{n} p_{y_{j-1}y_j} = p_{y_0}^{(0)}\prod_{i=1}^{k}\prod_{j=1}^{k} p_{ij}^{n_{ij}}$$

(donde n_{ij} denota la frecuencia en la muestra de la transición en un paso del estado i al estado j). El logaritmo de la función de verosimilitud será, entonces, de la forma $\ell = \log\left(p_{y_0}^{(0)}\right) + \sum_i\sum_j \log\left(p_{ij}\right)$. Maximizando esta función con respecto a p_{ij} sujeta a la restricción $\sum_{j=1}^{k} p_{ij} = 1$, se obtienen los estimadores:

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{k} n_{ij}} = \frac{n_{ij}}{n_i} \left(i, j = 0,1,2,...,k\right)$$

(debe tenerse en cuenta que, en este caso, se ha supuesto que la distribución de las probabilidades iniciales $\left\{p_i^{(0)}\right\}$ no proporciona ninguna información acerca de las probabilidades de transición)[94].

1.7.1.2.- Clasificación de los estados

Los estados de una cadena pueden ser clasificados, según las características de las probabilidades de transición que los involucran, de la siguiente forma:

1) Se dice que el estado j es **efímero**, si se verifica que $p_{ij} = 0 \left(i = 0,1,2,...\right)$. Una cadena con un estado efímero se caracteriza porque en su matriz de probabilidades de transición, la columna correspondiente a dicho estado está compuesta exclusivamente por ceros[95]:

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0,j-1} & 0 & p_{0,j+1} & \cdots \\ p_{10} & p_{11} & \cdots & p_{1,j-1} & 0 & p_{1,j+1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

2) Se dice que el estado j **está comunicado** con el estado $i \left(i \rightarrow j\right)$ o que el estado j **es accesible** desde el estado i o que el estado i **conduce** al estado j o que el estado j es un **consecuente** del estado i, si existe un valor entero $n > 0$, tal que $p_{ij}^{(n)} \neq 0 \left(i, j = 0,1,2,...; i \neq j\right)$.

3) Se dice que los estados i y j **están doblemente comunicados** o **comunicados en ambos sentidos** o que entre los estados i y j existe una **relación simétrica** ($i \leftrightarrow j$), si existen dos valores de n : $n_1 > 0$ y $n_2 > 0$, tales que $p_{ij}^{(n_1)} \neq 0$ y $p_{ji}^{(n_2)} \neq 0 \left(i, j = 0,1,2,...; i \neq j\right)$. A partir de la definición, se concluye en forma inmediata que la relación de doble comunicación cumple las propiedades idéntica ($i \leftrightarrow i$) y recíproca (si $i \leftrightarrow j$, entonces, $j \leftrightarrow i$). Además, dados tres estados, i, j y k, tales que $i \leftrightarrow j$ y $j \leftrightarrow k$, es posible hallar dos valores, $n_1 \geq 0$ y $n_2 \geq 0$, tales que $p_{ik}^{(n_1+n_2)} = \sum_{h=0}^{\infty} p_{ih}^{(n_1)} p_{hk}^{(n_2)} \geq p_{ij}^{(n_1)} p_{jk}^{(n_2)} > 0$, lo cual demuestra que también cumple la propiedad de transitividad (si $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces, $i \leftrightarrow k$).

4) Se dice que el estado j es **esencial** si está comunicado en ambos sentidos con todos los estados desde los

---

[94]. Dado que esta sección está referida fundamentalmente a la definición estricta de ciertos procesos, el problema de la estimación de los coeficientes será tratado sólo en forma introductoria. Para un análisis más detallado sobre el particular, ver Billingsley, P. (1961)(1961), Martin, J.I. (1967), Lee, T.C.; Judge, G.G.; Zellner, A. (1970), Basawa, I.V. (1972), Toe, D.T. (1976).

[95]. Una cadena sólo puede encontrarse en un estado efímero en el instante inicial.

cuales es accesible; es decir, si la relación $j \to k$ implica la relación $k \to j$, para $k = 0,1,2,...,k$.

5) Se dice que j es un estado **periódico de período** T_j, si se verifica que T_j es el máximo común divisor del conjunto $n = 1,2,...$ para el cual $p_{jj}^{(n)} > 0$ ($T_j = MCD\left\{n \geq 1: p_{jj}^{(n)} > 0\right\}$). Esta condición implica que, si $p_{jj}^{(n)} > 0$, entonces n es un entero múltiplo de T_j y T_j es el mayor entero que posee esta propiedad. Obviamente las recurrencias al estado j sólo se podrán verificar en períodos múltiplos de T_j. Si se verifica que $\left\{n > 1: p_{jj}^{(n)} > 0\right\} = \varnothing$, entonces $T_j = 1$, lo cual implica que el estado j es **no-periódico**.

6) Dado un estado j, el conjunto de estados de la cadena comunicados con él define su **clase comunicante**, $C(j)$ [96]. Si, dada una clase $C \neq \varnothing$, se verifica que, a partir de un estado cualquiera de C, es imposible alcanzar un estado externo a la misma $p_{ij}^{(n)} = 0\ (i \in C; j \notin C; n = 1,2,...)$, se dice que la clase C es **cerrada**. En este caso, la matriz de probabilidades de transición $P_C = \left[p_{ij}; i \in C, j \in C\right]$ es una matriz estocástica. Debe tenerse en cuenta que en una cadena de Markov puede haber una cantidad infinita de clases cerradas y que estas no tienen por qué ser necesariamente disjuntas.

7) Dada una clase comunicante cerrada, C, se denomina **clausura** al menor conjunto de estados que contiene dicha clase.

8) Si una clausura está formada solamente por un estado j, se dice que ese estado es **absorbente**. En este caso, una vez que la cadena accede al mismo, no lo puede abandonar más ($p_{jj} = 1$)[97].

Como corolario de las dos últimas definiciones, se puede concluir que la probabilidad de absorción por una clausura C está dada por la suma de las probabilidades de absorción de cada uno de sus estados[98].

1.7.1.3.- Probabilidades de recurrencia

Sea $f_{ij}^{(n)}$ la probabilidad de que una cadena, que se encuentra en el estado i en el instante inicial, **arribe por primera vez al estado j al cabo de n transiciones**:

---

[96]. Si j es un estado efímero, su clase comunicante es vacía.

[97]. Supóngase una cadena de Markov con un dominio de los estados definido por el conjunto $\{0,1,2,3\}$, en el cual los estados 0 y 3 son absorbentes, es decir, tales que $p_{00} = 1$, $p_{0j} = 0\ (j = 1,2,3)$, $p_{33} = 1$, $p_{3j} = 0\ (j = 0,1,2)$. Entonces el conjunto $\{1,2\}$ constituye una clase, pero no una clase cerrada.

[98]. Esta clasificación de los estados de una cadena no es exhaustiva y será completada en la próxima sección al considerar la frecuencia de retorno del proceso a cada estado.

$$f_{ij}^{(n)} = p\left\{\left[(E_1 \neq j)\cap(E_2 \neq j)\cap\ldots\cap(E_{n-1} \neq j)\cap(E_n = j)\right]/(E_0 = i)\right\} =$$

$$= \sum_{\substack{h_\ell \neq j \\ \ell=1,2,\ldots,n-1}} p_{ih_1}p_{h_1h_2}\cdots p_{h_{n-1}j} \qquad (i,j = 0,1,2,\ldots)$$

Entonces, la probabilidad de que una cadena, que se encuentra en el estado i en el momento inicial, **arribe alguna vez al estado** j, queda definida por:

$$f_{ij} = p\left[\left(E_n = j, para\ algun\ n > 0\right)/(E_0 = i)\right] = p\left[\left(\bigcup_{n=1}^{\infty} E_n = j\right)/(E_0 = i)\right] = \sum_{n=1}^{\infty} f_{ij}^{(n)} \qquad (i,j = 0,1,2,\ldots)$$

De la misma forma, $f_{jj}^{(n)}$ representa la probabilidad de que una cadena, que se encuentra en el estado j en el momento inicial, **retorne por primera vez a dicho estado al cabo de** n **transiciones (probabilidad de recurrencia al estado** j **al cabo de n transiciones)**, y $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$ $(j = 0,1,2,\ldots)$ define la probabilidad de que una cadena, que se encuentra en el estado j en el momento inicial, **retorne alguna vez a dicho estado (probabilidad de recurrencia al estado** j**).**

Para un proceso que se encuentra inicialmente en el estado i, la relación entre la probabilidad de arribar al estado j por primera vez en n pasos, y la probabilidad de encontrarse en el estado j al cabo de n pasos, puede ser expresada de la siguiente forma:

$$p_{ij}^{(n)} = f_{ij}^{(1)}p_{jj}^{(n-1)} + f_{ij}^{(2)}p_{jj}^{(n-2)} + \ldots + f_{ij}^{(n-1)}p_{jj} + f_{ij}^{(n)} \qquad (i \neq j; i,j = 0,1,2,\ldots; n = 1,2,\ldots)$$

Con un razonamiento similar se obtiene que la relación entre la probabilidad de retornar a un estado j por primera vez en n pasos y la probabilidad de encontrarse nuevamente en un estado j al cabo de n pasos, puede ser expresada de la siguiente forma:

$$p_{jj}^{(n)} = f_{jj}^{(1)}p_{jj}^{(n-1)} + f_{jj}^{(2)}p_{jj}^{(n-2)} + \ldots + f_{jj}^{(n-1)}p_{jj} + f_{jj}^{(n)} \qquad (j = 0,1,2,\ldots; n = 1,2,\ldots)$$

Si se verifica que $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)} = 1$, se dice que el estado j es **recurrente** (los estados recurrentes también suelen ser denominados **persistentes**, ver Feller, W. (1957)). Es decir que la condición necesaria y suficiente para que el estado j sea recurrente es que $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$. En este caso, n_j puede ser considerada como una variable aleatoria con función de probabilidades $f_{jj}^{(n)}$. De modo que $E(n_j) = \sum_n n f_{jj}^{(n)}$ define el denominado **tiempo esperado de recurrencia al estado** j. Es decir, el tiempo que, dada la distribución de probabilidades $f_{jj}^{(n)}$, se debe esperar que tarde el proceso en retornar al estado j. Si $E(n_j)$ es finito, se dice que j es un estado **recurrente-positivo**, si $E(n_j) = \infty$, se dice que es **recurrente-nulo**.

Si $f_{jj} < 1$, es decir, si existe una probabilidad no-nula de que la cadena no retorne al estado j

($1 - f_{jj} > 0$), entonces se dice que j es un estado **no-recurrente**. La condición necesaria y suficiente para que el estado j sea no-recurrente es que $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$ [99].

Dados dos estados i y $j\,(i,j = 0.1.2....;i \neq j)$ doblemente comunicados, se verifica que, si i es recurrente, entonces j también lo es. Luego, si $i \leftrightarrow j$, entonces la condición necesaria y suficiente para que i sea no-recurrente es que j sea no-recurrente. De la misma forma, dados dos estados i y j $(i,j = 0,1,2,...;i \neq j)$ tales que $i \rightarrow j$, si el estado i es recurrente, entonces j también es recurrente[100].

A partir de la sucesión de probabilidades de recurrencia al estado j en n pasos ($\left\{f_{jj}^{(n)}\right\}$) y de la sucesión $p_{jj}^{(n)}$, es posible construir las siguientes funciones generatrices:

$$G_j^{(f)}(w) = \sum_{n=0}^{\infty} f_{jj}^{(n)} w^n \qquad (0 < w < 1)$$

$$G_j^{(p)}(w) = \sum_{n=0}^{\infty} p_{jj}^{(n)} w^n \qquad (0 < w < 1)$$

(obsérvese que $G_j^{(p)}(w)$ no puede ser considerada una función generatriz de probabilidades). De acuerdo con las relaciones entre las probabilidades $f_{jj}^{(n)}$ y $p_{jj}^{(n)}$, mencionadas en la página precedente, se demuestra fácilmente que $G_j^{(p)}(w) = \dfrac{1}{1 - G_j^{(f)}(w)}$ (ver Resnick, S. (1994).

Como corolario de este resultado se puede concluir, entonces, que la condición necesaria y suficiente para que el estado j sea recurrente es que:

$$\lim_{w \to 1} G_j^{(p)}(w) = \lim_{w \to 1} \frac{1}{1 - G_j^{(f)}(w)} = \infty$$

(debe tenerse en cuenta que, si j es recurrente, $G_j^{(f)}(1) = 1$).

Dados dos estados i y j $(i,j = 0,1,2,...;i \neq j)$, tales que $i \leftrightarrow j$, si i es recurrente con período T_i, entonces se verifica que j es recurrente con período $T_j = T_i$.

Se dice que una cadena de Markov es **no-reducible** cuando no se la puede descomponer en **clases comunicantes**. Es decir, cuando todos sus estados son doblemente comunicados ($f_{ij} > 0$, $i,j = 0,1,2,...$). En

---

[99]. De acuerdo con estas definiciones, se puede concluir que todo estado recurrente es esencial y que todo estado no-esencial es no-recurrente. Pero, debe tenerse en cuenta que un estado esencial puede ser no-recurrente (ver Kolmogorov, A.N. (1936)).

[100]. La condición de recurrencia (y, obviamente, la de no-recurrencia) pertenece a la clase de las llamadas propiedades solidarias, las cuales son tales que, dada una clase C caracterizada por una propiedad de este tipo, si el estado $i \in C$ posee esa propiedad, entonces todo estado $j \in C$ posee la misma propiedad,

otros términos, cuando toda la cadena constituye una clase comunicante[102].

Por el contrario, si la cadena es **reducible**, su dominio de los estados (S) puede ser descompuesto de la siguiente forma $S = \Im \cup C_1 \cup C_2 \cup \ldots$, donde $\Im$ denota el conjunto de los estados no-recurrentes[103] y $C_1, C_2, \ldots$ son clases cerradas disjuntas de estados recurrentes y, por lo tanto, tales que, dado un estado $i \in C_\alpha$ $(\alpha = 1,2,\ldots)$, entonces:

$$f_{ij} = \begin{cases} 1 & si \quad j \in C_\alpha \\ 0 & si \quad j \notin C_\alpha \end{cases}$$

Reordenando los estados del proceso de modo que C_1 contenga los estados con menores subíndices y C_2 los siguientes y, así sucesivamente, la matriz de probabilidades de transición puede ser redefinida de la siguiente forma:

$$P = \begin{bmatrix} P_1 & 0 & 0 & \ldots & 0 \\ 0 & P_2 & 0 & \ldots & 0 \\ 0 & 0 & P_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Q_1 & Q_2 & Q_3 & \ldots & \ldots \end{bmatrix}$$

donde: **i)** P_i $(i = 1,2,\ldots)$ denota la matriz cuadrada formada por las probabilidades de transición entre los estados pertenecientes a C_i $(i = 1,2,\ldots)$ y **ii)** Q_i $(i = 1,2,\ldots)$ denota la matriz de probabilidades de transición entre los estados pertenecientes al conjunto $\Im$.

Supóngase que el dominio de los estados sea finito, $S = \{1,2,\ldots,k\}$, y que todos los estados sean no-recurrentes, entonces se verificará que $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$, para todo i y, en consecuencia, que $\lim_{n\to\infty} p_{ij}^{(n)} = 0$. Pero, estos supuestos conducen a la siguiente paradoja $1 = \sum_{j=1}^{k} p_{ij}^{(n)} \to 0$. Luego, se puede concluir que, dada una cadena con dominio de los estados finito, no todos los estados pueden ser no-recurrentes,

---

De la misma forma en que se ha definido la probabilidad de recurrencia por primera vez a un estado j, es posible definir la **probabilidad de recurrencia por segunda vez** a dicho estado, es decir, la probabilidad de que una cadena, que se encuentra en un estado recurrente, j en el instante inicial, retorne a dicho estado por segunda vez, al cabo de n transiciones:

$$f_{jj}^{(2,n)} = f_{jj}^{(1)} f_{jj}^{(n-1)} + f_{jj}^{(2)} f_{jj}^{(n-2)} + \ldots + f_{jj}^{(n-1)} f_{jj}^{(1)} = \left\{ f_{jj}^{(n)} \right\}^{*2} \qquad (n = 1,2,3,\ldots)$$

---

[102]. Luego, de acuerdo con la propiedad de solidaridad de la condición de recurrencia, se puede concluir que, dada una cadena no-reducible, sus estados o son todos recurrentes o ninguno es recurrente.

[103]. $\Im$ no define necesariamente una clase.

Esta convolución de segundo orden permite deducir que la variable **tiempo de recurrencia por segunda vez al estado** j ($n_j^{(2)}$), está definida por la suma de dos variables aleatorias independientes, $n_j^{(2)} = n_{j,1} + n_{j,2}$ (donde $n_{j,i}\ (j = 0,1,2,...;\ \ i = 1,2)$ denota el número de transiciones a ocurrir entre la $(i-1)$-ésima y la i-ésima recurrencias al estado j), ambas con la misma distribución de probabilidades ($f_{jj}^{(n)}$)[104].

En general, la **probabilidad de que la cadena retorne por k-ésima vez al estado recurrente** j, **al cabo de n transiciones**, está definida por:

$$f_{jj}^{(k,n)} = f_{jj}^{(1)} f_{jj}^{(k-1,n-1)} + f_{jj}^{(2)} f_{jj}^{(k-1,n-2)} + ... + f_{jj}^{(n-k-1)} f_{jj}^{(k-1,1)}$$

$$\left\{ f_{jj}^{(k,n)} \right\} = \left\{ f_{jj}^{(n)} \right\} * \left\{ f_{jj}^{(k-1,n)} \right\} = ... = \left\{ f_{jj}^{(n)} \right\}^{*k} \qquad (k = 1,2,...; n = k, k+1,...)$$

En consecuencia, se puede concluir que la variable aleatoria $n_j^{(k)}$ **(tiempo de recurrencia por k-ésima vez al estado j)** está formada por la suma de las k variables aleatorias independientes:

$$n_j^{(k)} = n_{j,1} + n_{j,2} + ... + n_{j,k} \qquad (k = 1,2,...,n)$$

todas con la misma distribución de probabilidades, $f_{jj}^{(n)}$.

Denotando por n_j^* a la variable que representa el número de recurrencias al estado $j\ (= 0,1,2...)$, para $n = 1,2,...$, se demuestra que: **i)** si j es no-recurrente, entonces n_j^* se distribuye de acuerdo con una función geométrica de la forma $p\left(n_j^* = n\right) = \left(1 - f_{jj}\right) f_{jj}^{(n)}\ (n = 0,1,2,...)$ y **ii)** si j es recurrente, se verifica que $p\left(n_j^* = \infty\right) = 1$.

1.7.1.4.- La condición de estacionariedad y las probabilidades en el límite

Sea una cadena de Markov y sea $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}\ (i, j = 0,1,2,...)$ (tal que: **i)** $\pi_j \geq 0\ (j = 0,1,2,...)$ y **ii)** $\sum_{j=0}^{\infty} \pi_j = 1$) su distribución de probabilidades de equilibrio. La condición necesaria y suficiente para que esta distribución sea estacionaria es que $\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}\ (j = 0,1,2,...)$. Se dice que una cadena de Markov es **estacionaria** cuando su distribución de probabilidades no-condicionada es estacionaria.

Se dice que una cadena de Markov es **reversible** con respecto a la distribución $\{\pi_j\}$ si se verifica que $\pi_i p_{ij} = \pi_j p_{ji}\ (i, j = 0,1,2,...)$ y se demuestra[105] que, si se cumple esta condición, la distribución $\{\pi_j\}$ es

---

[104]. Ver Landro, A.H. (2002).

[105]. Ver Rosenthal, J.S. (2006).

estacionaria para dicha cadena.

En particular, dada una cadena es no-reducible y no-periódica, entonces se demuestra que[106]: **i)** también es recurrente-positiva; **ii)** existe una distribución de probabilidades de equilibrio estacionaria $\pi' = \pi' P$; **iii)** esta distribución de probabilidades de equilibrio es única y tal que $\pi_j = \frac{1}{E(n_j)}$ (donde, como se mencionó en la Sec. 1.7.1.3, $E(n_j)$ denota el tiempo esperado de recurrencia al estado j)[107].

Dada una cadena no-reducible y no-periódica, la condición de ser recurrente-positiva es necesaria y suficiente para asegurar la existencia de una distribución límite estacionaria (a las cadenas no-reducibles, no-periódicas y recurrentes-positivas se las denomina **ergódicas** o **regulares**). En general, se demuestra que la condición suficiente, pero no necesaria, para que una cadena de Markov no-reducible sea recurrente, es que posea una distribución de probabilidades de equilibrio estacionaria. Es decir, que es posible hallar cadenas no-reducibles recurrentes que no poseen una distribución de equilibrio estacionaria.

Dada una cadena no-reducible, no-periódica y recurrente-nula o no-recurrente, se verifica que $\lim_{n\to\infty} p_{ij}^{(n)} = 0 \; (i, j = 0,1,2,...)$. Luego, como corolario se puede concluir que una cadena no-reducible y no-periódica, con dominio de los estados finito, no puede estar formada exclusivamente por estados recurrentes-nulos. Es decir que una cadena con estas características es recurrente-positiva y, por lo tanto, posee una distribución de probabilidades de equilibrio.

Por otra parte, dada una cadena recurrente con período T , se verifica que:

$$\lim_{n\to\infty} p_{ij}^{(nT_j)} = \frac{T_j}{E(n_j)} \qquad (j = 0,1,2,...)$$

1.7.1.5.- Probabilidades de permanencia

Sea $\nu_j^{(n)}$ la variable aleatoria que representa el **tiempo a transcurrir por el proceso en el estado j al cabo de n transiciones (tiempo de ocupación del estado j al cabo de n transiciones)**. A partir del concepto de tiempo de ocupación, la probabilidad $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ puede ser interpretada, también, como la probabilidad de que el tiempo a transcurrir en el estado j , al cabo de infinitas transiciones, sea mayor que cero, siendo i el estado inicial de la cadena, $f_{ij} = p\left[\left(\nu_j^{(\infty)} > 0\right) / \left(E_0 = i\right)\right]$ (las variables **tiempo de recurrencia por k-ésima vez al estado j**, $n_j^{(k)}$, y **tiempo de ocupación del estado j al cabo de n transiciones**, $\nu_j^{(n)}$,

---

[106]. Ver Billingsley, P. (1986).

[107]. Como se vio en la Sec. 1.7.1.1, para el caso particular de una cadena con un dominio de los estados $\{0,1\}$, se obtiene que:

$$\lim_{n\to\infty} \begin{bmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{bmatrix}^n = \begin{bmatrix} \frac{p_{10}}{p_{01}+p_{10}} & \frac{p_{01}}{p_{01}+p_{10}} \\ \frac{p_{10}}{p_{01}+p_{10}} & \frac{p_{01}}{p_{01}+p_{10}} \end{bmatrix}.$$

se encuentran, obviamente, vinculadas por la relación $p\left(n_j^{(k)} \le n\right) = p\left(\nu_j^{(n)} \ge k\right)$).

Sea $p_{ij}\left(\nu_j^{(\infty)}\right) = p\left[\left(\nu_j^{(\infty)} = \infty\right)/\left(E_0 = i\right)\right]$ la **probabilidad de que la cadena ocupe el estado** j **un tiempo infinito,** si el estado inicial es i. Entonces, se puede escribir:

$$p_{ij}\left(\nu_j^{(n)}\right) = p\left[\left(E_1 = j\right)\cap\left(E_\ell = j \text{ para } \inf\text{initos valores de } \ell > 1\right)/\left(E_0 = i\right)\right] +$$

$$+ p\left[\left(E_1 \ne j\right)\cap\left(E_2 = j\right)\cap\left(E_\ell = j \text{ para } \inf\text{initos valores de } \ell > 2\right)/\left(E_0 = i\right)\right] +$$

$$+ p\left[\left(E_1 \ne j\right)\cap\left(E_2 \ne j\right)\cap\left(E_3 = j\right)\cap\left(E_\ell = j \text{ para } \inf\text{initos valores de } \ell > 3\right)/\left(E_0 = i\right)\right] + \ldots +$$

$$+ p\left[\left(E_1 \ne j\right)\cap\ldots\cap\left(E_{m-1} \ne j\right)\cap\left(E_\ell = j \text{ para } \inf\text{initos valores de } \ell > m\right)/\left(E_0 = i\right)\right] + \ldots =$$

$$= p\left[\left(E_1 = j\right)/\left(E_0 = i\right)\right]p_{jj}\left(\nu_j^{(\infty)}\right) + p\left[\left(E_1 \ne j\right)\cap\left(E_2 = j\right)/\left(E_0 = i\right)\right]p_{jj}\left(\nu_j^{(\infty)}\right) +$$

$$+ p\left[\left(E_1 \ne j\right)\cap\left(E_2 \ne j\right)\cap\left(E_3 = j\right)/\left(E_0 = i\right)\right]p_{jj}\left(\nu_j^{(\infty)}\right) + \ldots + p\left[\left(E_1 \ne j\right)\cap\left(E_2 \ne j\right)\cap\ldots\cap\left(E_m = j\right)\right]p_{jj}\left(\nu_j^{(\infty)}\right) + \ldots =$$

$$= f_{ij}^{(1)}p_{jj}\left(\nu_j^{(\infty)}\right) + f_{ij}^{(2)}p_{jj}\left(\nu_j^{(\infty)}\right) + \ldots =$$

$$= p_{jj}\left(\nu_j^{(\infty)}\right)\sum_{n=1}^{\infty} f_{ij}^{(n)} = f_{ij}p_{jj}\left(\nu_j^{(\infty)}\right)$$

Por otra parte, se verifica que:

$$p\left[\left(\nu^{(\infty)} \ge k\right)/\left(E_0 = j\right)\right] = \sum_{m=1}^{\infty} f_{jj}^{(n)} p\left[\left(\nu_j^{(\infty)} \ge k-1\right)/\left(E_m = j\right)\right] =$$

$$= p\left[\left(\nu_j^{(\infty)} \ge k-1\right)/\left(E_0 = j\right)\right]\sum_{m=1}^{\infty} f_{jj}^{(m)} = p\left[\left(\nu_j^{(\infty)} \ge k-1\right)/\left(E_0 = i\right)\right]f_{jj} =$$

$$= p\left[\left(\nu_j^{(\infty)} \ge k-2\right)/\left(E_0 = i\right)\right]\left(f_{jj}\right)^2 = \ldots = \left(f_{jj}\right)^k \qquad \left(j = 0,1,2,\ldots; k = 1,2,\ldots\right)$$

Por lo tanto, será:

$$p_{jj}\left(\nu_j^{(\infty)}\right) = p\left[\left(\nu_j^{(\infty)} = \infty\right)/\left(E_0 = i\right)\right] = \lim_{k\to\infty} p\left[\left(\nu_j^{(\infty)} \ge k\right)/\left(E_0 = i\right)\right] = \lim_{k\to\infty}\left(f_{jj}\right)^k$$

En consecuencia, si el estado j es recurrente ($f_{jj} = 1$), entonces, con probabilidad igual a uno, la cadena retornará a dicho estado infinitas veces ($p_{jj}\left(\nu_j^{(\infty)}\right) = 1$) y si es no-recurrente, con probabilidad igual a uno, retornará, como máximo, un número finito de veces ($p_{jj}\left(\nu_{jj}^{(\infty)}\right) = 0$). En otros términos, si el estado j es recurrente, se verificará que $p\left[\left(Y_t = j\right)i.o./\left(Y_{t-n} = j\right)\right] \equiv p\left[\left(Y_t = j\right) \inf\text{initely often}/\left(Y_{t-n} = j\right)\right] = 1$ (donde $\left[\left(Y_t = j\right)i.o./\left(Y_{t-n} = j\right)\right]$ denota el evento "que la recurrencia al estado j se produzca infinitas veces"

$\left[\left(Y_t = j\right) i.o./\left(Y_{t-n} = j\right)\right] = \limsup_{n\to\infty}\left[\left(Y_t = j\right)/\left(Y_{t-n} = j\right)\right] = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}\left[\left(Y_t = j\right)/\left(Y_{t-m} = j\right)\right]$). De la misma forma, si el estado j es no-recurrente, se verificará que $p\left[\left(Y_t = j\right) i.o./\left(Y_{t-n} = j\right)\right] = 0$.

1.7.1.6.- El "random walk" (RW)

El proceso **random walk-unidimensional** es un caso especial de cadena de Markov en el que el dominio de la variables está definido por un subconjunto de los números enteros y cuyas transiciones desde un estado i $(= 0,1,2,...)$ admiten solamente tres posibilidades: **i)** de permanecer en dicho estado (con probabilidad $1 - p_i - q_i$), **ii)** pasar al estado $i+1$ (con probabilidad p_i) o **iii)** pasar al estado $i-1$ (con probabilidad q_i).

Supóngase, por razones de comodidad, que el dominio de las variables del RW está definido por el subconjunto de los enteros no-negativos. Las probabilidades de transición quedan, entonces, definidas de la siguiente forma:

$$p_{i,i+1} = p_i > 0 \qquad (i = 0,1,2,...)$$

$$p_{i,i-1} = q_i > 0 \qquad (i = 1,2,...)$$

$$p_{i,j} = 1 - p_i - q_i \qquad (i = 0,1,2,...)$$

Si se verifica que $p_{0,1} = p_0 = 1$ y, por lo tanto, que $p_{0,0} = 1 - p_0 = 0$, entonces se puede concluir que, en el estado 0 el proceso presenta una **barrera reflectora** (es decir, que una vez asumido el estado cero, en la transición siguiente el proceso retorna automáticamente al estado 1). De la misma forma, si se verifica que $p_{0,1} = p_0 = 0$ y, por lo tanto, que $p_{0,0} = 1 - p_0 = 1$, se puede concluir que en el estado cero el proceso presenta una **barrera absorbente** (es decir, que una vez asumido el estado cero, el proceso no lo puede abandonar nunca más). Finalmente, si $p_{0,1} = p_0 > 0$ y , en consecuencia, $p_{0,0} = 1 - p_0 > 0$, se puede concluir que, en el estado cero, el proceso presenta una **barrera parcialmente reflectora**.

Si bien originariamente el proceso RW fue diseñado para describir el recorrido aleatorio que efectúa -como expresa la literatura sobre el particular- una "persona intoxicada"[108], puede ser asimilado a la descripción del comportamiento de la variable que representa el estado de la fortuna de un jugador, al cabo de una serie de partidas. Supóngase un jugador A con una fortuna W_A que compite con un contrincante, B, que posee una fortuna ilimitada. Sean: **i)** p_i $(i = 1,2,...)$ la probabilidad de A de ganar una partida cuando su fortuna es igual a i (es decir, la probabilidad de que la fortuna de A pase del estado i al estado $i+1$) y **ii)** $q_i = 1 - p_i$ $(i = 1,2,...)$ la probabilidad de A de perder una partida cuando su fortuna es igual i (es decir, la probabilidad de que la fortuna de A pase del estado i al estado $i-1$). Obsérvese que, en $i = 0$, el proceso posee un estado absorbente. Se puede concluir, entonces, que el proceso $\{Y_j\}$, que representa la fortuna de A al cabo de j partidas (y al que se conoce como **de la ruina de los jugadores**) es un RW.

Sean las probabilidades $\pi_0 = 1$ y $\pi_j = \dfrac{p_0 p_1 \cdots p_{j-1}}{q_1 q_2 \cdots q_j}$ $(j = 1,2,...)$. A partir de la condición de recurrencia

---

[108]. En el barrio lo denominarían el "paseo del chorrabo".

enunciada en la Sec. 1.7.1.4, se demuestra[109] que la condición necesaria y suficiente para que este proceso RW sea estacionario en sentido estricto es que $\sum_{j=0}^{\infty}\pi_j < \infty$. Como corolario de esta condición se obtienen las siguientes condiciones necesarias y suficientes:

i) Si se verifica que $\sum_{j=0}^{\infty}\frac{1}{p_j\pi_j} = 0$, entonces el proceso RW es **recurrente.**

ii) Si, además de la condición precedente, se verifica que $\sum_{j=0}^{\infty}\pi_j = \infty$, entonces el proceso RW es **recurrente-nulo.**

iii) Si, además de la condición expresada en **i)**, se verifica que $\sum_{j=0}^{\infty}\pi_j < \infty$, entonces el proceso RW es **recurrente-positivo.**

iv) Si se verifica que $\sum_{j=0}^{\infty}\frac{1}{p_j\pi_j} < \infty$, entonces el proceso RW es **no-recurrente.**

En particular, si $p_i = p$ y $q_i = 1-p = q\ (i = 1,2,...)$, con $p > q$, queda definido un proceso con ventaja para el jugador A en cada partida individual. Se demuestra fácilmente que existe una probabilidad $\left(\frac{q}{p}\right)^{W_A}$ de que A termine arruinado y, por lo tanto, una probabilidad $1-\left(\frac{q}{p}\right)^{W_A}$ de que su fortuna se haya incrementado sin límite en el largo plazo. Para $p \le q$, se puede asegurar con certeza de que A terminará arruinado.

Otro caso particularmente interesante se presenta si el jugador B posee una fortuna limitada, W_B. Las probabilidades de transición correspondientes al proceso $\{Y_i; i = 1,2,...,W_A + W_B - 1\}$ (que representa la fortuna de A) serán de la forma:

$$p_{i,i+1} = p_i \qquad (i = 1,2,...,W_A + W_B - 1)$$
$$p_{i,i-1} = q_i \qquad (i = 1,2,...,W_A + W_B - 1)$$
$$p_{ii} = 1 - p_i - q_i \qquad (i = 1,2,...,W_A + W_B - 1)$$

Obsérvese que este proceso posee barreras absorbentes en los estados 0 (se arruina el jugador A) y $W_A + W_B$ (se arruina el jugador B).

Generalizando esta propiedad, se demuestra que en todo proceso RW con un número finito de estados

---

[109]. Ver Karlin, S.; Taylor, H.M. (1975).

($i = 0,1,2,...W$), los estados 0 y W pueden ser, independientemente y en cualquier combinación, barreras reflectoras, absorbentes o parcialmente reflectoras.

Se denomina **RW-simétrico** a todo proceso "random walk" cuyas probabilidades de transición son de la forma $p_{i,i+1} = p > 0$; $p_{i,i-1} = p$ y $p_{ii} = 1 - 2p \geq 0$ $(i = 0,\pm1,\pm2,...)$ [109]. Este proceso es tal que puede pasar de un estado cualquiera a cualquier otro estado con probabilidad igual a uno, es decir que, con probabilidad igual a uno, se verifica que $\limsup_{n\to\infty} Y_n = \infty$ y $\liminf_{n\to\infty} Y_n = -\infty$. Esta condición está relacionada con la ley del logaritmo iterado (ver Sec. 1.5): Sea $Y_n = X_1 + \ + X_2 + ... + X_n$ un RW, donde las variables X_i son iid y tales que $E(X_i) = 0$ y $\sigma^2(X_i) = 1$ $(i = 1,2,...,n)$, entonces, con probabilidad igual a uno, se verifica que $\limsup_{n\to\infty} \frac{Y_n}{\sqrt{2n\log\log(n)}} = 1$. En otras palabras, se verifica que:

$$p\left\{\left[Y_n \geq (1+\varepsilon)\sqrt{2n\log\log(n)}\right] i.o.\right\} = 0$$

y que:

$$p\left\{\left[Y_n \geq (1-\varepsilon)\sqrt{2n\log\log(n)}\right] i.o.\right\} = 1$$

(ver Billingsley, P. (1995).

1.7.2.- El proceso de Poisson

1.7.2.1.- El proceso de Poisson homogéneo o con función intensidad de ocurrencia constante

Sea $\{Y(t)\}$ $(t \geq 0)$ un proceso $AR(1)$ continuo en el dominio del tiempo y discreto en el dominio de las variables, tal que las variables $Y(t)$ pueden asumir valores enteros no-negativos, de acuerdo con el número de ocurrencias de un evento dado, E, al momento $t \geq 0$ (o, lo que es lo mismo, con el número de "saltos" de medida uno ocurridos al momento t)[110]. Siendo este evento tal que: **i)** sus ocurrencias para intervalos de tiempo

---

[109]. Algunos autores reservan el calificativo de simétrico para aquellos RW para los cuales $p = \frac{1}{2}$. Este proceso es la versión discreta en el dominio del tiempo y discreta en el dominio de las variables del movimiento Browniano (ver Sec. 4.4).

[110]. Este proceso puede ser considerado como un miembro (el más elemental) de la familia de los denominados **procesos con salto**. Como todo proceso unidimensional continuo en el dominio del tiempo, este proceso pertenece a la clase de los procesos **CADLAG ("continu à droite, limite à gauche"**, también denominados como **RCLL** en la impronunciable nomenclatura inglesa), los cuales se caracterizan por ser continuos por la derecha y con límite por la izquierda. Es decir, tales que: **i)** para todo $t \in T$, los límites $\lim_{\substack{\tau\to t \\ (\tau<t)}} Y(\tau) = Y(t-)$ y $\lim_{\substack{\tau\to t \\ (\tau>t)}} Y(\tau) = Y(t+)$ existen; **ii)** $Y(t) = Y(t+)$. La segunda condición implica que, si el proceso presenta un salto en el momento t , su ocurrencia no puede ser predicha a partir de la observación de su trayectoria hasta el momento t . Con un razonamiento similar se puede definir una clase de procesos **CAGLAD ("continu à gauche, limite à droite"**), tales que: **i)** los

disjuntos son independientes lo cual implica que, dado un valor $Y(t_0)$ conocido, la distribución de probabilidades de cualquier variable $Y(t)\,(t > t_0)$ no se ve afectada por el conocimiento que se pueda tener acerca de cualquier realización del proceso para $t < t_0$ (como se vio en la Sec. 1.3, ésta es la **condición de Markov** o **de autorregresividad de primer orden**); **ii)** es imposible la presentación simultánea de dos o más de sus ocurrencias (**condición de linealidad**)[112].

Las correspondientes **probabilidades de transición** del estado x, en el momento t_1, al estado y, en el momento $t_2\,(t_2 \geq t_1)$ poseen las siguientes propiedades: **i)** $0 \leq p_{xy}(t_1,t_2) \leq 1\ (x,y = 0,1,2,...)$ y **ii)** $\sum_{h=0}^{\infty} p_{xh}(t_1,t_2) \leq 1\,(x = 0,1,2,...)$[113].

La **ecuación de Kolmogorov-Chapman** asume, en este caso, la siguiente forma:

$$p\{[Y(t_2) = y]/[Y(t_1) = x]\} = p_{xy}(t_1,t_2) = \sum_{h=x}^{y} p_{xh}(t_1,t)\,p_{hy}(t,t_2) \qquad (t_1 \leq t \leq t_2)$$

Seleccionando un valor Δt, arbitrariamente pequeño, la probabilidad de que se produzca una ocurrencia del evento E en un intervalo $[t,t+\Delta t)$ puede ser definida de la siguiente forma:

$$p\{[Y(t+\Delta t) = y+1]/[Y(y) = y]\} = p_{y,y+1}(t,t+\Delta t) = \rho\,\Delta t \qquad (y = 0.,1,2...)$$

donde ρ denota la **tasa instantánea de ocurrencia del evento E** (en general, a cada estado le corresponde una función continua con respecto a la variable t, no-negativa, denominada **función intensidad de ocurrencia del evento E**, definida como $\lim_{\Delta t \to 0} \frac{1 - p_{yy}(t,t+\Delta t)}{\Delta t} = \rho(t)\,(y = 0,1,2,...)$, en el caso tratado en esta sección se ha supuesto que dicha función es independiente del tiempo)[114]. La probabilidad de que, en un intervalo $[t,t+\Delta t)$, se produzcan $k\,(= 2,3,...)$ ocurrencias del evento E, está dada por:

$$p\{[Y(t+\Delta t) = y+k]/[Y(t) = y]\} = p_{y,y+k}(t,t+\Delta t) = o(\Delta t) \qquad (y = 0,1,2,...;k = 2,3,4,...)$$

---

límites $\lim_{\substack{\tau \to t \\ (\tau<t)}} Y(\tau) = Y(t-)$ y $\lim_{\substack{\tau \to t \\ (\tau>t)}} Y(\tau) = Y(t+)$ existen y **ii)** $Y(t) = Y(t-)$, es decir que, si el proceso presenta un salto en el momento t, su ocurrencia puede ser predicha a partir de la observación de su trayectoria hasta el momento t.

[112]. El proceso de Poisson es un **proceso puro de salto**. En realidad se denomina proceso con salto a una suma de: **i)** una condición inicial no-aleatoria; **ii)** una integral de Itô con respecto a un movimiento Browniano; **iii)** una integral de Riemann con respecto a t y **iv)** un proceso puro de salto. Los procesos con "salto" serán tratados en la Sec. 4.4.

[113]. Si en esta condición se verifica la igualdad para todo x y todo $t \in T$, se dice que el proceso es **honesto**. Por el contrario, si para algún estado x y algún valor de t, se verifica la desigualdad, se dice que el proceso es **deshonesto**.

[114]. Como se verá inmediatamente, desde el punto de vista de las aplicaciones, la tasa ρ puede ser interpretada como el valor al cual converge la frecuencia relativa del número de ocurrencias del evento E a medida que el número de observaciones aumenta indefinidamente.

donde $o(\Delta t)$ denota una función de orden asintótico menor que Δt [115]tal que $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$ [116]. En consecuencia, la probabilidad de que en el intervalo $[t, t+\Delta t)$ se produzca más de una ocurrencia del evento E, será $\sum_{k \geq 2} p_{y,y+k}(t, t+\Delta t) = o(\Delta t)$. Finalmente. la probabilidad de que, en el intervalo $[t, t+\Delta t)$, no se produzca ninguna ocurrencia del evento E será:

$$p\{[Y(t+\Delta t) = y] / [Y(t) = y]\} = p_{yy}(t, t+\Delta t) = 1 - \rho \Delta t - o(\Delta t) \qquad (y = 0,1,2,...)$$

Las probabilidades definidas precedentemente componen el **sistema de probabilidades de transición infinitesimal** [117].

Supóngase un instante inicial $t = 0$ en el cual el proceso se encuentra en el estado $Y(0) = 0$. De acuerdo con la condición de Kolmogorov-Chapman y el sistema de probabilidades de transición infinitesimal definido precedentemente, la probabilidad de que al momento $t + \Delta t$ se hayan producido y ocurrencias del evento E, puede ser expresada de la siguiente forma:

$$p\{[Y(t+\Delta t) = y] / [Y(0) = 0]\} \equiv p_{0y}(0, t+\Delta t) \equiv p_y(t+\Delta t) = \sum_{j=0}^{y} p_{y-j}(t) p_{y-j,y}(t, t+\Delta t) =$$

$$= p_y(t)(1 - \rho \Delta t - o(\Delta t)) + p_{y-1}(t) \rho \Delta t + \sum_{j \geq 2} p_{y-j}(t) o(\Delta t)$$

Es decir, se puede escribir:

$$\frac{p_y(t+\Delta t) - p_y(t)}{\Delta t} = -\rho p_y(t) + \rho p_{y-1}(t) + \sum_{j \geq 2} p_{y-j}(t) \frac{o(\Delta t)}{\Delta t}$$

Considerando, en esta última ecuación, el límite cuando $\Delta t \to 0$, se obtiene el sistema de **ecuaciones estocásticas diferenciales-en diferencias de Kolmogorov**[118]:

---

[115]. Se dice que una función $f(x)$ es de orden asintótico menor que $g(x)$ (y se denota por $f(x) = o(g(x))$) si $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$. Si se verifica que $0 < \lim_{x \to \infty} \frac{f(x)}{g(x)} < \infty$, se dice que la función $f(x)$ es de orden asintótico igual a $g(x)$ y se denota de la siguiente forma $f(x) = O(g(x))$.

[116]. Lo cual implica suponer que, si la longitud Δt es seleccionada en forma adecuada, es posible asegurar la quasi imposibilidad de más de una ocurrencia de E en un intervalo $[t, t+\Delta t)$.

[117]. Los procesos Markovianos con estructura de probabilidades de transición estacionaria (que satisfacen una forma muy general de la condición de Doeblin) son uniformemente ϕ-mixing con coeficientes $\phi(t)$ geométricamente decrecientes (ver Doob, J.L. (1953), Rosenblatt, M. (1971)).

[118]. La denominación de **ecuaciones estocásticas diferenciales-en diferencias** para estas ecuaciones diferenciales con variables desfasadas se debe a Bellman, R.; Cooke, K.L. (1963). La calificación de estocásticas es para distinguirlas de las ecuaciones diferenciales-en diferencias ordinarias. La aleatoriedad de estas ecuaciones se origina en la habitual aleatoriedad inherente a la

$$\lim_{\Delta t \to 0} \frac{p_y(t+\Delta t) - p_y(t)}{\Delta t} = \frac{d}{dt} p_y(t) = -\rho\, p_y(t) + \rho\, p_{y-1}(t) \qquad (y = 0,1,2,...)$$

cuya solución particular es de la forma[119]:

$$p\{[Y(t) = y] / (Y(0) = 0)\} = p_{0y}(0,t) = p_y(t) = \frac{(\rho t)^y}{y!} e^{-\rho t} \qquad (y = 0,1,2,...)$$

Lo que demuestra que, para cada valor de $t > 0$, la variable "número de ocurrencias del evento E" tiene una distribución de probabilidades de Poisson con parámetro ρt ($Y(t):\wp(\rho t)$)[120].

Una forma alternativa de resolución del sistema de ecuaciones de Kolmogorov se obtiene utilizando la función generatriz de probabilidades, $G_Y(w,t) = \sum_{Y=0}^{\infty} p_Y(t) w^Y \ (|w| \le 1; 0 \le t < \infty)$. Multiplicando m. a m. la ecuación $\frac{d}{dt} p_Y(t) = -\rho p_Y(t) + \rho p_{Y-1}(t) \ (Y = 0,1,2,...)$ por w^Y y sumando con respecto a Y, se obtiene la siguiente ecuación en derivadas parciales[121]:

$$\sum_{Y=0}^{\infty} \left[\frac{d}{dt} p_Y(t)\right] w^Y = -\rho \sum_{Y=0}^{\infty} p_Y(t) w^Y + \rho w \sum_{Y=0}^{\infty} p_{Y-1}(t) w^{Y-1}$$

---

condición inicial. Debe tenerse en cuenta que, si se adapta la solución a una realización particular de la condición inicial, su solución no requiere de cálculo estocástico, es decir, se pueden utilizar los métodos analíticos clásicos. En particular, las ecuaciones aquí definidas pertenecen a la clase de las denominadas **ecuaciones "forward"**. Tomando en consideración el intervalo $(s - \Delta s, t]$ (para $t > s$) se puede escribir $p_{yx}(s - \Delta s, t) = p[X(t) = x \,/\, X(s - \Delta s) = y] = \sum_j p[X(s) = j \,/\, X(s - \Delta s) = y] p[X(t) = x \,/\, X(s) = j] = \sum_j p_{sj}(s - \Delta s, s) p_{js}(s,t)$. A partir de esta expresión se obtienen las correspondientes ecuaciones "backward" $\frac{d}{dt} p_{YY}(s,t) = \rho\, p_{YY}(s,t)$ y $\frac{d}{dt} p_{XY}(s,t) = \rho [p_{YX}(s-t) - p_{Y+1,X}(s,t)] \ (X > Y)$. Como se verá en la Sec. 1.6.2.4, a partir de estas ecuaciones (debidas a Kolmogorov, A.N. (1931)) queda definida la misma distribución de probabilidades $p_{XY}(t)$ que la proporcionada por las ecuaciones "forward". Para una análisis más detallado, ver Chung, K.L. (1960), Doob, J.L. (1953).

[119]. Debe tenerse en cuenta que, en estas soluciones, $\frac{d}{dt} p_0(t) = -\rho\, p_0(t)$.

[120]. Para una definición constructivista de este proceso de Poisson, ver Beśka, M. (1978).

[121]. Dado que $|p_Y(t) w^Y| < |w|^Y$, se puede concluir que, para $|w| < 1$, la serie $\sum_{Y=0}^{\infty} p_Y(t) w^Y$ es uniformemente convergente en t. De modo que se puede asegurar que la serie $\sum_{Y=0}^{\infty} \frac{d}{dt} p_Y(t) w^Y$ converge uniformemente en t y, por lo tanto, que se justifica la diferenciación término a término para $|w| < 1$. Por otra parte, dado que $\lim_{w \to 1} G_Y(w,t) = G_Y(1,t)$, esta condición se verifica también para $w = 1$.

$$\frac{\partial}{\partial t}G_Y(w,t) = -\rho(1-w)G_Y(w,t)$$

cuya solución general es de la forma $G_Y(w,t) = c(w)e^{-\rho t(1-w)}$. Teniendo en cuenta la condición inicial $G_Y(w,0) = \sum_{Y=0}^{\infty} p_Y(0)w^Y = p_0(0) = 1$, se obtiene la siguiente solución particular:

$$G_Y(w,t) = e^{-\rho t(1-w)} = \sum_{Y=0}^{\infty}\left[\frac{(\rho t)^Y}{Y!}e^{-\rho t}\right]w^Y$$

El proceso $\{Z_t\} = \{Y(t+\tau) - Y(t)\}$ formado por los incrementos del proceso $\{Y_t\}$, que representa el número de ocurrencias del evento E en el intervalo $(t;t+\tau]$ condicionado por el número de saltos ocurridos hasta el momento t, tiene la misma distribución de probabilidades que $\{Y_t\}$ [122] y es estacionario en sentido estricto (es decir, $\{Y(t)\}$ es un proceso con incrementos independientes estrictamente estacionarios) pero, como se comentó en la Sec. 1.3.1, esta condición no implica que $\{Y(t)\}$ sea estacionario (en realidad, se puede concluir que ningún proceso no-constante con incrementos independientes estacionarios es estacionario). En general se demuestra que el proceso formado por los incrementos $\{Y(t) - Y(\tau)\}$ $(0 \le \tau < t)$ es un proceso de Poisson con parámetro $\rho(t-\tau)$, es decir, tal que:

$$p[Y(t) - Y(\tau) = y] = \frac{[\rho(t-\tau)]^y}{y!}e^{-\rho(t-\tau)} \qquad (y = 0,1,2,\ldots)$$

Como corolario de los resultados anteriores se pueden obtener las siguientes definiciones:

i) Para $t_1 < t_2$ e $y_1 < y_2$ se verifica que:

$$p\{[Y(t_1) = y_1]/[Y(t_2) = y_2]\} = \frac{p\{[Y(t_1) = y_1]\cap[Y(t_2) - Y(t_1) = y_2 - y_1]\}}{p[Y(t_2) = y_2]} =$$

$$= \frac{\left(\frac{t_1^{y_1}}{y_1!}\right)\left(\frac{(t_2 - t_1)^{y_2 - y_1}}{(y_2 - y_1)!}e^{-\rho(t_2 - t_1)}\right)}{\frac{t_2^{y_2}}{y_2!}e^{-\rho t_2}} =$$

$$= \binom{y_2}{y_1}\frac{t_1^{y_1}(t_2 - t_1)^{y_2 - y_1}}{t_2^{y_2}}$$

---

[122]. Como se verá inmediatamente, a esta propiedad se la conoce como **ausencia de memoria**.

ii) Sean dos procesos independientes $\{Y_1(t)\}:\wp(\rho_1 t)$ y $\{Y_2(t)\}:\wp(\rho_2 t)$. Para $y_2 > y_1$, se verifica que:

$$p\{[Y_1(t) = y_1]/[Y_1(t) + Y_2(t) = y_2]\} = \frac{p\{[Y_1(t) = y_1] \cap [Y_2(t) - Y_1(t) = y_2 - y_1]\}}{p[Y_1(t) + Y_2(t) = y_2]} =$$

$$= \frac{\left(\frac{(\rho_1 t)^{y_1}}{y_1!} e^{-\rho_1 t}\right)\left(\frac{(\rho_2 t)^{y_2 - y_1}}{(y_2 - y_1)!} e^{-\rho_2 t}\right)}{\frac{[(\rho_1 + \rho_2)t]^{y_1}}{y_2!} e^{-(\rho_1 + \rho_2)t}} =$$

$$= \binom{y_2}{y_1}\left(\frac{\rho_1}{\rho_1 + \rho_2}\right)^{y_1}\left(\frac{\rho_2}{\rho_1 + \rho_2}\right)^{y_2 - y_1}$$

es decir, que $[(Y_1(t) = y_1)/(Y_1(t) + Y_2(t) = y_2)]:b^{-1}\left(y_2 - y_1; \frac{\rho_2}{\rho_1 + \rho_2}\right)$.

Sea un proceso $\{Y_t\}:\wp(\rho t)$. El proceso $\{Y^*(t)\} = \{Y(t) - \rho t\}$ se conoce como **de Poisson compensado** y, para $0 \le \tau < t$, es tal que:

$$E[Y^*(t)/Y(\tau)] = E[Y^*(t) - Y^*(\tau)/Y(\tau)] + E[Y(\tau)] =$$
$$= E[Y(t) - Y(\tau) - \rho(t - \tau)/Y(\tau)] + Y^*(\tau) =$$
$$= E[Y(t) - Y(\tau)] - \rho(t - \tau) + Y^*(\tau) = Y^*(\tau)$$

Lo que demuestra que, como todo proceso con incrementos independientes, cumple la condición de la **martingala**[123].

Trasladando, ahora, el análisis del proceso de Poisson al dominio del tiempo, sea T_{01} la variable aleatoria que representa el tiempo a transcurrir hasta la primera ocurrencia del evento E; es decir, el tiempo a transcurrir por el proceso en el estado inicial. Entonces, será:

$$p_0(t + \Delta t) = p(T_{01} > t + \Delta t) =$$
$$= p\{(T_{01} > t) \cap (\textit{no se produzca ninguna ocurrencia de E en el intervalo } (t, t + \Delta t))\} =$$
$$= p_0(t)[1 - \rho \Delta t - o(\Delta t)]$$

---

[123]. Esta condición, que implica que, a partir de la información $Y(\tau)$, la mejor predicción del estado del proceso $\{Y(t)\}$ para el momento t es su estado al momento $\tau \le t$, será tratada en la Sec. 4.3.

Luego, haciendo $\Delta t \to 0$ y teniendo en cuenta la condición inicial $p_0(0) = p(T_{01} > 0) = 1$, se obtiene que $p_0(t) = p(T_{01} > t) = e^{-\rho t}$. De modo que, la función de distribución de probabilidades de la variable T_{01} será de la forma $F_{T_{01}}(t) = p(T_{01} \le t) = 1 - e^{-\rho t}$ [124]. Por lo tanto, la correspondiente función de densidad queda definida por $f_{T_{01}}(t) = \rho e^{-\rho t}$ $(t \ge 0)$. Como corolario de este resultado se demuestra que:

$$p(T_{01} > t_1 + t_2 / T_{01} > t_1) = \frac{\int_{t_1+t_2}^{\infty} \rho e^{-\rho t}\, dt}{\int_{t_1}^{\infty} \rho e^{-\rho t}\, dt} = e^{-\rho t_2} = p(T_{01} > t_2)$$

De lo que se puede concluir que, transcurrido un tiempo $[0;t_1]$ en el estado inicial, la probabilidad de que el proceso permanezca en dicho estado al momento t_2 $(> t_1)$ es igual a la probabilidad de que, desde el momento inicial ($t = 0$) hasta el momento t_2, el evento E no se haya producido ninguna vez. A esta propiedad se la conoce como **ausencia de memoria** y se verifica solamente para las variables exponenciales[125].

Obviamente, todas las variables independientes $T_{i,i+1}$ $(i = 01,2,...)$, que representan los tiempos a transcurrir el proceso entre la i-ésima y la $(i+1)$-ésima ocurrencia del evento E (es decir, el tiempo a transcurrir por el proceso en el estado i $(= 0,1,2,...)$) se distribuyen de acuerdo con la misma función de densidad: $f_{T_{i,i+1}}(t) = \rho e^{-\rho t}$ $(i = 0,1,2,...;t \ge 0)$.

Este resultado puede conducir a la conclusión que la distribución exponencial de la variable tiempo a transcurrir entre la i-ésima y la (i+1)-ésima ocurrencia del evento E ($T_{i,i+1}$) implica que $\{Y(t)\}$ es un proceso de Poisson. Pero, debe teners e en cuenta que, si bien ésta es una condición necesaria no es suficiente para identificar a un proceso de Poisson. Esto se debe a que: **i)** la distinción entre aleatoriedad y caos no es necesariamente clara y en muchos casos es imposible y **ii)** habitualmente, dada un función de densidad de una variable unidimensional, es posible construir un número infinito de ecuaciones en-diferencias (de orden finito) cuyas iteraciones son caóticas y que obedecen a dicha función de densidad (se puede demostrar, por ejemplo, que las iteraciones sucesivas de la ecuación $T_{i,i+1} = -\frac{1}{\rho}\ln\left(\left|1 - 2e^{-\rho T_{i-1,i}}\right|\right)$ posee una función de densidad estable de la forma $f_{T_{i,i+1}}(t) = \rho e^{-\rho t}$). De modo que, en general, se puede concluir que, si bien para ciertos procesos dinámicos (caóticos determinísticos o estocásticos) es posible definir en forma unívoca su distribución de probabilidades asociada, el proceso inverso, consistente en caracterizar a un proceso dinámico a partir de una función de densidad, no garantiza la obtención de una solución única.

---

[124]. Obsérvese que la función $F_{T_{01}}(t)$ es invertible y que $F_{T_{01}}^{-1}(t) = -\frac{1}{\rho}\ln(1-t)$ $(0 \le t \le 1)$. Lo cual permite concluir que, dada una variable uniformemente distribuida en el intervalo $[0;1]$, la variable $-\frac{1}{\rho}\ln(U)$ se distribuye de acuerdo con una función exponencial con parámetro ρ.

[125]. En general, se puede asegurar que toda variable aleatoria que cumpla la condición $p[(X > x_1 + x_2) / (X > x_1)] = p(X > x_2)$ $(x_1, x_2 > 0)$ se distribuye de acuerdo con una función de probabilidades exponencial.

Con un razonamiento similar al realizado en la página precedente, se obtiene la siguiente relación:

$$p_Y(t) = p\left(\sum_{i=0}^{Y-1} T_{i,i+1} \le t < \sum_{i=0}^{Y} T_{i,i+1}\right)$$

De modo que el tiempo que se debe esperar que el proceso transcurra en el estado i, será:

$$E\left(T_{i,i+1}\right) = \int_0^\infty \rho t e^{-\rho t}\, dt = \frac{1}{\rho} \qquad (i = 0,1,2,...)$$

y la varianza de las variables $T_{i,i+1}$ será $\sigma^2\left(T_{i,i+1}\right) = \frac{1}{\rho^2}\,(i = 0,1,2,...)$. Luego, la función de distribución conjunta de la variable n-dimensional ($T_{01}, T_{12}, ..., T_{n-1,n}$) será de la forma:

$$F_{T_{01},T_{12},...,T_{n-1,n}}\left(t_1,t_2,...,t_n\right) = \rho^n \exp\left(\sum_{j=1}^{n} t_j\right)$$

Sea $T_i = \sum_{j=1}^{i} T_{j-1,j}\ (i \ge j)$ la variable aleatoria que representa el tiempo a transcurrir por el proceso hasta la ocurrencia por i -ésima vez del evento E. Se verificará, entonces, que $F_{T_i}(t) = p\left(T_i \le t\right)$, de lo que se puede concluir que la variable T_i se distribuye de acuerdo con una función de probabilidades Gamma de la forma:

$$f_{T_i}(t) = \rho e^{-\rho t} \frac{(\rho t)^{i-1}}{(i-1)!} \qquad (t > 0)$$

Dadas las propiedades de independencia y estacionariedad de los incrementos del proceso, se demuestra fácilmente que la distribución de probabilidades de la **variable tiempo a transcurrir hasta la ocurrencia del evento E en el intervalo** $[0,t]$ **, condicionada por el supuesto que el evento E ha ocurrido una vez en dicho intervalo**, es uniforme en el intervalo $[0,t]$:

$$p\left[\left(T_{1,2} < \tau\right) / \left(Y(t) = 1\right)\right] = \frac{p\left\{\left[Y(\tau) = 1\right] \cap \left[Y(t) - Y(\tau) = 0\right]\right\}}{p\left[Y(t) = 1\right]} = \frac{\tau}{t} \qquad (\tau \le t)$$

Generalizando este resultado, la distribución de probabilidades de los **tiempos de ocurrencia,** $T_1, T_2, ..., T_n$ **condicionada por el supuesto que el evento E ha ocurrido n veces en el intervalo** $[0,t]$, es de la forma:

$$p\Big\{\big[(t_1 \le T_1 \le t_1 + h_1)\cap(t_2 \le T_2 \le t_2 + h_2)\cap\ldots\cap(t_n \le T_n \le t_n + h_n)\big] / \big[Y(t) = n\big]\Big\} =$$

$$= \frac{1}{p\big[Y(t) = n\big]}\Big(p\Big\{\big[\textit{que } E \textit{ ocurra una vez en el intervalo } [t_i, t_i + h_i](i = 1,2,\ldots,n)\big]\cap$$

$$\cap\big[\textit{que } E \textit{ no ocurra en el resto del intervalo } [0,t]\big]\Big\} =$$

$$= \frac{\big(\rho h_1 e^{-\rho h_1}\big)\ldots\big(\rho h_n e^{-\rho h_n}\big)e^{-\rho(t-h_1-h_2-\ldots-h_n)}}{\frac{\rho t}{n!}} e^{-\rho t} =$$

$$= \frac{n!}{t^n} h_1 h_2 \ldots h_n \qquad \big(0 < t_1 < t_2 < \ldots < t_n < t_{n+1} \equiv t; t_i + h_i < t_{i+1}; i = 1,2,\ldots,n\big)$$

De modo que $\dfrac{p\left\{\left[\bigcap_{i=1}^{n}(t_i \le T_i \le t_i + h_i)\right] / \big[Y(t) = n\big]\right\}}{h_1 h_2 \ldots h_n} = \dfrac{n!}{t^n}$. Luego, será:

$$\lim_{\substack{h_1 \to 0 \\ \ldots \\ h_n \to 0}} \frac{p\left\{\left[\bigcap_{i=1}^{n}(t_i \le T_i \le t_i + h_i)\right] / \big[Y(t) = n\big]\right\}}{h_1 h_2 \ldots h_n} =$$

$$= f\big\{(T_1, T_2, \ldots, T_n) / \big(Y(t) = n\big)\big\} = \frac{n!}{t^n} \qquad \big(0 < t_1 < t_2 < \ldots < t_n < t\big)$$

Conocida como **distribución de Dirichlet** (en términos compactos, $\big[T_1, T_2, \ldots, T_n / \big(Y(t) = n\big)\big] : D\big(n, [0,t]\big)$)[126].

El problema que se plantea ahora es el de la estimación del parámetro ρ a partir de una muestra sobre las realizaciones del proceso $\{Y(t)\}$ [127]. Para este caso, en que $\{Y(t)\}$ es un proceso de Poisson homogéneo, existen distintas formas de diseño del experimento muestral, entre las que cabe mencionar:

1) La muestra se obtiene a partir de la **observación continua sobre un período fijo**: Sea $(0, T^*)$ el período

---

[126]. Este resultado podría ser interpretado de la siguiente manera: Sean n observaciones independientes sobre una variable con distribución uniforme en el intervalo $[0;t]$ y sea $T^{(1)} \le T^{(2)} \le \ldots \le T^{(n)}$, el arreglo en orden creciente de dichas observaciones. Entonces la variable n -dimensional $\big\{T^{(1)}, T^{(2)}, \ldots, T^{(n)}\big\} : D\big(n, [0;t]\big)$.

[127]. Como se mencionó en la Sec 1.7.1.1, dado que esta sección está referida a la definición estricta de ciertos procesos, el problema de la estimación de los coeficientes será tratado sólo en forma introductoria. Para un análisis más detallado, ver , Moran, P.A.P. (1953), Billingsley, P. (1961) , Prakasa Rao, B.L.S. (1974) (1978), Aalen, O. (1975)(1978), Basawa, I.V.; Scott, D.J. (1977), Basawa, I.V.; Prakasa, Rao, B.L.S. (1980).

de observación, sea n el número de veces que el evento E ocurrió en dicho intervalo y sean $t_1,t_2,...,t_n$ los momentos en que se produjeron dichas ocurrencias. Dado que, como se demostró en la página precedente, las variables $T_{i-1,i} = t_i - t_{i-1}$ $(i = 1,2,...,n)$ son iid con distribución de probabilidades exponencial con parámetro ρ y, teniendo en cuenta que se supuso que $Y(0) = 0$, la función de verosimilitud de la muestra será de la forma:

$$L = \rho^n \exp\left(-\rho\sum_{j=1}^{n} T_{i-1,i}\right)\exp\left[-\rho\left(T^* - t_n\right)\right] = \rho^n \exp\left(-\rho T^*\right)$$

De modo que el estimador máximo-verosímil de ρ será de la forma $\hat{\rho} = \frac{n}{T^*}$. Dado que, como se demostró precedentemente, $n : \wp\left(\rho T^*\right)$, se verifica que $E(\hat{\rho}) = \rho$ y $\sigma^2(\hat{\rho}) = \frac{\rho}{T^*}$ y, teniendo en cuenta que n es un estadístico suficiente y completo para ρ, de acuerdo con el teorema de Rao-Blackwell se puede concluir que $\hat{\rho}$ es un estimador insesgado de varianza mínima[127].

2) La muestra se obtiene a partir de la **observación continua sobre un período aleatorio**: Supóngase que: **i)** la observación comienza en $t = 0$ y continúa hasta que se produzcan n^* ocurrencias del evento E y **ii)** las ocurrencias de E se producen en los puntos $0 < t_1 < t_2 <...< t_{n^*}$. La función de verosimilitud será entonces, de la forma:

$$L = \rho^{n^*} \exp\left(-\rho T_{n^*}\right)$$

(donde $T_{n^*} = \sum_{j=1}^{n^*} T_{j-1,j}$) y el estimador máximo-verosímil de ρ será $\hat{\rho} = \frac{n^*}{T_{n^*}}$. Ahora bien dado que, según se demostró en la página precedente, T_{n^*} se distribuye de acuerdo a una función Gamma y que $E\left(\frac{1}{T_{n^*}}\right) = \frac{\rho}{n^*-1}\left(n^* > 1\right)$, se obtiene que $E(\hat{\rho}) = \frac{n^*}{n^*-1}\rho$. A fin de corregir el sesgo del estimador máximo-verosímil se define, entonces, el estimador $\hat{\rho}^* = \frac{n^*-1}{n^*}\hat{\rho}$. Como T_{n^*} es un estadístico suficiente y completo para ρ, se puede concluir que $\hat{\rho}^*$ es un estimador insesgado de varianza mínima de ρ, tal que $\sigma^2(\hat{\rho}^*) = \frac{\rho^2}{n^*-2}\left(n^* > 2\right)$. Teniendo en cuenta que $2\hat{\rho}T_{n^*} : \chi^2_{(2n^*)}$, se puede, en este caso, construir un intervalo de confiabilidad de la forma $p\left(\frac{\chi_I^2}{2T_{n^*}} \le \rho \le \frac{\chi_S^2}{2T_{n^*}}\right) = 1-\alpha$, donde los valores χ_I^2 y χ_S^2 son tales que $p\left(\chi^2_{(2n^*)} \le \chi_I^2\right) = \frac{\alpha}{2}$ y $p\left(\chi^2_{(2n^*)} \le \chi_S^2\right) = 1-\frac{\alpha}{2}$.

3) La muestra se obtiene a partir de **observaciones realizadas de acuerdo a un diseño temporal específico**:

---

[127]. Ver Rao, C.R. (1973).

Supóngase que las observaciones se realicen en n^* momentos, $t_1, t_2, ..., t_{n^*}$, seleccionados según un criterio específico y tales que las variables $T_{j-1,j} = t_j - t_{j-1}\ (j = 1,2,...,n^*; t_0 = 0)$ sean observables e iid con distribución de probabilidades $f_T(t/\lambda)$ independiente del proceso de Poisson $\{Y(t)\}$.

Sean $Y(t_1), Y(t_2), ..., Y(t_{n^*})$ los estados observados del proceso en dichos puntos del tiempo, de modo que las variables $Z(t_j) = Y(t_j) - Y(t_{j-1})\ (j = 1,2,...,n^*)$ representan el número de ocurrencias del evento E en el intervalo $T_{j-1,j}$. Entonces, condicionada por $(t_1, t_2, ..., t_{n^*})$, la función de verosimilitud queda definida de la siguiente forma:

$$L = \prod_{j=1}^{n^{**}} \frac{(\rho T_{j-1,j})^{Z(t_j)} e^{-\rho T_{j-1,j}}}{Z(t_j)!} = \rho^{n^{**}} e^{-\rho t_{n^*}} \prod_{j=1}^{n^{**}} \frac{T_{j-1,j}^{Z(t_j)}}{Z(t_j)!}$$

(donde: **i)** $n^{**} = \sum_{j=1}^{n^*} Z(t_j)$ denota el número total de ocurrencias del evento E al momento t_{n^*} y **ii)** t_{n^*} denota el tiempo total de observación). Luego, se puede concluir que: **i)** el estimador máximo-verosímil de ρ, dado t_{n^*}, coincide con el obtenido de acuerdo con el diseño mencionado en el punto **1)**, $\hat{\rho} = \frac{n^*}{T_{n^*}}$ y **ii)** la distribución no-condicionada de $\hat{\rho}$ (igual que la distribución de T_{n^*}) depende de la distribución de n^*, de modo que las propiedades de $\hat{\rho}$ dependerán de la forma de la función de densidad $f_{T_{n^*}}(t/\lambda)$.

En particular si, como se vio, $f_{T_{j-1,j}}(t/\lambda) = \lambda e^{-\lambda t}\ (t > 0)$, entonces T_{n^*} se distribuye de acuerdo con una función Gamma de la forma $f_{T_{n^*}}(t/\lambda) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n^*-1}}{n^*-1}\ (t > 0)$ y, en consecuencia, se obtiene que $E(\hat{\rho}) = E\left[E(\rho/T_{n^*})\right] = \rho$ y $\sigma^2(\hat{\rho}) = E\left[\sigma^2(\rho/T_{n^*})\right] + \sigma^2\left[E(\rho/T_{n^*})\right] = E\left(\frac{\rho}{T_{n^*}}\right) = \frac{\lambda}{n^*-1}\rho$. Lo cual demuestra que la precisión del estimador $\hat{\rho}$ depende de la selección de λ. Obsérvese que, para n^* suficientemente grande, se verifica que $\sigma^2(\hat{\rho}) \approx \frac{\rho}{E(T_{n^*})}$.

4) La muestra se obtiene a partir de las **observaciones realizadas en momentos seleccionados al azar**: Supóngase que las observaciones se realicen en puntos aleatorios $t_1 < t_2 < ... < t_{n^*}$ y que los intervalos $T_{j-1,j} = t_j - t_{j-1}\ (j = 1,2,...,n^*; t_0 = 0)$ sean variables aleatorias no-observables, independientes y con distribución exponencial con parámetro λ. Sean $Y(t_1), Y(t_2), ..., Y(t_{n^*})$ los valores de las observaciones realizadas en los punto aleatorios $t_j\ (j = 1,2,...,n^*)$, de modo que las variables $Z(t_j) = Y(t_j) - Y(t_{j-1})$, que representan el número de ocurrencias del evento E en el intervalo $T_{j-1,j}$, son iid con función de densidad de la forma:

$$p\left[Z(t_j)=z\right]=\int_0^{\infty}\frac{(t\rho)^z}{z!}e^{-\rho t}e^{-\lambda t}\,dt=\left(1-\frac{\rho}{\lambda+\rho}\right)^{n^*}\left(\frac{\rho}{\lambda+\rho}\right)^{n^{**}}$$

donde $n^{**}=\sum_{j=1}^{n^*}Z(t_j)$ denota el número total de ocurrencias observadas del evento E. Entonces, el estimador máximo-verosímil de ρ será de la forma $\hat{\rho}=\frac{\lambda n^{**}}{n^*}$. Dado que n^{**} se distribuye de acuerdo con una función binomial-negativa con momentos de la forma $E(n^{**})=\frac{n^*\rho}{\lambda}$ y $\sigma^2(n^{**})=\frac{n^*\rho(\lambda+\rho)}{\lambda^2}$, se obtiene que $E(\hat{\rho})=\rho$ y que $\sigma^2(\hat{\rho})=\frac{\rho(\lambda+\rho)}{n^*}$.

1.7.2.2.- El proceso de Poisson compuesto

Sea un proceso $\{Y(t),t\geq 0\}$, continuo en el dominio del tiempo y discreto en el dominio de las variables, definido de la siguiente forma: $Y(t)=\sum_{j=1}^{N(t)}X_j$. Donde $\{N(t),t\geq 0\}$ denota un proceso de Poisson de la forma $\wp(\rho t)$ y $\{X_i\}$ $(i=1,2,...)$ denota una familia de variables aleatorias iid, siendo el proceso $\{N(t)\}$ y las variables $\{X_j\}$ independientes entre sí. Se dice, entonces, que $\{Y(t)\}$ es un **proceso de Poisson-compuesto.** Debe tenerse en cuenta que los "saltos" en $\{Y(t)\}$ ocurren simultáneamente con los "saltos" en $\{N(t)\}$ pero, mientras los "saltos" en $\{N(t)\}$ son de medida uno, la medida de los "saltos" en $\{Y(t)\}$ está representada por la variable X_i.

Como en el caso del proceso de Poisson homogéneo, el proceso definido por los incrementos de $\{Y(t)\}$, $Y(t)-Y(\tau)=\sum_{i=N(\tau)+1}^{N(t)}X_i$ $(0\leq\tau<t)$, está formado por variables iid con distribución similar a la de las variables $Y(t-\tau)$ [129].

La función generatriz de momentos del proceso $\{Y(t)\}$ es de la forma:

$$M_Y(w,t)=E\left[e^{Y(t)w}\right]=\sum_{n=0}^{\infty}E\left\{e^{Y(t)w}/\left[N(t)=n\right]\right\}e^{-\rho t}\frac{(\rho t)^n}{n!}$$

Donde, dada la supuesta independencia de las variables X_j $(j=1,2,...)$ y de $\{X_j\}$ con $\{N(t)\}$, será:

---

[129]. Debe tenerse en cuenta que $N(t)-N(\tau)\overset{d}{=}N(t-\tau)$.

$$E\left\{e^{Y(t)w} / \left[N(t) = n\right]\right\} = E\left\{\exp\left[w\left(X_1 + ... + X_n\right)\right] / \left[N(t) = n\right]\right\} =$$

$$= E\left\{\exp\left[w\left(X_1 + ... + X_n\right)\right]\right\} =$$

$$= \left[E\left(e^{wX}\right)\right]^n = \left[M_X(w)\right]^n$$

De modo que:

$$M_Y(w,t) = \sum_{n=0}^{\infty}\left[M_X(w)\right]^n e^{-\rho t}\frac{(\rho t)^n}{n!} = e^{-\rho t}\sum_{n=0}^{\infty}\frac{\left[(\rho t)M_X(w)\right]^n}{n!} = e^{\rho t\left[M_X(w)-1\right]}$$

A partir de este resultado, se obtiene que $E\left[Y(t)\right] = M_Y'(0,t) = \rho t\, E(X) = E\left[N(t)\right]E(X)$ (es decir que el número esperado de saltos de medida esperada $E(X)$ en el intervalo $[0,t]$, es ρt) y que $\sigma^2\left[Y(t)\right] = \rho t\, E\left(X^2\right)$.

Si las variables X_i no fueran aleatorias y asumieran el valor constante $X_i = x$, entonces el proceso de Poisson compuesto sería de la forma $Y(t) = xN(t)$ con funciones generatrices de momentos $M_X(w) = e^{wx}$ y $M_Y(w,t) = M_{xN}(w,t) = E\left[e^{wxN(t)}\right] = \exp\left[\rho t\left(e^{wx} - 1\right)\right]$ (obsérvese que si, en particular, $x = 1$, queda definida la función generatriz de momentos del proceso de Poisson homogéneo).

En particular, si se supone que las variables X_j^* que representan las medidas de los saltos, existen en un dominio finito $\Omega_{X^*} = \left\{x_1, x_2, ..., x_n\right\}\left(x_j \neq 0; j = 1,2,...,m\right)$, con función de probabilidades $p\left(X_j^* = x_j\right) = p_j\ \left(j = 1,2,...,m\right)$ y, en consecuencia, con función generatriz de momentos de la forma $M_{X^*}(w) = \sum_{j=1}^{m} p_j e^{wx_j}$, la función generatriz de momentos del proceso de Poisson compuesto $Y^*(t) = \sum_{j=1}^{m} x_j N_j(t)$ será de la forma:

$$M_{Y^*}(w,t) = \exp\left(\rho t\sum_{j=1}^{m} p_j e^{wx_j} - 1\right) = \exp\left[\rho t\sum_{j=1}^{m} p_j\left(e^{wx_j} - 1\right)\right] =$$

$$= \exp\left[\left(\rho p_1\right)t\left(e^{wx_1} - 1\right)\right]\exp\left[\left(\rho p_2\right)t\left(e^{wx_2} - 1\right)\right]...\left[\left(\rho p_m\right)t\left(e^{wx_m} - 1\right)\right] =$$

$$= \prod_{j=1}^{m}\exp\left\{\left(\rho p_j\right)t\left[M_{X_j^*}(w) - 1\right]\right\}$$

Es decir que la función generatriz de momentos del proceso $\left\{Y^*(t)\right\}$ queda expresada como el producto de m funciones generatrices de momentos correspondientes a procesos de Poisson con funciones intensidad de

ocurrencia $(\rho p_j)\,(j = 1,2,...,m)$ [130].

Con un razonamiento similar al empleado en la sección precedente, se demuestra que para el **proceso de Poisson compuesto compensado,** $\{Y(t) - \rho tE(X)\}$, se verifica que:

$$E\{[Y(t) - \rho tE(X)] / Y(\tau)\} = \rho(t - \tau)E(X) + Y(\tau) - \rho tE(X) = Y(\tau) - \rho\tau E(X)$$

es decir que $\{Y(t) - \rho t\}$ cumple la condición de la martingala (ver Sec. 4.1).

1.7.2.3.- El proceso de Poisson ponderado

Los procesos de Poisson analizados en las secciones precedentes pueden ser interpretados como representaciones de las frecuencias relativas de la ocurrencia de un evento E para un individuo cualquiera de una población infinita, considerando que el **coeficiente de riesgo de ocurrencia de E** (que podría ser interpretado, por ejemplo, como la **propensión a sufrir un accidente**), ρ, es constante para todos los individuos de la población. Una forma de generalizar estos procesos consiste en considerar que el riesgo de ocurrencia del evento E es variable para cada uno de los individuos observados, de acuerdo con una distribución de probabilidades $f(\rho)$. De modo que la probabilidad de que un individuo de la población seleccionado al azar experimente $Y\,(= 0,1,2,...)$ ocurrencias del evento E en un intervalo $[0,t]$, queda expresada de la siguiente forma:

$$p_Y(t) = \int_0^\infty p_{Y/\rho}(t) f(\rho) d\rho = \int_0^\infty \frac{(\rho t)^Y}{Y!} e^{-\rho t} f(\rho) d\rho$$

(donde $p_{Y/\rho}(t)$ denota la probabilidad de que un individuo sometido al riesgo de ocurrencia ρ, experimente Y ocurrencias del evento E).

Supóngase, por ejemplo, que la variable ρ se distribuya de acuerdo con una función Gammma de la forma:

$$f(\rho) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} \rho^{\alpha-1} e^{-\beta\rho} & (\rho > 0; \alpha > 0; \beta > 0) \\ 0 & (\rho \le 0) \end{cases}$$

Entonces, se verifica que[131], para cada t, la variable $Y(t)$ se distribuye de acuerdo con una función

---

[130]. A partir de este resultado se puede concluir que existen dos formas de definir un proceso de Poisson compuesto que admite un número finito de saltos: **i)** como un proceso de Poisson en el cual la medida unitaria de los saltos es reemplazada por saltos de medida aleatoria y **ii)** como una suma de procesos de Poisson independientes en cada uno de los cuales los saltos de medida unitaria son reemplazados por saltos de medida fija (ver Protter, P. (2004)).

[131]. Ver Landro, A.H. (2002).

$b^{-1}\left(\alpha;\frac{\beta}{\beta+t}\right)$:

$$p_{Y/\rho}(t) = \frac{\beta^{\alpha}t^{Y}}{\Gamma(\alpha)Y!}\int_{0}^{\infty}\rho^{Y+\alpha-1}e^{-\rho(\beta+t)}\,d\rho = \frac{\Gamma(Y+\alpha)}{Y!\Gamma(\alpha)}\beta^{\alpha}t^{Y}(\beta+t)^{-(Y+\alpha)} =$$

$$= \binom{Y+\alpha}{Y}\left(\frac{t}{\beta+t}\right)^{Y}\left(\frac{\beta}{\beta+t}\right)^{\alpha} \qquad (Y = 0,1,2,...)$$

con valor esperado $E[Y(t)] = \frac{\alpha t}{\beta}$ y varianza $\sigma^2[Y(t)] = \frac{\alpha t}{\beta}\left(1+\frac{t}{\beta}\right)$ (para un análisis de las aplicaciones de este proceso a problemas relacionados con la propensión a sufrir accidentes y con la morbilidad, ver Lundberg, O. (1940), Bates, G.E.; Neyman, J. (1952), Chiang, Ch.L. (1966)) .

1.7.2.4.- El proceso de Poisson generalizado (o no-homogéneo)

Otra generalización del proceso de Poisson se obtiene de sustituir la tasa ρ por una función intensidad de ocurrencia del evento E continua con respecto a t , $\rho(t)$ [132]. En ese caso, a partir del **sistema de ecuaciones "forward"** de Kolmogorov, se obtiene que la función generatriz de probabilidades del proceso obedece a la siguiente ecuación en derivadas parciales:

$$\frac{\partial}{\partial t}G_Y(w,t) = \rho(t)(w-1)G_Y(w,t)$$

Cuya solución es:

$$G_Y(w,t) = \sum_{y=0}^{\infty}w^{Y}p_y(t) = \exp\left[(w-1)\int_0^t\rho(\tau)d\tau\right] = \sum_{y=0}^{\infty}\frac{1}{y!}\left[\int_0^t\rho(\tau)d\tau\right]^{y}\exp\left[-\int_0^t\rho(\tau)d\tau\right]w^{y}$$

De lo que se concluye que $Y(t)$ se distribuye de acuerdo con una función de probabilidades de Poisson de la forma[133]:

---

[132]. Ver Sec. 1.7.2.1.

[133]. De acuerdo con la condición de Kolmogorov-Chapman, es posible hallar una función $f_{yx}(s,t) = \sum_{j=y}^{x}p_{yj}(s,\tau)p_{jx}(\tau,t)$ $(x>y;s<\tau<t)$ tal que $\frac{d}{dt}f_{yx}(s,t) = -\rho(t)p_{yx}(s,t) + \rho(t)p_{y,x-1}(s,t)$ $(x>y)$. Por otra parte, a partir de las soluciones de los sistemas "backward" y "forward", se obtiene que $p_{yy}(s,\tau)p_{yy}(\tau,t) = p_{yy}(s,t)$ y, por lo tanto, que $f_{yy}(s,t) = p_{yy}(s,t)$ y que $f_{yx}(s,t) = p_{yx}(s,t)$ $(x \geq y)$. Teniendo en cuenta que, de acuerdo con la definición que figura más arriba, $\frac{d}{ds}f_{yx}(s,t) = \sum_{j=y}^{x}\frac{d}{ds}p^{*}_{yj}(s,\tau)p_{jx}(\tau,t) =$ $= \rho(t)\left[p^{*}_{yx}(s,t) - p^{*}_{y+1,x}(s,t)\right]$, se puede concluir que $f_{yx}(s,t) = p^{*}_{yx}(s,t)$ $(x \geq y)$ y, en consecuencia, que también en este caso los dos sistemas de ecuaciones proporcionan como solución la misma función de probabilidades.

$$p_y(t) = \frac{1}{y!}\left[\int_0^t \rho(\tau)d\tau\right]^y \exp\left[-\int_0^t \rho(\tau)d\tau\right] \qquad (y = 0,1,2,...)$$

con valor esperado $E[Y(t)] = \int_0^t \rho(\tau)d\tau$.

De la misma forma que en la Sec. 1.6.2.1, las variables aleatorias $T_{i-1,i}$ $(j = 1,2,...)$, que representan el tiempo a transcurrir por el proceso entre la $(i-1)$-ésima y la i-ésima ocurrencia del evento E, son independientes con distribución de probabilidades exponencial de la forma:

$$f_{T_{i-1,i}}(t) = \rho(t_i)\exp\left(-\int_{i-1}^{i} \rho(\tau)d\tau\right) \qquad (i = 1,2,...)$$

Supóngase: **i)** que la intensidad de ocurrencia del evento E esté definida por una función exponencial de la forma $\rho(t) = c(\eta)\exp[h(\theta),k(t)]$ (donde θ y η denotan los parámetros de la distribución) y **ii)** que, a fin de estimar los verdaderos valores de θ y η, se construye una muestra obtenida a partir de la observación continua del proceso sobre un intervalo fijo $[0,T^*]$, entonces la función de verosimilitud será de la forma:

$$L = \left[\prod_{j=1}^{n} \rho(t_j)\right]\exp\left[-\int_0^{T^*} \rho(\tau)d\tau\right] = [c(\eta)]^n \exp\left[h(\theta)\sum_{j=1}^{n} k(t_j) - \int_0^{T^*} \rho(\tau)d\tau\right]$$

(donde: **i)** n denota el número de ocurrencias del evento E al momento T^* y **ii)** $t_1,t_2,...,t_n$ denotan los momentos en que se produjeron dichas ocurrencias). Obsérvese que: **i)** $\left(n;\sum_{j=1}^{n} k(t_j)\right)$ son estadísticos suficientes para (θ,η) ; **ii)** para un valor dado de θ, n es un estadístico suficiente para η y **iii)** para un valor dado de n, $\sum_{j=1}^{n} k(t_j)$ es un estadístico suficiente para θ .

Supóngase que se desee estimar el coeficiente θ, la distribución de probabilidades conjunta de la variable $(0 < t_1 < t_2 < ... < t_{n^*} \le T^*)$ condicionada por el supuesto $n = n^*$, será de la forma:

$$f(t_1,t_2,...,t_{n^*},T^*) = \frac{n^*!}{\left[\int_0^{T^*} \rho(\tau)d\tau\right]^{n^*}} \prod_{j=1}^{n^*} \rho(t_j)$$

y, por lo tanto, la función de verosimilitud condicionada será de la forma:

$$L^* = \frac{\prod_{j=1}^{n^*} \rho(t_j)}{\left[\int_0^{T^*} \rho(\tau)\,d\tau\right]} = \frac{\exp\left[h(\theta)\sum_{j=1}^{n^*} k(t_j)\right]}{\left[\int_0^{T^*} e^{h(\theta)k(\tau)}\,d\tau\right]^{n^*}}$$

(independiente del coeficiente η).

El caso más general de proceso de Poisson no-homogéneo es aquél en el que la **función intensidad de ocurrencia** del evento E depende de t y del estado del proceso al momento t, $\rho_Y(t)$. Se genera, entonces, un tipo de procesos que pueden ser resueltos utilizando argumentos similares a los del caso precedente, pero en los que las probabilidades de transición en el intervalo $(t, t+\Delta t]$ dependen del estado del proceso al momento t. El sistema de ecuaciones diferenciales-en diferencias de Kolmogorov, en este caso queda definido de la siguiente forma[134]:

$$\frac{d}{dt}p_Y(t) = -\rho_Y(t)p_Y(t) + \rho_{Y-1}(t)p_{Y-1}(t) \qquad (Y = 0,1,2,\ldots; t \geq 0)$$

Cuya solución es de la forma[135]:

$$p_0(t) = \exp\left[-\int_0^t \rho_0(\tau)\,d\tau\right]$$

$$p_Y(t) = \int_0^t \rho_{Y-1}(\tau)p_{y-1}(\tau)e^{\int_\tau^t \rho_y(s)ds}\,d\tau \qquad (Y = 1,2,\ldots; t \geq 0)$$

Sumando m. a m. el sistema de ecuaciones de Kolmogorov e integrando con respecto a t, se obtiene que $\sum_{j=0}^{y} p_j(t) - 1 = -\int_0^t \rho_y(\tau)p_y(\tau)d\tau$. Donde la integral que figura en el segundo miembro es no-creciente respecto a Y, y tal que $0 \leq \int_0^t \rho_y(\tau)p_y(\tau) < 1 \; (y \geq 0; t \geq 0)$, de modo que se puede concluir

---

[134]. En este caso, a cada estado le corresponde una función continua respecto de la variable t, $\rho_Y(t) \geq 0$, tal que $\lim_{\Delta t \to 0} \frac{1 - p_{YY}(t, t+\Delta t)}{\Delta t} = \rho_Y(t)$.

[135]. Obsérvese que, a partir de la ecuación correspondiente a $p_0(t)$, la función intensidad de ocurrencia puede ser expresada de la siguiente forma: $\rho_Y(t) = -\frac{\frac{d^{Y+1}}{dt^{Y+1}}p_0(t)}{\frac{d^Y}{dt^Y}p_0(t)} = \rho_{Y-1}(t) - \frac{1}{\rho_{Y-1}(t)}\frac{d\rho_{Y-1}(t)}{dt} \; (Y = 1,2,\ldots)$.

fácilmente que $0 < \sum_{j=0}^{y} p_j(t) \le 1 \,(y \ge 0; t \ge 0)$. Obsérvese que, cuando Y aumenta indefinidamente y el límite de la integral es igual a cero, esta suma tiende a 1. Supóngase que $\lim_{Y\to\infty} \int_0^t \rho_Y(\tau) p_Y(\tau) d\tau = \ell$, se verificará, entonces, que:

$$\int_0^t p_y(\tau)d\tau \ge \int_0^t \frac{\rho_y(\tau)}{\underset{0\le\tau\le t}{Max}\,\rho_y(t)} p_y(\tau)d\tau \ge \frac{\ell}{\underset{0\le\tau\le t}{Max}\,\rho_y(\tau)} \qquad (y = 0,1,2,...)$$

y, sumando m. a m. esta expresión con respecto a Y, se obtiene que $t \ge \int_0^t \sum_{y=0}^{\infty} p_y(\tau)d\tau \ge \sum_{y=0}^{\infty} \frac{\ell}{\underset{0\le\tau\le t}{Max}\,\rho_y(\tau)}$, de modo que, si se verifica que $\sum_{Y=0}^{\infty} \frac{1}{\underset{0\le\tau\le t}{Max}\,\rho_Y(\tau)} = \infty$, entonces, debe ser $\ell = 0$. A partir de estos resultados se puede concluir que la condición suficiente para que se verifique la igualdad $\sum_{Y=0}^{\infty} p_Y(t) = 1$ (y, por lo tanto, para poder asegurar que la solución del sistema de ecuaciones de Kolmogorov es única), es que la sucesión $\sum_{Y=0}^{\infty} \frac{1}{\underset{0\le\tau\le t}{Max}\,\rho_Y(\tau)}$ sea divergente para todo intervalo finito $[0,t]$. De la misma forma, a partir de la inecuación $\int_0^t \rho_Y(\tau) p_Y(\tau) d\tau < 1$ se puede escribir:

$$\int_0^t p_y(\tau)d\tau \le \int_0^t \frac{\rho_y(\tau)}{\min_{0\le\tau\le t}\rho_y(\tau)} p_y(\tau)d\tau < \frac{1}{\min_{0\le\tau\le t}\rho_y(\tau)} \qquad (y = 0,1,2,...)$$

Supóngase que exista un valor t_0 tal que $\sum_{Y=0}^{\infty} \frac{1}{\min_{0\le\tau\le t}\rho_Y(\tau)} < t_0$. Entonces, será:

$$\int_0^{t_0} \sum_{y=0}^{\infty} p_y(\tau)d\tau < \sum_{y=0}^{\infty} \frac{1}{\min_{0\le\tau\le t}\rho_y(\tau)} < t_0$$

Pero, si se verifica que $\sum_{Y=0}^{\infty} p_Y(\tau) = 1 \,(0 \le \tau \le t)$, se obtendrá que $t_0 < t_0$, de modo que se puede concluir que la condición necesaria para poder asegurar que $p_Y(t)$ es una función de probabilidades es que la sucesión $\sum_{Y=0}^{\infty} \frac{1}{\min_{0\le\tau\le t}\rho_Y(\tau)}$ sea divergente para todo intervalo finito $[0,t]$.

Por otra parte, se concluye fácilmente que la solución del sistema de ecuaciones de Kolmogorov-Chapman puede ser expresada de la siguiente forma:

$$p_{0Y}(t) = \int_0^{t'} \rho_Y(\tau) p_{Y-1}(\tau) e^{-\int_\tau^t \rho_Y(x)dx} d\tau + \int_{t'}^{t} \rho_{Y-1}(\tau) p_{Y-1}(\tau) e^{-\int_\tau^t \rho_Y(x)dx} d\tau \qquad (t > t' > 0)$$

Con relación al primer sumando del segundo miembro, se puede escribir:

$$\int_0^{t'} p_{Y-1}(\tau)\rho_Y(\tau) e^{-\int_\tau^t \rho_Y(x)dx} d\tau < e^{-\int_{t'}^t \rho_Y(x)dx} \int_0^{t'} p_{Y-1}(\tau)\rho_{Y-1}(\tau)d\tau < e^{-\int_{t'}^t \rho_Y(x)dx}$$

Si se supone que , para valores suficientemente grandes de t, $\int_0^\infty \rho_Y(t)dt = \infty$, se obtiene que $\exp\left(-\int_{t'}^t \rho_Y(x)dx\right) < \frac{\varepsilon}{2}$. Asimismo, si se tiene en cuenta que, para un valor dado de Y, $f(t) = \int_0^t p_{Y-1}(\tau)\rho_{Y-1}(\tau)d\tau$ $(t \geq 0)$ es una función no-decreciente con respecto a t y tal que $\lim_{t\to\infty} f(t) \leq 1$, siempre será posible hallar un valor t' tal que $\int_{t'}^\infty p_{Y-1}(\tau)\rho_{Y-1}(\tau)d\tau < \frac{\varepsilon}{2}$, de modo que:

$$\int_{t'}^{t} p_{Y-1}(\tau)\rho_{Y-1}(\tau) e^{-\int_\tau^t \rho_Y(x)dx} d\tau < \frac{\varepsilon}{2} \qquad (t > t')$$

Luego, se puede concluir que, si se supone que $\int_0^\infty \rho_Y(t)dt = \infty$, para valores suficientemente grandes de t, se verificará que $p_Y(t) < \varepsilon$. Es decir, queda demostrado que la condición necesaria y suficiente para que $\lim_{t\to\infty} p_Y(t) = 0$ (suponiendo que $p_Y(t)$ no sea idénticamente nula para todo $t > 0$), es que la integral $\int_0^\infty \rho_Y(t)dt$ sea divergente[135].

1.7.3.- El proceso de nacimiento-fallecimiento

1.7.3.1.- El proceso de nacimiento-fallecimiento con funciones intensidad dependientes del tamaño de la población

Sea $\{n(t)\}$ un proceso que representa el tamaño de una población al momento t . Supóngase que dicha población admita ingresos por nacimiento y egresos por fallecimiento, y que sus integrantes estén

---

[135]. En esta sección se han tratado solamente los procesos que admiten un número finito de "saltos" en un intervalo de tiempo finito.

sometidos, en forma individualmente independiente, a los mismos riesgos de "reproducción"[137] y de "fallecimiento" los cuales son, a su vez, independientes del comportamiento histórico de cada individuo y de la "edad" que representa la evolución del proceso. Supóngase que el tamaño inicial de la población sea igual a n_0, es decir, que:

$$p_{n_0,n}(0) = \begin{cases} 1 & para \quad n = n_0 \\ 0 & para \quad n \neq n_0 \end{cases}$$

Sean: **i)** $p\{[n(t+\Delta t) = n+1]/[n(t) = n]\} = p_{n,n+1}(t,t+\Delta t) = \lambda_n \Delta t + o(\Delta t)$ (donde λ_n denota la **función intensidad de nacimiento** en una población de tamaño n) la probabilidad de que una población pase del tamaño n en el momento t al tamaño $n+1$ en el momento $t+\Delta t$; **ii)** $p\{[n(t+\Delta t) = n-1]/[n(t) = n]\} = p_{n,n-1}(t,t+\Delta t) = \mu_n \Delta t + o(\Delta t)$ (donde μ_n denota la **función intensidad de fallecimiento** en una población de tamaño n) la probabilidad de que una población pase del tamaño n en el momento t al tamaño $n-1$ en el momento $t+\Delta t$; **iii)** $p\{[n(t+\Delta t) = n \pm k]/[n(t) = n]\} = p_{n,n\pm k}(t,t+\Delta t) = o(\Delta t)\ (k = 2,3,...)$ la probabilidad de que una población pase del tamaño n en el momento t al tamaño $n \pm k\ (k \geq 2)$ en el momento $t+\Delta t$ y **iv)** $p\{[n(t+\Delta t) = n]/[n(t) = n]\} = p_{nn}(t,t+\Delta t) = 1-(\lambda_n + \mu_n)\Delta t - o(\Delta t)$ la probabilidad de que una población de tamaño n en el momento t conserve el mismo tamaño al momento $t+\Delta t$.

La matriz de probabilidades de transición infinitesimal será, entonces, de la forma[138]:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

A partir de estas probabilidades , se obtiene que las probabilidades de transición para cualquier intervalo de tiempo finito $p_{n_0,n}(0,t) \equiv p_{n_0,n}(t)\ \left(n_0 \geq 0\,;\, n = 0,1,2,...;p_{n_0,n}(t) = 0 \ para\ todo\ n < 0\right)$ se comportan de acuerdo al siguiente sistema de ecuaciones diferenciales-en diferencias de Kolmogorov:

$$\frac{d}{dt}p_{n_0,n}(t) = -(\lambda_n + \mu_n)p_{n_0,n}(t) + \lambda_{n-1}p_{n_0,n-1}(t) + \mu_{n+1}p_{n_0,n+1}(t)$$

---

[137]. El término "reproducción" debe ser interpretado en este contexto como "subdivisión binaria", lo cual implica que cada individuo, al reproducirse, da origen a un nuevo individuo que se agrega a la población original. La hipótesis de independencia implica suponer que las subpoblaciones generadas por cada individuo evolucionan en forma independiente. Se supone, además, que todos los individuos son capaces de reproducción.

[138]. Esta matriz, conocida como **matriz Q de Doob**, está formada por los coeficientes de las ecuaciones "forward" y "backward" y constituye el elemento de transferencia entre los procesos determinístico y estocástico (ver Doob, J.L.(en Grenander, U. (ed.) (1959))).

De acuerdo con el teorema de Reuter-Lederman[139], se demuestra que una cualquiera de las condiciones que se enumeran a continuación es suficiente para poder asegurar que la solución $p_{n_0,n}(t)$ define una distribución de probabilidades y, en consecuencia, que es la solución única del sistema de ecuaciones de Kolmogorov [140]:

* Que , para algún $n \geq 1$, $\lambda_n = 0$.

* Que, para $n \geq m$, $\lambda_n = 0$ y que la serie:

$$\sum_{n=m}^{\infty}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_{n-1}\lambda_n}+\frac{\mu_{n-1}\mu_n}{\lambda_{n-2}\lambda_{n-1}\lambda_n}+\ldots+\frac{\mu_{m+1}\mu_{m+2}\cdots\mu_n}{\lambda_m\lambda_{m+1}\cdots\lambda_n}+\frac{\mu_m\mu_{m+1}\cdots\mu_n}{\lambda_m\lambda_{m+1}\cdots\lambda_n}\right)$$

sea divergente.

* Que , para $n \geq m$, $\lambda_n > 0$ y que la serie $\sum_{n=m}^{\infty}\lambda_n^{-1}$ sea divergente.

* Que, para $n \geq m$, $\lambda_n > 0$ y que la serie $\sum_{n=m}^{\infty}\left(\prod_{j=m}^{n}\frac{\mu_j}{\lambda_j}\right)$ sea divergente.

* Que, para $n \geq m' \geq m$, $\mu_n > 0$ y que $\lim_{n\to\infty}\left(\prod_{j=m'}^{n}\frac{\mu_j}{\lambda_j}\right) = \begin{cases}0\\ \\1\end{cases}$, suponiendo que $\lambda_n > 0$, para $n \geq m$.

Se demuestra, además, que, si $\lambda_n > 0$, para $n > m$, es suficiente que se cumpla alguna de las siguientes condiciones para que exista una probabilidad no-nula de que se produzca un número infinito de variaciones en la población, en un intervalo de tiempo finito[141]:

* que la serie $\sum_{n=m}^{\infty}\left(\frac{\mu_n}{\lambda_n}+\frac{\mu_n}{\lambda_{n-1}\lambda_n}+\ldots+\frac{\mu_{m+1}\cdots\mu_n}{\lambda_m\ldots\lambda_n}+\frac{\mu_m\cdots\mu_n}{\lambda_m\ldots\lambda_n}\right)$ sea convergente y que:

$$\mu_{n+1}\left(\frac{1}{\lambda_n}+\frac{\mu_n}{\lambda_{n-1}\lambda_n}+\ldots+\frac{\mu_{m+1}\cdots\mu_n}{\lambda_m\ldots\lambda_n}+\frac{\mu_m\cdots\mu_n}{\lambda_m\cdots\lambda_n}\right) = \begin{cases}0\\1\end{cases}$$

* que, para $n \geq m^j \geq m$, $\frac{\mu_{n+1}}{\lambda_n} \leq k < 1$ y que la serie $\sum_{n=1}^{\infty}\lambda_n^{-1}$ sea divergente.

---

[139]. Ver Reuter, G.E.H.; Lederman, W. (1953).

[140]. Este resultado se debe a Feller, W. (1940). Para un análisis más detallado de estas condiciones, ver Reuter, G.E.H.; Lederman, W. (1953), Bharucha-Reid, A.T. (1953)(1960), Koopman, B.O. (1975).

[141]. Es decir, para que se verifique que $\sum_{n} p_{n_0,n}(t) < 1$.

La condición necesaria y suficiente para que se verifique la propiedad de unicidad de la solución del sistema de ecuaciones de Kolmogorov, es que la serie $\sum_{n=0}^{\infty}\left(\frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}+\frac{\mu_1\mu_2\ldots\mu_n}{\lambda_0\lambda_1\ldots\lambda_n}\right)$ sea divergente.

Dado un proceso de nacimiento-muerte como el que se ha definido aquí, se dice que un estado n es **recurrente** si se verifica que la probabilidad de que el proceso, que se encuentra inicialmente en el estado n, retorne a dicho estado, es igual a la unidad. Es decir, si se verifica que $\int_0^{\infty} dF_{nn}(t)dt = 1$, donde,

$$F_{nn}(t) = p\left\{\left(\left[n(\tau_1)\neq n\right]\cap\left[n(\tau_2)=n\right], \textit{para algun par de valores } \tau_1,\tau_2 \textit{, tales que } 0<\tau_1<\tau_2\leq t\right)/\left[n(0)=n\right]\right\}$$

El estado n será **recurrente-positivo** o **recurrente-nulo** según que el **tiempo esperado de recurrencia** sea finito o infinito. En caso contrario, si se verifica que $\int_0^{\infty} dF_{nn}(t)dt < 1$, se dice que el estado n, es **no-recurrente.**

Como se mencionó en la Sec. 1.7.1.3, un proceso será recurrente-positivo, recurrente-nulo o no-recurrente, según que cada uno de sus estados sea, respectivamente, recurrente-positivo, recurrente-nulo o no-recurrente. Se demuestra que la condición necesaria y suficiente para que el proceso con funciones intensidad λ_n y μ_n sea recurrente es que la serie $1+\frac{\mu_1}{\lambda_0\lambda_1}+\frac{\mu_1\mu_2}{\lambda_0\lambda_1\lambda_2}+\frac{\mu_1\mu_2\ldots\mu_n}{\lambda_0\lambda_1\ldots\lambda_n}+\ldots$ sea divergente. Para que el proceso sea recurrente-positivo es necesario y suficiente que se verifique, además, que la serie $1+\frac{\lambda_0}{\mu_1}+\frac{\lambda_0\lambda_1}{\mu_1\mu_2}+\ldots+\frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}+\ldots$ sea convergente. En caso contrario, el proceso será recurrente-nulo.

Las condiciones necesarias y suficientes para que el proceso sea no-recurrente son las siguientes:

* que la serie $1+\frac{\mu_1}{\lambda_0\lambda_1}+\frac{\mu_1\mu_2}{\lambda_0\lambda_1\lambda_2}+\ldots+\frac{\mu_1\mu_2\ldots\mu_n}{\lambda_0\lambda_1\ldots\lambda_n}+\ldots$ sea convergente.

* que la serie $1+\frac{\lambda_0}{\mu_1}+\frac{\lambda_0\lambda_1}{\mu_1\mu_2}+\ldots+\frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}+\ldots$ sea divergente.

(para un mayor detalle acerca de las condiciones de recurrencia, ver Karlin, S.; McGregor, J. (1957)). Debe tenerse en cuenta que afirmar que un proceso es no-recurrente significa admitir que existe una probabilidad igual a uno de que la población alcance un tamaño infinito ($p\left[\lim_{t\to\infty} n(t) = \infty / n_0\right] = 1$). Será no-recurrente de **tipo 1** ó de **tipo 2** según que la probabilidad de alcanzar un tamaño infinito en un intervalo de tiempo finito sea, respectivamente, positiva o nula. En otros términos, se dice que el proceso es **no-recurrente de tipo 1**, si existe un estado m para el cual $\sum_{n=0}^{\infty} p_{m,n}(t) < 1\,(t>0)$. Por el contrario, si para todo m, se verifica que $\sum_{n=0}^{\infty} p_{m,n}(t) = 1\,(t>1)$, se dice que el proceso es **no-recurrente de tipo 2.** Esta última condición es

equivalente a $\sum_{n=0}^{\infty} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \sum_{j=n}^{\infty} \frac{\mu_1 \mu_2 \ldots \mu_j}{\lambda_0 \lambda_1 \ldots \lambda_j} = \infty$ ..

Dado un proceso recurrente, es posible demostrar que, cuando t aumenta indefinidamente, existe una **distribución de probabilidades de equilibrio** estacionaria, $\lim_{t \to \infty} p_{n_0,n}(t) = \pi_n$, independiente de n_0, que satisface las ecuaciones:

$$-(\lambda_n + \mu_n)\pi_n + \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1} = 0 \qquad (n = 0,1,2,\ldots)$$

Operando iterativamente en este sistema, se obtiene que $\pi_n = \frac{\prod_{j=0}^{n-1} \lambda_j}{\prod_{j=1}^{n} \mu_j} \pi_0 \; (n = 1,2,\ldots)$. A partir de esta expresión, se puede concluir en forma inmediata, que:

$$\sum_{n=1}^{\infty} \pi_n = \pi_0 \sum_{n} \frac{\prod_{j=0}^{n-1} \lambda_j}{\prod_{j=1}^{n} \mu_j}$$

En consecuencia, será:

$$\pi_n = \frac{\prod_{j=0}^{n-1} \lambda_j}{\prod_{j=1}^{n} \mu_j} \left(\sum_{n} \frac{\prod_{j=0}^{n-1} \lambda_j}{\prod_{j=1}^{n} \mu_j} \right)^{-1} \qquad (n = 1,2,\ldots)$$

Por el contrario, si el proceso es no-recurrente, entonces la serie que figura en el denominador de la expresión precedente será divergente y, por lo tanto $\pi_n = 0 \; (n = 1,2,\ldots)$. Es decir, no existe una distribución de probabilidades límite estacionaria.

Una derivación particularmente interesante de este proceso es aquélla que presenta estados absorbentes en $n(t) = 0$ y $n(t) = N$. En forma compacta, el correspondiente sistema de ecuaciones de Kolmogorov-Chapman puede ser expresado de la siguiente forma:

$$\frac{d}{dt} P(t) = P(t)Q$$

(con la condición inicial $P(0) = I_{(N+1)}$) donde: **i)** $P(t) = \left[p_{m,n}(t); m,n = 0,1,2,\ldots,N \right]$ y **ii)** Q es la matriz de Doob de probabilidades de transición infinitesimal:

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \dots & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -(\lambda_{N-1}+\mu_{N-1}) & \lambda_{N-1} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

La solución de este sistema de ecuaciones es de la forma:

$$P(t) = e^{-Q^* t} = I_{(N+1)} - \frac{t}{1!}Q^* + \frac{t^2}{2!}(Q^*)^2 - \frac{t^3}{3!}(Q^*)^3 + \dots$$

donde:

$$Q^* = -Q = \begin{bmatrix} 0 & 0' & 0 \\ -D & Q^{**} & -B \\ 0 & 0' & 0 \end{bmatrix}$$

$$D' = \begin{bmatrix} \mu_1 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$B' = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & \lambda_{N-1} \end{bmatrix}$$

y Q^{**} es la matriz de probabilidades de transición infinitesimal entre los estados no-recurrentes:

$$Q^{**} = \begin{bmatrix} \lambda_1+\mu_1 & -\lambda_1 & 0 & \dots & 0 & 0 \\ -\mu_2 & \lambda_2+\mu_2 & -\lambda_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\mu_{N-1} & \lambda_{N-1}+\mu_{N-1} \end{bmatrix}$$

(debe tenerse en cuenta que, si $\lambda_i, \mu_i > 0 \; (i = 1,2,\dots,N-1)$, entonces esta matriz es positiva-regular). La matriz Q^* es tal que:

$$(Q^*)^j = \begin{bmatrix} 0 & 0' & 0 \\ -(Q^{**})^{j-1}D & (Q^{**})^j & -(Q^{**})^{j-1}B \\ o & 0' & 0 \end{bmatrix}$$

Luego, será:

$$P(t) = \sum_{j=0}^{\infty}(-1)^j \frac{t^j}{j!}(Q^*)^j = \begin{bmatrix} 1 & 0' & 0 \\ -\sum_{j=1}^{\infty}(-1)^j \frac{t^j}{j!}(Q^{**})^j D & \sum_{j=1}^{\infty}(-1)^j \frac{t^j}{j!}(Q^{**})^j & -\sum_{j=1}^{\infty}(-1)^j \frac{t^j}{j!}(Q^{**})^{j-1}B \\ 0 & 0' & 1 \end{bmatrix}$$

Teniendo en cuenta que:

$$-\sum_{j=1}^{\infty}(-1)^j\frac{t^j}{j!}(Q^{**})^{j-1}D=-(Q^{**})^{-1}\left[\sum_{j=0}^{\infty}(-1)^j\frac{t^j}{j!}(Q^{**})^j-I_{(N-1)}\right]=(Q^{**})^{-1}\left(I_{(N-1)}-e^{-A^{**}t}\right)D$$

y que $-\sum_{j=1}^{\infty}(-1)^j\frac{t^j}{j!}(Q^{**})^{j-1}B=(Q^{**})^{-1}\left(I_{(N-1)}-e^{-A^{**}t}\right)B$, se obtiene que:

$$P(t)=\begin{bmatrix}1 & 0 & \cdots & 0\\ p_{01}(t) & p_{11}(t) & \cdots & p_{1N}(t)\\ \cdots & \cdots & \cdots & \cdots\\ p_{N-1,0}(t) & p_{N-1,1}(t) & \cdots & p_{N-1,N}(t)\\ 0 & 0 & \cdots & 1\end{bmatrix}=$$

$$=\begin{bmatrix}1 & 0' & 0\\ (Q^{**})^{-1}\left(I_{(N-1)}-e^{-A^{**}t}\right)D & e^{-Q^{**}t} & (Q^{**})^{-1}\left(I_{(N-1)}-e^{-Q^{**}t}\right)B\\ 0 & 0' & 1\end{bmatrix}$$

De modo que la probabilidad de que, al momento t , la población se haya extinguido, puede ser expresada de la siguiente forma:

$$P_0(t)=\begin{bmatrix}p_{10}(t)\\ p_{20}(t)\\ \cdots\\ p_{N-1,0}(t)\end{bmatrix}=(Q^{**})^{-1}\left[I_{(N-1)}-e^{-Q^{**}t}\right]D$$

Por otra parte, si se tiene en cuenta que $Q^{**}=\sum_{j=1}^{M}\delta_j c_j$, donde $\delta_1,\delta_2,\ldots,\delta_M$ denotan sus valores característicos y que: **i)** $c_j=\prod_{\substack{i=1\\ i\neq j}}^{M}\frac{1}{\delta_j-\delta_i}\left[Q^{**}-\delta_i I_{(N-1)}\right]$; **ii)** $\sum_{j=1}^{M}c_j=I_{(N-1)}$; **iii)** $c_j^2=c_j$ y **iv)** $c_i c_j=0\ \ (i\neq j)$, se puede concluir que $(Q^{**})^k=\sum_{j=1}^{M}\delta_j^k c_j$. De modo que se puede escribir:

$$P_0(t)=\sum_{j=1}^{M}\delta_j^{-1}\left(1-e^{-\delta_j t}\right)c_j D$$

En consecuencia, la probabilidad de extinción final de la población estará dada por:

$$\pi_0=\lim_{t\to\infty}P_0(t)=(Q^{**})^{-1}D=\sum_{j=1}^{M}\delta_j^{-1}c_j D$$

Teniendo en cuenta que el elemento correspondiente a la i -ésima fila y la j -ésima columna de la matriz

$(Q^{**})^{-1} = [\alpha_{ij}]\,(i,j = 1,2,\ldots,N-1)$ es de la forma $\alpha_{ij} = \dfrac{\left(\sum_{k=1}^{\min(i,j)} \prod_{h=1}^{k-1}\mu_h \prod_{m=k}^{j-1}\lambda_m\right)\left(\sum_{n=1}^{N(i,j)} \prod_{h=1}^{N-n}\mu_h \prod_{m=N-n+1}^{N-1}\lambda_m\right)}{\sum_{n=1}^{N}\prod_{h=1}^{N-n}\mu_h \prod_{m=N-n+1}^{N-1}\lambda_m}$ (donde:

i) $N(i,j) = (N-i) - \max(i-j,0)$ y **ii)** $\prod_h \mu_h = 1$, si $h < j$), la probabilidad de que un proceso que se encuentra en el estado n, se extinga, puede ser expresada de la siguiente forma:

$$\pi_{n,0} = \mu_1\alpha_{n1} = \frac{\sum_{m=1}^{N-n}\prod_{i=1}^{N-m}\mu_i \prod_{j=N-m+1}^{N-1}\lambda_j}{\sum_{m=1}^{N}\prod_{i=1}^{N-n}\mu_i \prod_{j=N-m+1}^{N-1}\lambda_j} \qquad (n = 1,2,\ldots,N-1)$$

Con un razonamiento similar, se obtiene: **i)** la probabilidad de que, al momento t, el proceso haya sido absorbido por el estado N:

$$P_N(t) = \begin{bmatrix} P_{1N}(t) \\ P_{2N}(t) \\ \cdots \\ P_{N-1,N}(t) \end{bmatrix} = (Q^{**})^{-1}\left[I_{(N-1)} - e^{-Q^{**}t}\right]B = \sum_{j=1}^{M}\delta_j^{-1}\left(1 - e^{-\delta_j t}\right)c_j B$$

(donde $\{\delta_1,\delta_2,\ldots,\delta_M\}$ denota el conjunto de los valores característicos de la matriz Q^{**}); **ii)** la probabilidad de absorción final por el estado N:

$$\pi_N = \lim_{t\to\infty} P_N(t) = (Q^{**})^{-1}B = \sum_{j=1}^{M}\delta_j^{-1}c_j B$$

y **iii)** la probabilidad de que, si el proceso se encuentra en el estado n, sea absorbido por el estado N:

$$\pi_{n,N} = \lambda_{N-1}\alpha_{n,N-1} = \frac{\sum_{m=1}^{n}\prod_{i=1}^{m-1}\mu_i \prod_{j=m}^{N-1}\lambda_j}{\sum_{m=1}^{N}\prod_{i=1}^{N-m}\mu_i \prod_{j=N-m+1}^{N-1}\lambda_j} \qquad (n = 1,2,\ldots,N-1)$$

A partir de estos resultados, se obtiene que:

1) La probabilidad de absorción final del proceso (es decir, la probabilidad de extinción o de absorción por los estados 0 ó N) está dada por $\pi_0 + \pi_N = (Q^{**})^{-1}(D+B)$ y como $D+B = Q^{**}I_{(N-1)}$ (donde $I_{(N-1)}$ es un vector columna), se puede concluir que $\pi_0 + \pi_N = I_{(N-1)}$. Es decir, se puede asegurar que un proceso de nacimiento-fallecimiento con barrerass absorbentes en los estados 0 y N, que se encuentra inicialmente en un estado $n\,(= 1,2,\ldots,N-1)$, finalmente será absorbido.

2) La condición necesaria y suficiente para poder asegurar que, para un proceso que se encuentra en el

estado n, la probabilidad de absorción final por el estado N es mayor que la probabilidad de extinción (es decir que $\pi_{n,0} < \pi_{n,N}$) es que:

$$\sum_{m=n+1}^{N}\prod_{i=1}^{m-1}\mu_i\prod_{j=m}^{N-1}\lambda_j < \sum_{m=1}^{n}\prod_{i=1}^{m-1}\mu_i\prod_{j=m}^{N-1}\lambda_j \qquad (n=1,2,\ldots,N-1)$$

3) Denotando por:

$$f_{n,0}(t,t+\Delta t) = f_{n,0}(t)\Delta t = \sum_{m=1}^{N-1} p_{nm}(t) f_{m,0}(t)\Delta t$$

$$f_{n,N}(t,t+\Delta t) = f_{n,N}(t)\Delta t = \sum_{m=1}^{N-1} p_{nm}(t) f_{m,n}(t)\Delta t$$

a las probabilidades de que el proceso, que se encuentra en el estado n en el momento t, sea absorbido por los estados 0 ó N, respectivamente, en el intervalo $(t,t+\Delta t]$ [142], se puede escribir $p_{n0}(t+\Delta t) = p_{n0}(t) + f_{n,0}(t)\Delta t$ y $p_{nN}(t+\Delta t) = p_{nN}(t) + f_{n,N}(t)\Delta t$. Es decir $f_{n,0}(t) = \frac{d}{dt}p_{n0}(t)$ y $f_{n,N}(t) =$ $= \frac{d}{dt}p_{nN}(t)$. En consecuencia, las probabilidades de absorción por los estados 0 ó N al momento t quedan definidas, respectivamente, por los vectores $f_0(t) = \frac{d}{dt}P_0(t)$ y $f_N(t) = \frac{d}{dt}P_N(t)$ y la probabilidad total de absorción al momento t asume la forma:

$$f_{(0\cup N)}(t) = \frac{d}{dt}\left[P_0(t) + P_N(t)\right] = \frac{d}{dt}\left\{(Q^{**})^{-1}\left[I_{(N-1)} - e^{-Q^{**}t}\right]D + (Q^{**})^{-1}\left[I_{(N-1)} - e^{-Q^{**}t}\right]B\right\} =$$

$$= \frac{d}{dt}\left\{(Q^{**})^{-1}\left[I_{(N-1)} - e^{-Q^{**}t}\right](D+B)\right\} =$$

$$= \frac{d}{dt}\left\{(Q^{**})^{-1}\left[I_{(N-1)} - e^{-Q^{**}t}\right]Q^{**}I_{(N-1)}\right\} =$$

$$= \frac{d}{dt}\left\{\left[-\sum_{j=1}^{\infty}(-1)^j\frac{t^j}{j!}(Q^{**})^{j-1}\right]Q^{**}I_{(N-1)}\right\} =$$

$$= \frac{d}{dt}\left[(Q^{**})^{-1}\left(-e^{-Q^{**}t}\right)Q^{**}I_{(N-1)}\right] =$$

$$= \frac{d}{dt}\left[-e^{-Q^{**}t}I_{(N-1)}\right] = Q^{**}e^{-Q^{**}t}I_{(N-1)} = \sum_{j=1}^{M}\delta_j e^{-\delta_j t}c_j I_{(N-1)}$$

---

142. Obsérvese que, de acuerdo a las definiciones que figuran en la Sec. 1.7.1.3, $f_{m,n}(t)$ representa la probabilidad de que el proceso que se encuentra inicialmente en el estado m, arribe por primera vez al estado n en el momento t.

4) Como corolario del resultado precedente y denotando por $T_{(0\cup N)}$ a la variable **tiempo a transcurrir por el proceso hasta su absorción** y por $T_{(0\cup N)/n}\ (n = 1,2,...,N-1)$ a la variable **tiempo a transcurrir hasta su absorción para un proceso que se encuentra en el estado n**, el **tiempo esperado de absorción** queda definido por:

$$E\left(T_{(0\cup N)}\right) = \begin{bmatrix} E\left(T_{(0\cup N)/1}\right) \\ \cdots \\ E\left(T_{(0\cup N)/N-1}\right) \end{bmatrix} = \int_0^\infty t\, f(t)dt =$$

$$= \int_0^\infty t\, d\left[-e^{-Q^{**}t} I_{(N-1)}\right] = \int_0^\infty e^{-Q^{**}t} I_{(N-1)}\, dt = \left(Q^{**}\right)^{-1} I_{(N-1)} =$$

$$= \sum_{j=1}^{M} \delta_j^{-1} c_j I_{(N-1)}$$

Luego, será:

$$E\left(T_{(0\cup N)/n}\right) = \frac{\sum_{s=1}^{N-1}\left(\sum_{m=1}^{\min(s,t)}\prod_{i=1}^{m-1}\mu_i\prod_{j=m}^{s-1}\lambda_j\right)\left(\sum_{m=1}^{N(s,n)}\prod_{i=s+1}^{N-m}\mu_i\prod_{j=N-m+1}^{N-1}\lambda_j\right)}{\sum_{n=1}^{N}\prod_{i=1}^{N-m}\mu_i\prod_{j=N-m+1}^{N-1}\lambda_j} \qquad (n = 1,2,...,N-1)$$

(donde $N(s,n) = N - s - \max(n-s,0)$).

Por su parte, la varianza de la variable $T_{(0\cup N)}$ será de la forma:

$$\sigma^2\left(T_{(0\cup N)}\right) = \int_0^\infty t^2 f(t)dt - \left[E\left(T_{(0\cup N)}\right)\right]^{(2)} =$$

$$= \int_0^\infty t^2\left[Q^{**}e^{-Q^{**}t} I_{(N-1)}\right]dt - \left[E\left(T_{(0\cup N)}\right)\right]^{(2)} =$$

$$= 2\left(Q^{**}\right)^{-1} E\left(T_{(0\cup N)}\right) - \left[E\left(T_{(0\cup N)}\right)\right]^{(2)} = 2\left(Q^{**}\right)^{-1} I_{(N-1)} - \left[E\left(T_{(0\cup N)}\right)\right]^{(2)} =$$

$$= 2\sum_{j=1}^{M} \delta_j^{-1} c_j I_{(N-1)} - \left[E\left(T_{(0\cup N)}\right)\right]^{(2)}$$

donde $\left[E\left(T_{(0\cup N)}\right)\right]^{(2)} = \begin{bmatrix} \left[E\left(T_{(0\cup N)/1}\right)\right]^2 \\ \cdots \\ \left[E\left(T_{(0\cup N)/N-1}\right)\right]^2 \end{bmatrix}$.

Si, en particular, el proceso presenta un único estado absorbente en $n(t)=0$, las probabilidades de absorción final satisfacen el siguiente sistema de ecuaciones:

$$\begin{cases} \pi_{n,0} = \dfrac{\lambda_n}{\lambda_n+\mu_n}\pi_{n+1,0} + \dfrac{\mu_n}{\lambda_n+\mu_n}\pi_{n-1,0} & (n=1,2,...) \\ \pi_{0,0} = 1 \end{cases}$$

A partir del cual se obtiene que:

$$\pi_{n+1,0} - \pi_{n,0} = \frac{\mu_n}{\lambda_n}\left(\pi_{n,0} - \pi_{n-1,0}\right) = \frac{\mu_n\mu_{n-1}}{\lambda_n\lambda_{n-1}}\left(\pi_{n-1,0} - \pi_{n-2,0}\right) =$$

$$= ... = \frac{\mu_n\mu_{n-1}\cdots\mu_1}{\lambda_n\lambda_{n-1}\cdots\lambda_1}\left(\pi_{1,0} - 1\right) \qquad (n=1,2,3,...)$$

y, por lo tanto, que:

$$\sum_{j=1}^{n-1}\left(\pi_{j+1,0} - \pi_{j0}\right) = \pi_{n,0} - \pi_{1,0} = \left(\pi_{10} - 1\right)\sum_{j=1}^{n-1}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j} \qquad (n=1,2,...)$$

Entonces, si se verifica que $\lim_{n\to\infty}\sum_{j=1}^{n}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j} = \infty$, se obtiene que $\pi_{n,0}=1$ $(n=1,2,...)$. Asimismo, si $\pi_{1,0}<1$, será $\lim_{n\to\infty}\sum_{j=1}^{n}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j} < \infty$ y, $\lim_{n\to\infty}\pi_{n,0} = 0 = \pi_{1,0}\left(\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}+1\right) - \sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}$, es decir, se verificará que $\pi_{1,0} = \dfrac{\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}}{\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}+1}$. En consecuencia, la probabilidad de extinción para un proceso que se encuentra en el estado n será de la forma:

$$\pi_{n0} = \frac{\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}}{\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda\ \ldots\lambda_j}}\left(\sum_{j=1}^{n-1}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}+1\right) - \sum_{j=1}^{n-1}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j} = \frac{\sum_{j=n}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}}{\sum_{j=1}^{\infty}\frac{\mu_1\cdots\mu_j}{\lambda_1\cdots\lambda_j}+1} \qquad (n=1,2,...)$$

Según se vio, la variable "tiempo a transcurrir por el proceso en el estado n" está distribuida de acuerdo con la función $f_{T_{n,n}}(t) = \left(\lambda_n+\mu_n\right)e^{-(\lambda_n+\mu_n)t}$ $(n=1,2,3,...)$. De modo que el tiempo esperado a transcurrir por el proceso en el estado n es $E\left(T_{n,n}\right) = \dfrac{1}{\lambda_n+\mu_n}$ $(n=1,2,3,...)$. El tiempo esperado para que el proceso, que se encuentra en el estado n, pase al estado $n+1$, está definido por:

$$E\left(T_{n,n+1}\right) = \frac{\lambda_n}{\lambda_n + \mu_n}\frac{1}{\lambda_n + \mu_n} + \frac{\mu_n}{\lambda_n + \mu_n}\left[\frac{1}{\lambda_n + \mu_n} + E\left(T_{n-1,n}\right) + E\left(T_{n,n+1}\right)\right] =$$

$$= \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n}E\left(T_{n-1,n}\right) \qquad (n = 1,2,...)$$

Es decir, resulta que:

$$E\left(T_{12}\right) = \frac{1}{\lambda_1}$$

$$E\left(T_{23}\right) = \frac{1}{\lambda_2} + \frac{\mu_2}{\lambda_1\lambda_2}$$

.....

$$E\left(T_{n,n+1}\right) = \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_{n-1}\lambda_n} + ... + \frac{\mu_2 ... \mu_n}{\lambda_2 ... \lambda_n}$$

De la misma forma, el tiempo esperado hasta la extinción en una población de tamaño n, está dado por:

$$E\left(T_{n,0}\right) = \frac{1}{\lambda_n + \mu_n} + \frac{\lambda_n}{\lambda_n + \mu_n}E\left(T_{n+1,0}\right) + \frac{\mu_n}{\lambda_n + \mu_n}E\left(T_{n-1,,0}\right) \qquad (n = 1,2,...)$$

Luego, será:

$$\frac{\lambda_n}{\lambda_n + \mu_n}\left[E\left(T_{n,0}\right) - E\left(T_{n+1,0}\right)\right] = \frac{1}{\lambda_n + \mu_n}\left[E\left(T_{n-1,0}\right) - E\left(T_{n,0}\right)\right]$$

y, por lo tanto:

$$E\left(T_{n,0}\right) - E\left(T_{n+1,0}\right) = \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_n + \mu_n}\left[E\left(T_{n-1,0}\right) - E\left(T_{n,0}\right)\right] =$$

$$= \frac{1}{\lambda_n} + \frac{\mu_n}{\lambda_{n-1}\lambda_n} + \frac{\mu_{n-1}\mu_n}{\lambda_{n-1}\lambda_n}\left[E\left(T_{n-2,0}\right) - E\left(T_{n-1,0}\right)\right] = ... =$$

$$= \left(\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j}\right)\sum_{i=1}^{n}\frac{\lambda_1 ... \lambda_{i-1}}{\mu_1 ... \mu_i} - \left(\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j}\right)E\left(T_{1,0}\right)$$

(donde, por convención, $T_{0,0} = 0$). En consecuencia, se puede concluir que: **i)** si $\sum_{i=1}^{\infty}\frac{\lambda_1 ... \lambda_{i-1}}{\mu_1 ... \mu_i} = \infty$, entonces se verifica que $E\left(T_{1,0}\right) = \infty$ y **ii)** si esta sucesión es convergente, se verifica que:

$$E\left(T_{1,0}\right) = \sum_{j=1}^{\infty}\frac{\lambda_1 ... \lambda_{i-1}}{\mu_1 ... \mu_i} - \lim_{n\to\infty}\left(\prod_{j=1}^{n}\frac{\lambda_j}{\mu_j}\right)\left[E\left(T_{n,0}\right) - E\left(T_{n+1,0}\right)\right] = \sum_{i=1}^{\infty}\frac{\lambda_1 ... \lambda_{i-1}}{\mu_1 ... \mu_i}$$

(se demuestra que $\lim_{n\to\infty}\left(\prod_{j=1}^{n}\frac{\lambda_j}{\mu_j}\right)\left[E\left(T_{n,0}\right) - E\left(T_{n+1,0}\right)\right] = 0$).

1.7.3.2.- El proceso puro de nacimiento con función intensidad dependiente del tamaño de la población

Si en el proceso de nacimiento-fallecimiento definido en la sección precedente se verifica que $\mu_n = 0\ (n = 1,2,...)$, queda definido un **proceso puro de nacimiento** cuya solución es de la forma[143]:

$$p_{n_0,n}(t) = (-1)^{n-n_0} \prod_{j=n_0}^{n-1} \lambda_j \left[\sum_{i=n_0}^{n} \frac{e^{-\lambda_i t}}{\prod_{\substack{h=n_0 \\ h \neq i}}^{n} (\lambda_i - \lambda_h)} \right] =$$

$$= (-1)^{n-n_0} \prod_{j=n_0}^{n-1} \lambda_j \frac{\begin{vmatrix} 1 & \lambda_{n_0} & \dots & \lambda_{n_0}^{n-n_0-1} & e^{-\lambda_{n_0} t} \\ 1 & \lambda_{n_0+1} & \dots & \lambda_{n_0+1}^{n-n_0-1} & e^{-\lambda_{n_0+1} t} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \lambda_n & \dots & \lambda_n^{n-n_0-1} & e^{-\lambda_n t} \end{vmatrix}}{V_n} \qquad (\lambda_i \neq \lambda_h; i \neq h; i,h = n_0, n_0+1,...)$$

donde V_n denota el determinante de Vandermonde de orden n:

$$V_n = \begin{vmatrix} 1 & \lambda_{n_0} & \lambda_{n_0}^2 & \dots & \lambda_{n_0}^n \\ 1 & \lambda_{n_0+1} & \lambda_{n_0+1}^2 & \dots & \lambda_{n_0+1}^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \lambda_n & \lambda_n^2 & \dots & \lambda_n^n \end{vmatrix}$$

Como se mencionó en la sección precedente, es necesario determinar si la sucesión $\{p_{n_0,n}(t); n = n_0, n_0+1, ...; t \geq 0\}$ define una distribución de probabilidades, es decir, si se verifica que $\sum_{n=n_0}^{\infty} p_{n_0,n}(t) = 1\ (t \geq 0)$ o si, por el contrario $\sum_{n=n_0}^{\infty} p_{n_0,n}(t) < 1\ (t \geq 0)$ (obsérvese que, en este caso, la diferencia $1 - \sum_{n=n_0}^{\infty} p_{n_0,n}(t) > 0$ representa la probabilidad de que se produzca un número infinito de nacimientos en un intervalo finito, es decir, que el proceso sea **de crecimiento divergente** o **explosivo**).

Sea la suma parcial $S_n(t) = \sum_{j=n_0}^{n} p_{n_0 j}(t)$, se verificará, entonces, que:

$$\frac{d}{dt} S_n(t) = \sum_{j=n_0}^{n} \frac{d}{dt} p_{n_0 j}(t) = -\lambda_{n_0} p_{n_0 n_0}(t) - \lambda_{n_0+1} p_{n_0,n_0+1}(t) + \lambda_{n_0} p_{n_0 n_0}(t) -$$

---

[143]. Esta solución fue obtenida por McKendrick, A.G. (1914) y posteriormente generalizada por Yule, G.U. (1924) y Lundberg, O. (1940).

$$-\lambda_{n_0+2}p_{n_0,n_0+2}(t)+\lambda_{n_0+1}p_{n_0,n_0+1}(t)-\ldots-\lambda_n p_{n_0 n}(t)+\lambda_{n-1}p_{n_0,n-1}(t)=$$
$$=-\lambda_n p_{n_0 n}(t)=\lambda_n\left[S_n(t)-S_{n-1}(t)\right]$$

(donde $S_n(t)=0$, para $n<n_0$). Aplicando la transformación de Laplace m. a m. a esta expresión, se obtiene que:

$$L\left\{\frac{d}{dt}S_n(t)\right\}=sL\{S_n(t)\}-S_n(0)=sL\{S_n(t)\}-1=-\lambda_n L\{S_n(t)\}+\lambda_n L\{S_{n-1}(t)\}$$

es decir, que:

$$L\{S_{n_0}(t)\}=\frac{1}{s+\lambda_{n_0}}<\frac{1}{\lambda_{n_0}}$$

$$L\{S_n(t)\}=\frac{1+\lambda_n L\{S_{n-1}(t)\}}{s+\lambda_n}<\frac{1}{\lambda_n}+L\{S_{n-1}(t)\}<\sum_{j=n_0}^{n}\frac{1}{\lambda_j}\qquad (s>0;n>n_0)$$

De modo que $\lim_{n\to\infty}L\{S_n(t)\}=\lim_{n\to\infty}L\left\{\sum_{j=n_0}^{n}p_{n_0,j}(t)\right\}<\sum_{j=n_0}^{\infty}\frac{1}{\lambda_j}$. Luego, para que se verifique que $\sum_{n=n_0}^{\infty}p_{n_0,n}(t)=1\,(t\geq 0)$ se debe verificar que $\lim_{n\to\infty}S_n(t)=1\,(t\geq 1)$ o, lo que es lo mismo, que $\lim_{n\to\infty}L\{S_n(t)\}=\frac{1}{s}$. Es decir, debe ser $s\geq\frac{1}{\sum_{n=n_0}^{\infty}\lambda_n}$ y, por lo tanto, $\sum_{n=n_0}^{\infty}\lambda_n^{-1}=\infty$. En caso contrario, para valores de:

$$s<\frac{1}{\sum_{n=n_0}^{\infty}\lambda_n}$$

se daría la siguiente contradicción: $\lim_{n\to\infty}L\{S_n(t)\}>\sum_{n=n_0}^{\infty}\lambda_n^{-1}$. Luego, se puede concluir que la condición necesaria y suficiente para que $\{p_{n_0,n}(t)\}$ $(n=n_0,n_0+1,\ldots;t\geq 0)$ pueda ser considerada una distribución de probabilidades es que $\sum_{n=n_0}^{\infty}\lambda_n^{-1}$ sea divergente (este resultado se conoce como teorema de Feller-Lundberg, ver Feller, W. (1940), Lundberg, O. (1940)). Teniendo en cuenta que λ_n^{-1} define el tiempo que se espera ha de transcurrir el proceso en el estado $(n-1)$, que la serie $\sum_{nn_0}^{\infty}\lambda_n^{-1}$ sea divergente implica que el tiempo esperado total que el proceso ha de transcurrir en los estados $n_0,n_0+1,\ldots,\sum_{n=n_0}^{\infty}E\left(T_{n-1,n}\right)$ es infinito (es lógico suponer que, si las intensidades de nacimiento λ_n crecen demasiado rápidamente,

existirá una probabilidad no-nula de que la población alcance un tamaño infinito en un intervalo de tiempo finito).

El proceso de nacimiento de crecimiento divergente más simple es aquél en el cual $\lambda_n = n^2\lambda$ (conocido como **proceso de colisión molecular**, debido a Volterra, V. (1938)). La probabilidad de que ocurran infinitos nacimientos en un intervalo de tiempo finito está dada, en este caso, por una función **Theta de Jacobi** de la forma $\Theta_4\left(0;e^{-\lambda t}\right) = 1 - 2\sum_{n=1}^{\infty}(-1)^n e^{-n^2\lambda t}$.

1.7.3.3.- El proceso nacimiento-fallecimiento lineal

En particular, si en el proceso de nacimiento-fallecimiento se supone que las funciones intensidad de nacimiento y fallecimiento son lineales con respecto al tamaño de la población, $\lambda_n = n\lambda$ y $\mu_n = n\mu$ [143], la matriz Q de probabilidades de transición infinitesimal será de la forma:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \ldots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \ldots \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & 0 & \ldots \\ 0 & 0 & 3\mu & -3(\lambda+\mu) & 3\lambda & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

(obsérvese que, en este caso, considerando a los eventos "nacimiento" y "fallecimiento" como los dos resultados posibles de un fenómeno binomial, la probabilidad de que ocurra un nacimiento será $p = \frac{\lambda}{\lambda+\mu}$ y la probabilidad de que ocurra un fallecimiento será $1-p = \frac{\mu}{\lambda+\mu}$) y las probabilidades de transición satisfacen el siguiente sistema de ecuaciones diferenciales-en diferencias de Kolmogorov:

$$\frac{d}{dt}p_{n_0,n}(t) = -n(\lambda+\mu)p_{n_0,n}(t) + (n-1)p_{n_0,n-1}(t) + (n+1)p_{n_0,n+1}(t)$$

($n_0 \geq 0; n = 0,1,2,\ldots; p_{n_0,n}(t) = 0$ *para todo* $n < 0$), a partir del cual se obtiene la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades de las variables $n(t)$:

$$\frac{\partial}{\partial t}G_n(w,t) - (\lambda w - \mu)(w-1)\frac{\partial}{\partial w}G_n(w,t) = 0$$

cuya solución particular, para $\lambda \neq \mu$, es de la forma:

$$G_n(w,t) = \left[\frac{\mu\left(e^{\rho t}-1\right)+\left(\lambda-\mu e^{\rho t}\right)w}{\left(\lambda e^{\rho t}-\mu\right)+\lambda\left(1-e^{\rho t}\right)w}\right]^{n_0} = \left[\frac{a(t)+b(t)w}{c(t)+d(t)w}\right]^{n_0} = \left[\frac{\alpha(t)+\beta(t)w}{1-\gamma(t)w}\right]^{n_0} =$$

---

[143]. Este proceso se debe a Feller, W. (1939).

$$= \left[\alpha(t)^{n_0} + \binom{n_0}{1} \alpha(t)^{n_0-1} \beta(t) w + \binom{n_0}{2} \alpha(t)^{n_0-2} \beta(t)^2 w^2 + \ldots + \beta(t)^{n_0} w^{n_0} \right] \bullet$$

$$\bullet \left[1 + \binom{n_0}{1} \gamma(t) w + \binom{n_0+1}{2} \gamma(t)^2 w^2 + \binom{n_0+2}{3} \gamma(t)^3 w^3 + \ldots \right] =$$

$$= \sum_{n=0}^{n_0} \alpha(t)^{n_0-n} \left\{ \sum_{m=0}^{n} \binom{n_0}{n-m} \binom{n_0+m-1}{m} \left[\alpha(t)\gamma(t) \right]^m \beta(t)^{n_0-m} \right\} w^n +$$

$$+ \sum_{n=n_0+1}^{\infty} \gamma(t)^{n-n_0} \left\{ \sum_{m=0}^{n_0} \binom{n-m}{n_0} \binom{n+m-n_0}{n+m-1} \left[\alpha(t)\gamma(t) \right]^m \beta(t)^{n_0-m} \right\} w^n \qquad (\lambda \neq \mu)$$

(donde $\rho = \lambda - \mu$ denota la **tasa instantánea de reproducción**). De modo que las probabilidades de transición quedan definidas de la siguiente forma:

$$p_{n_0,n}(t) = \begin{cases} \alpha(t)^{n_0-n} \left\{ \sum_{m=0}^{n} \binom{n_0}{n-m} \binom{n_0+m-1}{n_0-1} \left[\alpha(t)\gamma(t) \right]^m \beta(t)^{n-m} \right\} & (n \leq n_0) \\ \gamma(t)^{n-n_0} \left\{ \sum_{m=0}^{n_0} \binom{n_0}{n-m} \binom{n+m-1}{n_0-1} \left[\alpha(t)\gamma(t) \right]^m \beta(t)^{n-m} \right\} & (n > m) \end{cases}$$

Para $\lambda = \mu$ la solución compacta obtenida para la función generatriz de probabilidades asume la forma:

$$\lim_{\lambda \to \mu} G_n(w,t) = \left[\frac{\mu t + (1 - \mu t) w}{(1 + \mu t) - \mu t w} \right]^{n_0}$$

A partir de estos resultados, se puede concluir que la probabilidad de extinción de esta población al momento t, será:

$$p_{n_0,0}(t) = G_n(0,t) = \begin{cases} \left[\frac{\mu \left(e^{\rho t} - 1 \right)}{\lambda e^{\rho t} - \mu} \right]^{n_0} & (\lambda \neq \mu) \\ \left(\frac{\mu t}{1 + \mu t} \right)^{n_0} & (\lambda = \mu) \end{cases}$$

Por otra parte, se demuestra fácilmente que:

$$\lim_{t \to \infty} G_n(w,t) = \begin{cases} \left(\frac{\mu}{\lambda} \right)^{n_0} & (\lambda > \mu) \\ 1 & (\lambda \leq \mu) \end{cases}$$

Por lo que se puede concluir, inmediatamente, que la probabilidad de extinción final de esta población

($\lim_{t\to\infty} p_{n_0,0}(t) = \pi_{n_0,0} = \lim_{t\to\infty} G_n(0,t)$) tiende a la certeza, si $\lambda \leq \mu$.

Obsérvese que, con un razonamiento similar al empleado en la Sec. 1.7.2.1, se puede concluir que la variable tiempo a transcurrir por el proceso en el estado n_0 (T_{n_0}) (es decir, la variable que representa el tiempo a transcurrir hasta la primera ocurrencia de un nacimiento o fallecimiento) se distribuye de acuerdo con una función exponencial con parámetro $n_0(\lambda+\mu)$.

Un método para estimar las tasas λ y μ consiste en construir una muestra a partir de la observación continua del proceso $\{n(t)\}$ sobre un intervalo $[0,T^*]$. Sean $t_1 < t_2 < ... < t_m$ los puntos del tiempo en los cuales se produjeron transiciones (es decir, en los cuales se produjeron nacimientos o fallecimientos). La función de verosimilitud será, entonces, de la forma:

$$L = \{n_0\xi_0 \exp[-n_0(\lambda+\mu)t_1]\}\{n_1\xi_1 \exp[-n_1(\lambda+\mu)t_2]\}...$$

$$...\{n_{m-1}\xi_{m-1} \exp[-n_{m-1}(\lambda+\mu)t_m]\} \exp\left[-n_m(\lambda+\mu)\left(T^* - \sum_{j=1}^{m} t_j\right)\right] =$$

$$= \prod_{j=0}^{m-1} n_j \prod_{h=0}^{m-1} \xi_h \exp[-(\lambda+\mu)S_{T^*}]$$

(donde: **i)** $\xi_j = \lambda$ si la $(j+1)$-ésima transición se debe a un nacimiento y $\xi_j = \mu$ si la $(j+1)$-ésima transición se debe a un fallecimiento y **ii)** $S_{T^*} = \sum_{j=1}^{m} n_{j-1}t_j + n_m\left(T^* - \sum_{j=1}^{m} t_j\right) = \int_0^{T^*} n(t)dt$). Si no se toman en consideración los términos que no involucran a los parámetros λ y μ, el logaritmo de la función de verosimilitud puede ser expresada de la siguiente forma:

$$\ell = -(\lambda+\mu)S_{T^*} + N_{T^*}\log(\lambda) + M_{T^*}\log(\mu)$$

(donde N_{T^*} y M_{T^*} denotan, respectivamente, el número de nacimientos y fallecimientos ocurridos en el intervalo $[0,T^*]$). Luego, los estimadores máximo-verosímiles de las intensidades λ y μ serán $\hat{\lambda} = \frac{N_{T^*}}{S_{T^*}}$ y $\hat{\mu} = \frac{M_{T^*}}{S_{T^*}}$.

Dado que el proceso de nacimiento-fallecimiento, con $n(0) = n_0$, analizado en esta sección puede ser considerado como la suma de n_0 procesos de nacimiento-fallecimiento con $n(0) = 1$ y, de acuerdo con los postulados del teorema central del límite, para $n_0 \to \infty$, se verificará que[145]:

---

[145]. Ver Morán, P.A.P. (1953), Billingsley, P. (1961)(1968), Keiding, N. (1975), Prakasa Rao, B.L.S. (1974)(1978), Basawa, I.V.; Scott, D.J. (1977), Basawa, I.V.; Prakasa Rao, B.L.S. (1980).

$$\sqrt{n_0}\begin{bmatrix}\hat{\lambda}-\lambda\\ \hat{\mu}-\mu\end{bmatrix}\xrightarrow{d}N_2\left(\begin{bmatrix}0\\0\end{bmatrix};\begin{bmatrix}\sigma_\lambda^2 & 0\\ 0 & \sigma_\mu^2\end{bmatrix}\right)$$

donde: **i)** $N_2(\bullet)$ denota una distribución Normal bidimensional; **ii)** $\sigma_\lambda^2=\dfrac{\lambda(\lambda-\mu)}{e^{(\lambda-\mu)T^*}-1}$ y **iii)** $\sigma_\mu^2=$

$$=\frac{\mu(\lambda-\mu)}{e^{(\lambda-\mu)T^*}-1}.$$

Se demuestra, además, que para n_0 fijo, cuando $T^*\to\infty$, los estimadores máximo-verosímiles λ y μ son fuertemente consistentes.

Si en este proceso de nacimiento-fallecimiento lineal se verifica que $\mu=0$, queda definido un **proceso puro de nacimiento lineal** o **proceso de Yule-Furry**[146] con matriz Q de probabilidades de transición infinitesimal de la forma:

$$\begin{bmatrix}0 & 0 & 0 & 0 & 0 & \ldots\\ 0 & -\lambda & \lambda & 0 & 0 & ..\\ 0 & 0 & -2\lambda & 2\lambda & 0 & \ldots\\ 0 & 0 & 0 & -3\lambda & 3\lambda & \ldots\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\end{bmatrix}$$

y función generatriz de probabilidades de la forma:

$$G_n(w,t)=\left[\frac{we^{-\lambda t}}{1-w\left(1-e^{-\lambda t}\right)}\right]^{n_0}=e^{-n_0\lambda t}\sum_{n=n_0}^{\infty}\binom{n-1}{n-n_0}\left(1-e^{-\lambda t}\right)^{n-n_0}w^n$$

Luego, se puede concluir que las variables $n(t)$ se distribuyen de acuerdo con una función $b^{-1}\left(n_0,e^{-\lambda t}\right)$ de la forma:

$$p_{n_0,n}(t)=\binom{n-1}{n-n_0}e^{n_0\lambda t}\left(1-e^{-\lambda t}\right)^{n-n_0}\quad\left(n=n_0,n_0+1,\ldots;n_0=1,2,\ldots;t\geq 0\right)$$

A fin de estimar el coeficiente λ se pueden considerar los siguientes diseños del experimento

---

[146]. Este proceso fue propuesto por Yule, G.U. (1924) en su estudio de la teoría matemática de la evolución, para representar la multiplicación de especies en un ámbito determinado (según la nomenclatura de Yule, $n\lambda\Delta t$ denota la probabilidad de que, en un ámbito en el que existen n especies, se genere por mutación una nueva especie en el intervalo $\left(t,t+\Delta t\right]$). El modelo se basa en los supuestos que las especies no pueden "morir", excepto por la ocurrencia de un cataclismo (circunstancia no considerada en las hipótesis iniciales) y que no existen limitaciones por restricciones en el entorno al crecimiento del número de individuos de cada especie, ni interferencias o competencias entre las distintas especies. En forma independiente, Furry, W.H. (1937) desarrolló un modelo de este tipo para describir la multiplicación de partículas en una lluvia de rayos cósmicos. Posteriormente, Arley, N. (1948) propuso, con el mismo fin, un modelo de nacimiento-fallecimiento con intensidad de nacimiento constante e intensidad de fallecimiento directamente proporcional al tiempo, $\mu(t)=kt$ (donde k denota una constante positiva), para representar la disminución de la energía de cada partícula a causa de sus colisiones con otras partículas.

muestral:

1) La muestra se obtiene a partir de **observaciones realizadas en puntos del tiempo fijos y equidistantes**[147]: Sea $\left[0,T^*\right]$ el período fijo total de observación. Supóngase que este intervalo sea dividido en n^* partes iguales de medida $h = \frac{T^*}{n^*}$ y que se realicen observaciones sobre el proceso $\{n(t)\}$ (con $n(0) = 1$) en los puntos $0, h, 2h,..., n^*h = T^*$. Sean $\sim\sim$ los estados observados del proceso en los momentos $t = jh\,(j = 0,1,2,...,n^*)$. Entonces, si no se toman en consideración los términos que no involucran al coeficiente λ, el logaritmo de la función de verosimilitud puede ser expresado de la siguiente forma:

$$\ell = -\sum_{j=1}^{n^*} n_{(j-1)h}\lambda h + \sum_{j=1}^{n^*}\left(n_{jh} - n_{(j-1)h}\right)\log\left(1 - e^{-\lambda h}\right)$$

de modo que el estimador máximo-verosímil queda definido por $\hat{\lambda} = \frac{1}{h}\log\left(\frac{\sum_{j=1}^{n^*} n_{jh}}{\sum_{j=0}^{n^*-1} n_{jh}}\right)$ o, lo que es lo mismo, por $\exp\left(\hat{\lambda}h\right) = \frac{\sum_{j=1}^{n^*} n_{jh}}{\sum_{j=0}^{n^*-1} n_{jh}}$.

Si se supone un tamaño inicial de la población $n(0) = m$, entonces el proceso $\{n(t)\}$ puede ser considerado como la suma de m procesos de nacimientos independientes con $n(0) = 1$.

Denotando por $n_{jh}^{(s)}$ $(s = 1,2,...,m)$ a una observación obtenida del s-ésimo proceso con $n(0) = 1$, se obtiene el siguiente estimador máximo-verosímil del coeficiente λ_m: $\exp\left(\hat{\lambda}_m h\right) = \frac{\sum_{s=1}^{m}\sum_{j=1}^{n^*} n_{jh}^{(s)}}{\sum_{s=1}^{m}\sum_{j=0}^{n^*-1} n_{jh}^{(s)}}$. Para n^* fijo y cuando m aumenta indefinidamente, se obtiene que:

$$\sqrt{m}\left(\hat{\lambda} - \lambda\right) \xrightarrow{d} N\left\{0; \left[\frac{senh\left(\frac{\lambda h}{2}\right)}{\frac{\lambda h}{2}}\right]^2 \left[\frac{\lambda^2}{e^{n^*h\lambda} - 1}\right]\right\}$$

---

147. Este método se debe a Kendall, D.G. (1949).

2) La muestra se obtiene a partir de una **observación continua sobre un período fijo**: Supóngase, en particular, que $n(0) = 1$. Sea $[0, T^*]$ el período total fijo de observación, sea $n^* - 1$ el número de nacimientos ocurridos en dicho intervalo, sean $t_1 < t_2 < ... < t_{n^*-1} \leq T^*$ los puntos del tiempo en que se produjeron los nacimientos y sean $T_{j-1,j} = t_j - t_{j-1}\left(j = 1,2,...,n^*\right)$ las n^* variables que representan el tiempo a transcurrir por el proceso entre el $(j-1)$-ésimo y el j-ésimo nacimiento. Entonces, la función de verosimilitud de la muestra será de la forma:

$$L = \prod_{j=1}^{n^*-1} j\lambda \exp\left(-j\lambda T_{j-1,j}\right) \exp\left[-\lambda n^*\left(T^* - t_{n^*-1}\right)\right]$$

Si no se toman en consideración los términos que no involucran al coeficiente λ, el logaritmo de la función de verosimilitud puede ser expresado de la siguiente forma:

$$\ell = (n^* - 1)\log(\lambda) - \lambda \sum_{j=1}^{n^*-1} jT_{j-1,j} - n^*\lambda T^* + n^*\lambda t_{n^*-1}$$

De modo que el estimador máximo-verosímil de λ será de la forma:

$$\hat{\lambda} = \frac{n^* - 1}{\sum_{j=1}^{n^*-1} jT_{j-1,j} + n^*\left(T^* - t_{n^*-1}\right)} = \frac{n^* - 1}{\int_0^{T^*} n(t)\,dt}$$

(debe tenerse en cuenta que el denominador del segundo miembro en esta expresión es igual al área comprendida bajo la curva $n(t)$ en el intervalo $[0, T^*]$).

En general, si el proceso $\{n(t)\}$ es tal que $n(0) = n_0$, el estimador máximo-verosímil de λ será de la forma:

$$\hat{\lambda}_{n_0} = \frac{n^* - n_0}{\int_0^{T^*} n(t)\,dt}$$

Para un valor de T^* fijo, como el considerado en este caso, cuando $n_0 \to \infty$, se verifica que:

$$\sqrt{n_0}\left(\hat{\lambda}_{n_0} - \lambda\right) \xrightarrow{d} N\left(0, \sigma^{*2}\left(\hat{\lambda}\right)\right)$$

donde, como en el caso precedente, $\sigma^{*2}\left(\hat{\lambda}\right)$ denota la varianza asintótica de $\hat{\lambda}$ la cual, haciendo $n^* \to \infty, h \to 0$ y $n^* h \to T^*$, puede ser expresada de la siguiente forma: $\sigma^{*2}\left(\hat{\lambda}\right) = \frac{\lambda^2}{e^{\lambda T^*} - 1}$.

La medida de la eficiencia relativa de este método respecto del anterior está dada por:

$$\frac{\sigma^{*2}\left(\hat{\lambda}\right)}{\sigma^2\left(\hat{\lambda}\right)} = \left[\frac{\frac{\lambda h}{2}}{senh\left(\frac{\lambda h}{2}\right)}\right]^2$$

Se demuestra que, para $n_0 = 1$ y $T^* \to \infty$: **i)** $\hat{\lambda} \xrightarrow{ccc} \lambda$; **ii)** $\sqrt{S_{T^*}}\left(\hat{\lambda} - \lambda\right) \xrightarrow{d} N(0,\lambda)$ (donde $S_{T^*} = \int_0^{T^*} n(t)\,dt$) y **iii)** $\sqrt{E\left(S_{T^*}\right)}\left(\hat{\lambda} - \lambda\right) \xrightarrow{d} t_{(2)}$.

3) La muestra se obtiene a partir de la **observación continua sobre un período aleatorio**[148]: Supóngase que: **i)** la observación del proceso $\{n(t)\}$ (con $n(0) = 1$) comience en $t = 0$ y continúe hasta que se produzcan n^* nacimientos y **ii)** los nacimientos ocurren en los puntos $t_1 < t_2 < ... < t_{n^*}$. La función de verosimilitud será, entonces, de la forma $L = \prod_{j=1}^{n^*} j\lambda \exp\left(- j\lambda T_{j-1,j}\right)$ (donde $T_{j-1,j} = t_j - t_{j-1}$) y, por lo tanto, el estimador máximo-verosímil de λ estará dado por $\hat{\lambda} = \frac{n^*}{\sum_{j=1}^{n^*} jT_{j-1,j}}$. El denominador de esta expresión es una variable aleatoria distribuida de acuerdo con una función Gamma con función característica de la forma $C(w) = \left(1 - \frac{1}{\lambda} iw\right)^{-n}$, a partir de la cual se obtiene que $E\left(\hat{\lambda}\right) = \frac{n^*}{n^* - 1}\lambda$. De modo que un estimador insesgado de λ puede ser obtenido de la siguiente forma:

$$\hat{\lambda}^* = \frac{n^* - 1}{n^*}\lambda = \frac{n^* - 1}{\sum_{j=1}^{n^*} jT_{j-1,j}}$$

cuya varianza será $\sigma^2\left(\hat{\lambda}^*\right) = \frac{\lambda^2}{n^* - 2}\left(n^* > 2\right)$.

Si en el proceso de nacimiento-fallecimiento lineal se verifica que $\lambda = 0$, queda definido un **proceso puro de fallecimiento lineal**, cuya función generatriz de probabilidades es de la forma:

$$G_n(w,t) = \left[\left(1 - e^{-\mu t}\right) + we^{-\mu t}\right]^{n_0} = \sum_{n=0}^{n_0} \binom{n_0}{n} e^{-n\mu t}\left(1 - e^{-\mu t}\right)^{n_0 - n} w^n$$

De modo que se puede concluir que las variables $n(t)$ se distribuyen de acuerdo con una función $b\left(n_0, e^{-\mu t}\right)$ de la forma $p_{n_0,n}(t) = \binom{n_0}{n}\left(e^{-\mu t}\right)^n\left(1 - e^{-\mu t}\right)^{n_0 - n}\left(n = 0,1,2,...,n_0\right)$.

---

[148]. Este método se debe a Morán, P.A.P. (1951).

Una derivación particularmente interesante del proceso de nacimiento-fallecimiento lineal es aquél en el que el proceso presenta, además del estado $n(t) = 0$, un segundo estado absorbente en $n(t) = N$.

En ese caso, la matriz Q de probabilidades de transición infinitesimal es de la forma:

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & \dots & 0 & 0 & 0 \\ 0 & 2\mu & -2(\lambda+\mu) & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & (N-1)\mu & -(N-1)(\lambda+\mu) & (N-1)\lambda \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}$$

y la distribución de probabilidades correspondiente a la variable tamaño de la población satisface el siguiente sistema de ecuaciones de Kolmogorov:

$$\begin{cases} \dfrac{d}{dt}p_{n_0,0}(t) = \mu\, p_{n_0,1}(t) \\ \dfrac{d}{dt}p_{n_0,1}(t) = -(\lambda+\mu)p_{n_0,1}(t) + 2\mu\, p_{n_0,2}(t) \\ \dfrac{d}{dt}p_{n_0,n}(t) = -n(\lambda+\mu)p_{n_0,n}(t) + (n-1)\lambda\, p_{n_0,n-1}(t) + (n+1)p_{n_0,n+1}(t) \\ \dfrac{d}{dt}p_{n_0,N-1}(t) = -(N-1)(\lambda+\mu)p_{n_0,N-1}(t) + (N-2)\lambda\, p_{n_0,N-2}(t) \\ \dfrac{d}{dt}p_{n_0,N}(t) = -(N-1)\lambda\, p_{n_0,N-1}(t) \end{cases}$$

a partir del cual y considerando la condición inicial $G_n(w,t) = \sum_{n=0}^{N} w^n p_{n_0,n}(0) = w^{n_0}$, se obtiene la siguiente solución para la función generatriz de probabilidades:

$$G_n(w,t) = -N\rho^2(1-w)(\lambda-\mu w)\int_0^t p_{n_0,N}(\tau)e^{\rho(t-\tau)}\frac{\left\{\mu\left[1-e^{\rho(t-\tau)}\right]-\left[\lambda-\mu e^{\rho(t-\tau)}\right]w\right\}^{N-1}}{\left\{\left[\mu-\lambda e^{\rho(t-\tau)}\right]-\lambda\left[1-e^{\rho(t-\tau)}\right]w\right\}^{N+1}}d\tau +$$

$$+\left[\frac{\mu\left(1-e^{\rho t}\right)-\left(\lambda-\mu e^{\rho t}\right)w}{\left(\mu-\lambda e^{\rho t}\right)-\lambda\left(1-e^{\rho t}\right)w}\right]^{n_0}$$

Denotando por $f_{n,N}(t)$ la probabilidad de que un proceso que se encuentra inicialmente en el estado n sea absorbido por el estado N en el intervalo $(0,t]$, de acuerdo a lo ya demostrado se obtiene, entonces, el siguiente sistema de ecuaciones "backward":

$$\begin{cases} \dfrac{d}{dt} f_{n,N}(t) = n(\lambda+\mu) f_{n,N}(t) + n\mu f_{n-1,N}(t) + n\lambda f_{n+1,N}(t) & (n = 1,2,\ldots,N-1) \\ f_{0,N}(t) = f_{N,N}(t) = 0 \end{cases}$$

A partir del cual queda definida la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades:

$$\frac{\partial}{\partial t} G_{f_N}(w,t) + \left[\mu w^2 - (\lambda+\mu)w + \lambda\right] \frac{\partial}{\partial w} G_{f_N}(w,t) = (\lambda+\mu-2\mu w) G_{f_N}(w,t) + N\mu w^{N-1} f_{N-1,N}(t)$$

cuya solución particular es de la forma:

$$G_{f_N}(w,t) = -N\mu \int_0^t f_{N-1,N}(\tau) e^{\rho(t-\tau)} \frac{\left\{\lambda\left[1-e^{-\rho(t-\tau)}\right] - \left[\mu-\lambda e^{-\rho(t-\tau)}\right]w\right\}^{N-1}}{\left\{\left[\lambda-\mu e^{-\rho(t-\tau)}\right] - \mu\left[1-e^{-\rho(t-\tau)}\right]w\right\}^{N+1}} d\tau +$$

$$+ (N+1)\lambda\rho^2 e^{-\rho t} \frac{\left[\lambda\left(1-e^{-\rho t}\right) - \left(\mu-\lambda e^{-\rho t}\right)w\right]^{N-2}}{\left[\left(\lambda-\mu e^{-\rho t}\right) - \mu\left(1-e^{-\rho t}\right)w\right]^{N}} \qquad (\lambda \neq \mu)$$

Desarrollando esta expresión en serie de potencias de w y, para valores suficientemente grandes de N, se obtiene que:

$$f_{nN}(t) = \frac{\lambda\left(1-\frac{\mu}{\lambda}\right)}{(n-1)!} e^{-\rho t} \frac{\left(1-e^{-\rho t}\right)^{N-n-1}}{\left(1-\frac{\mu}{\lambda}\right)^{N+n}} \sum_{k=0}^{n-1} \left(1-\frac{\mu}{\lambda}\right)^{2k} \left(\frac{\mu}{\lambda}\right)^{n-k-1} e^{-k\rho t} \sum_{j=k}^{n-1} (-1)^{n+j-1} \cdot$$

$$\cdot \left[\left(1-\frac{\mu}{\lambda}\right) - \frac{n}{N+j}\left(1-\frac{\mu}{\lambda}e^{-\rho t}\right)\right] \binom{n-1}{j} \binom{j}{k} n! \left[\sum_{m=1}^{j} \binom{j}{m}\binom{N}{n-m}\right] \cdot$$

$$\cdot \sum_{h=0}^{n-k-1} \sum_{\ell=0}^{n-k-1} (-1)^{h+\ell} \binom{n-k-1}{h} \binom{n-k-1}{\ell} e^{-(h+\ell)\rho t} \left(\frac{\mu}{\lambda}\right)^{-h+\ell} \qquad (N >> 1; n = 1,2,\ldots,N-1)$$

Otra derivación del proceso de nacimiento-fallecimiento lineal, de aparición bastante frecuente en las aplicaciones, se presenta cuando se supone que la población inicial no está dada por un valor fijo, sino que es una variable aleatoria con una distribución de Poisson con valor esperado n_0. En este caso, la condición inicial de la ecuación en derivadas parciales correspondiente a la función generatriz de probabilidades, será $G_n(w,0) = \exp\left[n_0(w-1)\right]$, de modo que su solución particular asumirá la forma:

$$G_n(w,t) = \exp\left[\frac{n_0(\lambda-\mu)(w-1)e^{\rho t}}{\lambda(w-1)e^{\rho t} - (\lambda w - \mu)}\right]$$

1.7.3.4.- El proceso de nacimiento-fallecimiento con funciones intensidad dependientes del tiempo y del tamaño de la población[149]

Supóngase un proceso de nacimiento-fallecimiento como el definido en la Sec. 1.7.3.1 pero con funciones intensidad de nacimiento y fallecimiento $\lambda_n(t)$ y $\mu_n(t)$ que, además de depender del tamaño de la población al momento t, sean funciones arbitrarias del tiempo.

El sistema de probabilidades de transición infinitesimal será de la forma:

$$p_{n,n+1}(t,t+\Delta t) = \lambda_n(t)\Delta t + o(\Delta t)$$
$$p_{n,n-1}(t,t+\Delta t) = \mu_n(t)\Delta t + o(\Delta t)$$
$$p_{n,n\pm k}(t,t+\Delta t) = o(\Delta t) \qquad (k = 2,3,...)$$
$$p_{n,n}(t,t+\Delta t) = 1-\left[\lambda_n(t)+\mu_n(t)\right]\Delta t - o(\Delta t)$$

A partir de estas expresiones y de acuerdo con la condición de Kolmogorv-Chapman, se obtiene un sistema de ecuaciones en diferencias de la forma:

$$p_n(t+\Delta t) = p_n(t)\left[1-\lambda_n(t)\Delta t-\mu_n(t)\Delta t\right]+p_{n-1}(t)\lambda_{n-1}(t)\Delta t + p_{n+1}(t)\mu_{n+1}(t)\Delta t + o(\Delta t)$$

del cual resulta el siguiente sistema de ecuaciones de Kolmogorov:

$$\begin{cases} \dfrac{d}{dt}p_0(t) = -\left[\lambda_0(t)+\mu_0(t)\right]p_0(t)+\mu_1(t)p_1(t) \\ \dfrac{d}{dt}p_n(t) = -\left[\lambda_n(t)+\mu_n(t)\right]p_n(t)+\lambda_{n-1}(t)p_{n-1}(t)+\mu_{n+1}(t)p_{n+1}(t) \qquad (n = 1,2,...) \end{cases}$$

Obviamente la solución particular de este sistema depende de la forma que asuman las funciones $\lambda_n(t)$ y $\mu_n(t)$. Si se supone, por ejemplo, que como en el caso tratado en la Sec. 1.7.3.3, las probabilidades de transición infinitesimal de nacimiento y fallecimiento son lineales con respecto al tamaño de la población, la matriz de probabilidades de transición infinitesimal es de la forma:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & ... & 0 \\ \mu(t) & -\left[\lambda(t)+\mu(t)\right] & \lambda(t) & 0 & ... & 0 \\ 0 & 2\mu(t) & -2\left[\lambda(t)+\mu(t)\right] & 2\lambda(t) & ... & 0 \\ 0 & 0 & 3\mu(t) & -3\left[\lambda(t)+\mu(t)\right] & ... & 3\lambda(t) \\ ... & ... & ... & ... & ... & ... \end{bmatrix}$$

y las probabilidades de transición obedecen al siguiente sistema de ecuaciones diferenciales-en diferencias de Kolmogorov:

$$\frac{d}{dt}p_{n_0,n}(t) = -n\left[\lambda(t)+\mu(t)\right]p_{n_0,n}(t)+(n-1)\lambda(t)p_{n_0n-1}(t)+(n+1)\mu(t)p_{n_0n}(t) \qquad (n = 0,1,2,...;t \geq 0)$$

---

[149]. La solución de este caso, en el que las intensidades de nacimiento y fallecimiento son funciones arbitrarias del tiempo, se debe a Kendall, D.G. (1948).

a partir del cual se obtiene la siguiente función generatriz de probabilidades:

$$G_n(w,t) = \left[\frac{w-1}{e^{-\rho(t)} - (w-1)\int_0^t \lambda(\tau) e^{-\rho(\tau)}\, d\tau} + 1 \right]^{n_0} = \left[\frac{a(t) + b(t)w}{c(t) - d(t)w} \right]^{n_0} = \left[\frac{\alpha(t) + \beta(t)w}{1 - \gamma(t)w} \right]^{n_0}$$

donde: **i)** $\rho(t) = \lambda(t) - \mu(t)$ denota la función intensidad de reproducción; **ii)** $a(t) = e^{-\rho(t)} - \left[1 - \int_0^t \lambda(\tau) e^{-\rho(\tau)}\, d\tau \right]$; **iii)** $b(t) = 1 - \int_0^t \lambda(\tau) e^{-\rho(\tau)}\, d\tau$; **iv)** $c(t) = e^{-\rho(t)} + \int_0^t \lambda(\tau) e^{-\rho(\tau)}\, d\tau$; **v)** $d(t) = \int_0^t \lambda(\tau) e^{-\rho(\tau)}\, d\tau$; **vi)** $\alpha(t) = \frac{a(t)}{c(t)}$; **vii)** $\beta(t) = \frac{b(t)}{c(t)}$ y **viii)** $\gamma(t) = \frac{d(t)}{c(t)}$. Desarrollando esta forma compacta de la solución en una serie de potencias de w, se obtiene que[150]:

$$G_n(w,t) = \left[\alpha(t)^{n_0} + \binom{n_0}{1} \alpha(t)^{n_0 - 1} \beta(t) w + \binom{n_0}{2} \alpha(t)^{n_0 2} \beta(t)^2 w^2 + \ldots + \beta(t)^{n_0} w^{n_0} \right] \cdot$$

$$\cdot \left[1 + \binom{n_0}{1} \gamma(t) w + \binom{n_0 + 1}{2} \gamma(t)^2 w^2 + \binom{n_0 + 2}{3} \gamma(t)^3 w^3 + \ldots \right] =$$

$$= \sum_{n=0}^{n_0} \alpha(t)^{n_0 - n} \left\{ \sum_{m=0}^{n} \binom{n_0}{n-m} \binom{n_0 + m - 1}{m} \left[\alpha(t)\gamma(t) \right]^m \left[\beta(t) \right]^{n-m} \right\} w^n +$$

$$+ \sum_{n=n_0+1}^{\infty} \gamma(t)^{n - n_0} \left\{ \sum_{m=0}^{n_0} \binom{n-m}{n_0} \binom{n+m-n_0}{n+m-1} \left[\alpha(t)\gamma(t) \right]^m \left[\beta(t) \right]^{n_0 - m} \right\} w^n$$

es decir, que:

$$p_{n_0,n}(t) = \begin{cases} \alpha(t)^{n_0 - n} \left\{ \sum_{m=0}^{n} \binom{n_0}{n-m} \binom{n_0 + m - 1}{n_0 - 1} \left[\alpha(t)\gamma(t) \right]^m \left[\beta(t)^{n-m} \right] \right\} & (n \le n_0) \\ \gamma(t)^{n - n_0} \left\{ \sum_{m=0}^{n_0} \binom{n_0}{n-m} \binom{n+m-1}{n_0 - 1} \left[\alpha(t)\gamma(t) \right]^m \left[\beta(t) \right]^{n-m} \right\} & (n > n_0) \end{cases}$$

Derivando sucesivamente la función generatriz de probabilidades se obtienen los momentos[151]:

---

[150]. La definición explícita de esta distribución para el caso con λ y μ constantes planteado por W. Feller, fue obtenida por Palm, C. (1943).

[151]. En particular, para $n_0 = 1, E\left[n(t)/1\right] = e^{\rho(t)}$, define la variación esperada del tamaño de la población, desde el momento inicial hasta el momento t. A partir de esta expresión es posible determinar las funciones intensidad de nacimiento y fallecimiento de la siguiente forma $\lambda(t) - \mu(t) = \frac{d}{dt} \ln\left\{ E\left[n(t)/1\right] \right\}$. Los valores óptimos de $\lambda(t)$ y $\mu(t)$ serán aquéllos que minimicen $\sigma^2\left[n(t)/1\right]$ (ver Arley, N. (1948)).

$$E\left[n(t)/n_0\right] = n_0 e^{\rho(t)} = n_0 \exp\left\{\int_0^t \left[\lambda(\tau) - \mu(\tau)\right] d\tau\right\}$$

$$\sigma^2\left[n(t)/n_0\right] = \frac{\lambda(t)+\mu(t)}{\lambda(t)-\mu(t)} n_0 e^{\rho(t)}\left[e^{\rho(t)} - 1\right]$$

A partir de la forma compacta de la solución para la función generatriz de probabilidades obtenida en la página precedente, se obtiene la probabilidad de que, al momento t, la población se haya extinguido:

$$p_{n_0,0}(t) = G_n(0,t) = \left[\frac{e^{-\rho(t)} + \int_0^t \lambda(\tau) e^{-\rho(\tau)} d\tau - 1}{e^{-\rho(\tau)} + \int_0^t \lambda(\tau) e^{-\rho(\tau)} d\tau}\right]^{n_0}$$

Ahora bien, teniendo en cuenta que $\int_0^t \left[\lambda(\tau) - \mu(\tau)\right] d\tau = \rho(t) \neq 0$, resulta que:

$$\int_0^t \left[\lambda(\tau) - \mu(\tau)\right] e^{-\rho(\tau)} d\tau = -\int_0^t d\left(e^{-\rho(\tau)}\right) = 1 - e^{-\rho(t)}$$

de modo que se puede escribir $e^{-\rho(t)} + \int_0^t \lambda(\tau) e^{-\rho(\tau)} d\tau = 1 + \int_0^t \mu(\tau) e^{-\rho(\tau)} d\tau$ y, por lo tanto, se obtiene que:

$$p_{n_0,0}(t) = \left[\frac{\int_0^t \mu(\tau) e^{-\rho(\tau)} d\tau}{1 + \int_0^t \mu(\tau) e^{-\rho(\tau)} d\tau}\right]^{n_0}$$ .Dado que el integrando, $\mu(\tau)e^{-\rho(\tau)}$, es no-negativo, cuando $t \to \infty$, la integral que figura en el segundo miembro de esta expresión asumirá un valor finito o tenderá a $+\infty$, luego, se puede concluir que la probabilidad de extinción siempre asumirá un valor significativo. La condición necesaria y suficiente para que esta probabilidad tienda a la certeza es que dicha integral sea divergente (una condición necesaria, pero no suficiente, para que se verifique esta propiedad es que $\int_0^t \mu(\tau) d\tau$ sea divergente). En este caso la variable **tiempo a transcurrir hasta la extinción** (esta variable es equivalente a la **edad del proceso al momento de la extinción**) es tal que $F_{T_{n_0,0}}(t) = p\left(T_{n_0,0} \leq t\right) = p_{n_0,0}(t)$, de modo que su distribución de probabilidades queda definida de la siguiente forma:

$$f_{T_{n_0,0}}(t)dt = d\left(p_{n_0,0}(t)\right) = n_0\left[p_{n_0,0}(t)\right]^{n_0-1} \frac{\mu(t)e^{-\rho(t)}}{1 + \int_0^t \mu(\tau) e^{-\rho(\tau)} d\tau} dt$$

Si en este proceso se verifica que: **i)** $\mu_n(t) = 0\,(n = 1,2,\ldots; t \geq 0)$ y **ii)** $\lambda_n(t) = n\lambda(t)\,(n = n_0,$

$n_0+1, n_0+2, \ldots)$, queda definido un **proceso puro de nacimiento con función intensidad dependiente del tiempo y lineal con respecto al tamaño de la población**, con matriz Q de probabilidades de transición infinitesimal de la forma:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & -\lambda(t) & \lambda(t) & 0 & 0 & \ldots \\ 0 & 0 & -2\lambda(t) & 2\lambda(t) & 0 & \ldots \\ 0 & 0 & 0 & -3\lambda(t) & 3\lambda(t) & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

y cuyas probabilidades transición obedecen al siguiente sistema de ecuaciones de Kolmogorov:

$$\frac{d}{dt}p_{n_0,n}(t) = -n\lambda(t)p_{n_0,n}(t) + (n-1)\lambda(t)p_{n_0,n-1}(t) \qquad (n = n_0, n_0+1, \ldots; t \geq 0)$$

a partir del cual se obtiene la siguiente función generatriz de probabilidades:

$$G_n(w,t) = w^{n_0}\left\{\frac{\exp\left[-\int_0^t \lambda(\tau)d\tau\right]}{1-\left(1-\exp\left[-\int_0^t \lambda(\tau)d\tau\right]\right)w}\right\}^{n_0}$$

De modo que se puede concluir que las variables $n(t)$ se distribuyen de acuerdo con una función $b^{-1}\left(n_0; \exp\left[-\int_0^t \lambda(\tau)d\tau\right]\right)$.

En particular, si en este proceso se verifica que $\lambda_n(t) = \frac{\lambda + \lambda\alpha n}{1+\lambda\alpha t}\ (\lambda \geq 0; \alpha \geq 0)$, queda definido el **proceso de Greenwood-Yule-Eggenberger-Pólya**[152], cuya matriz Q de probabilidades de transición infinitesimal es de la forma:

$$\begin{bmatrix} -\frac{\lambda}{1+\lambda\alpha t} & \frac{\lambda}{1+\lambda\alpha t} & 0 & 0 & 0 & \ldots \\ 0 & -\frac{\lambda+\lambda\alpha}{1+\lambda\alpha t} & \frac{\lambda+\lambda\alpha}{1+\lambda\alpha t} & 0 & 0 & \ldots \\ 0 & 0 & -\frac{\lambda+2\lambda\alpha}{1+\lambda\alpha t} & \frac{\lambda+2\lambda\alpha}{1+\lambda\alpha t} & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

y cuyas probabilidades de transición obedecen al siguiente sistema de ecuaciones diferenciales-en diferencias:

---

[152]. Ver Greenwood, M.; Yule, G.U. (1915)(1920), Eggenberger, F.; Pólya, G. (1923), Eggenberger, F. (1924), Pólya, G. (1931), Pólya, G.; Szegö, G. (1964).

$$\frac{d}{dt}p_{n_0,n}(t) = -\frac{\lambda+\lambda\alpha n}{1+\lambda\alpha t}p_{n_0,n}(t) + \frac{\lambda+\lambda\alpha(n-1)}{1+\lambda\alpha t}p_{n_0,n-1}(t) \qquad (n = n_0, n_0+1,\ldots; t \geq 0)$$

a partir del cual se obtiene la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades:

$$(1+\lambda\alpha t)\frac{\partial}{\partial t}G_n(w,t) + \lambda\alpha w(1-w)\frac{\partial}{\partial w}G_n(w,t) = -\lambda(1-w)G_n(w,t)$$

cuya solución general es de la forma:

$$G_n(w,t) = w^{-1/\alpha}\Phi\left[\frac{1-w}{w}(1+\lambda\alpha t)\right]$$

(donde Φ es una función diferenciable arbitraria). Teniendo en cuenta la condición inicial $G_n(w,0) = w^{n_0}$, se obtiene entonces la siguiente solución particular:

$$G_n(w,t) = w^{n_0}\left[\frac{\frac{1}{1+\lambda\alpha t}}{1-\left(\frac{\lambda\alpha t}{1+\lambda\alpha t}\right)w}\right]^{n_0+1/\alpha} = \frac{w^{n_0}}{\left[1+\lambda\alpha t(1-w)\right]^{n_0+1/\alpha}} =$$

$$= \sum_{n=n_0}^{\infty}\binom{n_0+\frac{1}{\alpha}-1}{n-n_0}(\lambda\alpha t)^{n-n_0}(1-\lambda\alpha t)^{-(n+1/\alpha)}w^n$$

De modo que se puede concluir que las variables $n(t)$ se distribuyen de acuerdo con una función $b^{-1}\left(n_0+\frac{1}{\alpha};\frac{1}{1+\lambda\alpha t}\right)$ de la forma $p_{n_0,n}(t) = \binom{n_0+\frac{1}{\alpha}-1}{n-n_0}(\lambda\alpha t)^{n-n_0}(1-\lambda\alpha t)^{-(n+1/\alpha)}$ $(n = n_0, n_0+1,\ldots)$, con momentos de primero y segundo orden:

$$E\left[n(t)/n_0\right] = n_0\left(n_0+\frac{1}{\alpha}\right)\lambda\alpha t$$

$$\sigma^2\left[n(t)/n_0\right] = \left(n_0+\frac{1}{\alpha}\right)\lambda\alpha t\left[2\left(n_0+\frac{1}{\alpha}\right)+\lambda\alpha t\right]$$

De los resultados anteriores se obtiene en forma inmediata que $\lim_{\alpha\to\infty}G_n(w,t) = w^{n_0}e^{-(1-w)\lambda t}$. Luego, se puede concluir que: **i)** cuando $\alpha \to \infty$, si no se tiene en cuenta la constante aditiva, este proceso converge al proceso de Poisson y **ii)** igual que en el caso del proceso de Poisson ponderado[153], en este proceso, las variables $n(t)$ se distribuyen de acuerdo con una función binomial negativa ($b^{-1}(n,p)$) donde el parámetro p es una función lineal inversa de t. Esta última conclusión implica la imposibilidad de

---

[153]. Ver Sec. 1.7.2.3.

distinguir entre los procesos de Poisson y binomial negativo a partir de observaciones sobre las variables $n(t)$ exclusivamente.

La distribución de probabilidades correspondiente a las variables de este proceso fue obtenida, en 1923, por F. Eggenberger y G. Pólya[154], tomando como punto de partida un esquema según el cual, de una urna que contiene N_1 bolillas blancas y N_2 bolillas negras, se realizan extracciones al azar de forma tal que, una vez realizada una extracción se repone la bolilla extraída y se agregan a la urna m bolillas de ese color. De modo que, excepto en el caso en que $m = 0$ (en el que queda definido un esquema de Bernoulli), los eventos definidos por los resultados de cada una de las extracciones no son independientes entre sí[155]. Supóngase que al cabo de ν extracciones se hayan obtenido n bolillas blancas, la probabilidad de extraer una bolilla blanca en la $(\nu+1)$ -ésima extracción:

$$\frac{N_1 + nm}{(N_1 + nm) + N_2 + (\nu - n)m} = \frac{N_1 + nm}{(N_1 + N_2) + \nu m} = \frac{\alpha + \beta n}{1 + \beta\nu}$$

depende (**efecto contagio**) del número de bolillas blancas obtenidas en las extracciones anteriores y del coeficiente $\beta = \frac{m}{N_1 + N_2}$ (que caracteriza la denominada **potencia del contagio**). La probabilidad de obtener X bolillas blancas al cabo de N extracciones será, entonces, de la forma:

$$p(X) = \frac{N_1(N_1 m)...[N_1 + (X-1)m]N_2(N_2 + m)...[N_2 + (N - X - 1)m]}{(N_1 + N_2)(N_1 + N_2 m)...[N_1 + N_2 + (N-1)m]}\binom{N}{X} =$$

$$= \frac{\alpha(\alpha+\beta)...[\alpha + (X-1)\beta](1-\alpha)(1-\alpha+\beta)...[1-\alpha+(N-X-1)\beta]}{1(1+\beta)...[1+(N-1)\beta]}\binom{N}{X} \qquad (X = 0,1,...N)$$

Los momentos de la variable X están definidos por:

$$E(X) = N\frac{N_1}{N_1 + N_2}$$

$$\sigma^2(X) = N\frac{N_1}{N_1 + N_2}\frac{N_2}{N_1 + N_2}\frac{(N_1 + N_2) + Nm}{(N_1 + N_2) + m} = N\alpha(1-\alpha)\frac{1 + N\beta}{1 + \beta}$$

Si se verifica que $\lim_{N\to\infty} N\alpha = \lambda t$ y $\lim_{N\to\infty} N\beta = \lambda kt$, se obtiene que:

$$\lim_{N\to\infty} p(X,t) = \begin{cases} \dfrac{1}{(1+\lambda kt)^{1/\lambda k}} & para\ X = 0 \\ \dfrac{1(1+\lambda k)...[1+(X-1)\lambda k]}{X!}\left(\dfrac{t}{1+\lambda kt}\right)^X\left(\dfrac{1}{1+\lambda kt}\right)^{1/\lambda k} & para\ X = 1,2,...N \end{cases}$$

---

154. Ver Eggenberger, F.; Pólya, G. (1923), Eggenberger, F. (1924), Pólya, G. (1931).

155. El resultado eventual de una extracción dada depende del número de "éxitos" y de "fracasos" obtenidos con anterioridad.

De modo que los momentos de la variable X, cuando N aumenta indefinidamente y se verifican las condiciones enunciadas más arriba, son $E(X,t) = \lambda t$ y $\sigma^2(X,t) = \lambda t(1 + \lambda k t)$ (obsérvese que, cuando $k \to \infty$, la distribución se transforma en una función de Poisson).

Los distintos límites a los que puede converger la distribución $p(X)$ dependen de los valores que asuman los coeficientes α y β, cuando N aumenta indefinidamente. Pólya G. (1931) define a un evento como "usual", en este esquema, cuando se verifica que $\lim_{N\to\infty} \alpha < \infty$ y $\lim_{N\to\infty} E(X) = \infty$. Por el contrario, define como "raro" a todo evento tal que $\lim_{N\to\infty} E(X) < \infty$ (de modo que $\lim_{N\to\infty} \alpha = 0$). A su vez, cuando $\lim_{N\to\infty} \beta = 0$, de modo que $N\beta$ tienda a un valor constante, se dice que el esquema es **de contagio débil**. En caso contrario, se dice que el esquema es **de contagio fuerte**. El caso que aquí se trata es el que corresponde a eventos raros con contagio débil.

Si a cada extracción, en una serie de N extracciones, se le asocia un intervalo de tiempo $\Delta t = \frac{t}{N}$ y expresando los coeficientes α y β como $\alpha = \lambda \frac{t}{N} = \lambda \Delta t$ y $\beta = \lambda k \frac{t}{N} = \lambda k \Delta t$, se obtiene que $\frac{\alpha + \beta n}{1 + \beta \nu} = \frac{\lambda + \lambda k n}{1 + \lambda k \tau} \Delta t$, donde $\tau = \nu \frac{t}{N} = \nu \Delta t$ denota el punto del tiempo hasta el cual se han producido ν extracciones.

Greenwood, M.; Yule, G.U. (1920) habían arribado a un resultado similar al de Eggenberger-Pólya basándose en la hipótesis de que el contagio es sólo aparente, que los eventos son independientes, debiéndose las variaciones en las probabilidades a la falta de homogeneidad de la población. Su solución se desarrolla a partir de una distribución de probabilidades de Poisson de la forma:

$$p(X,\lambda) = \begin{cases} \dfrac{\lambda^X e^{-\lambda}}{X!} & \text{para } X = 0,1,2,... \\ 0 & \text{para cualquier otro valor de } X \end{cases}$$

en la que el coeficiente λ denota una variable aleatoria con función de probabilidades Gamma de la forma:

$$g(\lambda,\alpha,\beta) = \begin{cases} \dfrac{\lambda^{\alpha-1} e^{-\lambda/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & \text{para } \lambda \geq 0; \alpha,\beta > 0 \\ 0 & \text{para cualquier otro caso} \end{cases}$$

Woodbury, M.A. (1949) estudió una distribución del tipo **de contagio verdadero** en la que hace depender la probabilidad de "éxito" exclusivamente del número de "éxitos" obtenidos en las pruebas anteriores (a partir de un esquema de urnas en el que, si el resultado obtenido en una extracción dada es un "fracaso", se repone la bolilla extraída y, si el resultado es un "éxito", se realiza una recomposición de la urna reemplazando, por ejemplo, bolillas "fracaso" por bolillas "éxito"), de modo que, denotando por X_n a la variable que representa el número de "éxitos" a obtener en n pruebas y por p_i $(i = 0,1,2,...,n)$ a la probabilidad de obtener "éxito" en una prueba dada, si ya se obtuvieron i "éxitos" en las pruebas

anteriores, será:

$$p(X_n = 0) = (1 - p_0)^n$$

$$p(X_n = x) = \left(\prod_{j=0}^{x-1} p_j\right) \sum_{i_0=0}^{n-x} (1 - p_0)^{i_0} \sum_{i_1=0}^{n-x-i_0} (1 - p_1)^{i_1} \sum_{i_2=0}^{n-x-i_0-i_1} (1 - p_2)^{2_2} \ldots \sum_{I_{x-2}=0}^{n-x-i_0-\ldots-i_{x-3}} (1 - p_{x-2})^{i_{x-2}} \cdot$$

$$\cdot \sum_{i_{x-1}=0}^{n-x-\ldots i_{x-2}} (1 - p_{x-1})^{i_{x-2}} (1 - p_x)^{n-x-i_0-\ldots-i_{x-1}} \qquad (x = 1,2,\ldots,n)$$

A partir de estas expresiones es posible definir la siguiente relación de recurrencia: $p\left[(X+1)_{n+1}\right] =$ $= p(X_n)p_X + p\left[(X+1)_n\right](1 - p_{X+1})$ y, efectuando el siguiente cambio de variables $p(X_n) =$ $= F(X_n)p_0p_1\ldots p_{X-1}$, se puede escribir:

$$F\left[(X+1)_{n+1}\right] = F(X_n) + F\left[(X+1)_n\right](1 - p_{X+1})$$

Si se multiplica m. a m. esta expresión por t^n y se suma con respecto a n, se obtiene que:

$$\sum_{n=X}^{\infty} F\left[(X+1)_{n+1}\right] w^n = \sum_{n=X}^{\infty} F(X_n) w^n + (1 - p_{X+1}) \sum_{n=X}^{\infty} F\left[(X+1)_n\right] w^n$$

$$\frac{1}{t} G_{X+1}(w) = G_X(w) + (1 - p_{X+1}) G_{X+1}(w)$$

(donde $G_X(w) = \sum_{n=X}^{\infty} F(X_n) w^n$ define la función generatriz de la sucesión $F(X_n)$). Luego, operando en forma iterativa, resulta que:

$$G_1(w) = \frac{w}{1 - (1 - p_1)w} G_0(w)$$

$$G_2(w) = \frac{w^2}{\left[1 - (1 - p_1)w\right]\left[1 - \left[1 - p_2\right]w\right]} G_0(w)$$

.....

$$G_X(w) = \frac{t^X}{\prod_{j=1}^{X} \left[1 - (1 - p_j)w\right]} G_0(w)$$

(donde $G_0(w) = F(0_0) + F(0_1)w + F(0_2)w^2 + \ldots = 1 + (1 - p_0)w + (1 - p_0)^2 w^2 + \ldots = \frac{1}{1 - (1 - p_0)w}$). Luego, se puede escribir:

$$G_X(w) = \frac{w^X}{\prod_{j=0}^{X}\left[1-\left(1-p_j\right)w\right]} = w^X \sum_{h=0}^{X} \frac{\left(1-p_h\right)^X}{\prod_{\substack{j=0 \\ j\neq h}}^{X}\left(p_j-p_h\right)\left[1-\left(1-p_h\right)\right]} =$$

$$= w^X \sum_{h=0}^{X} \frac{\left(1-p_h\right)^X}{\prod_{\substack{j=0 \\ j\neq h}}^{X}\left(p_j-p_h\right)}\left[1+\left(1-p_h\right)w+\left(1-p_h\right)^2 w^2+\ldots\right]$$

Expresión de la que se obtiene que $F(X_n) = \sum_{h=0}^{X} \frac{\left(1-p_h\right)^X}{\prod_{\substack{j=0 \\ j\neq h}}^{X}\left(p_j-p_h\right)}$ y, en consecuencia, resultas la siguiente expresión alternativa para la distribución de la variable X_n:

$$p(X_n) = \left[\prod_{i=0}^{X-1}\left(1-p_i\right)\right] \sum_{h=0}^{X} \frac{\left(1-p_h\right)^X}{\prod_{\substack{j=0 \\ j\neq h}}^{X}\left(p_j-p_h\right)}$$

Neyman, J. (1939) desarrolló, a su vez, una función del tipo **de contagio** para estudiar la distribución de larvas en un plano dividido en zonas de áreas iguales, suponiendo que las larvas pueden provenir de familias diferentes y estadísticamente independientes. Si se denota por $\pi(n)\,(n = 0,1,2,\ldots)$ a la probabilidad de que una familia cuente con n individuos sobrevivientes, entonces, la probabilidad de que una familia en particular cuente con n sobrevivientes, en una zona seleccionada al azar, estará definida por:

$$g(n,\alpha) = \sum_{j=0}^{\infty} \pi(j)\, p(n,\alpha / j) = \sum_{j=0}^{\infty} \pi(j) \binom{j}{n} \alpha^j (1-\alpha)^{j-n} \qquad (n = 0,1,2,\ldots; n \leq j)$$

donde $0 < \alpha < 1$ es un coeficiente que caracteriza a cada familia y al que se asocia una función $h(\alpha)$ que posee las propiedades de una función de distribución de probabilidades. Luego, la probabilidad de encontrar, en la zona seleccionada, n larvas sobrevivientes pertenecientes a la misma familia, será $g(n) = \int_0^1 g(n,\alpha)h(\alpha)\,d\alpha \;(n = 0,1,2,\ldots)$. Si se supone que la variable aleatoria que representa el número de familias que pueden existir en la zona seleccionada se distribuye de acuerdo con una función de Poisson, entonces la probabilidad de hallar n larvas en dicha zona, está definida por una función de Poisson generalizada de la forma $f(n) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda}\left[g(n)\right]^{k^*} \;(n = 0,1,2,\ldots)$. En particular:

i) si se verifica que $\pi(j) = \frac{\beta^j}{j!} e^{-\beta}\,(j = 0,1,2,\ldots)$ y $\alpha = \alpha_0$ es un valor constante, será $h(\alpha) = 0$

$\left(0 < \alpha < \alpha_0 ; \alpha_0 \leq 1\right)$, entonces:

$$g(n) = \sum_{j=n}^{\infty} \frac{\beta^j}{j!} e^{-\beta} \binom{j}{n} \alpha_0^n (1-\alpha_0)^{j-n} = e^{-\beta} \frac{(\alpha_0\beta)^n}{n!} \sum_{j=0}^{\infty} \frac{\left[(1-\alpha_0)\beta\right]^j}{j!} = \frac{(\alpha_0\beta)^n}{n!} e^{-\alpha_0\beta}$$

obteniéndose, en consecuencia, una distribución de Neyman-tipo A, cuya función generatriz de momentos es de la forma:

$$M_n^{(A)}(w) = \exp\left\{\lambda\left[\exp\left[\delta\left(e^w - 1\right)\right] - 1\right]\right\} = \exp\left[\lambda \sum_{j=0}^{\infty} \frac{\delta^{j+1}\left(e^w - 1\right)^{j+1}}{(j+1)!}\right]$$

ii) si se verifica que $h(\alpha) = \frac{1}{2\alpha_0}\left(0 < \alpha < 2\alpha_0 ; \alpha_0 \leq \frac{1}{2}\right)$, resulta que:

$$g(n) = \frac{1}{n!}\frac{1}{2\alpha_0} \int_0^{2\alpha_0} (\alpha\beta)^n e^{-\alpha\beta}\, d\alpha =$$

$$= \frac{1}{2\alpha_0\beta n!}\Big[-\left(2\alpha_0\beta\right)^n e^{-2\alpha_0\beta} - n\left(2\alpha_0\beta\right)^{n-1} e^{-2\alpha_0\beta} - n(n-1)\left(2\alpha_0\beta\right)^{n-2} e^{-2\alpha_0\beta} - \ldots -$$

$$- n(n-1)\ldots 4.3\left(2\alpha_0\beta\right)^2 e^{-2\alpha_0\beta} - n(n-1)\ldots 3.2\left(2\alpha_0\beta\right) e^{-2\alpha_0\beta} - n(n-1)\ldots 2.1\left(2\alpha_0\beta\right)^0 e^{-2\alpha_0\beta} + n!\Big] =$$

$$= \frac{1}{2\alpha_0\beta} - \frac{e^{-2\alpha_0\beta}}{2\alpha_0\beta} \sum_{j=-1}^{n-1} \frac{\left(2\alpha_0\beta\right)^{j+1}}{(j+1)!} + e^{2\alpha_0\beta} \sum_{j=n}^{\infty} \frac{\left(2\alpha_0\beta\right)^j}{(j+1)!} =$$

$$= e^{-2\alpha_0\beta} \sum_{j=n}^{\infty} \frac{\left(2\alpha_0\beta\right)^j}{(j+1)!}$$

obteniéndose una distribución de Neyman-tipo B, cuya función generatriz de momentos está definida por:

$$M_n^{(B)}(w) = \exp\left\{\frac{\lambda}{\delta\left(e^w - 1\right)}\left[e^{\delta\left(e^w - 1\right)} - 1 - \delta\left(e^w - 1\right)\right]\right\} = \exp\left\{\lambda \sum_{j=0}^{\infty} \frac{\delta^{j+1}\left(e^w - 1\right)^{j+1}}{(j+2)!}\right\}$$

iii) si se verifica que $h(\alpha) = \frac{2}{9\alpha_0^2}\left(3\alpha_0 - \alpha\right)\left(0 < \alpha < 3\alpha_0 ; \alpha_0 \leq \frac{1}{3}\right)$, resulta que:

$$g(n) = \frac{2}{3\alpha_0 n!} \int_0^{3\alpha_0} (\alpha\beta)^n e^{-\alpha\beta}\, d\alpha - \frac{2\beta^n}{9\alpha_0 n!} \int_0^{3\alpha_0} \alpha^{n+1} e^{-\alpha\beta}\, d\alpha = 2\exp\left(3\alpha_0\beta\right)\left[\sum_{j=n}^{\infty} \frac{\left(3\alpha_0\beta\right)^j}{(j+1)!} - \frac{n+1}{3\beta} \sum_{j=n+1}^{\infty} \frac{\left(3\alpha_0\beta\right)^j}{(j+1)!}\right]$$

obteniéndose una distribución de Neyman-tipo C, cuya función generatriz de momentos es de la forma:

$$M_m^{(C)}(w) = \exp\left\{\frac{2\lambda}{\delta^2\left(e^w-1\right)^2}\left[\exp\left(\delta\left(e^w-1\right)\right)-1-\delta\left(e^w-1\right)\right]-\frac{\delta^2}{2!}\left(e^w-1\right)\right\} =$$

$$= \exp\left[2\lambda\sum_{j=0}^{\infty}\frac{\delta^{j+1}\left(e^w-1\right)^{j+1}}{(j+3)!}\right]$$

Obsérvese que las funciones asociadas a las variables aquí estudiadas, constituyen casos particulares de una función general de la forma[156] $h(\alpha) = \frac{s}{(s+1)^s\alpha_0^s}\left[(s+1)\alpha_0-\alpha\right]^{s-1}$ $\left(0<\alpha<(s+1)\alpha_0;\alpha_0\le\frac{1}{s+1}\right)$, a partir de la cual es posible definir una forma general de la distribución de contagio de Neyman, cuya función generatriz de momentos está dada por la expresión:

$$M_n(w) = \exp\left[s!\lambda\sum_{j=0}^{\infty}\frac{\delta^{j+1}\left(e^w-1\right)^{j+1}}{(j+s+1)!}\right] = \exp\left[s!\lambda\sum_{j=0}^{\infty}\frac{\delta^j\left(e^w-1\right)^j}{(j+s)!}-\lambda\right]$$

Si en el proceso de nacimiento-fallecimiento definido en esta sección se verifica que: **i)** $\lambda_n(t) = 0\ (n = 0,1,2,...,n_0;t\ge 0)$ y **ii)** $\mu_n(t) = n\mu(t)\ (n = 1,2,...,n_0;t\ge 0)$ [157], queda definido un **proceso puro de fallecimiento con función intensidad dependiente del tiempo y lineal con respecto al tamaño de la población**, con matriz Q de probabilidades de transición infinitesimal de la forma[158]:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \mu(t) & -\mu(t) & 0 & 0 & \dots & 0 & 0 \\ 0 & 2\mu(t) & -2\mu(t) & 0 & \dots & 0 & 0 \\ 0 & 0 & 3\mu(t) & -3\mu(t) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & n_0\mu(t) & -n_0\mu(t) \end{bmatrix}$$

y cuyas probabilidades de transición son de la forma:

$$p_{n_0,n}(t) = \binom{n_0}{n}\left\{\exp\left[\int_0^t\mu(\tau)d\tau\right]\right\}^n\left\{1-\exp\left[\int_0^t\mu(\tau)d\tau\right]\right\}^{n_0-n} \qquad \left(n = 0,1,2,...,n_0;t\ge 0\right)$$

---

156. Obsérvese que, para $s = 0,1,2$, se obtienen, respectivamente, las funciones $h(\alpha)$ correspondientes a las distribuciones de Neyman tipo A, B y C.

157. A todo conjunto de individuos vivos al momento $t = 0$ y sujetos, en forma independiente, al mismo riesgo de fallecimiento se lo denomina **cohorte**.

158. La variable t se interpreta, en este caso, como representativa de la edad.

A partir de este resultado se puede concluir en forma inmediata que $n(t) : b\left(n_0; \exp\left[\int_0^t \mu(\tau)d\tau\right]\right)$. En particular, si $n_0 = 1$, $p_{11}(t) = \exp\left[-\int_0^t \mu(\tau)d\tau\right]$ $(t \geq 0)$ define la probabilidad de que un individuo de la población con las características mencionadas más arriba, permanezca vivo a la edad t, $p_{11}(t) = p(\mathfrak{I} > t)$ $= 1 - p(\mathfrak{I} \leq t) = 1 - F_{\mathfrak{I}}(t)$ (donde $\mathfrak{I}$ denota la variable que representa la **duración de la vida** de dicho individuo). Se puede concluir, entonces que la duración de la vida de un individuo en las condiciones mencionadas, se distribuye de acuerdo con una función de probabilidades de la forma $f_{\mathfrak{I}}(t) = \frac{d}{dt}F_{\mathfrak{I}}(t) = \mu(t)\exp\left[-\int_0^t \mu(\tau)d\tau\right]$ $(t \geq 0)$.

Gompertz, B. (1825) propuso la función $\frac{1}{\mu(t)}$ como una medida de la **"resistencia" al fallecimiento** de un individuo y, por lo tanto, concluyó que $\frac{d}{dt}\left[\frac{1}{\mu(t)}\right] = -\frac{h}{\mu(t)}$ (donde h denota una constante positiva). Expresión de la cual se obtiene que $\log\left[\frac{1}{\mu(t)}\right] = -ht + k$ y, en consecuencia, que $\mu(t) = Bc^t$. Sustituyendo este resultado en la expresión correspondiente a la función de densidad de la variable $\mathfrak{I}$, se obtiene que $f_{\mathfrak{I}}(t) = Bc^t \exp\left[\frac{-Bc^t}{\log(c)}\right]$.

Alternativamente a este planteo, Makeham, W.M. (1860) propuso la siguiente **ley de mortalidad**[159] $\mu(t) = A + Bc^t$. Si se supone que la función intensidad de fallecimiento se comporta de acuerdo con una función exponencial de la forma $\mu(t) = \mu\alpha t^{\alpha-1}$, se verifica que $f_{\mathfrak{I}}(t) = \mu\alpha t^{\alpha-1}\exp\left(-\mu t^{\alpha}\right)$ (ver Weibull, W. (1939)). Si se supone que $\mu(t) = \mu$, se obtiene que $f_{\mathfrak{I}}(t) = \mu e^{-\mu t}$ (ver Sec. 1.7.3.3).

1.7.4.- El proceso de nacimiento-migración-fallecimiento

1.7.4.1.- El proceso de nacimiento-inmigración-emigración-fallecimiento lineal unidimensional

Sea $\{n(t), t \geq 0\}$ un proceso que representa el tamaño de una población al momento t. Supóngase: **i)** que dicha población admita ingresos por nacimiento e inmigración y egresos por fallecimiento y emigración; **ii)** que sus integrantes están sometidos, en forma individualmente independiente, a los mismos riesgos de "reproducción" y fallecimiento; **iii)** que estos riesgos son, a su vez, independientes del comportamiento histórico de cada individuo y de la "edad" que representa la evolución del proceso y **iv)**

---

[159]. Especialmente apropiada para describir el comportamiento de la mortalidad en individuos mayores de 20 años.

que el tamaño de la población inicial es n_0 [160].

A partir de estas condiciones se obtiene el siguiente sistema de probabilidades de transición infinitesimal:

i) $p_{n,n+1}(t,t+\Delta t) = n\lambda\Delta t + \xi\Delta t + o(\Delta t)\,(n = 0,1,2,...)$ (donde ξ denota la **tasa instantánea de inmigración**, independiente del tamaño de la población);

ii) $p_{n,n-1}(t,t+\Delta t) = n\mu\Delta t + \theta\Delta t + o(\Delta t)\,(n = 0,1,2,...)$ (donde θ denota la **tasa instantánea de emigración**, independiente del tamaño de la población);

iii) $p_{n,n+k}(t,t+\Delta t) = p_{n,n-k}(t,t+\Delta t) = o(\Delta t)\,(k = 2,3,...;n = 0,1,2,...)$;

iv) $p_{nn}(t,t+\Delta t) = 1 - \left[n(\lambda+\mu) + (\xi+\theta)\right]\Delta t - o(\Delta t)\,(n = 0,1,2,...)$

y el siguiente sistema de ecuaciones diferenciales-en diferencias de Kolmogorov:

$$\frac{d}{dt}p_{n_0,n}(t) = -\left[n(\lambda+\mu) + (\xi+\theta)\right]p_{n_0,n}(t) + \left[(n-1)\lambda + \xi\right]p_{n_0,n-1}(t) +$$
$$+ \left[(n+1)\mu + \theta\right]p_{n_0,n+1}(t) \qquad (n = 0,1,2,...;t \geq 0)$$

a partir del cual se obtiene la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades:

$$\frac{\partial}{\partial t}G_n(w,t) - \left[\lambda w^2 - (\lambda+\mu)w + \mu\right]\frac{\partial}{\partial w}G_n(w,t) = \left[-(\xi+\theta) + \xi w + \frac{\theta}{w}\right]G_n(w,t)$$

cuya solución general es de la forma:

$$G_n(w,t) = w^{-\theta/w}(\lambda w - \mu)^{-\left(\frac{\xi}{\lambda}-\frac{\theta}{\mu}\right)} f\left[\frac{w-1}{\lambda w - \mu}e^{(\lambda-\mu)t}\right] \qquad (|w| < 1)$$

Teniendo en cuenta la condición inicial:

$$G_n(w,0) = w^{n_0} = w^{-\theta/w}(\lambda w - \mu)^{-\left(\frac{\xi}{\lambda}-\frac{\theta}{\mu}\right)} f\left(\frac{w-1}{\lambda w - \mu}\right)$$

se obtiene la siguiente solución particular:

---

[160]. Este tipo de procesos fue introducido por Kendall, D.G. (1948) como una versión estocástica del modelo de Fisher-Corbett-Williams (ver Fisher, R.A.; Corbett, A.S.; Williams, C.B. (1943)), para describir la evolución de una colonia hipotética de organismos cuya variación es atribuible a la reproducción (subdivisión binaria), fallecimiento e inmigración. La emigración figuraba incluida en el factor mortalidad, debido a la imposibilidad de diferenciarla matemáticamente. Ver, además, Bartlett, M.S. (1949), Patil, V.T. (1957), Chiang, Ch.L. (1964).

$$G_n(w,t) = w^{-\theta/\mu}(\lambda-\mu)^{\left(\frac{\xi}{\lambda}-\frac{\theta}{\mu}\right)} \frac{\left\{\mu\left[e^{(\lambda-\mu)t}-1\right]+\left[\lambda-\mu e^{(\lambda-\mu)t}\right]w\right\}^{n_0+\theta/\mu}}{\left\{\left[\lambda e^{(\lambda-\mu)t}-\mu\right]-\lambda\left[e^{(\lambda-\mu)t}-1\right]w\right\}^{n_0+\xi/\lambda}} =$$

$$= \frac{(\lambda-\mu)^{\frac{\xi}{\lambda}-\frac{\theta}{\mu}}}{w^{\theta/\mu}} \frac{\left[a(t)+b(t)w\right]^{n_0+\theta/\mu}}{\left[c(t)-d(t)w\right]^{n_0+\xi/\lambda}} =$$

$$= \frac{(\lambda-\mu)^{\frac{\xi}{\lambda}-\frac{\theta}{\mu}}}{w^{\theta/\mu}} \frac{\left[a(t)+b(t)w\right]^{n_0+\theta/\mu}}{c(t)^{n_0+\xi/\lambda}\left[1-\gamma(t)w\right]^{n_0+\xi/\lambda}} =$$

$$= \frac{(\lambda-\mu)^{\alpha-\beta}}{c(t)^{n_0+\alpha}}\left\{\sum_{n=0}^{n_0}\left[a(t)\right]^{n_0-n}\left[\sum_{i=0}^{n+\beta}\binom{n_0+\beta}{n+\beta-i}\binom{(n_0+\alpha)+i-1}{(n+\alpha)-1}\left[\alpha(t)\gamma(t)\right]^i\left[b(t)\right]^{(n+\beta)-i}\right]w^n + \right.$$

$$\left. + \sum_{n=n_0+1}^{\infty}\left[\gamma(t)\right]^{n-n_0}\left[\sum_{i=0}^{n_0+\beta}\binom{n_0+\beta}{(n_0+\beta)-i}\binom{(n+\alpha)+i-1}{(n_0+\alpha)-1}\left[a(t)\gamma(t)\right]^i\left[b(t)\right]^{n_0+\beta-i}\right]w^n\right\} \qquad (\lambda \neq \mu)$$

(donde $\alpha = \frac{\xi}{\lambda}$ y $\beta = \frac{\theta}{\mu}$), a partir de la cual se puede concluir en forma inmediata que:

$$p_{n_0 m}(t) = \begin{cases} \dfrac{(\lambda-\mu)^{\alpha-\beta}}{\left[c(t)\right]^{n_0+\alpha}}\left[a(t)\right]^{n_0-n}\displaystyle\sum_{i=0}^{n+\beta}\binom{n_0+\beta}{n+\beta-i}\binom{(n_0+\alpha)+i-1}{(n+\alpha)-1}\left[a(t)\gamma(t)\right]^i\left[b(t)\right]^{(n+\beta)-i} & (n \leq n_0; \lambda \neq \mu) \\ \dfrac{(\lambda-\mu)^{\alpha-\beta}}{\left[c(t)\right]^{n_0+\alpha}}\left[\gamma(t)\right]^{n-n_0}\displaystyle\sum_{i=1}^{n+\beta}\binom{n_0+\beta}{(n_0+\beta)-i}\binom{(n+\alpha)+i-1}{(n_0+\alpha)-1}\left[a(t)\gamma(t)\right]^i\left[b(t)\right]^{(n_0+\beta)-i} & (n > n_0; \lambda \neq \mu) \end{cases}$$

En consecuencia, el tamaño esperado de la población al momento t, está dado en este caso por:

$$E\left[n(t)/n_0\right] = \left(n_0 + \frac{\theta}{\mu}\right)e^{(\lambda-\mu)t} + \frac{\xi-\theta}{\lambda-\mu}\left[e^{(\lambda)\mu t}-1\right]$$

Para $\lambda = \mu$, se obtiene que:

$$\lim_{\lambda\to\mu} G_n(w,t) = \frac{1}{w^{\theta/\mu}}\left[\frac{1}{(1+\mu t)-\mu t w}\right]^{\frac{\xi}{\lambda}-\frac{\theta}{\mu}}\left[\frac{\mu t+(1-\mu t)w}{(1+\mu t)-\mu t w}\right]^{n_0+\frac{\theta}{\mu}}$$

y, en consecuencia, que $E\left[n(t)/n_0\right] = n_0 + (\xi-\theta)t \ (\lambda = \mu)$.

A partir de estos resultados, cuando $t \to \infty$, se obtiene en forma inmediata que:

$$\lim_{t\to\infty} E\left[n(t)/n_0\right] = \begin{cases} \dfrac{\xi-\theta}{\mu-\lambda} & (\lambda<\mu;\xi>0) \\ 0 & (\lambda<\mu;\xi\le\theta) \\ \infty & (\lambda=\mu;\xi>\theta) \\ n_0 & (\lambda=\mu;\xi=\theta) \\ 0 & (\lambda=\mu;\xi<\theta) \\ \infty & (\lambda>\mu;\xi\ge\theta) \\ 0 & \left(\lambda>\mu;\theta-\xi>(\lambda-\mu)\left(n_0+\dfrac{\theta}{\mu}\right)\right) \end{cases}$$

Si, en particular, $n_0 = 0$, es decir, si se trata de una población formada exclusivamente a partir del aporte inmigratorio, de modo que en un momento t dado estará integrada por un cierto número de familias independientes, cada una descendiente de un ancestro inmigrante, la función generatriz de probabilidades será de la forma $G_n(w,t) = \left(\dfrac{\lambda-\mu}{\lambda e^{(\lambda-\mu)t}-\mu}\right)^{\xi/\lambda}\left\{1-\dfrac{\lambda\left[1-w^{(\lambda-\mu)t}\right]}{\lambda e^{(\lambda)\mu}-\mu}w\right\}^{-\xi/\lambda} \quad (\lambda\neq\mu)$. Luego, se puede concluir que $n(t) = b^{-1}\left\{\dfrac{\xi}{\lambda};\dfrac{\lambda-\mu}{\lambda\left[e^{(\lambda-\mu)t}-1\right]}\right\}$ y, por lo tanto, que el tamaño esperado de la población al momento t será $E\left[n(t)/n_0\right] = \dfrac{\xi\left[e^{(\lambda-\mu)t}-1\right]}{\lambda-\mu}\ (\lambda\neq\mu)$.

1.7.4.2.- El proceso de nacimiento-migración-fallecimiento bidimensional

Sea $\{n(t)\}$ un proceso bidimensional que representa el tamaño de una población formada por dos subpoblaciones interconectadas, representadas por los procesos $\{n_1(t)\}$ y $\{n_2(t)\}$.

Supóngase: **i)** que las intensidades de nacimiento son proporcionales a los tamaños esperados de las respectivas poblaciones marginales ($E(n_1(t))$ y $E(n_2(t))$)[161]; **ii)** que las intensidades de fallecimiento son proporcionales a los tamaños de las poblaciones marginales; **iii)** que las intensidades de migración son proporcionales a los tamaños de las poblaciones de origen y **iv)** que los tamaños de las subpoblaciones

---

[161]. Renshaw, E. (1973) propuso esta modificación a la estructura de los nacimientos a fin de lograr una aproximación más simple a la solución de la ecuación en derivadas parciales para la función generatriz de probabilidades.

iniciales son $n_1(0) = n_{10}$ y $n_2(0) = n_{20}$.

De acuerdo con estas hipótesis: **i)** la probabilidad de que se produzca un nacimiento en la subpoblación n_i $(i = 1,2)$ en el intervalo infinitesimal $(t, t + \Delta t]$, será de la forma:

$$p\left\{\left[n_{ik}(t + \Delta t) = n_i + 1; n_j(t + \Delta t) = n_j\right] / \left[n_i(t) = n_i; n_j(t) = n_j\right]\right\} = p_{(n_i, n_j);(n_i+1, n_j)}(t, t + \Delta t) =$$

$$= \lambda E\left[n_i(t)\right]\Delta t + o(\Delta t) \qquad (i, j = 1,2,...; i \neq j)$$

ii) la probabilidad de que se produzca un fallecimiento en la subpoblación n_i $(i = 1,2,)$, en el intervalo infinitesimal $(t, t + \Delta t]$, será de la forma:

$$p\left\{\left[n_i(t + \Delta t) = n_i - 1; n_j(t + \Delta t) = n_j\right] / \left[n_i(t) = n_i; n_j(t) = n_j\right]\right\} = p_{(n_i, n_j)(n_i-1, n_j)}(t, t + \Delta t) =$$

$$= n_i \mu \Delta t + o(\Delta t) \qquad (i, j = 1,2; i \neq j)$$

iii) la probabilidad de que se produzca una migración de la subpoblación n_j $(j = 1,2)$ a la subpoblación n_i $(i = 1,2)$, en un intervalo infinitesimal $(t, t + \Delta t]$, será de la forma:

$$p\left\{\left[n_i(t + \Delta t) = n_i + 1; n_j(t + \Delta t) = n_j - 1\right] / \left[n_i(t) = n_i; n_j(t) = n_j\right]\right\} = p_{(n_i, n_j)(n_i+1, n_j-1)}(t, t + \Delta t) =$$

$$= n_j \frac{\theta}{2} \Delta t + o(\Delta t) \qquad (i, j = 1,2; i \neq j)$$

iv) la probabilidad de que en la subpoblación n_i $(i = 1,2)$ se produzca más de un evento (nacimiento, fallecimiento o migración) en un intervalo infinitesimal $(t; t + \Delta t]$ será asimilable a una función $o(\Delta t)$ (condición de linealidad);

v) la probabilidad de que las supoblaciones, al momento $t + \Delta t$ se mantengan en el estado en que se encontraban al momento t, será entonces de la forma:

$$p\left\{\left[n_i(t + \Delta t) = n_i; n_j(t + \Delta t) = n_j\right] / \left[n_i(t) = n_i; n_j(t) = n_j\right]\right\} = p_{(n_i, n_j)(n_i, n_j)}(t, t + \Delta t) =$$

$$= 1 - \left\{\lambda\left[E(n_1(t)) + E(n_2(t))\right] + (n_1 + n_2)\left(\mu + \frac{\theta}{2}\right)\right\}\Delta t - o(\Delta t) \qquad (i, j = 1,2; i \neq j)$$

A partir de este sistema de probabilidades de transición infinitesimal, se obtiene el siguiente sistema de ecuaciones de Kolmogorov:

$$\frac{d}{dt} p_{(n_{10}, n_{20})(n_1+n_2)}(t) = -\left\{\lambda\left[E(n_1(t)) + E(n_2(t))\right] + \left(\mu + \frac{\theta}{2}\right)(n_1 + n_2)\right\} p_{(n_{10}, n_{20})(n_1, n_2)}(t) +$$

$$+ \lambda E(n_1(t)) p_{(n_{10}, n_{20})(n_1-1, n_2)}(t) + \lambda E(n_2(t)) p_{(n_{10}, n_{20})(n_1, n_2-1)}(t) +$$

$$+ (n_1 + 1)\mu p_{(n_{10}, n_{20})(n_1+1, n_2)}(t) + (n_2 + 1)\mu p_{(n_{10}, n_{20})(n_1, n_2+1)}(t) +$$

$$+\left(n_2+1\right)\frac{\theta}{2}p_{(n_{10},n_{20})(n_1-1,n_2+1)}(t)+\left(n_1+1\right)\frac{\theta}{2}p_{(n_{10},n_{20})(n_1+1,n_2-1)}(t)\qquad \left(n_1,n_2=0,1,2,\ldots;t\geq 0\right)$$

Multiplicando m. a m. por $w_1^{n_1}$ y $w_2^{n_2}$ y sumando con respecto a n_1 y n_2, se obtiene la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades:

$$\frac{\partial}{\partial t}G_{n_1,n_2}(w_1,w_2,t)-\left[\mu(1-w_1)+\frac{\theta}{2}(w_2-w_1)\right]\frac{\partial}{\partial w_1}G_{n_1,n_2}(w_1,w_2,t)-$$

$$-\left[\mu(1-w_2)+\frac{\theta}{2}(w_1-w_2)\right]\frac{\partial}{\partial w_2}G_{n_1,n_2}(w_1,w_2,t)=$$

$$=\left[\lambda(w_1-1)E(n_1(t))+\lambda(w_2-1)E(n_2(t))\right]G_{n_1,n_2}(w_1,w_2,t)$$

(donde $G_{n_1,n_2}(w_1,w_2,t)=\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}w_1^{n_1}w_2^{n_2}p_{(n_{10},n_{20})(n_1,n_2)}(t)$). De acuerdo con las condiciones iniciales planteadas, la solución particular de esta ecuación es de la forma:

$$G_{n_1,n_2}(w_1,w_2,t)=\left[1+\frac{1}{2}(w_1+w_2-2)e^{-\mu t}+\frac{1}{2}(w_1-w_2)e^{-(\mu+\theta)t}\right]^{n_{10}}\cdot$$

$$\cdot\exp\left\{\frac{n_{10}}{2}\left(e^{\lambda t}-1\right)\left[(w_1+w_2-2)e^{-\mu t}+(w_1-w_2)e^{-(\mu+\theta)t}\right]\right\}+$$

$$+\left[1+\frac{1}{2}(w_1+w_2-2)e^{-\mu t}+\frac{1}{2}(w_2-w_1)e^{-(\mu+\theta)t}\right]^{n_{20}}\cdot$$

$$\cdot\exp\left\{\frac{n_{20}}{2}\left(e^{\lambda t}-1\right)\left[(w_1+w_2-2)e^{-\mu t}+(w_2-w_1)e^{-(\mu+\theta)t}\right]\right\}$$

A partir de la esta expresión se obtienen la funciones generatrices de probabilidades de los procesos marginales:

$$G_{n_i}(w_1,w_2,t)=\left[1+\frac{1}{2}(w_i+w_j-2)e^{-\mu t}+\frac{1}{2}(w_i-w_j)e^{-(\mu+\theta)t}\right]\cdot$$

$$\cdot\exp\left\{\frac{e^{\lambda t}-1}{2}\left[(w_i+w_j-2)e^{-\mu t}+(w_i-w_j)e^{-(\mu+\theta)t}\right]\right\}\qquad (i,j=1,2;i\neq j)$$

En particular, si se supone que $n_{10}=1$ y $n_{20}=0$, se obtiene: **i)** la siguiente distribución de probabilidades conjunta:

$$p_{(1,0)(n_1,n_2)}(t)=\frac{e^{-(n_1+n_2)\mu t}}{n_1!n_2!}\left(\frac{1-e^{-\theta t}}{2}\right)^{n_1}\left(\frac{1-e^{-\theta t}}{2}\right)^{n_2}\left(e^{\lambda t}-1\right)^{n_1+n_2-1}\cdot$$

$$\cdot\left[\left(e^{\lambda t}-1\right)\left(1-e^{-\mu t}\right)+n_1+n_2\right]\exp\left[-\left(e^{\lambda t}-1\right)e^{-\mu t}\right]$$

y **ii)** las siguientes distribuciones de probabilidades marginales:

$$p_{(1,0)(n_1,\bullet)}(t) = \frac{e^{-\mu t}\left(1+e^{-\theta t}\right)}{2n_1!}\left[e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1+e^{-\theta t}}{2}\right]^{n_1-1} \bullet$$

$$\bullet\left[\left(e^{\lambda t}-1\right)\left(n_1+1\right)-e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1+e^{-\theta t}}{2}\right]\exp\left[-e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1+e^{-\theta t}}{2}\right]$$

$$p_{(1,0)(\bullet,n_2)}(t) = \frac{e^{-\mu t}\left(1-e^{-\theta t}\right)}{2n_2!}\left[e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1-e^{-\theta t}}{2}\right]^{n_2-1} \bullet$$

$$\bullet\left[\left(e^{\lambda t}-1\right)\left(n_2+1\right)-e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1-e^{-\theta t}}{2}\right]\exp\left[-e^{-\mu t}\left(e^{\lambda t}-1\right)\frac{1-e^{-\theta t}}{2}\right]$$

Para valores de t suficientemente grandes la distribución de probabilidades conjunta puede ser aproximada por una función de Poisson bidimensional de la forma:

$$p_{(1,0)(n_1,n_2)}(t) = \frac{\left[\frac{1}{2}e^{(\lambda-\mu)t}\right]^{n_1+n_2}}{n_1!n_2!}\exp\left[-e^{(\lambda-\mu)t}\right] \qquad (n_1,n_2 = 0,1,2,...)$$

Por otra parte, haciendo $w_1 = w_2 = 0$ en la forma compacta de la función $G_{n_1,n_2}(w_1,w_2,t)$, se obtiene la probabilidad de extinción de ambas subpoblaciones:

$$p_{(1,0)(0,0)}(t) = \left(1-e^{-\mu t}\right)\exp\left[-e^{-\mu t}\left(e^{\lambda t}-1\right)\right]$$

Si se verifica que $\lambda > \mu$, entonces los valores esperados $E[n_1(t)]$ y $E[n_2(t)]$ crecen exponencialmente para todo $t > 0$ y, por lo tanto, la probabilidad de extinción tiende rápidamente a cero, $\lim_{t\to\infty} p_{(1,0)(0,0)}(t) =$ $= \pi_{(1,0)(0,0)} = 0$. Por el contrario, si $\lambda < \mu$, los valores esperados decrecen exponencialmente y la probabilidad de extinción tiende rápidamente a la certeza, $\pi_{(1,0)(0,0)} = 1$. Si $\lambda = \mu$ queda definido un **proceso de inmigración-fallecimiento** en el cual $E[n_1(t)] \approx \frac{1}{2}$ y $E[n_2(t)] \approx \frac{1}{2}$ y, en consecuencia, $\pi_{(1,0)(0,0)} = e^{-1}$.

Como una generalización de este proceso, supóngase que las tasas de nacimiento y fallecimiento correspondientes a cada subpoblación sean λ_i y μ_i $(i = 1,2)$, respectivamente. Teniendo en cuenta que las funciones generatrices de probabilidades de los procesos marginales satisfacen las ecuaciones de Puri[162], se obtiene que:

$$\frac{d}{dt}G_{n_1}(w_1,w,t) = \lambda_1\left[G_{n_1}(w_1,w_2,t)\right]^2 - (\lambda_1+\mu_1)G_{n_1}(w_1,w_2,t) + \frac{\theta}{2}\left[G_{n_2}(w,w_2,t) - G_{n_1}(w_1,w_2t)\right] + \mu_1$$

$$\frac{d}{dt}G_{n_2}(w_1,w_2,t) = \lambda_2\left[G_{n_2}(w_1,w_2t)\right]^2 - (\lambda_2+\mu_2)G_{n_2}(w_1,w_2,t) + \frac{\theta}{2}\left[G_{n_1}(w_1,w_2,t) - G_{n_2}(w_1,w_2,t)\right] + \mu_2$$

---

[162]. Ver Puri, P.S. (1968).

Las soluciones particulares de estas ecuaciones son de la forma:

$$G_{n_1}(w_1,w_2,t) = \sum_{n=0}^{\infty} A_n(w_1,w_2,t)\left(\frac{\theta}{2}\right)^n$$

$$G_{n_2}(w_1,w_2,t) = \sum_{n=0}^{\infty} B_n(w_1,w_2,t)\left(\frac{\theta}{2}\right)^n$$

Donde A_n y B_n son independientes de $\frac{\theta}{2}$ y tales que:

$$A_0(w_1,w_2,t) = \frac{\mu_{1_1}(1-w_1)-(\mu_1-\lambda_1 w_1)e^{-(\lambda_1-\mu_1)t}}{\lambda_1(1-w_1)-(\mu_1-\lambda_1 w_1)e^{-(\lambda_1-\mu_1)t}}$$

$$A_n(w_1,w_2,t) = \int_0^t \left[B_{n-1}(w_1,w_2,\tau) - A_{n-1}(w_1,w,\tau) + \lambda_1 \sum_{j=1}^{\infty} A_j(w_1,w_2,\tau) A_{n-j}(w_1,w_2,\tau) \right].$$

$$.\exp\left\{ \int_\tau^t \left[2\lambda_1 A_0(w_1,w_2,x) - (\lambda_1 + w_1) \right] dx \right\} d\tau \qquad (n = 1,2,...)$$

$$B_0(w_1,w_2,t) = \frac{\mu_2(1-w_2)-(\mu_2-\lambda_2 w_2)e^{-(\lambda_2-\mu)t}}{\lambda_2(1-w_2)-(\mu_2-\lambda_2 w_2)e^{-(\lambda_2-\mu_2)t}}$$

$$B_n(w_1,w_2,t) = \int_0^t \left[A_{n-1}(w_1,w_2,\tau) - B_{n-1}(w_1,w_2,\tau) + \lambda_2 \sum_{j=1}^{n-1} B_j(w_1,w_2,\tau) B_{n-j}(w_1,w_2,\tau) \right].$$

$$.\exp\left\{ \int_\tau^t \left[2\lambda_2 B_0(w_1,w_2,x) - (\lambda_2 + \mu_2) \right] dx \right\} d\tau \qquad (n = 1,2,...)$$

1.7.5.- El proceso de nacimiento-fallecimiento lineal bidimensional

Sea $\{n(t)\} = \{f(t) + v(t)\}$ un proceso bidimensional que representa el tamaño de una población y sean $\{f(t)\}$ y $\{v(t)\}$ los procesos marginales, que representan los tamaños de las subpoblaciones femenina y masculina, respectivamente.

Sean: **i)** $\lambda_f \Delta t = \alpha\lambda\Delta t$ la probabilidad de que un individuo de sexo femenino dé origen a un nuevo individuo de sexo femenino en un intervalo $(t, t+\Delta t]$ (donde α denota la proporción de nacimientos femeninos sobre el total de nacimientos, es decir, el **índice de femineidad**); **ii)** $\lambda_v \Delta t = \beta\lambda\Delta t$ la probabilidad de que un individuo de sexo femenino dé origen a un nuevo individuo de sexo masculino en un intervalo $(t, t+\Delta t]$ (donde β denota la proporción de nacimientos masculinos sobre el total de

nacimientos, es decir, el **índice de masculinidad**)[163]; **iii)** $\mu_f \Delta t$ la probabilidad de que un individuo de sexo femenino fallezca en un intervalo $(t, t+\Delta t]$ y **iv)** $\mu_v \Delta t$ la probabilidad de que un individuo de sexo masculino fallezca en un intervalo $(t, t+\Delta t]$ [164].

Supóngase, además: **i)** que se verifican las condiciones de linealidad (es decir, que no se pueden producir dos o más eventos en un mismo instante t) y de autorregresividad de primer orden (es decir que las ocurrencias de los eventos son independientes para intervalos de tiempo disjuntos) y **ii)** que los tamaños iniciales de las subpoblaciones son $f(0) = f_0$ y $v(0) = v_0$.

A partir de estas condiciones se obtienen: **i)** el siguiente sistema de probabilidades de transición infinitesimal:

$$p_{(f,v)(f+1,v)}(t, t+\Delta t) = f\lambda_f \Delta t + o(\Delta t)$$

$$p_{(f,v)(f,v+1)}(t, t+\Delta t) = f\lambda_v \Delta t + o(\Delta t)$$

$$p_{(f,v)(f-1,v)}(t, t+\Delta t) = f\mu_f \Delta t + o(\Delta t)$$

$$p_{(f,v)(f,v-1)}(t, t+\Delta t) = v\mu_v \Delta t + o(\Delta t)$$

$$p_{(f,v)(f\pm k_f,v)}(t, t+\Delta t) = o(\Delta t) \qquad (k_f = 2,3,4,\ldots)$$

$$p_{(f,v)(f,v\pm k_v)}(t, t+\Delta t) = o(\Delta t) \qquad (k_v = 2,3,4,\ldots)$$

$$p_{(f,v)(f,v)}(t, t+\Delta t) = 1 - f(\lambda_f + \lambda_v + \mu_f)\Delta t - v\mu_v \Delta t - o(\Delta t)$$

ii) el siguiente sistema de ecuaciones estocásticas diferenciales en-diferencias de Kolmogorov:

$$\frac{d}{dt} p_{(f_0,v_0)(f,v)}(t) = \left[- f(\lambda_f + \lambda_v + \mu_f) - v\mu_v \right] p_{(f_0,v_0)(f,v)}(t) + (f-1)\lambda_f p_{(f_0,v_0)(f-1,v)}(t) +$$
$$+ f\lambda_v p_{(f_0,v_0)(f,v-1)}(t) + (f+1)\mu_f p_{(f_0,v_0)(f+1,v)}(t) +$$
$$+ (v+1)\mu_v p_{(f_0,v_0)(f,v+1)}(t) \qquad (f,v = 0,1,2,\ldots; t \geq 0)$$

iii) la siguiente ecuación en derivadas parciales para la función generatriz de probabilidades:

$$\frac{\partial}{\partial t} G_{f,v}(w_1, w_2, t) = \left[- (\lambda_f + \lambda_v + \mu_f) w_1 + \lambda_f w_1^2 + \lambda_v w_1 w_2 + \mu_f \right] \frac{\partial}{\partial w_1} G_{f,v}(w_1, w_2, t) +$$
$$+ \mu_v (w_2 - 1) \frac{\partial}{\partial w_2} G_{f,v}(w_1, w_2, t) \qquad (|w_1|, |w_2| \leq 1)$$

---

[163]. Obviamente se verifica que $\lambda_f + \lambda_v = \lambda$.

[164]. Obviamente se verifica que $\mu_f + \mu_v = \mu$.

(donde $G_{f,v}(w_1,w_2,t)=\sum_{f=0}^{\infty}\sum_{v=0}^{\infty}w_1^f w_2^v p_{(f_0,v_0)(f,v)}(t)$) y **iv)** la siguiente ecuación en derivadas parciales para la función generatriz de momentos:

$$\frac{\partial}{\partial t}M_{f,v}(w_1,w_2,t)=\left[-\left(\lambda_f+\lambda_v+\mu_f\right)+\lambda_f e^{w_1}+\lambda_v e^{w_2}+\mu_f e^{-w_1}\right]\frac{\partial}{\partial w_1}M_{f,v}(w_1,w_2,t)+$$

$$+\mu_v\left(e^{-w_2}-1\right)\frac{\partial}{\partial w_2}M_{f,v}(w_1,w_2,t)\qquad\left(|w_1|,|w_2|\le 1\right)$$

(donde $M_{f,v}(w_1,w_2,t)=\sum_{f=0}^{\infty}\sum_{v=0}^{\infty}e^{w_1 f}e^{w_2 v}p_{(f_0,v_0)(f,v)}(t)$).

De esta última ecuación se obtienen los momentos[165]:

$$m_{10}(t)=E\left[f(t)/f_0\right]=f_0e^{\left(\lambda_f-\mu_f\right)t}$$

$$m_{01}(t)=E\left[v(t)/f_0,v_0\right]=\frac{\lambda_v f_0}{\lambda_f-\mu_f+\mu_v}\left[e^{\left(\lambda_f-\mu_f\right)t}-e^{-\mu_v t}\right]+v_0e^{-\mu_v t}$$

$$m_{20}(t)=E\left[f^2(t)/f_0\right]=\frac{\lambda_f+\mu_f}{\lambda_f-\mu_f}f_0e^{\left(\lambda_f-\mu_f\right)t}\left[e^{\left(\lambda_f-\mu_f\right)t}-1\right]+f_0^2e^{2\left(\lambda_f-\mu_f\right)t}$$

$$m_{02}(t)=E\left[v^2(t)/f_0,v_0\right]=\frac{\lambda_v f_0}{\mu_v\left(\lambda_f-\mu_f+\mu_v\right)^2}\left\{\left(\frac{\lambda_f+\mu_f}{\lambda_f-\mu_f}-f_0\right)\left[e^{2\left(\lambda_f-\mu_f\right)t}-e^{-2\mu_v t}\right]-\right.$$

$$\left.-\left[\left(\lambda_f+\mu_f\right)+\mu_v\left(\lambda_f v_0-\lambda_v f_0\right)-\mu_v\left(\mu_f-\mu_v\right)v_0\right]e^{-\mu_v t}\left[e^{\left(\lambda_f-\mu_f\right)t}-e^{-\mu_v t}\right]\right\}+$$

$$+\frac{\lambda_f v_0}{\mu_v\left(\lambda_f-\mu_f+\mu_v\right)}\left[\mu_v-\frac{2\left(\lambda_f+\mu_f\right)\left(\lambda_f-\mu_f+\mu_v\right)}{\left(\lambda_f-\mu_f\right)\left(\lambda_f-\mu_f+2\mu_v\right)}\right]\left[e^{\left(\lambda_f-\mu_f\right)t}-2e^{-2\mu_v t}\right]+$$

$$+\left[\left(v_0-\frac{\lambda_v f_0}{\lambda_f-\mu_f+\mu_v}\right)\left(1-e^{-\mu_v t}\right)+v_0^2e^{-\mu_v t}\right]e^{-\mu_v t}$$

$$m_{11}(t)=E\left[f(t)v(t)/f_0,v_0\right]=\left\{\frac{\left(\lambda_f+\mu_f\right)\lambda_v f_0}{\lambda_v\left(\lambda_f-\mu_f\right)\left(\lambda_f-\mu_f+\mu_v\right)}\left[\mu_v e^{\left(\lambda_f-\mu_f\right)t}+\left(\lambda_f-\mu_f\right)e^{-\mu_v t}-\right]\right.$$

$$\left.-\left(\lambda_f-\mu_f+\mu_v\right)\right]+f_0m_{01}(t)\Big\}e^{\left(\lambda_f-\mu_f\right)t}$$

---

[165]. Obsérvese que las intensidades de nacimiento dependen exclusivamente de la población femenina. Por esta razón a este tipo de procesos se los denomina **con dominancia femenina**. Para un proceso sin dominancias, la ecuación en derivadas parciales para la función generatriz de probabilidades bidimensional será de la forma $\frac{\partial}{\partial t}G_{f,v}(w_1,w_2t)=\left[-\left(1+\lambda_f\right)w_1+\lambda_f w_1^2+\lambda_f w_1w_2+\mu_f\right]\cdot\frac{\partial}{\partial w_1}G_{f,v}(w_1,w_2,t)+\left[-\left(1+\lambda_v\right)w_2+\lambda_v w_2^2+\lambda_v w_1w_2+\mu_v\right]\frac{\partial}{\partial w_2}G_{f,v}(w_1,w_2,t)$.

A partir de los valores esperados de los procesos marginales se obtiene la proporción esperada de mujeres en la población, al momento t (es decir, el **índice esperado de femineidad al momento** t):

$$I_f(t) = \frac{e\left[f(t)/f_0\right]}{E\left[f(t)+v(t)/f_0,v_0\right]} =$$

$$= f_0\left(\lambda_f - \mu_f + \mu_v\right)e^{\left(\lambda_f-\mu_f+\mu_v\right)t}\left\{f_0\left(\lambda_f - \mu_f + \mu_v\right)e^{\left(\lambda_f-\mu_f+\mu_v\right)t} + \lambda_v f_0\left[e^{\left(\lambda_f-\mu_f+\mu_v\right)t} - 1\right] + v_0\left(\lambda_f - \mu_f + \mu_v\right)\right\}^{-1}$$

Luego, el índice esperado de femineidad en el límite estará dado por:

$$\lim_{t\to\infty} I_f(t) = \frac{\lambda_f - \mu_f + \mu_v}{\left(\lambda_f + \lambda_v\right) - \left(\mu_f - \mu_v\right)} \qquad \left(\lambda_f + \lambda_v > \mu_f\right)$$

Por otra parte, se concluye en forma inmediata que:

$$\lim_{t\to\infty} E\left[f(t)/f_0\right] = \begin{cases} 0 & si \quad \lambda_f < \mu_f \\ f_0 & si \quad \lambda_f = \mu_f \\ \infty & si \quad \lambda_f > \mu_f \end{cases}$$

$$\lim_{t\to\infty} E\left[v(t)/f_0,v_0\right] = \begin{cases} 0 & si \quad \lambda_f \le \mu_f \\ \infty & si \quad \lambda_f > \mu_f \end{cases}$$

1.7.6.- El proceso de enfermedad-fallecimiento

Sea $\{n(t)\}$ un proceso estocástico que representa el tamaño de una población multidimensional, formada por subpoblaciones correspondientes a los **estados de enfermedad**, e_1 y e_2 , y a los **estados de fallecimiento**, $f_1, f_2, \ldots, f_k$ [166]. Se supone que un individuo puede abandonar un estado de enfermedad por recuperación, fallecimiento o por contraer otra enfermedad, es decir que los estados de enfermedad son no-recurrentes y los estados de fallecimiento son absorbentes. En muchas aplicaciones el estado e_1 puede ser considerado como un estado de salud y e_2 como un estado de enfermedad, de modo que la transición de e_1 a e_2 puede ser considerada como una enfermedad o una recaída y la transición de e_2 a e_1 puede ser considerada como una recuperación.

Sean: **i)** $\xi_{ij}\ (i,j = 1,2)$ la intensidad de transición de un estado de enfermedad e_i a otro estado de enfermedad e_j (en general, para cada estado inicial e_i, existe una función continua $\xi_{ii}(t) \le 0$ tal que $\lim_{\Delta t\to 0} \frac{1 - p_{ii}(t,t+\Delta t)}{\Delta t} = -\xi_{ii}(t)$ y, para cada par de estados e_i y $e_j\ (i \ne j)$,existe una función continua

---

[166]. Este proceso se debe a Fix, E.; Neyman, J. (1951), quienes estudiaron las probabilidades de recaída, recuperación y fallecimiento en pacientes afectados por cáncer.

$\xi_{ij}(t) \geq 0$, tal que $\lim_{\Delta t \to 0} \frac{p_{ij}(t, t+\Delta t)}{\Delta t} = \xi_{ij}(t)$, denominadas **funciones intensidad de transición en los estados de enfermedad)** y **ii)** $\mu_{ih}\ (i = 1,2; h = 1,2,...,k)$ la intensidad de transición de un estado de enfermedad e_i a un estado de fallecimiento f_h [167]. De modo que: **i)** la probabilidad de que un individuo, que se encuentra en el estado de enfermedad e_i en el momento t, pase al estado de enfermedad e_j en un intervalo $(t, t+\Delta t]$ será de la forma $\xi_{ij}\Delta t\ (i, j = 1,2; i \neq j)$; **ii)** la probabilidad de que un individuo, que se encuentra en un estado de enfermedad e_i en el momento t, pase a un estado de fallecimiento f_h en un intervalo $(t, t+\Delta t]$, será de la forma $\mu_{ih}\Delta t\ (i = 1,2; h = 1,2,...,k)$; **iii)** la probabilidad de que un individuo, que se encuentra en un estado de enfermedad e_i en el momento t, permanezca en dicho estado al momento $t+\Delta t$, será de la forma $1 + \xi_{ii}\Delta t\ (i = 1,2)$, donde la intensidad de permanencia en el estado e_i puede ser expresada de la siguiente forma: $\xi_{ii} = -\left(\xi_{ij} + \sum_{h=1}^{k} \mu_{ih}\right)$ (obsérvese que, de acuerdo con esta expresión, se obtiene que $\xi_{ii} \leq 0$)[168].

Sean: **i)** $p_{ij}(t_0, t)\ (i, j = 1,2; t > t_0)$ la probabilidad de que un individuo, que se encuentra en el estado $e_i\ (i = 1,2)$ en el momento t_0, se encuentre en el estado $e_j\ (j = 1,2; j \neq i)$ en el momento t **(probabilidad de transición en los estados de enfermedad); ii)** $q_{ih}(t_0, t)\ (i = 1,2; h = 1,2,...,k; t > t_0)$ la probabilidad de que un individuo, que se encuentra en el estado $e_i\ (i = 1,2)$ en el momento t_0, se encuentre en el estado $f_h\ (h = 1,2,...,k)$ en el momento t **(probabilidad de transición a los estados de fallecimiento)**.

Estas probabilidades satisfacen las siguientes condiciones iniciales:

$$p_{ij}(t_0, t_0) = \begin{cases} 1 & para \quad j = i; i = 1,2 \\ 0 & para \quad j \neq i; i, j = 1,2 \end{cases}$$

$$q_{ih}(t, t) = 0 \qquad (i = 1,2; h = 1,2,...,k)$$

De acuerdo con la condición de autorregresividad de primer orden, la probabilidad de que un individuo, que se encuentra en el estado $e_i\ (i = 1,2)$ en el momento t_0, se encuentre en el estado $e_j\ (j = 1,2)$ en el momento τ y se encuentre en el estado $e_h\ (h = 1,2)$ en el momento $t\ (t_0 \leq \tau \leq t)$, puede ser expresada de la siguiente forma $p_{ij}(t_0, \tau)p_{jh}(\tau, t)\ (i, j, h = 1,2)$ y, en consecuencia, a partir de la condicción de Kolmogorov-Chapman, se verifica que $p_{ij}(t_0, t) = \sum_{j=1}^{2} p_{ij}(t_0, \tau)p_{jh}(\tau, t)$. Esta relación permite definir el siguiente sistema de ecuaciones de transición infinitesimal en los estados de enfermedad:

---

[167]. En este caso en particular se ha supuesto que estas intensidades son independientes de t.

[168]. Si se verifica que $\xi_{ii} = 0$, entonces el estado e_i es absorbente. En ese caso, el estado e_i debe ser considerado como un estado de fallecimiento.

$$p_{ii}(t_0,t+\Delta t) = p_{ii}(t_0,t)(1+\xi_{ii}\Delta t) + p_{ij}(t_0,t)\xi_{ji}\Delta t + o(\Delta t) \qquad (i,j = 1,2)$$

$$p_{ij}(t_0,t+\Delta t) = p_{ii}(t_0,t)\xi_{ij}\Delta t + p_{ij}(t_0,t)(1+\xi_{jj}\Delta t) + o(\Delta t) \qquad (i,j = 1,2; i \neq j)$$

a partir del cual se obtiene el siguiente sistema de ecuaciones diferenciales de primer orden lineales homogéneas:

$$\frac{\partial}{\partial t} p_{ii}(t_0,t) = p_{ii}(t_0,t)\xi_{ii} + p_{ij}(t_0,t)\xi_{ji} \qquad (i,j = 1,2; i \neq j)$$

$$\frac{\partial}{\partial t} p_{ij}(t_0,t) = p_{ii}(t_0,t)\xi_{ij} + p_{ij}(t_0,t)\xi_{jj} \qquad (i,j = 1,2; i \neq j)$$

cuyas soluciones son de la forma $p_{ii}(t_0,t) = c_{ii}e^{\rho t}$ y $p_{ij}(t_0,t) = c_{ij}e^{\rho t}$. Sustituyendo estas expresiones en las ecuaciones anteriores, se obtiene que:

$$\rho c_{ii}e^{\rho t} = c_{ii}e^{\rho t}\xi_{ii} + c_{ij}e^{\rho t}\xi_{ji}$$

$$\rho c_{ij}e^{\rho t} = c_{ii}e^{\rho t}\xi_{ij} + c_{ij}e^{\rho t}\xi_{jj}$$

o, lo que es lo mismo, que:

$$(\rho - \xi_{ii})c_{ii} - \xi_{ji}c_{ij} = 0$$

$$-\xi_{ij}c_{ii} + (\rho - \xi_{jj})c_{ij} = 0$$

La condición necesaria y suficiente para que este sistema de ecuaciones proporcione soluciones no-triviales para c_{ii} y c_{ij} es que su ecuación característica sea tal que $\begin{vmatrix} \rho - \xi_{ii} & -\xi_{ji} \\ -\xi_{ij} & \rho - \xi_{jj} \end{vmatrix} = 0$. De modo que:

$$\rho_1 = \frac{1}{2}\left[\xi_{ii} + \xi_{jj} + \sqrt{(\xi_{ii} - \xi_{jj})^2 + 4\xi_{ij}\xi_{ji}}\right]$$

$$\rho_2 = \frac{1}{2}\left[\xi_{ii} + \xi_{jj} - \sqrt{(\xi_{ii} - \xi_{jj})^2 + 4\xi_{ij}\xi_{ji}}\right]$$

(obsérvese que, dado que $\xi_{ij} > 0$ y $\xi_{ji} > 0$, el discriminante de estas expresiones es estrictamente positivo y, por lo tanto, las raíces son reales, distintas y estrictamente negativas).

Del sistema de ecuaciones anterior se obtienen las siguientes relaciones para los coeficientes c_{ii} y c_{ij} correspondientes a cada raíz ρ_h $(h = 1,2)$:

$$\frac{c_{ii,h}}{c_{ij,h}} = \frac{\xi_{ij}}{\rho_h - \xi_{ii}} = \frac{\rho_i - \xi_{jj}}{\xi_{ij}} \qquad (i,j = 1,2; i \neq j)$$

De modo que las soluciones generales para el sistema de ecuaciones diferenciales pueden ser expresadas de la siguiente forma:

$$p_{ii}(t_0,t) = \sum_{h=1}^{2} \frac{c_{ij,h}}{\xi_{ij}}\left(\rho_h - \xi_{jj}\right)e^{\rho_h t} \qquad (i,j = 1,2; i \neq j)$$

$$p_{ij}(t_0,t) = \sum_{h=1}^{2} c_{ij,h}e^{\rho_h t} \qquad (i,j = 1,2; i \neq j)$$

Entonces, haciendo $t_0 = t$, será:

$$p_{ii}(t,t) = \sum_{h=1}^{2} \frac{c_{ij,h}}{\xi_{ij}}\left(\rho_h - \xi_{jj}\right)e^{\rho_h t} = 1$$

$$p_{ij}(t,t) = \sum_{h=1}^{2} c_{ij,h}e^{\rho_h t} = 0$$

Luego, resulta que $\frac{c_{ij,h}}{\xi_{ij}} = \frac{1}{\rho_h - \rho_\ell}e^{-\rho_h t}$ $(h,\ell = 1,2; h \neq \ell)$ y, por lo tanto, se obtienen las siguientes soluciones particulares:

$$p_{ii}(t_0,t) = \sum_{h=1}^{2} \frac{\rho_h - \xi_{jj}}{\rho_h - \rho_\ell}e^{\rho_h(t-t_0)}$$

$$p_{ij}(t_0,t) = \sum_{h=1}^{2} \frac{\xi_{ij}}{\rho_h - \rho_\ell}e^{\rho_h(t-t_0)}$$

De la misma forma, la probabilidad de que un individuo, que en el momento inicial se encuentra en el estado e_i $(i = 1,2)$, alcance el estado f_h $(h = 1,2,...,k)$ en un intervalo $(t,t+\Delta t]$, puede ser expresada como una suma de probabilidades de eventos incompatibles $p_{ii}(t)\mu_{ih}dt + p_{ij}(t)\mu_{jh}dt$ y, por lo tanto, se puede escribir:

$$q_{ih}(t) = \int_0^t p_{ii}(\tau)\mu_{ih}d\tau + \int_0^t p_{ij}(\tau)\mu_{jh}d\tau =$$

$$= \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell(\rho_\ell - \rho_m)}\left[\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}\right] \qquad (\ell \neq m; i \neq j; i,j = 1,2; h = 1,2,...,k)$$

Además, dado que $\sum_{j=1}^{2} p_{ij}(t) + \sum_{h=1}^{k} q_{ih}(t) = 1$, se obtiene, en forma inmediata, que:

$$\sum_{h=1}^{k} q_{ih}(t) = \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell(\rho_\ell - \rho_m)}\left[\left(\rho_\ell - \xi_{jj}\right)\sum_{h=1}^{k}\mu_{ih} + \xi_{ij}\sum_{h=1}^{k}\mu_{jh}\right] =$$

$$= \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell(\rho_\ell - \rho_m)}\left[-\left(\rho_\ell - \xi_{jj}\right)\left(\xi_{ii} + \xi_{ij}\right) - \xi_{ij}\left(\xi_{ji} + \xi_{jj}\right)\right] =$$

$$= \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell(\rho_\ell - \rho_m)}\left(-\rho_\ell^2 + \rho_\ell\xi_{jj} - \rho_\ell\xi_{ij}\right) =$$

$$= 1 - \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t}}{\rho_\ell - \rho_m}\left[\left(\rho_\ell - \xi_{jj}\right) + \xi_{ij}\right]$$

Si en las definiciones anteriores, se hace $t_0 = 0$, teniendo en cuenta que, como se mencionó en la página anterior, las raíces ρ_ℓ $(\ell = 1,2)$ son negativas, se obtienen las siguientes probabilidades en el límite:

$$\lim_{t\to\infty} p_{ii}(t) = p_{ii}(\infty) = \lim_{t\to\infty} \sum_{\ell=1}^{2} \frac{\rho_\ell - \xi_{jj}}{\rho_\ell - \rho_m} e^{\rho_\ell t} = 0 \qquad (i,j = 1,2)$$

$$\lim_{t\to\infty} p_{ij}(t) = p_{ij}(\infty) = \lim_{t\to\infty} \sum_{\ell=1}^{2} \frac{\xi_{ij}}{\rho_\ell - \rho_m} e^{\rho_\ell t} = 0 \qquad (i,j = 1,2; i \neq j)$$

$$\lim_{t\to\infty} q_{ih}(t) = q_{ih}(\infty) = \lim_{t\to\infty} \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell\left(\rho_\ell - \rho_m\right)}\left[\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}\right] =$$

$$= \sum_{\ell=1}^{2} \frac{-\left[\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}\right]}{\rho_\ell\left(\rho_\ell - \rho_m\right)} \qquad (i = 1,2; h = 1,2,\ldots,k)$$

$$\sum_{h=1}^{2} q_{ih}(\infty) = \sum_{h=1}^{k}\sum_{\ell=1}^{2} \frac{-\left[\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}\right]}{\rho_\ell\left(\rho_\ell - \rho_m\right)} = 1$$

Sea, ahora, un individuo que en el instante inicial se encuentra en el estado e_i y sean las siguientes funciones indicador:

$$I_{ij}(\tau) = \begin{cases} 1 \text{ } \textit{si dicho individuo se encuentra en el estado } e_j \textit{ al momento } \tau \ (0 \le \tau \le t) \\ \\ 0 \text{ } \textit{en cualquier otro caso} \end{cases}$$

$$J_{ih}(\tau) = \begin{cases} 1 \text{ } \textit{si dicho individuo se encuentra en el estado } f_h \textit{ al momento } \tau \ (0 \le \tau \le t) \\ \\ 0 \text{ } \textit{en cualquier otro caso} \end{cases}$$

(donde $i,j = 1,2; h = 1,2,\ldots,k$), cuyos valores esperados son $E\left[I_{ij}(\tau)\right] = p_{ij}(\tau)$ y $E\left[J_{ih}(\tau)\right] = q_{ih}(\tau)$. Se obtiene, entonces, que el tiempo esperado de permanencia de dicho individuo en el estado e_j $(j = 1,2)$ en el intervalo $[0,t]$, está dado por $E_{ij}^{(e)}(t) = E\left[\int_0^t I_{ij}(\tau)\,d\tau\right] = \int_0^t p_{ij}(\tau)\,d\tau$. Sustituyendo en esta expresión la definición obtenida en páginas anteriores para las probabilidades de transición en los estados de enfermedad, resulta que:

$$E_{ii}^{(e)}(t) = \int_0^t \sum_{\ell=1}^{2} \frac{\rho_\ell - \xi_{jj}}{\rho_\ell - \rho_m} e^{\rho_\ell \tau}\,d\tau = \sum_{\ell=1}^{2} \frac{\rho_\ell - \xi_{jj}}{\rho_\ell\left(\rho_\ell - \rho_m\right)}\left(e^{\rho_\ell t} - 1\right) \qquad (i,j = 1,2; i \neq j)$$

$$E_{ij}^{(e)}(t) = \int_0^t \sum_{\ell=1}^{2} \frac{\xi_{ij}}{\rho_\ell - \rho_m} e^{\rho_\ell \tau} d\tau = \sum_{\ell=1}^{2} \frac{\xi_{ij}}{\rho_\ell(\rho_\ell - \rho_m)}\left(e^{\rho_\ell t} - 1\right) \qquad (i, j = 1,2; i \neq j)$$

De la misma forma, el tiempo esperado de demora en arribar al estado f_h $(h = 1,2,...,k)$ para un individuo que se encuentra inicialmente en el estado e_i $(i = 1,2)$, está dado por:

$$E_{ih}^{(f)}(t) = E\left[\int_0^t J_{ih}(\tau)d\tau\right] = \int_0^t q_{ih}(\tau)d\tau =$$

$$= \int_0^t \sum_{\ell=1}^{2} \frac{e^{\rho_\ell t} - 1}{\rho_\ell(\rho_\ell - 1)}\left[\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}\right]d\tau =$$

$$= \sum_{\ell=1}^{2}\left[\frac{1}{\rho_\ell}\left(e^{\rho_\ell t} - 1\right) - t\right]\frac{\left(\rho_\ell - \xi_{jj}\right)\mu_{ih} + \xi_{ij}\mu_{jh}}{\rho_\ell(\rho_\ell - \rho_m)} \qquad (i, j = 1,2; h = 1,2,...,k; i \neq j)$$

Supóngase ahora una población inicial:

$$n(0) = n_1^e(0) + n_2^e(0) = n_1^e(t) + n_2^e(t) + n_1^f(t) + ... + n_k^f(t)$$

donde $n_j^e(t) = n_{1j}^e(t) + n_{2j}^e(t)$ y $n_h^f(t) = n_{1h}^f(t) + n_{2h}^f(t)$ $(h = 1,2,...,k)$ (siendo: **i)** $n_j^e(t)$ $(j = 1,2)$ el tamaño de la población que se encuentra en el estado e_j al momento t ; **ii)** $n_{ij}^e(t)$ $(i, j = 1,2)$ la cantidad de individuos que se encontraban inicialmente en el estado e_i y se encuentran en el estado e_j al momento t; **iii)** $n_h^f(t)$ $(h = 1,2,...,k)$ el tamaño de la población que se encuentra en el estado f_h al momento t y **iv)** $n_{ih}^f(t)$ $(i = 1,2,; h = 1,2,...,k)$ la cantidad de individuos que se encontraban inicialmente en el estado e_i y se encuentran en el estado f_h al momento t). Se verificará, entonces, que:

$$n_i^e(0) = n_{i1}^e(t) + n_{i2}^e(t) + n_{i1}^f(t) + ... + n_{ik}^f(t) \qquad (i = 1,2)$$

donde el segundo miembro es una variable multinomial con función generatriz de probabilidades de la forma:

$$G_{n_{i1}^e(t),n_{i2}^e(t),n_{i1}^f(t),...,n_{ik}^f(t)}(w_1, w_2, z_1, ..., z_k, t) = E\left[w_1^{n_{i1}^e(t)} w_2^{n_{i2}^e(t)} z_1^{n_{i1}^f(t)} ... z_k^{n_{ik}^f(t)} / n_i^e(0)\right] =$$

$$= \left[p_{i1}(t)w_1 + p_{i2}(t)w_2 + q_{i1}(t)z_1 + ... + q_{ik}z_k\right]^{n_i^e(0)} \qquad (i = 1,2)$$

En consecuencia, la función generatriz de probabilidades de la variable multinomial que representa el tamaño de la población total al momento t, será de la forma:

$$G_{n_1^e(t),n_2^e(t),n_1^f(t),...,n_k^f(t)}(w_1, w_2, z_1, ... z_k, t) = E\left[w_1^{n_1^e(t)} w_2^{n_2^e(t)} z_1^{n_1^f(t)} ... z_k^{n_k^f(t)} / n_1^e(0), n_2^e(0)\right] =$$

$$= \prod_{i=1}^{2}\left[p_{i1}(t)w_1 + p_{i2}(t)w_2 + q_{i1}(t)z_1 + ... + q_{ik}(t)z_k\right]^{n_i^e(t)}$$

De modo que la distribución de probabilidades conjunta de esta variable será de la forma:

$$p\left\{\left[n_1^e(t)=n_1^e\right]\cap\left[n_2^e(t)=n_2^e\right]\cap\left[n_1^f(t)=n_1^f\right]\cap..\cap\left[n_k^f(t)=n_k^f\right]\right\}=$$

$$=\sum_{n_{ij}^e,n_{ih}^f}\prod_{i=1}^{2}\frac{n_i^e(0)!}{n_{i1}^e!n_{i2}^e!n_{i1}^f!...n_{ik}^f!}\left[p_{i1}(t)\right]^{n_{i1}^e}\left[p_{i2}(t)\right]^{n_{i2}^e}\left[q_{i1}(t)\right]^{n_{i1}^f}...\left[q_{ik}(t)\right]^{n_{ik}^f}$$

(donde $\sum_{n_{ij}^e,n_{ih}^f}$ denota la suma sobre todos los posibles valores de n_{ij}^e y n_{ih}^f $(i,j=1,2;h=1,2,...,k)$). En consecuencia, el tamaño esperado de las poblaciones en cada uno de los estados de enfermedad estará dado por:

$$E\left[n_j^e(t)/n_1^e(0),n_2^e(0)\right]=n_1^e(0)p_{1j}(t)+n_2^e(0)p_{2j}(t) \qquad (j=1,2)$$

y el tamaño esperado de las poblaciones en cada uno de los estados de fallecimiento, estará dado por:

$$E\left[n_h^f(t)/n_1^e(0),n_2^e(0)\right]=n_1^e(0)q_{ih}(t)+n_2^e(0)q_{2h}(t)$$

Las correspondientes varianzas serán de la forma:

$$\sigma^2\left[n_j^e(t)\right]=\sum_{i=1}^{2}n_i^e(0)p_{ij}(t)\left[1-p_{ij}(t)\right] \qquad (j=1,2)$$

$$\sigma^2\left[n_h^f(t)\right]=\sum_{i=1}^{2}n_i^e(0)q_{ih}(t)\left[1-q_{ih}(t)\right] \qquad (h=1,2,...,k)$$

De los resultados anteriores se obtiene en forma inmediata que la probabilidad de extinción de la población al momento t, está dada por:

$$p\left\{\left[n_1^e(t)=0,n_2^e(t)=0\right]/n_1^e(0),n_2^e(0)\right\}=\prod_{i=1}^{2}\left[q_{i1}(t)+...+q_{ik}(t)\right]^{n_i^e(0)}=$$

$$=\prod_{i=1}^{2}\left\{1-\sum_{\ell=1}^{2}\frac{e^{\rho_\ell t}}{\rho_\ell-\rho_m}\left[\left(\rho_\ell-\xi_{jj}\right)+\xi_{ij}\right]\right\}^{n_i(0)}$$

Además, dado que:

$$\lim_{t\to\infty}G_{n_1^e(t),n_2^e(t),n_1^f(t),...,n_k^f(t)}\left(w_1,w_2,z_1,...,z_k,t\right)=\prod_{i=1}^{2}\left[q_{i1}(\infty)z_1+...+q_{ik}(\infty)z_k\right]^{n_i^e(0)}$$

se puede concluir en forma inmediata que el límite de la distribución de probabilidades conjunta está definida por la convolución de dos distribuciones multinomiales independientes (ver Landro, A.H. (2002)), con probabilidades $q_{1h}(\infty)$ y $q_{2h}(\infty)$, respectivamente.

CAPITULO 2

Los procesos estacionarios de parámetro discreto

"El tiempo es como un círculo que girara infinitamente:
El arco que desciende es el pasado, el que asciende es el porvenir; arriba hay un punto indivisible que toca la tangente y es el ahora".

(Borges, J.L.: "Nueva refutación del tiempo")

2.1.- Definiciones

Como se mencionó en la Sec. 1.3, se define como **serie cronológica** $\{Y_t\}$ a todo proceso estocástico discreto en el dominio del tiempo, continuo en el dominio de las variables, con la única restricción -la cual no significa pérdida de generalidad- de que sus variables aleatorias se generen a intervalos regulares de tiempo.

Asimismo, de acuerdo a lo expresado en la misma sección, también se denomina serie cronológica a la sucesión finita y ordenada en el tiempo de las **realizaciones** de las variables aleatorias que integran el proceso[1].

En general, el proceso estocástico definido por la serie cronológica se denota por $\{Y_t\}_{-\infty}^{\infty}$, en tanto que la realización de dicho proceso (que define su **trayectoria** en el tiempo) se denota por $\{y_t\}_1^n$. A partir de la condición de ergodicidad, cualquier análisis sobre $\{Y_t\}$ se desarrollará asimilando el conjunto $(-\infty,\infty)$ del proceso al conjunto $[1;n]$ de su realización, de modo que la relación entre **proceso** y **realización** puede ser considerada como similar a la relación entre **población** y **muestra**, en la inferencia clásica. En el texto, dado que no existe posibilidad de confusión, se utilizará $\{Y_t\}$ para denotar tanto al proceso como a su realización[2].

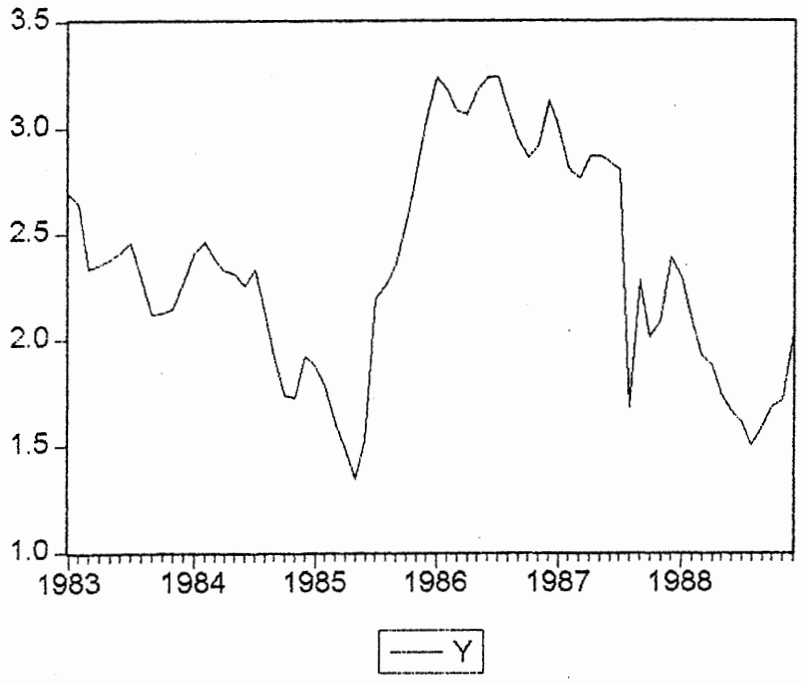
En el **GRAFICO 2.1** figura , a modo de ejemplo, la trayectoria del proceso $\{Y_t\}$, que representa la evolución de la demanda monetaria real expresada en períodos mensuales en el intervalo [ENE-1983;DIC-1988].

---

[1]. Debe tenerse en cuenta que las observaciones sobre las realizaciones de un proceso estocástico siempre son discretas en el tiempo. Es decir que, como se verá en la Sec. 4.4, los procesos de series cronológicas pueden ser consideradas como necesarias discretizaciones de ecuaciones diferenciales estocásticas. "*La realidad es continua y constante, pero para que su percepción sea posible debemos dividirla en momentos separados, imaginarla como una serie infinita de momentos separados de los que para nosotros existe sólo uno*" (Ouspensky, P.D. (1982)).

[2]. El proceso $\{Y_t\}$ puede ser interpretado como la hipotética población formada por sus infinitas trayectorias posibles.

GRAFICO 2.1: Trayectoria del proceso $\{Y_t\}$ en el período [ENE-1983;DIC-1988]



2.2.- La función de autocovarianzas

De acuerdo a lo establecido en la Sec.1.4.1 $\{Y_t; t \in T\}$ $(T = 1,2,...)$ es un proceso estacionario en las covarianzas, si sus momentos de primero y segundo orden son independientes del tiempo:

$$E(Y_t) = m \qquad (t = 1,2,...)$$

$$\Gamma(Y) = \left\{\text{cov}(Y_t, Y_{t+j})\right\} = \left\{E\left[(Y_t - m)(Y_{t+j} - m)\right]\right\} = \left\{\gamma_j(Y)\right\} \qquad (t = 1,2,...; j = 0,\pm 1,\pm 2,...)$$

Donde las autocovarianzas poseen las siguientes propiedades: **i)** de acuerdo con la **desigualdad de Cauchy-Schwartz**[3] son tales que $\left|\gamma(Y_t, Y_{t+j})\right| = \left|\gamma_j(Y)\right| \leq \sigma(Y_t)\sigma(Y_{t+j}) = \gamma_0(Y)$ $(j = 0,\pm 1,\pm 2,...)$ y **ii)** definen una función par, es decir, tal que $\gamma_j(Y) = \gamma_{-j}(Y)$ $(j = 0,1,2,...)$.

Sea, en particular, un proceso $\{Y_t\}$ estacionario de segundo orden, si su filtración natural está formado por una sucesión de n variables aleatorias observables $\{Y_{t-(n-1)}, Y_{t-(n-2)}, ..., Y_t\}$, la función de autocovarianzas puede ser ordenada de acuerdo con una matriz de Toeplitz de la forma[4]:

---

[3]. Ver Landro, A.H. (2002).

[4]. Ver Grenander, U.; Szego, G. (1958).

$$\Gamma_n(Y) = \begin{bmatrix} \gamma_0(Y) & \gamma_1(Y) & \cdots & \gamma_{n-1}(Y) \\ \gamma_1(Y) & \gamma_0(Y) & \cdots & \gamma_{n-2}(Y) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1}(Y) & \gamma_{n-2}(Y) & \cdots & \gamma_0(Y) \end{bmatrix}$$

Sea, por otra parte, la siguiente combinación lineal de estas n variables aleatorias:

$$Z_t = a_0 Y_t + a_1 Y_{t-1} + \ldots + a_{n-1} Y_{t-(n-1)}$$

Dado que la función de autocovarianzas es par, se verificará que $\sigma^2(Z_t) = \sum_{i=0}^{n}\sum_{j=0}^{n} a_i a_j \gamma_{|i-j|}(Y)$. Obviamente, salvo que los coeficientes a_i $(i = 0,1,2,\ldots,n-1)$ sean todos nulos, esta varianza será positiva, por lo que se puede concluir que la matriz $\Gamma_n(Y)$ es definida positiva y que la función de autocovarianzas $\gamma_j(Y)$ $(j = 0,1,2,\ldots,n)$ también es definida positiva para todo n entero y para cualquier conjunto de coeficientes $\{a_0, a_1, \ldots, a_{n-1}\}$.

Como se mencionó en la Sec. 1.5, las autocovarianzas de un proceso lineal estacionario y ergódico de segundo orden son tales que $\sum_{j=-\infty}^{\infty} |\gamma_j(Y)| < \infty$ y pueden ser obtenidas utilizando la **función generatriz de autocovarianzas**, $\gamma_Y(w) = \sum_{j=-\infty}^{\infty} \gamma_j(Y) w^j$, donde w es un escalar complejo.

Sean dos procesos estocásticos lineales unidimensionales, $\{Y_{1t}\}$ e $\{Y_{2t}\}$, estacionarios en las covarianzas e incorrelacionados entre sí y sea el proceso bidimensional $\{Y_t\} = \{Y_{1t}\} + \{Y_{2t}\}$. Dado que $\gamma_j(Y) = \gamma_j(Y_1 + Y_2) = \gamma_j(Y_1) + \gamma_j(Y_2)$, de acuerdo con la definición de la función generatriz de autocovarianzas, se puede concluir en forma inmediata que $\sum_{j=-\infty}^{\infty} \gamma_j(Y) w^j = \sum_{j=-\infty}^{\infty} \gamma_j(Y_1) w^j + \sum_{j=-\infty}^{\infty} \gamma_j(Y_2) w^j$. Es decir, que $\gamma_Y(w) = \gamma_{Y_1}(w) + \gamma_{Y_2}(w)$.

2.3.- La estimación del valor medio

Sea $\{Y_t\}$ $(t = 1,2,\ldots,n)$ la serie cronológica formada por una realización del proceso $\{Y_t\}$ estacionario en las covarianzas. De acuerdo con los teoremas de ergodicidad comentados en la Sec.1.5, si $\{Y_t\}$ es asintóticamente incorrelacionado, entonces se puede asegurar que el estimador $\overline{Y}_{(n)} = \frac{1}{n}\sum_{t=1}^{n} Y_t$ es insesgado y consistente, con varianza definida como un promedio ponderado de las primeras $n-1$ autocovarianzas. De modo que $\lim_{n\to\infty} nE\left\{\left[\overline{Y}_{(n)} - m(Y)\right]^2\right\} = \gamma_0(Y) + 2\sum_{j=1}^{\infty}\gamma_j(Y) = \sum_{j=-\infty}^{\infty}\gamma_j(Y)$. Entonces, de acuerdo con los postulados del teorema central del límite se puede concluir que, cuando n aumenta indefinidamente, $\sqrt{n}\left(\overline{Y}_{(n)} - m(Y)\right) \xrightarrow{d} N\left[0; \sum_{j=-\infty}^{\infty}\gamma_j(Y)\right]$. En particular, si el proceso $\{Y_t\}$ fuera además Normal, entonces para

todo n finito, se verificará que[5]:

$$\sqrt{n}\left(\overline{Y}_{(n)} - m(Y)\right) : N\left[0; \gamma_0(Y) + \sum_{j=1}^{n-1} \frac{2(n-j)}{n} \gamma_j(Y)\right]$$

2.4.- La estimación de la función de autocovarianzas

Sea $\{Y_t\}$ un proceso estocástico estacionario en las covarianzas, con momento de primer orden $E(Y_t) = m(Y)\,(t = 1,2,...)$ conocido. Un estimador de las autocovarianzas es:

$$\hat{\gamma}_j^{**}(Y) = \frac{1}{n}\sum_{t=1}^{n-j}\left[Y_t - m(Y)\right]\left[Y_{t+j} - m(Y)\right] \qquad (j = 0,1,2,...,n-1)$$

Teniendo en cuenta que $E\left(\hat{\gamma}_j^{**}(Y)\right) = \gamma_j(Y) - \frac{j}{n}\gamma_j(Y)$, se puede concluir que este estimador es asintóticamente insesgado y que, para n finito, posee un sesgo de magnitud directamente proporcional al valor de j de orden $-\frac{j}{n}\gamma_j(Y)$.

Suponiendo conocido el valor medio del proceso, un estimador alternativo de las autocovarianzas que satisface la condición de ser insesgado, es:

$$\hat{\gamma}_j^{*}(Y) = \frac{1}{n-j}\sum_{t=1}^{n-j}\left[Y_t - m(Y)\right]\left[Y_{t+j} - m(Y)\right] \qquad (j = 0,1,2,...,n-1)$$

cuyos momentos de segundo orden son de la forma:

$$\gamma\left[\hat{\gamma}_i^{*}(Y), \hat{\gamma}_j^{*}(Y)\right] = E\left[\hat{\gamma}_i^{*}(Y)\hat{\gamma}_j^{*}(Y)\right] = \frac{1}{(n-i)(n-j)}\sum_{h=1}^{n-i}\sum_{\ell=1}^{n-j}\left[\gamma_i(Y)\gamma_j(Y) + \gamma_{h-i}(Y)\gamma_{h-i-\ell-j}(Y) + \right.$$

$$\left. + \gamma_{h-\ell-j}(Y)\gamma_{h-i-\ell}(Y) + \kappa_4(i, \ell-h, \ell-h+j)\right] \qquad (0 \le i < n; 0 \le j < n)$$

donde $\kappa_4(.,.,.)$ denota el cumulante de cuarto orden:

$$\kappa_4(h,i,j) = E\left[\left(Y_t - m(Y)\right)\left(Y_{t+h} - m(Y)\right)\left(Y_{t+i} - m(Y)\right)\left(Y_{t+j} - m(Y)\right)\right] -$$

$$- \left[\gamma_h(Y)\gamma_{i-j}(Y) + \gamma_i(Y)\gamma_{h-j}(Y) + \gamma_j(Y)\gamma_{h-i}(Y)\right] \qquad (t = 0,1,2,...)$$

En particular, su varianza puede ser expresada como:

---

[5]. Esta aproximación se basa en el teorema de Cesàro según el cual, si $\sum_{j=1}^{\infty} a_j < \infty$, entonces se verifica que $\lim_{n\to\infty}\left(1 - \frac{j}{n}\right)a_j = \sum_{j=1}^{\infty} a_j$.

$$\sigma^2\left(\hat{\gamma}_j^*(Y)\right) = \frac{1}{(n-j)^2}\left\{\sum_{h=-(n-j-1)}^{n-j-1}(n-j-|h|)\left[\gamma_h^2(Y)+\gamma_{h-j}(Y)\gamma_{h+j}(Y)+\kappa_4(j,-h,j-h)\right]\right\}$$

Por lo tanto, dado que $\sum_{j=-\infty}^{\infty}\gamma_j^2(Y)<\infty$ y que $\left|\sum_{h=-\infty}^{\infty}\kappa_4(j,-h,j-h)\right|<\infty$, se verificará que:

$$\lim_{j\to\infty}\lim_{n\to\infty}(n-j)\sigma^2\left(\hat{\gamma}_j^*(Y)\right) = \sum_{h=-\infty}^{\infty}\gamma_h^2(Y)+\lim_{j\to\infty}\left|\sum_{h=-\infty}^{\infty}\kappa_4(j,-h,j-h)\right|$$

Sea, ahora, un proceso $\{Y_t\}$ estacionario de segundo orden con momento de primer orden $m(Y)$ desconocido, un estimador de las autocovarianzas es de la forma:

$$\gamma_j''(Y) = \frac{1}{n-j}\sum_{t=1}^{n-j}\left(Y_t-\overline{Y}\right)\left(Y_{t+j}-\overline{Y}\right) \qquad (j=0,1,2,\ldots,n-1)$$

su valor esperado está definido por:

$$E\left(\gamma_j(Y)\right) = \gamma_j(Y)-\frac{1}{n(n-j)}\sum_{h=1}^{n-j}\sum_{i=1}^{n}\left[\gamma_{h-i}(Y)+\gamma_{h+j-i}(Y)\right]+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{h=1}^{n}\gamma_{h-i}(Y) \qquad (j\geq 0)$$

A partir de esta expresión se obtiene que:

$$n\left|E\left(\hat{\gamma}_j''(Y)-\gamma_j(Y)\right)\right| = \frac{2}{n-j}\left|\sum_{h=n-j+1}^{n-1}(j-h)\gamma_j(Y)-j\sum_{h=0}^{n-1}\gamma_h(Y)-\right.$$

$$\left.-2\sum_{h=n-j+1}^{n-1}(n-h)\gamma_h(Y)-2\sum_{h=n-j+1}^{n-1}(n-h)\gamma_h(Y)+2\sum_{h=1}^{n}\sum_{i=1}^{n}\gamma_{h-i}(Y)\right|$$

De modo que $\lim_{n\to\infty} n\left[E\left(\hat{\gamma}_j(Y)\right)-\gamma_j(Y)\right] = \sum_{h=-\infty}^{\infty}\gamma_h(Y)$. De lo que se puede concluir que este estimador está afectado por un sesgo de orden $\frac{1}{n}$.

Un estimador alternativo (más general) de la función de autocovarianzas, para el caso en que el valor medio sea desconocido, es:

$$\hat{\gamma}_j'(Y) = \frac{1}{n-j}\sum_{t=1}^{n-j}\left(Y_t-\overline{Y}_t\right)\left(Y_{t+j}-\overline{Y}_{t+j}\right)$$

Donde $\overline{Y}_t = \frac{1}{n-j}\sum_{t=1}^{n-j}Y_t$ y $\overline{Y}_{t+j} = \frac{1}{n-j}\sum_{t=1}^{n-j}Y_{t+j}$. El valor esperado de este estimador es de la forma $E\left(\hat{\gamma}_j'(Y)\right) = \gamma_j(Y)-\frac{1}{(n-j)^2}\sum_{h=1}^{n-j}\sum_{i=j+1}^{n}\gamma_{h-i}(Y)$, de modo que $\lim_{n\to\infty} n\left[E\left(\hat{\gamma}_j'(Y)\right)-\gamma_j(Y)\right] = \sum_{j=-\infty}^{\infty}\gamma_j(Y)$. Es decir que, igual que en el caso precedente, este estimador está afectado por un sesgo de orden $\frac{1}{n}$. Sus momentos de

segundo orden son de la forma:

$$\gamma\left[\hat{\gamma}_i'(Y),\hat{\gamma}_j'(Y)\right] =$$

$$= \frac{1}{(n-i)(n-j)}\sum_{h=1}^{n-i}\sum_{\ell=1}^{n-j}\left[\gamma_{h-\ell}(Y)\gamma_{h+i-\ell-j}(Y)+\gamma_{h-\ell-j}(Y)\gamma_{h+i-\ell}(Y)+\kappa_4(i,\ell-h,\ell-h+j)\right]-$$

$$-\frac{1}{(n-i)(n-j)^2}\sum_{h=1}^{n-i}\sum_{\ell=1}^{n-j}\sum_{s=1}^{n-j}\left[\gamma_{h-\ell}(Y)\gamma_{h+i-s-j}(Y)+\gamma_{h-s-j}(Y)\gamma_{h+i-\ell}(Y)+\kappa_4(i,\ell-h,s-h+j)\right]-$$

$$-\frac{1}{(n-i)^2(n-j)}\sum_{h=1}^{n-i}\sum_{\ell=1}^{n-i}\sum_{s=1}^{n-j}\left[\gamma_{h-s}(Y)\gamma_{\ell+i-s-j}(Y)+\gamma_{h-s-j}(Y)\gamma_{\ell+i-s}(Y)+\kappa_4(\ell+i-h,s-h,s+j-h)\right]+$$

$$+\frac{1}{(n-i)^2(n-j)^2}\sum_{h=1}^{n-i}\sum_{\ell=1}^{n-i}\sum_{r=1}^{n-j}\sum_{s=1}^{n-j}\left[\gamma_{h-r}(Y)\gamma_{\ell+i-s-j}(Y)+\gamma_{h-s-j}(Y)\gamma_{\ell+i-r}(Y)+\kappa_4(\ell+i-r,s-r,s+j-r)\right]$$

Otro estimador asintóticamente insesgado de las autocovarianzas para el caso de valor medio desconocido es:

$$\hat{\gamma}_j(Y) = \frac{1}{n}\sum_{t=1}^{n-j}\left(Y_t-\overline{Y}\right)\left(Y_{t+j}-\overline{Y}\right) = \frac{n-j}{n}\hat{\gamma}_j''(Y) \qquad (j = 0,1,2,\ldots,n-1)$$

Este es el estimador considerado al tratar las condiciones de ergodicidad de segundo orden (ver Sec. 1.5) y el que será utilizado en los sucesivo en las aplicaciones que figuran en el texto. La utilización del denominador n (en vez de $n-j$) garantiza que la matriz de varianzas y covarianzas del estimador sea definida-positiva

---

Sea un proceso $\{Y_t\}$ estacionario de segundo orden y ergódico, que admite una representación de la forma $Y_t = m(Y)+\sum_{j=-\infty}^{\infty}\gamma_j(Y)\varepsilon_{t-j}$, donde el proceso $\{\varepsilon_t\}$:$WN-N\left(0,\sigma_\varepsilon^2\right)$, con momentos de cuarto orden finitos, $E\left(\varepsilon_t^4\right) = 3\sigma_\varepsilon^4+\kappa_4 < \infty$. Se demuestra, entonces, que, si n aumenta indefinidamente, los estimadores:

$$\left\{\sqrt{n}\left(\hat{\gamma}_j^{**}(Y)-\gamma_j(Y)\right), j = 0,1,2,\ldots,n\right\}$$

$$\left\{\sqrt{n}\left(\hat{\gamma}_j^{*}(Y)-\gamma_j(Y)\right), j = 0,1,2,\ldots,n\right\}$$

$$\left\{\sqrt{n}\left(\hat{\gamma}_j''(Y)-\gamma_j(Y)\right), j = 0,1,2,\ldots,n\right\}$$

$$\left\{\sqrt{n}\left(\hat{\gamma}_j'(Y)-\gamma_j(Y)\right), j = 0,1,2,\ldots,n\right\}$$

$$\left\{\sqrt{n}\left(\hat{\gamma}_j(Y)-\gamma_j(Y)\right), j = 0,1,2,\ldots,n\right\}$$

convergen en-distribución a una función Normal con valor medio nulo y momentos de segundo orden de

la forma[6] $\lim_{n\to\infty} n\gamma\left(\hat{\gamma}_i(Y),\hat{\gamma}_j(Y)\right) = \sum_{h=-\infty}^{\infty}\left[\gamma_{h+i}(Y)\gamma_{h+j}(Y) + \gamma_{h-j}(Y)\gamma_{h+i}(Y)\right] + \frac{\kappa_4}{\sigma_\varepsilon^2}\gamma_i(Y)\gamma_j(Y)$.

2.5.- La función de densidad espectral

Sea $\{Y_t\}$ un proceso estacionario de segundo orden. Su función de autocovarianzas (que, como ya se demostró, es acotada) de acuerdo con los postulados de los teoremas Wiener-Khinchine y de Herglotz admite una **representación espectral** de la forma

$$\gamma_j(Y) = \int_{-n}^{n}\omega_Y(\alpha)e^{i\alpha j}\,d\alpha = \int_{-\pi}^{\pi}e^{i\alpha j}\,dW_Y(\alpha) \qquad \left(j = 0,\pm1,\pm2,\ldots;-\pi < \alpha < \pi\right)$$

que permite la descomposición de $\{Y_t\}$ en sus componentes de distintas fecuencias[7] y donde $\omega_Y(\alpha)$ denota la llamada **función de densidad espectral** (o, simplemente, el **espectro**) de $\{Y_t\}$ y:

$$W_Y(\alpha) = \int_{-\infty}^{\alpha}\omega_Y(u)du$$

que es monótona no-decreciente, acotada[8] y tal que $W_Y(-\pi) = 0$ y $W_Y(\pi) = \sigma^2(Y)$, denota la **función de distribución espectral** (o **espectro integrado**) asociada al proceso $\{Y_t\}$.

En particular, se puede escribir $\gamma_0(Y) = \sigma_Y^2 = \int_{-\pi}^{\pi}\omega_Y(\alpha)d\alpha$, de modo que $w_Y(\alpha)d\alpha$ puede ser

---

[6]. Ver Anderson, T.W. (1964).

[7]. El teorema de Herglotz, que demuestra que las clases de las funciones definidas no-negativas y de las funciones características son coincidentes (ver Brockwell, P.J.; Davis, R.A. (1987)), permite concluir que la existencia de esta forma de representación espectral es condición necesaria y suficiente para que una función $\gamma_j(Y)$ sea definida no-negativa.

[8]. Como se verá en la Sec. 4.2.3, si se verifica que la función de densidad espectral es no-acotada en las frecuencias bajas, $\lim_{\alpha\to0}\omega_Y(\alpha) = \infty$, entonces se dice que $\{Y_t\}$ es un **proceso de memoria larga** (de acuerdo con la nomenclatura de Robinson, P.M. (1994) y Baillie, R.I. (1996)). Sea $\{Y_t\}$ un proceso formado por variables centradas y sea $\{Y_t^*\}$ el proceso que se obtiene de su acumulación progresiva $Y_t^* = \sum_{\tau=1}^{t} Y_\tau$, si se verifica: **i)** que la varianza $\sigma_{Y^*}^2 = \lim_{n\to\infty}\frac{1}{n}E\left(Y_n^{*2}\right)$ existe y es no-nula y **ii)** que, de acuerdo con el **principio de invariancia**, es posible hallar por lo menos un valor de $\alpha \in [0;1]$ tal que se cumpla la siguiente convergencia débil $\frac{1}{\sqrt{n}}Y_{[\alpha n]}^* \Rightarrow \sigma_{Y^*}B(\alpha)$ (donde $[\alpha n]$ denota la parte entera de αn y $B(\alpha)$ denota un proceso movimiento Browniano), entonces se dice que el proceso $\{Y_t\}$ es de **memoria corta** (estas condiciones se conocen como condiciones de Lee, D.; Schmidt, P. (1996), ver Sec. 4.1). Obsérvese que la condición de que un proceso sea estacionario de segundo orden no es necesaria para que sea de memoria corta. Muchos autores consideran como condiciones necesarias para la memoria corta: **i)** la existencia de los momentos absolutos de orden $s > 2$; **ii)** el cumplimiento de condiciones α-mixing con coeficientes α_j tales que $\sum_{j=1}^{\infty}\alpha_j^{1-2/s} < \infty$ (conocidas como condiciones de Phillips, P.C.B. (1987)).

considerada como la contribución a la variación total de $\{Y_t\}$ de las frecuencias comprendidas en el intervalo $(\alpha, \alpha + d\alpha)$ (ver Jenkins, G.M.; Watts, D.G. (1969), Dhrymes, P.J. (1971)(1974), Chatfield, C. (1975), Bloomfield, P. (1976), Granger, C.W.J.; Newbold, P. (1977), Priestley, B. (1981)).

Teniendo en cuenta que $e^{i\alpha j} = \cos(\alpha j) + i\,sen(\alpha j)$, se puede escribir :

$$\gamma_j(Y) = \begin{cases} \int_{-\pi}^{\pi} \omega_Y(\alpha)\left[\cos(\alpha j) + i\,sen(\alpha j)\right] d\alpha \\ \int_{-\pi}^{\pi} \omega_Y(\alpha)\left[\cos(\alpha j) - i\,sen(\alpha j)\right] d\alpha \end{cases}$$

y, dado que $\gamma_j(Y)$ es una función par, se obtiene (debe tenerse en cuenta que, si $\{Y_t\}$ es no-estacionario, su función de densidad espectral no puede ser expresada como una suma de senos y cosenos): **i)** que $\gamma_j(Y) = \int_{-\pi}^{\pi} \omega_Y(\alpha)\cos(\alpha j)\,d\alpha$ $(j = 0,\pm 1,\pm 2,...)$ y **ii)** que la función $\omega_Y(\alpha)$ también es par (por analogía con la función de autocorrelaciones, es posible definir una función de la forma $\omega_Y^*(\alpha) = \frac{\omega_Y(\alpha)}{\gamma_0(Y)} = \frac{1}{2\pi}\left[1 + 2\sum_{j=1}^{\infty}\rho_j(Y)\cos(\alpha j)\right]$ la cual, obviamente, posee muchas de las propiedades de la función de densidad espectral):

$$\omega_Y(\alpha) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j(Y)e^{-i\alpha j} = \frac{1}{2\pi}\left[\gamma_0(Y) + \sum_{j=1}^{\infty}\gamma_j(Y)\left(e^{-i\alpha j} + e^{i\alpha j}\right)\right] =$$

$$= \frac{1}{2\pi}\left[\gamma_0(Y) + 2\sum_{j=1}^{\infty}\gamma_j(Y)\cos(\alpha j)\right]$$

2.6.- La estimación de la función de densidad espectral

Si en la definición dada en la sección precedente se sustituyen las autocovarianzas por sus estimadores, por ejemplo, por:

$$\hat{\gamma}_j(Y) = \frac{1}{n}\sum_{t=1}^{n-j}\left(Y_t - \bar{Y}\right)\left(Y_{t+j} - \bar{Y}\right) \qquad (j = 1,2,...,n-1)$$

se obtiene el siguiente estimador autorregresivo de la función de densidad espectral:

$$\hat{\omega}_Y(\alpha) = \frac{1}{2\pi}\left[\gamma_0(Y) + 2\sum_{j=1}^{n-1}\hat{\gamma}_j(Y)\cos(\alpha j)\right]$$

que no es consistente[9]. A fin de resolver este inconveniente se han propuesto modificaciones no-

[9]. Ver Akaike, H. (1969), Parzen, E. (1974).

paramétricas de este estimador, cuya expresión general es de la forma:

$$\hat{\omega}_Y(\alpha) = \frac{1}{2\pi}\left[\lambda_0\hat{\gamma}_0(Y) + 2\sum_{j=1}^{N_n}\lambda_j\gamma_j(Y)\cos(\alpha j)\right]$$

Donde el conjunto $\lambda_0, \lambda_1, ..., \lambda_{N_n}$ define la llamada **ventana del estimador** y $N_n < n$ denota el **punto de truncamiento**[10].

2.7.- La función de autocorrelaciones

Asociada a la función de autocovarianzas existe la **función de autocorrelaciones**, definida de la siguiente forma: $\rho_j(Y) = \frac{\gamma_j(Y)}{\gamma_0(Y)}\ (j = 1,2....)$. Obviamente, la función de autocorrelaciones (que hereda muchas de las propiedades de la función de autocovarianzas) es una función par, $\rho_j(Y) = \rho_{-j}(Y)\ \ (j = 1,2,...)$. Además, a partir de su definición, se puede concluir en forma inmediata que $|\rho_j(Y)| \le 1\ \ (j = 1,2,...)$ y que $\rho_0(Y) = 1$ [11].

De la misma forma que en el caso de las autocovarianzas, para un conjunto dado de n variables aleatorias observables, $\left\{Y_{t-(n-1)}, Y_{t-(n-2)}, ..., Y_t\right\}$, los coeficientes de autocorrelación pueden ser ordenados en una matriz de la forma:

$$\mathrm{P}_n(Y) = \frac{1}{\gamma_0(Y)}\Gamma_n(Y) = \frac{1}{\gamma_0(Y)}\begin{bmatrix} \gamma_0(Y) & \gamma_1(Y) & \cdots & \gamma_{n-1}(Y) \\ \gamma_1(Y) & \gamma_0(Y) & \cdots & \gamma_{n-2}(Y) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1}(Y) & \gamma_{n-2}(Y) & \cdots & \gamma_0(Y) \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & \rho_1(Y) & \rho_2(Y) & \cdots & \rho_{n-1}(Y) \\ \rho_1(Y) & 1 & \rho_1(Y) & \cdots & \rho_{n-2}(Y) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n-1}(Y) & \rho_{n-2}(Y) & \rho_{n-3}(Y) & \cdots & 1 \end{bmatrix}$$

conocida como **matriz de Laurent** la cual, igual que la matriz de autocovarianzas, es definida positiva.

Una interesante propiedad de la función de autocorrelaciones, que constituye el fundamento de

---

[10]. Por ejemplo, para $\lambda_j = \begin{cases} 1 - 6\left(\frac{j}{N_n}\right)^2 + 6\left(\frac{j}{N_n}\right)^3 & \left(j = 0,1,...,\frac{N_n}{2}\right) \\ 2\left(1 - \frac{j}{N_n}\right)^3 & \left(j = \frac{N_n}{2}, \frac{N_n}{2}+1, ..., N_n\right) \end{cases}$, se obtiene la conocida como ventana de Parzen la cual, si se verifica que $\lim_{n\to\infty} N_n = \infty$ y $\lim_{n\to\infty}\frac{N_n}{n} = 0$, permite obtener un estimador consistente de $\omega_Y(\alpha)$ (ver Parzen, E. (1967)). Otras ventanas pueden ser halladas en Hannan, E.J. (1970), Neave, H.R. (1972), Priestley, M.B. (1981).

[11]. Si se verifica que las variables $Y_t : N\left(m(Y), \sigma_Y^2\right)$, la única forma de dependencia posible es la de primer orden.

un gran número de tests de Normalidad, indica que, dado un proceso $\{Y_t\}$ Normal, estacionario, la función generatriz de momentos correspondiente a un par de sus variables, (Y_{t-j}, Y_t), cualesquiera será de la forma $M_{Y_{t-j},Y_t}(w_1,w_2) = \exp\left[\frac{\sigma_Y^2}{2}\left(w_1^2 + w_2^2 + 2w_1w_2\rho_j(Y)\right)\right]$. A partir de esta función se obtienen los siguientes momentos absolutos:

$$m_2\left(Y_t^2\right) = E\left(Y_t^2\right) = \frac{\partial^2}{\partial w_1^2\,\partial w_2^2} M_{Y_{t-j},Y_t}(w_1,w_2)\Big|_{w_1=w_2=0} = \sigma_Y^2$$

$$m_4\left(Y_t\right) = E\left(Y_t^4\right) = \frac{\partial^4}{\partial w_2^4} M_{Y_{t-j},Y_t}(w_1,w_2)\Big|_{w_1=w_2=0} = 3\sigma_Y^4$$

$$m_{2,2}\left(Y_{t-j},Y_t\right) = E\left(Y_{t-j}^2,Y_t^2\right) = \frac{\partial^8}{\partial w_1^2\,\partial w_2^2} M_{Y_{t-j},Y_t}(w_1,w_2)\Big|_{w_1=w_2=0} = \left[1+2\rho_j^2(Y)\right]\sigma_Y^2$$

y, en consecuencia, las siguientes medidas:

$$\sigma^2\left(Y^2\right) = \gamma_0\left(Y^2\right) = 2\sigma_y^2$$

$$\gamma\left(Y_{t-j}^2,Y_t^2\right) = \gamma_j\left(Y^2\right) = m_{2,2}\left(Y_{t-j},Y_t\right) - m_2\left(Y_{t-j}\right)m_2\left(Y_t\right) = 2\rho_j^2(Y)\sigma_Y^4$$

$$\rho\left(Y_{t-j}^2,Y_t^2\right) = \rho_j\left(Y^2\right) = \rho_j^2(Y)$$

2.8.- La estimación de la función de autocorrelaciones

Sea $\{Y_t\}$ un proceso estacionario en las covarianzas. De acuerdo con las definiciones dadas para los estimadores de la función de autocovarianzas se puede concluir que: **i)** si $m(Y)$ es conocido, el estimador $\hat{\rho}_j^{**}(Y) = \dfrac{\sum_{t=1}^{n-j}\left(Y_t - m(Y)\right)\left(Y_{t+j} - m(Y)\right)}{\sum_{t=1}^{n}\left(Y_t - m(Y)\right)^2}$, será tal que $\lim_{n\to\infty}\sqrt{n}\left[\hat{\rho}_j^{**}(Y) - \rho_j(Y)\right] = 2\sum_{i=-\infty}^{\infty}\left[\rho_i^2(Y)\rho_{i+j}(Y) - \right.$

$\left. - \rho_i(Y)\rho_{i-j}(Y)\right]$ y, por lo tanto, que está afectado por un sesgo de magnitud $O\left(n^{-1}\right)$ y **ii)** si, como ocurre habitualmente en las aplicaciones, $m(Y)$ es desconocido, $\hat{\rho}_j(Y) = \dfrac{\sum_{t=1}^{n-j}\left(Y_t - \bar{Y}\right)\left(Y_{t+j} - \bar{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2}$ es un estimador consistente de la función $\rho_j(Y)$, con matriz de varianzas y covarianzas $\Gamma\left(\hat{\rho}(Y)\right) = \left[\gamma\left(\hat{\rho}_i(Y),\hat{\rho}_j(Y)\right)\right]$ $\left(i,j = 1,2,\ldots,n\right)$, donde:

$$\gamma\left(\hat{\rho}_i(Y),\hat{\rho}_j(Y)\right) = \sum_{h=-\infty}^{\infty}\left(\rho_{h+i}(Y)\rho_{h+j}(Y) + \rho_{h+i}(Y)\rho_{h+j}(Y) - \right.$$

$$\left. - 2\rho_j(Y)\rho_h(Y)\rho_{h+i}(Y) - 2\rho_i(Y)\rho_h(Y)\rho_{h+j}(Y) + 2\rho_i(Y)\rho_j(Y)\rho_h^2(Y)\right) \qquad \left(i,j = 1,2,\ldots,n; i \neq j\right)$$

$$\hat{\sigma}\left(\hat{\rho}_i(Y)\right) = \sum_{h=-\infty}^{\infty}\left[\rho_h^2(Y) + \rho_{h-i}(Y)\rho_{h+i}(Y) - 4\rho_i(Y)\rho_h(Y)\rho_{h+i}(Y) + 2\rho_i^2(Y)\rho_h^2(Y)\right]$$

(debe tenerse en cuenta que estas igualdades se verifican también para el caso en que $m(Y)$ sea conocido)[12]

De todo lo anterior se puede concluir que, si el proceso $\{Y_t\}$ es estacionario de segundo orden: **i)** la matriz de los estimadores de las autocorrelaciones:

$$\hat{\mathrm{P}}_n(Y) = \frac{1}{\hat{\gamma}_0(Y)}\hat{\Gamma}_n(Y) = \frac{1}{\hat{\gamma}_0(Y)}\begin{bmatrix} \hat{\gamma}_0(Y) & \hat{\gamma}_1(Y) & \dots & \hat{\gamma}_{n-1}(Y) \\ \hat{\gamma}_1(Y) & \hat{\gamma}_0(Y) & \dots & \hat{\gamma}_{n-2}(Y) \\ \dots & \dots & \dots & \dots \\ \hat{\gamma}_{n-1}(Y) & \hat{\gamma}_{n-2}(Y) & \dots & \hat{\gamma}_0(Y) \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & \hat{\rho}_1(Y) & \hat{\rho}_2(Y) & \dots & \hat{\rho}_{n-1}(Y) \\ \hat{\rho}_1(Y) & 1 & \hat{\rho}_1(Y) & \dots & \hat{\rho}_{n-2}(Y) \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{n-1}(Y) & \hat{\rho}_{n-2}(Y) & \hat{\rho}_{n-3}(Y) & \dots & 1 \end{bmatrix}$$

es definida positiva y **ii)** la varianza de $\hat{\rho}_j(Y)$ depende de todas las autocorrelaciones del proceso, lo cual, obviamente, impide su cálculo a partir de una serie cronológica finita.

En particular, si el proceso es tal que su función de autocorrelaciones se anula para todo $j > q$, las definiciones anteriores asumen la forma [13]:

$$\sigma^2\left(\hat{\rho}_j(Y)\right) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=-(j-1)}^{j-1}\rho_i^2(Y) = \frac{1}{n}\left[1 + 2\sum_{i=1}^{j-1}\rho_i^2(Y)\right] & (j \le q) \\ \dfrac{1}{n}\displaystyle\sum_{i=-q}^{q}\rho_i^2(Y) = \frac{1}{n}\left[1 + 2\sum_{i=1}^{q}\rho_i^2(Y)\right] & (j > q) \end{cases}$$

$$\gamma\left(\hat{\rho}_j(Y), \hat{\rho}_{j+k}(Y)\right) = \frac{1}{n}\sum_{i=-q}^{q}\rho_i(Y)\rho_{i+k}(Y) \qquad (k > q)$$

si $\{Y_t\}$:$WN - N\left(0, \sigma_\varepsilon^2\right)$ (condición suficiente para que $q = 0$), se verifica que $E\left[\hat{\rho}_j(Y)\right] = -\dfrac{n-j}{n(n-1)}$ y la varianza anterior puede ser aproximada de la siguiente forma: $\sigma^2\left(\hat{\rho}_j(Y)\right) \approx \dfrac{1}{n}$.

Sustituyendo los $\rho_j(Y)$ por sus estimadores muestrales se obtiene el siguiente estimador de la

---

12. Ver Mann, H.B.; Wald, A. (1943), Anderson, T.W.; Walker, A.M. (1964), Anderson, T.W. (1971).

13. Es decir, si $\{Y_t\}$ admite una representación $MA(q)$ (que será tratada en la Sec. 2.11) de la forma $Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0, \sigma_\varepsilon^2\right)$), formado por variables con momentos de cuarto orden proporcionales a σ_ε^4 y momentos de sexto orden finitos.

varianza de los estimadores de los coeficientes de autocorrelación:

$$\hat{\sigma}^2\left(\hat{\rho}_j(Y)\right) \approx \frac{1}{n}\left[1+2\sum_{i=1}^{q}\rho_i^2(Y)\right] \qquad (j > q)$$

Bajo el supuesto de que la hipótesis nula $H_0{:}\rho_j(Y) = 0$ es verdadera, de la misma forma (y en condiciones menos estrictas) que en el caso de los estimadores de las autocovarianzas, se demuestra que $\sqrt{n}\hat{\rho}_j(Y) \xrightarrow{d} N(0;1)$ [14].

Obviamente, a partir de una serie cronológica $\{Y_1, Y_2, ..., Y_n\}$ no se pueden estimar autocovarianzas ni autocorrelaciones para órdenes mayores o iguales a n, y es de suponer que para órdenes inferiores, pero próximos a n, las estimaciones resultarán muy poco confiables, debido a lo reducido del conjunto de observaciones utilizado (habrá muy pocos pares de valores (Y_t, Y_{t+j}) disponibles para realizar la estimación). Algunos autores sugieren que las estimaciones "confiables" serán las de orden menor o igual que $\frac{n}{4}$, para $n \geq 50$ [15], otros indican que el test resultará más confiable si $j < \ln(n)$ [16].

A fin de decidir cuándo, en un proceso $\{Y_t\}$ con función de autocorrelaciones que se anula para todo $j > q$, un coeficiente $\rho_j(Y)$ puede ser considerado nulo y cuándo significativamente distinto de cero, a partir de la información que proporciona la serie cronológica $\{Y_t\}$ $(t = 1,2,...,n)$, se podría construir un test en el que $H_0{:}\rho_j(Y) = 0$ (para $j > q$). Bajo el supuesto que la hipótesis nula es verdadera, utilizando la distribución asintótica Normal conjunta y las aproximaciones mencionadas más arriba y considerando un nivel de confiabilidad del 95%, los límites de la zona crítica estarán dados, aproximadamente, por los valores:

$$\pm\frac{2}{\sqrt{n}}\sqrt{\sum_{i=-(j-1)}^{j-1}\hat{\rho}_i^2(Y)} = \pm\frac{2}{\sqrt{n}}\sqrt{1+2\sum_{i=1}^{j-1}\hat{\rho}_i^2(Y)} \qquad (para\ j \leq q)$$

y,

$$\pm\frac{2}{\sqrt{n}}\sqrt{1+2\sum_{i=1}^{q}\hat{\rho}_i^2(Y)} \qquad (para\ j > q)$$

(la representación gráfica de los estimadores $\hat{\rho}_j(Y)$ y de los límites de las zonas de aceptación de las hipótesis nulas de este test forman el **correlograma**[17]).

No obstante estas consideraciones, debe tenerse en cuenta que, para órdenes próximos a cero, los

---

[14]. Ver Anderson, T.W. (1971), Brockwell, P.J.; Davis, R.A. (1991).

[15]. Ver Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. (1994).

[16]. Debe tenerse en cuenta que estas condiciones son válidas si el proceso no está afectado por factores estacionales.

[17]. Denominación debida a Wold, H. (1949).

estimadores $\hat{\rho}_j(Y)$ y $\hat{\rho}_{j+h}(Y)$ son altamente correlacionados y con varianza apreciablemente menor que $\frac{1}{n}$ y que, por lo tanto, el análisis de la significatividad individual de los coeficientes $\rho_j(Y)$ no tiene sentido, dado que el test está afectado por un sesgo hacia la aceptación de la hipótesis nula. En consecuencia, la única consideración que se puede hacer sobre la función de autocorrelaciones se refiere a su comportamiento global y se basa en un test "port manteau" que considera las primeras $J\left(=1,2,...,\frac{n}{4}\right)$ autocorrelaciones en conjunto, contrastando la hipótesis nula:

$$H_0: \rho_1(Y) = \rho_2(Y) = ... = \rho_J(Y) = 0$$

con respecto a la alternativa:

$$H_1: \textit{que por lo menos un } \rho_j(Y)\,(j = 1,2,...,J) \textit{ sea significativo}$$

Bajo el supuesto de que la hipótesis nula sea verdadera, Box, G.E.P.; Pierce, D.A. 1(970) propusieron el estadístico $Q^{(BP)}(J) = n\sum_{i=1}^{J}\hat{\rho}_i^2(Y) \xrightarrow{d} \chi^2_{(J)}$. Ahora bien, las experiencias obtenidas de la utilización de métodos de simulación[18] demostraron que, bajo el supuesto de que la hipótesis nula que $\{Y_t\}: WN - N\left(0,\sigma_Y^2\right)$ es verdadera, la distribución χ^2 no constituye una buena aproximación de la distribución del estimador de $Q_j^{(BP)}$, cuyos valores tienden a ser menores que los esperados a partir de la distribución χ^2. Con el fin de superar esta falencia (y, en consecuencia, de aumentar la potencia del test), Ljung, G.M.; Box, G.E.P. (1978), basándose en la consideración de que las varianzas de los estimadores $\hat{\rho}_j(Y)$ resultan mejor aproximadas por $\frac{n-j}{n(n+2)}$ que por $\frac{1}{n}$, propusieron una modificación del estadístico de Box-Pierce de la forma:

$$Q^{(LB)}(J) = n(n+2)\sum_{i=1}^{J}\frac{\hat{\rho}_i^2(Y)}{n-i} \xrightarrow{d} \chi^2_{(J)}$$

(debe tenerse en cuenta que esta convergencia en-distribución se verifica bajo el supuesto que la hipótesis nula que $\{Y_t\}: WN - N\left(0,\sigma_Y^2\right)$ es verdadera)[19].

Una cuestión importante que surge de la aplicación de este test radica en la selección del valor J: si se toma un valor muy bajo se corre el riesgo de dejar de considerar autocorrelaciones significativas de orden superior; si se toma un valor demasiado alto el test puede perder potencia debido a la inclusión de autocorrelaciones despreciables de orden alto[20].

---

18. Ver Davies, N.; Triggs, C.M.; Newbold, P. (1977).

19. Obsérvese que éste pertenece a la familia de los denominados tests de diagnóstico, ya que no requiere de la definición de una hipótesis alternativa específica (la hipótesis nula se contrasta con respecto a un conjunto de alternativas posibles).

20. Debe tenerse en cuenta que, en presencia de heterocedasticidad condicionada, se modifican las propiedades asintóticas de las covarianzas muestrales y, en consecuencia, los límites de la zona crítica. Por ejemplo, la hipótesis nula $H_0: \rho_1(Y) = 0$ será aceptada con un 95% de confiabilidad, si se verifica que $\left|\rho_1(Y)\right| < \frac{2}{\sqrt{n}}$. Es decir, si $\left|\gamma_j(Y)\right| < \frac{2\gamma_0(Y)}{2\sqrt{n}} = \frac{2}{\sqrt{n}}E\left(Y_t^2\right)$. Pero, en el caso de existencia de heterocedasticidad condicionada la hipótesis nula debería ser aceptada, con un 95% de confiabilidad, si se verificara

Ejemplo 2.1:

Sea un proceso $\{Y_t\}$, expresado en términos mensuales, cuya trayectoria en el período [ENE-1999;Dic-2003] figura en el **GRAFICO 2.2** y cuya función de autocorrelaciones está representada en el correlograma que figura en el **CUADRO 2.1**. Obsérvese que el valor del estadístico $Q^{(LB)}(j)$ para los 15 primeros valores de j permite concluir que no existen razones suficientes para rechazar la hipótesis nula $H_0:\{Y_t\}:WN-N(0,\sigma_Y^2)$.

2.9.- Las expresiones MA y AR de orden infinito como dos formas equivalentes de representación lineal de un proceso estocástico

2.9.1.- La representación MA

De acuerdo a los postulados del teorema de descomposición analizado en la Sec. 1.6, todo proceso estocástico débilmente estacionario, admite una representación definida por la combinación lineal de infinitos "shocks" aleatorios incorrelacionados (**representación en promedios móviles** o **representación MA**) de la forma[21]:

$$Y_t = \lim_{j\to\infty}\left(\varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2}+...+b_j\varepsilon_{t-j} + Y_{(t-j-1)}\right) =$$

$$= \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2}+... = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{\infty}\psi_j B^j\varepsilon_t = \Psi(B)\varepsilon_t \qquad (\psi_0 = 1)$$

donde: **i)** Y_t denota una variable centrada; **ii)** B denota un **operador backward** tal que $B\varepsilon_t = \varepsilon_{t-1}$, $B^2\varepsilon_t = B(B\varepsilon_t) = B\varepsilon_{t-1} = \varepsilon_{t-2}$,... y, en general, $B^j\varepsilon_t = \varepsilon_{t-j}\ (j = 0,1,2,...)$ (siendo por B^0 el elemento identidad en este dominio de los operadores)[22] y **iii)** el **operador de promedios móviles de orden infinito (operador MA)**, $\Psi(B) = 1+\psi_1 B+\psi_2 B^2+...$, representa la **función de transferencia** del sistema lineal que vincula a los procesos $\{Y_t\}$ y $\{\varepsilon_t\}$.

Se puede concluir, entonces, que la condición necesaria y suficiente para que el sistema $Y_t = \Psi(B)\varepsilon_t$ permanezca **estable**, es decir, para que el operador $\Psi(B)$ sea convergente para $|B| \le 1$ y, por lo tanto, para que el proceso $\{Y_t\}$, de acuerdo con el supuesto planteado al comienzo de esta sección, sea **estacionario de segundo orden**, es que la sucesión de ponderaciones $\{\psi_j\}$ (que define la función de

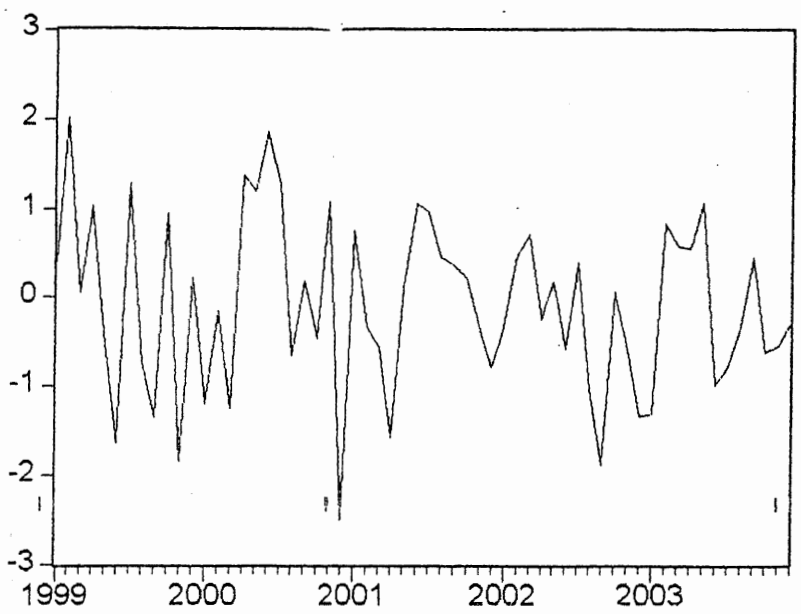
---

que $|\gamma_1(Y)| < <\frac{2}{\sqrt{n}}\sqrt{E(Y_{t-1}^2Y_t^2)}$ (debe tenerse en cuenta que la raíz $\sqrt{Y_{t-1}^2Y_t^2}$ puede ser distinta de 1).

[21]. Debe tenerse en cuenta que en esta representación del proceso $\{Y_t\}$, definida como una suma ponderada, el total de las ponderaciones no necesariamente debe ser igual a la unidad.

[22].A partir de esta definición se obtiene, en forma inmediata, que: **i)** $B^{j_1+j_2} = B^{j_1}B^{j_2} = B^{j_2}B^{j_1}\ (j_1, j_2 > 0)$ y **ii)** $\sum_{j=0}^{k-1}B^j = \frac{1-B^k}{1-B}$. Asimismo, definiendo por $R(B)$ al conjunto de todas las combinaciones lineales finitas de elementos $B^j\ (j = 0,1,2,...)$, se demuestra que el álgebra de operadores $R(B)$ es isomórfica con respecto al álgebra de funciones polinomiales $R(X)$, en el dominio de los números reales.

GRAFICO 2.2: Trayectoria del proceso $\{Y_t\}$



CUADRO 2.1: Correlograma correspondiente al proceso $\{Y_t\}$, cuya trayectoria figura en el GRAFICO 2.2

Sample: 1999:01 2003:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.013	0.013	0.0110	0.917
		2	0.129	0.129	1.0783	0.583
		3	-0.044	-0.048	1.2070	0.751
		4	-0.055	-0.072	1.4088	0.843
		5	-0.129	-0.118	2.5388	0.771
		6	-0.162	-0.150	4.3434	0.630
		7	-0.101	-0.079	5.0639	0.652
		8	-0.135	-0.119	6.3709	0.606
		9	0.026	0.017	6.4220	0.697
		10	-0.172	-0.199	8.6358	0.567
		11	-0.012	-0.096	8.6465	0.654
		12	-0.018	-0.056	8.6729	0.731
		13	0.141	0.076	10.257	0.673
		14	0.070	0.011	10.657	0.713
		15	0.104	0.014	11.546	0.713

respuesta del sistema a un impulso) sea de cuadrado sumable, $\sum_{j=0}^{\infty}\psi_j^2 < \infty$ [23].

Como se vio en la Sec.1.5, en términos de la condición de ergodicidad, resultan de particular interés aquellos procesos en los cuales la condición anterior puede ser sustituida por la más fuerte de ser **absolutamente sumable**, $\sum_{j=0}^{\infty}|\psi_j| < \infty$ [24].

Con un razonamiento inverso, a partir del teorema de Wold, se dice que un proceso $\{Y_t\}$ es **estacionario** si puede ser representado, a partir de un proceso $\{\varepsilon_t\}$:*WN - debil* , mediante la aplicación del **filtro lineal**:

$$\Psi(B) = \sum_{j=0}^{\infty}\psi_j B^j \quad (\psi_0 = 1)$$

donde cada ψ_j representa la medida del impacto que ejerce sobre Y_t , un "shock" aleatorio ocurrido en el período $t - j$ [25], y la suma de las ponderaciones ($\Psi(1) = \sum_{j=0}^{\infty}\psi_j$) define la medida de la **persistencia** de la influencia de un "shock" aleatorio sobre el nivel del proceso[26].

Teniendo en cuenta que[27]:

---

[23]. La forma de decrecimiento de las ponderaciones ψ_j permite concluir que todo proceso estacionario tiende a retornar a su valor medio y a fluctuar a su alrededor en un rango más o menos constante (propiedad conocida como **de retorno al valor medio**). Como se verá inmediatamente, la velocidad de este retorno está relacionada con el comportamiento de las autocovarianzas (si los coeficientes $\gamma_j(Y)$ son todos nulos, el proceso es WN y el retorno al valor medio es instantáneo, si asume valores significativos pero decrece rápidamente, dicho retorno es rápido y, si su decrecimiento es lento, el retorno también es lento).

[24]. Debe tenerse en cuenta que la condición de absolutamente sumable implica la de cuadrado sumable, pero que la implicación inversa no se verifica necesariamente.

[25]. Sea $\{\varepsilon_t\}$:*WN - estricto* y sea $\{Y_t\}$ un proceso estrictamente estacionario en el que las variables Y_t son centradas. Se demuestra, entonces, que $\{Y_t\}$ admite una representación en términos de un desarrollo de Volterra de la forma $Y_t = \sum_{j=0}^{\infty}\psi_j^*\varepsilon_{t-j}^* + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_{ij}^*\varepsilon_{t-i}^*\varepsilon_{t-j}^* + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_{ijk}^*\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-k} + \ldots$ (ver Nisio, M. (1960)(1961)). Ahora bien, si se Supone además que el proceso $\{Y_t\}$ es débilmente estacionario de segundo orden, de acuerdo a lo expresado en el texto, también admite una representación MA de la forma $Y_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$. Estas dos representaciones coincidirán sólo cuando los términos de orden mayor que uno en el desarrollo de Volterra se anulen, es decir, cuando $\psi_{ij}^* = \psi_{ijk}^* = \ldots = 0$. Luego, se puede concluir que si bien todo proceso débilmente estacionario de segundo orden admite una representación MA en términos de un proceso WN-débil, en general no admite una representación MA en términos de un proceso WN-estricto.

[26]. El problema de la persistencia será tratado en el Cap. 5.

[27]. La condición que $\{\varepsilon_t\}$:*WN - debil* implica que el proceso estacionario $\{Y_t\}$ es incorrelacionado con $\{\varepsilon_\tau\}$ $(\tau > t)$. Se dice, entonces, que $\{Y_t\}$ es un **proceso causal** o **autorregresivo o independiente del futuro** (ver Brockwell, P.J.; Davis, R.A. (1987)).

$$\gamma_j(Y) = E\left(Y_t Y_{t+j}\right) = E\left(\sum_{i=0}^{\infty}\sum_{h=0}^{\infty}\psi_i\psi_h\varepsilon_{t-i}\varepsilon_{t+j-h}\right) = \sigma_\varepsilon^2\sum_{i=0}^{\infty}\psi_i\psi_{i+j} \qquad (\psi_0 = 1)$$

y, por lo tanto, que $\gamma_0(Y) = \sigma_Y^2 = \sigma_\varepsilon^2\sum_{i=0}^{\infty}\psi_i^2$ $(\psi_0 = 1)$, se puede concluir que, si el proceso es estable, su varianza es finita, es decir que la condición $\sigma_Y^2 < \infty$ es necesaria y suficiente para poder asegurar que $\{Y_t\}$ es estacionario en las covarianzas.

Se demuestra, además, que si la sucesión de ponderaciones $\{\psi_j\}$ es absolutamente sumable, entonces $\sum_{i=0}^{\infty}\sum_{h=0}^{\infty}|\psi_i\psi_h| = \sum_{i=0}^{\infty}|\psi_i|\sum_{h=0}^{\infty}|\psi_h| < \infty$ y, por lo tanto, las autocovarianzas $\gamma_j(Y)$ son absolutamente sumables, $\sum_{j=0}^{\infty}|\gamma_j(Y)| < \infty$ y viceversa (de acuerdo a lo demostrado en la Sec. 1.5, si se cumple esta condición, se puede asegurar que $\{Y_t\}$ es un proceso ergódico)[28].

Como se mencionó en la Sec.2.2, en general las autocovarianzas de un proceso lineal estacionario de segundo orden pueden ser obtenidas utilizando la función generatriz de autocovarianzas, $\gamma_Y(B) = \sum_{j=-\infty}^{\infty}\gamma_j(Y)B^j$ (donde $B^{-j} = F^j$, denotando F un operador "forward" tal que $FY_t = Y_{t+1}$). Reemplazando en esta definición $\gamma_j(Y)$ por la expresión obtenida en la página precedente resulta:

$$\gamma_Y(B) = \sum_{j=-\infty}^{\infty}\left[\sigma_\varepsilon^2\sum_{i=0}^{\infty}\psi_i\psi_{i+j}\right]B^j = \sigma_\varepsilon^2\sum_{j=-\infty}^{\infty}\sum_{i=0}^{\infty}\psi_i\psi_{i+j}B^j$$

y, teniendo en cuenta que, para $i < 0$ se puede escribir $\gamma_Y(B) = \sigma_\varepsilon^2\sum_{i=0}^{\infty}\sum_{j=-i}^{\infty}\psi_i\psi_{i+j}B^j$, aplicando la sustitución $i + j = h$, de modo que $j = h - i$, se verificará que, cuando $j = -i$, entonces $h = 0$ y cuando $j = \infty$, $h = \infty$. Es decir, será:

$$\gamma_Y(B) = \sigma_\varepsilon^2\sum_{i=0}^{\infty}\sum_{h=0}^{\infty}\psi_i\psi_h B^{h-i} = \sigma_\varepsilon^2\left(\sum_{h=0}^{\infty}\psi_h B^h\right)\left(\sum_{i=0}^{\infty}\psi_i B^{-i}\right) = \sigma_\varepsilon^2\left(\sum_{h=0}^{\infty}\psi_h B^h\right)\left(\sum_{i=0}^{\infty}\psi_i F^i\right) = \sigma_\varepsilon^2\Psi(B)\Psi(F)$$

Asimismo, si el proceso $\{Y_t\}$ es estacionario con función de densidad espectral continua, entonces, admite una **representación bilateral** convergente de la forma $Y_t = \sum_{j=-\infty}^{\infty}\psi_j\varepsilon_{t-j}$, donde $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$ (en realidad, la condición de Normalidad no es esencial, las variables ε_t podrían ser, por ejemplo, uniformemente distribuidas).

---

[28]. Como se verá en la Sec. 5.2.2, de acuerdo con la definición de McLeod, A.I.; Hipel, K.W. (1978) (ver Resnick, S.I. (1987), Robinson, P.M.; Baillie, R.T. (1996)), si la función de autocovarianzas no es absolutamente sumable, entonces el proceso es puede ser considerado como de memoria larga.

Se dice que el proceso $\{Y_t\}$ admite una **representación polinomial bilateral** si puede ser expresado por un desarrollo convergente casi con certeza, de la forma:

$$Y_t = \sum_{p=1}^{\infty} \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} \cdots \sum_{i_p=-\infty}^{\infty} \psi_{i_1,i_2,\ldots,i_p} \prod_{\nu=1}^{p} \varepsilon_{t+i_\nu}$$

Luego, se puede concluir que: **i)** si $p = 1$, la representación anterior es lineal e $\{Y_t\}$ es una **serie cronológica lineal** (lo que demuestra que todo proceso Normal estacionario admite una representación bilateral) y **ii)** si $p > 1$, $\{Y_t\}$ es un **proceso no-lineal**.

Si el proceso $\{Y_t\}$ es no-Normal, estacionario y su función de densidad espectral es positiva en casi todo su dominio, entonces se puede asegurar que también admite una **representación lineal bilateral**. Si a la condición de no-negatividad de la función de densidad espectral se agrega la condición $\int_{-\infty}^{\infty} \ln[w_Y(\alpha)]d\alpha > -\infty$, entonces, se puede asegurar que existe una función $\varphi(\alpha)$ tal que:

$$\omega_Y(\alpha) = |\varphi_Y(\alpha)|^2 = \varphi_Y(\alpha)\varphi_Y^*(\alpha)$$

la cual puede ser desarrollada de acuerdo con una serie de Fourier de la forma $\varphi_Y(\alpha) = \sum_{j=0}^{\infty} \psi_j e^{-i\alpha j}$ y, por lo tanto, que el proceso $\{Y_t\}$ no-Normal estacionario admite una representación lineal de la forma $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ $(\psi_0 = 1)$.

Como se vio en la Sec.1.4.1, un proceso $\{Y_t\}$ lineal estacionario es **reversible en el dominio del tiempo**, si admite las representaciones alternativas $Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t+j} = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$. Esta expresión permite deducir las siguientes propiedades sobre los coeficientes ψ_j de esta **representación lineal bilateral**: **i)** es posible hallar un valor $k < \infty$ y un $s \in (0;1)$, tal que $\psi_{2k+j} = (-1)^s \psi_{2k-j}$ y **ii)** si se verifica que $\psi_0 = 1$ y $\psi_j = 0$, para $j < 0$ ó $j > q$, entonces, será $\psi_q = -1$ y $\psi_{q-j} = -\psi_j$ $(j = 1,2,\ldots,q-1)$ o $\psi_q = 1$ y $\psi_{q-j} = \psi_j$ $(j = 1,2,\ldots,q-1)$ [29].

2.9.2.- La representación AR

Haciendo ahora $j \to 0$ en la expresión:

$$Y_t = \eta_{t-j} + Y_{(t-j-1)} = \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_j\varepsilon_{t-j} + a_{j+1}Y_{t-j-1} + a_{j+2}Y_{t-j-2} + \ldots$$

correspondiente al teorema de descomposición, queda definida una representación del proceso $\{Y_t\}$ como una combinación lineal de sus realizaciones pasadas, más un "shock" aleatorio que representa la capacidad

---

[29]. Obsérvese que éste es un proceso no-previsible a partir de su filtración natural porque requiere información sobre su comportamiento en instantes posteriores a t.

de innovación del proceso (**representación autorregresiva** o **representación AR**):

$$Y_t = -\pi_1 Y_{t-1} - \pi_2 Y_{t-2} - ... + \varepsilon_t$$

$$\left(1 + \pi_1 B + \pi_2 B^2 + ...\right) Y_t = \varepsilon_t$$

$$\Pi(B) Y_t = \varepsilon_t$$

donde: **i)** Y_t denota una variable centrada; **ii)** $\{\varepsilon_t\}: WN - N\left(0, \sigma_\varepsilon^2\right)$ y **iii)** $\Pi(B)$ denota el **operador autorregresivo de orden infinito**[30].

Se denomina **estructura temporal** o **esqueleto** de una representación AR a la expresión que se obtiene de considerar a las innovaciones, ε_t, como constantes (o, lo que es equivalente, iguales a cero), para todo t, es decir, a la parte del comportamiento de $\{Y_t\}$ debida exclusivamente a su autorregresividad[31].

Si la sucesión de ponderaciones $\{\pi_j\}$ es tal que $\sum_{j=0}^{\infty} |\pi_j| < \infty$; es decir, si el operador $\Pi(B)$ es convergente para $|B| \leq 1$, se dice que el proceso $\{Y_t\}$ es **invertible**.

Aplicando m.a m. el operador $\Psi(B)$ en la representación AR, se obtiene que:

$$\Psi(B)\Pi(B) Y_t = \Psi(B)\varepsilon_t = Y_t$$

De lo que se concluye que $\Psi(B)\Pi(B) = 1$. Es decir, que $\Pi(B) = \left[\Psi(B)\right]^{-1}$. De acuerdo con esta igualdad, se puede concluir que la estacionariedad resume el conjunto de condiciones necesarias para poder transformar una representación AR en una representación MA y, viceversa, la invertibilidad define el conjunto de condiciones necesarias para transfomar una representación MA en una representación AR estacionaria[32].

---

30. Fue recién a partir de 1925 que E.Slutzsky (ver Slutzsky, E. (1925)(1927)(1929)(1934)(1937)) y G.U. Yule (ver Yule, G.U. (1927)), mediante el uso de ecuaciones estocásticas en-diferencias lineales, introdujeron las (sin ninguna connotación económica) nociones de sistema lineal y esquemas autorregresivos aplicables a procesos pseudo-periódicos estacionarios con innovaciones aleatorias, logrando una interpretación de la vinculación entre los esquemas autorregresivos y sus correspondientes representaciones MA (en sistemas lineales) propuesta en el teorema de Wold (ver Wold, H. (1949)(1953)(1954) (1956), Abril, J.C. (2004)). Debe tenerse en cuenta que la metodología expuesta aquí sobre procesos estocásticos estacionarios, filtros y componentes periódicos variables son posteriores y se deben fundamentalmente a A. Khinchin, A.N. Kolmogorov y N. Wiener.

31. Ver Tong, H. (237).

32. La condición de invertibilidad (que es atribuible aún a procesos lineales no-estacionarios) es independiente de la de estacionariedad. Igual que la estacionariedad, la invertibilidad no es una propiedad de $\{Y_t\}$ exclusivamente,sino de la relación entre los procesos $\{Y_t\}$ y $\{\varepsilon_t\}$. Granger, C.W.J.; Andersen, A. (1978) introdujeron una definición generalizada de invertibilidad aplicable a cualquier proceso de series cronológicas (aún aquellos no-lineales): Sea $\{Y_t\}$ un proceso que admite una representación de la forma $Y_t = f\left(t, Y_{t-1}, Y_{t-2}, ..., Y_{t-p}; \varepsilon_{t-1}, ..., \varepsilon_{t-q}\right) + \varepsilon_t$. A partir de una realización $\{Y_t, t = 1,2,...,n\}$ de este proceso, es posible obtener una estimación de las innovaciones ε_t de la forma $\hat{\varepsilon}_t = Y_t - f\left(t, Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, \hat{\varepsilon}_{t-1}, ..., \hat{\varepsilon}_{t-q}\right)$. Se dice que el proceso $\{Y_t\}$ es **invertible en el sentido de Granger-Andersen**, si se verifica que $\lim_{t \to \infty} \sigma^2\left(\hat{\varepsilon}_t - \varepsilon_t\right) = 0$, para cualquier conjunto de estimaciones iniciales $\hat{\varepsilon}_0, \hat{\varepsilon}_{-1}, ..., \hat{\varepsilon}_{-q+1}$.

2.10.- Los procesos autorregresivos de orden p

2.10.1.- Las condiciones de estacionariedad e invertibilidad

Dado que las representaciones autorregresivas y en promedios móviles presentadas en la sección precedente, involucran un número infinito de coeficientes (obviamente) no resultan útiles en las aplicaciones. En consecuencia, es necesario aproximar las representaciones de orden infinito mediante representaciones que incluyan un conjunto finito de coeficientes significativos.

Supóngase, por ejemplo, un proceso $\{Y_t\}$ que admita una representación autorregresiva en la cual sólo los p primeros coeficientes sean significativos, de la forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$), se dice entonces que $\{Y_t\}$ es un **proceso autorregresivo de orden p** ($\{Y_t\}$:$AR(p)$).

Utilizando el operador "backward" definido en la Sec.2.9.1, se puede escribir:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) Y_t = \Phi(B) Y_t = (1 - G_1^{-1} B)(1 - G_2^{-1} B)...(1 - G_p^{-1} B) Y_t = \varepsilon_t$$

donde $\Phi(B)$ denota el **operador autorregresivo de orden p**[33] y las G_i^{-1} $(i = 1,2,...,p)$ denotan las inversas de las raíces de la **ecuación característica del AR(p)**:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p = 0$$

La correspondiente **representación en promedios móviles del proceso AR(p)** será, entonces, de la forma:

$$Y_t = [\Phi(B)]^{-1} \varepsilon_t = \frac{1}{\prod_{i=1}^{p} (1 - G_i^{-1} B)} \varepsilon_t = \sum_{i=1}^{p} \frac{A_i}{1 - G_i^{-1} B} \varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} = \Psi(B) \varepsilon_t$$

Como se vio en el apartado precedente, la condición de estacionariedad está relacionada con la convergencia del operador $\Psi(B)$ para $|B| \leq 1$, y para que esa convergencia se cumpla, debe verificarse que $|G_i^{-1}| < 1 \, (i = 1,2,...,p)$. Es decir, que la condición necesaria y suficiente para el cumplimiento de la condición de estacionariedad de un $AR(p)$ es que todas las raíces de su ecuación característica sean, en módulo, mayores que la unidad[34].

---

[33]. Sea un proceso $\{Y_t^*\}$ estacionario de segundo orden, con valor medio $m(Y^*)$, su representación $AR(p)$ será de la forma $\Phi(B)(Y_t^* - m(Y^*)) = (1 - \phi_1 B - ... - \phi_p B^p)(Y_t^* - m(Y^*)) = \varepsilon_t$. Es decir, se puede escribir $\Phi(B) Y_t - (1 - \phi_1 - ... - \phi_p) m(Y^*) = \varepsilon_t$ o, lo que es lo mismo $\Phi(B) Y_t^* = \Phi(1) m(Y^*) + \varepsilon_t = \phi_0 + \varepsilon_t$.

[34]. En este caso, cuando $t \to \infty$, $\{Y_t\}$ tiende a una constante finita única e independiente del valor inicial -que define su atractor puntual. Si se verifica que la ecuación característica tiene una raíz, en módulo, menor que la unidad, entonces, cuando $t \to \infty$, $\{|Y_t|\}$ tiende a infinito, generando una situación inestable. Si se verifica que la ecuación característica tiene algunas raíces iguales a la

Una condición necesaria para que se verifique que $|G_i| > 1\,(i = 1,2,...,p)$ es que $\phi_1 + \phi_2 + ... + \phi_p < 1$. Como los ϕ_j pueden ser positivos o negativos, una condición suficiente para esta propiedad es que $|\phi_1| + |\phi_2| + ... + |\phi_p| < 1$. Si se verifica que $\phi_1 + \phi_2 + ... + \phi_p = 1$, se puede asegurar que la ecuación en-diferencias que rige el comportamiento de la estructura temporal de $\{Y_t\}$, $Y_t - \phi_1 Y_{t-1} - ... - \phi_p Y_{t-p} = 0$, posee al menos una raíz unitaria[35].

Por otra parte, dado que la representación $AR(p)$ del proceso $\{Y_t\}$ involucra sólo un número finito de coeficientes, se puede asegurar que la condición de invertibilidad se cumple en todos los casos.

2.10.2.- La función de autocorrelaciones

Dado que se supuso que las variables Y_t son centradas, multiplicando m. a m. la representación $AR(p)$ del proceso $\{Y_t\}$ por Y_{t-j}, y aplicando m. a m. el operador esperanza matemática, se obtiene el siguiente sistema de ecuaciones en diferencias para la función de autocovarianzas[36]:

$$\gamma_j(Y) - \phi_1\gamma_{j-1}(Y) - \phi_2\gamma_{|j-2|}(Y) - ... - \phi_p\gamma_{|j-p|}(Y) =$$

$$= \left(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_P B^p\right)\gamma_j(Y) = \Phi(B)\gamma_j(Y) = 0 \qquad (j = 1,2,...)$$

Para $j = 0$, se obtiene que $E(Y_t\varepsilon_t) = \sum_{j=0}^{\infty}\psi_j E(\varepsilon_{t-j}\varepsilon_t) = \sigma_\varepsilon^2$, por lo tanto, será:

$$\gamma_0(Y) = \sigma_Y^2 = \phi_1\gamma_1(Y) + \phi_2\gamma_2(Y) + ... + \phi_p\gamma_p(Y) + \sigma_\varepsilon^2$$

Para la función de autocorrelaciones se obtiene, a su vez, el siguiente sistema de ecuaciones en-diferencias:

$$\rho_j(Y) = \phi_1\rho_{j-1}(Y) + \phi_2\rho_{|j-2|}(Y) + ... + \phi_p\rho_{|j-p|}(Y) \qquad (j = 1,2,...)$$

Estas ecuaciones en diferencias son del mismo orden y con los mismos coeficientes que las correspondientes a la estructura temporal del proceso $\{Y_t\}$, por lo tanto, si el proceso es estacionario y las raíces de la ecuación característica son todas distintas entre sí, la función de autocorrelaciones[37]:

---

unidad, se genera una **situación neutralmente estable**, caracterizada por un comportamiento de $\{Y_t\}$ eventualmente oscilante en un conjunto de puntos cuyos valores dependen del valor inicial (debe tenerse en cuenta que las ecuaciones en-diferencias lineales no admiten soluciones periódicas estables independientes del valor inicial).

35. Como se mencionó en la sección precedente, esta representación del proceso $\{Y_t\}$ como una ecuación en-diferencias (en prinicipio desprovista de cualquier interpretación probabilística) se debe a G.U. Yule (ver Yule, G.U. (1911)(1927), Yule, G.U.; Kendall, M.G. (1950)).

36. Debe tenerse en cuenta que $E(Y_{t-j}\varepsilon_t) = \sum_{i=0}^{\infty}\psi_j E(\varepsilon_{t-j-i}\varepsilon_t) = 0\,(j \geq 1, \psi_0 = 1)$.

37. La única diferencia entre la estructura temporal del proceso $\{Y_t\}$ y su función de autocorrelaciones es que, mientras la primera puede asumir cualquier valor real, la segunda está restringida a límites muy precisos. No obstante, dado que los verdaderos valores de los coeficientes $\rho_j(Y)$ son desconocidos para el observador y que, por lo tanto, sus propiedades deben ser analizadas a partir de sus estimadores, debe tenerse en cuenta que la observación de un cierto comportamiento en la serie cronológica no necesariamente

$$\rho_j(Y) = B_1\left(G_1^{-1}\right)^j + B_2\left(G_2^{-1}\right)^j + \ldots + B_p\left(G_p^{-1}\right)^j \qquad (j = 1,2,\ldots)$$

(donde $B_i = \dfrac{\left(G_i^{-1}\right)^{p-1}}{\prod_{\substack{h=1 \\ h\neq i}}^{p}\left(G_i^{-1} - G_h^{-1}\right)}$), estará definida por una suma de funciones exponenciales decrecientes (originadas por las raíces reales de la ecuación característica $\Phi(B) = 0$) y por una suma de funciones sinusoidales amortiguadas (originadas en las raíces complejas de dicha ecuación). Se puede concluir, entonces, que la función de autocorrelaciones de un proceso $AR(p)$ queda unívocamente determinada por los p primeros valores de $\rho_j(Y)$ [38].

A partir de la definición obtenida para $\gamma_0(Y)$ resulta que:

$$1 = \phi_1\rho_1(Y) + \phi_2\rho_2(Y) + \ldots + \phi_p\rho_p(Y) + \frac{\sigma_\varepsilon^2}{\sigma_Y^2}$$

y, por lo tanto, que $\sigma_\varepsilon^2 = \sigma_Y^2\left[1 - \phi_1\rho_1(Y) - \phi_2\rho_2(Y) - \ldots - \phi_p\rho_p(Y)\right]$, es decir, que:

$$\sigma_Y^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1(Y) - \phi_2\rho_2(Y) - \ldots - \phi_p\rho_p(Y)}$$

Como se verá más adelante, esta expresión conduce a una descomposición de la varianza de acuerdo con las "contribuciones" correspondientes a las distintas autorregresividades que afectan el comportamiento de $\{Y_t\}$.

2.10.3.- La estimación de los coeficientes

Tomando las p primeras ecuaciones del sistema obtenido en la Sec. 2.10.2 para la función de autocorrelaciones, suponiendo que p sea conocido o, en caso contrario, que haya sido estimado en forma consistente (ver Sec. 2.10.5), se obtiene el denominado **sistema de Yule-Walker**, cuya solución, a partir de la sustitución de los coeficientes $\rho_j(Y)$ por sus estimadores:

$$\begin{cases} \hat{\rho}_1(Y) = \phi_1 + \phi_2\hat{\rho}_1(Y) + \ldots + \phi_p\hat{\rho}_{p-1}(Y) \\ \hat{\rho}_2(Y) = \phi_1\hat{\rho}_1(Y) + \phi_2 + \ldots + \phi_p\hat{\rho}_{p-2}(Y) \\ \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \hat{\rho}_p(Y) = \phi_1\hat{\rho}_{p-1}(Y) + \phi_2\hat{\rho}_{p-2}(Y) + \ldots + \phi_p \end{cases}$$

---

debe conducir a una única representación óptima del mecanismo generados de dichos datos. Puede haber otras representaciones compatibles con la misma serie cronológica.

[38]. En general, se verifica que, para aquellos órdenes para los cuales la función de autocorrelaciones es marcadamente significativa, el comportamiento del proceso puede ser aproximado mediante modelos lineales. Para los órdenes más alejados de cero con bajas autocorrelaciones es más apropiada una aproximación no-lineal

proporciona los **estimadores de Yule-Walker** de los coeficientes $\left(\hat{\phi}_1^{(YW)},\hat{\phi}_2^{(YW)},...,\hat{\phi}_p^{(YW)}\right)$ que permiten construir el modelo[39]:

$$\hat{Y}_t = \hat{\phi}_1^{(YW)}Y_{t-1} + \hat{\phi}_2^{(YW)}Y_{t-2}+...+\hat{\phi}_p^{(YW)}Y_{t-p}$$

En forma matricial este sistema puede ser expresado de la siguiente forma:

$$\hat{\rho}_p(Y) = \hat{\mathrm{P}}_p(Y)\Phi_p$$

donde $\hat{\rho}'_p(Y) = \left[\hat{\rho}_1(Y) \quad \hat{\rho}_2(Y) \quad ... \quad \hat{\rho}_p(Y)\right]$, $\Phi'_p = \left[\phi_1 \quad \phi_2 \quad ... \quad \phi_p\right]$ y, como se vio en la Sec. 2.7:

$$\hat{\mathrm{P}}_p(Y) = \begin{bmatrix} 1 & \hat{\rho}_1(Y) & \hat{\rho}_2(Y) & ... & \hat{\rho}_{p-1}(Y) \\ \hat{\rho}_1(Y) & 1 & \hat{\rho}_1(Y) & ... & \hat{\rho}_{p-2}(Y) \\ ... & ... & ... & ... & ... \\ \hat{\rho}_{p-1}(Y) & \hat{\rho}_{p-2}(Y) & \hat{\rho}_{p-3}(Y) & ... & 1 \end{bmatrix}$$

es una matriz no-singular. De modo que $\hat{\Phi}_p^{(YW)} = \hat{\mathrm{P}}_p^{-1}(Y)\hat{\rho}_p(Y)$ y,

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0(Y)\left(1 - \hat{\Phi}_p^{(YW)'}\hat{\rho}_p(Y)\right) = \hat{\gamma}_0(Y)\left(1 - \hat{\rho}'_p(Y)\hat{\Phi}_p^{(YW)}\right) = \hat{\gamma}_0(Y)\left[1 - \hat{\rho}'_p(Y)\hat{\mathrm{P}}_p^{-1}(Y)\hat{\rho}_p(Y)\right]$$

Estos estimadores (dado que se obtienen de igualar las autocovarianzas a sus correspondientes estimaciones) son asimilables a aquellos obtenidos por el método de los momentos y, suponiendo que el orden p sea conocido, son asintóticamente eficientes y convergen en-distribución a la función Normal cuando n aumenta indefinidamente.

Los **estimadores mínimo-cuadráticos** de los coeficientes se obtienen de aplicar el criterio de optimización de los cuadrados mínimos ordinarios (CMO):

$$f\left(\hat{\phi}_1^{CM},\hat{\phi}_2^{CM},...,\hat{\phi}_p^{CM}\right) = \sum_{t=p+1}^{n}\left(Y_t - \hat{\phi}_1^{CM}Y_{t-1} - \hat{\phi}_2^{CM}Y_{t-2} - ... - \hat{\phi}_p^{CM}Y_{t-p}\right)^2 = \sum_{t=p+1}^{n}\hat{\varepsilon}_t^2 = \min$$

(donde $\{\hat{\varepsilon}_t\}$ denota el **proceso residual**, que constituye la estimación del proceso representativo de las innovaciones, $\{\varepsilon_t\}$, y representa la parte de $\{Y_t\}$ no explicada por el modelo $AR(p)$) como aquellos valores que anulan las derivadas parciales de primer orden:

$$\frac{\partial}{\partial\hat{\phi}_j^{CM}} f\left(\hat{\phi}_1^{CM},\hat{\phi}_2^{CM},...,\hat{\phi}_p^{CM}\right) = 0 \qquad (j = 1,2,...,p)$$

Estos estimadores -que son asintóticamente equivalentes a los de Yule-Walker- sólo son formalmente análogos a los obtenidos en el modelo lineal general, ya que en los modelos dinámicos no se verifican algunos de los supuestos clásicos (por ejemplo, no se puede concebir a $Y_{t-1},Y_{t-2},...,Y_{t-p}$ como variables determinísticas exógenas, y a Y_t como una variable aleatoria). Si se verifica que $\{\varepsilon_t\}:WN - N\left(0,\sigma_\varepsilon^2\right)$, se demuestra que: **i)** los estimadores $\hat{\phi}_j^{(CM)}$ son asintóticamente eficientes; **ii)** de

---

[39]. A este tipo de modelos se los califica como **estocásticos** para distinguirlos de aquellos modelos de existencia puramente ideal en el ámbito de los fenómenos fácticos -llamados **determinísticos**- que proporcionan predicciones ciertas.

acuerdo con los postulados del teorema de Mann-Wald (ver Mann, H.B.; Wald, A. (1943)), la variable $\left\{\sqrt{n}\left(\hat{\phi}_1^{CM}-\phi_1\right);\sqrt{n}\left(\hat{\phi}_2^{CM}-\phi_2\right);\ldots;\quad \sqrt{n}\left(\hat{\phi}_p^{CM}-\phi_p\right)\right\}$ converge en-distribución a una variable Normal p-dimensional con valor esperado nulo y varianzas y covarianzas finitas y **iii)** los estimadores $\hat{\phi}_j^{(CM)}$ son consistentes[40].

En el caso en que $\{Y_t\}$ sea un proceso Normal, se puede obtener una estimación máximo-verosímil condicionada por los valores de las realizaciones $Y_1,Y_2,\ldots,Y_n$. A partir de la distribución de probabilidades conjunta de la variable $\{\varepsilon_{p+1},\varepsilon_{p+2},\ldots,\varepsilon_n\}$:

$$dF\left(\varepsilon_{p+1},\varepsilon_{p+2},\ldots,\varepsilon_n\right)=\frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{(n-p)/2}}\exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=p+1}^{n}\varepsilon_t^2\right)d\varepsilon_{p+1}\ldots d\varepsilon_n$$

se obtiene la función de probabilidades conjunta de la variable $\{Y_{p+1},Y_{p+2},\ldots,Y_n\}$:

$$dF\left(Y_{p+1},\ldots,Y_n\right)=\frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{(n-p)/2}}\exp\left[-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=p+1}^{n}\left(Y_t-\hat{\phi}_1^{MV}Y_{t-1}-\ldots-\hat{\phi}_p^{MV}Y_{t-p}\right)^2\right]dY_{p+1}\ldots dY_n$$

Los estimadores máximo-verosímiles se obtienen, entonces, de minimizar la función:

$$\sum_{t=p+1}^{n}\left(Y_t-\hat{\phi}_1Y_{t-1}-\hat{\phi}_2Y_{t-2}-\ldots-\hat{\phi}_pY_{t-p}\right)^2$$

por lo que se puede concluir que, en las condiciones mencionadas, los estimadores máximo-verosímiles son equivalentes a los mínimo-cuadráticos[41].

En las aplicaciones económicas, en las cuales las series cronológicas no suelen ser suficientemente largas, los más convenientes son los estimadores máximo-verosímiles, tomando los estimadores de Yule-Walker como valores iniciales en los algoritmos de optimización[42].

La matriz de varianzas y covarianzas de los estimadores es de la forma:

$$\Gamma_p\left(\hat{\Phi}\right)=\begin{bmatrix}\sigma^2\left(\hat{\phi}_1\right) & \gamma\left(\hat{\phi}_1,\hat{\phi}_2\right) & \ldots & \gamma\left(\hat{\phi}_1,\hat{\phi}_p\right)\\ \ldots & \ldots & \ldots & \ldots\\ \gamma\left(\hat{\phi}_1,\hat{\phi}_p\right) & \gamma\left(\hat{\phi}_2,\hat{\phi}_p\right) & \ldots & \sigma^2\left(\hat{\phi}_p\right)\end{bmatrix}=\frac{1}{n}\sigma_\varepsilon^2\Gamma_p^{-1}(Y)$$

---

40. Si la hipótesis $\{\varepsilon_t\}:WN-N\left(0,\sigma_\varepsilon^2\right)$ no se verifica y, además $\{Y_t\}$ presenta "outliers", es preferible utilizar métodos de estimación robusta (ver Martin, R.D. (1981), Martin, R.D.; Samorov, A.; Vandaele, W. (1981), Birch, J.B.; Martin, R.D. (1981)). Para más consideraciones acerca de los métodos de estimación de los cuadrados mínimos, de máxima verosmilitud y de los momentos, ver Cap. 6.

41. Ver Newbold, P. (1974), Granger, C.W.J.; Newbold, P. (1977).

42. Ver Ansley, C.F.; Newbold, P. (1980).

De acuerdo a lo demostrado en las páginas precedentes, será:

$$\Gamma_p\left(\hat{\Phi}\right) \approx \frac{1}{n}\gamma_0(Y)\left(1 - \rho_p^{'}(Y)\Phi_p\right)\Gamma_p^{-1}(Y) = \frac{1}{n}\left(1 - \rho_p^{'}(Y)\Phi_p\right)\mathrm{P}_p^{-1}(Y)$$

Sustituyendo en esta ecuación los coeficientes por sus estimadores, se obtiene el estimador de la matriz de varianzas y covarianzas de los estimadores, $\hat{\Gamma}_p\left(\hat{\Phi}\right) \approx \frac{1}{n}\left(1 - \hat{\rho}_p^{'}(Y)\hat{\Phi}_p\right)\hat{\mathrm{P}}_p^{-1}(Y)$. Luego, a partir del supuesto que $\{\varepsilon_t\}: N\left(0, \sigma_\varepsilon^2\right)$, se verificará que:

$$\hat{\Phi}_p \xrightarrow{d} N\left(\Phi_p, \frac{1}{n}\left(1 - \rho_p^{'}(Y)\Phi_p\right)\mathrm{P}_p^{-1}(Y)\right)$$

donde $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$.

A fin de decidir acerca de la significatividad individual de los coeficientes ϕ_j, lo habitual es utilizar la misma metodología que en los modelos clásicos de regresión lineal mediante la construcción de un test "t" y utilizando un estadístico de la forma $T\left(\hat{\phi}_j\right) = \frac{\hat{\phi}_j}{\hat{\sigma}\left(\hat{\phi}_j\right)} : t_{(n-p)}$ $(j = 1,2,...,p)$ (si el modelo posee término independiente, la distribución del estadístico T será $t_{(n-p-1)}$). Ahora bien, dado que, obviamente, estos estadísticos, bajo el supuesto que la hipótesis nula $H_0: \phi_j = 0$ en verdadera, no tienen distribución "t", este test es útil sólo si los mismos resultan "marcadamente significativos". Si su significatividad es débil no se debe confiar demasiado en el test, dado que las variables Y_t pueden muy bien no ser Normales. Debe tenerse en cuenta, además, que en las proximidades de la no-estacionariedad, los estadísticos T son muy sesgados.

2.10.4.- La función de autocorrelaciones parciales

La **función de autocorrelaciones parciales**, $\phi_{jj}(Y)\,(j = 1,2,...)$, es una función de los coeficientes de autocorrelación $\rho_j(Y)$ que contribuye a seleccionar "a priori" el orden p adecuado en una representación autorregresiva, a partir de las $p\,(= 1,2,...)$ primeras ecuaciones del sistema de infinitas ecuaciones en diferencias que, como se vio en la Sec. 2.10.2, define el comportamiento de la función de autocorrelaciones.

La **autocorrelación parcial de orden j** ($\phi_{jj}(Y)$) puede ser definida como el grado de relación lineal entre las variables Y_{t-j} e Y_t depurada de las influencias que sobre ambas ejercen las variables intermedias, $Y_{t-1}, Y_{t-2}, ..., Y_{t-(j-1)}$.

Como se vio en la Sec.2.10.1, de la aplicación a las variables Y_t e Y_{t-j} de representaciones $AR(j-1)$ de la forma:

$$Y_t = \phi_{1,j-1}Y_{t-1} + \phi_{2,j-1}Y_{t-2} + ... + \phi_{j-1,j-1}Y_{t-(j-1)} + e_t$$

$$Y_{t-j} = \phi^*_{1,j-1}Y_{t-(j-1)} + \phi^*_{2,j-1}Y_{t-(j-2)} + ... + \phi^*_{j-1,j-1}Y_{t-1} + e_{t-j}$$

se obtienen las perturbaciones e_{t-j} y e_t. La autocorrelación parcial de orden j puede ser definida, entonces, como la correlación entre las perturbaciones e_{t-j} y e_t:

$$\phi_{jj}(Y) = \rho\left(Y_{t-j}, Y_t / Y_{t-1}, Y_{t-2}, ..., Y_{t-(j-1)}\right) = \rho\left(e_{t-j}, e_t\right)$$

Si, como en los casos que se han tratado hasta ahora, el proceso $\{Y_t\}$ no es reversible, la relación causal entre la variable Y_{t-j} y las variables posteriores ("futuras" para la variable Y_{t-j}) $Y_{t-1}, Y_{t-2}, ..., Y_{t-(j-1)}$ será nula, de modo que $Y_{t-j} = e_{t-j}$ y, por lo tanto, el coeficiente $\phi_{jj}(Y)$ puede ser definido como la correlación entre la variable Y_{t-j} y la perturbación obtenida de aplicar al proceso $\{Y_t\}$ una representación $AR(j-1)$, $\phi_{jj}(Y) = \rho\left(Y_{t-j}, e_t\right)$.

A partir de las ecuaciones:

$$E\left[\left(Y_t - \sum_{i=1}^{p} \phi_{pi} Y_{t-i}\right) Y_{t-j}\right] = E\left(Y_{t-j} Y_t\right) - \sum_{i=1}^{p} \phi_{pi} E\left(Y_{t-i} Y_{t-j}\right) = 0 \qquad (j = 1,2,...,p)$$

(donde ϕ_{pj} denota el j-ésimo coeficiente de un proceso-autorregresivo de orden p), se obtiene el siguiente sistema de ecuaciones de Yule-Walker:

$$\rho_j(Y) = \phi_{p1}\rho_{j-1}(Y) + \phi_{p2}\rho_{j-2}(Y) + ... + \phi_{p,p-1}\rho_{j-p+1}(Y) + \phi_{pp}\rho_{j-p}(Y) \qquad (j = 1,2,...,p)$$

cuya solución, para $p = 1$, proporciona el **coeficiente de autocorrelación parcial de primer orden**, $\phi_{11}(Y) = \rho_1(Y)$. De la misma forma, para $p = 2$, el sistema de ecuaciones de Yule-Walker asume la forma:

$$\begin{cases} \rho_2(Y) = \phi_{21}\rho_1(Y) + \phi_{22} \\ \rho_1(Y) = \phi_{21} + \phi_{22}\rho_1(Y) \end{cases}$$

o, utilizando notación matricial, $\begin{bmatrix} \rho_2(Y) \\ \rho_1(Y) \end{bmatrix} = P_2(Y) \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix}$, donde $P_2(Y) = \begin{bmatrix} \rho_1(Y) & 1 \\ 1 & \rho_1(Y) \end{bmatrix}$. De modo que:

$$\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = P_2^{-1}(Y) \begin{bmatrix} \rho_2(Y) \\ \rho_1(Y) \end{bmatrix}$$

De la solución de este sistema de ecuaciones se obtiene el **coeficiente de autocorrelación parcial de segundo orden**[43]:

$$\phi_{22}(Y) = \rho\left(Y_{t-2}, Y_t / Y_{t-1}\right) = \frac{\rho_2(Y) - \rho_1^2(Y)}{1 - \rho_1^2(Y)}$$

Para $p = 3$, será:

---

43. A partir de la interpretación que figura en el texto, se obtiene que $\phi_{22}(Y) = \rho\left(Y_{t-2} - \phi_1 Y_{t-1}, Y_t - \phi_1 Y_{t-1}\right) =$

$$= \rho\left(Y_{t-2} - \rho_1(Y) Y_{t-1}, Y_t - \rho_1(Y) Y_{t-1}\right) = \frac{\gamma_2(Y) - 2\rho_1(Y)\gamma_1(Y) - \rho_1^2(Y)\gamma_0(Y)}{\gamma_0(Y) + \rho_1^2(Y)\gamma_1(Y) - \rho_1(Y)\gamma_1(Y)} = \frac{\rho_2(Y) - \rho_1^2(Y)}{1 - \rho_1^2(Y)}$$

$$\begin{cases} \rho_3(Y) = \phi_{31}\rho_2(Y) + \phi_{32}\rho_1(Y) + \phi_{33} \\ \rho_2(Y) = \phi_{31}\rho_1(Y) + \phi_{32} + \phi_{33}\rho_1(Y) \\ \rho_1(Y) = \phi_{31} + \phi_{32}\rho_1(Y) + \phi_{33}\rho_2(Y) \end{cases}$$

A partir de las dos últimas ecuaciones de este sistema se puede obtener una expresión para los coeficientes ϕ_{31} y ϕ_{32} en función de ϕ_{33} :

$$\phi_{31}\rho_1^2(Y) = \rho_1(Y)\rho_2(Y) - \left[\rho_1(Y) - \phi_{31} - \phi_{33}\rho_2(Y)\right] - \phi_{33}\rho_1^2(Y)$$

$$\phi_{32}\rho_1^2(Y) = \rho_1^2(Y) - \left[\rho_2(Y) - \phi_{32} - \phi_{33}\rho_1(Y)\right] - \phi_{33}\rho_1^2(Y)$$

Es decir,

$$\phi_{31} = -\frac{\rho_1(Y)\rho_2(Y) - \phi_{33}\rho_1^2(Y) - \rho_1(Y) + \phi_{33}\rho_2(Y)}{1 - \rho_1^2(Y)} = -\frac{\rho_1(Y)\left[\rho_2(Y) - \phi_{33}\rho_1(Y)\right] - \left[\rho_1(Y) - \phi_{33}\rho_2(Y)\right]}{1 - \rho_1^2(Y)}$$

$$\phi_{32} = -\frac{-\rho_2(Y) + \phi_{33}\rho_1(Y) + \rho_1^2(Y) - \phi_{33}\rho_1(Y)\rho_2(Y)}{1 - \rho_1^2(Y)} = -\frac{-\left[\rho_2(Y) - \phi_{33}\rho_1(Y)\right] + \rho_1(Y)\left[\rho_1(Y) - \phi_{33}\rho_2(Y)\right]}{1 - \rho_1^2(Y)}$$

o, utilizando notación matricial:

$$\begin{bmatrix} \phi_{31} \\ \phi_{32} \end{bmatrix} = -\frac{1}{1 - \rho_1^2(Y)} \begin{bmatrix} \rho_1(Y) & -1 \\ -1 & \rho_1(Y) \end{bmatrix} \begin{bmatrix} \rho_2(Y) - \phi_{33}\rho_1(Y) \\ \rho_1(Y) - \phi_{33}\rho_2(Y) \end{bmatrix} = \mathrm{P}_2^{-1}(Y) \begin{bmatrix} \rho_2(Y) - \phi_{33}\rho_1(Y) \\ \rho_1(Y) - \phi_{33}\rho_2(Y) \end{bmatrix} =$$

$$= \mathrm{P}_2^{-1}(Y) \begin{bmatrix} \rho_2(Y) \\ \rho_1(Y) \end{bmatrix} - \phi_{33}\mathrm{P}_2^{-1}(Y) \begin{bmatrix} \rho_1(Y) \\ \rho_2(Y) \end{bmatrix} = \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} - \phi_{33} \begin{bmatrix} \phi_{22} \\ \phi_{21} \end{bmatrix}$$

Es decir, se puede escribir:

$$\phi_{31} = \phi_{21} - \phi_{22}\phi_{33}$$

$$\phi_{32} = \phi_{22} - \phi_{21}\phi_{33}$$

Reemplazando estos resultados en la primera ecuación del sistema que figura en la página precedente, $\rho_3(Y) = \phi_{31}\rho_2(Y) + \phi_{32}\rho_1(Y) + \phi_{33}$, se obtiene el **coeficiente de autocorrelación parcial de tercer orden:**

$$\phi_{33}(Y) = \frac{\rho_3(Y) - \phi_{21}\rho_2(Y) - \phi_{22}\rho_1(Y)}{1 - \phi_{21}\rho_1(Y) - \phi_{22}\rho_2(Y)}$$

En general, la función de autocorrelaciones parciales puede ser expresada de la siguiente forma recursiva:

$$\phi_{jj}(Y) = \frac{\rho_j(Y) - \sum_{i=1}^{j-1} \phi_{j-1,i}\rho_{j-i}(Y)}{1 - \sum_{i=1}^{j-1} \phi_{j-1,i}\rho_i(Y)} \qquad (j = 2,3,\ldots)$$

(donde $\phi_{ji} = \phi_{j-1,i} - \phi_{jj}(Y)\phi_{j-1,j-i}$ $(i = 1,2,\ldots)$). Expresión conocida como **algoritmo de Durbin-Levinson**

(ver Durbin, J. (1970), Brockwell, P.J.; Davis, R.A. (1987)).

Utilizando notación matricial se puede escribir:

$$\phi_{jj}(Y) = \left| \frac{\mathrm{P}_j^*(Y)}{\mathrm{P}_j(Y)} \right| \qquad (j = 1,2,...)$$

donde:

$$\mathrm{P}_j^*(Y) = \begin{bmatrix} 1 & \rho_1(Y) & \rho_2(Y) & \dots & \rho_{j-2}(Y) & \rho_1(Y) \\ \rho_1(Y) & 1 & \rho_1(Y) & \dots & \rho_{j-3}(Y) & \rho_2(Y) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{j-1}(Y) & \rho_{j-2}(Y) & \rho_{j-3}(Y) & \dots & 1 & \rho_j(Y) \end{bmatrix}$$

y $P_j(Y)$ es la matriz de Laurent definida en la Sec. 2.7.

Para un proceso autorregresivo de orden p, la función de autocorrelaciones parciales asume valores significativos hasta el orden p, y después se anula. Lo cual implica que, obviamente, para $j > p$ no existe relación lineal directa entre las variables Y_{t-j} e Y_t, la única relación existente se debe a que las variables están vinculadas secuencialmente a través de las variables intermedias. Esta propiedad, como se mencionó al comienzo de esta sección, convierte a la función de autocorrelaciones parciales en una herramienta esencial en una selección "a priori" del orden más adecuado de la representación autorregresiva de un proceso estocástico.

En los cuadros **2.2** y **2.3** figuran las funciones de autocorrelaciones y de autocorrelaciones parciales típicas de los procesos $AR(1)$ estacionarios y, en los cuadros **2.4**, **2.5** y **2.6** figuran las funciones de autocorrelaciones y de autocorrelaciones parciales típicas de los procesos $AR(2)$ estacionarios (los procesos $AR(1)$ y $AR(2)$ serán tratados en particular, en las secciones 2.10.7 y 2.10.8).

Si en las expresiones anteriores se sustituyen los coeficientes $\rho_j(Y)$ por sus estimadores, se obtienen los estimadores mínimo-cuadráticos de los coeficientes $\phi_{jj}(Y)$ de un modelo $AR(j)$:

$$\hat{\phi}_{jj}(Y) = \left| \frac{\hat{\mathrm{P}}_j^*(Y)}{\hat{\mathrm{P}}_j(Y)} \right| \qquad (j = 1,2,...)$$

Quenouille, M. H. (1949) demostró que, bajo el supuesto que la hipótesis nula $H_0: AR(p)$ es verdadera (es decir, bajo el supuesto que la hipótesis nula $H_0: \phi_{jj}(Y) = 0 \; (j = p+1, p+2,...)$ es verdadera), los estimadores $\hat{\phi}_{jj}(Y) \; (j = p+1, p+2,...)$ son aproximadamente independientes, con varianzas $\sigma^2\left(\hat{\phi}_{jj}(Y)\right) \approx \frac{1}{n}$ y asintóticamente Normales [44]. Es decir, que:

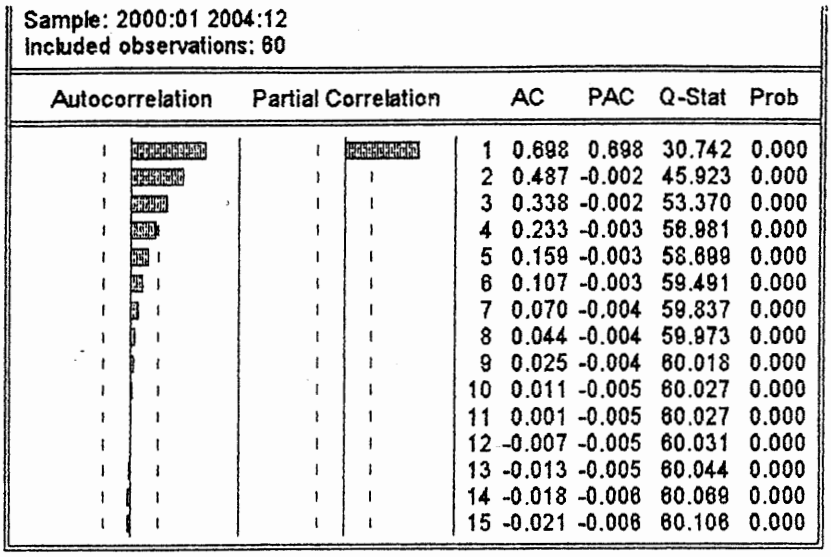
$$\hat{\phi}_{jj}(Y) \xrightarrow{d} N\left(0; \frac{1}{n}\right) \qquad (j = p+1, p+2,...)$$

Debe tenerse en cuenta que cuando una o más raíces de la ecuación característica sean próximas

---

[44]. Ver también Jenkins, G.M. (1954)(1956), Jenkins, G.M.; Watts, D.G. (1969), Kendall, M.G. (1973), Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. (1994).

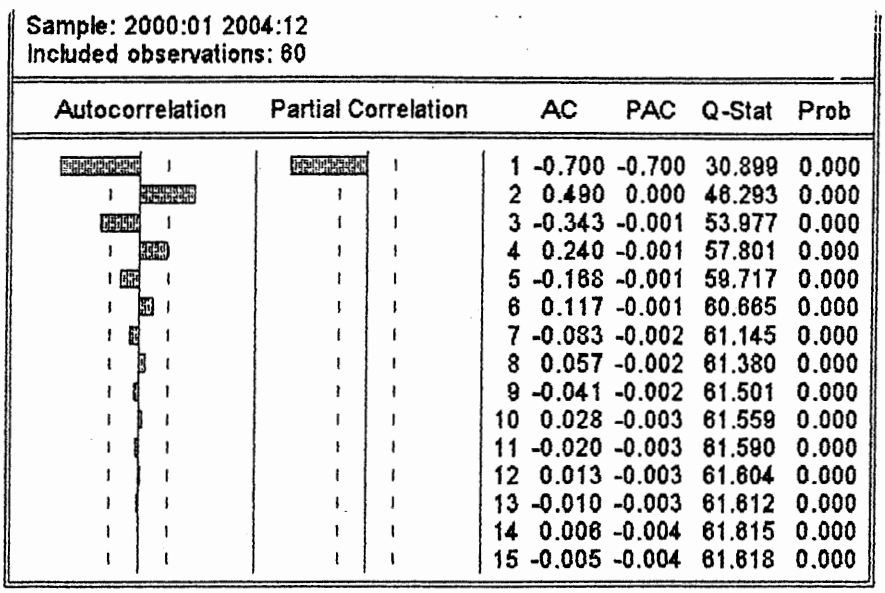
CUADRO 2.2: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación AR(1) con φ_1=0.7



Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.698	0.698	30.742	0.000
		2	0.487	-0.002	45.923	0.000
		3	0.338	-0.002	53.370	0.000
		4	0.233	-0.003	56.981	0.000
		5	0.159	-0.003	58.699	0.000
		6	0.107	-0.003	59.491	0.000
		7	0.070	-0.004	59.837	0.000
		8	0.044	-0.004	59.973	0.000
		9	0.025	-0.004	60.018	0.000
		10	0.011	-0.005	60.027	0.000
		11	0.001	-0.005	60.027	0.000
		12	-0.007	-0.005	60.031	0.000
		13	-0.013	-0.005	60.044	0.000
		14	-0.018	-0.006	60.069	0.000
		15	-0.021	-0.006	60.106	0.000

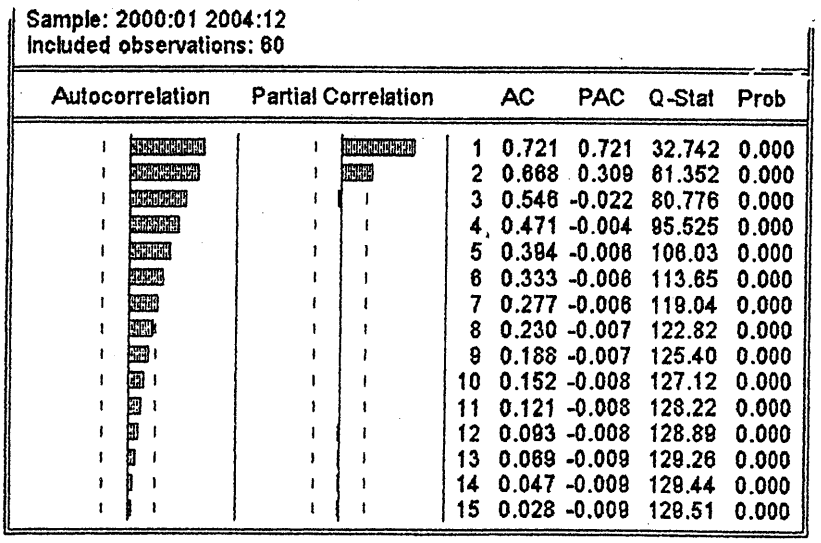
CUADRO 2.3: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación AR(1) con φ_1=-0.7



Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.700	-0.700	30.899	0.000
		2	0.490	0.000	46.293	0.000
		3	-0.343	-0.001	53.977	0.000
		4	0.240	-0.001	57.801	0.000
		5	-0.168	-0.001	59.717	0.000
		6	0.117	-0.001	60.665	0.000
		7	-0.083	-0.002	61.145	0.000
		8	0.057	-0.002	61.380	0.000
		9	-0.041	-0.002	61.501	0.000
		10	0.028	-0.003	61.559	0.000
		11	-0.020	-0.003	61.590	0.000
		12	0.013	-0.003	61.604	0.000
		13	-0.010	-0.003	61.612	0.000
		14	0.006	-0.004	61.615	0.000
		15	-0.005	-0.004	61.618	0.000

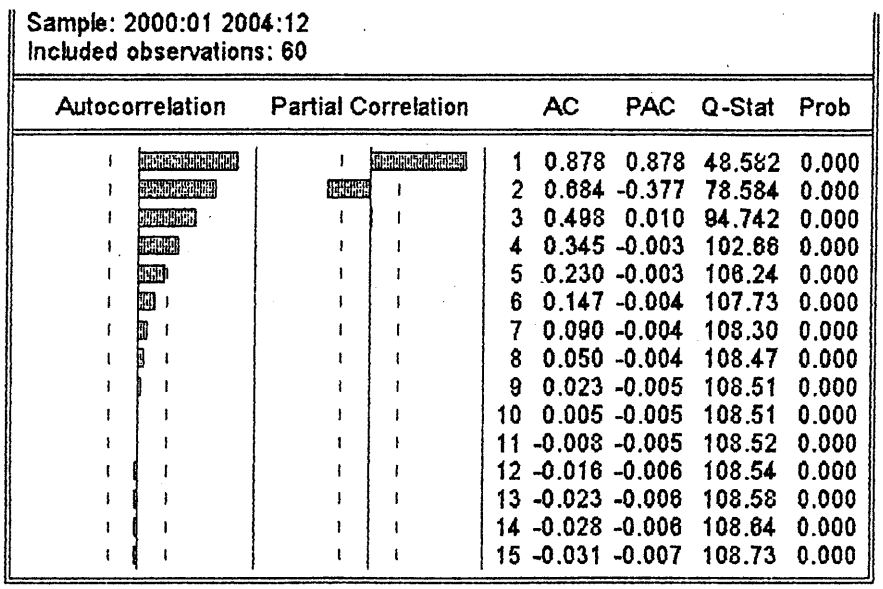
CUADRO 2.4: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación AR(2) con φ_1=0.45 y φ_2=0.35



Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.721	0.721	32.742	0.000
		2	0.668	0.309	61.352	0.000
		3	0.546	-0.022	80.776	0.000
		4	0.471	-0.004	95.525	0.000
		5	0.394	-0.006	106.03	0.000
		6	0.333	-0.006	113.65	0.000
		7	0.277	-0.006	118.04	0.000
		8	0.230	-0.007	122.82	0.000
		9	0.188	-0.007	125.40	0.000
		10	0.152	-0.008	127.12	0.000
		11	0.121	-0.008	128.22	0.000
		12	0.093	-0.008	128.89	0.000
		13	0.069	-0.009	129.26	0.000
		14	0.047	-0.009	129.44	0.000
		15	0.028	-0.009	129.51	0.000

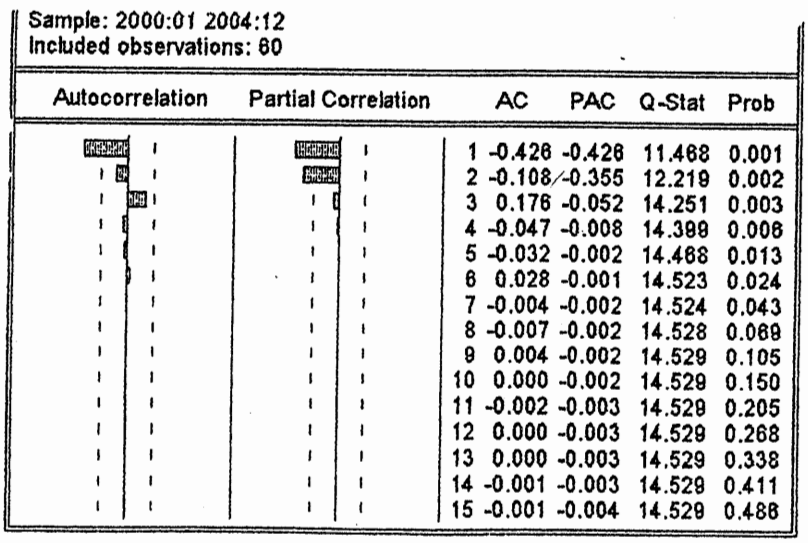
CUADRO 2.5: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación AR(2) con φ_1=1.18 y φ_2=-0.35



Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.878	0.878	48.582	0.000
		2	0.684	-0.377	78.584	0.000
		3	0.498	0.010	94.742	0.000
		4	0.345	-0.003	102.66	0.000
		5	0.230	-0.003	106.24	0.000
		6	0.147	-0.004	107.73	0.000
		7	0.090	-0.004	108.30	0.000
		8	0.050	-0.004	108.47	0.000
		9	0.023	-0.005	108.51	0.000
		10	0.005	-0.005	108.51	0.000
		11	-0.008	-0.005	108.52	0.000
		12	-0.016	-0.006	108.54	0.000
		13	-0.023	-0.006	108.58	0.000
		14	-0.028	-0.006	108.64	0.000
		15	-0.031	-0.007	108.73	0.000

CUADRO 2.6: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación AR(2) con φ_1=-0.45 y φ_2=-0.30



Sample: 2000:01 2004:12
Included observations: 80

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.426	-0.426	11.468	0.001
		2	-0.108	-0.355	12.219	0.002
		3	0.176	-0.052	14.251	0.003
		4	-0.047	-0.008	14.398	0.006
		5	-0.032	-0.002	14.468	0.013
		6	0.028	-0.001	14.523	0.024
		7	-0.004	-0.002	14.524	0.043
		8	-0.007	-0.002	14.528	0.069
		9	0.004	-0.002	14.529	0.105
		10	0.000	-0.002	14.529	0.150
		11	-0.002	-0.003	14.529	0.205
		12	0.000	-0.003	14.529	0.268
		13	0.000	-0.003	14.529	0.338
		14	-0.001	-0.003	14.529	0.411
		15	-0.001	-0.004	14.529	0.488

a la unidad, estas propiedades de los estimadores $\hat{\phi}_{jj}(Y)$ se debilitan .

Ejemplo 2.2:

Sea un proceso $\{Y_t\}$ expresado en términos mensuales, cuya trayectoria para el período [ENE-1997;Dic-2003] figura en el **GRAFICO 2.3** y cuyos estimadores de las funciones de autocorrelaciones y autocorrelaciones parciales están representadas en el correlograma que figura en el **CUADRO 2.7**. De la observación de los estadísticos Q_j $(j = 1,2,...,20)$ se puede concluir que existen razones suficientes para rechazar la hipótesis nula $H_0:\{Y_t\}:AR(0)$ a partir de $j = 1$.

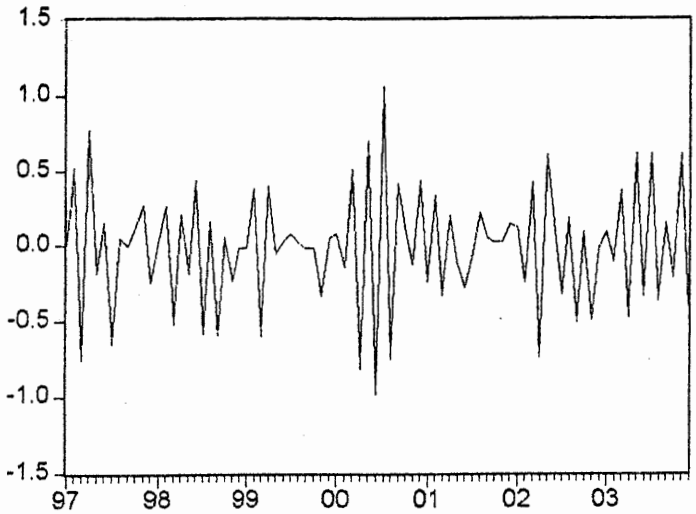
La consecuente observación de los estimadores individuales de la función de autocorrelaciones parciales permite concluir que el coeficiente $\phi_{11}(Y)$ puede ser considerado significativo con una confiabilidad del 95% y que, a partir de $j = 2$ y con el mismo nivel de confiabilidad, no existen razones suficientes para rechazar la hipótesis de nulidad individual de los $\phi_{jj}(Y)$ (para $\alpha = 0.05$, los límites de la zona crítica están dados por $\pm \frac{2}{\sqrt{84}} = \pm 0.218$).

Luego, considerando la forma de decrecimiento de las funciones de autocorrelaciones y de autocorrelaciones parciales, se puede concluir que, en principio, el proceso $\{Y_t\}$ admite una representación $AR(1)$.

2.10.5.- La selección del orden de autorregresividad

El problema de la selección del orden de autorregresividad en los modelos AR constituye un caso particular de la cuestión relacionada con la selección de modelos en general (estáticos con múltiples variables explicativas, de funciones de transferencia, vectores autorregresivos, autorregresivos con

GRAFICO 2.3: Trayectoria del proceso $\{Y_t\}$ para el período ENE-2001;DIC-2007)



CUADRO 2.7: Correlograma correspondiente al proceso $\{Y_t\}$ cuya trayectoria figura en el GRAFICO 2.3

Sample: 1997:01 2003:12
Included observations: 84

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.739	-0.739	47.583	0.000
		2	0.542	-0.011	73.456	0.000
		3	-0.369	0.062	85.614	0.000
		4	0.195	-0.120	89.047	0.000
		5	-0.055	0.087	89.327	0.000
		6	-0.072	-0.100	89.810	0.000
		7	0.101	-0.087	90.769	0.000
		8	-0.126	-0.034	92.276	0.000
		9	0.104	-0.052	93.310	0.000
		10	-0.064	0.015	93.715	0.000
		11	-0.046	-0.177	93.922	0.000
		12	0.030	-0.211	94.010	0.000
		13	0.009	0.068	94.018	0.000
		14	0.002	0.047	94.019	0.000
		15	0.031	0.037	94.119	0.000
		16	-0.032	0.036	94.230	0.000
		17	0.014	-0.108	94.253	0.000
		18	0.078	0.158	94.912	0.000
		19	-0.148	-0.052	97.346	0.000
		20	0.184	-0.004	101.16	0.000

heterocedasticidad condicionada)[45].

El criterio habitualmente aceptado es que cuantas más variables explicativas se incorporen al modelo (es decir, cuánto más complejo sea el modelo), mayor será su grado de representatividad del comportamiento del proceso. Los métodos de selección intentan, precisamente, optimizar la relación entre la complejidad y el grado de explicabilidad del modelo mediante la combinación de **funciones de distancia** al modelo óptimo y **funciones de penalización** por el aumento de la complejidad del modelo[46].

Históricamente. el primer criterio de selección de modelos fue el coeficiente de determinación ajustado:

$$\overline{R}^2_{Y/p} = 1 - \frac{n-1}{n-p} \frac{\sum_{t=p+1}^{n} \hat{\varepsilon}_t^2}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2}$$

cuyas limitaciones en los modelos dinámicos serán comentadas en las secciones 2.10.7 y 2.10.8 y analizadas con más detalle en la Sec. 6.1.7. Debe recordarse que el criterio de determinación:

$$R^2_{Y/p} = 1 - \frac{\sum_{t=p+1}^{n} \hat{\varepsilon}_t^2}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2}$$

aumenta su valor siempre que se agrega al modelo una variable explicativa, sin tomar en consideración su contribución a la representación de $\{Y_t\}$.

Un segundo método de selección que no pertenece a la familia de los criterios con término de penalización, debido a Quenouille, M.H. (1947), se basa en los tests de significatividad conjunta de los coeficientes de una representación.

Por último, a fines de la década de 1960 y comienzos de la década de 1970 surge una nueva metodología basada inicialmente en los trabajos de Akaike, H. (1969) y Mallows, C.L. (1973) y continuada (en forma acorde con el desarrollo de la teoría de la información y en función de la medida de distancia dirigida de Kullbac-Leibler) en los trabajos de Akaike, H. (1973)(1974). Estas propuestas fueron completadas por las publicaciones de Bhansali, R.J.; Downham, D.Y. (1977), Akaike, H. (1978), Schwarz, G. (1978), Hannan, E.J.: Quinn, B.G. (1979). Geweke, J.; Meese, R. (1981), Shibata, R. (1981), Hurvich, C.M.; Tsai, C.L. (1989).

2.10.5.1.- Criterios eficientes y criterios consistentes

La clasificación de los criterios de selección del orden de autorregresividad en **consistentes** y **eficientes** se basa, respectivamente, en el supuesto de que la **verdadera representación** del proceso pueda ser considerada o no como perteneciente al conjunto de los **modelos candidatos**[47].

---

[45]. Estos modelos serán tratados en las partes segunda y tercera de este trabajo.

[46]. Con respecto al significado de la condición de óptimo, debe tenerse en cuenta la opinión de G. Box según el cual "*All models are wrong, some models are useful*".

[47]. Obsérvese que la clasificación de un criterio como consistente implica suponer que el proceso es **anticipatorio** (de acuerdo con la nomenclatura de Rosen, R. (1985)), en otros términos que, con respecto a este proceso, la **tesis de Turing-Church** es falsa. Esta tesis, desarrollada por el lógico Alonzo Church a partir del trabajo de Alan Turing, supone que la complejidad de la estructura causal de un proceso (uni o multidimensional) involucra más vínculos que cualquier expresión sintáctica (modelo) de su

Sea un proceso $\{Y_t\}$ $(t = 1,2,...)$ estacionario de segundo orden y tal que las variables Y_t sean centradas. Si se supone que su verdadera representación AR es de orden infinito, entonces se puede concluir que el conjunto de las representaciones AR candidatas de orden finito no contiene a la verdadera. En este caso el método de selección consiste en elegir aquella representación de orden finito que se aproxime más a la verdadera. Obviamente esta idea de mejor aproximación implica la minimización de una **función de distancia** (o **discrepancia**[48] o **información**) entre la verdadera representación y cada una de las representaciones candidatas.

Se dice que un criterio de selección del orden p es **asintóticamente eficiente**[49] si, para una serie cronológica suficientemente larga, la función a minimizar es la distribución del error medio cuadrático que se genera al aplicar un modelo distinto del verdadero o, alternativamente, el error medio cuadrático de la predicción que proporciona el modelo candidato[50].

Supóngase que la verdadera representación del proceso $\{Y_t\}$ sea de orden finito y que, por lo tanto, esté incluida en el conjunto de las representaciones candidatas, entonces el método de selección consiste en identificar, en este conjunto, la verdadera representación[51]. Este criterio de selección, que identifica asintóticamente el verdadero modelo con probabilidad igual a uno, se dice que es **consistente**[52] (obsérvese que, mientras en los criterios eficientes existe una probabilidad no-nula de estimación errónea del orden p, en los criterios consistentes, esta probabilidad es asintóticamente nula).

La estimación errónea implica una subestimación o una sobre-estimación del orden p. Se dice que una representación está subsestimada (sobre-estimada) si su orden de autorregresividad es menor (mayor) que el de la verdadera representación.

El problema que se presenta habitualmente en las aplicaciones se reduce a seleccionar el mejor de un conjunto de modelos candidatos. La regla en estos casos es que el mejor modelo es aquél al que le corresponde el menor valor del criterio de selección. Sean $CSM(p_1)$ y $CSM(p_2)$ los valores asumidos por un criterio de selección para dos modelos $AR(p_1)$ y $AR(p_2)$, respectivamente, siendo $p_2 > p_1$, entonces, si se verifica que $CSM(p_1) < CSM(p_2)$, se considerará que el primer modelo es preferible al segundo. La esperanza matemática de la variable que representa la diferencia entre ambos valores, $E[CSM(p_2) - CSM(p_1)]$, define una medida denominada **señal del criterio de selección**, $S[CSM(p_1, p_2)]$ (que, en series cortas, depende fundamentalmente de la importancia relativa de la función de penalización del criterio), y su desvío estándar, $\sigma[CSM(p_2) - CSM(p_1)]$, define otra medida denominada **ruido del**

---

comportamiento (ver Davis, M. (1958), Kolmogorov, A.N. (1965), Rosen, R. (1962)(en Day, D.H.; Groves, T. (eds.) (1975), (en Mintz, S.L.; Perlmutter, A. (eds.) (1985))(1988), Davis, M.; Weymuker, E. (1983), Chaitin, G. (1987)(1988)(1990), Wang, H. (1987), Casti, L. (1991), Casti, L.; Karlqvist, A. (1991).

48. De acuerdo con la denominación de Haldane, J.B.S. (1951).

49. Denominación debida a Shibata, R. (1980).

50. Como se verá en las secciones siguientes, entre estos criterios cabe mencionar el PRESS de Allen, D.M. (1974) y el FPE de Akaike, H. (1969). Para un análisis de la relación entre los criterios de predicción y la propiedad de eficiencia asintótica, ver Shibata, R. (1980).

51. El supuesto de la existencia, en todo caso, de una verdadera representación que pertenece al conjunto de las representaciones candidatas es, obviamente, una restricción muy fuerte y justificable únicamente en aquellos ámbitos que poseen sólidas estructuras teóricas como justificación del comportamiento de los procesos. Se puede concluir, entonces, que la preferencia entre estas dos clases de criterios de selección, si bien puede verse condicionada por factores tales como la complejidad de la representación, se basa en razones fundamentalmente subjetivas.

52. Obsérvese que estos criterios consistentes, en virtud de su definición, no utilizan ninguna medida de discrepancia.

criterio de selección, $W(p_1,p_2)$ (que depende de las distribuciones de probabilidades de las sumas de los cuadrados de los residuos y de sus diferencias). A partir de estos momentos de la variable diferencia es posible calcular el **cociente señal-ruido del criterio de selección** como:

$$SW\left[CSM(p_1,p_2)\right] = \frac{S\left[CSM(p_1,p_2)\right]}{W\left[CSM(p_1,p_2)\right]} = \frac{E\left[CSM(p_2) - CSM(p_1)\right]}{\sigma\left[CSM(p_2) - CSM(p_1)\right]} \qquad (p_2 > p_1)$$

Si se verifica que la función de penalización por el aumento de la complejidad de un modelo es más débil que el ruido correspondiente, entonces al criterio de selección le corresponderá una señal débil, un cociente SW débil (decir que un coeficiente SW es débil significa decir que asumirá valores positivos próximos a cero o negativos ($SW < 0.5$)) y estará afectado por un sesgo hacia la sobre-estimación. Por el contrario, cuando la función de penalización es más fuerte que el ruido, el cociente SW asume valores positivos marcadamente significativos ($SW > 2$). No obstante, debe tenerse en cuenta que en este caso, si la función de penalización es demasiado fuerte, el cociente SW se volverá débil y el criterio de selección estará afectado por un sesgo hacia la subestimación.

Para series cronológicas de dimensión $n \leq 100$, se puede asegurar, en general, que cuanto más fuerte sea el cociente SW menor será la probabilidad de selección de un orden de autorregresividad equivocado y, viceversa, si el cociente SW es poco significativo, en particular, si es negativo.

2.10.5.2.- Las medidas de distancia

Como se mencionó en la sección precedente: **i)** si la verdadera representación pertenece al conjunto de las representaciones candidatas, la forma natural de comparar los métodos de selección consistentes es a partir de la comparación de sus probabilidades de elegir la verdadera representación y **ii)** si la verdadera representación no pertenece al conjunto de las representaciones candidatas, el método de selección consiste en minimizar la distancia entre cada una de las representaciones candidatas y aquélla que más se aproxime a la verdadera. Esta distancia está definida por una función (o métrica) $D(u,v)$ (donde u y v pueden denotar escalares o vectores) que debe satisfacer las siguientes condiciones: **i)** $D(u,v) > 0$ (para $u \neq v$) y $D(u,v) = 0$ (para $u = v$); **ii) de simetría**, $D(u,v) = D(v,u)$ y **iii) de desigualdad triangular**, $D(u,w) \leq D(u,v) + D(v,w)$. En lo sucesivo, a los efectos de la aplicación de los criterios de selección, se considerará solamente la primera de estas condiciones.

Del conjunto de funciones de distancia que satisfacen esta condición[53] se tomarán en consideración solamente las siguientes: **1)** la **distancia dirigida de Kullback-Leibler (KL)**[54] basada en el cociente de verosimilitudes, y definida como:

$$KL\left(\hat{Y}_n^*, \hat{Y}_m\right) = E\left\{\left[\log\left(\frac{f_n^*(y)}{f_m(y)}\right)\right] / f_n^*(y)\right\}$$

---

[53]. Para un análisis de las propiedades que debe reunir la familia de funciones que cumple la condición **i)**, ver Linhart, H.; Zucchini, W. (1986).

[54]. Esta función, conocida también como **coeficiente de información de Kullback-Leibler** o **discrepancia de Kullback-Leibler**, es la medida de distancia más importante en la metodología de selección de modelos. Originalmente fue definida para medir la distancia entre dos distribuciones de probabilidades, $p = \{p_1, p_2, \ldots, p_n\}$ y $q = \{q_1, q_2, \ldots, q_n\}$, como $D_{KL}(p,q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right)$ (ver Kullback, S.; Leibler, R.A. (1951)) y constituye un caso particular de la familia de los **principios directos de optimización de la entropía**. El concepto de entropía está relacionado aquí con el concepto de incertidumbre, en particular con la incertidumbre asociada a las probabilidades de las posibles realizaciones de un fenómeno de resultado eventual. Obviamente, la cantidad de incertidumbre contenida en una distribución de probabilidades es función de la cantidad de información con que cuenta el observador, de modo que cualquier incremento en el conjunto de información producirá una disminución de su incertidumbre.

(donde: **i)** $\hat{Y}_n^*$ denota el verdadero modelo, con función de densidad $f_n^*(\bullet)$ y **ii)** $\hat{Y}_m$ denota el modelo candidato, con función de densidad $f_m(\bullet)$); **2)** la **distancia** L_2 definida como:

$$L_2 = \frac{1}{n-p}\sum_{t=p+1}^{n}\left[\hat{Y}_t - m\left(Y_t^*\right)\right]^2$$

(donde: **i)** $\hat{Y}_t$ denota el valor de Y_t estimado a partir de la aplicación del modelo candidato y **ii)** $m\left(Y_t^*\right)(t = 1,2,...,n)$ denota la estructura temporal del verdadero modelo, $Y_t = m\left(Y_t^*\right) + \varepsilon_t$), siendo $\left\{\varepsilon_t\right\}: WN - N\left(0,\sigma_\varepsilon^2\right)$ (debe tenerse en cuenta que, a fin de permitir la comparación de distintos modelos candidatos que pueden haber sido estimados a partir de series de tamaños diferentes, la medida L_2 está normalizada por $n-p$, valor determinado por el modelo candidato).

La distancia L_2 es particularmente apropiada para aplicar a procesos autorregresivos, además posee la ventaja de depender exclusivamente de los valores medios de las dos distribuciones y no de sus funciones de densidad, por lo cual puede ser aplicada aún cuando las perturbaciones ε_t no estén Normalmente distribuidas.

Supóngase: **i)** que $\hat{Y}_n^p$ sea el modelo $AR(p)$ más aproximado al verdadero en términos de L_2 y **ii)** que, aplicado un criterio de selección de modelos, el modelo elegido sea un $AR(p_1)$, $\hat{Y}_m^{p_1}$. Obviamente se verificará que $L_2\left(\hat{Y}^{p_1}\right) \geq L_2\left(\hat{Y}^{p}\right)$. De modo que se puede definir una **medida de la eficiencia observada de un criterio de selección en términos de** L_2 de la siguiente forma:

$$EO(L_2) = \frac{L_2\left(\hat{Y}^{p}\right)}{L_2\left(\hat{Y}^{p_1}\right)}$$

$(0 < EO(L_2) \leq 1)$.

Por su parte, Shibata, R. (1980) propuso utilizar como medida alternativa de eficiencia el valor esperado de L_2, condicionado por el supuesto que $\hat{Y}_n^*$ es el verdadero modelo con función de densidad $f_n^*(y)$ ($E_{f_n^*}(L_2)$). Supóngase: **i)** que $\hat{Y}_n^p$ sea el modelo $AR(p)$ más aproximado al verdadero en términos de la esperanza matemática condicionada de L_2 ($E_{f_n^*}\left[L_2\left(\hat{Y}^{p}\right)\right]$) y **ii)** que, aplicado un criterio de selección de modelos, el modelo elegido sea un $AR(p_1)$, $\hat{Y}_m^{p_1}$. Obviamente se verificará que:

$$E_{f_n^*}\left[L_2\left(\hat{Y}^{p_1}\right)\right] \geq E_{f_n^*}\left[L_2\left(\hat{Y}^{p}\right)\right]$$

De modo que se puede definir una **medida de la eficiencia esperada de un criterio de selección en términos de** L_2 de la siguiente forma:

$$EE(L_2) = \frac{E_{f_n^*}\left[L_2\left(\hat{Y}^{p}\right)\right]}{E_{f_n^*}\left[L_2\left(\hat{Y}^{p_1}\right)\right]}$$

$(0 < EE(L_2) \leq 1)$.

Como corolario de esta definición, se dice que un criterio de selección es **asintóticamente eficiente en términos de** L_2 si se verifica que el cociente de esperanzas matemáticas cumple con la ley fuerte de los grandes números, es decir, si $p\left\{\lim_{n\to\infty}\left(\frac{E_{f_n^*}\left[L_2\left(\hat{Y}^p\right)\right]}{E_{f_n^*}\left[L_2\left(\hat{Y}^{p_1}\right)\right]}-1\right)=0\right\}=1$.

De la misma forma, se pueden definir: **i)** una **medida de eficiencia observada en términos de la distancia de Kullback-Leibler** de la forma:

$$EO(KL)=\frac{KL\left(\hat{Y}^p\right)}{KL\left(\hat{Y}^{p_1}\right)}$$

(donde $\hat{Y}_n^p$ denota el modelo AR más aproximado al verdadero en términos de la distancia de Kullback-Leibler e $\hat{Y}_m^{p_1}$ denota un modelo $AR(p_1)$, elegido a partir de la aplicación de un criterio de selección dado) y **ii)** una **medida de eficiencia esperada en términos de la distancia de Kullback-Leibler** de la forma:

$$EE(KL)=\frac{E_{f_n^*}\left[KL\left(\hat{Y}^p\right)\right]}{E_{f_n^*}\left[KL\left(\hat{Y}^{p_1}\right)\right]}$$

(donde $E_{f^*}\left[KL(\bullet)\right]$ denota el valor esperado de la distancia de Kullback-Leibler, condicionado por el supuesto que $\hat{Y}_n^p$ es el verdadero modelo con función de densidad $f_n^*(y)$).

Con un criterio similar al anterior, se dice que un criterio de selección es **asintóticamente eficiente en términos de la distancia de Kullback-Leibler** si se verifica que:

$$p\left\{\lim_{n\to\infty}\left(\frac{E_{f_n^*}\left[KL\left(\hat{Y}^p\right)\right]}{E_{f_n^*}\left[KL\left(\hat{Y}^{p_1}\right)\right]}-1\right)=0\right\}=1$$

2.10.5.3.- Criterios de selección eficientes basados en la medida de Kullback-Leibler

Bajo el supuesto de Normalidad del verdadero modelo, su función de densidad asume la forma:

$$f\left(\hat{Y}^*\right)=\frac{1}{\left(\sigma_\varepsilon^*\sqrt{2\pi}\right)^{n-p}}\exp\left\{\frac{1}{2\sigma_\varepsilon^{*2}}\sum_{t=p+1}^{n}\left[Y_t-m\left(Y_t^*\right)\right]^2\right\}$$

y la función de verosimilitud del modelo candidato ($\hat{Y}_t^p$) será de la forma:

$$f\left(\hat{Y}^p\right)=\frac{1}{\left(\sigma_{p\varepsilon}\sqrt{2\pi}\right)^{n-p}}\exp\left[\frac{1}{2\sigma_{p\varepsilon}^2}\sum_{t=p+1}^{n}\left(Y_t-\phi_1Y_{t-1}-\ldots-\phi_pY_{t-p}\right)^2\right]$$

(obsérvese que el tamaño de la muestra está determinado por el orden, p, del modelo candidato), por lo tanto, la distancia de Kullback-Leibler asumirá la forma:

$$KL = \frac{2}{n-p} E\left[\ln\left(\frac{f\left(\hat{Y}^*\right)}{f\left(\hat{Y}^p\right)} \right) / f\left(\hat{Y}^*\right) \right] =$$

$$= \ln\left(\frac{\sigma_{p\varepsilon}^2}{\sigma_\varepsilon^{*2}} \right) + E\left\{ \left[-\frac{1}{2\sigma_\varepsilon^{*2}} \sum_{t=p+1}^{n} \left(Y_t - m\left(Y_t^*\right) \right)^2 - \frac{1}{2\sigma_{p\varepsilon}^2} \sum_{t=p+1}^{n} \left(Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} \right)^2 \right] / f\left(Y^*\right) \right\} =$$

$$= \ln\left(\frac{\sigma_{p\varepsilon}^2}{\sigma_\varepsilon^{*2}} \right) + \frac{\sigma_\varepsilon^{*2}}{\sigma_{p\varepsilon}^2} + \frac{1}{\sigma_{p\varepsilon}^2} \frac{1}{n-p} \sum_{t=p+1}^{n} \left[m\left(\hat{Y}_t^*\right) - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} \right]^2 - 1$$

De modo que, sustituyendo en esta expresión los coeficientes ϕ_j, σ_ε^{*2} y $\sigma_{p\varepsilon}^2$ por sus estimadores máximo-verosímiles, se obtiene que:

$$KL = \ln\left(\frac{\hat{\sigma}_{p\varepsilon}^{(MV)2}}{\hat{\sigma}_\varepsilon^{*(MV)2}} \right) + \frac{\hat{\sigma}_\varepsilon^{*(MV)2}}{\sigma_{p\varepsilon}^{(MV)2}} + \frac{L_2}{\hat{\sigma}_{p\varepsilon}^{(MV)2}} - 1$$

A partir de esta definición se consideran los siguientes criterios de selección:

1) AIC

El primer criterio de selección de modelos basado en la función de información de Kullback-Leibler fue el **AIC**[55], propuesto por Akaike, H. (1973):

$$AIC(p) = -2\ln\left(verosimilitud\left(\hat{Y}_t^p\right) \right) + 2\left(numero\ de\ coeficientes\ del\ \mathrm{mod}elo\ \hat{Y}_t^p \right)$$

(donde $\hat{Y}_t^p$ denota el modelo candidato). Suponiendo la Normalidad de $\hat{Y}_t^p$, se obtiene que:

$$-2\ln\left(verosimilitud\left(\hat{Y}_t^p\right) \right) = (n-p)\ln(2\pi) + (n-p)\ln\left(\sigma_{p\varepsilon}^{(MV)2}\right) + \frac{1}{\sigma_{p\varepsilon}^2} \sum_{t=p+1}^{n} \left(Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} \right)^2$$

de modo que se puede escribir:

$$AIC(p) = (n-p)\ln(2\pi) + (n-p)\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right) + (n-p) + 2(p+1) =$$

$$= (n-p)\left[\ln(2\pi) + \ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right) + 1 \right] + 2(p+1)$$

Dado que el término $\ln(2\pi)+1$ no cumple ningún papel en la selección de los modelos, la expresión anterior puede ser resumida de la siguiente forma:

$$AIC(p) = (n-p)\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right) + 2(p+1)$$

o expresada en forma normalizada:

---

[55]. Abreviatura correspondiente a la denominación **"Akaike information criterion"**.

$$AIC(p) = \ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right) + \frac{2(p+1)}{n-p}$$

donde $\frac{2(p+1)}{n-p}$ denota la función de penalización por el aumento del orden de autorregresividad.

Debe tenerse en cuenta, curiosamente, que este criterio, que es asintóticamente insesgado con respecto a la distancia de Kullback-Leibler fue desarrollado a partir de la hipótesis que el verdadero modelo pertenece al conjunto de modelos candidatos[56].

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ $(p_2 > p_1)$ y sean $\{\hat{\varepsilon}_{1t}\}$ y $\{\hat{\varepsilon}_{2t}\}$ las correspondientes series residuales. Se demuestra que: **i)** $\frac{1}{\sigma_\varepsilon^{*2}}\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - \sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right) : \chi^2_{2(p_2-p_1)}$ (donde σ_ε^{*2} denota la varianza de las perturbaciones del verdadero modelo); **ii)** $\frac{1}{\sigma_\varepsilon^{*2}}\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 : \chi^2_{n-2p_1}$; **iii)** $\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - \sum_{t=p_2+1}^{n}\varepsilon_{2t}^2\right) \perp \sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2$; **iv)** $Q = \frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{2t}^2} : \frac{\chi^2_{n-2p_2}}{\chi^2_{n-2p_2} + \chi^2_{2(p_2-p_1)}}$; **v)** teniendo en cuenta que las distribuciones del numerador y del denominador son independientes, $Q : B\left(\frac{n-2p_2}{2}; p_2 - p_1\right)$ (donde $B(\bullet,\bullet)$ denota una distribución Beta) y **vi)** $\log(Q) : \log\text{-}B\left(\frac{n-2p_2}{2}; p_2 - p_1\right)$ (donde $\log\text{-}B(\bullet,\bullet)$ denota una distribución log-Beta).

Por otra parte, se verifica que:

$$E\left[\left(a\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - b\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right)/\sigma_\varepsilon^{*2}\right] = \left[a(n-2p_1) - b(n-2p_2)\right]\sigma_\varepsilon^{*2}$$

(donde a y b son constantes) y que:

$$\sigma^2\left[\left(a\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - b\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right)/\sigma_\varepsilon^{*2}\right] =$$

$$= \sigma^2\left[\left(a\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - a\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 + a\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 - b\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right)/\sigma_\varepsilon^{*2}\right] =$$

$$= \sigma^2\left\{\left[a\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - \sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right) + (a-b)\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2\right]/\sigma_\varepsilon^{*2}\right\} =$$

$$= 4a^2(p_2 - p_1)\sigma_\varepsilon^{*4} + 2(a-b)^2(n-2p_2)\sigma_\varepsilon^{*4}$$

---

56. Este criterio se basa en una generalización del principio de máxima verosimilitud y consiste en aplicar el principio de entropía cruzada mínima de Kullback (conocido como el **principio minx-ent**), para obtener una estimación de la medida de distancia de Kullback-Leibler (para una análisis detallado de la derivación de este criterio, ver Linhart, H.; Zucchini, W. (1986)).

Asimismo, se demuestra que[57]:

$$E\left[\ln\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right)/\sigma_\varepsilon^{*2}\right]=\ln\left(\sigma_\varepsilon^{*2}\right)+\ln(2)+\psi\left(\frac{n-2p_1}{2}\right)$$

A partir del desarrollo de Taylor de $\ln\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right)$ en el entorno de:

$$E\left[\ln\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right)/\sigma_\varepsilon^{*2}\right]$$

se obtiene una aproximación de la forma:

$$E\left[\ln\left(\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right)/\sigma_\varepsilon^{*2}\right]=\ln\left(\sigma_\varepsilon^{*2}\right)+\ln(n-2p_1)-\frac{1}{n-2p_1}$$

Es decir, resulta que:

$$E\left[\ln\left(\hat{\sigma}_{p_2\varepsilon}^{(MV)2}\right)/\sigma_\varepsilon^{*2}\right]=\ln\left(\sigma_\varepsilon^{*2}\right)+\ln(n-2p_1)-\frac{1}{n-2p_1}-\ln(n-p_1)$$

De modo que la señal correspondiente al criterio $AIC(p)$ queda definida de la siguiente forma:

$$S\left(AIC(p_1,p_2)\right)=E\left\{\left[AIC(p_2)-AIC(p_1)\right]/\sigma_\varepsilon^{*2}\right\}=$$

$$=E\left\{\left[\ln\left(\hat{\sigma}_{p_2\varepsilon}^{(MV)2}\right)+\frac{2(p_2+1)}{n-p_2}\right]-\left[\ln\left(\hat{\sigma}_{p_1\varepsilon}^{(MV)2}\right)+\frac{2(p_1+1)}{n-p_1}\right]\right\}=$$

$$=\ln(n-p_2)-\frac{1}{n-2p_2}-\ln(n-p_2)+\frac{2(p_2+1)}{n-p_2}-\left[\ln(n-2p_1)-\frac{1}{n-2p_1}-\ln(n-p_1)+\frac{2(p_1+1)}{n-p_1}\right]=$$

---

57. Se demuestra que $\int_0^\infty z^{w-1}e^{-vz}\ln(z)dz=\frac{1}{v^w}\Gamma(u)\left[\psi(u)-\ln(v)\right]\ (u>0,v>0)$ (donde: **i)** $\psi(u)=-c-\sum_{j=0}^{\infty}\left(\frac{1}{j+u}-\frac{1}{j+1}\right)=\psi(u-1)+\frac{1}{u-1}\ (u>1)$ denota la función de Euler; **ii)** $\psi\left(\frac{1}{2}\right)=-c-2\ln(2)$; **iii)** $\psi(1)=-c$ y **iv)** $c=0.577215664901$ es la constante de Euler). Para una variable $Z:\chi^2_{(m)}$, se verifica que $E\left[\ln(Z)\right]=\frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)}\int_0^\infty\ln(z)z^{m/2-1}e^{-z/2}dz=\log(2)+\psi\left(\frac{m}{2}\right)$ (ver Gradsztein, S.; Ryzhik, I.M. (1980)).

$$= \ln\left(\frac{n-2p_2}{n-2p_1}\right) - \ln\left(\frac{n-p_2}{n-p_1}\right) - \frac{2(p_2-p_1)}{(n-2p_1)(n-2p_2)} + \frac{2(n+1)(p_2-p_1)}{(n-p_1)(n-p_2)}$$

De acuerdo con lo expresado más arriba:

$$\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right) :\log- B\left(\frac{n-2p_2}{2}, p_2-p_1\right)$$

de modo que $\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right] = \psi'\left(\frac{n-2p_2}{2}\right) - \psi'\left(\frac{n-2p_1}{2}\right)$ y utilizando el desarrollo de Taylor, esta expresión puede ser aproximada de la siguiente forma: $\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right] \approx \frac{4(p_2-p_1)}{(n-2p_2)(n-2p_1+2)}$. Es decir que el ruido del criterio AIC queda definido de la siguiente forma[58]:

$$W\left(AIC(p_1,p_2)\right) = \sigma\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right] = \frac{2\sqrt{p_2-p_1}}{\sqrt{(n-2p_2)(n-2p_1+2)}}$$

De modo que el cociente *SW* correspondiente al criterio AIC, con respecto a la sobre-estimación, puede ser expresado de la siguiente forma:

---

[58]. Se demuestra que $\int_0^1 x^{v-1}(1-x^r)^{u-1}\ln(x)\,dx = \frac{1}{r^2}B\left(\frac{v}{r}+u,u\right)\left[\psi\left(\frac{v}{r}\right) - \psi\left(\frac{v}{r}+u\right)\right]$ y que $\int_0^1 x^{v-1}(1-x^r)^{u-1}\ln^2(x)\,dx = B\left(\frac{v}{r}+u,u\right)\left\{\left[\psi\left(\frac{v}{r}\right) - \psi\left(\frac{v}{r}+u\right)\right]^2 + \psi'(v) - \psi'(u+v)\right\}$ (donde $\psi'(u) = \sum_{j=0}^{\infty}\frac{1}{(j+u)^2}, u>0, v>0$). En particular, para una variable $Q: B\left(\frac{h}{2},\frac{k}{2}\right)$, se verifica que: **i)** $E\left[\ln(Q)\right] = \int_0^1 \ln(x)\,B^{-1}\left(\frac{h}{2},\frac{k}{2}\right)x^{h/2-1}(1-x)^{k/2-1}\,dx = \psi\left(\frac{h}{2}\right) - \psi\left(\frac{h}{2},\frac{k}{2}\right)$; **ii)** $E\left[\ln^2(Q)\right] = \int_0^1 \ln^2(x)\,B^{-1}\left(\frac{h}{2},\frac{k}{2}\right)x^{h/2-1}(1-x)^{k/2-1}\,dx = \left[\psi\left(\frac{h}{2}\right) - \psi\left(\frac{h}{2}+\frac{k}{2}\right)\right]^2 + \psi'\left(\frac{h}{2}\right) - \psi'\left(\frac{h}{2}+\frac{k}{2}\right)$ y, por lo tanto, que $\sigma^2\left[\ln(Q)\right] = \sigma^2\left[\ln(Q)\right] = \psi'\left(\frac{h}{2}\right) - \psi'\left(\frac{h}{2}+\frac{k}{2}\right)$ (ver Gradszteyn, I.S.; Ryzhik, I.M. (1980)).

$$SW\left(AIC\left(p_1,p_2\right)\right)=\frac{1}{2}\sqrt{\frac{\left(n-2p_2\right)\left(n-2p_1+2\right)}{p_2-p_1}}\left\{\ln\left[\frac{\left(n-p_1\right)\left(n-p_2\right)}{\left(n-2p_1\right)\left(n-2p_2\right)}\right]-\right.$$

$$\left.-2\left(p_2-p_1\right)\left[\frac{1}{\left(n-2p_1\right)\left(n-2p_2\right)}-\frac{n+1}{\left(n-p_1\right)\left(n-p_2\right)}\right]\right\}$$

Para valores pequeños de $\left(p_2-p_1\right)$ el coeficiente $SW(AIC)$ crece, pero cuando $p_2\to n$, $SW(AIC)\to-\infty$. Esta circunstancia pone en evidencia el sesgo del criterio AIC hacia la sobre-estimación, para series cortas. Este es un defecto que afecta, no sólo al AIC, sino a todos los criterios de la forma $\ln\left(\hat{\sigma}^2_{p\varepsilon}\right)+\frac{\alpha\, p}{n-p}$ (donde α es una constante), debido a que poseen una función de penalización poco significativa (que genera un cociente SW muy débil lo cual, como ya se mencionó, implica una alta probabilidad de sobre-estimación)[59]. Sugiura, N. (1978) y Hurvich, C.M.; Tsai, C.L. (1989) demostraron que, para evitar esta falencia, la función de penalización debe ser de orden mayor que el lineal con respecto a p_1, de modo que su primera derivada sea positiva y la función de penalización sea no-acotada para $p_1\le n$. Esta condición asegura que la señal del criterio crecerá rápidamente cuando aumente la medida de la sobre-estimación, p_2-p_1.

Por otra parte, se verifica que:

$$\lim_{n\to\infty}SW\left(AIC\left(p_1,p_2\right)\right)=\lim_{n\to\infty}\frac{n}{2\sqrt{p_2-p_1}}\left[\frac{p_2-p_1}{n}-\frac{2\left(p_2-p_1\right)}{n}\right]=\frac{\sqrt{p_2-p_1}}{2}$$

Bajo el supuesto que el modelo más aproximado al verdadero sea $AR\left(p_1\right)$, la probabilidad de que el criterio AIC sobre-estime la selección, es decir, la probabilidad de que se verifique que $AIC\left(AR\left(p_2\right)\right)<AIC\left(AR\left(p_1\right)\right)\left(p_2>p_1\right)$, queda definida de la siguiente forma:

$$p\left[AIC\left(p_2\right)<AIC\left(p_1\right)\right]=$$

$$=p\left[\ln\left(\hat{\sigma}^{(MV)2}_{p_2\varepsilon}\right)+\frac{2\left(p_2+1\right)}{n-p_2}<\ln\left(\sigma^{(MV)2}_{p_1\varepsilon}\right)+\frac{2\left(p_1+1\right)}{n-p_1}\right]=$$

$$=p\left[\ln\left(\sum_{t=p_{2+1}+1}^{n}\varepsilon^2_{2t}\right)-\ln\left(n-p_2\right)+\frac{2\left(p_2+1\right)}{n-p_2}<\log\left(\sum_{t=p_1+1}^{n}\varepsilon^2_{1t}\right)-\log\left(n-p_1\right)+\frac{2\left(p_1+1\right)}{n-p_1}\right]=$$

$$=p\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}^2_{2t}}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}^2_{2t}}\right)<\log\left(\frac{n-p_1}{n-p_2}\right)+\frac{2\left(p_1+1\right)}{\left(n-p_1\right)}-\frac{2\left(p_2+1\right)}{n-p_2}\right]=$$

---

[59]. Por el contrario, un valor de SW marcadamente positivo permite suponer una probabilidad de sobre-estimación próxima a cero.

$$= p\left[\ln\left(\frac{\sum_{t=p_1+1}^{n}\varepsilon_{1t}^2}{\sum_{t=p_2+1}^{n}\varepsilon_{2t}^2}\right) > \log\left(\frac{n-p_1}{n-p_2}\right) + \frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] =$$

$$= p\left\{\frac{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2} - 1 > \frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right\} =$$

$$= p\left\{\frac{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - \sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2} > \frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right\} =$$

$$= p\left\{\frac{\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2p_2}} > \frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right\} =$$

$$= p\left\{\Im_{2(p_2-p_1),n-2p_2} > \frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right)\right\}$$

(obsérvese que la probabilidad de sobre-estimación en criterios como éste, que poseen función de penalización débil, disminuye a medida que n crece). A partir de esta expresión se puede escribir:

$$\frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right) =$$

$$= \frac{n-2p_2}{2(p_2-p_1)}\left\{\frac{n-p_1}{n-p_2}\left[1 + \frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)} + O\left(\frac{1}{n^2}\right)\right] - 1\right\} =$$

$$= \frac{n-2p_2}{2(p_2-p_1)}\left[\frac{p_2-p_1}{n-p_2} + \frac{2(p_2-p_1)(n+1)}{(n-p_2)^2} + O\left(\frac{1}{n^2}\right)\right]$$

De modo que:

$$\lim_{n\to\infty}\frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)(n+1)}{(n-p_1)(n-p_2)}\right] - 1\right) =$$

$$= \frac{n}{2(p_2-p_1)}\left[\frac{p_2-p_1}{n} + \frac{2(p_2-p_1)}{n}\right] = 1.5$$

Asimismo, de acuerdo con el teorema de Slutsky, se verifica que $\lim_{n\to\infty}\Im_{2(p_2-p_1),n-2p_2} = \frac{\chi_{2(p_2-p_1)}}{2(p_2-p_1)}$

Por lo tanto, se puede concluir que la probabilidad asintótica de que, utilizando el criterio AIC, el

verdadero modelo $AR(p_1)$ resulte sobre-estimado por un modelo $AR(p_2)$ ($p_2 > p_1$), será:

$$\lim_{n\to\infty} p\left[AIC(p_2) < AIC(p_1)\right] = p\left[\chi_{2(p_2-p_1)} > 3(p_2 - p_1)\right]$$

2) AIC-corregido

A fin de resolver el problema del sesgo hacia la sobre-estimación del criterio AIC en series cortas, Sugiura, N. (1978) y Hurvich, C.M.; Tsai, C.L. (1989), basándose (como en el caso anterior) en el supuesto que el verdadero modelo $AR(p^*)$ pertenece al conjunto de los modelos candidatos y en la esperanza matemática condicionada de la distancia de Kullback-Leibler propusieron el criterio conocido como **AICc**[60].

$$AICc(p) = E\left[KL / f\left(\hat{Y}^*\right)\right] = E\left\{\left[\ln\left(\frac{\hat{\sigma}_{p\varepsilon}^{(MV)2}}{\sigma_\varepsilon^{*2}}\right) + \frac{\sigma_\varepsilon^{*2}}{\hat{\sigma}_{p\varepsilon}^2} + \frac{L_2}{\hat{\sigma}_{p\varepsilon}^2} - 1\right] / f\left(\hat{Y}^*\right)\right\} =$$

$$= E\left\{\left[\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right)\right] / f\left(\hat{Y}^*\right)\right\} - \ln\left(\sigma_\varepsilon^{*2}\right) + \frac{(n-p)+p}{(n-p)-p-2} - 1$$

(donde p denota el orden de autorregresividad del modelo candidato y es tal que $p > p^*$).

Teniendo en cuenta que $\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right)$ es un estimador insesgado de $E\left\{\left[\ln\left(\sigma_{p\varepsilon}^2\right)\right] / f\left(\hat{Y}^*\right)\right\}$, se puede escribir:

$$AICc(p) = E\left[KL / f\left(\hat{Y}^*\right)\right] = \ln\left(\sigma_{p\varepsilon}^{(MV)2}\right) + \frac{n}{n-2p-2} - \left[\ln\left(\sigma_\varepsilon^{*2}\right) + 1\right]$$

y, dado que el término $-\left[\ln\left(\hat{\sigma}_\varepsilon^{*2}\right) + 1\right]$ no influye en la selección del modelo, se obtiene la siguiente definición:

$$AICc(p) = \ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right) + \frac{n}{n-2(p+1)}$$

Como se verá inmediatamente, Hurvich, C.M.; Tsai, C.L. (1989) demostraron que AICc mejora la estimación de AIC en series cortas, pero que es asintóticamente equivalente AICc en series suficientemente largas.

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ ($p_2 > p_1$) y sean $\{\hat{\varepsilon}_{1t}\}$ y $\{\hat{\varepsilon}_{2t}\}$ las correspondientes series residuales. Utilizando un razonamiento similar al empleado en el criterio AIC se obtiene que:

$$S\left(AICc(p_1, p_2)\right) = E\left\{\left[AICc(p_2) - AICc(p_1)\right] / \sigma_{1\varepsilon}^2\right\} =$$

---

60. El agregado de la letra c a la denominación de los criterios indica que en las respectivas definiciones fueron corregidas las funciones de penalización.

$$= \ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right] - \frac{2(p_2-p_1)}{(n-p_1)(n-2p_2)} + \frac{2(p_2-p_1)}{(n-2p_1-2)(n-2p_2-2)}$$

y que:

$$\sigma^2\left[\log\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right] = \frac{4(p_2-p_1)}{(n-2p_1+2)(n-2p_2)}$$

De modo que el cociente SW con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$SW\big(AIC(p_1,p_2)\big) = \frac{1}{2}\sqrt{\frac{(n-2p_1+2)(n-2p_2)}{p_2-p_1}}\cdot$$

$$\cdot\left\{\ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right] - \frac{2(p_2-p_1)}{(n-p_1)(n-2p_2)} + \frac{2(p_2-p_1)}{(n-2p_1-2)(n-2p_2-2)}\right\}$$

De esta expresión se puede concluir en forma inmediata que, cuando $p_2 - p_1$ crece, el cociente SW también crece, de modo que su comportamiento con respecto a la sobre-estimación es mejor que el del AIC.

Por otra parte, igual que para el criterio AIC, se verifica que:

$$\lim_{n\to\infty} SW\big(AICc(p_1,p_2)\big) = \frac{\sqrt{p_2-p_1}}{2}$$

Bajo el supuesto que el modelo más aproximado al verdadero es $AR(p_1)$, la probabilidad de que, en series cortas, el criterio AICc sobre-estime la selección, es decir, la probabilidad de que se verifique que $AICc\big(AR(p_2)\big) < AICc\big(AR(p_1)\big)\,(p_2 > p_1)$ queda definida de la siguiente forma:

$$p\big[AICc(p_2) < AICc(p_1)\big] =$$

$$= p\left[\ln\left(\sigma_{p_2\varepsilon}^{(MV)2}\right) + \frac{n}{n-2p_2-2} < \ln\left(\sigma_{p_1\varepsilon}^{(MV)2}\right) + \frac{n}{n-2p_1-2}\right] =$$

$$= p\left[\log\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right) < \ln\left(\frac{n-p_2}{n-p_1}\right) + \frac{n}{n-2p_1-2} - \frac{n}{n-2p_2-2}\right] =$$

$$= p\left\{\frac{\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2p_2-2}} > \frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right] - 1\right\} =$$

$$= p\left\{ \mathfrak{I}_{2(p_2-p_1,\,n-2p_2-2)} > \frac{n-2p_2-2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)} \right] - 1 \right) \right\}$$

(las probabilidades de sobre-estimación en criterios como éste, con funciones de penalización fuertes, aumentan a medida que n crece). A partir de la expresión anterior, se puede escribir:

$$\frac{n-2p_2-2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)}{(n-2p_1-2)(n-2p_2-2)} \right] - 1 \right) =$$

$$= \frac{n-2p_2}{2(p_2-p_1)}\left\{ \frac{n-p_1}{n-p_2}\left[1 + \frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)} + O\left(\frac{1}{n^2}\right) \right] - 1 \right\}$$

De modo que:

$$\lim_{n\to\infty} \frac{n-2p_2-2}{2(p_2-p_1)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)} \right] - 1 \right) =$$

$$= \lim_{n\to\infty} \frac{n}{2(p_2-p_1)}\left[\frac{p_2-p_1}{n} + \frac{2(p_2-p_1)}{n} \right] = 1.5$$

Es decir, se puede concluir que la probabilidad asintótica de sobre-estimación para el criterio AICc, igual que para el AIC, será:

$$\lim_{n\to\infty} p\left[AICc(p_2) < AICc(p_1) \right] = p\left[\chi^2_{2(p_2-p_1)} > 3(p_2-p_1) \right]$$

3) AIC-insesgado

Otra forma de corregir el sesgo del criterio AIC hacia la sobre-estimación en series cortas consiste en sustituir: **i)** la estimación máximo-verosímil de $\sigma^2_{p\varepsilon}$ ($\hat{\sigma}^{(MV)2}_{p\varepsilon} = \frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\varepsilon}^2_t$) por su estimación insesgada ($\hat{\sigma}^2_{p\varepsilon} = \frac{1}{n-2p}\sum_{t=p+1}^{n}\hat{\varepsilon}^2_t$) y **ii)** la función de penalización $\frac{2(p+1)}{n-p}$ por $\frac{2(p+1)}{n-p-2}$. De modo que, eliminadas las constantes que no influyen en la selección y sumando 1 (por comodidad), se obtiene el siguiente criterio[61]:

$$AICu(p) = \ln\left(\hat{\sigma}^2_{p\varepsilon}\right) + \frac{n}{n-2(p+1)}$$

debido a Akaike, H. (1969)[62]. Obsérvese que los criterios AICc y AICu son asintóticamente equivalentes.

---

[61]. El agregado de la letra u a la denominación de los criterios indica, que las correspondientes definiciones fueron obtenidas utilizando el estimador insesgado de $\sigma^2_{p\varepsilon}$.

[62]. Como se mencionó al comienzo de la Sec. 2.10.5, éste fue cronológicamente, el primero de los criterios de selección basados en la distancia de Kullback-Leibler.

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ ($p_2 > p_1$). Utilizando un razonamiento similar al de los casos anteriores, se obtiene que:

$$S\left(AICu(p_1,p_2)\right) = E\left\{\left[AICu(p_2) - AICu(p_1)\right] / \sigma^2_{1\varepsilon}\right\} =$$

$$= -\frac{2(p_2 - p_1)}{(n-2p_1)(n-2p_2)} + \frac{2(p_2 - p_1)n}{(n-2p_1-2)(n-2p_2-2)}$$

y que:

$$\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}^2_{1t}}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}^2_{2t}}\right)\right] = \frac{4(p_2 - p_1)}{(n-2p_1+2)(n-2p_2)}$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma[63]:

$$SW\left(AICu(p_1,p_2)\right) = \frac{1}{2}\sqrt{\frac{(n-2p_1+2)(n-2p_2)}{p_2 - p_1}} \cdot$$

$$\cdot\left[-\frac{2(p_2 - p_1)}{(n-2p_1)(n-2p_2)} + \frac{2(p_2 - p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right]$$

Se demuestra que, para $n \geq 5$, $0 < p_2 - p_1 < n - p_2 - 2$ y suponiendo una sobre-estimación del tipo $0 < p_1 \leq p_2 < n - p_2 - 3$, se verifica que, cuando $p_2 - p_1$ crece, $SW\left(AICu(p_1,p_2)\right)$ también crece y es mayor que $SW\left(AIC(p_1,p_2)\right)$ (ver McQuarrie, A.D.R.; Tsai, C.L. (2007)).

Por otra parte, se verifica que:

$$\lim_{n\to\infty} SW\left(AICu(p_1,p_2)\right) = \sqrt{p_2 - p_1}$$

Bajo el supuesto que el modelo más aproximado al verdadero es $AR(p_1)$, la probabilidad de que, en series cortas, el criterio AICu sobre-estime la selección, es decir, la probabilidad de que se verifique que $AICu\left(AR(p_2)\right) < AICu\left(AR(p_1)\right)(p_2 > p_1)$ queda definida de la siguiente forma:

$$p\left[AICu(p_2) < AICu(p_1)\right] =$$

$$= p\left[\ln\left(\hat{\sigma}^2_{p_2\varepsilon}\right) + \frac{n}{n-2p_2-2} < \ln\left(\hat{\sigma}^2_{p_1\varepsilon}\right) + \frac{n}{n-2p_1-2}\right] =$$

---

63. Ver McQuarrie, A.D.R.; Shumway, R.H.; Tsai, C.L. (1997).

$$= p\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right) < \ln\left(\frac{n-2p_2}{n-2p_1}\right) + \frac{n}{n-2p_1-2} - \frac{n}{n-2p_2-2}\right] =$$

$$= p\left\{\frac{\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2p_2}} > \frac{n-2p_1}{n-2p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right] - 1\right\} =$$

$$= p\left\{\Im_{2(p_2-p_1),n-2p_2} > \frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-2p_1}{n-2p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right] - 1\right)\right\}$$

(obsérvese que, igual que en el AICc, la probabilidad de sobre-estimación aumenta a medida que n crece). A partir de la expresión anterior se puede escribir:

$$\frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-2p_1}{n-2p_2}\exp\left[\frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right] - 1\right) =$$

$$= \frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-2p_1}{n-2p_2}\left[1 + \frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)} + O\left(\frac{1}{n^2}\right)\right] - 1\right)$$

De modo que:

$$\lim_{n\to\infty}\frac{n-2p_2}{2(p_2-p_1)}\left(\frac{n-2p_1}{n-2p_2}\left[1 + \frac{2(p_2-p_1)n}{(n-2p_1-2)(n-2p_2-2)}\right] - 1\right) =$$

$$= \lim_{n\to\infty}\frac{n}{2(p_2-p_1)}\left[\frac{2(p_2-p_1)}{n} + \frac{2(p_2-p_1)}{n}\right] = 2$$

es decir, se puede concluir que la probabilidad asintótica de sobre-estimación para el criterio AICu, será:

$$\lim_{n\to\infty} p\left[AICu(p_2) < AICu(p_1)\right] = p\left[\chi^2_{2(p_2-p_1)} > 4(p_2-p_1)\right]$$

2.10.5.4.- Criterios de selección eficientes basados en la medida L_2

1) FPE[64]

Este criterio propuesto por Akaike, H. (1969)(1970)(1973)[65] fue el primero en basarse en la

---

[64]. Abreviatura correspondiente a la denominación **"final prediction error"**.

[65]. En realidad este criterio fue propuesto originalmente por Davisson, L.D. (1965) para el análisis de datos del tipo señal-más-ruido. La atribución de este resultado a Akaike probablemente se deba a que sus publicaciones, realizadas en el ámbito de la teoría de las series cronológicas, fueron más conocidas.

minimización de la varianza del error de predicción[66].

Sean: **i)** un proceso $\{Y_t\}$ estacionario y tal que $m(Y) = 0$, respecto del cual se cuenta con una serie cronológica $\{Y_1, Y_2, ..., Y_n\}$ que permite construir un modelo $AR(p)$ de la forma:

$$\hat{Y}_t = \hat{\phi}_{Y1} Y_{t-1} + \hat{\phi}_{Y2} Y_{t-2} + ... + \hat{\phi}_{Yp} Y_{t-p}$$

(donde los $\hat{\phi}_{Yj}$ $(j = 1,2,...,p)$ denotan los estimadores máximo-verosímiles de los correspondientes coeficientes); **ii)** una serie cronológica $\{X_1, X_2, ..., X_n\}$ obtenida como una nueva realización, independiente de la anterior, del mismo proceso $\{Y_t\}$ y **iii)** $\hat{X}_{n+1}$ el estimador de la variable X_{n+1}, obtenido de la aplicación del modelo calculado a partir de la serie $\{Y_t, t = 1,2,...,n\}$:

$$\hat{X}_{n+1} = \hat{\phi}_{Y1} X_n + \hat{\phi}_{Y2} X_{n-2} + ... + \hat{\phi}_{Yp} X_{n-p}$$

Akaike, H. (1969) demostró que el error medio cuadrático de esta predicción, para n suficientemente grande, es de la forma:

$$E\left[\left(X_{n+1} - \hat{\phi}_{Y1} X_n - \hat{\phi}_{Y2} X_{n-2} - ... - \hat{\phi}_{Yp} X_{n-p+1}\right)^2\right] =$$

$$= E\left\{\left[\left(X_{n+1} - \hat{\phi}_{Y1} X_n - ... - \hat{\phi}_{Yp} X_{n-p+1}\right) - \left(\hat{\phi}_{Y1} - \phi_{Y1}\right) X_n - ... - \left(\hat{\phi}_{Yp} - \phi_{Yp}\right) X_{n-p+1}\right]^2\right\} =$$

$$= \sigma_\varepsilon^2 + E\left\{\left[\left(\hat{\phi}_{Y1} - \phi_{Y1}\right) \quad ... \quad \left(\hat{\phi}_{Yp} - \phi_{Yp}\right)\right] \begin{bmatrix} Y_n Y_n & Y_n Y_{n-1} & ... & Y_n Y_{n+1-p} \\ ... & ... & ... & ... \\ Y_{n+1-p} Y_n & Y_{n+1-p} Y_{n-1} & ... & Y_{n+1-p} Y_{n+1-p} \end{bmatrix} \begin{bmatrix} \hat{\phi}_{Y1} - \phi_{Y1} \\ ... \\ \hat{\phi}_{Yp} - \phi_{Yp} \end{bmatrix}\right\} =$$

$$= \sigma_\varepsilon^2 + E\left[\left(\hat{\phi}_Y - \phi_Y\right)'(YY)\left(\hat{\phi}_Y - \phi_Y\right)\right]$$

Expresando el segundo término como la esperanza matemática del valor esperado condicionado por $\{Y_1, Y_2, ..., Y_n\}$ y teniendo en cuenta que $\{Y_t\} \perp \{X_t\}$, se puede escribir:

$$E\left[\left(X_{n+1} - \hat{X}_{n+1}\right)^2\right] = E\left[\left(X_{n+1} - \hat{\phi}_{Y1} X_n - \hat{\phi}_{Y2} X_{n-1} - ... - \hat{\phi}_{Yp} X_{n-p+1}\right)^2\right] =$$

$$= \sigma_\varepsilon^2 + E\left[\left(\hat{\phi}_Y - \phi_Y\right)' \Gamma_p \left(\hat{\phi}_Y - \phi_Y\right)\right]$$

Bajo el supuesto que $\sqrt{n}\left(\hat{\phi}_Y - \phi_Y\right) \xrightarrow{d} N\left(0, \sigma_\varepsilon^2 \Gamma_p^{-1}\right)$, para n suficientemente grande, se obtiene la aproximación $E\left[\left(X_{n+1} - \hat{X}_{n+1}\right)^2\right] \approx \sigma_\varepsilon^2\left(1 + \frac{p}{n-p}\right)$. Suponiendo que el verdadero modelo sea un $AR(p^*)$ y que $p > p^*$, la esperanza matemática del estimador máximo-verosímil del error es de la forma:

$$E\left(\hat{\sigma}_\varepsilon^{(MV)2}\right) = \left(1 - \frac{p}{n-p}\right)\sigma_\varepsilon^2$$

---

66. Originalmente este criterio fue diseñado por Akaike para ser aplicado exclusivamente a modelos autorregresivos

De modo que $\dfrac{\hat{\sigma}_{\varepsilon}^{(MV)2}}{1-\dfrac{p}{n-p}}$ es un estimador insesgado de σ_{ε}^2. Sustituyendo en la expresión original σ_{ε}^2 por este estimador, se obtiene un estimador insesgado del error medio cuadrático de la predicción:

$$FPE(p) = \hat{\sigma}_{p\varepsilon}^{(MV)2}\frac{n}{n-2p}$$

Como ya se mencionó, se concluye en forma inmediata que el mejor modelo en términos de predicción es el que minimiza esta expresión.

Obsérvese que $\ln[FPE(p)] - AIC(p) = o\left(\dfrac{1}{n^2}\right)$, por lo que se puede concluir que ambos criterios son asintóticamente equivalentes.

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ ($p_2 > p_1$). Entonces, se obtiene que:

$$\begin{aligned}
S(FPE(p_1,p_2)) &= E\left\{\left[FPE(p_2) - FPE(p_1)\right]/\sigma_{\varepsilon}^{*2}\right\} = \\
&= E\left\{\left(\left[\frac{n}{n-2p_2}\hat{\sigma}_{p_2\varepsilon}^2\right] - \left[\frac{n}{n-2p_1}\hat{\sigma}_{p_1\varepsilon}^2\right]\right)/\sigma_{\varepsilon}^{*2}\right\} = \\
&= E\left[\left(\frac{n}{(n-p_2)(n-2p_2)}\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 - \frac{n}{(n-p_1)(n-2p_1)}\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right)/\sigma_{\varepsilon}^{*2}\right] = \\
&= \left[\frac{n}{(n-p_2)(n-2p_2)}(n-2p_2) - \frac{n}{(n-p_1)(n-2p_1)}\right]\sigma_{\varepsilon}^{*2} = \\
&= \frac{(p_2-p_1)n}{(n-p_1)(n-p2)}\sigma_{\varepsilon}^{*2}
\end{aligned}$$

(donde $\sigma_{\varepsilon}^{*2}$ denota la varianza de las perturbaciones del verdadero modelo) y que:

$$\begin{aligned}
&\sigma^2\left\{\left[\frac{n}{(n-p_2)(n-2p_2)}\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 - \frac{n}{(n-p_1)(n-2p_1)}\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right]/\sigma_{\varepsilon}^{*2}\right\} = \\
&= \frac{4n^2(p_2-p_1)}{(n-p_2)^2(n-2p_2)^2}\sigma_{\varepsilon}^{*4} + 2\left[\frac{n}{(n-p_2)(n-2p_2)} - \frac{n}{(n-p_1)(n-2p_1)}\right]^2(n-2p_2)\sigma_{\varepsilon}^{*4} = \\
&= \frac{4(p_2-p_1)n^2(n-2p_2)(n-p_2)^2 + 2(p_2-p_1)^2\left[3n^2 - 4np_1 + 2(p_2-p_1)^2\right]}{(n-2p_2)(n-p_2)^2(n-2p_1)^2(n-p_1)^2}\sigma_{\varepsilon}^{*4}
\end{aligned}$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$SW\left(FPE\left(p_1,p_2\right)\right)=\frac{n\left(n-2p_1\right)\left(p_2-p_1\right)\sqrt{n-2p_2}}{\sqrt{4\left(p_2-p_1\right)n^2\left(n-2p_2\right)\left(n-p_2\right)^2+2\left(p_2-p_1\right)^2\left[3n^2-4np_1+2\left(p_2-p_1\right)^2\right]}}$$

De esta expresión se puede concluir: **i)** que, para valores pequeños de p_2-p_1, este cociente tiende a cero; **ii)** que, en general, para series cortas, $SW(FPE)$ es débil y iii) que $\lim_{n\to\infty} FPE\left(p_1,p_2\right)=\frac{1}{2}\sqrt{p_2-p_1}$.

Suponiendo que $AR\left(p_1\right)$ sea el que más se aproxima al verdadero modelo, la probabilidad de que, en series cortas, el criterio FPE sobre-estime la selección, es decir, la probabilidad de que se verifique que $FPE\left(AR\left(p_2\right)\right)<FPE\left(AR\left(p_1\right)\right)\left(p_2>p_1\right)$ queda definida de la siguiente forma:

$$p\left[FPE\left(p_2\right)<FPE\left(p_1\right)\right]=p\left(\frac{n}{n-2p_2}\hat{\sigma}^2_{2\varepsilon}<\frac{n}{n-2p_1}\hat{\sigma}^2_{1\varepsilon}\right)=$$

$$=p\left[\frac{n}{\left(n-p_2\right)\left(n-2p_2\right)}\sum_{t=p_2+1}^{n}\hat{\varepsilon}^2_{2t}<\frac{n}{\left(n-p_1\right)\left(n-2p_1\right)}\sum_{t=p_1+1}^{n}\hat{\varepsilon}^2_{1t}\right]=$$

$$=p\left[\frac{\chi^2_{2\left(p_2-p_1\right)}}{\chi^2_{n-2p_2}}>\frac{\left(n-p_1\right)\left(n-2p_1\right)}{\left(n-p_2\right)\left(n-2p_2\right)}-1\right]=$$

$$=p\left\{\Im_{2\left(p_2-p_1\right),n-2p_2}>\frac{n-2p_2}{2\left(p_2-p_1\right)}\left[\frac{\left(n-p_1\right)\left(n-2p_1\right)}{\left(n-p_2\right)\left(n-2p_2\right)}-1\right]\right\}=$$

$$=p\left[\Im_{2\left(p_2-p_1\right),n-2p_2}>1+\frac{n-2p_1}{2\left(n-p_1\right)}\right]$$

(obsérvese que, en este caso, la probabilidad de sobre-estimación decrece a medida que n crece). A partir de la expresión anterior se puede concluir que $\lim_{n\to\infty}\left[1+\frac{n-2p_1}{2\left(n-p_1\right)}\right]=1.5$. Es decir que la probabilidad asintótica de sobre-estimación será de la forma[67]:

$$\lim_{n\to\infty} p\left[FPE\left(p_2\right)<FPE\left(p_1\right)\right]=p\left[\chi^2_{2\left(p_2-p_1\right)}>3\left(p_2-p_1\right)\right]$$

2) FPE-insesgado

Si en la definición del FPE se reemplaza la estimación máximo-verosímil de σ^2_ε por su estimación

---

[67]. Bhansali, R.J.; Downham, D.Y. (1977) analizaron la función de penalización del criterio FPE con respecto a su probabilidad de sobre-estimación. Denotando por $FPE_\alpha=\frac{n}{n-\alpha p}$ y basándose en métodos de simulación, propusieron como más adecuado el criterio FPE_4. Otros autores demostraron que para $1.5\le\alpha\le 5$, se obtienen resultados aceptables.

insesgada, $\hat{\sigma}_\varepsilon^2 = \dfrac{1}{n-2p}\sum_{t=p+1}^{n}\varepsilon_t^2$, se obtiene el criterio:

$$FPEu(p) = \hat{\sigma}_\varepsilon^2 \frac{n}{n-2p}$$

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ ($p_2 > p_1$). Se obtiene, entonces, que la correspondiente señal es de la forma:

$$\begin{aligned}
S\left(FPEu(p_1,p_2)\right) &= E\left\{\left[FPEu(p_2) - FPEu(p_1)\right]/\sigma_\varepsilon^{*2}\right\} = \\
&= E\left\{\left[\frac{n}{n-2p_2}\sigma_{2\varepsilon}^2 - \frac{n}{n-2p_1}\sigma_{1\varepsilon}^2\right]/\sigma_\varepsilon^{*2}\right\} = \\
&= E\left\{\left[\frac{n}{(n-2p_2)^2}\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 - \frac{n}{(n-2p_1)^2}\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right]/\sigma_\varepsilon^{*2}\right\} = \\
&= \frac{2(p_2-p_1)n}{(n-2p_1)(n-2p_2)}\sigma_\varepsilon^{*2}
\end{aligned}$$

(donde σ_ε^{*2} denota la varianza de las perturbaciones de la verdadera representación). También se obtiene que:

$$\begin{aligned}
&\sigma^2\left\{\left[\frac{n}{(n-2p_2)^2}\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2 - \frac{n}{(n-2p_1)^2}\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2\right]/\sigma_\varepsilon^{*2}\right\} = \\
&= \frac{4(p_2-p_1)n^2(n-2p_2)^3 + 32n^2(p_2-p_1)^2(n-p_1-p_2)^2}{(n-2p_1)^4(n-2p_2)^3}\sigma_\varepsilon^{*2}
\end{aligned}$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$SW\left(FPEu(p_1,p_2)\right) = \frac{(n-2p_1)\sqrt{(p_2-p_1)(n-p_2)}}{\sqrt{(n-2p_2)^3 + 8(p_2-p_1)(n-p_1-p_2)^2}}$$

De esta expresión se puede concluir que, contrariamente a lo que ocurre con el criterio AICu, cuando $p_2 - p_1$ crece, $SW\left(FPEu(p_1,p_2)\right)$ no es estrictamente creciente. Se verifica que:

$$\lim_{(p_2-p_1)\to(n-p_1)}(n-2p_1)\sqrt{(p_2-p_1)(n-p_2)} = 0$$

y

$$\lim_{(p_2-p_1)\to(n-p_1)}\sqrt{(n-2p_2)^2 + 8(p_2-p_1)(n-p_1-p_2)} > 0$$

De modo que $\lim_{(p_2-p_1)\to(n-p_1)} SW\left(FPEu(p_1,p_2)\right) = 0$. Lo cual, si bien implica que existe un sesgo hacia la sobre-estimación, permite demostrar[68] que, para $n \geq 5$ y $0 < p_2 - p_1 < n - p_1 - 2$, y para el caso de sobre-estimación en el cual $0 < p^* < p_1 < n-3$, se verifica que $SW(FPEu) \geq SW(FPE)$. Se obtiene, además, que:

$$\lim_{n\to\infty} FPEu(p_1,p_2) = \sqrt{p_2 - p_1}$$

Suponiendo que $AR(p_1)$ sea el que más se aproxima al verdadero modelo, la probabilidad de que, para series cortas, el criterio FPEu sobre-estime la selección, es decir, la probabilidad de que se verifique que $FPEu\left(AR(p_2)\right) < FPEu\left(AR(p_1)\right)(p_2 > p_1)$, queda definida de la siguiente forma:

$$p\left[FPEu(p_2) < FPEu(p_1)\right] = p\left[\frac{n}{n-2p_2}\sigma^2_{2\varepsilon} < \frac{n}{n-2p_1}\sigma^2_{1\varepsilon}\right] =$$

$$= p\left[\frac{n}{(n-2p_2)^2}\sum_{t=p_2+1}^{n}\hat{\varepsilon}^2_{2t} < \frac{n}{(n-2p_1)^2}\sum_{t=p_1+1}^{n}\hat{\varepsilon}^2_{1t}\right] =$$

$$= p\left[\frac{\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2p_2}} > \frac{(n-2p_1)^2}{(n-2p_2)^2} - 1\right] =$$

$$= p\left\{\Im_{2(p_2-p_1),n-2p_2} > \frac{n-2p_2}{2(p_2-p_1)}\left[\frac{(n-2p_1)^2}{(n-2p_2^2)^2} - 1\right]\right\} =$$

$$= p\left[\Im_{2(p_2-p_1),n-2p_2} > 1 + \frac{n-2p_1}{n-2p_2}\right]$$

Teniendo en cuenta que $\lim_{n\to\infty}\left(1 + \frac{n-2p_1}{n-2p_2}\right) = 2$, de la expresión anterior se obtiene que la probabilidad asintótica de sobre-estimación será de la forma:

$$\lim_{n\to\infty} p\left[FPEu(p_2) < FPEu(p_1)\right] = p\left[\chi^2_{2(p_2-p_1)} > 4(p_2 - p_1)\right]$$

Se puede concluir, entonces, que, en series largas, se debe esperar que FPEu produzca menos sobre-estimaciones que FPE[69].

3) Cp

Mallows, C.L. (1973) adoptó una aproximación diferente a los criterios basados en la distancia

---

[68]. Ver McQuarrie, A.D.R.; Tsai, C.L. (2007).

[69]. Ver Howrey, E.P. (en Zarembka, P. (ed.) (1974)).

L_2[70]. Sea la función:

$$J_p = \frac{1}{\sigma_\varepsilon^{*2}} \sum_{t=p+1}^{n} \left[\left(\hat{\phi}_1 Y_{t-1} + ... + \hat{\phi}_p Y_{t-p} \right) - \left(\phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} \right) \right]^2 = \frac{1}{\sigma_\varepsilon^{*2}} \sum_{t=p+1}^{n} \left(\hat{Y}_t - Y_t \right)^2 = \frac{n}{\sigma_\varepsilon^{*2}} L_2$$

(donde σ_ε^{*2} denota la varianza de la verdadera representación). Se demuestra que $E(J_p) = V_p + \frac{B_p}{\sigma_\varepsilon^{*2}}$ (donde $V_p = p$ denota la varianza y B_p denota el sesgo)[71].

Teniendo en cuenta que $E\left(\sum_{t=p+1}^{n} \hat{\varepsilon}_t^2 \right) = (n-2p)\sigma_\varepsilon^{*2} + B_p$ y que $E\left[\frac{1}{\sigma_\varepsilon^{*2}} \sum_{t=p+1}^{n} \hat{\varepsilon}_t^2 - (n-3p) \right] =$

$= p + \frac{B_p}{\sigma_\varepsilon^{*2}} = V_p + \frac{B_p}{\sigma_\varepsilon^{*2}} = E(J_p)$, se puede concluir que $\frac{1}{\sigma_\varepsilon^{*2}} \sum_{t=p+1}^{n} \hat{\varepsilon}_t^2 - (n-3p)$ es un estimador insesgado de $E(J_p)$. Sustituyendo en esta expresión σ_ε^{*2} por su estimador insesgado (calculado a partir del modelo candidato de mayor orden de autorregresividad), se obtiene el criterio de selección[72]:

$$Cp = \frac{1}{\sigma_{p\varepsilon}^2} \sum_{t=p+1}^{n} \hat{\varepsilon}_t^2 - (n-3p)$$

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ ($p_2 > p_1$) y un modelo $AR(P)$, que es el de mayor orden de autorregresividad entre los posibles modelos candidatos, se obtiene, entonces, que la correspondiente señal es de la forma:

$$S(Cp(p_1, p_2)) = E\left\{ \left[Cp(p_2) - Cp(p_1) \right] / \sigma_\varepsilon^2 \right\} =$$

$$= E\left\{ \left(\left[(n-2P) \frac{\sum_{t=p_2+1}^{n} \hat{\varepsilon}_{2t}^2}{\sum_{t=P+1}^{n} \hat{\varepsilon}_{Pt}^2} - n + 3p_2 \right] - \left[(n-2P) \frac{\sum_{t=p_1+1}^{n} \hat{\varepsilon}_{1t}^2}{\sum_{t=P+1}^{n} \hat{\varepsilon}_{Pt}^n} - n + 3p_1 \right] \right) / \sigma_\varepsilon^{*2} \right\} =$$

$$= E\left\{ \left[(n-2P) \frac{\sum_{t=p_2+1}^{n} \hat{\varepsilon}_{2t}^2 - \sum_{t=p_1+1}^{n} \hat{\varepsilon}_{1t}^2}{\sum_{t=P+1}^{n} \hat{\varepsilon}_{Pt}^2} + 3(p_2 - p_1) \right] / \sigma_\varepsilon^{*2} \right\} =$$

---

[70]. Ver también Mallows, C.L. (1995).

[71]. Ver Mallows, C.L. (1973).

[72]. Otro criterio de selección basado en este concepto es el conocido como **de la función de transferencia autorregresiva,** $CAT(P) = \frac{1}{n^2} \sum_{j=1}^{P} \frac{n-j}{\hat{\sigma}_{j\varepsilon}^2} - \frac{n-P}{n\hat{\sigma}_{P\varepsilon}^2}$, debido a Parzen, E. (1974).

$$= (n-2P)\,E\left[\frac{-\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2P-2}}\right] + 3(p_2-p_1) =$$

$$= \frac{-2(p_2-p_1)(n-2P)}{n-2P-2} + 3(p_2-p_1)$$

(donde σ_ε^{*2} denota la varianza de las perturbaciones de la verdadera representación) y se obtiene que:

$$\sigma^2\left\{\left[Cp(p_2)-Cp(p_1)\right]/\sigma_\varepsilon^{*2}\right\} = (n-2P)^2\sigma^2\left[\frac{-\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2P-2}}\right] =$$

$$= \frac{-2(p_2-p_1)(n-2P)}{n-2P-2} + 3(p_2-p_1)$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$SW\left(Cp(p_1,p_2)\right) = \frac{\dfrac{-2(p_2-p_1)(n-2P)}{n-2P-2} + 3(p_2-p_1)}{(n-2P)\sqrt{\dfrac{4\left[(p_2-p_1)+(p_2-p_1)^2\right]}{(n-2P-2)(n-2P-4)} - 4\left(\dfrac{p_2-p_1}{n-2P-2}\right)^2}}$$

De las correspondientes expresiones se puede concluir que, cuando $p_2 - p_1$ aumenta: **i)** el signo crece en forma lineal; **ii)** el ruido crece en forma aproximadamente lineal y **iii)** el cociente SW crece.

Si bien el cociente $SW(Cp)$ habitualmente es débil (lo cual implica la posibilidad de sobre-estimación), como su ruido es superior, se debe esperar que el riesgo de sobre-estimación sea menor que en el criterio FPE.

Debe tenerse en cuenta que, dado que $SW(Cp)$ depende del orden P, correspondiente al modelo de mayor orden entre el conjunto de posibles modelos candidatos, ante una variación de P el criterio Cp debe ser recalculado para todos los modelos candidatos.

Además, se obtiene en forma inmediata que:

$$\lim_{n\to\infty} SW\left(Cp(p_1,p_2)\right) = \sqrt{\frac{p_2-p_1}{2}}$$

De la observación de los resultados anteriores se puede concluir que los cocientes SW correspondientes a todos los criterios eficientes, son asintóticamente equivalentes.

Suponiendo que el $AR(p_1)$ sea el que más se aproxima al verdadero modelo, la probabilidad de que, en series cortas, el criterio Cp sobre-estime la selección, es decir, la probabilidad de que se verifique que $Cp\left(AR(p_2)\right) < Cp\left(AR(p_1)\right)\,(p_2 > p_1)$ (teniendo en cuenta que $AR(P)$ denota el modelo de mayor orden de autorregresividad entre los modelos candidatos), queda definida de la siguiente forma:

$$p\left[Cp(p_2) < Cp(p_1)\right] =$$

$$= p\left[(n-2P)\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=P+1}^{n}\hat{\varepsilon}_{Pt}^2} - n + 3p_2 < (n-2P)\frac{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}{\sum_{t=P+1}^{n}\hat{\varepsilon}_{Pt}^2} - n + 3p_1\right] =$$

$$= p\left[(n-2P)\frac{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2 - \sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=P+1}^{n}\hat{\varepsilon}_{Pt}^2} > 3(p_2 - p_1)\right] =$$

$$= p\left[(n-2P)\frac{\chi^2_{2(p_2-p_1)}}{\chi^2_{n-2P}} > 3(p_2 - p_1)\right] =$$

$$= p\left[\Im_{2(p_2-p_1),n-2P} > 1.5\right]$$

Teniendo en cuenta que $\lim_{n\to\infty} \Im_{2(p_2-p_1),n-2P} = \frac{1}{2(p_2-p_1)}\chi^2_{2(p_2-p_1)}$, de la expresión anterior se obtiene que la probabilidad asintótica de sobre-estimación será de la forma:

$$\lim_{n\to\infty} p\left[Cp(p_2) < Cp(p_1)\right] = p\left[\chi^2_{2(p_2-p_1)} > 3(p_2 - p_1)\right]$$

De las definiciones anteriores se puede concluir que, para series suficientemente largas, el criterio FPEu tiene menos probabilidad de sobre-estimación que FPE y Cp.

2.10.5.5.- Criterios de selección consistentes

1) SIC[73]

Como se mencionó en la Sec. 2.10.5.2, el fundamento sobre el que se basan estos criterios es la hipótesis de que el verdadero modelo pertenece al conjunto de los modelos candidatos. Definir un criterio consistente implica, entonces, en identificar al verdadero modelo con una probabilidad asintótica igual a la unidad. Akaike, H. (1978) y Schwarz, G. (1978) propusieron criterios de selección aproximadamente equivalentes, basados en una interpretación Bayesiana[74].

Akaike propuso el criterio conocido como **BIC**[75], aplicable a modelos lineales de regresión (modelos estáticos)[76]. La propuesta de Schwarz es más general, supone que las observaciones provienen de un proceso de la familia de Koopman-Darmois, con distribución de probabilidades de la forma $f_Y(y,\theta) = \exp\left[\theta y - b(\theta)\right]$ (donde: **i)** $\theta \in \Theta$, siendo Θ un subconjunto convexo en R_p y **ii)** Y es un

---

73. La abreviatura SIC corresponde a la denominación **"Schwarz information criterion"**.

74. Ver también Risanen, J. (1978), Stone, M.B. (1979), Priestley, M.B. (1981).

75. Abreviatura que corresponde a la denominación **"Bayesian information criterion"**.

76. Estos modelos serán tratados en la segunda parte de este trabajo.

estadístico

p -dimensional suficiente de Θ). Schwarz propuso un criterio de la forma $\sum \alpha_p \mu_p$ (donde: **i)** α_p denota la probabilidad "a priori" del modelo $AR(p)$ y **ii)** μ_p denota la distribución de θ, dado el modelo $AR(p)$) y estableció una penalización fija para la selección del modelo equivocado.

La solución Bayesiana al problema de la selección consiste en elegir aquel modelo con la mayor probabilidad "a posteriori" de ser el correcto. En series suficientemente largas, esta probabilidad "a posteriori" puede ser aproximada por un desarrollo de Taylor cuyos dos primeros términos son de la forma:

$$(n-p)\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right)+p\ln(n-p)$$

(donde $\hat{\sigma}_{p\varepsilon}^{(MV)2}$ denota el estimador máximo-verosímil de la varianza de las perturbaciones del modelo $AR(p)$) y los términos restantes son acotados, de modo que no ejercen ninguna influencia en series suficientemente largas[77]. De modo que, normalizando la expresión de Schwarz, se obtiene el criterio[78]:

$$SIC\left(AR(p)\right)=\ln\left(\sigma_{p\varepsilon}^{(MV)2}\right)+\frac{p\ln(n-p)}{n-p}$$

Sean dos modelos candidatos $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ $(p_2 > p_1)$. Aplicando un razonamiento similar al utilizado en el AIC, se obtiene que:

$$S\left(SIC(p_1,p_2)\right)=E\left\{\left[SIC(p_2)-SIC(p_1)\right]/\sigma_\varepsilon^{*2}\right\}=$$

$$=E\left\{\left[\left(\ln\left(\hat{\sigma}_{p_2\varepsilon}^{(MV)2}\right)+\frac{p_2\ln(n-p_2)}{n-p_2}\right)-\left(\ln\left(\hat{\sigma}_{p_1\varepsilon}^{(MV)2}\right)+\frac{p_1\ln(n-p_1)}{n-p_1}\right)\right]/\sigma_\varepsilon^{*2}\right\}=$$

$$=\ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right]-\frac{2(p_2-p_1)}{(n-2p_1)(n-2p_2)}+\frac{p_2\ln(n-p_2)}{n-p_2}-\frac{p_1\ln(n-p_1)}{n-p_1}$$

y que:

$$\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right]=\frac{4(p_2-p_1)}{(n-2p_1+2)(n-2p_2)}$$

De modo que el cociente SW , con respecto a la sobre-estimación, queda definido de la siguiente forma:

---

77. Ver Schwarz, G. (1978).

78. Debe tenerse en cuenta que las propiedades asintóticas de los estimadores de los coeficientes ϕ_j de una representación $AR(p)$ se mantienen, aún cuando el valor de p sea desconocido, si su estimador es consistente.

$$SW\left(SIC\left(p_1,p_2\right)\right)=\frac{1}{2}\sqrt{\frac{\left(n-2p_1+2\right)\left(n-2p_2\right)}{p_2-p_1}}\left\{\ln\left[\frac{\left(n-p_1\right)\left(n-2p_2\right)}{\left(n-2p_1\right)\left(n-p_2\right)}\right]-\frac{2\left(p_2-p_1\right)}{\left(n-2p_1\right)\left(n-2p_2\right)}+\right.$$

$$\left.+\frac{p_2\ln\left(n-p_2\right)}{n-p_2}-\frac{p_1\ln\left(n-p_1\right)}{n-p_1}\right\}$$

A partir de esta expresión se puede concluir que este cociente adolece de los mismos defectos que el $SW(AIC)$, en la medida que está afectado por un sesgo hacia la sobre-estimación, para series cortas (ver Shibata, R. (1976), Jones, R.H. (1975)).

Por otra parte, se obtiene en forma inmediata que:

$$\lim_{n\to\infty} SW\left(SIC\left(p_1,p_2\right)\right)=$$

$$=\lim_{n\to\infty}\frac{n}{2\sqrt{p_2-p_1}}\left[-\frac{2\left(p_2-p_1\right)}{n}+\frac{p_2-p_1}{n}-\frac{2\left(p_2-p_1\right)}{n^2}+\frac{\left(p_2-p_1\ln(n)\right)}{n}\right]=$$

$$=\lim_{n\to\infty}\frac{1}{2\sqrt{p_2-p_1}}\ln(n)=\infty$$

Debe tenerse en cuenta que, como se verá en las páginas siguientes, esta característica de poseer un cociente SW asintóticamente igual a infinito, se verifica para todos los criterios consistentes.

La probabilidad de que, para series cortas, el criterio SIC sobre-estime la selección, es decir, la probabilidad de que se verifique que $SIC\left(AR(p_2)\right)<SIC\left(AR(p_1)\right)\left(p_2>p_1\right)$, queda definida de la siguiente forma:

$$p\left[SIC\left(p_2\right)<SIC\left(p_1\right)\right]=$$

$$=p\left[\ln\left(\hat{\sigma}_{2\varepsilon}^2\right)+\frac{p_2\ln\left(n-p_2\right)}{n-p_2}<\ln\left(\hat{\sigma}_{1\varepsilon}^2\right)+\frac{p_1\ln\left(n-p_1\right)}{n-p_1}\right]=$$

$$=p\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)<\ln\left(\frac{n-p_2}{n-p_1}\right)+\frac{p_1\ln\left(n-p_1\right)}{n-p_1}-\frac{p_2\ln\left(n-p_2\right)}{n-p_2}\right]=$$

$$=p\left\{\Im_{2(p_2-p_1),n-2p_2}>\frac{n-2p_2}{2\left(p_2-p_1\right)}\left(\frac{n-p_1}{n-p_2}\exp\left[\frac{p_2\ln\left(n-p_2\right)}{n-p_2}-\frac{p_1\ln\left(n-p_1\right)}{n-p_1}\right]-1\right)\right\}$$

De esta expresión se obtiene en forma inmediata que:

$$\lim_{n\to\infty}\frac{n-2p_2}{2\left(p_2-p_1\right)}\left\{\frac{n-p_1}{n-p_2}\exp\left[\frac{p_2\ln\left(n-p_2\right)}{n-p_2}-\frac{p_1\ln\left(n-p_1\right)}{n-p_1}\right]-1\right\}=$$

$$=\lim_{n\to\infty}\frac{n-2p_2}{2\left(p_2-p_1\right)}\left\{\frac{n-p_1}{n-p_2}\left[1+\frac{p_2\ln\left(n-p_2\right)}{n-p_2}-\frac{p_1\ln\left(n-p_1\right)}{n-p_1}+O\left(\frac{1}{n^2}\right)\right]-1\right\}=$$

$$= \lim_{n\to\infty} \frac{n}{2(p_2 - p_1)} \left[\frac{p_2 - p_1}{n} + \frac{(p_2 - p_1)\ln(n)}{n} \right] = \infty$$

Este resultado demuestra formalmente la propiedad comentada en la Sec. 2.10.5.2, que para el criterio SIC (como para todos los criterios consistentes) la probabilidad de sobre-estimación es asintóticamente nula[79].

2) HQ

Hannan, E.J.; Quinn, B.G. (1979) propusieron un criterio basado en las condiciones asintóticas de los procesos AR del tipo:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \qquad (t = p+1, p+2, ..., n)$$

(donde: **i)** $E(\varepsilon_t / \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_1) = 0$; **ii)** $E(\varepsilon_t^2 / \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_1) = \sigma_\varepsilon^2$ y **iii)** $E(\varepsilon_t^4 / \varepsilon_{t-1}, ..., \varepsilon_1) < \infty$). Suponiendo un modelo verdadero $AR(p^*)$ y un modelo candidato $AR(p)$ $(p > p^*)$, Hannan-Quinn demostraron que, para valores de n suficientemente grandes, se verifica que:

$$\hat{\phi}_p = b_p(n-p)\sqrt{\frac{2\ln[\ln(n-p)]}{n-p}}$$

(donde $b_p(n-p) \in [0;1]$). A partir de este resultado propusieron un criterio de la forma $\ln(\hat{\sigma}_{p\varepsilon}^2) + pC_{(n-p)}$ y demostraron que la propiedad de consistencia se verifica para $C_{(n-p)} = \frac{2c}{n-p}\ln(n-p)$ $(c \geq 1)$. Lo cual condujo a un criterio fuertemente consistente de la forma[80]:

$$\psi(p) = \ln(\hat{\sigma}_{p\varepsilon}^{(MV)2}) + 2c\frac{p\ln[\ln(n-p)]}{n-p} \qquad (c \geq 1)$$

De modo que, para $c = 1$, el criterio de selección queda definido por:

$$HQ(p) = \ln(\hat{\sigma}_{p\varepsilon}^{(MV)2}) + \frac{2p\ln[\ln(n-p)]}{n-p}$$

Sean dos modelos candidatos, $AR(p_1)$ (considerado el más próximo al verdadero modelo) y $AR(p_2)$ $(p_2 > p_1)$. La señal correspondiente al criterio HQ será de la forma:

$$S(HQ(p_1, p_2)) = E\{[HQ(p_2) - HQ(p_1)] / \sigma_\varepsilon^{*2}\} =$$

$$= E\left\{\left[\left(\ln(\hat{\sigma}_{p_2\varepsilon}^{(MV)2}) + \frac{2p_2\ln(\ln(n-p_2))}{n-p_2}\right) - \left(\ln(\hat{\sigma}_{p_1\varepsilon}^{(MV)2}) + \frac{2p_1\ln(\ln(n-p_1))}{n-p_1}\right)\right] / \sigma_\varepsilon^{*2}\right\} =$$

---

79. Ver Hannan, E.J. (1980).

80. Ver Hannan, e.J. (1980).

$$= \ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right] - \frac{2(p_2-p_1)}{(n-2p_1)(n-2p_2)} + \frac{2p_2\ln(\ln(n-p_2))}{n-p_2} - \frac{2p_1\ln(\ln(n-p_1))}{n-p_1}$$

Por otra parte, se obtiene que:

$$\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right] = \frac{4(p_2-p_1)}{(n-2p_1+2)(n-2p_2)}$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$SW(HQ(p_1,p_2)) = \frac{1}{2}\sqrt{\frac{(n-2p_1+2)(n-2p_2)}{p_2-p_1}}\left\{\ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right] - \right.$$

$$\left. - \frac{2(p_2-p_1)}{(n-2p_1)(n-2p_2)} + \frac{2p_2\ln(\ln(n-p_2))}{n-p_2} - \frac{2p_1\ln(\ln(n-p_1))}{n-p_1}\right\}$$

A partir de este resultado se puede concluir que este cociente, como el correspondiente al AIC, crece con $p_2 - p_1$ para valores pequeños de $p_2 - p_1$ y luego tiende a infinito. Es decir que, para series cortas, presenta un sesgo hacia la sobre-estimación.

Por otra parte, se obtiene en forma inmediata que:

$$\lim_{n\to\infty} SW(HQ(p_1,p_2)) =$$

$$= \lim_{n\to\infty} \frac{n}{2\sqrt{p_2-p_1}}\left[-\frac{2(p_2-p_1)}{n} + \frac{p_2-p_1}{n} - \frac{2(p_2-p_1)}{n^2} + \frac{2(p_2-p_1)\ln(\ln(n))}{n}\right] = \infty$$

La probabilidad de que el criterio HQ sobre-estime la selección, es decir, la probabilidad de que se verifique que $HQ(AR(p_2)) < HQ(AR(p_1))\ (p_2 > p_1)$, queda definida de la siguiente forma:

$$p[HQ(p_2) < HQ(p_1)] =$$

$$= p\left[\ln(\hat{\sigma}_{2\varepsilon}^2) + \frac{2p_2\ln[\ln(n-p_2)]}{n-p_2} < \ln(\hat{\sigma}_{1\varepsilon}^2) + \frac{2p_1\ln[\ln(n-p_1)]}{n-p_1}\right] =$$

$$= p\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right) < \ln\left(\frac{n-p_2}{n-p_1}\right) + \frac{2p_1\ln[\ln(n-p_1)]}{n-p_1} - \frac{2p_2\ln[\ln(n-p_1)]}{n-p_1}\right] =$$

$$= p\left\{\Im_{2(p_2-p_1),n-2p_2} > \frac{n-2p_2}{2(p_2-p_1)}\left[\frac{n-p_1}{n-p_2}\exp\left(\frac{2p_2\ln[\ln(n-p_2)]}{n-p_2} - \frac{2p_1\ln[\ln(n-p_1)]}{n-p_1}\right)\right]\right\}$$

De esta expresión se obtiene en forma inmediata que:

$$\lim_{n\to\infty}\frac{n-2p_2}{2(p_2-p_1)}\left[\frac{n-p_1}{n-p_2}\exp\left(\frac{2p_2\ln[\ln(n-p_2)]}{n-p_2}-\frac{2p_1\ln[\ln(n-p_1)]}{n-p_1}\right)-1\right]=$$

$$=\lim_{n\to\infty}\frac{n-2p_2}{2(p_2-p_1)}\left[\frac{n-p_1}{n-p_2}\left(1+\frac{2p_2\ln[\ln(n-p_2)]}{n-p_2}-\frac{2p_1\ln[\ln(n-p_1)]}{n-p_1}+O\left(\frac{1}{n^2}\right)\right)-1\right]=$$

$$=\lim_{n\to\infty}\frac{n}{2(p_2-p_1)}\left[\frac{p_2-p_1}{n}+\left(\frac{2(p_2-p_1)\ln[\ln(n)]}{n}\right)\right]=\infty$$

Es decir, que la probabilidad de sobre-estimación es asintóticamente nula.

Si bien las probabilidades de sobre-estimación para los criterios de selección consistentes son decrecientes, dado que la función de penalización $\ln(n-p)$ es mucho más fuerte que la función $2\ln[\ln(n-p)]$, se verifica que para el SIC decrece mucho más rápidamente que para el HQ.

Obsérvese que, si bien el AIC es asintóticamente eficiente y el HQ es asintóticamente consistente, ambos poseen estructuras de penalización similares y ambos poseen un sesgo hacia la sobre-estimación en series cortas.

3) HQ-corregido

Este criterio se basa en la corrección de la función de penalización del HQ en forma similar a la aplicada en el AICc. De modo que queda definido de la siguiente forma:

$$HQc(p)=\ln\left(\hat{\sigma}_{p\varepsilon}^{(MV)2}\right)+\frac{2p\ln[\ln(n-p)]}{n-2(p+1)}$$

Sean dos modelos candidatos, $AR(p_1)$ (considerado el más aproximado al verdadero modelo) y $AR(p_2)\,(p_2>p_1)$. La señal correspondiente a este criterio será de la forma:

$$S(HQc(p_1,p_2))=E\left\{\left[HQc(p_2)-HQc(p_1)\right]/\sigma_\varepsilon^{*2}\right\}=$$

$$=E\left\{\left[\left(\ln\left(\hat{\sigma}_{p_2\varepsilon}^{(MV)2}\right)+\frac{2p_2\ln[\ln(n-p_2)]}{n-p_2-2}\right)-\left(\ln\left(\hat{\sigma}_{p_1\varepsilon}^{(MV)2}\right)+\frac{2p_1\ln[\ln(n-p_1)]}{n-p_1-2}\right)\right]/\sigma_\varepsilon^{*2}\right\}=$$

$$=\ln\left[\frac{(n-p_1)(n-2p_2)}{(n-2p_1)(n-p_2)}\right]-\frac{2(p_2-p_1)}{(n-2p_1)(n-2p_2)}+\frac{2p_2\ln[\ln(n-p_2)]}{n-p_2-2}-\frac{2p_1\ln[\ln(n-p_1)]}{n-p_2-2}$$

Por otra parte, se obtiene que:

$$\sigma^2\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}_{2t}^2}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}_{1t}^2}\right)\right]=\frac{4(p_2-p_1)}{(n-2p_1+2)(n-2p_2)}$$

De modo que el cociente SW, con respecto a la sobre-estimación, queda definido de la siguiente forma:

$$\lim_{n\to\infty} SW\left(HQc\left(p_1,p_2\right)\right) =$$

$$= \lim_{n\to\infty} \frac{n}{2\sqrt{p_2 - p_1}}\left[-\frac{2\left(p_2 - p_1\right)}{n} + \frac{p_2 - p_1}{n} - \frac{2\left(p_2 - p_1\right)}{n^2} + \frac{2\left(p_2 - p_1\right)\ln\left(\ln\left(n\right)\right)}{n}\right] = \infty$$

A partir de este resultado se demuestra que, para: **i)** $n \geq 6$; **ii)** $p_2 - p_1 < n - p_1 - 2$ y **iii)** $0 < p^* \leq p < n-3$, a medida que $p_2 - p_1$ crece, $SW\left(HQc\right)$ crece y es mayor que $SW\left(HQ\right)$ (ver McQuarrie, A.D.R.; Tsai, C.L. (2007)).

Por otra parte, se obtiene en forma inmediata que:

$$SW\left(HQc\left(p_1,p_2\right)\right) = \frac{1}{2}\sqrt{\frac{\left(n-2p_1+2\right)\left(n-2p_2\right)}{p_2 - p_1}}\left\{\ln\left[\frac{\left(n-p_1\right)\left(n-2p_2\right)}{\left(n-2p_1\right)\left(n-p_2\right)}\right] - \frac{2\left(p_2 - p_1\right)}{\left(n-2p_1\right)\left(n-2p_2\right)} + \right.$$

$$\left. + \frac{2p_2\ln\left[\ln\left(n-p_2\right)\right]}{n-p_2-2} - \frac{2p_1\ln\left[\ln\left(n-p_1\right)\right]}{n-p_1-2}\right\}$$

La probabilidad de que, para series cortas, el criterio HQc sobre-estime la selección, es decir, la probabilidad de que se verifique que $HQc\left(AR\left(p_2\right)\right) < HQc\left(AR\left(p_1\right)\right)\left(p_2 > p_1\right)$, queda definida de la siguiente forma:

$$p\left[HQc\left(p_2\right) < HQc\left(p_1\right)\right] =$$

$$= p\left[\ln\left(\hat{\sigma}^2_{2\varepsilon}\right) + \frac{2p_2\ln\left(\ln\left(n-p_2\right)\right)}{n-2p_2-2} < \ln\left(\hat{\sigma}^2_{1\varepsilon}\right) + \frac{2p_1\ln\left(\ln\left(n-p_1\right)\right)}{n-2p_1-2}\right] =$$

$$= p\left[\ln\left(\frac{\sum_{t=p_2+1}^{n}\hat{\varepsilon}^2_{2t}}{\sum_{t=p_1+1}^{n}\hat{\varepsilon}^2_{1t}}\right) < \ln\left(\frac{n-p_2}{n-p_1}\right) + \frac{2p_1\ln\left(\ln\left(n-p_1\right)\right)}{n-2p_1-2} - \frac{2p_2\ln\left(\ln\left(n-p_2\right)\right)}{n-2p_2-2}\right] =$$

$$= p\left\{\Im_{2\left(p_2-p_1\right),n-2p_2} > \frac{n-2p_2}{2\left(p_2-p_1\right)}\left[\frac{n-p_1}{n-p_2}\exp\left(\frac{2p_2\ln\left(\ln\left(n-p_2\right)\right)}{n-2p_2-2} - \frac{2p_1\ln\left(\ln\left(n-p_1\right)\right)}{n-2p_1-2}\right) - 1\right]\right\}$$

De esta expresión se obtiene en forma inmediata que:

$$\lim_{n\to\infty}\frac{n-2p_2}{2\left(p_2-p_1\right)}\left[\frac{n-p_1}{n-p_2}\exp\left(\frac{2p_2\ln\left(\ln\left(n-p_2\right)\right)}{n-2p_2-2} - \frac{2p_1\ln\left(\ln\left(n-p_1\right)\right)}{n-2p_1-2}\right) - 1\right] =$$

$$= \lim_{n\to\infty}\frac{n-2p_2}{2\left(p_2-p_1\right)}\left[\frac{n-p_1}{n-p_2}\left(1 + \frac{2p_2\ln\left(\ln\left(n-p_2\right)\right)}{n-2p_2-2} - \frac{2p_1\ln\left(\ln\left(n-p_1\right)\right)}{n-2p_1-2} + O\left(\frac{1}{n^2}\right)\right) - 1\right] =$$

$$= \lim_{n\to\infty} \frac{n}{2(p_2 - p_1)}\left[\frac{p_2 - p_1}{n} + \frac{2(p_2 - p_1)\ln(\ln(n))}{n}\right] = \infty$$

Cabe destacar que, si bien el HQc es asintóticamente consistente, puede considerarse un criterio de selección adecuado tanto bajo el supuesto que el verdadero modelo pertenece al conjunto de los modelos candidatos, como bajo el supuesto contrario.

De los resultados anteriores se puede concluir que HQ y HQc son asintóticamente equivalentes.

Con un razonamiento similar al empleado para tratar la sobre-estimación, se pueden analizar las propiedades de los criterios de selección para series cortas con respecto a subestimación del orden de autorregresividad. A este respecto, debe tenerse en cuenta que, en este caso: **i)** $\sum_{t=p^*+1}^{n} \hat{\varepsilon}_t^{*2} : \chi^2_{n-p^*}$ (donde $\left\{\varepsilon_t^{*2}\right\}$ $\left(t = p^* + 1,\ldots,n\right)$ denota la sucesión de residuos obtenidos de aplicar a la serie $\left\{Y_t\right\}$ $\left(t = 1,2,\ldots,n\right)$ el verdadero modelo $AR\left(p^*\right)$) ; **ii)** $\sum_{t=p+1}^{n} \hat{\varepsilon}_{pt}^2 : \chi^2_{n-p^*}(\lambda^*)$ $(p < p^*)$ (donde λ^* denota el parámetro representativo de la no-centralidad de la distribución) y **iii)** $\frac{\sum_{t=p^*+1}^{n} \hat{\varepsilon}_t^{*2}}{\sum_{t=p+1}^{n} \hat{\varepsilon}_{pt}^2} : \frac{\chi^2_{n-p^*}}{\chi^2_{n-p^*} + \chi_{(p^*-p)}(\lambda)}$ (debe tenerse en cuenta que, contrariamente a lo que ocurre en el análisis de la sobre-estimación, en este caso las distribuciones son no-centradas).

Por otra parte, se demuestra que:

$$KL = \ln\left(\frac{\hat{\sigma}^2_{p\varepsilon}}{\hat{\sigma}^{*2}_{\varepsilon}}\right) + \frac{\sigma^{*2}_{\varepsilon}}{\hat{\sigma}^2_{p\varepsilon}} + \frac{L_2}{\hat{\sigma}^2_{p\varepsilon}} - 1 \qquad (p < p^*)$$

$$E(KL) = \ln(2) - \ln(n) - 1 + \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{\lambda}{2}\right)^j e^{-\lambda/2}\left[\psi\left(\frac{n-2p}{2} + j\right) + \frac{n+2p+\lambda}{n-2p-2+2j}\right] \qquad (p < p^*)$$

A partir de estos resultados se obtiene que la esperanza matemática de la distancia de Kullback-Leibler entre un modelo subestimado y el verdadero modelo, tiende a ser menor que el valor esperado de la norma L_2, lo cual implica que los criterios de selección basados en la distancia KL penalizan la subestimación en forma menos fuerte que los que se basan en la medida L_2 (la medida KL es particularmente recomendable en aquellos criterios con funciones de penalización fuertes, en las cuales la subestimación es insignificante).

Se demuestra, además, que para modelos autorregresivos de orden mayor que 5, la eficiencia esperada de los criterios de selección en términos de KL es menor que la de los criterios en términos de L_2 y viceversa, para modelos autorregresivos de orden menor que 5, la eficiencia esperada de los criterios de selección en términos de KL es mayor que la de los criterios que se basan en la medida L_2. A los criterios con las funciones de penalización más fuertes y con mayores probabilidades de subestimación les corresponde la mayor eficiencia observada en términos de KL. Asimismo, los criterios con un marcado sesgo hacia la sobre-estimación poseen una menor eficiencia observada en series cortas que aquéllos con sesgo hacia la subestimación.

Por otra parte cabe recordar que, como se mencionó en la Sec. 2.10.5.1, a los criterios con

cociente *SW* significativamente positivo, les corresponden probabilidades relativamente pequeñas de seleccionar el modelo en forma equivocada (tanto por sobre-estimación como por subestimación) y, a los criterios con cociente *SW* próximo a cero (y, en particular, negativo), les corresponden mayores probabilidades de selección equivocada. Para criterios con cociente *SW* fuertemente negativo, las probabilidades de selección equivocada son aproximadamente iguales a uno.

En relación a los criterios de selección considerados en forma individual, cabe destacar que: **i)** a AIC le corresponde un cociente *SW* negativo con respecto a la subestimación (y, como ya se comentó, un *SW* débil con respecto a la sobre-estimación) y, por consiguiente, que posee un sesgo hacia la subestimación (y, sólo ocasionalmente, un sesgo hacia la sobre-estimación excesiva): **ii)** dada su fuerte función de penalización, AICu posee algún sesgo hacia la subestimación cuando los coeficientes del modelo candidato no son suficientemente significativos; **iii)** a AICu le corresponde el mayor cociente *SW* para órdenes de autorregresividad $p < p^*$, lo cual indica que posee un sesgo hacia la subestimación mayor que cualquier otro criterio de selección; **iv)** AICc, AICu y HQc poseen funciones de penalización superlineales que eliminan el riesgo de sobre-estimación excesiva, pero están afectadas por una alto riesgo de subestimación en series cortas; **v)** las funciones de penalización de AIC, SIC y HQ pueden provocar sobre-estimaciones excesivas en series cortas; **vi)** si bien AIC es asintóticamente eficiente y HQ es asintóticamente consistente, ambos poseen estructuras de penalización similares y ambos sobre-estiman excesivamente en series cortas; **vii)** en SIC y FPEu el riesgo de subestimación es mayor que el de sobre-estimación; **viii)** dada su función de penalización débil, AIC observa un mejor comportamiento con respecto a la subestimación asintótica; **ix)** dadas sus fuertes funciones de penalización, AICc y HQc poseen un sesgo hacia la subestimación, pero presentan un balance adecuado entre sobre y subestimación; **x)** AIC, HQ, Cp y FPE también poseen un sesgo hacia la subestimación, pero menos importante que en los criterios anteriores.

2.10.5.6.- El método de Quenouille

Un método de selección que no pertenece a la familia de los criterios con término de penalización, basado en los tests de significatividad conjunta de los coeficientes de una representación, es el debido a Quenouille, M.H. (1947).

De acuerdo con este criterio, dado un proceso $\{Y_t\}$ estacionario de segundo orden. Una forma de determinar un valor de p adecuado para su representación consiste en contrastar, en forma secuencial, las hipótesis nulas:

$$H_0: \phi_p = 0$$

$$H_0: \phi_{p-1} = \phi_p = 0$$

$$H_0: \phi_{p-2} = \phi_{p-1} = \phi_p = 0$$

$$\ldots\ldots\ldots\ldots$$

$$H_0: \phi_1 = \phi_2 = \ldots = \phi_p = 0$$

con respecto a la alternativa H_1: *que al menos uno de los coeficientes* ϕ_j *sea significativo* .

Obsérvese que si en esta secuencia se considera que existen razones suficientes para rechazar la j-ésima hipótesis nula , $H_0: \phi_{p-(j-1)} = \phi_{p-(j-2)} = \ldots = \phi_p = 0$, con un nivel de confiabilidad dado, entonces se rechazarán todas las hipótesis nulas siguientes (es decir que la secuencia se interrumpirá cuando una hipótesis nula sea rechazada)[81].

---

[81]. Obviamente la estimación de p dependerá del valor de α seleccionado. Debe tenerse en cuenta que la medida del error de tipo-1 correspondiente al j-ésimo test α_j^* $(j = 1,2,\ldots,p)$, de una secuencia es distinto de aquél correspondiente a dicho j-ésimo test, tomado en forma individual α . De acuerdo con Anderson, T.W. (1971), la relación entre ambos niveles de significación es de la

El estadístico correspondiente está definido por:

$$\Delta_j\left(\hat{\sigma}_\varepsilon^2\right) = n\left\{\ln\left[\hat{\sigma}_\varepsilon^2\left(AR(p-j)\right) - \hat{\sigma}_\varepsilon^2\left(AR(p)\right)\right]\right\}$$

(donde $\hat{\sigma}_\varepsilon^2\left(AR(p)\right) = \dfrac{1}{n-p}\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2$ denota el estimador máximo-verosímil de σ_ε^2 cuando al proceso $\{Y_t\}$ se le ha aplicado una representación $AR(p)$) y se demuestra que $\Delta_j\left(\hat{\sigma}_\varepsilon^2\right) \xrightarrow{d} \chi_{(j)}^2$.

Con un razonamiento similar, las hipótesis nulas anteriores pueden ser contrastadas utilizando los estadísticos de Quenouille[82]:

$$Q_1 = n\hat{\phi}_{pp}^2(Y)$$

$$Q_2 = n\left[\hat{\phi}_{p-1,p-1}^2(Y) + \hat{\phi}_{pp}^2(Y)\right]$$

.....

$$Q_j = n\left[\hat{\phi}_{p-j+1,p-j+1}(Y) + \hat{\phi}_{p-j,p-j}^2(Y) + ... + \hat{\phi}_{pp}^2(Y)\right]$$

.....

$$Q_p = n\left[\hat{\phi}_{11}^2(Y) + \hat{\phi}_{22}^2(Y) + ... + \hat{\phi}_{pp}^2(Y)\right]$$

los cuales, bajo el supuesto de que la correspondiente hipótesis nula es verdadera, se distribuyen asintóticamente de acuerdo con funciones $\chi_{(i)}^2$ $(i = 1,2,...,p)$, respectivamente[83].

2.10.6.- La condición de linealidad

Como se señaló desde el comienzo de este trabajo, los modelos estudiados en este capítulo se basan en el supuesto que el proceso $\{Y_t\}$ es débilmente linealmente $AR(p)$, es decir, tal que:

$$E\left(Y_t / Y_{t-1},...,Y_{t-p}\right) = \phi_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p}$$

Luego, construido un modelo, surge el problema de decidir si esta hipótesis de linealidad realizada "a priori" puede ser considerada válida o no.

---

forma $\alpha_j^* = \begin{cases} \alpha & (j=1) \\ 1-(1-\alpha)^j & (j=2,3,...,p) \end{cases}$. En general, si el nivel de significatividad individual del j-ésimo test es α_j $(j=1,2,...,p)$, entonces la relación anterior puede ser expresada mediante la siguiente fórmula de recurrencia:

$$\alpha_j^* = \begin{cases} \alpha_j & (j=1) \\ \alpha_j\left(1-\alpha_j^*\right) + \alpha_{j-1} & (j=2,3,...,p) \end{cases}.$$

82. Ver Quenouille, M.H. (1949).

83. Ver Anderson, T.W. (1971).

Una forma particular de un test debido a Ramsey, J.B. (1969) (denominado habitualmente como RESET o RASET)[84] considera el contraste de la hipótesis nula de linealidad con respecto a una hipótesis alternativa no-específica de no-linealidad, representada por una función de la forma:

$$h\left(Y_{t-1},...,Y_{t-p}\right) = b_2 m^2\left(Y_t\right) + b_3 m^3\left(Y_t\right) + ... + b_k m^k\left(Y_t\right)$$

(donde $m\left(Y_t\right) = E\left(Y_t / Y_{t-1},...,Y_{t-p}\right)$). De modo que la hipótesis nula puede ser expresada de la siguiente forma $H_0 : b_2 = b_3 = ... = b_k = 0$ y la hipótesis alternativa como H_1: *existe por lo menos un* $b_j \neq 0$ $\left(j = 2,3,...,k\right)$. El test consiste, entonces, en construir un modelo de la forma:

$$\hat{\varepsilon}_t = \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} + ... + \hat{\phi}_p Y_{t-p} + \hat{b}_2 \hat{Y}_t^2 + \hat{b}_3 \hat{Y}_t^3 + ... + \hat{b}_k \hat{Y}_t^k$$

donde $\left\{\hat{\varepsilon}_t\right\}$ $\left(t = p+1,...,n\right)$ denota la serie residual que surge de la aplicación del modelo:

$$\hat{Y}_t = \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1} + ... + \hat{\phi}_p Y_{t-p}$$

A partir de este modelo se obtiene el correspondiente coeficiente de determinación (R_ε^2) que permite definir el estadístico nR_ε^2 el cual, bajo el supuesto que la hipótesis nula es verdadera, converge en-distribución a una función $\chi^2_{(k+p-1)}$ [85].

Tsay, R.S. (1986) propuso una generalización de este test a partir de la construcción de un modelo de la forma:

$$\hat{\varepsilon}_t = \hat{\phi}_1 Y_{t-1} + ... + \hat{\phi}_p Y_{t-p} + \sum_{i=1}^{p}\sum_{j=1}^{p} \hat{\phi}_{ij}^{*} Y_{t-i} Y_{t-j}$$

y considerando la hipótesis nula $H_0 : \phi_{ij}^{*} = 0\left(i,j = 1,2,...,p\right)$ [86]. Este test de Tsay es más potente que el RESET con respecto a una gran variedad de modelos no-lineales. Pero el RESET posee la ventaja de involucrar un número menor de coeficientes.

Con un razonamiento similar, Teräsvirta, T.C.; Lin, F.; Granger, C.W.J. (1993) propusieron un test a partir de la construcción de un modelo de la forma [87]:

$$\hat{\varepsilon}_t = \hat{\phi}_1 Y_{t-1} + ... + \hat{\phi}_p Y_{t-p} + \sum_{i=1}^{p} \hat{\phi}_{ij}^{*} Y_{t-i} Y_{t-j} + \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{h=1}^{p} \hat{\phi}_{ijh}^{**} Y_{t-i} Y_{t-j} Y_{t-h}$$

---

84. "Regression error specification test" o "Ramsey's error specification test". Su forma general será tratada en el Cap. 6.

85. En particular, si $k = 2$, queda definido el test de linealidad de Keenan, D.M. (1985). El test RESET también puede ser interpretado como un test LM con respecto a una alternativa específica (ver Teräsvirta, T. (1990), Granger, C.W.J.; Teräsvirta, T. (1993)).

86. Obviamente la generalización de este test al caso multivariable es inmediata.

87. Una generalización interesante de estos tests de linealidad se basa en la aplicación de modelos de redes neuronales (ver Lee, T.; White, H.; Granger, C.W.J. (1993), Teräsvirta, T.C.; Lin, F.; Granger, C.W.J. (1993)).

y considerando la hipótesis nula $H_0{:}\phi_{ij}^{*} = \phi_{ijh}^{**} = 0\,(i,j,h = 1,2,\ldots,p)$.

2.10.7.- El proceso AR(1)

Sea, en particular, un proceso $\{Y_t\}$ **autorregresivo de primer orden (AR(1))** o **proceso de Markov** que admite una representación de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}{:}WN - N(0,\sigma_\varepsilon^2)$).

De acuerdo a lo expresado en la Sec. 1.3, la función de densidad conjunta para cualquier sucesión finita de sus variables puede ser expresada de la siguiente forma:

$$f_{Y_1,\ldots,Y_t}(y_1,\ldots,y_t) = f_{Y_1}(y_1)\prod_{\tau=2}^{t} f(Y_\tau / y_{\tau-1})$$

Su representación en-promedios móviles será la forma:

$$Y_t = (1-\phi_1 B)^{-1}\varepsilon_t = \sum_{j=0}^{\infty}\phi_1^j B^j \varepsilon_t = \sum_{j=0}^{\infty}\psi_j \varepsilon_{t-j} \qquad (\psi_0 = 1)$$

De esta expresión se concluye que el proceso cumplirá la condición de estacionariedad siempre que la única raíz de la ecuación-característica, ϕ_1^{-1}, sea, en valor absoluto, mayor que la unidad; es decir, siempre que se verifique que $|\phi_1| < 1$ [88].

Si se tratara de un proceso $\{Y_t\}$ tal que $E(Y_t) = E(Y_{t-1}) = m(Y)$, entonces la representación $AR(1)$ que figura más arriba incluirá un término independiente de la forma $\phi_0 = m(Y)(1-\phi_1)$ [89].

La función de autocovarianzas asume la forma[90]:

---

[88]. Cuando $|Y_t|$ asume valores grandes, si $|\phi_1| < 1$, el proceso converge al origen anulando la influencia del término aleatorio. Por el contrario, cuando $|Y_t|$ asume valores pequeños, el término aleatorio desestabiliza el sistema. Luego, la condición $|\phi_1| < 1$ asegura que el proceso posee un atractor puntual en $Y = 0$.

[89]. De lo que se puede concluir que $m(Y) = \dfrac{\phi_0}{1-\phi_0}$.

[90]. Supóngase, en particular, que el proceso $\{Y_t\}$ sea continuo en el dominio del tiempo y sus variables se distribuyan Normalmente. La función de autocovarianzas será, entonces, de la forma $\gamma(Y(s),Y(t)) = \gamma_{|t-s|}(Y) = \dfrac{\gamma(Y(t_1),Y(s))\gamma(Y(t_1),Y(t))}{\gamma_0(Y)} =$

$= \dfrac{\gamma_{|s-t_1|}(Y)\gamma_{|t-t_1|}(Y)}{\gamma_0(Y)}$ $(s,t \in R)$. Expresión que puede ser escrita de la siguiente forma: $\gamma_0(Y)\gamma_{|t-s|}(Y) = \gamma_0(Y)\gamma_{t+s}(Y) =$

$= \gamma_{|s-t_1|}(Y)\gamma_{|t-t_1|}(Y)$ $(s,t \in R_+)$. La única solución de esta ecuación (conocida como **ecuación de Cauchy**) es de la forma

$$\gamma_j(Y) = \phi_1^j \gamma_0(Y) \qquad (j = 1,2,...)$$

y la función de autocorrelaciones satisface el siguiente sistema de ecuaciones en-diferencias:

$$\rho_j(Y) = \phi_1 \rho_{j-1}(Y) \qquad (j = 1,2,3,...)$$

cuya solución, $\rho_j(Y) = \rho_1^j(Y)\,(j = 1,2,3,...)$, permite concluir que la sucesión $\{\rho_j(Y)\}$ decrece exponencialmente en valor absoluto.

La función de autocorrelaciones parciales será de la forma:

$$\phi_{jj}(Y) = \begin{cases} \rho_1(Y) & para \quad j = 1 \\ 0 & para \quad j = 2,3,... \end{cases}$$

De acuerdo con la expresión general obtenida en la Sec.2.10.2, la varianza no-condicionada de Y_t será de la forma $\sigma_Y^2 = \dfrac{\sigma_\varepsilon^2}{1 - \phi_1^2}$.

Del correspondiente sistema de ecuaciones de Yule-Walker se obtiene, en forma inmediata, que $\hat{\phi}_1^{(YW)} = \hat{\rho}_1(Y)$.

El estimador mínimo-cuadrático del coeficiente ϕ_1 se obtiene de minimizar la función:

$$f\left(\hat{\phi}_1^{(CM)}\right) = \sum_{t=2}^{n}\left(Y_t - \hat{\phi}_1^{(CM)} Y_{t-1}\right)^2 = \sum_{t=2}^{n} \hat{\varepsilon}_t^2$$

Es decir, será:

$$\hat{\phi}_1^{(CM)} = \frac{\sum_{t=2}^{n} Y_{t-1} Y_t}{\sum_{t=2}^{n} Y_{t-1}^2}$$

Por su parte, dado que $\{\varepsilon_t\}: WN - N\left(0, \sigma_\varepsilon^2\right)$, de modo que $Y_t: N\left(0, \dfrac{\sigma_\varepsilon^2}{1 - \phi_1^2}\right)$, queda definida la distribución de probabilidades conjunta de la variable n-dimensional $(Y_1, \varepsilon_2, \varepsilon_3, ..., \varepsilon_n)$ (donde las variables marginales son independientes entre sí)[91].

$\gamma_\tau(Y) = \gamma_=(Y) e^{\delta\tau}\ (\delta > 0)$ (Ver Eichorn, W. (1978)). Comparando esta expresión con la correspondiente a los procesos $AR(1)$ de parámetro discreto, se puede concluir que $\delta = \ln\left(\phi_1^{-1}\right)$.

[91]. Además, se verificará que $\left(Y / Y_{t-1}\right): N\left(\phi_0 + \phi_1 Y_{t-1}; \sigma^2\left(Y_t / Y_1, ..., Y_{t-1}\right)\right)$, donde $\sigma^2\left(Y_t / Y_1, ..., Y_{t-1}\right) = \gamma_0(Y) - \dfrac{\gamma_1^2(Y)}{\gamma_0(Y)}$.

$$dF(Y_1,\varepsilon_2,...,\varepsilon_n) = \frac{\sqrt{1-\phi_1^2}}{\left(\sigma_\varepsilon\sqrt{2\pi}\right)^n}\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[\left(1-\phi_1^2\right)Y_1^2+\sum_{t=2}^{n}\varepsilon_t^2\right]\right\}dY_1\,d\varepsilon_2...d\varepsilon_n$$

Teniendo en cuenta que $\varepsilon_t = Y_t - \phi_1 Y_{t-1}$ $(t = 2,3,...,n)$ y que el Jacobiano de la transformación es igual a la unidad:

$$\left|\frac{\partial(Y_1,\varepsilon_2,...,\varepsilon_n)}{\partial(Y_1,Y_2,...,Y_n)}\right| = \begin{vmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ -\phi_1 & 1 & 0 & ... & 0 & 0 \\ 0 & -\phi_1 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -\phi_1 & 1 \end{vmatrix} = 1$$

se obtiene la función de verosimilitud:

$$L = dF(Y_1,Y_2,...,Y_n) = \frac{\sqrt{1-\phi_1^2}}{\left(\sigma_\varepsilon\sqrt{2\pi}\right)^n}\exp\left[-\frac{1}{2\sigma_\varepsilon^2}\left(\sum_{t=1}^{n}Y_t^2+\phi_1^2\sum_{t=2}^{n-1}Y_{t-1}^2-2\phi_1\sum_{t=2}^{n}Y_{t-1}Y_t\right)\right]dY_1\,dY_2...dY_n$$

cuyo logaritmo será de la forma:

$$\ell = \ln(L) = \frac{1}{2}\ln\left(1-\phi_1^2\right)-n\ln\left(\sigma_\varepsilon\sqrt{2\pi}\right)-\frac{1}{2\sigma_\varepsilon^2}\left(\sum_{t=1}^{n}Y_t^2+\phi_1^2\sum_{t=2}^{n-1}Y_{t-1}^2-2\phi_1\sum_{t=2}^{n}Y_{t-1}Y_t\right)dY_1\,dY_2...dY_n$$

Por lo tanto:

$$\frac{\partial\ell}{\partial\phi_1} = -\frac{\phi_1}{1-\phi_1^2}-\frac{1}{\sigma_\varepsilon^2}\left(\phi_1\sum_{t=2}^{n-1}Y_t^2-\sum_{t=2}^{n}Y_{t-1}Y_t\right)$$

$$\frac{\partial\ell}{\partial\sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2}-\frac{1}{2\sigma_\varepsilon^4}\left(\sum_{t=1}^{n}Y_t^2+\phi_1^2\sum_{t=2}^{n-1}Y_t^2-2\sum_{t=2}^{n}Y_{t-1}Y_t\right)$$

Igualando a cero la primera ecuación y sin tomar en consideración el primer sumando (que es de orden $O\left(\frac{1}{n}\right)$), se obtiene el **estimador máximo-verosímil exacto** de ϕ_1, $\hat{\phi}_1^{(MV)} = \dfrac{\sum_{t=2}^{n}Y_{t-1}Y_t}{\sum_{t=2}^{n-1}Y_t^2}$ (que es asintóticamente equivalente al estimador mínimo-cuadrático el cual, a su vez, coincide con el estimador máximo-verosímil condicionado por el valor de la realización Y_1).

En este caso particular en el que $\{Y_t\}$ admite una representación $AR(1)$, fue posible obtener una estimación máximo-verosímil exacta en forma relativamente simple. En el caso general del $AR(p)$ se puede optar por la alternativa más simple de una estimación máximo-verosímil condicionada.

Sin considerar los términos de orden $O\left(\frac{1}{n}\right)$, se obtiene que

$$E\left(\frac{\partial^2 \ell}{\partial \phi_1^2}\right) = -\frac{E\left(\sum_{t=2}^{n-1} Y_t^2\right)}{\sigma_\varepsilon^2} = -\frac{n}{1-\phi_1^2}$$

Es decir, que $\sigma^2\left(\hat{\phi}_1\right) = \frac{1-\phi_1^2}{n}$. La función de densidad espectral será, entonces, de la forma:

$$\omega_Y(\alpha) = \frac{1}{2\pi}\left[\gamma_0(Y) + \gamma_0(Y)\sum_{j=1}^{\infty}\phi_1^j e^{-i\alpha j} + \gamma_0(Y)\sum_{j=1}^{\infty}\phi_1^j e^{i\alpha j}\right] =$$

$$= \frac{\gamma_0(Y)}{2\pi}\left(1 + \frac{\phi_1 e^{-i\alpha}}{1-\phi_1 e^{-i\alpha}} + \frac{\phi_1 e^{i\alpha}}{1-\phi_1 e^{i\alpha}}\right) =$$

$$= \frac{\gamma_0(Y)}{2\pi}\frac{\left(1-\phi_1 e^{-i\alpha}\right)\left(1-\phi_1 e^{i\alpha}\right) + \phi_1 e^{-i\alpha}\left(1-\phi_1 e^{i\alpha}\right) + \phi_1 e^{i\alpha}\left(1-\phi_1 e^{-i\alpha}\right)}{\left(1-\phi_1 e^{-i\alpha}\right)\left(1-\phi_1 e^{i\alpha}\right)} =$$

$$= \frac{\gamma_0(Y)}{2\pi}\frac{1-\phi_1^2}{1+\phi_1^2 - 2\phi_1\cos(\alpha)}$$

A pesar de su capacidad para representar adecuadamente muchos fenómenos reales, los procesos autorregresivos de primer orden presentan algunos inconvenientes originados en su propia simplicidad: su estructura temporal tiende a cero en modo oscilante o monótono, según que ϕ_1 sea, respectivamente, menor o mayor que cero, por lo tanto, no sirven para representar las hipótesis de fenómenos alternantes en el tiempo alrededor de una tendencia determinística con una variabilidad más o menos constante o con una estructura cíclica rigurosamente periódica.

Teniendo en cuenta la expresión correspondiente a σ_ε^2 dada en la Sec. 2.10.2 y recurriendo a la fórmula del coeficiente de determinación que se utiliza en los modelos estáticos[92], se puede escribir:

$$R_1^2 = 1 - \frac{\hat{\sigma}_\varepsilon^2\left(AR(1)\right)}{\hat{\sigma}_Y^2} = \hat{\rho}_1(Y)\hat{\phi}_1 = \hat{\rho}_1^2(Y) = \hat{\phi}_1^2$$

Es decir, se puede obtener una estimación de la capacidad de predicción del modelo a partir del estimador de $\rho_1(Y)$.

Obsérvese que: **i)** en este caso de procesos estacionarios $AR(1)$, el R_1^2 se obtiene de comparar la variación de $\{Y_t\}$ explicada por su autorregresividad de orden p con la variación total de $\{Y_t\}$, esta última puede ser interpretada como la parte de la variación de $\{Y_t\}$ que se obtiene de aplicar como modelo el valor medio de las observaciones ($\bar{Y}$); **ii)** si $\phi_1 = 1$, entonces $R_1^2 = 1$ (es decir que, en los límites de la región de estacionariedad, el coeficiente R_1^2 pierde su utilidad como medida de la capacidad explicativa del modelo $AR(1)$ respecto del comportamiento del proceso $\{Y_t\}$).

---

[92]. Estos modelos serán tratados en al Cap. 6.

Ejemplo 2.3:

Sea un proceso $\{Y_t\}$ expresado en términos trimestrales, para cuyo análisis se cuenta con una serie cronológica que abarca el período [3er.TRIM-1996;4to.TRIM-2006] y cuya trayectoria figura en el **GRAFICO 2.5.**

A partir de estos datos se obtuvieron los siguientes estimadores de la función de autocorrelaciones:

j	$\hat{\rho}_j(Y)$	j	$\hat{\varphi}_{jj}(Y)$
1	0,570	6	-0,048
2	0,352	7	0,009
3	0,247	8	-0,053
4	0,161	9	-0,048
5	0,014	10	-0,074

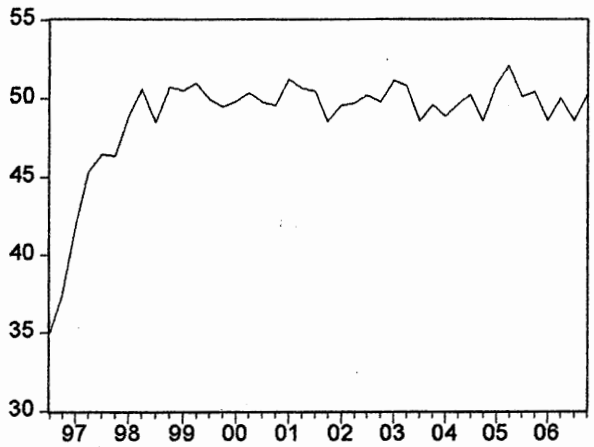
La comparación de los correspondientes valores del estadístico $Q^{(LBP)}$ que figura en el siguiente cuadro:

j	$Q_j^{(LBP)}$	j	$Q_j^{(LBP)}$
1	14.322	6	24.114
2	19.922	7	24.118
3	22.753	8	24.271
4	23.989	9	24.397
5	23.999	10	24.708

con los valores críticos que proporciona la tabla $\chi^2_{(j)}(0.05)$:

j	$\chi^2_{(j)}$	j	$\chi^2_{(j)}$
1	3.481	6	12.590
2	5.991	7	14.070
3	7.815	8	15.510
4	9.488	9	16.920
5	11.070	10	18.310

GRAFICO 2.5: Trayectoria del proceso $\{Y_t\}$



permite concluir, con un 95% de confiabilidad, que el proceso $\{Y_t\}$ presenta autorregresividades significativas.

A partir de los valores estimados de los $\rho_j(Y)$, de acuerdo con el sistema de ecuaciones de Yule-Walker y utilizando el algoritmo de Durbin-Levinson, se obtienen los siguientes estimadores de la función autocorrelaciones parciales:

$$\hat{\phi}_{11}(Y) = \hat{\rho}_1(Y) = 0.570$$

$$\hat{\phi}_{22}(Y) = \frac{\hat{\rho}_2(Y) - \rho_1^2(Y)}{1 - \rho_1^2(Y)} = \frac{0.352 - 0.325}{1 - 0.235} = 0.040$$

Del mismo sistema de Yule-Walker se obtiene, además, que:

$$\hat{\phi}_{21} = \frac{\hat{\rho}_1(Y) - \hat{\rho}_1(Y)\hat{\rho}_2(Y)}{1 - \rho_1^2(Y)} = 0.547$$

De modo que:

$$\hat{\phi}_{33}(Y) = \frac{\hat{\rho}_3(Y) - \hat{\phi}_{21}\hat{\rho}_2(Y) - \hat{\phi}_{22}(Y)\rho_1(Y)}{1 - \hat{\phi}_{21}\hat{\rho}_1(Y) - \hat{\phi}_{22}(Y)\hat{\rho}_2(Y)} = 0.046$$

De la misma forma, se obtiene que:

$$\hat{\phi}_{44}(Y) = \frac{\hat{\rho}_4(Y)\hat{\phi}_{31}\hat{\rho}_3(Y) - \hat{\phi}_{32}\hat{\rho}_2(Y) - \hat{\phi}_{33}(Y)\hat{\rho}_1(Y)}{1 - \hat{\phi}_{31}\hat{\rho}_1(Y) - \hat{\phi}_{32}\hat{\rho}_2(Y) - \hat{\phi}_{33}(Y)\hat{\rho}_3(Y)} = -0.008$$

Donde:

$$\hat{\phi}_{31} = \hat{\phi}_{21} - \hat{\phi}_{22}(Y)\hat{\phi}_{33}(Y) = 0.547 - (0.046)(0.040) = 0.545$$

$$\hat{\phi}_{32} = \hat{\phi}_{22}(Y) - \hat{\phi}_{21}\hat{\phi}_{33}(Y) = 0.040 - (0.046)(0.547) = 0.015$$

Aplicando un criterio similar, se obtiene que:

j	$\hat{\phi}_{jj}(Y)$	j	$\hat{\phi}_{jj}(Y)$
5	-0.135	8	-0.102
6	-0.028	9	0.036
7	0.103	10	-0.069

Teniendo en cuenta que los valores críticos (para un $\alpha = 0.05$) correspondientes al test de significatividad individual de los coeficientes $\phi_{jj}(Y)$ son $\pm \frac{2}{\sqrt{42}} = 0.309$, se puede concluir (con un 5% de confiabilidad) que el único coeficiente significativo es $\phi_{11}(Y)$.

Si se tiene en cuenta, además, que la función de autocorrelaciones decrece más lentamente que la de autocorrelaciones parciales, se puede concluir que, aparentemente, el proceso $\{Y_t\}$ admite una representación $AR(1)$ de la forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$$

A partir de la serie cronológica se obtuvo el siguiente estimador mínimo cuadrático de esta representación:

$$\hat{Y}_t = 49.867 + (0.574)Y_{t-1}$$

2.10.8.- El proceso AR(2)

Sea $\{Y_t\}$ un **proceso autorregresivo de segundo orden (AR(2))** o **proceso de Yule** que admite una representación de la forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$.

La característica fundamental de estos procesos es su capacidad para generar estructuras muy flexibles las cuales, además de poseer el comportamiento inercial de los $AR(1)$, proporcionan soluciones alternativas a partir de la utilización de dos parámetros solamente.

La condición de estacionariedad del $AR(2)$ se verifica cuando las raíces de su ecuación característica:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$$

son, en módulo, mayores que la unidad. Es decir cuando los coeficientes de la representación cumplen con el siguiente sistema de inecuaciones: $\phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1$ y $|\phi_2| < 1$ que define una **región de**

estacionariedad de forma triangular en el plano (ϕ_1,ϕ_2) (ver **GRAFICO 2.6**).

La estructura temporal de este proceso puede ser expresada de la siguiente forma:

$$S(t) = A_1\left(G_1^{-1}\right)^t + A_2\left(G_2^{-1}\right)^t$$

(donde G_1 y G_2 denotan las raíces de la ecuación característica y A_1 y A_2 denotan las constantes calculadas a partir de las condiciones iniciales, las cuales pueden ser expresadas en función de las soluciones G_1 y G_2 y de los valores iniciales Y_1 e Y_2), por lo que se puede concluir que el comportamiento de $S(t)$ depende de la naturaleza de las raíces de la ecuación característica las cuales, dentro de la región de estacionariedad, pueden ser reales y distintas, reales y coincidentes, o complejas, de acuerdo con el valor que asuma el discriminante $D = \phi_1^2 + 4\phi_2$.

La parábola convexa que aparece en el **GRAFICO 2.6** describe el conjunto de valores a asumir por este discriminante y marca el límite de la subclase de parámetros ϕ_1 y ϕ_2 que dan origen a las raíces complejas.

Si **D>0**, el correspondiente proceso presenta un comportamiento monótono decreciente, suma de dos exponenciales decrecientes. Si una de las raíces es negativa y la componente correspondiente a esta raíz supera a la correspondiente a la raíz positiva, entonces el comportamiento será exponencialmente decreciente, pero de forma alternada (a este caso se lo denomina de **raíz negativa dominante**).

Si **D=0**, es decir, si $\phi_1 = \sqrt{-4\phi_2}$, las raíces de la ecuación característica son reales y coincidentes, entonces, la estructura temporal del proceso queda expresada de la siguiente forma:

$$S(t) = \left(A_1 + tA_2\right)\left(G^{-1}\right)^t$$

Donde $G^{-1} = G_1^{-1} = G_2^{-1} = \dfrac{\phi_1}{2}$. Es decir, $S(t) = \left(A_1 + tA_2\right)\left(\dfrac{\phi_1}{2}\right)^t$

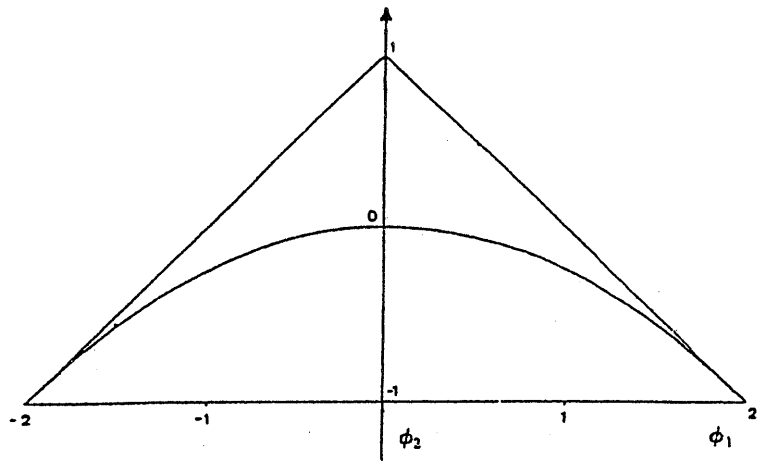
Suponiendo, por comodidad, la condición $Y_0 = 1$, será:

$$S(t) = \left[1 + \left(Y_1\frac{2}{\phi_1} - 1\right)t\right]\left(\frac{\phi_1}{2}\right)^t$$

En este caso, la estructura temporal puede ser asimilada al comportamiento de una función compuesta por una función lineal de la forma $(1+\alpha t)$ (donde $\alpha = Y_1\dfrac{2}{\phi_1} - 1$) creciente o decreciente con respecto a t según que α sea positivo o negativo (es decir, según que ϕ_1 sea menor o mayor que $2Y_1$) y por una función exponencial-decreciente de la forma $\left(\dfrac{\phi_1}{2}\right)^t = \phi^t$, cuyo decrecimiento será monótono o alternado, según que ϕ sea positivo o negativo (debe tenerse en cuenta que, de acuerdo con las condiciones de estacionariedad ya demostradas, $|\phi| = \left|\dfrac{\phi_1}{2}\right| < 1$).

Obsérvese que, de acuerdo con la condición considerada más arriba, $S(0) = 1$ para cualquier valor de α y $S(t) = (1+\alpha t)\left(\dfrac{\phi_1}{2}\right)^t = 0$ para $t^* = -\dfrac{1}{\alpha}$, solución que será positiva o negativa según que α, sea

GRAFICO 2.6: Región de estacionariedad para los coeficientes Φ_1 y Φ_2



menor o mayor que cero[93]. Además se verifica que:

$$\lim_{t\to-\infty} S(t) = \begin{cases} -\infty & si \quad \alpha > 0 \\ +\infty & si \quad \alpha < 0 \end{cases}$$

y que

$$\lim_{t\to+\infty} S(t) = 0$$

para cualquier valor de α (el límite será desde valores positivos o negativos, según que α sea mayor o menor que cero)[94].

Se demuestra en forma inmediata que las derivadas sucesivas de la función $S(t)$ son de la forma:

$$S^{(k)}(t) = \left(\frac{\phi_1}{2}\right)^t \left[\ln\left(\frac{\phi}{2}\right)^{k-1}\left(k\alpha + (1+\alpha t)\ln\left(\frac{\phi_1}{2}\right)\right)\right] \qquad (k = 0,1,2...)$$

y que, por lo tanto, $S(t)$ presenta un extremo en el valor de t para el cual se anula la derivada primera:

$$\left(\frac{\phi_1}{2}\right)^t \left[\alpha + (1+\alpha t)\ln\left(\frac{\phi_1}{2}\right)\right] = 0$$

---

[93]. Otra solución de esta ecuación se obtiene para $\phi_1 = 2$ pero, en este caso la representación $AR(2)$ será no-estacionaria.

[94]. De la observación del **GRAFICO 2.7** se puede concluir que la única hipótesis atendible desde el punto de vista económico es que $\alpha > 0$.

La única solución finita de esta ecuación es[95]:

$$t_0 = -\frac{1}{\alpha} - \frac{1}{\ln\left(\frac{\phi_1}{2}\right)} = -\frac{\alpha + \ln\left(\frac{\phi_1}{2}\right)}{\alpha \ln\left(\frac{\phi_1}{2}\right)}$$

Para $t = t_0$ el valor de la derivada segunda, $S^{(2)}(t_0) = \left(\frac{\phi_1}{2}\right)^{t_0} \alpha \ln\left(\frac{\phi_1}{2}\right)$, será positivo, nulo o negativo según que α sea negativo, nulo o positivo (es decir que en t_0 presentará un máximo o un mínimo, según que α sea negativo o positivo).

Sea $t^{(i)}$ el valor de t para el cual se anula la derivada segunda, si se verifica que la derivada tercera:

$$S^{(3)}\left(t^{(i)}\right) = \left(\frac{\phi_1}{2}\right)^{t^{(i)}} \left[\ln\left(\frac{\phi_1}{2}\right)\right]^2 \left[3\alpha + \left(1 + \alpha t^{(i)}\right) \ln\left(\frac{\phi_1}{2}\right)\right] \neq 0$$

entonces se puede concluir que para ese valor de t, la función $S(t)$ presenta un punto de inflexión. Asimismo, se demuestra en forma inmediata que:

$$t^{(i)} = -\frac{1}{\alpha} - \frac{2}{\ln\left(\frac{\phi_1}{2}\right)} = -\frac{2\alpha + \ln\left(\frac{\phi_1}{2}\right)}{\alpha \ln\left(\frac{\phi_1}{2}\right)} = t_0 - \frac{1}{\ln\left(\frac{\phi_1}{2}\right)} = t_0 + \phi^*$$

donde $\phi^* = \left(\ln\left(\frac{\phi}{2}\right)\right)^{-1} > 0$. De modo que se puede asegurar que el punto de inflexión está ubicado siempre a la derecha del punto extremo ($t^{(i)} > t_0$).

El valor de la derivada tercera en el punto $t^{(i)}$ será positivo o negativo, según que α sea mayor o menor que cero. Lo cual demuestra que la inflexión será ascendente o descendente, según que α sea positivo o negativo.

Debe tenerse en cuenta que si $\phi_1 < 2e^{-\alpha}$ $(\alpha > 0)$ entonces el máximo de $S(t)$ estará ubicado en el semieje correspondiente a los valores de $t < 0$. De modo que la correspondiente serie cronológica (que existe para todo $t > 0$) no observará el cambio de tendencia planteado más arriba. De la misma forma se demuestra que, si $\phi_1 < 2e^{-2\alpha}$ $(\alpha > 0)$, el punto de inflexión también estará ubicado en el semieje correspondiente a $t < 0$.

En los Gráficos 2.7 figuran las distintas formas de comportamiento de un proceso $AR(2)$ con discriminante nulo.

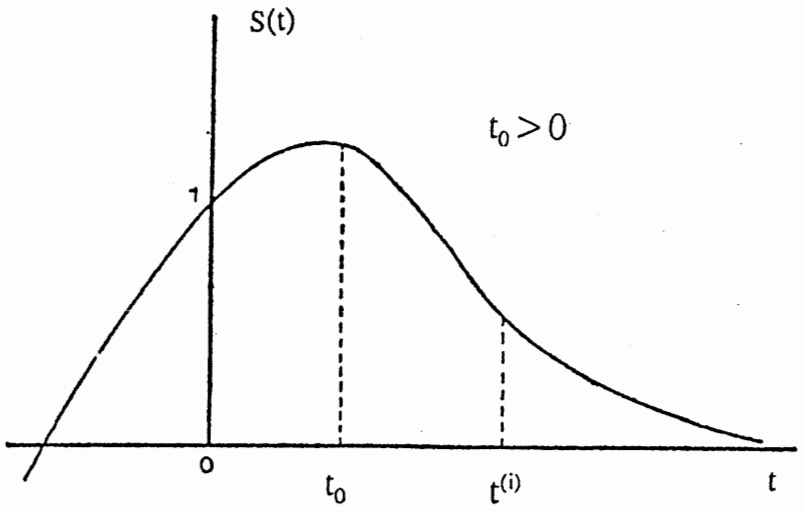
Si **D<0**, las dos raíces de la ecuación característica son complejas. Esta situación -cuya importancia radica en la posibilidad de expresar la estructura temporal del proceso bajo la forma de una

---

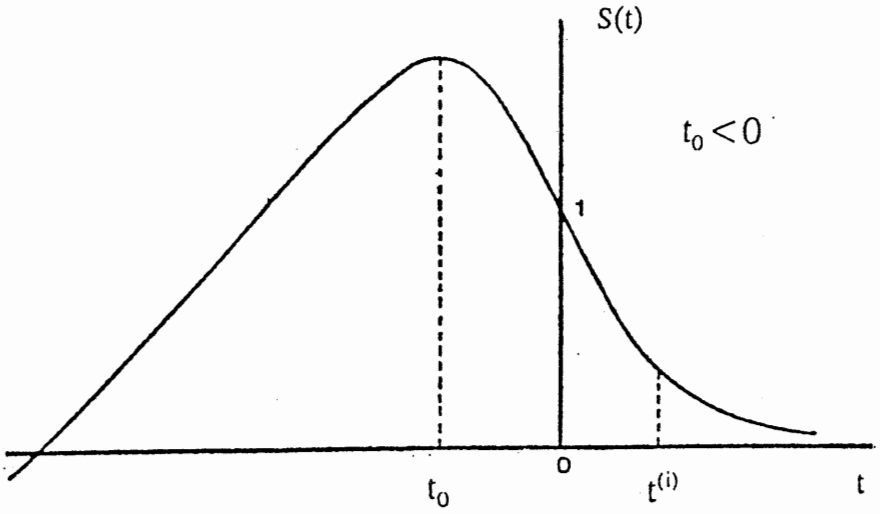
[95]. Se supondrá que $\frac{\phi_1}{2} > 0$, en caso contrario el factor $\ln\left(\frac{\phi_1}{2}\right)$ carecería de sentido.

GRAFICO 2.7.- Comportamiento de un proceso AR(2) con discriminante nulo

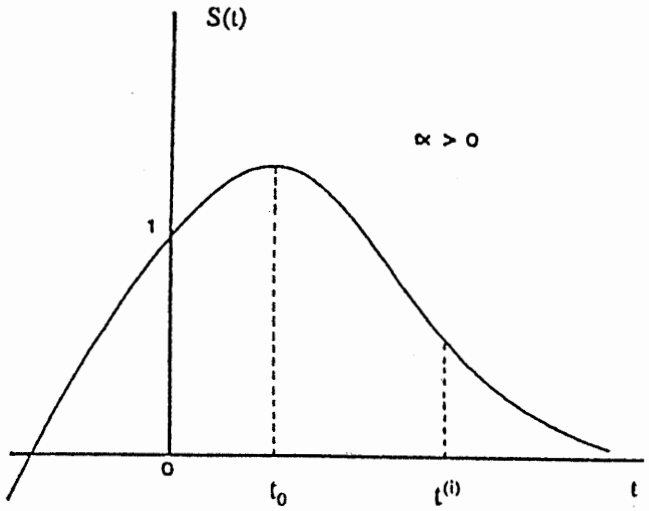
(a)



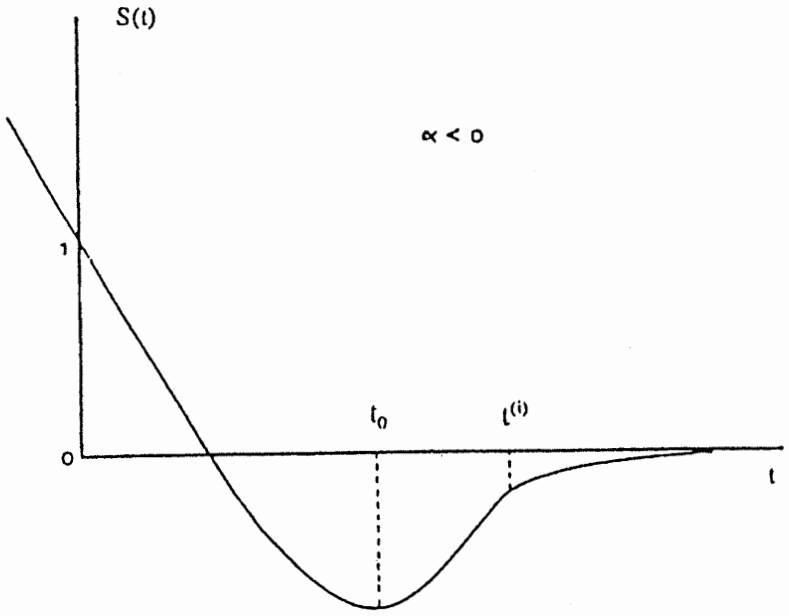
(b)



(c)



(d)



función sinusoidal[96]- se presenta cuando se verifican simultáneamente las relaciones $\phi_2 < 0$ y $|\phi_2| > \frac{\phi_1^2}{4}$.

En este caso, la estructura temporal del proceso puede ser expresada de la siguiente forma:

$$S(t) = A_1\left(G_1^{-1}\right)^t + A_2\left(G_2^{-1}\right)^t = A_1(\alpha + i\beta)^{-t} + A_2(\alpha - i\beta)^{-t} \qquad (t > 0)$$

(donde $\alpha = -\frac{\phi_1}{2\phi_2}$ y $\beta = \frac{\sqrt{-\left(\phi_1^2 + 4\phi_2\right)}}{2\phi_2}$). De acuerdo con la representación trigonométrica de los números complejos, se puede escribir:

$$S(t) = A_1\left[d\left(\cos(\theta t) + i\,sen(\theta t)\right)\right]^{-t} + A_2\left[d\left(\cos(\theta t) - i\,sen(\theta t)\right)\right]^{-t} =$$

$$= d^{-t}\left[\left(A_1 + A_2\right)\cos(\theta t) + i\left(A_2 - A_1\right)sen(\theta t)\right]$$

(donde, $d^2 = \alpha^2 + \beta^2 = -\frac{1}{\phi_2}$, $\theta = arco\,tg\left(-\frac{\beta}{\alpha}\right) = arco\,tg\left(\frac{\sqrt{-\left(\phi_1^2 + \phi_2\right)}}{|\phi_1|}\right) = arco\,tg\left(\frac{\sqrt{-D}}{|\phi_1|}\right)$, y, como ya se mencionó, $D = \phi_1^2 + 4\phi 2$ es el discriminante de la ecuación característica). El primer miembro de esta expresión debe ser real para todo t. La condición necesaria y suficiente para que el segundo miembro sea real para valores cualesquiera de Y_1 e Y_2, es que A_1 y A_2 sean complejos conjugados, $A_1 = a + ib$ y $A_2 = a - ib$. Luego, se puede escibir (ver Jenkins, G.M.; Watts, D.G. (1969), Kendall, M.G. (1973), Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. (1994)):

$$S(t) = d^{-t}\left[2a\cos(\theta t) - i2b\,sen(\theta t)\right] = \frac{\left[signo(\phi_1)\right]^t d^{-t}\,sen(2\pi\theta + \delta)}{sen(\delta)}$$

(donde, $signo(\phi_1) = 1$, *si* $\phi_1 > 0$ y $signo(\phi_1) = -1$, *si* $\phi_1 < 0$, θ es tal que $\cos(2\pi\theta) = \cos(\theta) = \frac{\phi_1}{2\sqrt{-\phi_2}}$ y δ es tal que $tg(\delta) = \frac{1 + d^{-2}}{1 - d^{-2}} tg(2\pi\theta^*)$, es decir, que $\delta = arco\,tg\left[\frac{1 + d^{-2}}{1 - d^{-2}} tg(2\pi\theta)\right]$). Se puede concluir, entonces que, en el caso en que raíces sean complejas, la estructura temporal del $AR(2)$ se comporta como una función sinusoidal amortiguada donde $d^{-1} = \sqrt{-\phi_2} < 1$ $(-1 < \phi_2 < 0)$ es el **factor de amortiguación** (debe tenerse en cuenta que, como $D < 0$, será

$$\frac{\phi_1}{2} < \sqrt{-\phi_2} = d^{-1}$$

y, en consecuencia que, a menos que $-\phi_2$ sea próximo a la unidad, este comportamiento periódico amortiguado tiende a anularse rápidamente.

Las partes componentes de esta función sinusoidal representativa de la estructura temporal son:

---

[96]. Es decir, la importancia de este caso radica en la capacidad que posee la estructura temporal del $AR(2)$ de contener una parte componente cíclica cuya amplitud, período y fase, como se verá más adelante, son funciones explícitas de los mismos parámetros del proceso.

i) La **amplitud:**

$$a(t) = \frac{\left(signo(\phi_1)\right)^t d^{-t}}{sen(\delta)}$$

que es decreciente en valor absoluto (con signo alternado si $\phi_1 < 0$).

ii) La **frecuencia**:

$$\theta = \frac{1}{2\pi} arco\ tg\left(\frac{\sqrt{-D}}{|\phi_1|}\right) = \frac{1}{2\pi} arco\cos\left(\frac{|\phi_1|}{2\sqrt{-\phi_2}}\right)$$

que depende exclusivamente de los coeficientes ϕ_1 y ϕ_2.

iii) El **período**, que es la inversa de la frecuencia (y que está expresado en unidades temporales: días, meses, años, etc):

$$T = \frac{1}{\theta} = \frac{2\pi}{\theta^*} = \frac{2\pi}{arco\cos\left(\frac{|\phi_1|}{2\sqrt{-\phi_2}}\right)}$$

Desde un punto de vista estrictamente teórico se debe verificar que $0 < \theta < \pi$ y, por lo tanto, que $0 < \theta^* < \frac{1}{2}$, es decir, que $T > 2$. Pero, teniendo en cuenta que, si $D < 0$, entonces $0 < \frac{|\phi_1|}{2\sqrt{-\phi_2}} < 1$ y, en consecuencia,

$0 < arco\cos\left(\frac{|\phi_1|}{2\sqrt{-\phi_2}}\right) = \theta < \frac{\pi}{2}$ entonces, resulta que $T > 4$. De acuerdo con la definición de período se comprueba que, si ϕ_1 se mantiene fijo, T disminuye a medida que $-\phi_2$ crece de cero a uno (es decir, que T varía en forma inversamente proporcional al factor de amortiguación d) y,. viceversa, si $-\phi_2$ se mantiene fijo, T aumenta en forma directamente proporcional a $|\phi_1|$.

Una cuestión que puede resultar interesante en las aplicaciones consiste en hallar el conjunto de pares de valores (ϕ_1,ϕ_2) incluidos en la clase $C(\phi_1,\phi_2)$ de coeficientes de la representación $AR(2)$ tales que $D < 0$ y que dan lugar a un período T dado. A partir de la definición del período, se obtiene que $\phi_1 = 2\sqrt{-\phi_2}\cos\left(\frac{2\pi}{T}\right)$, de modo que, asignando un valor a T y haciendo variar $-\phi_2$ en el intervalo $[0;1]$, se puede concluir que

$$0 < \phi_1 < 2\cos\left(\frac{2\pi}{T}\right)$$

Es decir que, dado un valor de T, la subclase de coeficientes $T(\phi_1,\phi_2) \subset C(\phi_1,\phi_2)$ que dan origen a dicho comportamiento periódico estará formada por los puntos $\phi_1 \in \left[0,2\cos\left(\frac{2\pi}{T}\right)\right]$ y $\phi_2 \in [-1;0]$ (se demuestra que, por ejemplo, para $T = 6$, debe ser $\phi_1 < 1$ y $\phi_2 = -\phi_1^2$).

iv) La **fase:**

$$\delta = arco\ tg\left(\frac{1-\phi_2}{1+\phi_2}\frac{\sqrt{-D}}{|\phi_1|}\right)$$

De esta expresión se puede concluir que: **i)** para que $\delta = 0$, se debe verificar que $\phi_2 = 1$ ó que $D = \phi_1^2 + 4\phi_2 = 0$, ahora bien, si $\phi_2 = 1$, entonces $D > 0$ y, en ese caso, la ecuación característica no posee raíces complejas y **ii)** si $D < 0$, necesariamente $\delta \neq 0$.

De acuerdo a lo demostrado en la Sec. 2.10.2, la función de autocorrelaciones del $AR(2)$ satisface el siguiente sistema de ecuaciones en diferencias:

$$\rho_j(Y) = \phi_1\rho_{j-1}(Y) + \phi_2\rho_{j-2}(Y) \qquad (j > 0)$$

De la solución del sistema de ecuaciones de Yule-Walker de orden dos:

$$\begin{cases}\rho_1(Y) = \phi_1 + \phi_2\rho_2(Y)\\ \rho_2(Y) = \phi_1\rho_1(Y) + \phi_2\end{cases}$$

se obtienen los valores iniciales $\rho_1(Y) = \frac{\phi_1}{1-\phi_2}$ y $\rho_2(Y) = \phi_2 + \frac{\phi_1^2}{1-\phi_2}$, que permiten expresar los valores de $\rho_j(Y)\,(j = 3,4,...)$ en función de los coeficientes de la representación $AR(2)$ de la siguiente forma:

$$\rho_3(Y) = \rho_1(Y)\rho_2(Y) + \phi_2\left[\rho_1(Y) - \rho_1(Y)\rho_2(Y)\right] = \frac{\phi_1^3 + \phi_1\phi_2(2-\phi_2)}{1-\phi_2}$$

$$\rho_4(Y) = \rho_1(Y)\rho_3(Y) + \phi_2\left[\rho_2(Y) - \rho_1(Y)\rho_3(Y)\right] = \frac{\phi_1^4 - \phi_2^3 + \phi_2^2(3-\phi_2)}{1-\phi_2}$$

.....

En forma resumida:

$$\rho_j(Y) = \rho_1(Y)\rho_{j-1}(Y) + \phi_2\left[\rho_{j-2}(Y) - \rho_1(Y)\rho_{j-1}(Y)\right] \qquad (j = 3,4,...)$$

Esta relación demuestra que la autocorrelación de este proceso puede ser desagregada en dos partes componentes: **i)** $\rho_1(Y)\rho_{j-1}(Y)$, que es la correspondiente a un proceso $AR(1)$ y **ii)** una proporción de una parte componente $\left(\rho_{j-2}(Y) - \rho_1(Y)\rho_{j-1}(Y)\right)$, que es la correspondiente la autorregresividad de orden 2, neta de la explicada por la correspondiente a la autorregresividad de orden 1[97]. Además se puede concluir que, manteniendo a ϕ_1 constante, los valores de $\rho_j(Y)$ son inversamente proporcionales a $(1-\phi_2)$, es decir, que aumentan con $-\phi_2$.

---

[97]. Obsérvese que, como se mencionó más arriba, las condiciones iniciales de la ecuación en-diferencias que rige el comportamiento de la función de autocorrelaciones están definidas por los coeficientes $\rho_1(Y)$ y $\rho_2(Y)$ exclusivamente.

Luego, la varianza de $\{Y_t\}$:

$$\sigma_Y^2 = \frac{\sigma_\varepsilon^2}{1-\phi_1\rho_1(Y)-\phi_2\rho_2(Y)} = \frac{\sigma_\varepsilon^2}{1-\rho_1^2(Y)(1-\phi_2)-\rho_2(Y)\phi_2} =$$

$$= \frac{\sigma_\varepsilon^2}{1-\rho_1^2(Y)-\phi_2\left[\rho_2(Y)-\rho_1^2(Y)\right]} = \frac{\sigma_\varepsilon^2}{1-\rho_1^2(Y)-\phi_2^2\left[1-\rho_1^2(Y)\right]} = \frac{\sigma_\varepsilon^2}{\left(1-\rho_1^2(Y)\right)\left(1-\phi_2^2\right)} =$$

$$= \sigma_\varepsilon^2\left(1+\rho_1^2(Y)+\rho_1^4(Y)+...\right)\left(1+\phi_2^2+\phi_2^4+...\right)$$

se puede descomponer en: **i)** una varianza debida a las innovaciones $\left(\sigma_\varepsilon^2\right)$; **ii)** una varianza debida a la autocorrelación de primer orden, $\left(1-\rho_1^2(Y)\right)^{-1}$ y **iii)** una varianza debida a la autocorrelación parcial de segundo orden, $\left(1-\phi_2^2\right)^{-1}$.

En lo que hace a la expresión de las condiciones de estacionariedad en función de los coeficientes de autocorrelación, teniendo en cuenta que $|\phi_2| = \left|\frac{\rho_2(Y)-\rho_1^2(Y)}{1-\rho_1^2(Y)}\right| < 1$, se obtienen las siguientes desigualdades: $|\rho_1(Y)| < 1; |\rho_2(Y)| < 1$ y $\rho_2(Y) > 2\rho_1^2(Y) - 1$. Esta última inecuación demuestra que si el proceso $AR(2)$ es estacionario, el intervalo de variación de $\rho_2(Y)$ no es $[-1;1]$, sino $\left[2\rho_1^2(Y)-1;1\right]$. En el **GRAFICO 2.8** figura la región de estacionariedad en función de los coeficientes de autocorrelación.

Retornando al sistema de ecuaciones de Yule-Walker de segundo orden que figura en la página precedente:

$$\begin{cases} \rho_1(Y) = \phi_1 + \phi_2\rho_1(Y) \\ \rho_2(Y) = \phi_1\rho_1(Y) + \phi_2 \end{cases}$$

se obtiene que:

$$\phi_1 = \frac{\rho_1(Y)-\rho_1(Y)\rho_2(Y)}{1-\rho_1^2(Y)}$$

$$\phi_2 = \frac{\rho_2(Y)-\rho_1^2(Y)}{1-\rho_1^2(Y)}$$

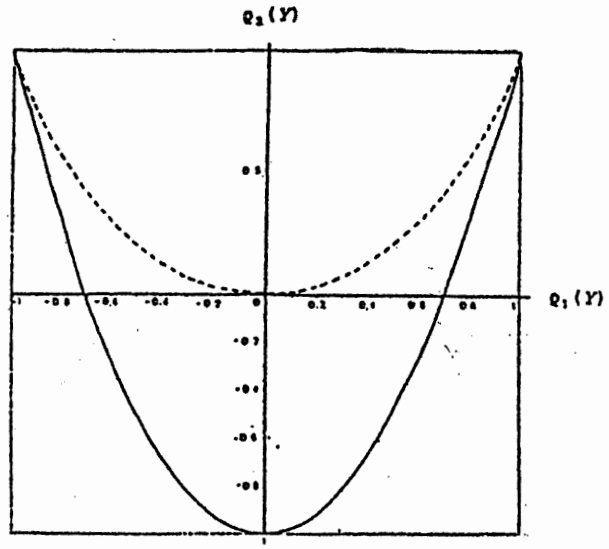
Reemplazando estos resultados en el discriminante de la ecuación característica, para $D < 0$, se obtiene la siguiente inecuación de segundo grado en $\rho_2(Y)$:

$$\phi_1^2 + 4\phi_1 = \left(\frac{\rho_1(Y)-\rho_1(Y)\rho_2(Y)}{1-\rho_1^2(Y)}\right)^2 + 4\left(\frac{\rho_2(Y)-\rho_1^2(Y)}{1-\rho_1^2(Y)}\right) =$$

$$= \rho_2^2(Y) + 2\rho_2(Y)\left(\frac{2}{\rho_1^2(Y)} - 3\right) + \left(4\rho_1^2(Y) - 3\right) < 0$$

que se satisface por los valores internos de sus raíces:

GRAFICO 2.8.- Región de estacionariedad para los coficientes de autocorrelación



$$\rho_2(Y) = -2\left(\frac{2}{\rho_1^2(Y)} - 3\right) \pm \sqrt{4\left(\frac{2}{\rho_1^2(Y)} - 3\right)^2 - 4\left(4\rho_1^2(Y) - 3\right)}$$

y permite demostrar que la región correspondiente a las raíces complejas es la comprendida entre las dos parábolas que figuran en el **GRAFICO 2.8** [98].

Obsérvese que: **i)** $\phi_2 < 0$ es condición necesaria pero no suficiente para que $D < 0$; **ii)** la condición $\rho_2(Y) < \rho_1^2(Y)$ (que deriva de la anterior) también es necesaria pero no suficiente para que la función de autocorrelaciones observe un comportamiento periódico amortiguado.

La función de autocorrelaciones parciales del proceso $AR(2)$ queda definida de la siguiente forma:

$$\phi_{00}(Y) = 1$$

$$\phi_{11}(Y) = \rho_1(Y) = \frac{\phi_1}{1 - \phi_2}$$

$$\phi_{22}(Y) = \frac{\rho_2(Y) - \rho_1^2(Y)}{1 - \rho_1^2(Y)} = \phi_2$$

$$\phi_{jj}(Y) = 0 \qquad (j > 2)$$

A partir de estas expresiones se puede escribir:

$$Y_t = \rho_1(Y)Y_{t-1} + \phi_2\left[Y_{t-2} - \rho_1(Y)Y_{t-1}\right] + \varepsilon_t$$

---

[98]. Obsérvese que ésta inecuación no constituye una condición necesaria para que el proceso $AR(2)$ sea estacionario.

lo que permite concluir que Y_t, en cierta forma, puede ser expresada como una proporción $\rho_1(Y)$ de Y_{t-1}, más una proporción ϕ_2 de Y_{t-2}, neta de la correlación de primer orden ya contenida en Y_{t-1}.

Las condiciones de estacionariedad expresadas en función de los coeficientes de autocorrelación parcial permiten definir las siguientes relaciones:

$$\begin{cases} \phi_{22}(Y) + \phi_{11}(Y) - \phi_{11}(Y)\phi_{22}(Y) < 1 \\ \phi_{22}(Y) - \phi_{11}(Y) + \phi_{11}(Y)\phi_{22}(Y) < 1 \\ |\phi_{22}(Y)| < 1 \end{cases}$$

Dado que las dos primeras condiciones son redundantes, la región de estacionariedad queda definida, entonces, por las desigualdades:

$$\begin{cases} |\phi_{11}(Y)| < 1 \\ |\phi_{22}(Y)| < 1 \end{cases}$$

En particular, la región correspondiente a las raíces complejas está definida por la ecuación:

$$D = \phi_1^2 + 4\phi_2 = \phi_{11}^2(Y)\left(1 - \phi_{22}(Y)\right)^2 + 4\phi_{22}(Y) = \phi_{22}^2(Y) - 2\phi_{22}(Y)\left(1 - \frac{2}{\phi_{11}^2(Y)}\right) + 1 < 0$$

que se satisface por los valores internos de sus raíces:

$$\phi_{22}(Y) = \left(1 - \frac{2}{\phi_{11}^2(Y)}\right) \pm \sqrt{\left(1 - \frac{2}{\phi_{11}^2(Y)}\right)^2 - 1}$$

De modo que la región de estacionariedad queda definida por las relaciones:

$$\left(1 - \frac{2}{\phi_{11}^2(Y)}\right) - \sqrt{\left(1 - \frac{2}{\phi_{11}^2(Y)}\right)^2 - 1} < \phi_{22}(Y) < \left(1 - \frac{2}{\phi_{11}^2(Y)}\right)^2 + \sqrt{\left(1 - \frac{2}{\phi_{11}^2(Y)}\right)^2 - 1}$$

$$1 - \frac{2}{\phi_{11}^2(Y)}\left(1 + \sqrt{1 - \phi_{11}^2(Y)}\right) < \phi_{22}(Y) < 1 + \frac{2}{\phi_{11}^2(Y)}\left(-1 + \sqrt{1 - \phi_{11}^2(Y)}\right)$$

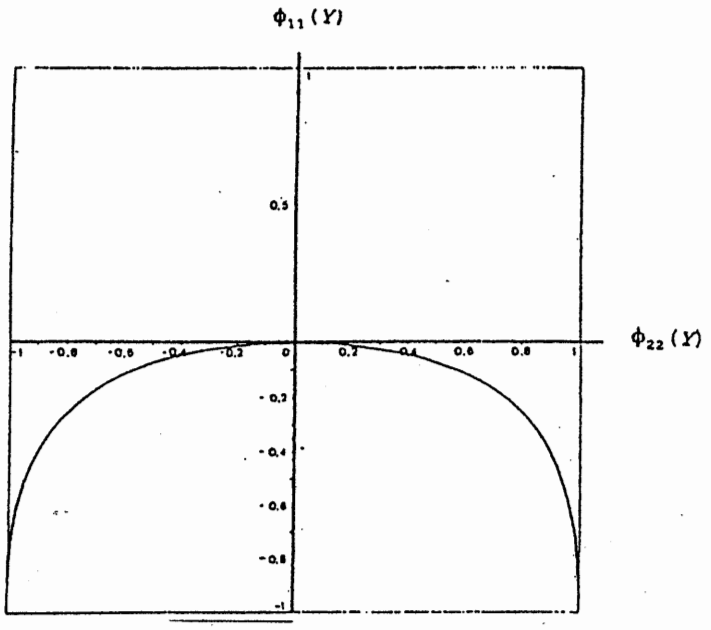
Gráficamente estas inecuaciones están representadas por la región del plano comprendida entre la recta $\phi_{22}(Y) = -1$ y la parábola definida por la ecuación $\phi_{22}^2(Y) - 2\phi_{22}(Y)\left(1 - \frac{2}{\phi_{11}^2(Y)}\right) + 1 = 0$ (ver **Gráfico 2.9**).

De acuerdo a lo demostrado en la Sec. 2.10.3, se obtienen los siguientes **estimadores de Yule-Walker** de los coeficientes de la representación:

$$\hat{\phi}_1^{(YW)} = \frac{\hat{\rho}_1(Y) - \hat{\rho}_1(Y)\hat{\rho}_2(Y)}{1 - \hat{\rho}_1^2(Y)}$$

$$\hat{\phi}_2^{(YW)} = \frac{\hat{\rho}_2(Y) - \hat{\rho}_1^2(Y)}{1 - \hat{\rho}_1^2(Y)}$$

GRAFICO 2.9.- Región de estacionariedad para los coeficientes de autorrelación-parcial



Los correspondientes **estimadores mínimo-cuadráticos** se obtienen de minimizar la expresión:

$$f\left(\hat{\phi}_1^{(CM)}\right) = \sum_{t=3}^{n} \bar{\varepsilon}_t^2 = \sum_{t=3}^{n} \left(Y_t - \hat{\phi}_1^{(CM)} Y_{t-1} - \hat{\phi}_2^{(CM)} Y_{t-2}\right)^2$$

Resolviendo las derivadas

$$\frac{\partial}{\partial \hat{\phi}_1^{(CM)}} f\left(\hat{\phi}_1^{(CM)}, \hat{\phi}_2^{(CM)}\right) = 0$$

y

$$\frac{\partial}{\partial \hat{\phi}_2^{(CM)}} f\left(\hat{\phi}_1^{(CM)}, \hat{\phi}_2^{(CM)}\right) = 0$$

se obtiene el siguiente sistema de ecuaciones:

$$\begin{cases} \hat{\phi}_1^{(CM)} \sum_{t=3}^{n} Y_{t-1}^2 + \hat{\phi}_2^{(CM)} \sum_{t=3}^{n} Y_{t-2} Y_{t-1} = \sum_{t=3}^{n} Y_{t-1} Y_t \\ \hat{\phi}_1^{(CM)} \sum_{t=3}^{n} Y_{t-2} Y_{t-1} + \hat{\phi}_2^{(CM)} \sum_{t=3}^{n} Y_{t-2}^2 = \sum_{t=3}^{n} Y_{t-2} Y_t \end{cases}$$

cuyas soluciones son:

$$\hat{\phi}_1^{(CM)} = \frac{\sum_{t=3}^{n} Y_{t-1}Y_t \sum_{t=3}^{n} Y_{t-2}^2 - \sum_{t=3}^{n} Y_{t-2}Y_t \sum_{t=3}^{n} Y_{t-2}Y_{t-1}}{\sum_{t=3}^{n} Y_{t-1}^2 \sum_{t=3}^{n} Y_{t-2}^2 - \left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1}\right)^2}$$

$$\hat{\phi}_2^{(CM)} = \frac{\sum_{t=3}^{n} Y_{t-2}Y_t \sum_{t=3}^{n} Y_{t-1}^2 - \sum_{t=3}^{n} Y_{t-1}Y_t \sum_{t=3}^{n} Y_{t-2}Y_{t-1}}{\sum_{t=3}^{n} Y_{t-1}^2 \sum_{t=3}^{n} Y_{t-2}^2 - \left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1}\right)^2}$$

Dividiendo en estas igualdades numerador y denominador por $n^2\gamma_0^2(Y)$, se obtienen las siguientes expresiones para los estimadores mínimo-cuadráticos:

$$\hat{\phi}_1^{(CM)} = \frac{\left(\hat{\rho}_1(Y) - \frac{Y_1Y_2}{n\hat{\gamma}_0(Y)}\right)\left(1 - \frac{Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)}\right) - \hat{\rho}_2(Y)\left(\hat{\rho}_1(Y) - \frac{Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)}\right)}{\left(1 - \frac{Y_1^2 + Y_n^2}{n\hat{\gamma}_0(Y)}\right)\left(1 - \frac{Y_{n-1}^2Y_n^2}{n\hat{\gamma}_0(Y)}\right) - \left(\hat{\rho}_1(Y) - \frac{Y_{n-1}Y_n}{n\gamma_0(Y)}\right)^2}$$

$$\hat{\phi}_2^{(CM)} = \frac{\hat{\rho}_2(Y)\left(1 - \frac{Y_1^2 + Y_n^2}{n\hat{\gamma}_0(Y)}\right) - \left(\hat{\rho}_1(Y) - \frac{Y_1Y_2}{n\hat{\gamma}_0(Y)}\right)\left(\hat{\rho}_1(Y) - \frac{Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)}\right)}{\left(1 - \frac{Y_1^2 + Y_n^2}{n\hat{\gamma}_0(Y)}\right)\left(1 - \frac{Y_{n-1}^2 + Y_n^2}{n\hat{\gamma}_0(Y)}\right) - \left(\hat{\rho}_1(Y) - \frac{Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)}\right)^2}$$

a partir de las cuales se concluye fácilmente que, para valores de n suficientemente grandes y suponiendo que el proceso $\{Y_t\}$ sea suficientemente homogéneo (es decir que los valores Y_1, Y_2, Y_{n-1} e Y_n no sean "demasiado" diferentes de los demás), los estimadores de Yule-Walker son aproximadamente iguales a los mínimo-cuadráticos.

Los estimadores máximo-verosímiles condicionados por los valores de las realizaciones $Y_1, Y_2, \ldots, Y_n$ (suponiendo que $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$ y no tomando en consideración los términos del tipo $O\left(\frac{1}{n}\right)$) pueden ser expresados aproximadamente de la siguiente forma:

$$\hat{\phi}_1^{(MV)} = \frac{\left(\frac{n}{n-1}\sum_{t=2}^{n} Y_{t-1}Y_t\right)\left(\frac{n}{n-4}\sum_{t=5}^{n} Y_{t-2}^2\right) - \left(\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-2}Y_t\right)\left(\frac{n}{n-3}\sum_{t=4}^{n} Y_{t-2}Y_{t-1}\right)}{\left(\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2\right)\left(\frac{n}{n-4}\sum_{t=5}^{n} Y_{t-2}^2\right) - \left(\frac{n}{n-3}\sum_{t=4}^{n} Y_{t-2}Y_{t-1}\right)^2} =$$

$$= \frac{\frac{n}{n-1}\left(\sum_{t=3}^{n} Y_{t-1}Y_t\right)\frac{n}{n-4}\left(\sum_{t=3}^{n} Y_{t-2}^2 - Y_1^2 - Y_2^2\right) - \frac{n}{n-2}\sum_{t=3}^{n} Y_{t-2}Y_t \frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right)}{\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2 \frac{n}{n-4}\left(\sum_{t=3}^{n} Y_{t-2}^2 - Y_1^2 - Y_2^2\right) - \left[\frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right)\right]^2}$$

$$\hat{\phi}_2^{(MV)} = \frac{\left(\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-2}Y_{t-1}\right)\left(\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2\right) - \left(\frac{n}{n-1}\sum_{t=2}^{n} Y_{t-1}Y_t\right)\left(\frac{n}{n-3}\sum_{t=3}^{n} Y_{t-2}Y_{t-1}\right)}{\left(\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2\right)\left(\frac{n}{n-4}\sum_{t=5}^{n} Y_{t-2}^2\right) - \left(\frac{n}{n-3}\sum_{t=4}^{n} Y_{t-2}Y_{t-1}\right)^2}$$

$$= \frac{\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-2}Y_t \frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2 - \frac{n}{n-1}\left(\sum_{t=3}^{n} Y_{t-1}Y_t + Y_1Y_2\right)\frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right)}{\frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2 \frac{n}{n-4}\left(\sum_{t=3}^{n} Y_{t-2}^2 - Y_1^2 - Y_2^2\right) - \left[\frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right)\right]^2}$$

Estas expresiones son la solución de un sistema de ecuaciones lineales de la forma:

$$\begin{cases} \hat{\phi}_1^{(MV)} \frac{n}{n-2}\sum_{t=3}^{n} Y_{t-1}^2 + \hat{\phi}_2^{(MV)} \frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right) = \frac{n}{n-1}\left(\sum_{t=3}^{n} Y_{t-1}Y_t + Y_1Y_2\right) \\ \hat{\phi}_1^{(MV)} \frac{n}{n-3}\left(\sum_{t=3}^{n} Y_{t-2}Y_{t-1} - Y_1Y_2\right) + \hat{\phi}_2^{(MV)} \frac{n}{n-4}\left(\sum_{t=3}^{n} Y_{t-2}^2 - Y_1^2 - Y_2^2\right) = \frac{n}{n-2}\sum_{t=3}^{n} Y_{t-2}Y_t \end{cases}$$

Luego, se puede concluir que, cuando n aumenta indefinidamente y para valores de Y_1 e Y_2 tales que $Y_1Y_2(1+\phi_2) \approx 0$ y $\frac{Y_1Y_2}{Y_1^2+Y_2^2} \approx -\frac{\phi_2}{\phi_1}$, los estimadores máximo-verosímiles se aproximan a los estimadores mínimo-cuadráticos.

De los resultados obtenidos de aplicar los métodos de estimación de Yule-Walker, de los cuadrados mínimos y de máxima verosimilitud se puede concluir que, si $1 - \frac{Y_1Y_2 + Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)} \approx 1$, los tres estimadores tienden a aproximarse y , si $Y_1Y_2 + Y_{n-1}Y_n = 0$ (por ejemplo, porque $Y_1 = Y_n = 0$), entonces los tres estimadores son sensiblemente diferentes.

La matriz de autocovarianzas del proceso, $\Gamma_2(Y) = \begin{bmatrix} \gamma_0(Y) & \gamma_1(Y) \\ \gamma_1(Y) & \gamma_0(Y) \end{bmatrix}$ es tal que:

$$\Gamma_2^{-1}(Y) = \frac{1}{\gamma_0^2(Y) - \gamma_1^2(Y)}\begin{bmatrix} \gamma_0(Y) & -\gamma_1(Y) \\ -\gamma_1(Y) & \gamma_0(Y) \end{bmatrix} = \frac{\gamma_0(Y)}{\gamma_0^2(Y) - \gamma_1^2(Y)}\begin{bmatrix} 1 & -\rho_1(Y) \\ -\rho_1(Y) & 1 \end{bmatrix}$$

De modo que:

$$\hat{\Gamma}_2(\hat{\Phi}) = \begin{bmatrix} \hat{\sigma}^2(\hat{\phi}_1) & \hat{\gamma}(\hat{\phi}_1, \hat{\phi}_2) \\ \hat{\gamma}(\hat{\phi}_1, \hat{\phi}_2) & \hat{\sigma}^2(\hat{\phi}_2) \end{bmatrix} \approx \frac{1}{n}\left(1 - \sum_{i=1}^{p} \hat{\phi}_i\hat{\rho}_i(Y)\right)\hat{\gamma}_0(Y)\hat{\Gamma}_2^{-1}(Y) =$$

$$= \frac{1}{n}\left[1 - \hat{\phi}_1\hat{\rho}_1(Y) - \hat{\phi}_2\hat{\rho}_2(Y)\right]\frac{\hat{\gamma}_0(Y)}{\hat{\gamma}_0^2(Y) - \hat{\gamma}_1^2(Y)}\begin{bmatrix} 1 & -\hat{\rho}_1(Y) \\ -\hat{\rho}_1(Y) & 1 \end{bmatrix} =$$

$$= \frac{1}{n}\left[1 - \hat{\phi}_1\hat{\rho}_1(Y) - \hat{\phi}_2\hat{\rho}_2(Y)\right]\frac{1}{1-\hat{\rho}_1^2(Y)}\begin{bmatrix} 1 & -\hat{\rho}_1(Y) \\ -\hat{\rho}_1(Y) & 1 \end{bmatrix} =$$

$$= \frac{1}{n}\left[1 - \hat{\phi}_1\hat{\rho}_1(Y) - \hat{\phi}_2\hat{\rho}_2(Y)\right]\hat{\mathrm{P}}_2^{-1}(Y)$$

Ejemplo 2.4:

Sea un proceso $\{Y_t\}$ de parámetro mensual cuya trayectoria para el período [ENE-1997;DIC-2003] figura en el **GRAFICO 2.10** y cuyos estimadores de las funciones de autocorrelaciones y autocorrelaciones parciales están representadas en el correlograma que figura en el **CUADRO 2.8**. De la observación del comportamiento de los estimadores de $\hat{\rho}_j(Y)$ y $\hat{\phi}_{jj}(Y)$ se puede concluir que el proceso $\{Y_t\}$ admite una representación $AR(2)$.

A partir de los estimadores $\hat{\rho}_1(Y) = 0.579$ y $\hat{\rho}_2(Y) = 0.595$ se obtuvieron los siguientes estimadores de los coeficientes del modelo: **i)** de Yule-Walker: $\hat{\phi}_1^{(YW)} = 0.352759$ y $\hat{\phi}_2^{(YW)} = 0.390760$; **ii)** mínimo-cuadráticos: $\hat{\phi}_1^{(CM)} = 0.362342$ y $\hat{\phi}_2^{(CM)} = 0.433400$ y **iii)** máximo-verosímiles: $\hat{\phi}_1^{(MV)} = 0.350004$ y $\hat{\phi}_2^{(MV)} = 0.435433$.

De la observación de las realizaciones del proceso $\{Y_t\}$, $Y_1 = 0$, $Y_2 = 0.5$, $Y_{83} = 0.272387$ y $Y_{84} = 0.259722$, se puede concluir que:

$$1 - \frac{Y_1Y_2 + Y_{n-1}Y_n}{n\hat{\gamma}_0(Y)} = 1 - \frac{Y_1Y_2 + Y_{83}Y_{84}}{(84)(0.150959)} = 0.99421 \approx 1$$

Resultado que justifica la aproximación de los tres estimadores calculados precedentemente.

Se obtuvo, además, la siguiente matriz de varianzas y covarianzas de los estimadores:

$$\Gamma_2\left(\hat{\phi}\right) = \frac{1}{84}\left[1 - (.352759)(0.579) - (0.390760)(0.595)\right]\frac{1}{1-(0.579)^2}\begin{bmatrix} 1 & -0.579 \\ -0.579 & 1 \end{bmatrix} =$$

$$= 0.010087\begin{bmatrix} 1 & -0.579 \\ -0.579 & 1 \end{bmatrix}$$

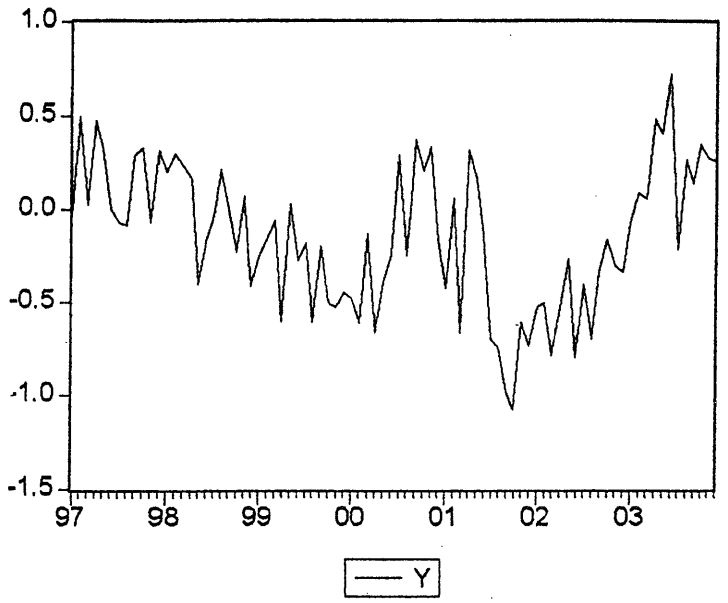
es decir, que $\sigma^2\left(\hat{\phi}_1\right) = \sigma^2\left(\hat{\phi}_2\right) = 0.10087$ y $\gamma\left(\hat{\phi}_1, \hat{\phi}_2\right) = -0.005840$.

Teniendo en cuenta que los estadísticos T correspondientes a los estimadores de Yule-Walker asumen los valores:

$$\frac{\hat{\phi}_1^{YW}}{\sigma\left(\hat{\phi}_1^{YW}\right)} = 3.512344 \qquad \frac{\hat{\phi}_2^{YW}}{\sigma\left(\hat{\phi}_2^{YW}\right)} = 3.890714$$

y que, para un 95% de confiabilidad, los límites de la zona crítica que proporciona la tabla t son aproximadamente iguales a ±2.0, se puede concluir que existen razones suficientes para rechazar la hipótesis de nulidad individual: $H_0: \phi_i = 0\ (i = 1,2)$ (debe tenerse en cuenta que, de acuerdo a lo comentado en la Sec. 2.10.3, este test es útil solamente cuando los estadísticos resultan marcadamente significativos).

GRAFICO 2.10: Trayectoria del proceso $\{Y_t\}$ en el período [ENE-1997;DIC-2003]



CUADRO 2.8: Correlograma correspondiente al proceso $\{Y_t\}$ cuya trayectoria figura en el GRAFICO 2.10

Sample: 1997:01 2003:12
Included observations: 84

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.579	0.579	29.140	0.000
		2	0.595	0.392	60.372	0.000
		3	0.425	-0.021	76.452	0.000
		4	0.401	0.040	90.979	0.000
		5	0.295	-0.026	98.936	0.000
		6	0.328	0.114	108.89	0.000
		7	0.276	0.046	116.04	0.000
		8	0.181	-0.159	119.14	0.000
		9	0.098	-0.128	120.07	0.000
		10	0.021	-0.079	120.11	0.000
		11	-0.027	-0.021	120.18	0.000
		12	-0.156	-0.189	122.81	0.000
		13	-0.149	-0.051	124.87	0.000
		14	-0.186	0.026	128.44	0.000
		15	-0.226	-0.057	133.78	0.000

Tomando en consideración los estimadores mínimo-cuadráticos de los coeficientes del modelo, queda definida la siguiente ecuación característica:

$$1-(0.362342)B-(0.433400)B^2=0$$

cuyas raíces, $\hat{G}_1 = 1.157439$ y $\hat{G}_2 = -1.993484$, sugieren que el proceso $\{Y_t\}$ es estacionario de segundo orden[99].

La estructura temporal del proceso $\{Y_t\}$ queda definida, entonces, por la solución de la ecuación en-diferencias:

$$Y_t-(0.363342)Y_{t-1}-(0.433400)Y_{t-2}=0$$

la cual será de la forma:

$$S(t)=A_1(0.863976)^t+A_2(-0.501634)^t$$

Teniendo en cuenta que $Y_1 = 0$ e $Y_2 = 0.5$, se obtiene el siguiente sistema de ecuaciones:

$$\begin{cases}(0.863976)A_1-(0.501634)A_2=0\\(0.746455)A_1+(0.251637)A_2=0.5\end{cases}$$

a partir del cual quedan definidos los valores de A_1 y A_2 que conducen a la solución particular:

$$S(t)=(0.423781)(0.863976)^t+(0.729888)(-0.501634)^t$$

Obsérvese (ver **GRAFICO 2.11**) que la presencia de una raiz negativa, $\hat{G}_2 = -1.993484$, hace que el comportamiento de $S(t)$ sea alternado para $t \leq 6$ y exponencialmente decreciente a partir de $t = 7$.

Ejemplo 2.5:

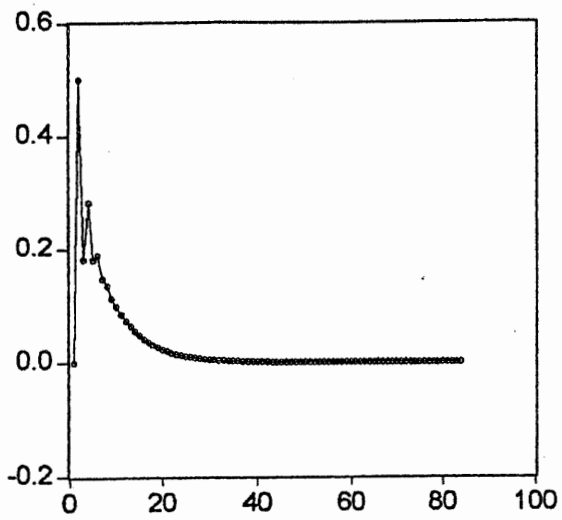
Sea un proceso $\{Y_t\}$ expresado en términos mensuales con valor medio nulo, para cuyo análisis se cuenta con una serie cronológica que abarca el período [ENE-1997;DIC-2003].

A partir de los estimadores $\hat{\rho}_1(Y) = 0.780$ y $\hat{\rho}_2(Y) = 0.517$, teniendo en cuenta que los estimadores de los dos únicos coeficientes de autocorrelación parcial significativos son $\hat{\phi}_{11}(Y) = 0.780$ y $\hat{\phi}_{22}(Y) = -0.236$ y que, por lo tanto, se puede concluir que $\{Y_t\}$ admite una representación $AR(2)$ estacionaria, se obtuvieron los siguientes estimadores de Yule-Walker de los coeficientes del modelo: $\hat{\phi}_1^{(YW)} = 0.962053$ y $\hat{\phi}_2^{(YW)} = -0.233401$, los estimadores mínimo-cuadráticos $\hat{\phi}_1^{(CM)} = 0.974023$ y $\hat{\phi}_2^{(CM)} = -0.245842$ y la siguiente matriz de varianzas y covarianzas de los estimadores:

---

[99]. Como se verá en el Cap. 3, estos valores constituyen los estimadores de los verdaderos valores de las raíces G_1 y G_2 obtenidos a partir de los datos proporcionados por la serie cronológica para el período [ENE-1997;DIC-2003]. Por lo tanto, el análisis de cualquier característica sobre las mismas requerirá la construcción de un test de hipótesis, considerando la información que proporcionan dichos estimadores.

GRAFICO 2.11: Estructura temporal del proceso $\{Y_t\}$



$$\Gamma_2\left(\hat{\phi}\right) = 0.011531\begin{bmatrix} 1 & -0.780 \\ -0.780 & 1 \end{bmatrix}$$

Dado que los estadísticos T asumen los valores:

$$\frac{\hat{\phi}_1^{(YW)}}{\sigma\left(\hat{\phi}_1^{(YW)}\right)} = 8.959123 \qquad \frac{\hat{\phi}_2^{(YW)}}{\sigma\left(\hat{\phi}_2^{(YW)}\right)} = -2.173558$$

y que, para un 95% de confiabilidad, los límites de la zona crítica que proporciona la tabla t son aproximadamente iguales a ±2.0, se puede concluir que existen razones para rechazar la hipótesis de nulidad individual: $H_0: \phi_i = 0 \; (i = 1;2)$.

Teniendo en cuenta que $Y_1 = 0$, $Y_2 = 0.3$, $Y_{83} = -0.008978$ e $Y_{84} = -0.014020$ y que, en consecuencia, $1 - \frac{Y_1 Y_2 + Y_{83} Y_{84}}{(84)\hat{\gamma}_0(Y)} \approx 1$, se justifica la aproximación de ambos estimadores.

A partir de los estimadores de Yule-Walker se obtiene la siguiente ecuación característica:

$$1 - (0.962053)B + (0.233401)B^2 = 0$$

cuyo discriminante es $\hat{D} = \hat{\phi}_1^2 + 4\hat{\phi}_2 = -0.008058 \approx 0$. Por lo que su estructura temporal puede ser expresada de la siguiente forma general:

$$S(t) = \left(A_1 + tA_2\right)(0.4810265)^t$$

y, considerando los ya mencionados valores iniciales de la serie cronológica ($Y_1 = 0$ e $Y_2 = 0.3$), la solución

particular de la ecuación en diferencias:

$$Y_t - (0.962053)Y_{t-1} + (0.233401)Y_{t-2} = 0$$

será de la forma:

$$S(t) = (1.296532)(-1+t)(0.4810265)^t$$

Las derivadas sucesivas de esta función son de la forma:

$$S^{(k)}(t) = (1.296532)(0.4810265)^t\left[\log(0.4810265)\right]^{k-1}\left[k + (-1+t)\log(0.4810265)\right] \quad (k = 1,2,.- - -)$$

Luego, se puede concluir que: **i)** la derivada primera se anula para $t_0 = 2.366432$; **ii)** para ese valor la derivada segunda es negativa, por lo que se puede asegurar que en t_0 la función $S(t)$ presenta un máximo; **iii)** la derivada segunda se anula para $t^{(i)} = 3.734717$, es decir que, para dicho valor, la función $S(t)$ presenta un punto de inflexión; **iv)** dado que, para ese valor $t^{(i)}$, la derivada tercera es negativa ($S^{(3)}\left(t^{(i)}\right) < 0$), se puede asegurar que la inflexión es negativa. El **GRAFICO 2.12** muestra el comportamiento de $S(t)$.

A partir de la estimación de los coeficientes del modelo se obtienen los estimadores de los componentes de la estructura temporal. En especial, para aquellos casos en que el discriminante de la ecuación característica es negativo $\left(D = \phi_1^2 + 4\phi_2 < 0\right)$ y basándose en el principio que, para obtener una "buena" estimación (aunque no siempre la más correcta) de una función de un parámetro, es recomendable usar la misma función del estimador[100], se pueden definir los siguientes estimadores de: **i)** la **amplitud**:

$$\hat{d}^{-1} = \sqrt{-\hat{\phi}_2}$$

ii) la **frecuencia**:

$$\hat{\theta}^* = \frac{1}{2\pi} arco\ tg\left(\frac{\sqrt{-\hat{D}}}{\left|\hat{\phi}_1\right|}\right) = \frac{1}{2\pi} arco\cos\left(\frac{\left|\hat{\phi}_1\right|}{2\sqrt{-\hat{\phi}_2}}\right)$$

iii) el **período**:

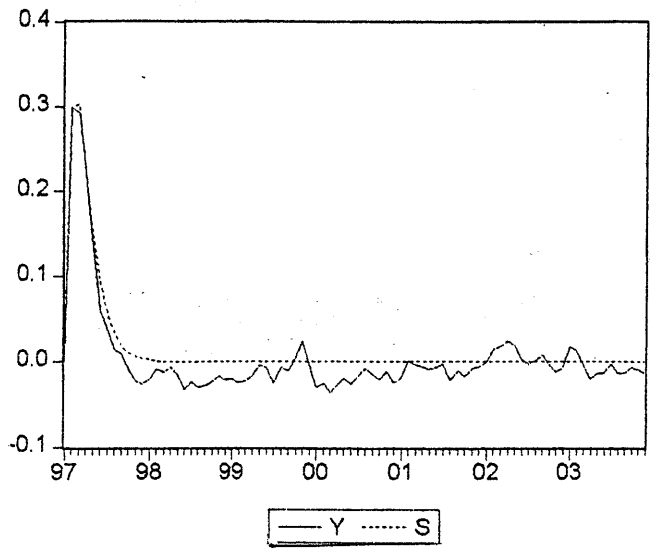
$$\hat{T} = \frac{1}{\hat{\theta}^*} = \frac{2\pi}{arco\cos\left(\frac{\left|\hat{\phi}_1\right|}{2\sqrt{-\hat{\phi}_2}}\right)}$$

y **iv)** la **fase**:

---

[100]. Ver Quenouille, M.H. (1947). Debe tenerse en cuenta que este principio no siempre proporciona la mejor estimación. Por ejemplo, de acuerdo con este principio, el mejor estimador de $\sigma^2(Y) = E\left[\left(Y - E(Y)\right)^2\right]$ debería ser $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ exclusivamente.

GRAFICO 2.12: Trayectoria y estructura temporal de $\{Y_t\}$



$$\hat{\delta} = arco\ tg\left(\frac{1-\hat{\phi}_2}{1+\hat{\phi}2}\frac{\sqrt{-\hat{D}}}{|\hat{\phi}_1|}\right)$$

(debe tenerse en cuenta que, para poder decidir acerca de la calidad de estos estimadores es necesario, obviamente, formular alguna hipótesis acerca de la distribución de probabilidades de las variables ε_t. En caso contrario, se debe optar por una solución aproximada).

Se demuestra que, dada una variable aleatoria k-dimensional $Z = \{X_1, X_2, ..., X_k\}$ y dada una función no-lineal de dicha variable $g(Z) = g(X_1, X_2, ..., X_k)$, la aproximación de segundo orden que proporciona el desarrollo de Taylor, en el entorno del punto $a = \{a_1, a_2, ..., a_k\}$ será de la forma:

$$g(X_1, X_2, ..., X_k) \approx g(a_1, a_2, ..., a_k) + \sum_{i=1}^{k}\left[\frac{\partial}{\partial X_i} g(X_1, X_2, ..., X_k)\right]_{X_1=a_1,...,X_k=a_k} (X_i - a_i) +$$

$$+ \frac{1}{2!}\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\frac{\partial^2}{\partial X_i \partial X_j} g(X_1, X_2, ..., X_k)\right]_{X_1=a_1,...,X_k=a_k} (X_i - a_i)(X_j - a_j)$$

Luego, para valores de $\sigma^2(X_i)$ $(i = 1,2,...,k)$ suficientemente pequeños y una función $g(\bullet)$ "bastante" regular, será[101]:

---

[101]. Ver Lindley, D.V. (1969), Jenkins, G.M.; Watts, D.G. (1969).

$$E\left[g(X_1,X_2,...,X_k)\right] \approx g(a_1,a_2,...,a_k) + \sum_{i=1}^{k}\left[\frac{\partial}{\partial X_i} g(X_1,X_2,...,X_k)\right]_{X_1=a_1,...,X_k=a_k} E(X_i - a_i) +$$

$$+ \frac{1}{2!}\sum_{i=1}^{k}\left[\frac{\partial^2}{\partial X_i^2} g(X_1,X_2,...,X_k)\right]_{X_1=a_1,...,X_k=a_k} E\left[(X_i - a_i)^2\right]$$

De la misma forma, teniendo en cuenta que si se toma sólo el término lineal:

$$g(X_1,X_2,...,X_k) - g\left[E(X_1),E(X_2),...,E(X_k)\right] \approx \sum_{i=1}^{k}\left[\frac{\partial}{\partial X_i} g(X_1,X_2,...,X_k)\right]_{X_1=E(X_1),...,X_k=E(X_k)} \left[X_i - E(X_i)\right]$$

se puede escribir:

$$\sigma^2\left[g(X_1,X_2,...,X_k)\right] \approx$$

$$\approx \sum_{i=1}^{k}\sum_{j=1}^{k}\left[\frac{\partial}{\partial X_i} g(X_1,X_2,...,X_k)\right]_{X_1=E(X_1),...,X_k=E(X_k)}\left[\frac{\partial}{\partial X_j} g(X_1,X_2,...,X_k)\right]_{X_1=E(X_1),...,X_k=E(X_k)} \gamma(X_i,X_j)$$

para $k = 1$ y $a = E(X)$, será:

$$E(Z) = E\left[g(X)\right] \approx g\left[E(X)\right] + \frac{\sigma^2(X)}{2!} g^{(2)}\left[E(X)\right]$$

y

$$\sigma^2(Z) = \sigma^2\left[g(X)\right] \approx \left[g^{(2)}\left(E(X)\right)\right]^2 \sigma^2(X).$$

A partir de estas expresiones es posible definir, entonces, los coeficientes $c_1(g) = E(X) - E\left[g(X)\right]$ $\approx \frac{\sigma^2(X)}{2} g^{(2)}\left[E(X)\right]$ y $c_2(g) = \frac{\sigma^2(Z)}{\sigma^2(X)} \approx 2\left[g^{(2)}\left(E(X)\right)\right]^2$. El primero mide aproximadamente el **grado de distorsión** que se introduce en una estimación en la que, de acuerdo con el criterio mencionado en la página precedente, el estimador se calcula utilizando la misma función que la del parámetro a estimar (según que sea negativo o positivo, este coeficiente permite concluir si el estimador subestima o sobreestima al parámetro). El segundo mide aproximadamente el **grado de aumento** o **disminución** de la varianza de dicho estimador (según que sea menor o mayor que la unidad, permite concluir si el estimador obtenido de aplicar la misma función $g(X)$ es más o menos confiable que la estimación de X).

En particular, para $k = 2, a_1 = E(X_1)$ y $a_2 = E(X_2)$, será:

$$E(Z) = E\left[g(X_1,X_2)\right] \approx$$

$$\approx g\left[E(X_1),E(X_2)\right] + \frac{1}{2}\frac{\partial^2}{\partial X_i^2} g\left[E(X_1,X_2)\right]\sigma^2(X_1) +$$

$$+ \frac{\partial^2}{\partial X_1 \partial X_2} g\left[E(X_1),E(X_2)\right]\gamma(X_1,X_2) + \frac{1}{2}\frac{\partial^2}{\partial X_2^2} g\left[E(X_1),E(X_2)\right]\sigma^2(X_2)$$

y,

$$\sigma^2(Z) = \sigma^2\left[g(X_1, X_2)\right] \approx$$

$$\approx \left\{\frac{\partial}{\partial X_1} g\left[E(X_1), E(X_2)\right]\right\}^2 \sigma^2(X_1) + 2\left\{\frac{\partial}{\partial X_1} g\left[E(X_1), E(X_2)\right]\right\}\left\{\frac{\partial}{\partial X_2} g\left[E(X_1), E(X_2)\right]\right\}\gamma(X_1, X_2) +$$

$$+ \left\{\frac{\partial}{\partial X_2} g\left[E(X_1), E(X_2)\right]\right\}^2 \sigma^2(X_2)$$

Suponiendo que $\sigma^2(X_1) = \sigma^2(X_2) = \sigma^2$, los coeficientes c_1 y c_2 definidos más arriba asumen la forma:

$$c_1(g) = E(Z) - g\left[E(X_1), E(X_2)\right] \approx$$

$$\approx \frac{1}{2}\left\{g_1^{(2)}\left[E(X_1), E(X_2)\right] + 2g_{1,2}^{(2)}\left[E(X_1), E(X_2)\right]\rho(X_1, X_2) + g_2^{(2)}\left[E(X_1), E(X_2)\right]\right\}$$

$$c_2(g) = \frac{\sigma^2(Z)}{\sigma^2(X)} \approx \left\{g_1^{(1)}\left[E(X_1), E(X_2)\right]\right\}^2 +$$

$$+ 2g_1^{(1)}\left[E(X_1), E(X_2)\right]g_2^{(1)}\left[E(X_1), E(X_2)\right]\rho(X_1, X_2) + \left\{g_2^{(1)}\left[E(X_1), E(X_2)\right]\right\}^2$$

Partiendo de la hipótesis de que los estimadores de $\hat{\phi}_1$ y $\hat{\phi}_2$ son insesgados y consistentes, es decir, tales que $E\left(\hat{\phi}_i\right) = \phi_i$ y $\sigma^2\left(\hat{\phi}_i\right) = \frac{1}{n}\left(1 - \phi_2^2\right)(i = 1,2)$. En realidad los estimadores $\hat{\phi}_1$ y $\hat{\phi}_2$ son sesgados, pero la condición de insesgado se plantea para demostrar que la distorsión de los estimadores de los parámetros de la estructura temporal puede ser alterada por la distorsión originaria de los estimadores de los coeficientes del modelo. Luego, a las aproximaciones inherentes al método utilizado, se le deben agregar los eventuales sesgos de los estimadores de los coeficientes del modelo.

Con respecto a la **amplitud**, se verifica que $c_1(d) \approx \frac{1}{8n}\frac{1 - \phi_2^2}{\phi_2\sqrt{-\phi_2}} < 0$ (debe tenerse en cuenta que en el caso en que $D < 0$ será $\phi_2 < 0$) y que:

$$c_2(d) \approx \frac{1}{-4\phi_2}\begin{cases} > 1 & si \quad -0.25 < \phi_2 < 0 \\ = 1 & si \quad \phi_2 = 0.25 \\ < 1 & si \quad -1 < \phi_2 < -0.25 \end{cases}$$

Luego, se puede concluir que, si se utiliza el estimador $\hat{d} = \sqrt{-\phi_2}$: **i)** el coeficiente d de la estructura temporal resulta subestimado en una cantidad que varía en forma inversa a $-\phi_2$; **ii)** para $-\phi_2 > 0.25$, la variación de $\sigma^2\left(\hat{d}\right)$ es lo suficientemente pequeña como para ser despreciada y **iii)**, para $-\phi_2 \leq 0.25$, la confiabilidad de $\hat{d}$ es superior a la de $\hat{\phi}_2$. Por lo tanto, se puede asegurar que, para $-\phi_2 > 0.25$, $\hat{d}$ es un "buen" estimador de d y, para $-\phi_2 \leq 0.25$, $\hat{d}$ debe ser corregido restándole $c_1(d)$.

En lo que respecta a la **frecuencia,** se demuestra que el coeficiente de distorsión del estimador

$\hat{\theta}^{**} = 2\pi\hat{\theta}^{*} = arco\cos\left(\frac{|\hat{\phi}_1|}{2\sqrt{-\hat{\phi}_2}}\right)$ puede ser expresado de la siguiente forma:

$$c_1(\theta^{**}) \approx \frac{1}{n}\frac{1-\phi_2}{D\sqrt{-D}}\left[(D+2\phi_2)\frac{-\phi_1}{4\phi_2^2} - \frac{2\phi_1}{1-\phi_2} + \frac{\phi_1}{2}\right]$$

(donde debe tenerse en cuenta que $D < 0$ y $\phi_2 < 0$). De lo que se concluye que, si $\phi_1 > 0$, en general, $c_1(\theta^{**})$ es negativo (si bien depende de ϕ_1 y ϕ_2, en general decrece con ϕ_2), por lo tanto, subestima al coeficiente θ^{**} y es asintóticamente insesgado.

De la misma forma se demuestra que:

$$c_2(\theta^{**}) \approx \frac{1}{-D}\left[1 + \frac{\phi_1^2}{(1-\phi_2)\phi_2} + \left(\frac{\phi_1}{2\phi_2}\right)^2\right]$$

por lo que se puede concluir que $\hat{\theta}^{**}$ es un estimador es consistente.

En lo que hace al **período** se demuestra que, su estimador, $\hat{T} = \frac{1}{\hat{\theta}^{*}} = \frac{2\pi}{\hat{\theta}^{**}}$, es tal que:

$$E(\hat{T}) = E\left(\frac{1}{\hat{\theta}^{*}}\right) \approx \frac{1}{\theta^{*}} + \frac{\sigma^2(\hat{\theta}^{*})}{2}\left[\frac{\partial^2}{\partial(\hat{\theta}^{*})^2}\frac{1}{\hat{\theta}^{*}}\right]_{\hat{\theta}^{*}=\theta^{*}}$$

$$\sigma^2(\hat{T}) = \sigma^2\left(\frac{1}{\hat{\theta}^{*}}\right) \approx \left[\left\{\frac{\partial}{\partial\hat{\theta}^{*}}\frac{1}{\hat{\theta}^{*}}\right\}^2\right]_{\hat{\theta}^{*}=\theta^{*}}$$

De modo que se puede concluir que $c_1(T) \approx \frac{1}{n}\frac{2\pi c_2(\hat{\theta}^{**})}{(\hat{\theta}^{**})^3}(1-\phi_2^2)$. Es decir que el coeficiente de distorsión de $\hat{T}$: **i)** disminuye rápidamente cuando $\hat{\theta}^{**}$ crece y **ii)** crece a medida que crece el período. Asimismo, se verifica que $c_2(T) \approx \frac{c_2(\hat{\theta}^{**})}{(\hat{\theta}^{**})^2}$. Por lo que se puede concluir que la confiabilidad de $\hat{T}$, con respecto a $\hat{\phi}_1$ y $\hat{\phi}_2$, decrece sensiblemente al crecer T. Es decir, la calidad del estimador disminuye a medida que el período del proceso crece[102].

La evaluación de la calidad del estimador de la **fase** ($\hat{\delta}$) implica desarrollos algebraicos sumamente complicados y, dado que su importancia en las aplicaciones es muy inferior a la de los estimadores analizados precedentemente, no será tratada en esta versión.

De todo lo expuesto en las páginas precedentes se puede concluir, en términos generales, que los

---

[102]. Lo que permite concluir que la estructura de un $AR(2)$ no logrará individualizar correctamente partes componentes periódicas "medianamente amplias" como, por ejemplo, factores estacionales.

estimadores de los elementos de la estructura temporal son asintóticamente correctos sólo si los estimadores $\hat{\phi}_1$ y $\hat{\phi}_2$ son asintóticamente correctos y que su calidad aumenta cuando n aumenta.

Ejemplo 2.6:

Sea un proceso $\{Y_t\}$ de parámetro mensual con valor medio nulo. Se cuenta, para su análisis, con una serie cronológica para el período [ENE-1997;DIC-2003] a partir de la cual se obtuvo el siguiente modelo $AR(2)$:

$$Y_t = (0.608661)Y_{t-1} - (0.175435)Y_{t-2}$$

Donde los estimadores de los coeficientes fueron obtenidos de la aplicación del criterio de los cuadrados mínimos y los correspondientes estadísticos T asumieron los valores $T\left(\hat{\phi}_1^{cm}\right) = 13.089630$ y $T\left(\hat{\phi}_2^{cm}\right) =$ $= -3.765485$, lo cual permite aceptar la significatividad de ϕ_1 y ϕ_2 para niveles de significación del 10%, 5% y 1%.

De la solución de la ecuación característica:

$$\Phi(B) = 1 - (0.608661)B + (0.175435)B^2 = 0$$

se obtuvieron las raíces

$$\hat{G}_1 = 1.734719 + (1.638984)i$$

y

$$\hat{G}_2 = 1.734719 - (1.638984)i$$

De modo que se puede concluir que la estructura temporal observa un comportamiento sinusoidal amortiguado, cuyas características están dadas por los siguientes estimadores:

i) Factor de amortiguación: $\hat{d}^{-1} = \sqrt{-\hat{\phi}_2} = 0.418850$. El coeficiente de distorsión r de $\hat{d}$, $c_1\left(\hat{d}\right) = -0.019628$, permite concluir que $\hat{d}$ subestima el verdadero valor de d . Por otra parte, dado que $-\phi_2 = -0.175435$, se puede asegurar que $\hat{d}$ es un estimador más confiable que $\hat{\phi}_1$.

ii) Frecuencia: $\hat{\theta}^* = 0.120553$ o, alternativamente, $\hat{\theta}^{**} = 2\pi\hat{\theta}^* = 0.757456$. En este caso, como se comentó en las páginas anteriores, dado que $\phi_1 > 0$, se verifica que $c_1\left(\hat{\theta}^{**}\right) < 0$, es decir, que este estimador subestima el verdadero valor de la frecuencia. Por otra parte, se verifica que el coeficiente $c_2\left(\hat{\theta}^{**}\right) > 1$.

iii) Período: $\hat{T} = \dfrac{1}{\hat{\theta}^*} = 8.295107$.

iv) Fase: $\hat{\delta} = 0.462819$

De los resultados anteriores se puede concluir que la estructura temporal del proceso $\{Y_t\}$ será de la forma:

$$S(t) = \frac{(0.418850)^t \, sen\left[2\pi(0.120553)t + 0.462819\right]}{sen(0.462819)}$$

En el **GRAFICO 2.13** figura el comportamiento de esta función.

Como se mencionó en páginas anteriores, un análisis acerca de la calidad de los estimadores de la estructura temporal requiere de la formulación de alguna hipótesis sobre la distribución de probabilidades de las variables ε_t que permita definir la distribución de probabilidades de $\hat{\phi}_1$ y $\hat{\phi}_2$. Ahora bien, como este procedimiento resulta obviamente muy complicado (sobre todo en los casos de muestras chicas, en las cuales no rigen los postulados del teorema central del límite) se puede aplicar un método aproximado que -en virtud de su calidad de funciones de realizaciones de un proceso estocástico- considera a los coeficientes ϕ_1 y ϕ_2 como variables marginales de una variable aleatoria bidimensional cuyo dominio, dado que $D < 0$, es de la forma $-1 < \phi_2 < 0$ y $|\phi_1| < 2\sqrt{-\phi_2}$ (ver **GRAFICO 2.14**). Obviamente la hipótesis más simple es la de ignorancia absoluta acerca de la distribución de las variables ϕ_1 y ϕ_2, la cual implica el supuesto de distribución uniforme de la variable bidimensional (ϕ_1, ϕ_2) en la región C del **GRAFICO 2.14**. Dada la simetría del dominio respecto de la variable ϕ_1 en la región C, este supuesto es equivalente a una distribución de probabilidades de la forma:

$$f\left(|\phi_1|, \phi_2\right) = \begin{cases} \dfrac{1}{area\ C'} & si \qquad \left(|\phi_1|, \phi_2\right) \in C' \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Es decir,

$$f\left(|\phi_1|, \phi_2\right) = \begin{cases} \dfrac{3}{4} & para \qquad \left(-1 < \phi_1 < 0\right) \cap \left(|\phi_1| < 2\sqrt{-\phi_2}\right) \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

De modo que la función de densidad de la variable

$$\theta^{**} = 2\pi\theta^{*} = arco\cos\left(\frac{|\phi|}{2\sqrt{-\phi_2}}\right)$$

puede ser construida de la siguiente forma: Sean las variables X y θ^{**}, definidas por las siguientes transformaciones de las variables ϕ_1 y ϕ_2:

$$X = \sqrt{-\phi_2} \quad \left(\Omega(X) = \{0 < X < 1\}\right)$$

y

$$\theta^{**} = arco\cos\left(\frac{|\phi_1|}{2\sqrt{-\phi_2}}\right) \quad \left(\Omega(\theta^{**}) = \left\{0 < \theta^{**} < \frac{\pi}{2}\right\}\right)$$

Aplicando la transformación inversa se obtiene que $|\phi_1| = 2X\cos\left(\theta^{**}\right)$ y $\phi_2 = -X^2$, cuyo Jacobiano, en valor absoluto, es:

GRAFICO 2.13: Trayectoria y estructura temporal del proceso $\{Y_t\}$

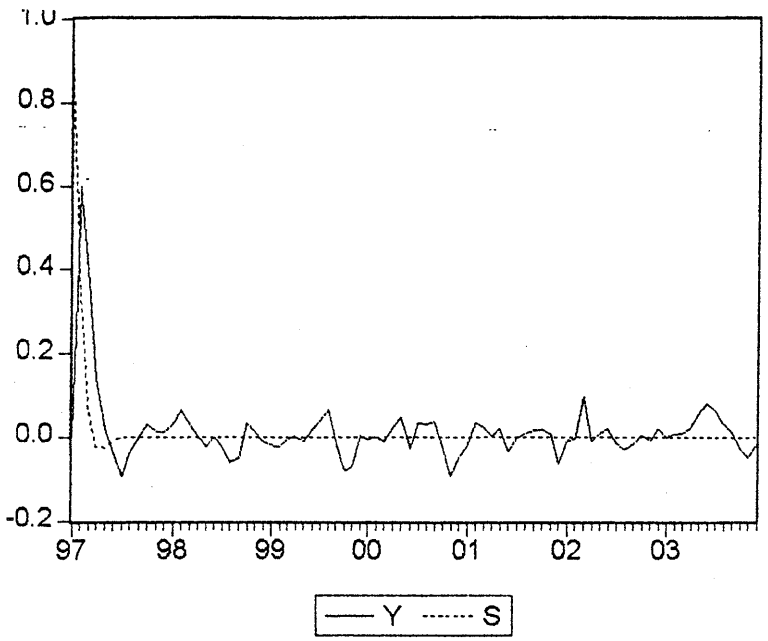
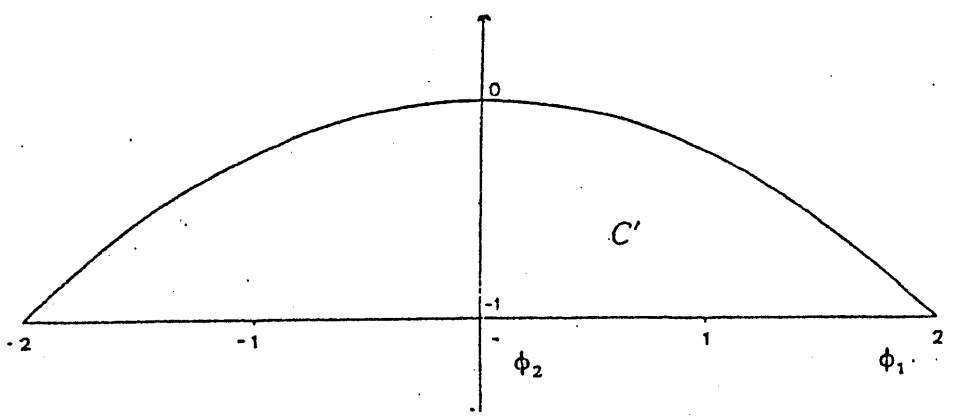


GRAFICO 2.14.- Región de estacionariedad para los coeficientes φ_1 y φ_2 que generan un AR(2) con raíces complejas.



$$|J| = \begin{vmatrix} 2\cos(\theta^{**}) & -2X\,sen(\theta^{**}) \\ -2X & 0 \end{vmatrix} = \left|-4X^2\,sen(\theta^{**})\right| = 4X^2\,sen(\theta^{**})$$

La función de densidad de la variable bidimensional (X,θ^{**}) será, entonces, de la forma:

$$f(X,\theta^{**}) = \begin{cases} 3X^2\,sen(\theta^{**}) & para \quad (0 < X < 1) \cap \left(0 < \theta^{**} < \dfrac{\pi}{2}\right) \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y, por lo tanto, la función de densidad de la variable marginal θ^{**} será:

$$f(\theta^{**}) = \int_0^1 3X^2\,sen(\theta^{**})dX = sen(\theta^{**}) \qquad \left(0 < \theta^{**} < \frac{\pi}{2}\right)$$

su función de distribución en consecuencia, será de la forma:

$$F(\theta^{**}) = \begin{cases} 0 & para \quad \theta^{**} \le 0 \\ 1-\cos(\theta^{**}) & para \quad 0 < \theta^{**} < \dfrac{\pi}{2} \\ 1 & para \quad \theta^{**} \ge \dfrac{\pi}{2} \end{cases}$$

y su función generatriz de momentos será:

$$M_{\theta^{**}}(w) = E\left(e^{w\theta^{**}}\right) = \int_0^{\pi/2} e^{w\theta^{**}} f(\theta^{**})\,d\theta^{**} = \int_0^{\pi/2} e^{w\theta^{**}} sen(\theta^{**})d\theta^{**} = \frac{1 - we^{w\pi/2}}{1+w^2}$$

A partir de la cual se obtienen los siguientes momentos absolutos: $m(\theta^{**}) = E(\theta^{**}) = 1$, $m_2(\theta^{**}) = \pi - 2$, $m_3(\theta^{**}) = \frac{3}{4}\left(\pi^2 - \theta\right)$ y $m_4(\theta^{**}) = \frac{1}{2}\left(n^3 - 24\pi + 48\right)$ y los momentos centrados $\mu_2(\theta^{**}) = \sigma^2(\theta^{**}) = \pi - 3$, $\mu_3(\theta^{**}) = \frac{3}{4}\left(\pi^2 - 12\pi + 8\right)$ y $\mu_4(\theta^{**}) = \frac{\pi^3}{2} - 3\pi^2 - 6\pi + 33$.

De la misma forma se obtiene la función de densidad de la variable:

$$d^{-1} = \sqrt{-\phi_2}\ \left(\Omega\left(d^{-1}\right) = \left(0 < d^{-1} < 1\right)\right.$$

que será de la forma:

$$f\left(d^{-1}\right) = \begin{cases} 3d^{-2} & para \quad d^{-1} \le 0 \\ 0 & para\ cualquier\ otro\ valor\ de\ d^{-1} \end{cases}$$

La correspondiente función de distribución será:

$$F\left(d^{-1}\right) = \begin{cases} 0 & para \quad d^{-1} \leq 0 \\ d^{-1} & para \quad 0 < d^{-1} < 1 \\ 1 & para \quad d^{-1} \geq 1 \end{cases}$$

Se obtienen, entonces, los siguientes momentos absolutos: $m\left(d^{-1}\right) = E\left(d^{-1}\right) = \frac{3}{4}$; $m_2\left(d^{-1}\right) = \frac{3}{5}$; $m_3\left(d^{-1}\right) = \frac{3}{6}$ y $m_4\left(d^{-1}\right) = \frac{3}{7}$ y los siguientes momentos centrados: $\mu_2\left(d^{-1}\right) = \sigma^2\left(d^{-1}\right) = \frac{3}{80}$; $\mu_3\left(d^{-1}\right) = -\frac{1}{160}$ y $\mu_4\left(d^{-1}\right) = \frac{6519}{8960}$.

Teniendo en cuenta la expresión correspondiente a σ_ε^2 dada en páginas anteriores y recurriendo a la fórmula del coeficiente de determinación que se utiliza en los modelos estáticos[103], se puede escribir:

$$R_2^2 = 1 - \frac{\sigma_\varepsilon^2\left(AR(2)\right)}{\sigma_Y^2} = \rho_1(Y) + \rho_2(Y)\phi_2$$

En general, el coeficiente de determinación correspondiente a una representación $AR(p)$ puede ser expresado utilizando la siguiente forma recursiva: $R_p^2 = \phi_{pp}\left(1 - R_{p-1}^2\right) + R_{p-1}^2 \ (p = 1,2,3,...)$ en la que el segundo sumando representa la proporción de la variación de $\{Y_t\}$ explicada por la autorregresividad de orden $p-1$ y el primer sumando mide la contribución de Y_{t-p} a la parte de la variación de $\{Y_t\}$ no explicada por la representación $AR(p-1)$ (ver Nelson, C.R. (1976)).

En particular, de acuerdo a lo ya demostrado, en el caso del $AR(2)$ se puede escribir:

$$R_2^2 = 2\rho_2(Y)\frac{\rho_1^2(Y)}{1-\rho_1^2(Y)} + \frac{\rho_1^2(Y)+\rho_2^2(Y)}{1-\rho_1^2(Y)} = \rho_2(Y) - \phi_2 + \rho_2(Y)\phi_2$$

Obsérvese que, si $\phi_1 = \phi_2 = 0$, entonces $R_2^2 = 0$ y si $\phi_2 + \phi_1 = 1$ o si $\phi_2 - \phi_1 = 1$, entonces $R_2^2 = 1$. Es decir que, en los límites de la región de estacionariedad, el R_2^2 pierde su utilidad como medida de las condiciones de explicabilidad del modelo $AR(2)$ respecto del comportamiento del proceso $\{Y_t\}$.

Sustituyendo $\rho_1(Y), \rho_2(Y), \phi_1$ y ϕ_2 por sus estimadores, se obtiene el estimador del coeficiente de determinación $\hat{R}_2^2 = \hat{\rho}_2(Y) - \hat{\phi}_2 + \hat{\rho}_2(Y)\hat{\phi}_2$.

A partir de las propiedades de la función de autocorrelaciones parciales, Quenouille, M.H. (1947)[104]

---

[103]. Estos modelos serán tratados en el Cap. 6.

[104]. Ver también Kendall, M.G. (1973).

propuso, para muestras suficientemente grandes, un test para decidir acerca de estas condiciones de explicabilidad de una representación $AR(p)$, utilizando un coeficiente de la forma:

$$W_j(Y) = \frac{1}{n\sigma_Y^2} E\left[\left(Y_{t+j} - \phi_1 Y_{t+j-1} - \ldots - \phi_p Y_{t+j-p}\right)\left(Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p}\right)\right] =$$

$$= \rho_j(Y) + \phi_1^* \rho_{j-1}(Y) + \ldots + \phi_p^* \rho_{j-2p} \qquad (j > p)$$

(donde $\phi_j^* = \sum_{i=1}^{j} \phi_i \phi_{j-i}$). En particular, para una representación $AR(2)$ este coeficiente asume la forma:

$$W_j(Y) = \rho_j(Y) - 2\phi_1 \rho_{j-1}(Y) + \left(\phi_1^2 - 2\phi_2\right)\rho_{j-2}(Y) + 2\phi_1\phi_2\rho_{j-3}(Y) + \phi_2^2 \rho_{j-4}(Y) \qquad (j > 2)$$

De modo que la representación $AR(2)$ se considerará aceptable, con un cierto grado de confiabilidad, si no existen razones suficientes para rechazar la hipótesis $H_0: W_j(Y) = 0$ con respecto a la alternativa $H_1: W_j(Y) \neq 0$. El estadístico a utilizar se obtiene de reemplazar los coeficientes ϕ_1 y ϕ_2 y los coeficientes de autocorrelación ($\rho_j(Y)$) por sus estimadores muestrales y se demuestra que, para muestras suficientemente grandes, los estadísticos $\hat{W}_j(Y)$ son variables aleatorias iid con distribución aproximable por una función Normal del tipo $N\left(0, \sigma^2_{W_j(Y)}\right)$, donde

$$\hat{\sigma}^2_{W_j}(Y) = \frac{1}{n}\left\{\frac{1+\hat{\phi}_2}{1-\hat{\phi}_2}\left[\left(1-\hat{\phi}_2\right)^2 - \hat{\phi}_1^2\right]\right\}^2 \qquad (j > 2)$$

expresión que, considerando los estimadores de Yule-Walker de los coeficientes ϕ_1 y ϕ_2, puede ser escrita de la siguiente forma:

$$\sigma^2_{W_j}(Y) = \frac{1}{n}\left\{\frac{\left[1 - 2\hat{\rho}_1^2(Y) + \hat{\rho}_2(Y)\right]\left[1 - \hat{\rho}_2(Y)\right]}{1 - \hat{\rho}_1^2(Y)}\right\}^2 = \frac{1}{n}\left\{1 + \hat{\phi}_2^{(YW)}\left[1 - \hat{\rho}_2(Y)\right]\right\}^2$$

2.11.- Los procesos de promedios móviles de orden q

2.11.1.- Las condiciones de estacionariedad e invertibilidad

Supóngase un proceso $\{Y_t\}$ para el cual sólo los q primeros coeficientes de su representación en promedios móviles sean significativos:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

(donde: **i)** Y_t denota una variable centrada y **ii)** el proceso representativo de las innovaciones $\{\varepsilon_t\}: WN - N\left(0, \sigma_\varepsilon^2\right)$). Este proceso se conoce como **de promedios móviles de orden q (MA(q)).**

Utilizando el operador "backward", se puede escribir:

$$Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right)\varepsilon_t = \Theta(B)\varepsilon_t$$

Donde $\Theta(B)$ denota el **operador de promedios móviles de orden q.**

Su correspondiente expresión autorregresiva será de la forma:

$$\left[\Theta(B)\right]^{-1} Y_t = \varepsilon_t$$

$$\frac{1}{\prod_{i=1}^{q}\left(1 - H_i^{-1}B\right)} Y_t = \sum_{i=1}^{q} \frac{A_i}{1 - H_i^{-1}B} Y_t = \sum_{i=1}^{\infty} \pi_i Y_{t-i} = \Pi(B) Y_t = \varepsilon_t$$

Donde las $H_i^{-1}\ (i = 1,2,...,q)$ denotan las inversas de las raíces de la ecuación característica, $\Theta(B) = 0$.

Dado que el operador de promedios móviles que corresponde a este proceso es de orden finito, no es necesario fijar ninguna condición para asegurar su estacionariedad. La condición de invertibilidad -que, como se vio, está vinculada con la convergencia del operador $\Pi(B)$, para $|B| \leq 1$ - se verificará cuando todas las raíces $H_i^{-1}\ (i = 1,2,...,q)$ sean, en módulo, menores que la unidad; es decir, cuando todas las raíces de la ecuación característica:

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q = 0$$

sean, en módulo, mayores que la unidad.

Una interpretación alternativa de las representaciones $MA(q)$ es a partir de la representación AR de orden infinito de $\{Y_t\}$ (Ver Sec. 2.9.2). Sea un proceso $\{Y_t\}$ que admite una representación AR de la forma $Y_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + ... = Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + ... = \varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$), entonces, se puede escribir $Y_{t-1} + \theta_1 Y_{t-2} + \theta_1^2 Y_{t-3} + ... = \varepsilon_{t-1}$. Multiplicando esta ecuación por θ_1 y restándola de la ecuación anterior, resulta la representación de $\{Y_t\}$ como un promedio ponderado de ε_t y ε_{t-1} de la forma $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$. Este resultado justifica la denominación de esta representación como promedio móvil. De una forma similar se obtiene la interpretación de la representación $MA(q)$.

2.11.2.- Las funciones de autocorrelaciones y de autocorrelaciones parciales

La función de autocovarianzas de este proceso se obtiene a partir de la siguiente expresión:

$$\gamma_j(Y) = E\left(Y_{t-j} Y_t\right) =$$

$$= E\left[\left(\varepsilon_{t-j} - \theta_1 \varepsilon_{t-j-1} - ... - \theta_q \varepsilon_{t-j-q}\right)\left(\varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}\right)\right] \qquad (j = 0,1,2,...)$$

Efectuando las operaciones correspondientes y aplicando las propiedades de la esperanza matemática, se obtiene que:

$$\gamma_j(Y) = \begin{cases} \left(-\theta_j + \theta_1\theta_{j+1} + ... + \theta_{q-j}\theta_q\right)\sigma_\varepsilon^2 & (j = 1,2,...,q) \\ 0 & (j > q) \end{cases}$$

En particular, para $j = 0$, será:

$$\gamma_0(Y) = \sigma_Y^2 = \left(1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2\right)\sigma_\varepsilon^2$$

De la función de autocovarianzas resulta la función de autocorrelaciones:

$$\rho_j(Y) = \begin{cases} \dfrac{-\theta_j + \theta_1\theta_{j+1} + ... + \theta_{q-j}\theta_q}{1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2} & (j = 1,2,...,q) \\ 0 & (j > q) \end{cases}$$

cuyo comportamiento permite concluir que la memoria de un proceso $MA(q)$ es de q períodos.

De la aplicación de estos coeficientes de autocorrelación a las expresiones obtenidas en la Sec. 2.10.4 surge, en forma inmediata, que la función de autocorrelaciones parciales es de dimensión infinita.

En los cuadros **2.9** y **2.10** figuran las funciones de autocorrelaciones y de autocorrelaciones parciales típicas de los procesos $MA(1)$ invertibles y, en el **CUADRO 2.11** figuran las funciones de autocorrelaciones y de autocorrelaciones parciales típicas de un proceso $MA(2)$ invertible (estas distintas modalidades de los procesos MA serán tratadas en las secciones 2.11.4 y 2.11.5).

Sean $\{Y_{1t}\}$ e $\{Y_{2t}\}$ dos procesos incorrelacionados entre sí que admiten sendas representaciones MA de la forma:

$$Y_{1t} = \varepsilon_t^{(1)} + \theta_1^{(1)}\varepsilon_{t-1}^{(1)} + ... + \theta_{q_1}^{(1)}\varepsilon_{t-q_1}^{(1)}$$

$$Y_{2t} = \varepsilon_t^{(2)} + \theta_1^{(2)}\varepsilon_{t-1}^{(2)} + ... + \theta_{q_2}^{(2)}\varepsilon_{t-q_2}^{(2)}$$

(donde: **i)** Y_{1t} e Y_{2t} son variables centradas; **ii)** $\left\{\varepsilon_t^{(1)}\right\}: WN - N\left(0,\sigma_1^2\right)$ y $\left\{\varepsilon_t^{(2)}\right\}: WN - N\left(0,\sigma_2^2\right)$ son incorrelacionados entr sí). Entonces, de acuerdo a lo expresado en la Sec. 2.2, la función de autocovarianzas del proceso bidimensional $\{Y_t\} = \{Y_{1t}\} + \{Y_{2t}\}$ será de la forma:

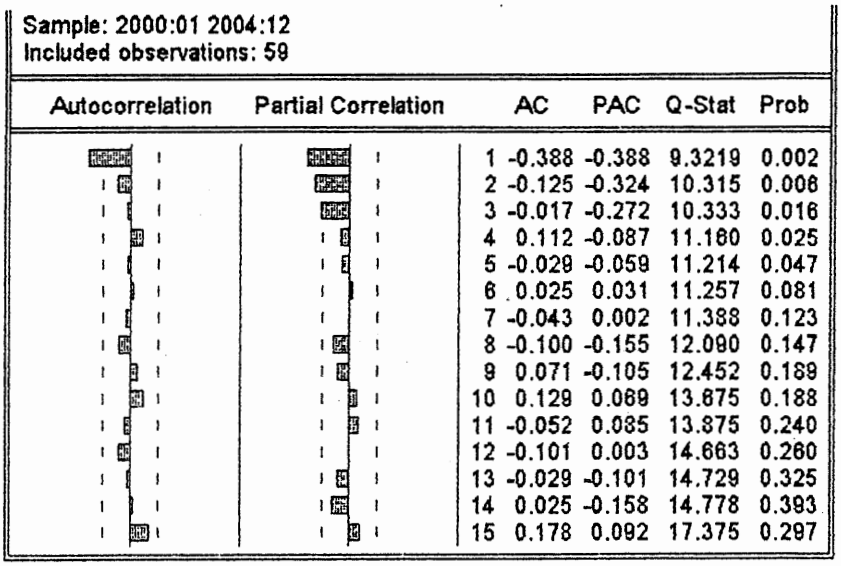
$$\gamma_j(Y) = \begin{cases} \gamma_j(Y_1) + \gamma_j(Y_2) & para \quad j = 0,1,2,...; Max(q_1,q_2) \\ 0 & para \quad j > Max(q_1,q_2) \end{cases}$$

Luego, se puede concluir que la suma de dos procesos incorrelacionados entre sí, que admiten representaciones de la forma $MA(q_1)$ y $MA(q_2)$, respectivamente, genera un proceso que admite una representación $MA\left(\max(q_1,q_2)\right)$.

2.11.3.- La estimación de los coeficientes

Reemplazando las autocorrelaciones $\rho_1(Y),\rho_2(Y),...,\rho_q(Y)$ por sus estimadores en el sistema de ecuaciones no-lineales que figura más arriba, se obtienen los **estimadores preliminares** de los coeficientes $\theta_1,\theta_2,...,\theta_q$. Asimismo, reemplazando en la expresión $\sigma_\varepsilon^2 = \dfrac{\gamma_0(Y)}{1 + \theta_1^2 + ... + \theta_q^2}$, $\theta_1,\theta_2,...,\theta_q$ por sus estimadores

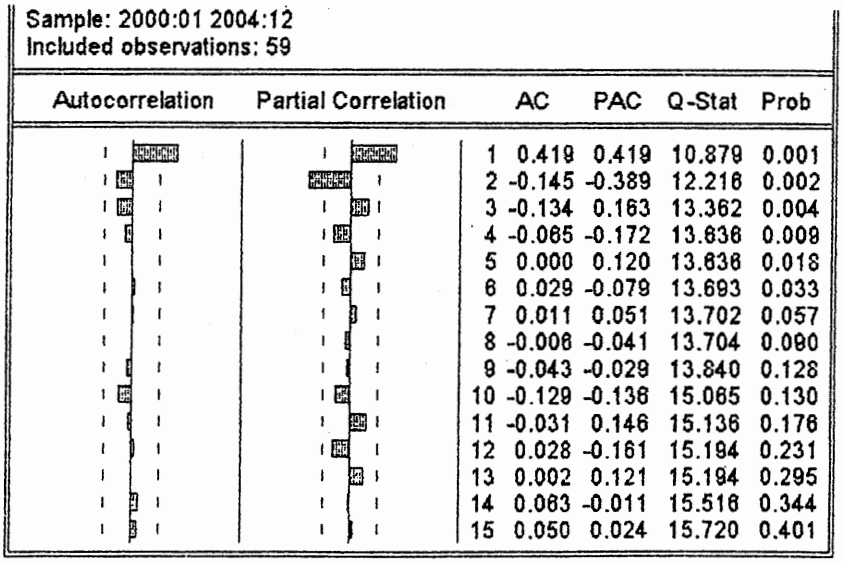
CUADRO 2.9: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación MA(1) con θ_1=0.75.



Sample: 2000:01 2004:12
Included observations: 59

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.388	-0.388	9.3219	0.002
		2	-0.125	-0.324	10.315	0.006
		3	-0.017	-0.272	10.333	0.016
		4	0.112	-0.087	11.160	0.025
		5	-0.029	-0.059	11.214	0.047
		6	0.025	0.031	11.257	0.081
		7	-0.043	0.002	11.388	0.123
		8	-0.100	-0.155	12.090	0.147
		9	0.071	-0.105	12.452	0.189
		10	0.129	0.089	13.875	0.188
		11	-0.052	0.085	13.875	0.240
		12	-0.101	0.003	14.663	0.260
		13	-0.029	-0.101	14.729	0.325
		14	0.025	-0.158	14.778	0.393
		15	0.178	0.092	17.375	0.297

CUADRO 2.10: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación MA(1) con θ_1=-0.75.

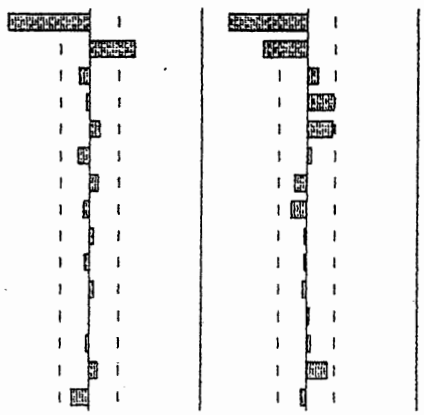


Sample: 2000:01 2004:12
Included observations: 59

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.419	0.419	10.879	0.001
		2	-0.145	-0.389	12.216	0.002
		3	-0.134	0.163	13.362	0.004
		4	-0.065	-0.172	13.636	0.009
		5	0.000	0.120	13.636	0.018
		6	0.029	-0.079	13.693	0.033
		7	0.011	0.051	13.702	0.057
		8	-0.006	-0.041	13.704	0.090
		9	-0.043	-0.029	13.840	0.128
		10	-0.129	-0.136	15.065	0.130
		11	-0.031	0.146	15.136	0.176
		12	0.028	-0.161	15.194	0.231
		13	0.002	0.121	15.194	0.295
		14	0.063	-0.011	15.516	0.344
		15	0.050	0.024	15.720	0.401

CUADRO 2.11: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación MA(2) con θ_1=1.08 y θ_2=0.70.

Sample: 2000:01 2004:12
Included observations: 58



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.769	-0.769	36.087	0.000
		2	0.416	-0.429	46.833	0.000
		3	-0.092	0.107	47.363	0.000
		4	-0.031	0.255	47.425	0.000
		5	0.093	0.254	47.993	0.000
		6	-0.100	0.038	48.666	0.000
		7	0.082	-0.116	49.124	0.000
		8	-0.054	-0.151	49.324	0.000
		9	0.037	-0.030	49.423	0.000
		10	-0.049	-0.021	49.598	0.000
		11	0.036	-0.042	49.696	0.000
		12	-0.007	0.025	49.700	0.000
		13	-0.024	0.046	49.746	0.000
		14	0.081	0.211	50.259	0.000
		15	-0.156	-0.051	52.235	0.000

preliminares y σ_Y^2 por el estimador de $\gamma_0(Y)$, se obtiene el estimador preliminar de σ_ε^2.

La solución del sistema de ecuaciones no-lineales puede ser obtenida:

i) Utilizando un **proceso iterativo linealmente convergente** que permite el cálculo de los estimadores de los coeficientes $\sigma_\varepsilon^2, \theta_q, \theta_{q-1}, ..., \theta_1$, en ese orden, a partir de la iteración:

$$\sigma_\varepsilon^2 = \frac{\gamma_0(Y)}{1+\theta_1^2+...+\theta_q^2}$$

$$\theta_q = -\frac{\gamma_q(Y)}{\sigma_\varepsilon^2}$$

$$\theta_{q-1} = -\left(\frac{\gamma_{q-1}(Y)}{\sigma_\varepsilon^2} - \theta_1\theta_q\right)$$

$$\theta_{q-2} = -\left(\frac{\gamma_{q-2}}{\sigma_\varepsilon^2} - \theta_1\theta_{q-1} - \theta_{q-1}\theta_q\right)$$

.....

$$\theta_1 = -\left(\frac{\gamma_1(Y)}{\sigma_\varepsilon^2} - \theta_1\theta_2 - \theta_2\theta_3 - ... - \theta_{q-1}\theta_q\right)$$

(con $\theta_0 = 0$). La primera iteración se resuelve haciendo $\theta_1 = \theta_2 = ... = \theta_q = 0$.

ii) Utilizando una forma del **algoritmo de Newton-Raphson** desarrollada por Wilson, G. (1969),

que genera un proceso iterativo con **convergencia cuadrática**. Sea el vector $\tau^T = \left[\tau_0 \quad \tau_1 \quad \ldots \quad \tau_q\right]$, donde $\tau_0^2 = \sigma_\varepsilon^2$ y $\theta_j = -\frac{\tau_j}{\tau_0}\,(j = 1,2,\ldots,q)$ y, denotando por $\tau^{(i)}$ al estimador de τ obtenido en la i-ésima iteración, se puede plantear la siguiente fórmula de recurrencia: $\tau^{(i+1)} = \tau^{(i)} - \left(T^{(i)}\right)^{-1} f^{(i)}$, donde,

$$\left(f^{(i)}\right)^T = \left[f_0^{(i)} \quad f_1^{(i)} \quad \ldots \quad f_q^{(i)}\right]$$

$$f_j^{(i)} = \sum_{h=0}^{q-j} \tau_h^{(i)} \tau_{h+j}^{(i)} - \hat{\gamma}_j(Y)$$

$$T^{(i)} = \begin{bmatrix} \tau_0^{(i)} & \tau_1^{(i)} & \ldots & \tau_{q-2}^{(i)} & \tau_{q-1}^{(i)} & \tau_q^{(i)} \\ \tau_1^{(i)} & \tau_2^{(i)} & \ldots & \tau_{q-1}^{(i)} & \tau_q^{(i)} & 0 \\ \tau_2^{(i)} & \tau_3^{(i)} & \ldots & \tau_q^{(ik)} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \tau_q^{(i)} & 0 & \ldots & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tau_0^{(i)} & \tau_1^{(i)} & \tau_2^{(i)} & \ldots & \tau_q^{(i)} \\ 0 & \tau_0^{(i)} & \tau_1^{(i)} & \ldots & \tau_{q-1}^{(i)} \\ 0 & 0 & \tau_0^{(i)} & \ldots & \tau_{q-2}^{(i)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \tau_0^{(i)} \end{bmatrix}$$

Los estimadores de los coeficientes $\theta_1, \theta_2, \ldots, \theta_q$ pueden ser calculados utilizando los τ_j obtenidos en cada iteración.

Si bien estos estimadores son asintóticamente consistentes, $\hat{\theta} \xrightarrow{p} \theta$ y $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$, a diferencia de lo que ocurre en los modelos AR, no son eficientes, pero proporcionan los valores iniciales del método iterativo que converge a los estimadores eficientes de máxima-verosimilitud.

Se verificará que $\hat{\sigma}_\varepsilon^2 = \frac{1}{n} S\left(\hat{\theta}\right)$, donde

$$S\left(\hat{\theta}\right) = \sum_{j=1}^{n} \frac{\left(Y_j - \hat{Y}_j\right)^2}{\rho_{j-1}(Y)} \quad \text{y} \quad \hat{\theta}' = \left[\hat{\theta}_1 \quad \hat{\theta}_2 \quad \ldots \quad \hat{\theta}_q\right]$$

denota el vector de los estimadores $\hat{\theta}$ que minimizan la función de verosimilitud reducida (o condicionada) la cual supone que $\varepsilon_t = 0$, para $t \leq 1$:

$$\ell\left(\hat{\theta}_q\right) = \ln\left[\frac{S\left(\hat{\theta}_q\right)}{n}\right] + \frac{1}{n}\sum_{j=1}^{n} \ln\left[\hat{\rho}_{j-1}(Y)\right]$$

Una segunda aproximación (más aconsejable que la anterior, especialmente en aquellos casos en que la representación MA se encuentre próxima a la zona de no-invertibilidad) se obtiene utilizando el método de máxima verosimilitud exacta, la cual considera a las perturbaciones ε_t (para $t \leq 1$) como coeficientes adicionales a la representación y que, por lo tanto, deben ser estimadas juntamente con los coeficientes θ_j [106].

Se demuestra que en procesos invertibles, para muestras moderadamente grandes, la matriz de varianzas y covarianzas de los estimadores:

---

106. Para valores de n suficientemente grandes, los resultados obtenidos de ambos métodos de estimación serán aproximadamente iguales (ver Brockwell, P.J.; Davis, R.A. (1987)).

$$\Gamma_q(\hat{\theta}) = \begin{bmatrix} \sigma^2(\hat{\theta}_1) & \gamma(\hat{\theta}_1,\hat{\theta}_2) & \dots & \gamma(\hat{\theta}_1,\hat{\theta}_q) \\ \dots & \dots & \dots & \dots \\ \gamma(\hat{\theta}_1,\hat{\theta}_q) & \gamma(\hat{\theta}_2,\hat{\theta}_q) & \dots & \sigma^2(\hat{\theta}_q) \end{bmatrix} \approx \frac{1}{n}\sigma_\varepsilon^2\Gamma_q^{-1}(Y)$$

es similar a la de los modelos $AR(p)$ y, si se supone que $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$, igual que en los modelos $AR(p)$, se verificará que $\hat{\theta}_q \xrightarrow{d} N\left(\theta_q, \frac{1}{n}\sigma_\varepsilon^2\Gamma_q^{-1}(Y)\right)$.

2.11.4.- El proceso MA(1)

En particular, para una representación **MA(1)** de la forma:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} = (1 - \theta_1 B)\varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$: $WN - N(0,\sigma_\varepsilon^2)$) se verifica que:

$$\gamma_0(Y) = \sigma_\varepsilon^2(1 + \theta_1^2)$$

$$\gamma_1(Y) = -\theta_1\sigma_\varepsilon^2$$

$$\gamma_j(Y) = 0 \qquad (j = 1,2,\dots)$$

de modo que la correspondiente función de autocorrelaciones será de la forma:

$$\rho_j(Y) = \begin{cases} -\dfrac{\theta_1}{1+\theta_1^2} & (j = 1) \\ 0 & (j = 2,3,\dots) \end{cases}$$

A partir de esta expresión se obtiene fácilmente que la autocorrelación de primer orden obedece a una ecuación de segundo grado de la forma:

$$\rho_1(Y)\theta_1^2 + \theta_1 + \rho_1(Y) = 0$$

Como $\theta_1 = \dfrac{-1 \pm \sqrt{1 - 4\rho_1^2(Y)}}{2\rho_1(Y)}$ sólo puede asumir valores reales, para que el proceso sea invertible necesariamente se tiene que verificar que $-\frac{1}{2} < \rho_1(Y) < \frac{1}{2}$. Asimismo, como el producto de las raíces es igual a la unidad, las soluciones de la ecuación serán θ_1 y θ_1^{-1}. Esto podría sugerir que es posible definir dos representaciones $MA(1)$ asociadas a la misma función de autocorrelaciones, pero debe tenerse en cuenta que la condición de invertibilidad exige que $|\theta_1| < 1$ y que, por lo tanto, hay una única representación $MA(1)$ invertible para cada función de autocorrelaciones (una forma general de esta condición de unicidad será demostrada en la próxima sección).

En términos más generales se puede asegurar que, para cada representación $MA(q)$, existe una única función de autocorrelaciones, pero la implicación inversa no siempre se verifica. Lo único que se puede asegurar es que a una función de autocorrelaciones dada sólo le corresponde una representación $MA(q)$

invertible.

Los coeficientes de autocorrelación parcial quedan definidos de la siguiente forma:

$$\phi_{11}(Y) = \rho_1(Y) = \frac{-\theta_1}{1+\theta_1^2} = \frac{-\theta_1(1-\theta_1^2)}{1-\theta_1^4}$$

$$\phi_{22}(Y) = \frac{-\rho_1^2(Y)}{1-\rho_1^2(Y)} = \frac{-\dfrac{\theta_1^2}{(1+\theta_1^2)^2}}{1-\dfrac{\theta_1^2}{(1+\theta_1^2)^2}} = \frac{-\theta_1^2(1-\theta_1^2)}{(1+\theta_1^2+\theta_1^4)(1-\theta_1^2)} = \frac{-\theta_1^2(1-\theta_1^2)}{1-\theta_1^6}$$

$$\phi_{33}(Y) = \frac{-\phi_{22}(Y)\rho_1(Y)}{1-[\phi_{11}(Y)-\phi_{11}(Y)\phi_{22}(Y)]\rho_1(Y)} = \frac{-\theta_1^3(1-\theta_1^2)}{(1+\theta_1^2)(1+\theta_1^2)(1-\theta_1^2)} = \frac{-\theta_1^3(1-\theta_1^2)}{(1+\theta_1^4)(1-\theta_1^4)}$$

Es decir que, en general, se puede escribir $\phi_{jj}(Y) = \dfrac{-\theta_1^j(1-\theta_1^2)}{1-\theta_1^{2(j+1)}}$ $(j = 1,2,3,....)$. Luego, se puede concluir en forma inmediata que $|\phi_{jj}| < \theta_1^j$ y, por lo tanto, que la función de autocorrelaciones parciales está dominada por una función exponencialmente decreciente. Si se verifica que $\rho_1(Y) > 0$ y, por lo tanto, que $\theta_1 < 0$, entonces los $\phi_{jj}(Y)$ serán de signo alternado; viceversa, si $\rho_1(Y) < 0$, la función de autocorrelaciones parciales será negativa.

La función de densidad espectral será de la forma:

$$\omega_Y(\alpha) = \frac{1}{2\pi}(1+\theta_1^2 - 2\theta_1\cos(\alpha))$$

La forma AR de esta representación $MA(1)$ será de la forma:

$$[\Theta(B)]^{-1}Y_t = (1-\theta_1 B)^{-1}Y_t = Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-1} + ... = \varepsilon_t$$

de modo que el estimador mínimo-cuadrático de θ_1 puede ser obtenido de minimizar la expresión[107]:

$$\sum_{t=1}^{n}\varepsilon_t^2 = \sum_{t=1}^{n}(Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + ... + \theta_1^{n-1}Y_1)^2$$

(obsérvese que éste es un problema de cuadrados mínimos no-lineales, solucionable mediante minimización numérica)[108]

2.11.5.- El proceso MA(2)

Sea una representación $MA(2)$ de la forma:

---

[107]. (obsérvese que éste es un problema de cuadrados mínimos no-lineales, solucionable mediante minimización numérica) (ver Judge, G.G.; Griffiths, W.E.; Carter-Hill, R.; Lütkepohl, H.; Lee, T. (1985)).

[108]. Ver Judge, G.G.; Griffiths, W.E.; Carter-Hill, R.; Lütkepohl, H.; Lee, T. (1985).

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

La condición de invertibilidad es que las raíces de la ecuación característica:

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 = 0$$

sean, en módulo, mayores que la unidad. Es decir, que los coeficientes sean tales que $\theta_2 + \theta_1 < 1$, $\theta_2 - \theta_1 < 0$ y $|\theta_2| < 1$.

La función de autocorrelaciones es de la forma:

$$\begin{cases} \rho_1(Y) = \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} \\ \rho_2(Y) = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} \\ \rho_j(Y) = 0 \qquad (j \geq 3) \end{cases}$$

De estas ecuaciones se concluye que, si el proceso es invertible, las autocorrelaciones de primero y segundo orden estarán incluidas en la región limitada por las curvas:

$$\begin{cases} \rho_2(Y) + \rho_1(Y) = -\dfrac{1}{2} \\ \rho_2(Y) - \rho_1(Y) = -\dfrac{1}{2} \\ \rho_1^2(Y) = 4\rho_2(Y)(1-2\rho_2(Y)) \end{cases}$$

Igual que la función de autocorrelaciones de un $AR(2)$, la función de autocorrelaciones parciales del $MA(2)$ está formada por la suma de dos funciones exponenciales o por una función sinusoidal amortiguada, según que las raíces de la ecuación característica sean reales o complejas.

De acuerdo a lo demostrado en la Sec.2.11.3, los valores iniciales de la estimación Máximo-verosímil de los coeficientes del $MA(2)$ pueden ser expresados de la siguiente forma:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\gamma}_0(Y)}{1+\hat{\theta}_1^2+\hat{\theta}_2^2}$$

$$\hat{\theta}_2 = -\frac{\hat{\gamma}_2(Y)}{\hat{\sigma}_\varepsilon^2}$$

$$\hat{\theta}_1 = -\left(\frac{\hat{\gamma}_1(Y)}{\hat{\sigma}_\varepsilon^2} - \hat{\theta}_1\hat{\theta}_2\right)$$

Ejemplo 2.7:

Sea un proceso $\{Y_t\}$ de parámetro mensual formado por variables centradas, para cuyo análisis se cuenta con una serie cronológica para el período [ENE-2000;DIC-2004] y cuya trayectoria figura en el

GRAFICO 2.15: Trayectoria de $\{Y_t\}$

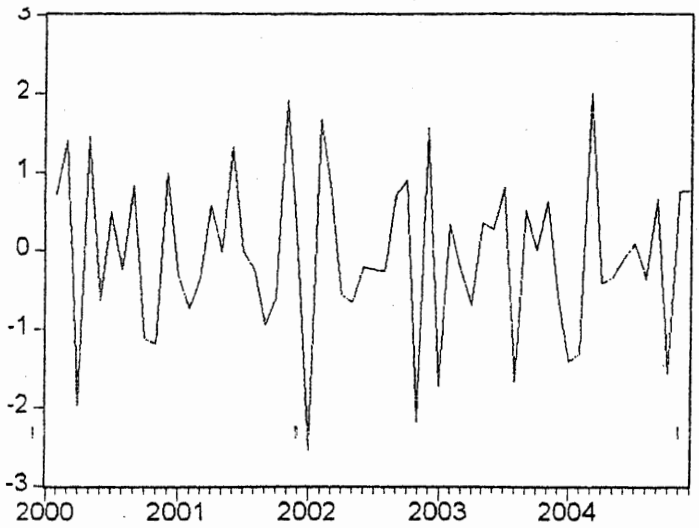


GRAFICO 2.15.

De la observación del correlograma se concluyó que $\{Y_t\}$ es estacionario de segundo orden y admite una representación $MA(2)$ de la forma:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

A partir de los datos proporcionados por la mencionada serie cronológica se obtuvieron los siguientes resultados: $\hat{\sigma}_Y^2 = 1.058$, $\hat{\rho}_1(Y) = -0.388$ y $\hat{\rho}_2(Y) = -0.105$.

Para obtener los estimadores preliminares de los coeficientes σ_ε^2 y θ_1 a partir del proceso iterativo lineal, se consideraron los valores iniciales $\hat{\theta}_1^{(0)} = \hat{\theta}_2^{(0)} = 0$. Obteniéndose en las iteraciones primera y segunda los siguientes resultados:

Primera iteración

$$\left[\sigma_\varepsilon^{(1)}\right]^2 = \hat{\gamma}_0(Y) = 1.058$$

$$\hat{\theta}_2^{(1)} = -\frac{\hat{\gamma}_2(Y)}{\left[\hat{\sigma}_\varepsilon^{(1)}\right]^2} = -\frac{-0.111}{1.058} = 0.105$$

$$\hat{\theta}_1^{(1)} = -\frac{\hat{\gamma}_1(Y)}{\left[\sigma_\varepsilon^{(1)}\right]^2} = -\frac{-0.411}{1.058} = 0.388$$

Segunda iteración:

$$\left[\hat{\sigma}_{\varepsilon}^{(2)}\right]^2 = \frac{\hat{\gamma}_0(Y)}{1+\left[\hat{\theta}_1^{(1)}\right]^2+\left[\hat{\theta}_2^{(1)}\right]^2} = \frac{1.058}{1.162} = 0.911$$

$$\hat{\theta}_2^{(2)} = -\frac{\hat{\gamma}_2(Y)}{\left[\hat{\sigma}_{\varepsilon}^{(2)}\right]^2} = -\frac{-0.111}{0.828} = 0.122$$

$$\hat{\theta}_1^{(2)} = -\left(\frac{\hat{\gamma}_1(Y)}{\left[\hat{\sigma}_{\varepsilon}^{(2)}\right]} - \hat{\theta}_1^{(1)}\hat{\theta}_2^{(1)}\right) = -(-0.451-0.040) = 0.491$$

En las iteraciones sucesivas se obtuvieron los siguientes resultados:

i	$[\hat{\sigma}_{\varepsilon}^{(i)}]^2$	$\hat{\theta}_2^{(i)}$	$\hat{\theta}_1^{(i)}$
3	0,842	0,132	0,547
4	0,803	0,138	0,583
5	0,778	0,143	0,608
6	0,761	0,146	0,626
7	0,749	0,148	0,640
8	0,739	0,150	0,650
9	0,732	0,152	0,659
10	0,726	0,153	0,665
11	0,722	0,154	0,671
12	0,718	0,155	0,675
13	0,715	0,155	0,678
14	0,713	0,156	0,681
15	0,711	0,156	0,684

Para la utilización del algoritmo de Newton-Raphson se consideraron los siguientes valores iniciales:

$$\tau_0^{(0)} = \hat{\sigma}_{\varepsilon} = \hat{\sigma}_Y = 1.029$$

$$\tau_1^{(0)} = 0$$

$$\tau_2^{(0)} = 0$$

$$\hat{\theta}_1^{(0)} = \frac{\tau_1^{(0)}}{\tau_0^{(0)}} = 0$$

$$\hat{\theta}_2^{(0)} = \frac{\tau_2^{(0)}}{\tau_0^{(0)}} = 0$$

$$f_0^{(0)} = \left(\tau_0^{(0)}\right)^2 + \left(\tau_1^{(0)}\right)^2 + \left(\tau_2^{(0)}\right)^2 - \hat{\gamma}_0(Y) = 0$$

$$f_1^{(0)} = \tau_0^{(0)}\tau_1^{(0)} + \tau_1^{(0)}\tau_2^{(0)} - \hat{\gamma}_1(Y) = 0.411$$

$$f_2^{(0)} = \tau_0^{(0)}\tau_2^{(0)} - \hat{\gamma}_2(Y) = 0.111$$

$$\left|T^{(0)}\right| = 2\left\{\left(\tau_0^{(0)}\right)^3 + \left(\tau_0^{(0)}\right)^2\tau_2^{(0)} - \tau_0^{(0)}\left(\tau_1^{(0)}\right)^2 + \left(\tau_1^{(0)}\right)^2\tau_2^{(0)} - \tau_0^{(0)}\left(\tau_2^{(0)}\right)^2 - \left(\tau_2^{(0)}\right)^3\right\} = 2(1.029)^3 = 2.179$$

$$\left[T^{(0)}\right]^{-1} f^{(0)} = \frac{1}{|T|}\begin{bmatrix} \tau_0^{(0)}\left(\tau_0^{(0)} + \tau_2^{(0)}\right)f_0^{(0)} - 2\tau_0^{(0)}\tau_1^{(0)}f_1^{(0)} + 2\left\{\left(\tau_1^{(0)}\right)^2 - \tau_2^{(0)}\left(\tau_0^{(0)} + \tau_2^{(0)}\right)\right\}f_2^{(0)} \\ \tau_1^{(0)}\left(\tau_2^{(0)} - \tau_0^{(0)}\right)f_0^{(0)} + 2\left[\left(\tau_0^{(0)}\right)^2 - \left(\tau_2^{(0)}\right)^2\right]f_1^{(0)} + 2\tau_1^{(0)}\left(\tau_2^{(0)} - \tau_0^{(0)}\right)f_2^{(0)} \\ -\tau_2^{(0)}\left(\tau_0^{(0)} + \tau_2^{(0)}\right)f_0^{(0)} + 2\tau_1^{(0)}t_2^{(0)}f_1^{(0)} + 2\left[\tau_0^{(0)}\left(\tau_0^{(0)} + \tau_2^{(0)}\right) - \left(\tau_1^{(0)}\right)^2\right]f_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.798 \\ 0.215 \end{bmatrix}$$

En las iteraciones sucesivas siguientes se obtuvieron los siguientes resultados:

Primera iteración

$$\tau_0^{(1)} = \tau_0^{(0)} - 0 = 1.029$$

$$\tau_1^{(1)}\tau_1^{(0)} - 0.399 = -0.399$$

$$\tau_2^{(1)} = \tau_2^{(0)} - 0.235 = -0.235$$

$$\hat{\theta}_1^{(1)} = \frac{\tau_1^{(1)}}{\tau_0^{(1)}} = 0.338$$

$$\hat{\theta}_2^{(1)} = \frac{\tau_2^{(1)}}{\tau_0^{(1)}} = 0.228$$

$$f_0^{(1)} = \left(\tau_0^{(1)}\right)^2 + \left(\tau_1^{(1)}\right)^2 + \left(\tau_2^{(1)}\right)^2 - \hat{\gamma}_0(Y) = 0.214$$

$$f_1^{(1)} = \tau_0^{(1)}\tau_1^{(1)} + \tau_1^{(1)}\tau_2^{(1)} - \hat{\gamma}_1(Y) = 0.094$$

$$f_2^{(1)} = \tau_0^{(1)}\tau_2^{(1)} - \hat{\gamma}_2(Y) = -0.131$$

$$\left|T^{(1)}\right| = 2\left[\left(\tau_0^{(1)}\right)^3 + \left(\tau_0^{(1)}\right)^2\tau_2^{(1)} - \tau_0^{(1)}\left(\tau_1^{(1)}\right)^2 + \left(\tau_1^{(1)}\right)^2\tau_2^{(1)} - \tau_0^{(1)}\left(\tau_2^{(1)}\right)^2 - \left(\tau_2^{(1)}\right)^3\right] = 1.192$$

$$\left[T^{(1)}\right]^{-1} f^{(1)} = \frac{1}{|T|}\begin{bmatrix} \tau_0^{(1)}\left(\tau_0^{(1)} + \tau_2^{(1)}\right)f_0^{(1)} - 2\tau_0^{(1)}\tau_1^{(1)}f_1^{(1)} + 2\left[\left(\tau_1^{(1)}\right)^2 - \tau_2^{(1)}\left(\tau_0^{(1)} + \tau_2^{(1)}\right)\right]f_2^{(1)} \\ \tau_1^{(1)}\left(\tau_2^{(1)} + \tau_0^{(1)}\right)f_0^{(1)} + 2\left[\left(\tau_0^{(1)}\right)^2 - \left(\tau_2^{(1)}\right)^2\right]f_1^{(1)} + 2\tau_1^{(1)}\left(\tau_2^{(1)} - \tau_0^{(1)}\right)f_2^{(1)} \\ -\tau_2^{(1)}\left(\tau_0^{(1)} + \tau_2^{(1)}\right)f_0^{(1)} + 2\tau_1^{(1)}\tau_2^{(1)}f_1^{(1)} + 2\left[\tau_0^{(1)}\left(\tau_0^{(1)} + \tau_2^{(1)}\right) - \left(\tau_1^{(1)}\right)^2\right]f_2^{(1)} \end{bmatrix} = \begin{bmatrix} 0.135 \\ 0.138 \\ 0.096 \end{bmatrix}$$

Segunda iteración:

$$\tau_0^{(2)} = \tau_0^{(1)} - 0.135 = 0.894$$

$$\tau_1^{(2)}\tau_1^{(1)} - 0.138 = -0.537$$

$$\tau_2^{(2)} = \tau_2^{(1)} + 0.096 = -0.139$$

$$\hat{\theta}_1^{(2)} = \frac{\tau_1^{(2)}}{\tau_0^{(2)}} = 0.601$$

$$\hat{\theta}_2^{(2)} = \frac{\tau_2^{(2)}}{\tau_0^{(2)}} = 0.155$$

$$f_0^{(2)} = \left(\tau_0^{(2)}\right)^2 + \left(\tau_1^{(2)}\right)^2 + \left(\tau_2^{(2)}\right)^2 - \hat{\gamma}_0(Y) = 0.047$$

$$f_1^{(2)} = \tau_0^{(2)}\tau_1^{(2)} + \tau_1^{(2)}\tau_2^{(2)} - \hat{\gamma}_1(Y) = 0.0056$$

$$f_2^{(2)} = \tau_0^{(2)}\tau_2^{(2)} - \hat{\gamma}_2(Y) = -0.013$$

$$\left|T^{(2)}\right| = 2\left[\left(\tau_0^{(2)}\right)^3 + \left(\tau_0^{(2)}\right)^2\tau_2^{(2)} - \tau_0^{(2)}\left(\tau_1^{(2)}\right)^2 + \left(\tau_2^{(1)}\right)^2\tau_2^{(2)} - \tau_0^{(2)}\left(\tau_2^{(2)}\right)^2 - \left(\tau_2^{(2)}\right)^3\right] = 0.583$$

$$\left[T^{(2)}\right]^{-1} f^{(2)} = \frac{1}{|T|}\begin{bmatrix} \tau_0^{(2)}\left(\tau_0^{(2)} + \tau_2^{(2)}\right)f_0^{(2)} - 2\tau_0^{(2)}\tau_1^{(2)}f_1^{(2)} + 2\left[\left(\tau_1^{(2)}\right)^2 - \tau_2^{(1)}\left(\tau_0^{(2)} + \tau_2^{(2)}\right)\right]f_2^{(2)} \\ \tau_1^{(2)}\left(\tau_2^{(2)} - \tau_0^{(2)}\right)f_0^{(2)} + 2\left[\left(\tau_0^{(2)}\right)^2 - \left(\tau_2^{(2)}\right)^2\right]f_1^{(2)} + 2\tau_1^{(2)}\left(\tau_2^{(2)} - \tau_0^{(2)}\right)f_2^{(2)} \\ -\tau_2^{(2)}\left(\tau_0^{(2)} + \tau_2^{(2)}\right)f_0^{(2)} + 2\tau_1^{(2)}\tau_2^{(2)}f_1^{(2)} + 2\left[\tau_0^{(2)}\left(\tau_0^{(2)} + \tau_2^{(2)}\right) - \left(\tau_1^{(2)}\right)^2\right]f_2^{(2)} \end{bmatrix} = \begin{bmatrix} 0.0559 \\ 0.0355 \\ -0.0074 \end{bmatrix}$$

Tercera iteración[109]:

$$\tau_0^{(3)} = \tau_0^{(2)} - 0.0554 = 0.838$$

$$\tau_1^{(3)}\tau_1^{(2)} - 0 - 0355 = -0.5725$$

$$\tau_2^{(3)} = \tau_2^{(2)} + 0.0074 = -0.132$$

$$\hat{\theta}_1^{(3)} = \frac{\tau_1^{(3)}}{\tau_0^{(3)}} = 0.684$$

$$\hat{\theta}_2^{(3)} = \frac{\tau_2^{(3)}}{\tau_0^{(3)}} = 0.157$$

2.12.- Los procesos autorregresivos de promedios móviles de orden (p,q)

2.12.1.- Las condiciones de estacionariedad e invertibilidad

Una representación que suele resultar útil en las aplicaciones, es la que considera una parte componente autorregresiva de orden p y una de promedios móviles de orden q :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}$$

---

109. La diferencia numérica entre los resultados obtenidos en ambos métodos se debe exclusivamente a la aproximación decimal utilizada.

(donde: **i)** $\{Y_t\}$ denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$). Este proceso -que se conoce como $ARMA(p,q)$- puede ser interpretado como un proceso $AR(p)$ de la forma $\Phi(B)Y_t = e_t$, donde $\{e_t\}$ admite, a su vez, una representación $MA(q)$ de la forma $e_t = \Theta(B)\varepsilon_t$. O, como un proceso $MA(q)$ de la forma $Y_t = \Theta(B)Z_t$, donde $\{Z_t\}$ admite una representación $AR(p)$ de la forma $\Phi(B)Z_t = \varepsilon_t$.

Utilizando el operador "backward", se puede escribir:

$$\left(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p\right)Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q\right)\varepsilon_t$$

o, en forma compacta:

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

donde $\Phi(B)$ y $\Theta(B)$ definen, respectivamente. sendos polinomios de grados p y q que no poseen raíces comunes (debe tenerse en cuenta que, en caso contrario, el orden (p,q) de la representación podría ser reducido).

La expresión MA de este proceso es de la forma:

$$Y_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \frac{1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p}\varepsilon_t = \frac{1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q}{\left(1 - G_1^{-1}B\right)\left(1 - G_2^{-1}B\right)...\left(1 - G_p^{-1}B\right)}\varepsilon_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \Psi(B)\varepsilon_t$$

donde, como ya se señaló en la Sec. 2.10.1, las G_i^{-1} $(i = 1,2,...,p)$ denotan las inversas de las raíces de la ecuación característica de la componente AR, $\Phi(B) = 0$ y los coeficientes ψ_j pueden ser obtenidos a partir de las siguientes fórmulas de recurrencia:

$$\psi_j = \begin{cases} \phi_1 - \theta_1 & (j = 1) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_{j-1}\psi_1 + \phi_j - \theta_j & \left(2 \le j \le \min(p,q)\right) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_p\psi_1 - \theta_j & \left(\min(p,q) < j \le Max(p,q)\right) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_p\psi_{j-p} & \left(j > Max(p,q)\right) \end{cases}$$

Obsérvese que, para $j > Max(p,q)$, los coeficientes ψ_j satisfacen la siguiente ecuación en-diferencias:

$$\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} - ... - \phi_p\psi_{j-p} = \Phi(B)\psi_j = \left(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p\right)\psi_j = 0$$

(debe tenerse en cuenta que, en este caso, el operador B afecta al subíndice j), lo cual indica que su comportamiento está definido por una combinación de funciones exponenciales decrecientes y sinusoidales amortiguadas.

Si los polinomios $\Phi(B)$ y $\Theta(B)$ no tienen raíces comunes, la **condición de estacionariedad** se verificará cuando todas las raíces de la ecuación característica de la componente AR sean, en módulo, mayores que la unidad, $|G_i| > 1\,(i = 1,2,...,p)$.

De la misma forma, la expresión AR del proceso $\{Y_t\}$ será:

$$\frac{\Phi(B)}{\Theta(B)}Y_t = \frac{1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p}{1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q}Y_t = \frac{1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p}{\left(1-H_1^{-1}B\right)\left(1-H_2^{-1}B\right)...\left(1-H_q^{-1}B\right)}Y_t = \Pi(B)Y_t = \varepsilon_t$$

Donde, como ya se señaló en la Sec. 2.11.1, las $H_i^{-1}\ (i = 1,2,...,q)$ denotan las inversas de las raíces de la ecuación característica de la componente MA, $\Theta(B) = 0$ y los coeficientes π_j pueden ser obtenidos a partir de las siguientes fórmulas de recurrencia:

$$\pi_j = \begin{cases} \phi_1 - \theta_1 & (j = 1) \\ \phi_j + \theta_1\pi_{j-1} + \theta_2\pi_{j-2}+...+\theta_{j-1}\pi_1 - \theta_j & \left(2 \le j \le \min(p,q)\right) \\ \phi_j + \theta_1\pi_{j-1} + \theta_2\pi_{j-2}+...+\theta_q\pi_1 - \theta_j & \left(\min(p,q) < j \le Max(p,q)\right) \\ \theta_1\pi_{j-1} + \theta_2\pi_{j-2}+...+\theta_q\pi_{j-q} & \left(j > Max(p,q)\right) \end{cases}$$

Obsérvese que, para $j > Max(p,q)$ los coeficientes π_j satisfacen la ecuación en-diferencias:

$$\pi_j - \theta_1\pi_{j-1}-...-\theta_q\pi_{j-q} = \Theta(B)\pi_j = \left(1-\theta_1 B-...-\theta_q B^q\right)\pi_j = 0$$

(debe tenerse en cuenta que, en este caso, el operador B afecta al subíndice j).

Si los polinomios $\Theta(B)$ y $\Phi(B)$ no tienen raíces comunes, la condición de invertibilidad se verificará cuando todas las raíces de la ecuación característica de la componente MA sean, en módulo, mayores que la unidad, $|H_i| > 1\ (i = 1,2,...,q)$.

A partir de la representación MA de orden infinito del proceso $ARMA(p,q)$ y suponiendo que se cuenta con información acerca de las realizaciones del proceso hasta el momento $t-n$, se puede plantear una **representación truncada** finita en términos de una combinación lineal de $n+1$ "shocks" aleatorios, $\varepsilon_t, \varepsilon_{t-1},...,\varepsilon_{t-n}$. Esta forma de representación truncada consiste en considerar a la representación $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ (donde: **i)** las variables Y_t son centradas y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$) como una ecuación en-diferencias cuya solución es de la forma:

$$Y_t = f^{(n)}\left(Y_{t-n}, Y_{t-n-1},...\right) + e(t-n,n)$$

donde: **i)** $f^{(n)}\left(Y_{t-n}, Y_{t-n-1},...\right)$, que denota la **función complementaria** (y que representa la parte de $\{Y_t\}$ conocida al momento $t-n$), es la solución general de la ecuación en-diferencias:

$$\Phi(B)f^{(n)}\left(Y_{t-n}, Y_{t-n-1},...\right) = 0$$

y, por lo tanto, asumirá la forma:

$$f^{(n)}\left(Y_{t-n}, Y_{t-n-1},...\right) = A_1\left(G_1^{-1}\right)^n + A_2\left(G_2^{-1}\right)^n+...+A_p\left(G_p^{-1}\right)^n$$

suponiendo que las raices sean todas distintas entre sí y **ii)** $e(t-n,n)$, que denota la **integral particular** (y

que representa la parte de $\{Y_t\}$ no-predecible al momento $t-n$), es cualquier función que satisfaga la ecuación $\Phi(B)e(t-n,n) = \Theta(B)\varepsilon_t$ y asumirá la forma:

$$e(t-n,n) = \begin{cases} 0 & (t \le t-n) \\ \varepsilon_t + \psi_1\varepsilon_{t-1} + \ldots + \psi_{n-1}\varepsilon_{t-n+1} & (t > t-n) \end{cases}$$

(donde los ψ_j se calculan a partir de la expresión iterativa que figura en las páginas anteriores).

En consecuencia, la expresión truncada de la representación MA de orden infinito del proceso $\{Y_t\}$: $ARMA(p,q)$ puede ser escrita de acuerdo con la siguiente descomposición agregativa:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots) \qquad (\psi_0 = 1)$$

Obsérvese que, de acuerdo con las definiciones anteriores, se puede escribir:

$$f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots) = f^{(n+n^*)}(Y_{t-n-n^*}, Y_{t-n-n^*-1}, \ldots) + \psi_n\varepsilon_{t-n} + \psi_{n+1}\varepsilon_{t-n-1} + \ldots + \psi_{n+n^*-1}\varepsilon_{t-n-n^*+1}$$

por lo que se puede concluir que la función complementaria varía con el origen $t-n$. Asimismo, se demuestra en forma inmediata que $E(Y_t / Y_{t-n}, Y_{t-n-1}, \ldots) = f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots)$.

2.12.2.- Las funciones de autocorrelaciones y de autocorrelaciones parciales

Multiplicando m. a m. la representación del $ARMA(p,q)$ por Y_{t-j} y aplicando m. a m. el operador esperanza matemática:

$$E(Y_{t-j}Y_t) = \phi_1 E(Y_{t-j}Y_{t-1}) + \phi_2 E(Y_{t-j}Y_{t-2}) + \ldots + \phi_p E(Y_{t-j}Y_{t-p}) + E(Y_{t-j}\varepsilon_t) - \theta_1 E(Y_{t-j}\varepsilon_{t-1}) - \ldots - \theta_q E(Y_{t-j}\varepsilon_{t-q})$$

se obtiene la función de autocovarianzas:

$$\gamma_0(Y) \equiv \sigma_Y^2 = \phi_1\gamma_1(Y) + \phi_2\gamma_2(Y) + \ldots + \phi_p\gamma_p(Y) + \sigma_\varepsilon^2 - \theta_1\gamma(Y_t, \varepsilon_{t-1}) - \ldots - \theta_q\gamma(Y_t, \varepsilon_{t-q})$$

$$\gamma_1(Y) = \phi_1\gamma_0(Y) + \phi_2\gamma_1(Y) + \ldots + \phi_p\gamma_{p-1}(Y) - \theta_1\sigma_\varepsilon^2 - \theta_2\gamma(Y_{t-1}, \varepsilon_{t-2}) - \ldots - \theta_q\gamma(Y_{t-1}, \varepsilon_{t-q})$$

.....

$$\gamma_q(Y) = \phi_1\gamma_{q-1}(Y) + \phi_2\gamma_{|q-2|}(Y) + \ldots + \phi_p\gamma_{|q-p|}(Y) - \theta_q\sigma_\varepsilon^2$$

$$\gamma_j(Y) = \phi_1\gamma_{j-1}(Y) + \phi_2\gamma_{|j-2|}(Y) + \ldots + \phi_p\gamma_{|j-p|}(Y) \qquad (j = q+1, q+2, \ldots)$$

De lo que se puede concluir que la función de autocovarianzas y, por lo tanto, la función de autocorrelaciones es infinita y, a partir del orden $q+1$, queda definida por una combinación de funciones exponenciales decrecientes y sinusoidales amortiguadas.

En el **CUADRO 2.12** figuran las funciones de autocorrelaciones y autocorrelaciones parciales típicas de un proceso $ARMA(1,1)$ estacionario e invertible.

Teniendo en cuenta que, de acuerdo a lo expresado en la Sec. 2.9.1, la función de autocovarianzas

puede ser expresada de la siguiente forma: $\gamma_j(Y) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}$ $(\psi_0 = 1; j = 0,1,2,...)$, los valores de $\gamma_j(Y)$ pueden ser calculados utilizando la relación $\Psi(B)\Phi(B) = \Theta(B)$, a partir de la cual se obtiene que:

$$\psi_j = \begin{cases} \phi_1 - \theta_1 & (j = 1) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_{j-1}\psi_1 + \phi_j - \theta_j & (2 \le j \le \min(p,q)) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_p\psi_1 - \theta_j & (\min(p,q) < j \le Max(p,q)) \\ \phi_1\psi_{j-1} + \phi_2\psi_{j-2} + ... + \phi_p\psi_{j-p} & (j > Max(p,q)) \end{cases}$$

Por otra parte, de acuerdo a lo mencionado en la Sec. 2.12.1, la representación MA de este proceso mixto es de orden infinito. Por lo tanto, su función de autocorrelaciones parciales será de extensión infinita, y tendrá un comportamiento de acuerdo con una combinación de funciones exponenciales o de funciones sinusoidales amortiguadas.

Contrariamente a lo que ocurre con las representaciones $AR(p)$ y $MA(q)$, las funciones de autocorrelaciones y autocorrelaciones parciales de un proceso ARMA no resultan útiles para especificar el orden (p,q). A fin de resolver este problema Tsay, R.S.; Tiao, G.C. (1984) propusieron un método que utiliza la **función de autocorrelaciones extendida**, basado en el principio que, si se puede obtener una estimación consistente de la parte componente AR de la representación ARMA, entonces se puede derivar la parte componente MA. La función de autocorrelaciones de esta parte componente MA permite identificar el orden de la representación ARMA.

De acuerdo a lo demostrado en la sección precedente, un proceso $ARMA(p,q)$ puede ser expresado, en función de las inversas de las raíces de las ecuaciones características, de la siguiente forma:

$$\prod_{i=1}^{p}(1 - G_i^{-1}B)Y_t = \prod_{i=1}^{q}(1 - H_i^{-1}B)\varepsilon_t$$

y la función generatriz de autocovarianzas del proceso puede ser expresada como:

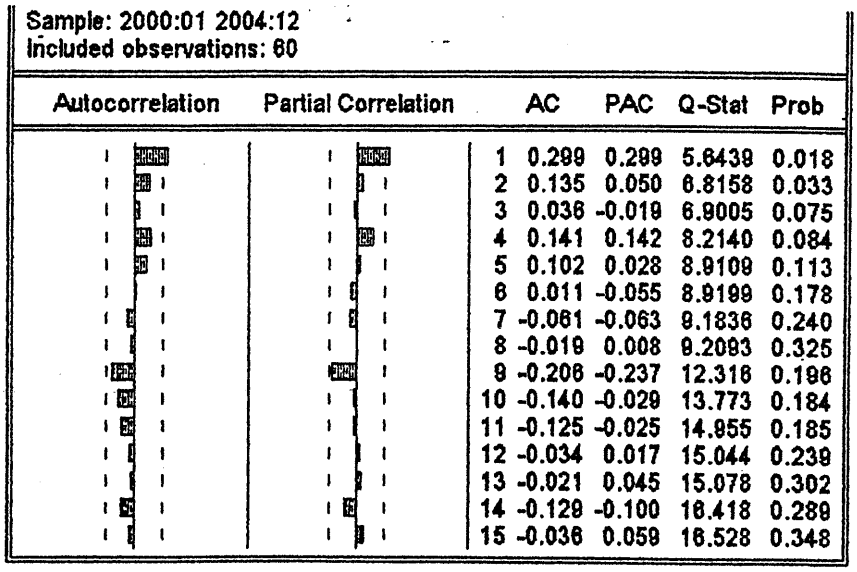
$$\gamma(B) = \sigma_\varepsilon^2 \prod_{i=1}^{p}(1 - G_i^{-1}B)^{-1}(1 - G_i^{-1}B^{-1})^{-1} \prod_{i=1}^{q}(1 - H_i^{-1}B)(1 - H_i^{-1}B^{-1}) =$$

$$= \frac{\prod_{i=1}^{q}(H_i^{-1})^2(H_i - B)(H_i - B^{-1})}{\prod_{i=1}^{p}(1 - G_i^{-1}B)(1 - G_i^{-1}B^{-1})}\sigma_\varepsilon^2 =$$

$$= \left[\prod_{i=1}^{q}(H_i^{-1})^2\right] \frac{\prod_{i=1}^{q}(1 - H_iB)(1 - H_iB^{-1})}{\prod_{i=1}^{p}(1 - G_i^{-1}B)(1 - G_i^{-1}B^{-1})}\sigma_\varepsilon^2 = \gamma_{Y^*}(B)$$

(donde $\gamma_{Y^*}(B)$ denota la función generatriz de autocovarianzas para un proceso $ARMA(p,q)$ representable

CUADRO 2.12: Correlograma correspondiente a un proceso $\{Y_t\}$ que admite una representación ARMA(1,1) con φ_1=0.6 y θ_1=0.70.



Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.299	0.299	5.6439	0.018
		2	0.135	0.050	6.8158	0.033
		3	0.036	-0.019	6.9005	0.075
		4	0.141	0.142	8.2140	0.084
		5	0.102	0.028	8.9109	0.113
		6	0.011	-0.055	8.9199	0.178
		7	-0.061	-0.063	9.1836	0.240
		8	-0.019	0.008	9.2093	0.325
		9	-0.206	-0.237	12.316	0.196
		10	-0.140	-0.029	13.773	0.184
		11	-0.125	-0.025	14.955	0.185
		12	-0.034	0.017	15.044	0.239
		13	-0.021	0.045	15.078	0.302
		14	-0.129	-0.100	16.418	0.289
		15	-0.036	0.059	16.528	0.348

de la siguiente forma[110]: $\prod_{i=1}^{q}\left(1-G_i^{-1}B\right)Y_t = \prod_{i=1}^{q}\left(1-H_iB\right)\varepsilon_t$. Luego, se puede concluir que a todos las representaciones de la forma $\prod_{i=1}^{p}\left(1-G_i^{-1}B\right)Y_t = \prod_{i=1}^{q}\left[1-\left(H_i^{-1}\right)^{\pm 1}B\right]\varepsilon_t$ les corresponde la misma función de autocovarianzas. Ahora bien, si se supone que el proceso $\{Y_t\}$ es estacionario e invertible, no será posible hallar ninguna raíz H_i^{-1} tal que $\left(H_i^{-1}\right)^{+1}$ y $\left(H_i^{-1}\right)^{-1}$ cumplan simultáneamente la condición de ser, en módulo, menores que la unidad. Es decir, que habrá una sola representación estacionaria e invertible para una función de autocovarianzas dada, y viceversa (esta correspondencia biunívoca entre la estructura de autocovarianzas y las representaciones ARMA estacionarias e invertibles es la que permite la utilización de la función de autocovarianzas en la determinación de los valores de p y q).

Sean $\{Y_{1t}\}$ e $\{Y_{2t}\}$ dos procesos unidimensionales, estacionarios, incorrelacionados entre sí que admiten representaciones AR de la forma:

$$Y_{1t} = \phi_1^{(1)}Y_{1,t-1} + \phi_2^{(1)}Y_{1,t-2} + \ldots + \phi_{p_1}^{(1)}Y_{1,t-p_1} + \varepsilon_t^{(1)}$$

---

[110]. Supóngase que $q=1$. Entonces será $\left(1-H_1^{-1}B\right)\left(1-H_1^{-1}B^{-1}\right) = \left(1-\frac{B}{H_1}\right)\left(1-\frac{1}{H_1B}\right) = 1-\frac{1}{H_1B}-\frac{B}{H_1}+\frac{1}{H_1^2}$. Por otra parte, $\left(1-H_1B\right)\left(1-H_1B^{-1}\right) = H_1^2\left(H_1^{-1}-B\right)\left(H_1^{-1}-B^{-1}\right) = H_1^2\left(\frac{1}{H_1}-B\right)\left(\frac{1}{H_1}-\frac{1}{B}\right) = H_1^2\left(\frac{1}{H_1^2}-\frac{1}{H_1B}-\frac{1}{H_1}+1\right)$. Es decir que $\left(1-H_1^{-1}B\right)\cdot\left(1-H_1^{-1}B^{-1}\right) = \left(H_1^{-1}\right)^2\left(1-H_1B\right)\left(1-H_1B^{-1}\right)$.

$$Y_{2t} = \phi_1^{(2)} Y_{2,t-1} + \phi_2^{(2)} Y_{2,t-2} + ... + \phi_{p_2}^{(2)} Y_{2,t-p_2} + \varepsilon_t^{(2)}$$

(donde: **i)** Y_{1t} e Y_{2t} denotan variables centradas y **ii)** $\left\{\varepsilon_t^{(1)}\right\}: WN - N\left(0, \sigma_{\varepsilon 1}^2\right)$ y $\left\{\varepsilon_t^{(2)}\right\}: WN - N\left(0, \sigma_{\varepsilon 2}^2\right)$ son incorrelacionados entre sí) y sea un proceso bidimensional $\left\{Y_t\right\} = \left\{Y_{1t}\right\} + \left\{Y_{2t}\right\}$.

Utilizando la expresión en función del operador "backward" y multiplicando m. a m. por $1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}$, la primera ecuación queda expresada de la siguiente forma:

$$\left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(1)} B^{p_1}\right)\left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}\right) Y_{1t} =$$

$$= \left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}\right) \varepsilon_t^{(1)}$$

De la misma forma, multiplicando m. a m. la segunda ecuación por $1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(1)} B^{p_1}$, se obtiene que:

$$\left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(1)} B^{p_1}\right)\left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}\right) Y_{2t} =$$

$$= \left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(1)} B^{p_1}\right) \varepsilon_t^{(2)}$$

Luego, sumando m. a m. ambas igualdades se obtiene que:

$$\left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(1)} B^{p_1}\right)\left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}\right)\left(Y_{1t} + Y_{2t}\right) =$$

$$= \left(1 - \phi_1^{(2)} B - \phi_2^{(2)} B^2 - ... - \phi_{p_2}^{(2)} B^{p_2}\right) \varepsilon_t^{(1)} + \left(1 - \phi_1^{(1)} B - \phi_2^{(1)} B^2 - ... - \phi_{p_1}^{(2)} B^{p_2}\right) \varepsilon_t^{(2)}$$

Efectuando las operaciones que figuran en el primer miembro y teniendo en cuenta que, de acuerdo a lo demostrado en la Sec.2.11.2 para la suma de dos procesos MA incorrelacionados entre sí, el segundo miembro de esta ecuación admite una representación $MA\left(Max\left(p_1, p_2\right)\right)$, se puede escribir:

$$\left(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_{p_1 + p_2} B^{p_1 + p_2}\right) Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_{Max\left(p_1, p_2\right)}\right) \varepsilon_t$$

(donde: **i)** $\phi_j = \sum_{i=0}^{j} \phi_i^{(1)} \phi_j^{(2)} \left(j = 0,1,2,..., p_1 + p_2 ; \phi_0^{(1)} = \phi_0^{(2)} = 0\right)$; **ii)** $Y_t = Y_{1t} + Y_{2t}$ y **iii)** $\varepsilon_t = \varepsilon_t^{(1)} + \varepsilon_t^{(2)}$. Lo que demuestra que la suma de dos procesos estacionarios, incorrelacionados entre sí, con representaciones $AR\left(p_1\right)$ y $AR\left(p_2\right)$, respectivamente, genera un proceso que admite una representación $ARMA\left(\left(p_1 + p_2\right), Max\left(p_1, p_2\right)\right)$.

El **proceso compuesto** $\left\{Y_t\right\}$ resume las características estocásticas de $\left\{Y_{1t}\right\}$ e $\left\{Y_{2t}\right\}$ de modo que, de acuerdo a lo demostrado, resulta obvio que su representación sea más compleja que las de los procesos marginales. Existen dos formas distintas de agregación de procesos estocásticos: **i) en el dominio de las variables** (que es la que se trata en el texto) la cual surge como una suma ponderada de elementos con experiencias históricas diferentes y **ii) en el dominio del tiempo**, la cual surge cuando, por ejemplo, se utilizan procesos anuales o trimestrales para analizar el comportamiento de fenómenos dinámicos afectados por perturbaciones que se producen en períodos más cortos.

2.12.3.- La estimación de los coeficientes

De la solución del sistema de ecuaciones:

$$\begin{bmatrix} \hat{\gamma}_{q+1}(Y) \\ \hat{\gamma}_{q+2}(Y) \\ \cdots \\ \hat{\gamma}_{q+p}(Y) \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_{q}(Y) & \hat{\gamma}_{q+1}(Y) & \cdots & \hat{\gamma}_{q+p-1}(Y) \\ \hat{\gamma}_{q+1}(Y) & \hat{\gamma}_{q}(Y) & \cdots & \hat{\gamma}_{q-p}(Y) \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\gamma}_{q+p-1}(Y) & \hat{\gamma}_{q+p-2}(Y) & \cdots & \hat{\gamma}_{q}(Y) \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \cdots \\ \hat{\phi}_p \end{bmatrix}$$

se obtienen los estimadores preliminares $\hat{\phi}_j \ (j = 1,2,...,p)$.

Igual que en el caso de los $MA(q)$, los estimadores preliminares $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_q$ pueden ser obtenidos:

i) Utilizando un **proceso iterativo linealmente convergente**, en el siguiente orden $\sigma_\varepsilon^2, \theta_q, \theta_{q-1}, ..., \theta_1$, a partir de la iteración:

$$\sigma_\varepsilon^2 = \frac{\gamma_0(e)}{1 + \theta_1^2 + ... + \theta_q^2}$$

$$\theta_q = -\frac{\gamma_q(e)}{\sigma_\varepsilon^2}$$

$$\theta_{q-1} = -\left(\frac{\gamma_{q-1}(e)}{\sigma_\varepsilon^2} - \theta_1 \theta_q \right)$$

$$\theta_{q-2} = -\left(\frac{\gamma_{q-2}(e)}{\sigma_\varepsilon^2} - \theta_1 \theta_{q-1} - \theta_{q-1} \theta_q \right)$$

.....

$$\theta_1 = -\left(\frac{\gamma_1(e)}{\sigma_\varepsilon^2} - \theta_1 \theta_2 - \theta_2 \theta_3 - ... - \theta_{q-1} \theta_q \right)$$

Donde $\gamma_j(e) \ (j = 0,1,2...,q)$ denota la función de autocovarianzas correspondiente al proceso que se obtiene de aplicar a $\{Y_t\}$ una representación $AR(p)$ de la forma $\Phi(B)Y_t = e_t$ y puede ser obtenida a partir de la siguiente expresión:

$$\hat{\gamma}_j(e) = \left(\sum_{i=0}^{p} \hat{\phi}_i^2 \right) \hat{\gamma}_j(Y) + \sum_{i=1}^{p} \left(\hat{\phi}_0 \hat{\phi}_i + \hat{\phi}_1 \hat{\phi}_{i+1} + ... + \hat{\phi}_{p-i} \hat{\phi}_p \right) \left[\hat{\gamma}_{j+i}(Y) - \hat{\gamma}_{j-i}(Y) \right] \quad \left(j = 0,1,2,...,q; \hat{\phi}_0 = -1 \right)$$

ii) Utilizando un **proceso iterativo con convergencia cuadrática** en el cual, de la i -ésima iteración se obtiene que:

$$\tau^{(i+1)} = \tau^{(i)} - \left(T^{(i)} \right)^{-1} f^{(i)}$$

donde:

$$\left(\tau^{(i)}\right)^T = \left[\tau_0^{(i)} \quad \tau_1^{(i)} \quad \ldots \quad \tau_q^{(i)}\right]$$

$$\left(\tau_0^{(i)}\right)^T = \sigma_\varepsilon^2$$

$$\hat{\theta}_j^{(i)} = -\frac{\tau_j^{(i)}}{\tau_0^{(i)}} \qquad (j = 1,2,\ldots,q)$$

$$\left(f^{(i)}\right)^T = \left[f_0^{(i)} \quad f_1^{(i)} \quad \ldots \quad f_q^{(i)}\right]$$

$$f_j^{(i)} = \sum_{h=0}^{q-j} \tau_h^{(i)} \tau_{h+j}^{(i)} - \bar{\gamma}_j(e)$$

$$T^{(i)} = \begin{bmatrix} \tau_0^{(i)} & \tau_1^{(i)} & \ldots & \tau_{q-2}^{(i)} & \tau_{q-1}^{(i)} & \tau_q^{(i)} \\ \tau_1^{(i)} & \tau_2^{(i)} & \ldots & \tau_{q-1}^{(i)} & \tau_q^{(i)} & 0 \\ \tau_2^{(i)} & \tau_3^{(i)} & \ldots & \tau_q^{(i)} & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \tau_q^{(i)} & 0 & \ldots & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tau_0^{(i)} & \tau_1^{(i)} & \tau_2^{(i)} & \ldots & \tau_q^{(i)} \\ 0 & \tau_0^{(i)} & \tau_1^{(i)} & \ldots & \tau_{q-1}^{(i)} \\ 0 & 0 & \tau_0^{(i)} & \ldots & \tau_{q-2}^{(i)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \tau_0^{(i)} \end{bmatrix}$$

A partir de estas estimaciones, que pueden considerarse consistentes, es decir, tales que $\hat{\phi} \xrightarrow{p} \phi$; $\hat{\theta} \xrightarrow{p} \theta$ y $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$, se obtienen los estimadores máximo-verosímiles, que satisfacen las condiciones $\hat{\sigma}_\varepsilon^2 = \frac{1}{n} S\left(\hat{\phi},\hat{\theta}\right)$ (donde: **i)** $S\left(\hat{\phi},\hat{\theta}\right) = \sum_{j=1}^{n} \frac{\left(Y_j - \hat{Y}_j\right)^2}{\hat{\rho}_{j-1}(y)}$ y **ii)** $\hat{\phi}' = \left[\hat{\phi}_1 \quad \hat{\phi}_2 \quad \ldots \quad \hat{\phi}_p\right]$ y $\hat{\theta}' = \left[\hat{\theta}_1 \quad \hat{\theta}_2 \quad \ldots \quad \hat{\theta}_q\right]$ denotan los estimadores de ϕ y θ que minimizan la función de verosimilitud reducida:

$$\ell\left(\hat{\phi},\hat{\theta}\right) = \ln\left(\frac{S(\phi,\theta)}{n}\right) + \frac{1}{n}\sum_{j=1}^{n} \ln\left[\hat{\rho}_{j-1}(Y)\right].$$

Se demuestra[111] que, para procesos estacionarios e invertibles, el vector de los estimadores máximo-verosímiles, $\hat{\beta} = \left[\hat{\phi}' \quad \hat{\theta}'\right] = \left[\hat{\phi}_1 \quad \hat{\phi}_2 \quad \ldots \quad \hat{\phi}_p \quad \hat{\theta}_1 \quad \hat{\theta}_2 \quad \ldots \quad \hat{\theta}_q\right]$, es tal que:

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, V(\beta)\right)$$

(donde: **i)** $V(\beta) = \begin{bmatrix} E\left(U_t U_t'\right) & E\left(U_t V_t'\right) \\ E\left(V_t U_t'\right) & E\left(V_t V_t'\right) \end{bmatrix}$; **ii)** $U_t' = \left[U_t \quad U_{t+1} \quad \ldots \quad U_{t-p+1}\right]$; **iii)** $V_t' = \left[V_t \quad V_{t+1} \quad \ldots \quad V_{t-q+1}\right]$; **iv)** $\{U_t\}$ denota un proceso autorregresivo que admite una representación de la forma $\Phi(B)U_t = \varepsilon_t$ y **v)** $\{V_t\}$ denota un proceso de promedios móviles que admite una representación de la forma $\Theta(B)V_t = \varepsilon_t$.

Ejemplo 2.8:

Sea un proceso $\{Y_t\}$ de parámetro mensual formado por variables centradas, para cuyo análisis se cuenta con una serie cronológica para el período [ENE-2000;DIC-2004] y cuya trayectoria figura en el

---

111. Ver Brockwell, P.J.; Davis, R.A. (1987).

GRAFICO 2.16: Trayectoria del proceso $\{Y_t\}$

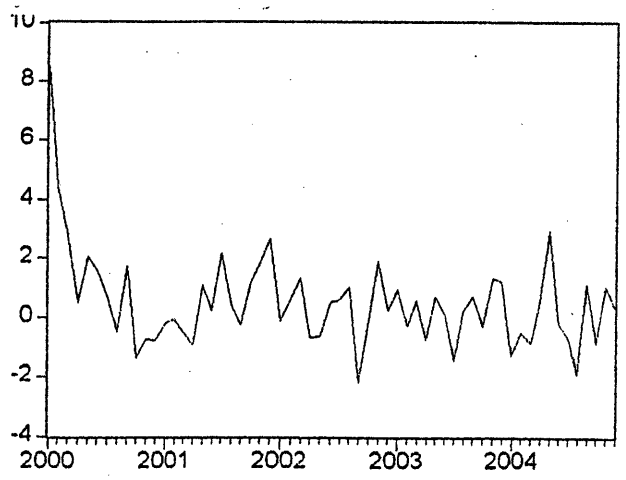


GRAFICO 2.16

El comportamiento del correlograma sugiere que $\{Y_t\}$ es estacionario de segundo orden y admite una representación $ARMA(p,q)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

De la mencionada serie cronológica se obtuvieron los siguientes resultados: $\hat{\gamma}_0(Y) = 1.588$; $\hat{\gamma}_1(Y) = 0.475$ y $\hat{\gamma}_2(Y) = 0.214$, a partir de los cuales se obtuvo el siguiente estimador de ϕ_1 :

$$\hat{\phi}_1 = \frac{\hat{\gamma}_2(Y)}{\hat{\gamma}_0(Y)} = 0.451 .$$

La función de autocovarianzas sobre el proceso residual $\{\hat{e}_t\}$, obtenida al aplicar el modelo $(1 - 0.451B)Y_t = \hat{e}_t$ es de la forma:

$$\gamma_0(\hat{e}) = \left(\hat{\phi}_0^2 + \hat{\phi}_1^2\right)\hat{\gamma}_0(Y) + \hat{\phi}_0\hat{\phi}_1\left[\hat{\gamma}_1(Y) + \hat{\gamma}_{-1}(Y)\right] = \left(\hat{\phi}_0^2 + \hat{\phi}_1^2\right)\hat{\gamma}_0(Y) + 2\hat{\phi}_0\hat{\phi}_1\hat{\gamma}_1(Y) = 1.483$$

$$\gamma_1(\hat{e}) = \left(\hat{\phi}_0^2 + \hat{\phi}_1^2\right)\hat{\gamma}_1(Y) + \hat{\phi}_0\hat{\phi}_1\left[\hat{\gamma}_2(Y) + \hat{\gamma}_0(Y)\right] = -0.241$$

Utilizando estos resultados, para calcular los estimadores preliminares de los coeficientes σ_ε^2 y θ_1 a partir del proceso iterativo lineal, se consideró el valor inicial $\hat{\theta}_1^{(0)} = 0$, obteniéndose de la primera iteración los siguientes resultados:

$$\left[\sigma_\varepsilon^{(1)}\right]^2 = \gamma_0(\hat{e}) = 1.483$$

$$\hat{\theta}_1^{(1)} = -\frac{\gamma_0(\hat{e})}{\left[\sigma_\varepsilon^{(1)}\right]^2} = -\frac{-0.241}{1.483} = 0.163$$

En las iteraciones sucesivas se obtuvieron los siguientes resultados:

Segunda iteración:

$$\left[\sigma_\varepsilon^{(2)}\right]^2 = \frac{\gamma_0(\hat{e})}{1+\left(\hat{\theta}_1^{(1)}\right)^2} = \frac{1.483}{1.026} = 1.448$$

$$\hat{\theta}_1^{(2)} = -\frac{\gamma_0(\hat{e})}{\left[\sigma_\varepsilon^{(2)}\right]^2} = -\frac{-0.241}{1.448} = 0.166$$

Tercera iteración:

$$\left[\sigma_\varepsilon^{(3)}\right]^2 = \frac{\gamma_0(\hat{e})}{1+\left(\theta_1^{(2)}\right)^2} = 1.443$$

$$\hat{\theta}_1^{(3)} = -\frac{\gamma_0(\hat{e})}{\left[\hat{\sigma}_\varepsilon^{(3)}\right]^2} = 0.167$$

Cuarta iteración:

$$\left[\sigma_\varepsilon^{(4)}\right]^2 = \frac{\gamma_0(\hat{e})}{1+\left(\theta_1^{(3)}\right)^2} = 1.443$$

$$\hat{\theta}_1^{(4)} = -\frac{\gamma_0(\hat{e})}{\left[\hat{\sigma}_\varepsilon^{(4)}\right]^2} = 0.167$$

Para la utilización del algoritmo de Newton-Raphson, se consideraron los siguientes **valores iniciales**:

$$\tau_0^{(0)} = \sqrt{\gamma_0(\hat{e})} = 1.218$$

$$\tau_1^{(0)} = 0$$

$$\hat{\theta}_1^{(0)} = 0$$

$$f_0^{(0)} = \left[\tau_0^{(0)}\right]^2 + \left[\tau_1^{(0)}\right]^2 - \gamma_0(\hat{e}) = 0$$

$$f_1^{(0)} = \tau_0^{(0)}\tau_1^{(0)} - \gamma_1(\hat{e}) = 0.241$$

$$\left[T^{(0)}\right]^{-1} f^{(0)} = \frac{1}{2\left[\left(\tau_0^{(0)}\right)^2 - \left(\tau_1^{(0)}\right)^2\right]}\begin{bmatrix} \tau_0^{(0)} f_0^{(0)} - 2\tau_1^{(0)} f_1^{(0)} \\ -\tau_1^{(0)} f_0^{(0)} + 2\tau_0^{(0)} f_1^{(0)} \end{bmatrix} = \begin{bmatrix} 0 \\ .198 \end{bmatrix}$$

Obteniéndose en la **primera iteración** los siguientes resultados:

$$\tau_0^{(1)} = \tau_0^{(0)} - 0 = 1.218$$

$$\tau_1^{(1)} = \tau_1^{(0)} - 0.198 = -0.198$$

$$\hat{\theta}_1^{(1)} = 0.163$$

$$f_0^{(1)} = \left[\tau_0^{(1)}\right]^2 + \left[\tau_1^{(1)}\right]^2 - \gamma_0(\hat{e}) = 0.039$$

$$f_1^{(1)} = \tau_0^{(1)}\tau_1^{(1)} - \gamma_1(\hat{e}) = 0$$

$$\left[T^{(1)}\right]^{-1} f^{(1)} = \frac{1}{2\left[\left(\tau_0^{(1)}\right)^2 - \left(\tau_1^{(1)}\right)^2\right]}\begin{bmatrix} \tau_0^{(1)} f_0^{(1)} - 2\tau_1^{(1)} f_1^{(1)} \\ -\tau_1^{(1)} f_0^{(1)} + 2\tau_0^{(1)} f_1^{(1)} \end{bmatrix} = \begin{bmatrix} 0.017 \\ 0.003 \end{bmatrix}$$

En la **segunda iteración** se obtuvieron los siguientes resultados:

$$\tau_0^{(2)} = \tau_0^{(1)} - 0.017 = 1.201$$

$$\tau_1^{(2)} = \tau_1^{(1)} - 0.003 = -0.201$$

$$\hat{\theta}_1^{(2)} = 0.167$$

Sea $\{Y_t\}$ un proceso estacionario e invertible que admite una representación $ARMA(p,q)$ de la forma $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$), cuya forma MA es:

$$Y_t = \left[\Phi(B)\right]^{-1}\Theta(B)\varepsilon_t = \Psi(B)\varepsilon_t .$$

Luego, teniendo en cuenta las expresiones obtenidas en la Sec.2.10.6, el coeficiente de determinación puede ser definido de la siguiente forma:

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_Y^2} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2} = \frac{\sum_{i=1}^{\infty} \psi_i^2}{\sum_{i=0}^{\infty} \psi_i^2} \qquad (\psi_0 = 1)$$

Como se ve, esta expresión no da lugar a interpretaciones tan interesantes como en el caso de las representaciones AR. Para su estimación parece recomendable utilizar la forma AR del $ARMA(p,q)$, considerando una representación AR de orden muy alto[112].

2.12.4.- La estimación de los órdenes p y q

Con un razonamiento similar al utilizado en la Sec. 2.10.5, al tratar las representaciones $AR(p)$, sean: **i)** un proceso $\{Y_t\}$ respecto del cual se cuenta con una serie cronológica $\{Y_1, Y_2, ..., Y_n\}$ que permite construir un modelo $ARMA(p,q)$ de la forma:

$$\hat{Y}_t = \hat{\phi}_{Y1}^{(MV)} Y_{t-1} + \hat{\phi}_{Y2}^{(MV)} Y_{t-2} + ... + \hat{\phi}_{Yp}^{(MV)} Y_{t-p} + \varepsilon_{Yt} - \hat{\theta}_{Y1}^{(MV)} \varepsilon_{Y,t-1} - ... - \hat{\theta}_{Yq}^{(MV)} \varepsilon_{Y,t-q}$$

(donde $\hat{\phi}_{Yj}^{(MV)}$ $(j = 1,2,...,p)$ y $\hat{\theta}_{Yj}^{(MV)}$ $(j = 1,2,...,q)$ denotan los estimadores máximo-verosímiles de los

---

[112]. Ver Nelson, C.R. (1976).

correspondientes coeficientes); **ii)** $\{X_t, t = 1,2,...,n\}$ una serie cronológica obtenida como una realización alternativa del mismo proceso, independiente de $\{Y_t\}$ y **iii)** $\hat{X}_{n+1}$ el estimador de la variable X_{n+1}, obtenido de la aplicación del modelo calculado a partir de la serie $\{Y_t\}$ $(1,2,...,n)$:

$$\hat{X}_{n+1} = \hat{\phi}_{Y1}^{(MV)} X_n + ... + \hat{\phi}_{Yp}^{(MV)} X_{n+1-p} + \hat{\varepsilon}_{Xn} - \hat{\theta}_{Y1}^{(MV)} \varepsilon_{X,n-1} - ... - \hat{\theta}_{Yq}^{(MV)} \varepsilon_{X,t-q}$$

Akaike, H. (1973) demostró que el error medio cuadrático de la predicción, para n suficientemente grande, puede ser aproximado de la siguiente forma: $E\left[\left(\hat{X}_{n+1} - X_{n+1}\right)^2\right] \approx \sigma_\varepsilon^2\left(1 + \frac{p+q}{n-p-q}\right)$ y que $\frac{\hat{\sigma}_\varepsilon^2}{1 + \frac{p+q}{n-p-q}}$ es un estimador insesgado de σ_ε^2, de modo que resulta el siguiente estimador insesgado del error medio cuadrático de la predicción, que define el criterio de información:

$$FPE(p,q) = \hat{\sigma}_{(p+q)\varepsilon}^2 \frac{n}{n - 2(p+q)}$$

De la misma forma, se obtienen los criterios alternativos: **i) AIC de Akaike**[113]:

$$AIC(p,q) = \ln\left[\hat{\sigma}_\varepsilon^2\left(ARMA(p,q)\right)\right] + \frac{2(p+q)}{n-p}$$

ii) SIC de Schwarz[114]:

$$SIC(p,q) = \ln\left[\hat{\sigma}_\varepsilon^2\left(ARMA(p,q)\right)\right] + \frac{(p+q)\ln(n-p)}{n-p}$$

iii) HQ de Hannan-Quinn[115]:

$$HQ(p,q) = \ln\left[\sigma_\varepsilon^2\left(ARMA(p,q)\right)\right] + \frac{2(p+q)\ln(n-p)}{n-p}$$

y sus correspondientes correcciones en las funciones de penalización y en los estimadores de σ_ε^2 (FPEu, AICc, AICu, HQc)[116].

Luego, dados dos conjuntos de órdenes posibles: $p = \{0,1,2,...,p_{Max}\}$ y $q = \{0,1,2,...,q_{Max}\}$ para $\Phi(B)$ y $\Theta(B)$, respectivamente, y un criterio de selección ($CSM(p,q)$), se elegirá como modelo más adecuado aquél para el cual se verifique que:

$$CSM(\hat{p},\hat{q}) = \min_{p,q}\left(CSM(p,q)\right)$$

---

113. Ver Akaike, H. (1974), Kitagawa, G. (1977), Ozaki, T. (1977).

114. Ver Schwarz, G. (1978), Akaike, H. (1978), Hannan, E.J.; Rissanen, J. (1982).

115. Ver Hannan, E.J.; Qinn, B.G. (1979).

116. Con respecto a estas generalizaciones, ver Howrey, E.P. (En Zarembka, P. (ed.)(1974)).

Cabe mencionar que, con respecto a las cualidades, defectos y ventajas comparativas de estos criterios, rigen las consideraciones realizadas en la Sec. 2.10.5, al tratar los procesos $AR(p)$.

Con respecto a los criterios de selección mencionados, puede suceder que existan modelos diferentes para los cuales los valores de los coeficientes de información sean "muy próximos" al mínimo y, por lo tanto, puedan ser considerados como alternativas válidas. En base a este concepto Poskitt, D.S.; Tremayne, A.R. (1987) introdujeron la definición de **cartera de modelos** como el conjunto de modelos para los cuales el coeficiente de comparación con relación a un modelo particular, $R = \exp\left\{\frac{n}{2}\left[AIC(p^*,q^*) - AIC(p,q)\right]\right\}$, sea menor que $\sqrt{10}$.

2.12.5.- El proceso ARMA(1,1)

Sea, en particular, una representación $ARMA(1,1)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Su expresión MA será:

$$Y_t = \frac{1-\theta_1 B}{1-\phi_1 B}\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad (\psi_0 = 1)$$

donde $\psi_0 = 1$ y $\psi_j = \phi_1^{j-1}(\phi_1 - \theta_1)$ $(j = 1,2,3,...)$.

La función de autocovarianzas será:

$$\gamma_0(Y) = \phi_1\gamma_1(Y) + \sigma_\varepsilon^2 - \theta_1\gamma(Y_t,\varepsilon_{t-1})$$
$$\gamma_1(Y) = \phi_1\gamma_0(Y) - \theta_1\sigma_\varepsilon^2$$
$$\gamma_j = \phi_1\gamma_{j-1}(Y) \qquad (j = 2,3,...)$$

donde $\gamma(Y_t,\varepsilon_{t-1}) = E\left[\left(\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\right)\varepsilon_{t-1}\right] = \psi_1\sigma_\varepsilon^2 = (\phi_1 - \theta_1)\sigma_\varepsilon^2$.

Del sistema de ecuaciones:

$$\begin{cases} \gamma_0(Y) - \phi_1\gamma_1(Y) = \left[1 - \theta_1(\phi_1 - \theta_1)\right]\sigma_\varepsilon^2 \\ -\phi_1\gamma_0(Y) + \gamma_1(Y) = -\theta_1\sigma_\varepsilon^2 \end{cases}$$

se obtienen las siguientes soluciones:

$$\gamma_0(Y) = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1-\phi_1^2}\sigma_\varepsilon^2$$

$$\gamma_1(Y) = \frac{(1-\phi_1\theta_1)(\phi_1 - \theta_1)}{1-\phi_1^2}\sigma_\varepsilon^2$$

Luego, la función de autocorrelaciones queda definida de la siguiente forma:

$$\rho_j(Y) = \begin{cases} \dfrac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} & (j=1) \\ \phi_1\rho_{j-1}(Y) & (j=2,3,\ldots) \end{cases}$$

(obsérvese que el decrecimiento exponencial de esta función se produce a partir del valor $\rho_1(Y)$ y no del valor $\rho_0(Y)=1$, como en la representación $AR(1)$). A partir de estos valores de $\rho_j(Y)$, se obtiene una función de autocorrelaciones parciales exponencialmente decreciente a partir del valor $\phi_{11}(Y)=\rho_1(Y)$.

2.13.- Las funciones de autocovarianzas y autocorrelaciones inversas

La **función de autocorrelaciones inversas** -introducida por Cleveland, W.S. (1972)- está formada por la asociación de las autocovarianzas con la expresión recíproca de la función de densidad espectral.

Sea $\{Y_t\}$ un proceso estacionario de segundo orden, con funciones generatrices de autocovarianzas y autocorrelaciones de la forma $\gamma_Y(w)=\sum_{j=-\infty}^{\infty}\gamma_j(Y)w^j$ y $\rho_Y(w)=\dfrac{\gamma_Y(w)}{\gamma_0(Y)}=\sum_{j=-\infty}^{\infty}\rho_j(Y)w^j$. Se define como **función generatriz de autocovarianzas inversas** ($\gamma_Y^{(I)}(w)$) a la inversa de la función $\gamma_Y(w)$ (el coeficiente que acompaña a la potencia w^j, en la función $\gamma_Y^{(I)}(w)$ define la **autocovarianza inversa de orden j** ($\gamma_j^{(I)}(Y)$))[117]. De modo que $\gamma_Y(w)\gamma_Y^{(I)}(w)=1$ (esta relación ortogonal entre la autocovarianzas y las autocovarianzas inversas, $\sum_{h=-\infty}^{\infty}\gamma_j(Y)\gamma_{j+h}^{(I)}(Y)=\delta_h$, donde δ_h denota la delta de Kronecker, permite considerar a la matriz de autocovarianzas inversas como una aproximación a la inversa de la matriz de autocovarianzas)[118].

Si $\{Y_t\}$:$WN-debil$, se puede escribir $\gamma_j^{(I)}(Y)=\dfrac{\delta_j}{\gamma_0(Y)}$ (donde δ_j denota la "delta de Kronecker" y $\gamma_0^{(I)}(Y)$ define la autocovarianza inversa de orden 0), en cualquier otro caso, de acuerdo con la desigualdad de Schwartz-Hölder, se verifica que:

$$1=\int_{-\pi}^{\pi}\sqrt{\omega_Y(\alpha)}\frac{1}{\sqrt{\omega_Y(\alpha)}}d\alpha \le \sqrt{\int_{-\pi}^{\pi}\omega_Y(\alpha)d\alpha\int_{-\pi}^{\pi}\frac{1}{\omega_Y(\alpha)}d\alpha}=\sqrt{\gamma_0(Y)\gamma_0^{(I)}(Y)}$$

---

[117]. De acuerdo a lo expresado en la Sec. 2.9.1, dado un proceso estocástico $\{Y_t\}$ estacionario de segundo orden, con función de densidad espectral estrictamente positiva de la forma $\omega_Y(\alpha)=\dfrac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j(Y)e^{-i\alpha j}$ $(-\pi<\alpha<\pi)$, la definición de Cleveland, W.S. (1972) de la función $\gamma_j^{(I)}(Y)$ está dada por $\gamma_j^{(I)}(Y)=\int_{-\pi/2}^{\pi/2}\dfrac{1}{\omega_Y(\alpha)}e^{i\alpha j}d\alpha$ $(j=0,\pm1,\pm2,\ldots)$. De modo que $\dfrac{1}{\omega_Y(\alpha)}=\sum_j\gamma_j^{(I)}(Y)e^{-i\alpha j}$.

[118]. Ver Shaman, P. (1975)(1976).

(donde $\omega_Y(\alpha)$ denota la función de densidad espectral del proceso $\{Y_t\}$). Entonces queda demostrado que $\gamma_0(Y)\gamma_0^{(I)}(Y) \geq 1$.

El **coeficiente de autocorrelación inversa de orden j** ($\rho_j^{(I)}(Y)$) queda definido, en consecuencia, de la siguiente forma:

$$\rho_j^{(I)}(Y) = \frac{\gamma_j^{(I)}(Y)}{\gamma_0^{(I)}(Y)} \qquad (j = 1,2,...)$$

y la **función generatriz de autocorrelaciones inversas,** está definida por:

$$\rho_Y^{(I)}(w) = \frac{\gamma_Y^{(I)}(w)}{\gamma_0^{(I)}(w)}$$

Se puede observar que esta definición de la función generatriz de autocorrelaciones inversas no necesita la incorporación de nuevos elementos respecto a la función de autocorrelaciones lo cual, indudablemente, constituye una ventaja de esta función con relación a la función de autocorrelaciones parciales en la identificación de modelos y en la estimación de sus coeficientes.

Sea, por ejemplo, un proceso $\{Y_t\}$, estacionario de segundo orden e invertible, que admite una representación $ARMA(p,q)$ de la forma $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ y cuya representación MA es de la forma:

$$Y_t = [\Phi(B)]^{-1}\Theta(B)\varepsilon_t = \Psi(B)\varepsilon_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + ...$$

De acuerdo a lo demostrado en la página precedente, la función generatriz de autocovarianzas inversas está definida por:

$$\gamma_Y^{(I)}(w) = \sum_{j=-\infty}^{\infty} \gamma_j^{(I)}(Y)w^j = \frac{1}{\sigma_\varepsilon^2}\Psi(w)\Psi(w^{-1}) = \frac{1}{\sigma_\varepsilon^2}\frac{\Phi(w)\Phi(w^{-1})}{\Theta(w)\Theta(w^{-1})}$$

A partir de esta expresión se demuestra fácilmente que la función de autocorrelaciones inversas de un proceso $ARMA(p,q)$ es igual a la función de autocorrelaciones de un proceso $\{Z_t\}$ que admite una representación $ARMA(q,p)$ de la forma $\Theta(B)Z_t = \Phi(B)\varepsilon_t$. La **representación MA del proceso inverso** será de la forma:

$$Z_t = \frac{1}{\sigma_\varepsilon^2}\frac{\Phi(B)\Phi(B^{-1})}{\Theta(B)\Theta(B^{-1})}Y_t = \frac{1}{\sigma_\varepsilon^2}\frac{\Phi(B^{-1})}{\Theta(B^{-1})}\frac{\Phi(B)}{\Theta(B)}Y_t = \frac{1}{\sigma_\varepsilon^2}\frac{\Phi(F)}{\Theta(F)}\varepsilon_t$$

(donde $B^{-1} = F$ denota el operador "forward"). Esto implica que el proceso $\{Z_t\}$ puede ser expresado como una función lineal de los "shocks" aleatorios, ε_t presentes y futuros:

$$Z_t = \frac{1}{\sigma_\varepsilon^2}(\varepsilon_t + \lambda_1\varepsilon_{t+1} + \lambda_2\varepsilon_{t+1} + ...)$$

Por ejemplo, si $\{Y_t\}$ es un proceso estacionario que admite una representación $AR(p)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

de acuerdo a lo demostrado en la Sec.2.10.2, su función de autocorrelaciones está definida como una combinación de funciones exponenciales y funciones sinusoidales amortiguadas, y su función de autocorrelaciones inversas será la correspondiente a un proceso $MA(p)$ de la forma:

$$Y_t = \varepsilon_t - \phi_1\varepsilon_{t-1} - \phi_2\varepsilon_{t-2} - \ldots - \phi_p\varepsilon_{t-p}$$

es decir, asume valores significativos hasta el orden p, y después se anula:

$$\rho_j^{(I)}(Y) = \begin{cases} 1 & (j = 0) \\ \dfrac{-\phi_j + \sum_{i=j}^{p} \phi_{i-j}\phi_i}{1 + \sum_{i=1}^{p} \phi_i^2} & (j = 1,2,\ldots,p) \\ 0 & (j = p+1, p+2, \ldots) \end{cases}$$

En particular, para un proceso $AR(1)$ de la forma $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ $(|\phi_1| < 1)$, será:

$$\rho_j^{(I)}(Y) = \begin{cases} 1 & (j = 0) \\ \dfrac{-\phi_1}{1+\phi_1^2} & (j = 1) \\ 0 & (j = 2,3,\ldots) \end{cases}$$

y, de acuerdo con las expresiones correspondientes a las funciones generatrices de autocovarianzas y autocorrelaciones dadas en páginas anteriores, se puede concluir que:

$$\gamma_Y^{(I)}(w) = \frac{\left(1 - \phi_1 w\right)\left(1 - \dfrac{\phi_1}{w}\right)}{\sigma_\varepsilon^2}$$

de modo que:

$$\gamma_0^{(I)}(Y) = \frac{1+\phi_1^2}{\sigma_\varepsilon^2}$$

y, teniendo en cuenta que $\gamma_0(Y) = \dfrac{\sigma_\varepsilon^2}{1-\phi_1^2}$, se concluye fácilmente que $\gamma_0^{(I)}(Y) \neq \dfrac{1}{\gamma_0(Y)}$.

La representación MA del proceso inverso, $\{Z_t\}$, es de la forma:

$$Z_t = \sum_j \gamma_j^{(I)}(Y) Y_{t-j} = -\phi_1 Y_{t-1} + \left(1+\phi_1^2\right) Y_t - \phi_1 Y_{t+1} = \Phi(F)\varepsilon_t = \varepsilon_t - \phi_1\varepsilon_{t+1}$$

Si $\{Y_t\}$ es un proceso invertible que admite una representación $MA(q)$, la función de autocorrelaciones inversas tiene la misma forma que la función de autocorrelaciones correspondiente al proceso inverso, $AR(q)$ (es decir, está definida como la combinación de funciones exponenciales y funciones

sinusoidales amortiguadas).

En particular, para un proceso $MA(1)$ de la forma:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} = (1 - \theta_1 B)\varepsilon_t$$

las autocovarianzas inversas son de la forma $\gamma_j^{(I)}(Y) = \dfrac{\theta_1^{|j|}}{1-\theta_1^2}\ (j = 1,2,...)$ y, por lo tanto, el proceso inverso, $\{Z_t\}$, puede ser representado como:

$$Z_t = \sum_{j=-\infty}^{\infty} \frac{\theta_1^{|j|}}{1-\theta_1^2} Y_{t-j} = \Theta^{-1}(F)\varepsilon_t = \sum_{j=0}^{\infty} \theta_1^j \varepsilon_{t+j}$$

De acuerdo con Cleveland, W.S. (1972), existen dos formas posibles de estimación de la función de autocorrelaciones inversas: **i)** una que consiste en estimar el espectro de la serie cronológica mediante el suavizado del periodograma, calcular la forma recíproca de este estimador y construir la transformación de Fourier apropiada y **ii)** otra que consiste en aproximar el proceso mediante una representación de orden suficientemente grande[119], obtener los estimadores de Yule-Walker de sus coeficientes y sustituir estos estimadores en la expresión[120]:

$$\rho_j^{(I)}(Y) = \frac{-\phi_j + \sum_{i=j}^{p} \phi_{i-j}\phi_i}{1 + \sum_{i=1}^{p} \phi_i^2}$$

Una aplicación de las autocovarianzas inversas -en particular, de la **varianza inversa**, $\gamma_0^{(I)}(Y)$ - está relacionada con la caracterización de la estructura de un proceso estacionario. De acuerdo con Battaglia, F. (1983), un **índice de determinismo lineal**, para un proceso $\{Y_t\}$, estacionario, puede ser definido como:

$$I_L(Y) = 1 - \frac{1}{\gamma_0(Y)\gamma_0^{(I)}(Y)}$$

En particular, de acuerdo a lo mencionado en páginas anteriores, dado un proceso $\{Y_t\}$:*WN - debil*, se verifica que $\gamma_0(Y)\gamma_0^{(I)}(Y) = 1$, por lo que el índice resulta igual a cero, lo cual significa que la variabilidad

---

[119]. De acuerdo con Hipel, K.W.; McLeod, A.I.; Lennox, W.C. (1977), deben considerarse órdenes de autorregresividad $10 \le p \le 40$. Debe tenerse en cuenta que, en este caso, la utilización de criterios de información no es apropiada, dado que la intención no es hallar el orden p óptimo, sino determinar un valor de p suficientemente grande para asegurar la convergencia a cero de la función de autocorrelaciones inversas. Por otra parte, debe tenerse en cuenta que un orden p demasiado grande conduce a estimadores con varianza muy alta.

[120]. Existe un tercer método de estimación asintóticamente eficiente, debido a Hannan, E.J. (1970) y McClave, J.T. (1978), pero tiene el inconveniente de presentar complicaciones formales excesivas con respecto a los mencionados en el texto.

de $\{Y_t\}$ no puede ser reducida por la aplicación de un filtro lineal y, si el proceso satisface, con probabilidad igual a uno, una restricción determinística lineal, entonces el índice es igual a la unidad.

Supóngase que el proceso $\{Y_t\}$ admita una representación $ARMA(p,q)$ de la forma $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ con varianza:

$$\sigma^2(Y) = \gamma_0(Y) = \int_{-1/2}^{1/2} \left| \frac{\Theta\left(e^{i2\pi\lambda}\right)}{\Phi\left(e^{i2\pi\lambda}\right)} \right|^2 d\lambda$$

Entonces, será:

$$I_L(Y) = 1 - \frac{1}{\int_{-1/2}^{1/2} \left| \frac{\Theta\left(e^{i2\pi\lambda}\right)}{\Phi\left(e^{i2\pi\lambda}\right)} \right|^2 d\lambda \int_{-1/2}^{1/2} \left| \frac{\Phi\left(e^{i2\pi\vartheta}\right)}{\Theta\left(e^{i2\pi\vartheta}\right)} \right|^2 d\vartheta} = 1 - \left(1 - R_Y^2\right)\left(1 - R_Z^2\right)$$

Donde $\{Z_t\}$ denota el proceso inverso, y R_Y^2 y R_Z^2 denotan los correspondientes coeficientes de determinación.

Dado un proceso $\{Y_t\}$ estacionario e invertible, su comportamiento puede ser desagregado de la siguiente forma: **i)** un término que representa la parte del proceso que puede ser "determinada" a partir de su historia pasada y futura:

$$D_t = Y_t - \frac{Z_t}{\gamma_0^{(I)}(Y)} = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t - \frac{1}{\gamma_0^{(I)}(Y)\sigma_\varepsilon^2}\frac{\Phi(F)}{\Theta(F)}\varepsilon_t$$

(donde: **i)** $\{\varepsilon_t\}$:$WN - debil$ y **ii)** el coeficiente que acompaña a ε_t es $1 - \left[\gamma_0^{(I)}(Y)\sigma_\varepsilon^2\right]^{-1}$), y **ii)** una parte que representa la influencia de las perturbaciones de $\{Y_t\}$ con respecto a $\{D_t\}$, $\eta_t = \frac{Z_t}{\gamma_0^{(I)}(Y)}$ (donde $E(D_t\eta_t) = 0$). Luego, se puede escribir:

$$\sigma^2(Y_t) = \sigma^2(D_t) + \sigma^2(\eta_t) = \left[\gamma_0(Y) - \frac{1}{\gamma_0^{(I)}(Y)}\right] + \frac{1}{\gamma_0^{(I)}(Y)}$$

Dado que $\{\varepsilon_t\}$:WN, será:

$$\sigma^2(D_t) = I_L(Y)\gamma_0(Y) = \gamma_0(Y) - \frac{1}{\gamma_0^{(I)}(Y)} =$$

$$= \sigma_\varepsilon^2\left[\int_{-1/2}^{1/2} \left| \frac{\Theta\left(e^{i2\pi\lambda}\right)}{\Phi\left(e^{i2\pi\lambda}\right)} \right|^2 d\lambda - 1\right] + \sigma_\varepsilon^2\left[1 - \frac{1}{\sigma_\varepsilon^2\gamma_0^{(I)}(Y)}\right]^2 + \sigma_\varepsilon^2\left[\frac{1}{\sigma_\varepsilon^2\gamma_0^{(I)}(Y)}\right]^2\left[\int_{-1/2}^{1/2} \left| \frac{\Phi\left(e^{i2\pi\lambda}\right)}{\Theta\left(e^{i2\pi\lambda}\right)} \right|^2 d\lambda - 1\right] =$$

$$= \left[\gamma_0(Y) - \sigma_\varepsilon^2\right] + \left[\sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2\left(\gamma_0^{(I)}(Y)\right)^2} - \frac{2}{\gamma_0^{(I)}(Y)}\right] + \left[\frac{1}{\gamma_0^{(I)}(Y)} - \frac{1}{\sigma_\varepsilon^2\left(\gamma_0^{(I)}(Y)\right)^2}\right]$$

(donde: **i)** el primer sumando , $\gamma_0(Y) - \sigma_\varepsilon^2$, representa la **contribución de** $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ ($\sigma_\varepsilon^2(pasado)$); **ii)** el segundo sumando $\sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2\left(\gamma_0^{(I)}(Y)\right)^2} - \frac{2}{\gamma_0^{(I)}(Y)}$ representa la **contribución de** ε_t ($\sigma_\varepsilon^2(presente)$) y **iii)** el tercero, $\frac{1}{\gamma_0^{(I)}(Y)} - \frac{1}{\sigma_\varepsilon^2\left(\gamma_0^{(I)}(Y)\right)^2}$, representa la **contribución de** $\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots$ ($\sigma_\varepsilon^2(futuro)$)). De modo que el índice de determinismo lineal puede ser desagregado de la siguuiente forma:

$$I_L(Y) = \frac{\sigma^2(D_t)}{\gamma_0(Y)} = \frac{\sigma_\varepsilon^2(pasado)}{\gamma_0(Y)} + \frac{\sigma_\varepsilon^2(presente) + \sigma_\varepsilon^2(futuro)}{\gamma_0(Y)} =$$

$$= \frac{\gamma_0(Y) - \sigma_\varepsilon^2}{\gamma_0(Y)} + \frac{\sigma_\varepsilon^2 - \frac{1}{\gamma_0^{(I)}(Y)}}{\gamma_0(Y)} =$$

$$= R^2(pasado) + R^2(presente + futuro)$$

(donde $R^2(pasado)$ denota el **coeficiente de determinación de** $\{Y_t\}$).

En particular, para un proceso $\{Y_t\}$, estacionario de segundo orden, que admite una representación $AR(1)$ de la forma $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$), será:

$$I_L(Y) = R^2(pasado) + R^2(presente + futuro) = \phi_1^2 + \phi_1^2 \frac{1 - \phi_1^2}{1 + \phi_1^2}$$

Para un proceso $\{Y_t\}: AR(2)$ estacionario de segundo orden y tal que $m(Y) = 0$, será:

$$I_L(Y) = R^2(pasado) + R^2(presente + futuro) =$$

$$= \left(\phi_2^2 + \phi_1^2 \frac{1 + \phi_2}{1 - \phi_2}\right) + \left\{\phi_2^2\left(\frac{1 - \phi_1^2\phi_2^2}{1 + \phi_1^2 - \phi_2^2}\right) + \frac{\phi_1^2}{1 - \phi_2}\left[1 + \phi_2 - 2\left(\phi_2 + \frac{\phi_1^2 + \phi_2^2}{1 + \phi_1^2 + \phi_2^2}\right)\right]\right\}$$

Se demuestra entonces que, para los procesos $AR(1)$ y $AR(2)$ estacionarios:

$$R^2(pasado) > R^2(presente + futuro) .$$

Para un proceso $\{Y_t\}: MA(1)$ invertible, que admite una representación de la forma $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ (donde $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$), se obtiene que:

$$I_L(Y) = R^2(pasado) + R^2(presente + futuro) = \frac{\theta_1^2}{1 + \theta_1^2} + \frac{\theta_1^2}{1 + \theta_1^2}$$

De lo que se concluye que $R^2(pasado) = R^2(presente + futuro)$.

Para un proceso $\{Y_t\}: MA(2)$ invertible y tal que $m(Y) = 0$, será:

$$I_L(Y) = R^2(pasado) + R^2(presente + futuro) = \frac{\theta_1^2 + \theta_2^2}{1 + \theta_1^2 + \theta_2^2} + \frac{\frac{\theta_1^2(1+\theta_2)}{1-\theta_2} + \theta_2^2}{1 + \theta_1^2 + \theta_2^2}$$

A partir de esta expresión, se demuestra fácilmente que, en los procesos $MA(2)$ el signo de la diferencia $R^2(presente + futuro) - R^2(pasado)$ es igual al signo de θ_2.

De la misma forma, para un proceso $\{Y_t\}: ARMA(1,1)$ estacionario e invertible, que admite una representación de la forma $Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$, se verifica que:

$$I_L(Y) = R^2(pasado) + R^2(presente + futuro) =$$

$$= \frac{(\phi_1 - \theta_1)^2}{1 - \phi_1^2 + (\phi_1 - \theta_1)^2} + \frac{(1 - \phi_1^2)(\phi_1 - \theta_1)^2}{\left[1 - \phi_1^2 + (\phi_1 - \theta_1)^2\right]\left[1 - \theta_1^2 + (\phi_1 - \theta_1)^2\right]}$$

Lo que demuestra que, si $0 < \phi_1 < \theta_1$ ó $\theta_1 < \phi_1 < 0$, entonces, se verifica que $R^2(pasado) < R^2(presente + + futuro)$.

De todo lo anterior se puede concluir que: **i)** Los procesos AR dependen más de las "innovaciones pasadas" que de las futuras (la variabilidad de $\{Y_t\}$ está más afectada por el pasado que por el presente o el futuro), y, viceversa, los procesos MA y ARMA dependen más de las innovaciones futuras que de las pasadas, según el valor de sus coeficientes[121]. **ii)** En los límites de las regiones de estacionariedad e invertibilidad, el coeficiente I_L es igual a la unidad.

2.14.- Análisis de la adecuacidad de un modelo

2.14.1.- Acerca de la existencia de autocorrelación en el proceso $\{\varepsilon_t\}$

Teniendo en cuenta que, en todos los casos, se supuso que $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$, uno de los aspectos a analizar para decidir acerca de la adecuacidad de un modelo consiste en estudiar la existencia de autocorrelaciones significativas en las perturbaciones, a partir de la información que proporciona la serie formada por los residuos del comportamiento de $\{Y_t\}$ no explicado por dicho modelo.

Sea $\{Y_t\}$ un proceso estacionario de segundo orden que admite una representación $ARMA(p,q)$ de la forma $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$) y sean $\hat{\phi}_j^{(MV)}$ $(i = 1,2,...,p)$ y $\hat{\theta}_j^{(MV)}$ $(j = 1,2,...,q)$ los estimadores máximo-verosímiles de los coeficientes ϕ_j y θ_j analizados en la Sec.2.12.3. Entonces, la serie de los residuos $\{\hat{\varepsilon}_t\}$ puede ser obtenida de la siguiente forma:

$$\hat{\varepsilon}_t = \hat{\Phi}(B)\left[\hat{\Theta}(B)\right]^{-1} Y_t = Y_t - \left(\hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} + ... + \hat{\phi}_p Y_{t-p}\right) + \left(\hat{\theta}_1 \hat{\varepsilon}_{t-1} + \hat{\theta}_2 \hat{\varepsilon}_{t-2} + ... + \hat{\theta}_q \hat{\varepsilon}_{t-q}\right)$$

---

121. "*...en el pasado, detrás de nosotros, se encuentra no sólo lo que ha sucedido realmente, sino lo que podría haber sido (...) Y en el futuro se encuentra no sólo lo que será, sino todo lo que puede ser*" (Ouspensky, P.D. (1982)).

$\left(Y_0 = 0; \hat{\varepsilon}_0 = 0; t = 1,2,\ldots,n\right)$ [121].

Para que el modelo pueda ser considerado adecuado se debe verificar que $\hat{\varepsilon}_t = \varepsilon_t + o\left(\frac{1}{\sqrt{n}}\right)$, de modo de poder asegurar que, a medida que n crece, los residuos se aproximan a un WN.

Si la verdadera representación $ARMA(p,q)$ fuera conocida[122], entonces se demuestra que, bajo el supuesto de que la hipótesis nula $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$ es verdadera, los estimadores de $\rho_j(\varepsilon)$ se distribuyen aproximadamente de acuerdo con una función Normal con varianza $\frac{1}{n}$. Por el contrario, si, como ocurre habitualmente en las aplicaciones, no se conoce el verdadero valor de los coeficientes, sino sus estimadores (y, por lo tanto, no se conocen los estimadores de $\rho_j(\varepsilon)$ sino los estimadores de $\rho_j(\hat{\varepsilon})$), es incorrecto testear la significatividad individual de los coeficientes de autocorrelación aproximando la varianza del estimador por $\frac{1}{n}$ [123] ya que, para órdenes próximos a cero, los estimadores de $\rho_j(\hat{\varepsilon})$ son altamente correlacionados y su varianza resulta apreciablemente menor que $\frac{1}{n}$ (efectos que desaparecen rápidamente a medida que j crece)[124]. Lo cual implica que este test genera, para los órdenes próximos a cero, un sesgo hacia la aceptación de la hipótesis nula.

Como se mencionó en la Sec. 2.8, una forma alternativa de análisis de la existencia de autocorrelación en las perturbaciones se basa en el test "portmanteau" de Ljung-Box, que considera la primeras J autocorrelaciones en conjunto, donde el valor de J es tal que los coeficientes ψ_j (para $j > J$) de la representación MA:

$$Y_t = \Psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad (\psi_0 = 0)$$

pueden ser considerados nulos. Debe tenerse en cuenta que, bajo el supuesto que la hipótesis nula que $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$ es verdadera el estadístico de Ljung-Box será tal que (algunos autores denominan al coeficiente $Q(j)$ como estadístico de McLeod-Li)[125]:

$$Q^{(LB)}(J) = n(n+2)\sum_{i=1}^{J} \frac{\hat{\rho}_i^2(Y)}{n-i} \xrightarrow{d} \chi^2_{(J-p-q)}$$

---

121. Alternativamente a este método condicional que considera valores iniciales nulos para Y_t y $\hat{\varepsilon}_t$, se podrían haber calculado sus valores exactos obtenidos a partir de la utilización del modelo.

122. Es decir, si se conocieran los verdaderos valores de $p,q,\phi_j\ (j = 1,2,\ldots,p)$ y $\theta_j\ (j = 1,2,\ldots,q)$.

123. Ver Durbin, J. (1979).

124. Ver Box, G.E.P.; Pierce, D.A. (1970).

125. En particular, Brockwell, P.J.; Davis, R.A. (1996) y Gouriéroux, C. (1997) recomiendan la utilización de este estadístico para contrastar la hipótesis nula de Normalidad de las perturbaciones ε_t.

(si el modelo posee término independiente, la distribución del estadístico de Ljung-Box será $\chi^2_{(J-p-q-1)}$).

2.14.2.- Acerca de la no-Normalidad del proceso $\{\varepsilon_t\}$

Sea $\left\{\hat{\varepsilon}_t\right\}(t = 1,2,...,n)$ la serie cronológica formada por los residuos obtenidos de aplicar un modelo $ARMA(p,q)$. Suponiendo que estos residuos puedan ser considerados como realizaciones de un proceso $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$, el análisis de la Normalidad consiste en "medir" la distancia entre la curva definida por la función de distribución Normal y la distribución (acumulada) de los valores de los residuos $\left\{\hat{\varepsilon}_t\right\}$ [127]. Esta distancia puede ser medida: **i)** en forma **vertical** en términos de los percentiles, mediante el **diagrama P-P** o **ii)** en forma **horizontal** en términos de los quantiles, mediante el **diagrama Q-Q**[128].

Sea $\left\{\varepsilon_{(t)}\right\}(t = 1,2,...,n)$ la serie cronológica formada por una sucesión de variables ε_t ordenadas de menor a mayor. Cualquiera sea la distribución de las variables $\varepsilon_{(t)}$ su función de distribución toma la forma $F_{\varepsilon_{(t)}}(\varepsilon) = p\left(\varepsilon_{(t)} \le \varepsilon\right) = \sum_{j=t}^{n}\binom{n}{j}\left[F(\varepsilon)\right]^j\left[1 - F(\varepsilon)\right]^{n-j}$. De modo que su correspondiente función de densidad será de la forma[129]:

$$f_{\varepsilon_{(t)}}(\varepsilon) = \frac{n!}{(t-1)!(n-t)!}\left[F(\varepsilon)\right]^{t-1}\left[1 - F(\varepsilon)\right]^{n-t}$$

Supóngase, en particular, que ε_t:$U(0;1)(t = 1,2,...,n)$. Entonces la variable $\varepsilon_{(t)}$ se distribuye de acuerdo con una función de probabilidades Beta con coeficientes $(t, n - t + 1)$ de la forma:

$$f_{\varepsilon_{(t)}}(\varepsilon) = \frac{n!}{(t-1)!(n-t)!}\varepsilon^{t-1}(1-\varepsilon)^{n-t}f(\varepsilon) \qquad (0 \le \varepsilon \le 1)$$

a partir de la cual se obtiene que:

$$E\left(\varepsilon_{(t)}\right) = \frac{t}{n+1} \qquad (t = 1,2,...,n)$$

$$\sigma^2\left(\varepsilon_{(t)}\right) = \frac{t(n-t+1)}{(n+1)^2(n+2)} \qquad (t = 1,2,...,n)$$

$$\gamma\left(\varepsilon_{(t)},\varepsilon_{(s)}\right) = \frac{t(n-s+1)}{(n+1)^2(n+2)} \qquad (s,t = 1,2...,n; s \ne t)$$

Luego, se puede concluir que el denominado diagrama P-P, en el cual se comparan los valores

---

127. Este análisis se basa en la comparación entre condiciones teóricas como independencia, incorrelación, distribución idéntica y Normalidad de las variables ε_t , con las características de las realizaciones que se suponen generadas de acuerdo con estas condiciones.

128. Debe recordarse que los quantiles y los percentiles constituyen una familia de medidas de posición de una variable aleatoria (la denominación se debe a Galton, F. (1885)).

129. Ver Wilk, B.; Gnadadesikan, R. (1968).

esperados $\frac{t}{n+1}$ con los valores $\varepsilon_{(t)}\,(t = 1,2,...,n)$, correspondientes a variables con distribución uniforme, definirá una recta que pasa por el origen de coordenadas. Dada, entonces, una serie cronológica $\left\{\hat{\varepsilon}_{(t)}\right\}(t = 1,2,...,n)$ formada por los residuos obtenidos de aplicar un modelo $ARMA(p,q)$, si el diagrama construido de la comparación de los valores $\frac{t}{n+1}$ y los valores $\hat{\varepsilon}_{(t)}\,(t = 1,2,...,n)$ presentara una forma aproximadamente lineal, se podría concluir que las variables ε_t están uniformemente distribuidas.

De acuerdo con el **lema de la transformación de la probabilidad integral** según el cual, dada una variable continua $\varepsilon_{(t)}$ con función de distribución de probabilidades $F_{\varepsilon_{(t)}}(\varepsilon)$, la variable aleatoria definida por $\varepsilon^*_{(t)} = F_{\varepsilon_{(t)}}(\varepsilon)$ tiene una distribución del tipo $U(0;1)$ [130], el diagrama P-P puede ser generalizado a otras distribuciones continuas. La comparación será, entonces, entre los valores $\frac{t}{n+1}$ y los valores $F_{\varepsilon_{(t)}}(\varepsilon)\,(t = 1,2,...,n)$.

En particular, como se mencionó al comienzo de esta sección, en el análisis de los residuos interesa considerar como distribución de referencia a la función Normal. En este caso el diagrama P-P estará dado por la comparación entre los valores $\frac{t}{n+1}\,(t = 1,2,...,n)$ y los valores asumidos por la función de distribución:

$$F^{(N)}_{\varepsilon_{(t)}}(\varepsilon) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\varepsilon} e^{-u^2/2}du \qquad (t = 1,2,...,n)$$

Las consideraciones acerca de la forma del diagrama P-P analizadas hasta aquí se basan en la transformación de las variables $\varepsilon_{(t)}$ en variables $\varepsilon^*_{(t)}:U(0,1)$ y en la consiguiente presunta distribución lineal de los puntos. De la misma forma, se puede realizar el análisis en sentido inverso, basado en la denominada **transformación quantil** la cual permite transformar una variable $\varepsilon^*_{(t)}:U(0,1)$ en cualquier otra variable $\varepsilon_{(t)} = F^{-1}_{\varepsilon^*_{(t)}}(\varepsilon)\,(0 < \varepsilon < 1)$, donde $\varepsilon_{(t)}$ denota una variable estandarizada y, suponiendo que exista, $F^{-1}_{\varepsilon^*_{(t)}}(\varepsilon) = \inf\left(\varepsilon_{(t)}: F_{\varepsilon_{(t)}}(\varepsilon) \geq \varepsilon^*_{(t)}; 0 \leq \varepsilon^*_{(t)} \leq 1\right)$. Como resultado el diagrama P-P (conocido como **diagrama P-P estandarizado**) generará una curva, producto de la comparación de los valores $\frac{t}{n+1}\,(t = 1,2,...,n)$ y de los valores de la correspondiente inversa de la función de distribución de referencia, $F^{-1}_{\varepsilon^*_{(t)}}(\varepsilon)$.

A continuación se consideran las expresiones analíticas de las inversas de las funciones de distribución para algunas variables que se presentan frecuentemente en la econometría del mercado financiero.

Supóngase que: **i)** $\varepsilon^*_{(t)}$ se distribuya de acuerdo con una **función uniforme** con función de

---

[130]. La función generatriz de momentos, $M_{\varepsilon^*_{(t)}}(w) = E\left(e^{w\varepsilon^*_{(t)}}\right) = E\left[\exp\left(wF_{\varepsilon_{(t)}}(\varepsilon)\right)\right] = \int_{-\infty}^{\infty}\exp\left[wF_{\varepsilon_{(t)}}(\varepsilon)\right]f(\varepsilon)d\varepsilon =$

$= \frac{1}{w}\exp\left[wF_{\varepsilon_{(t)}}\right]\Big|_{-\infty}^{\infty} = \frac{e^w - 1}{w}$, es la correspondiente a una variable $U(0;1)$.

distribución de la forma $F_{\varepsilon_{(t)}^*} = \frac{1}{\beta - \alpha}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \alpha + (\beta - \alpha)\varepsilon$; **ii)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función de Caùchy** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \frac{\beta}{\pi\left[\beta^2 + (\varepsilon - \alpha)^2\right]}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \alpha + \beta\, tg\left[\pi\left(\varepsilon - \frac{1}{2}\right)\right]$; **iii)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función exponencial** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \frac{1}{\beta}\exp\left(-\frac{\varepsilon}{\beta}\right)$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = -\beta\ln(1-\varepsilon)$; **iv)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función logística** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \frac{\exp\left(-\frac{\varepsilon - \alpha}{\beta}\right)}{\beta\left[1 + \exp\left(-\frac{\varepsilon - \alpha}{\beta}\right)\right]^2}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \alpha + \beta\ln\left(\frac{\varepsilon}{1-\varepsilon}\right)$; **v)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función de Pareto** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \alpha\beta^\alpha\varepsilon^{-(\alpha+1)}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \frac{\beta}{(1-\varepsilon)^{1/\alpha}}$; **vi)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función de Weibull** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \frac{\alpha}{\beta^\alpha}\varepsilon^{\alpha-1}\exp\left[-\left(\frac{\varepsilon}{\beta}\right)^\alpha\right]$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \beta\left[-\ln(1-\varepsilon)^{1/\alpha}\right]$; **vii)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función Beta** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \alpha\varepsilon^{\alpha-1}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = \varepsilon^{1/\alpha}$; **viii)** $\varepsilon_{(t)}^*$ se distribuya de acuerdo con una **función Beta(1,β)** con función de distribución de la forma $F_{\varepsilon_{(t)}^*}(\varepsilon) = \beta(1-\varepsilon)^{\beta-1}$, se verificará, entonces, que $F_{\varepsilon_{(t)}^*}^{-1}(\varepsilon) = 1 - (1-\varepsilon)^{1/\beta}$.

Alternativamente, para obtener un medida horizontal de la distancia entre dos distribuciones es conveniente tener en cuenta que: **i)** el valor $F_{\varepsilon_{(t)}^*}^{-1}\left(\frac{t}{n+1}\right)$ coincide con el $\frac{t}{n+1}$-ésimo quantil de la distribución $F_{\varepsilon_{(t)}^*}(.)$ y **ii)** la sucesión $\left\{\varepsilon_{(1)}, \varepsilon_{(2)}, ..., \varepsilon_{(n)}\right\}$ se vincula con los quantiles mediante la relación $F_{\varepsilon_{(t)}}^{-1}(Q) = \varepsilon_{(t)}\left(\frac{t-1}{n} < Q < \frac{t}{n}\right)$. Estas relaciones sugieren la consideración de una sucesión de cuantiles $\{Q_1, Q_2, ..., Q_n\}$, seleccionada de modo que se verifique que $F_{\varepsilon_{(t)}}^{-1}(Q_t) \approx E\left(\varepsilon_{(t)}\right)(t = 1,2,...,n)$. En consecuencia, graficando los pares ordenados $\left\{F_{\varepsilon_{(t)}}^{-1}(Q_t); \varepsilon_{(t)}\right\}(t = 1,2,...,n)$ se obtiene el denominado diagrama Q-Q. Si la sucesión $\{\varepsilon_t\}(t = 1,2,...,n)$ (formada por variables aleatorias independientes) presenta una distribución de la forma $F_{\varepsilon_{(t)}}(\varepsilon)$, entonces el diagrama Q-Q asumirá una forma lineal. La selección de los argumentos Q_{0t} depende de la naturaleza de la función de distribución de referencia $F_{\varepsilon_{(t)}}(Q_{0t})$. En general se suele utilizar la sucesión $Q_t = \frac{t-c}{n-2c+1}(0 \le c \le 1; t = 1,2,...,n)$. En particular, si la distribución de referencia es la Normal,

la sucesión es de la forma $Q_t = \frac{t - 0.375}{n + 0.25}\,(t = 1,2,...,n)$.

A partir de la forma en que se distribuyan los puntos alrededor de la recta $\left\{F^{-1}_{\varepsilon_{(t)}}(Q_t);\varepsilon_{(t)}\right\}$ $(t = 1,2,...,n)$ se pueden obtener la siguientes conclusiones: **i)** si los puntos del diagrama ocupan el mismo lugar geométrico que la recta $Q_t = Q_{0t}$, obviamente se puede concluir que las distribuciones coinciden; **ii)** si los puntos del diagrama observan un ordenamiento lineal paralelo a la recta, se puede concluir que las dos distribuciones difieren en su posición; **iii)** si los puntos del diagrama observan un ordenamiento lineal que pasa por el origen, pero que no es paralelo a la recta, se puede concluir que las dos distribuciones coinciden en su forma, pero difieren en su variabilidad; **iv)** si los puntos del diagrama observan un ordenamiento no-lineal caracterizado por un comportamiento descendente (por debajo de la recta) en el extremo superior de la distribución y ascendente (por encima de la recta) en el extremo inferior, se puede concluir que la distribución de $\{\varepsilon_t\}$ posee "colas" más largas que la distribución de referencia, viceversa, si los puntos del diagrama observan un ordenamiento no-lineal caracterizado por un comportamiento ascendente en el extremo superior de la distribución y descendente en el extremo inferior, se puede concluir que la distribución de $\{\varepsilon_t\}$ posee "colas" más cortas que la distribución de referencia[131]; **v)** si la distribución de referencia es simétrica y los puntos del diagrama observan una cierta curvatura en todo su recorrido, en particular, si observan un ordenamiento alrededor de una curva con pendiente creciente hacia la derecha, se puede concluir que la distribución de $\{\varepsilon_t\}$ es asimétrica hacia la derecha; **vi)** si el comportamiento de la sigmoide alrededor de la cual se agrupan las observaciones es más pronunciada que el lineal, se puede concluir que la distribución de $\{\varepsilon_t\}$ es leptokúrtica[132].

Debe tenerse en cuenta que la variabilidad del proceso (en particular, la presencia de valores extremos)[133] afecta a la asimetría y a la kurtosis y, por lo tanto, puede provocar desvíos de la linealidad en la distribución de las observaciones, aún cuando la hipótesis de referencia sea válida, especialmente en series cortas[134].

En general, el diagrama Q-Q es más apropiado para caracterizar la kurtosis de la distribución y el P-P es más apropiado para caracterizar su asimetría. Asimismo, el diagrama Q-Q es más sensible que el P-P para analizar el comportamiento de las "colas" y, viceversa, el P-P es más apropiado para considerar el comportamiento de la región central de la distribución.

Por otra parte, el diagrama Q-Q puede ser utilizado para estimar los parámetros de posición y escala

---

131. Como "cola" se suele definir (en forma arbitraria) la zona ubicada en los intervalos $\left(-\infty, m(\varepsilon) - 2\sigma(\varepsilon)\right)$ y $\left(m(\varepsilon) + 2\sigma(\varepsilon), \infty\right)$.

132. El término "kurtosis", como denominación de la medida de apuntamiento con respecto al comportamiento de la "colas", se debe a Pearson, K. (1906) y proviene del griego κύρτωση (literalmente: curvatura del lomo). De él derivan los calificativos **mesokúrtica**, del griego μέσος (que significa "medio", con referencia a la distribución para la cual $K = 3$) **platokúrtica**, del griego πλατυς (que significa "ancho", con referencia a la distribución para la cual $K < 3$) y **leptokúrtica**, del griego λεπτός (que significa "delgado", conreferencia a la distribución para la cual $K > 3$).

133. Los métodos de detección y análisis de la naturaleza de los valores extremos (**"outliers"** o **contaminantes**) serán tratados en el Cap. 4.

134. Debe tenerse en cuenta que las representaciones $\left\{\varepsilon_{(t)}; F^{-1}_{\varepsilon^*_{(t)}}(\varepsilon)\right\}(t = 1,2,...,n)$ y $\left\{\varepsilon_{(t)}; m + \sigma G^{-1}_{\varepsilon^*_{(t)}}(\varepsilon)\right\}(t = 1,2,...,n)$, donde $\varepsilon^*_{(t)}$ denota la variable de referencia, son equivalentes cuando $G_{\varepsilon^*_{(t)}}(\varepsilon) = F_{\varepsilon^*_{(t)}}(\varepsilon)$. De todas las consideraciones realizadas se puede concluir que la condición necesaria para la validez de este método de análisis es que las variables ε_t sean iid.

de la distribución de referencia a partir del término independiente y la pendiente de la recta $F^{-1}_{\varepsilon_{(t)}}(Q_t) = f(\varepsilon_{(t)})$.

A partir de los resultados obtenidos de los diagramas P-P y Q-Q, una forma de aproximar la distribución de una variable ε_t a la función Normal, es mediante la aplicación de la **familia de transformaciones de Johnson**[135], definida de la siguiente forma:

$$T(\varepsilon_t) = \gamma + \delta g\left(\frac{\varepsilon_t - \beta}{\eta}\right)$$

donde β, γ, δ y η son constantes a determinar y g asume las siguientes formas, de acuerdo a tres tipos de distribuciones[136]:

$$g(\varepsilon_t) = \log(\varepsilon_t) \qquad (sistema\ \log\text{-}\ Normal)$$

$$g(\varepsilon_t) = \log\left(\frac{\varepsilon_t}{1-\varepsilon_t}\right) \qquad (sistema\ S_B\ de\ Pearson)$$

$$g(\varepsilon_t) = senh^{-1}(\varepsilon_t) \qquad (sistema\ S_U\ de\ Pearson)$$

2.14.3.- Los tests para decidir acerca de la adecuacidad de una distribución

Estos tests fueron desarrollados a partir: **i)** de los momentos; **ii)** de la distribución de probabilidades y **iii)** de los estadísticos de orden de la función de referencia.

Una forma de decidir acerca de la adecuacidad de la función Normal para representar la distribución de la variables ε_t iid, es utilizando el **test de Jarque-Bera**, el cual se basa en las propiedades de los momentos centrados de tercero y cuarto orden y considera la hipótesis nula[137]:

$$H_0: las\ perturbaciones\ \varepsilon_t\ se\ distribuyen\ Normalmente$$

El estadístico a utilizar es de la forma:

$$JB = \frac{n}{6}\left[\hat{As}^2(\varepsilon) + \frac{1}{4}\left(\hat{K}(\varepsilon) - 3\right)^2\right]$$

donde $As(\hat{\varepsilon}) = \dfrac{\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^3}{\left(\sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2}\right)^3}$ y $K(\hat{\varepsilon}) = \dfrac{\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^4}{\left[\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2\right]^2}$ denotan los estimadores máximo-verosímiles de los

---

135. Ver Johnson, N.L. (1949).

136. Ver Landro, A.H. (2002).

137. Ver Jarque, C.M.; Bera, A.K. (1980)(1987), Bera, A.K.; Jarque, C.M. (1981).

coeficientes de asimetría y kurtosis de la sucesión $\{\hat{\varepsilon}_t\}$ $(t = 1,2,...,n)$, respectivamente[137].

Teniendo en cuenta que, bajo el supuesto de que la hipótesis de Normalidad es verdadera, estos estimadores convergen en-distribución a la función Normal[138]:,

$$\hat{As}(\varepsilon) \xrightarrow{d} N\left(0, \frac{6}{n}\right)$$

y

$$\hat{K}(\varepsilon) \xrightarrow{d} N\left(3; \frac{24}{n}\right)$$

y son asintóticamente independientes, se puede concluir que $JB \xrightarrow{d} \chi^2_{(2)}$ [139].

Obviamente la restricción más importante para la utilización de este test radica en la distribución asintótica del estadístico JB. A fin de resolver el problema del sesgo que se genera al utilizar muestras finitas, Doornik, J.A.; Hansen, H. (1994) propusieron un test basado en aproximaciones de las distribuciones de la asimetría y la kurtosis bajo los supuestos: **i)** de Normalidad; **ii)** de distribución Gamma de la kurtosis y **iii)** que $\hat{K}^2(\varepsilon) > 1 + \hat{As}^2(\varepsilon)$ (para un análisis de una aproximación a la distribución de la asimetría en muestras finitas, ver D'Agostino, R.B. (1970))[140]. En estas condiciones, demostraron que el estadístico de Wald:

$$\hat{W} = As^{*2}(\hat{\varepsilon}) + K^{*2}(\hat{\varepsilon}) \xrightarrow{d} \chi^2_{(2)}$$

donde $As^*(\hat{\varepsilon}) = \frac{1}{\sqrt{\log(w)}} \log\left(g + \sqrt{1+g^2}\right)$ y $K^*(\hat{\varepsilon}) = \left[\left(\frac{X}{2\alpha}\right)^{1/3} - 1 + \frac{1}{9\alpha}\right]\sqrt{9\alpha}$ denotan, respectivamente, los estimadores de la asimtería y de la kurtosis obtenidos a partir de una muestra finita, siendo: **i)** $g = As^2(\hat{\varepsilon})\sqrt{\frac{w^2-1}{2}\frac{(n+1)(n+3)}{6(n-2)}}$; **ii)** $w^2 = -1 + \sqrt{2(b_0 - 1)}$; **iii)** $X = 2b_1\left(K(\hat{\varepsilon}) - 1 - As^2(\hat{\varepsilon})\right)$; **iv)** $\alpha = b_2 + b_3 As^2(\hat{\varepsilon})$; **v)** $b_0 = \frac{3(n^2 - 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$; **vi)** $b_1 = \frac{(n+5)(n+7)(n^3 + 37n^2 + 11n - 313)}{12(n-3)(n+1)(n^2 + 15n - 4)}$; **vii)** $b_2 = \frac{(n-2)(n+5)(n+7)(n^2 + 27n - 70)}{6(n-3)(n+1)(n^2 + 15n - 4)}$ y **viii)** $b_3 = \frac{(n-7)(n+5)(n+7)(n^2 + 2n - 5)}{6(n-3)(n+1)(n^2 + 15n - 4)}$.

Por su parte, a fin de resolver el problema que surge de la estimación de la asimetría y la varianza

---

137. Estos estimadores fueron propuestos por Pearson, K. (1895). Dados los valores de sus varianzas asintóticas, estos estimadores resultan aceptables sólo para conjuntos de observaciones suficientemente grandes.

138. Ver Snedecor, G.W.; Cochran, W.G. (1980).

139. Para el caso en que las variables ε_t sean débilmente autocorrelacionadas, Bai, I.; Ng, S.N. (2005) propusieron distribuciones alternativas para los estimadores muestrales de la asimetría y la kurtosis y un test de Normalidad conjunta.

140. Existen tests de Normalidad previos a éste que consideran los coeficientes de asimetría y kurtosis separadamente y en forma conjunta (en particular, como una combinación lineal de ambos). Ver Pearson, E.S.; Hartley, H.O. (1982) y Pearson, E.S.; D'Agostino, R.B.; Browman, K.O. (1977).

el test de Jarque-Bera a partir de la utilización de estimadores del valor medio y de la varianza afectados por errores muestrales, Richardson, M.; Smith, T. (1993) y Harvey, C.R. (1995) propusieron, en forma independiente, la estimación conjunta de los momentos a partir de las siguientes condiciones de ortogonalidad del método generalizado de los momentos:

$$\begin{cases} e_{1,t} = \hat{\varepsilon}_t - \bar{\varepsilon} \\ e_{2,t} = \left(\hat{\varepsilon}_t - \bar{\varepsilon}\right)^2 - \hat{\sigma}_\varepsilon^2 \\ e_{3,t} = \dfrac{\left(\hat{\varepsilon}_t - \bar{\varepsilon}\right)^3}{\hat{\sigma}_\varepsilon^3} - \hat{As}(\varepsilon) \\ e_{4,t} = \dfrac{\left(\hat{\varepsilon}_t - \bar{\varepsilon}\right)^4}{\hat{\sigma}_\varepsilon^4} - \hat{K}(\varepsilon) \end{cases}$$

La solución de este sistema de ecuaciones proporciona el estimador conjunto ($\hat{\theta}$) de los cuatro primeros momentos, $\theta' = \left[m(\varepsilon) \quad \sigma^2(\varepsilon) \quad As(\varepsilon) \quad K(\varepsilon)\right]$ y una matriz de covarianzas asociada ($\Gamma(\hat{\varepsilon})$) y procura, a su vez, corregir la dependencia de la distribución asintótica de dichos momentos[142].

A partir de estos resultados, Richardson-Smith propusieron un test para contrastar la hipótesis nula $H_0: As(\varepsilon) = K(\varepsilon) - 3 = 0$ utilizando un estadístico de Wald de la forma:

$$W^* = n\left(G\hat{\theta}\right)'\left(G\Gamma(\varepsilon)G'\right)^{-1}G\hat{\theta}:\chi^2_{(2)}$$

Existen otros tests de características más generales[143] para decicir acerca de la adecuacidad de una función de probabilidades tomada como referencia para representar la distribución de un conjunto de observaciones $\left\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\right\}$, entre los que cabe mencionar:

1.- El **test de Kolmogorov-Smirnov**[144] el cual, basado en el comportamiento de las funciones de distribución de la variable $\varepsilon_t\left(F_{\varepsilon_t}(\varepsilon)\right)$ y de la variable de referencia, $F_0(\varepsilon,\theta)$, contrasta las siguientes

---

142. Harvey, C.R. (1995) propuso la estimación bajo el supuesto que la hipótesis nula es verdadera y utilizó las dos últimas condiciones de ortogonalidad como condiciones de sobre-identificación.

143. La literatura reconoce a estos métodos como "goddness of fit tests". En general estos tests intentan decidir acerca de la adecuacidad de una función de referencia para representar un conjunto cualquiera de funciones aleatorias definidas paramétricamente. Es decir, intentan decidir acerca de la aceptación de la hipótesis nula $H_0: F_{\varepsilon_t}(\varepsilon) = F_0(\varepsilon,\theta)$ (donde: **i)** $F_{\varepsilon_t}(\varepsilon)$ denota la distribución que genera la muestra $\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n$ y **ii)** $F_0(\varepsilon,\theta)$ denota la distribución de probabilidades de referencia). Si θ asume un valor fijo $\theta = \theta_0$, entonces $H_0: F_{\varepsilon_t}(\varepsilon) = F_0(\varepsilon,\theta_0)$ es una hipótesis simple y está completamente especificada. Ahora bien, en muchos casos los verdaderos valores de los parámetros, bajo el supuesto que la hipótesis H_0 es verdadera, no son conocidos y, por lo tanto, deben ser estimados a partir de las observaciones. Entonces la hipótesis nula (compuesta) a contrastar es que $F_{\varepsilon_t}(\varepsilon)$ es un miembro no-especificado de una familia de distribuciones de probabilidades. Como se verá, muchos de estos tests (paramétricos y no-paramétricos) son de distribución libre (es decir, la distribución del estadístico no depende de θ) y otros son de distribución-específica. Pueden estar referidos a distribuciones discretas o continuas, unidimensionales o multidimensionales. Cada test merece un análisis particular para cada alternativa y, no se puede asegurar que un test sea el mejor en todos los casos (ver Hall, P. (1985), Bickel, P.J.; Ritov, Y. (1992), Ingster, Y.I. (1993), Choi, S.; Hall, W.J.; Schick, A. (1996), Janssen, A. (2000), Bickel, P.J.; Ritov, Y.; Stocker, T.M. (2003), Claeskens, G.; Hjort, N.I. (2003)).

144. Ver Kolmogorov, A.N. (1933), Smirnov, N. (1935)(1939), Kotz, S.; Johnson, N.L. (eds.) (1983).

hipótesis: $H_0: F_{\varepsilon_t}(\varepsilon) = F_0(\varepsilon,\theta)$ y $H_1: F_{\varepsilon_t}(\varepsilon) \neq F_0(\varepsilon,\theta)$. Considera la función de distribución empírica $F^*_{\varepsilon_t}(\varepsilon)$ como estimador de $F_{\varepsilon_t}(\varepsilon)$ y el estadístico definido como la mayor distancia entre $F^*_{\varepsilon_t}(\varepsilon)$ y $F_0(\varepsilon,\theta)$:

$$D = Max\left(D^+, D^-\right) = \underset{\varepsilon}{Max}\left[F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right]$$

(donde $D^+ = Max\left[F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right]$ y $D^- = \min\left[F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right]$ y θ denota el vector de los parámetros de la distribución de referencia)[145].

Debe tenerse en cuenta que, dada una función monótona $g(\varepsilon_t)$, las variables D^+ y D^- no se modifican, lo que demuestra que las distribuciones de probabilidades de D^+ y D^- son independientes de $F_0(\varepsilon,\theta)$.

A partir de la propuesta de Kolmogorov, A.N. (1933), Smirnov, N.V. (1949) obtuvo la siguiente función de distribución asintótica del estadístico D:

$$\lim_{n\to\infty} p\left(\sqrt{n}\ \underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right| \leq z\right) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1} e^{-2j^2z^2} = L(z)$$

a partir de la cual se obtiene la siguiente función de densidad[146]:

$$f\left(\sqrt{n}\ \underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right| = z\right) = 8z\sum_{j=1}^{\infty}(-1)^{j-1} j^2 e^{-2j^2z^2}$$

Luego, teniendo en cuenta que, para $z > 1$, el primer término de la fórmula de Smirnov es mucho mayor que los demás, se puede considerar la siguiente aproximación[147]:

$$\lim_{n\to\infty} p\left(\sqrt{n}\ \underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right| \leq z\right) \approx 1 - 2e^{-2z^2}$$

---

[145]. Como una modificación al de Kolmogorov-Smirnov, Kuiper, N.H. (1960) propuso el estadístico $V = D^+ - D^-$ para testar observaciones sobre un círculo.

[146]. A partir de esta función de densidad se obtiene la siguiente expresión para los momentos absolutos: $m_s\left(\underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right|\right) = \dfrac{\Gamma\left(\frac{s}{2}+1\right)}{2^{(s/2)+1}}\sum_{j=1}^{\infty}(-1)^{j-1} j^{-s}\ (s = 1,2,\ldots)$, de los cuales se obtienen las siguientes medidas: $E\left(\underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right|\right) = \sqrt{\dfrac{\pi}{2}}\ln(2) \approx 0.8687$ y $\sigma^2\left(\underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right|\right) = \dfrac{\pi^2}{8} - \dfrac{\pi}{2}\left[\ln(2)\right]^2 \approx 0.6771$.

[147]. Debe tenerse en cuenta que, en las aplicaciones, esta aproximación está afectada por los errores que se originan : **i)** en la utilización de una distribución asintótica para un conjunto finito de observaciones y **ii)** en haber considerado sólo el primer término de la suma. Para valores pequeños de z se obtuvo la siguiente expresión para la función de distribución asintótica: $\lim_{n\to\infty} p\left(\sqrt{n}\ \underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)\right| \leq z\right) =$

$= \dfrac{\sqrt{2\pi}}{z}\sum_{j=1}^{\infty}\exp\left[-\dfrac{(2j-1)^2\pi^2}{8z^2}\right]$ y la correspondiente aproximación $\dfrac{\sqrt{2\pi}}{z}\exp\left(-\dfrac{n^2}{8z^2}\right)$.

De acuerdo a lo expuesto precedentemente, H_0 será verdadera si $E\left[F^*_{\varepsilon_t}(\varepsilon)\right] = F_0(\varepsilon,\theta)$, es decir, se puede considerar que existirán razones suficientes para rechazar la hipótesis nula solamente si el estadístico de D asume un valor mayor que el valor crítico z [147]. Luego, es necesario hallar el valor de z tal que $p\left[\left(D > \frac{z}{\sqrt{n}}\right) / H_0\right] = \alpha$. De modo que se puede aceptar H_0 si, y sólo si[148] $D < \sqrt{-\frac{1}{2n}\ln\left(\frac{\alpha}{2}\right)} = z$.

En el caso en que la distribución $F_0(\varepsilon,\theta)$ contenga coeficientes θ desconocidos, estos pueden ser estimados por máxima verosimilitud (en general se requiere que los estimadores de los coeficientes θ sean asintóticamente eficientes). Srinivasan, R. (1970) sugiere la utilización del teorema de Rao-Blackwell para mejorar la estimación de $F_0(\varepsilon,\theta)$ y luego calcular el estadístico D [149]. En particular, para las distribuciones de referencia Normal o exponencial, los correspondientes valores críticos fueron tabulados por Stephens, M.A. (1974)[150].

Con un razonamiento similar, el estadístico D^+ puede ser utilizado para testear la adecuacidad de la distribución $F_0(\varepsilon,\theta)$ con respecto a alternativas $F_A(\varepsilon)$ para las cuales -idealmente- se verifique que $F_A(\varepsilon) \geq F_0(\varepsilon,\theta)$ sobre la mayor parte del rango, lo cual implica que el valor medio de la variable de referencia es menor que el valor medio de la distribución de las observaciones. De la misma forma, el estadístico D^- puede ser utilizado con respecto a alternativas en las que el valor medio de la variable de referencia es mayor que el valor medio de la distribución de las observaciones[151].

Debe tenerse en cuenta que, dado que el estadístico D está calculado sobre una sola diferencia entre $F^*_{\varepsilon_t}(\varepsilon)$ y $F_0(\varepsilon,\theta)$, el test de Kolmogorov-Smirnov resulta en muchos casos menos potente que si se utilizara un estadístico que considerara las diferencias $F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta)$ para todos los valores observados, pero esta consideración genera un teoría de los estimadores excesivamente complicada[152] que hace prácticamente

---

147. Los valores críticos se encuentran tabulados.

148. Es decir, para un valor de z calculado de la forma indicada en el texto, se verificará que $p\left\{\underset{\varepsilon}{Max}\left|F^*_{\varepsilon_t}(\varepsilon).F_0(\varepsilon,\theta)\right| \leq \frac{z}{\sqrt{n}}\right\} = \quad = 1-\alpha$ para todo ε. De modo que se puede construir un intervalo de confiabilidad de la forma $F^*_{\varepsilon_t}(\varepsilon) - \frac{z}{\sqrt{n}} \leq F_0(\varepsilon,\theta) \leq F^*_{\varepsilon_t}(\varepsilon,\theta) + \frac{z}{\sqrt{n}}$. Con un razonamiento similar se puede construir un test para comparar dos distribuciones cualesquiera, a partir de las correspondientes distribuciones muestrales. Dadas dos muestras de tamaño n_1 y n_2, se puede utilizar el criterio $\lim_{n_1,n_2\to\infty} p(Z^* \leq z) = L(z) = \sum_{j=1}^{\infty}(-1)^j \exp\left(-2j^2z^2\right)\ (z>0)$. Ver Mardia, K.V. (1967), Friedman, J.H.; Rafsky, L.C. (1979)(1981), Wellner, J.A. (1979), Brown, B.M. (1982), van Putten, B. (1987), Henze, N. (1988), Neuhaus, G.; Zhu, L.X. (1999), Hall, P.; Tajvidi, N. (2002).

149. Debe tenerse en cuenta que, excepto que $F_0(\varepsilon,\theta)$ represente la función exponencial, la estimación máximo-verosímil proporciona expresiones sumamente complicadas.

150. Ver también Pearson, E.S.; Hartley, H.O. (1972), Kumar, A.; Pathak, P.K. (1977).

151. A partir de una aproximación diferente, Feller, W. (1948) demostró que $4n\left(D^+\right)^2 \xrightarrow{d} \chi^2_{(2)}$.

152. Para un análisis de la potencia de este test, ver Seier, E. (2002).

imposible su aplicación[154].

2.- El **test de Cramér-von Mises**[155] , el cual considera un estadístico de la forma:

$$W^2_{(n)} = n \int_{-\infty}^{\infty} \left[F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta) \right]^2 dF_0(\varepsilon,\theta)$$

(donde, como en el test de Kolmogorov-Smirnov, $F^*_{\varepsilon_t}(\varepsilon)$ denota la distribucióm empírica como estimador de $F_{\varepsilon_t}(\varepsilon)$, $F_0(\varepsilon,\theta)$ denota la función de distribución de la variable de referencia y θ denota el vector de los parámetros). La distribución de probabilidades de este estadístico es sumamente complicada[156], sus momentos son de la forma[157]:

$$m_1\left(W^2_{(n)}\right) = \frac{1}{6}$$

$$m_2\left(W^2_{(n)}\right) = \frac{4n-3}{180n}$$

$$m_3\left(W^2_{(n)}\right) = \frac{32n^2 - 61n + 30}{3780n^2}$$

$$m_4\left(W^2_{(n)}\right) = \frac{496n^3 - 1532n^2 + 1671n - 630}{75600n^3}$$

La distribución de este estadístico es independiente de la forma de la distribución de referencia $F_0(\varepsilon,\theta)$.

Para el caso en que los coeficientes θ sean desconocidos, no es posible el cálculo exacto de los

---

154. Algunos autores propusieron estadísticos vinculados con la relación $\frac{F^*_{\varepsilon_t}(\varepsilon)}{F_0(\varepsilon,\theta)}$ en general, con las restricciones $a \le F_0(\varepsilon,\theta) \le b$ ó $a \le F^*_{\varepsilon_t}(\varepsilon) \le b$ (Ver Renyi, A. (1967)).

155. Mencionado en la literatura también como **test de Smirnov** (ver Smirnov, N.V. (1949)).

156. Anderson, T.W.; Darling, D.A. (1952) construyeron, a partir de la expresión $\lim_{n\to\infty} p\left(W^2_{(n)} \le z\right) = \frac{1}{n\sqrt{z}} \sum_{j=0}^{\infty} (-1)^j \binom{-\frac{1}{2}}{j} \sqrt{4j+1}\ .\ e^{-z_j^*} K_{1/4}\left(z_j^*\right)$ (donde: **i)** $z_j^* = \frac{1}{z}\left(j+\frac{1}{4}\right)^2$ y **ii)** $K_{1/4}(\bullet)$ denota una función de Bessel modificada), una tabla de valores de z tales que $\lim_{n\to\infty} p\left(W^2_{(n)} \le z\right) = \alpha$ (ver Durbin, J. (1973)).

157. Ver Pearson, E.S. (1963). Riedwyl, H. (1967) propuso estadísticos relacionados con $D^-_{(n)}, D^+_{(n)}, D_{(n)}$ y $W^2_{(n)}$ a partir del comportamiento de los desvíos $d_t = F^*_{y_t} - F_0(y,\theta)$. Estos tests consideran, entre otros, los estadísticos $\sum_{t=1}^{n} d_t^2, \sum_{t=1}^{n} |d_t|, \sum_{t=1}^{n} d_t, \underset{t}{Max}(d_t)$ y $\underset{t}{Max}(|d_t|)$.

correspondientes valores críticos. Un cálculo asintótico, para las distribuciones Normal y exponencial, puede ser hallado en Durbin, J.; Knott, M.; Taylor, C.C. (1975)[158].

3.- El **test de Stephens**[159], el cual considera un estadístico de la forma:

$$U^2_{(n)} = n\int_{-\infty}^{\infty}\left\{F^*_{\varepsilon_t}(\varepsilon) - F_0(\varepsilon,\theta) - \int_{-\infty}^{\infty}\left[F^*_{\varepsilon_t}(u) - F_0(u,\theta)\right]dF_0(u,\theta)\right\}^2 dF_0(\varepsilon,\theta)$$

La distribución de probabilidades de este estadístico es sumamente complicada[160]. Sus momentos son de la forma[161]:

$$m\left(U^2_{(n)}\right) = \frac{1}{12}$$

$$m_2\left(U^2_{(n)}\right) = \frac{n-1}{360n}$$

$$m_3\left(U^2_{(n)}\right) = \frac{(n-1)(2n-3)}{7560n^2}$$

$$m_4\left(U^2_{(n)}\right) = \frac{(n-1)\left(19n^2 - 51n + 36\right)}{302400n^3}$$

También en este caso la distribución del estadístico $U^2_{(n)}$ es independiente de $F_0(\varepsilon,\theta)$.

4.- El **test de Sherman**, el cual considera un estadístico de la forma:

$$\frac{1}{2}\sum_{t=1}^{n+1}\left|F\left(\varepsilon_{(t)}\right) - F\left(\varepsilon_{(t-1)}\right)\right| - \frac{1}{n+1} \qquad \left(F\left(\varepsilon_{(0)}\right) = 0, F\left(\varepsilon_{(n+1)}\right) = 1\right)$$

basado en los estadísticos de orden $\varepsilon_{(1)} < \varepsilon_{(2)} < ... < \varepsilon_{(n)}$, a partir de los cuales las funciones $F\left(\varepsilon_{(1)}\right), F\left(\varepsilon_{(2)}\right), ..., F\left(\varepsilon_{(n)}\right)$, como se mencionó en la Sec. 2.14.2, pueden ser consideradas como variables aleatorias independientes, estandarizadas, con distribución uniforme, de modo que:

$$E\left[F\left(\varepsilon_t\right)\right] = \frac{t}{n+1}$$

$$E\left[F\left(\varepsilon_{(t)}\right) - F\left(\varepsilon_{(t-1)}\right)\right] = \frac{1}{n+1}$$

---

158. Algunas extensiones de estos resultados pueden ser halladas en Stephens, M.A. (1974)(1976).

159. Ver Stephens, M.A. (1963)(1964).

160. Para mayores detalles sobre la distribución de probabilidades de U^2, ver Durbin, J. (1973), Pearson, E.S.; Stephens, M.A. (1962), Watson, G.S. (1961). En Pearson, E.S.; Hartley, H.O. (1972) figuran las tablas de valores críticos correspondientes al test U^2.

161. Ver Stephens, M.A. (1963).

5.- El **test de Anderson-Darling**[162], el cual considera un estadístico de la forma:

$$\left(W^*_{(n)}\right)^2 = n \int_{-\infty}^{\infty} \frac{\left[F^*_{\varepsilon_i}(\varepsilon) - F_0(\varepsilon,\theta)\right]^2}{F_0(\varepsilon,\theta)\left[1 - F_0(\varepsilon,\theta)\right]} dF_0(\varepsilon,\theta)$$

cuya distribución de probabilidades asintótica es de la forma:

$$\lim_{n\to\infty} p\left[\left(W^*_{(n)}\right)^2 \le z\right] = \frac{\sqrt{2\pi}}{z} \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j} (4j-1) e^{-z^*_j} \int_0^{\infty} \exp\left[\frac{z}{8\left(u^2+1\right)} - z^*_j u^2\right] du$$

donde $z^*_j = \dfrac{n^2(4j+1)}{8z}$.

Obsérvese que este test es una modificación del test de Cramér-von Mises consistente en utilizar un sistema de ponderaciones $\psi(\varepsilon) = \left\{F_0(\varepsilon,\theta)\left[1 - F_0(\varepsilon,\theta)\right]\right\}^{-1}$. Como $E\left\{n\left[F^*_{\varepsilon_i}(\varepsilon) - F_0(\varepsilon,\theta)\right]^2\right\} = F_0(\varepsilon,\theta).\left[1 - F_0(\varepsilon,\theta)\right]$, se puede concluir que estas ponderaciones privilegian las "colas" de la distribución.

6.- El **test χ^2**, el cual se basa en la división del conjunto de n residuos, $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\}$ (para n suficientemente grande) en k clases, $A_1, A_2, ..., A_k$, cada una con una frecuencia absoluta f_i $(i = 1,2,...,k)$ y en el contraste de la hipótesis nula:

H_0: La probabilidad de que, dada una distribucion de probabilidades de referencia,

un residuo este incluido en una clase A_i $(i = 1,2,...,k)$ es p_i $(i = 1,2,...,k)$

(es decir, $H_0: f_i = p_1$ $(i = 1,2,...,k)$) con respecto a las hipótesis alternativas que consideran que la probabilidad de que una observación esté incluida en la clase A_i $(i = 1,2,...,k)$ asuma otros valores distintos de p_i $(i = 1,2,...,k)$ ($H_1: f_i \neq p_i$ $(i = 1,2,...,k)$). Con este fin, considera un cociente de verosimilitudes[163] (definido, a partir del exponente de la aproximación Normal k -dimensional de la distribución multinomial, como la suma de los desvíos entre las frecuencias observadas y sus valores esperados ponderados inversamente por las esperanzas matemáticas)[164] de la forma:

---

[162]. Ver Anderson, T.W.; Darling, D.A. (1952).

[163]. Este cociente fue propuesto por Pearson, K. (1900) y su expresión es de la forma $\Lambda = \prod_{i=1}^{k} \left(\dfrac{np_i(\theta)}{f_i}\right)^{f_i}$. Habitualmente es utilizado de acuerdo con la expresión $-2\ln\left(\Lambda(\theta)\right) = -2\sum_{i=1}^{k} f_i\left[\ln\left(np_i(\theta)\right) - \ln\left(f_i\right)\right] \xrightarrow{d} \chi^2_{(k-1)}$. Si el vector de los coeficientes $\theta = \left[\theta_1 \quad \theta_2 \quad ... \quad \theta_q\right]$ es desconocido y, por lo tanto, los coeficientes θ_j deben ser estimados, se verifica que, bajo el supuesto que la hipótesis nula sea verdadera, $-2\ln\left(\Lambda\left(\hat{\theta}\right)\right) \xrightarrow{d} \chi^2_{(k-q-1)}$, donde $\hat{\theta}$ denota el vector de los estimadores (por ejemplo, máximo-verosímiles) de los coeficientes θ_j $(j = 1,2,...,q)$.

[164]. Para un análisis de las formas generalizadas de este test, ver Patnaik, P.B. (1949).

$$X^2 = \sum_{i=1}^{k} \frac{\left(f_i - np_i(\theta)\right)^2}{np_i(\theta)} .$$

Considerando a las frecuencias relativas como asimilables a probabilidades , $\frac{f_i}{n} = \hat{p}_i \ \ (i = 1,2,...,k)$, de modo que $\sum_{i=1}^{k} \hat{p}_i = 1$, entonces, bajo el supuesto de que la hipótesis nula es verdadera, de acuerdo con el teorema central del límite, se verifica que:

$$\frac{\hat{p}_i - p_i}{\sqrt{\frac{p_i q_i}{n}}} \xrightarrow{d} N(0;1) \ (i = 1,2,...,k)$$

y, por lo tanto, se puede concluir que:

$$T = n\sum_{i=1}^{k} \frac{\left(\hat{p}_i - p_i\right)^2}{p_i} \xrightarrow{d} \chi^2_{(k-1)}$$

De modo que se considerará que no existen razones suficientes para rechazar H_0 si se verifica que $T < \chi^2_{(k-1)}$.

Debe tenerse en cuenta que estos tests pierden potencia cuando los valores $np_i \ (i = 1,2,...,k)$ son demasiado pequeños. La regla a considerar es: **i)** que se verifique que $np_i > 1 \ (i = 1,2,...,k)$ y **ii)** que el 80% de estos valores sean mayores o iguales a 5[165].

Debe tenerse en cuenta que, en el caso en que la distribución de referencia sea discreta, el test de Kolmogorov-Smirnov es más potente que el χ^2.

Si la distribución de referencia contuviera coeficientes desconocidos, estos pueden ser estimados, por ejemplo, por máxima verosimilitud. En general se demuestra que, si los estimadores de los coeficientes θ son asintóticamente eficientes, entonces el test conserva sus propiedades asintóticas[166].

7.- El **test de Moore**[167], el cual considera un estadístico de la forma:

$$T_n\left(\hat{\theta}_n\right) = V_n^T\left(\hat{\theta}_n\right)\left[I_{(n)} - B\left(\hat{\theta}_n\right)J^{-1}\left(\hat{\theta}_n\right)B^T\left(\hat{\theta}_n\right)\right]^{-1} V_n\left(\hat{\theta}_n\right) \xrightarrow{d} \chi^2_{(k-1)}$$

donde:

---

165. Ver Horn, S.D. (1977).

166. Supóngase, por ejemplo que en la distribución de referencia $p_i(\theta)$ el vector de los coeficientes θ sea desconocido y que se haya estimado por máxima verosimilitud. Entonces, bajo el supuesto de que la hipótesis $H_0 : f_i = p_i(\theta) \ (i = 1,2,...,k)$ es verdadera se verifica que $X^2\left(\hat{\theta}\right) = \sum_{j=1}^{k} \frac{f_j - np_j\left(\hat{\theta}\right)}{np_j\left(\hat{\theta}\right)} \xrightarrow{d} \chi^2_{(k-q-1)}$, donde q denota el número de coeficientes estimados. Para un tratamiento riguroso del test χ^2 , ver Cramér, H. (1945), Moore, D.S.; Spruill, M.C. (1975), Moore, D.S. (En Hogg, R.V. (ed.) (1979)).

167. Ver Moore, D.S. (1977).

$$V_n^T(\theta) = \left[\frac{f_1 - np_1(\theta)}{\sqrt{np_1(\theta)}} \quad \cdots \quad \frac{f_k - np_k(\theta)}{\sqrt{np_k(\theta)}} \right]$$

$$B(\theta) = \begin{bmatrix} \frac{1}{p_1^*(\theta)} \frac{\partial p_1^*(\theta)}{\partial \theta_1} & \cdots & \frac{1}{p_1^*(\theta)} \frac{\partial p_1^*(\theta)}{\partial \theta_q} \\ \cdots & \cdots & \cdots \\ \frac{1}{p_k^*(\theta)} \frac{\partial p_k^*(\theta)}{\partial \theta_1} & \cdots & \frac{1}{p_k^*(\theta)} \frac{\partial p_k^*(\theta)}{\partial \theta_q} \end{bmatrix}$$

$$p_j^*(\theta) = \int_{A_j} f_{\varepsilon_t}(\varepsilon, \theta) d\varepsilon$$

$J(\theta)$ denota la matriz de información para $F_0(\varepsilon, \theta)$ y $\hat{\theta}_j$ $(j = 1,2,...,q)$ denota el estimador máximo-verosímil θ_j.

8.- El **test de Shapiro-Wilk**, el cual se basa en un principio similar al utilizado al tratar el diagrama Q-Q[168]: dados los cuantiles de la variable de referencia ($Q_{0t}\left(\frac{t}{n+1}\right) = F_0^{-1}\left(\frac{t}{n+1}\right)$ para $t = 1,2,...,n$) y de los residuos ($Q_{\hat{\varepsilon}_t}\left(\frac{t}{n+1}\right) = \varepsilon_{(t)}$), la hipótesis de adecuacidad de la función de referencia como aproximación de la distribución de los residuos $\hat{\varepsilon}_t$ podrá ser aceptada, si se verifica que el estadístico W, definido como el coeficiente de correlación lineal:

$$W = \rho^2\left(\hat{Q}_{\hat{\varepsilon}_t}\left(\frac{t}{n+1}\right), Q_0\left(\frac{t}{n+1}\right)\right)$$

asume valores suficientemente próximos a la unidad.

A partir del estudio realizado mediantes simulaciones efectuadas utilizando el método de Monte Carlo, se concluye que el test es muy potente respecto de un conjunto amplio de distribuciones de referencia, en especial con respecto a hipótesis alternativas de "colas" cortas o muy largas, aún para muestras pequeñas si las razones para rechazar la hipótesis nula son muy significativas[169].

9.- El **test de Royston**[170] constituye una extensión del test de Shapiro-Wilk para contrastar la hipótesis de Normalidad de las perturbaciones ε_t cuando se cuenta con sucesiones de residuos de tamaño menor que 2000 y considera un estadístico de la forma $Z = \frac{w - \mu}{\sigma} : N(0;1)$, donde:

$$w = \ln\left(1 - w^*\right)$$

---

168. Este test fue propuesto originalmente por Shapiro, S.S.; Wilk, M.B. (1965) para contrastar la hipótesis de Normalidad, posteriormente (ver Shapiro, S.S.; Wilk, M.B. (1972)) fue generalizado, primero para el caso de la distribución exponencial y, posteriormente, para el resto de las distribuciones.

169. Ver Shapiro, S.S.; Wilk, M.B.; Chen, H.J. (1968), Pearson, E.S.; D'Agostino, R.B.; Browman, K.O. (1975), Seier, E. (2002).

170. Ver Royston, P. (1992).

$$W^* = \frac{\left(\sum_{t=1}^{n} \tilde{a}_t \hat{\varepsilon}_{(t)}\right)^2}{\sum_{t=1}^{n} \left(\hat{\varepsilon}_{(t)} - \bar{\hat{\varepsilon}}\right)^2}$$

$$\tilde{a}_t = \frac{\tilde{m}_t}{\sqrt{\Phi}} \qquad (t = 1,2,...,n-2; n > 5)$$

$$\tilde{m}_t = F_N^{-1}\left(\frac{t - \frac{3}{8}}{n + \frac{1}{4}}\right) \qquad (t = 1,2,...,n)$$

$$\Phi = \frac{\tilde{m}'\tilde{m} - 2\tilde{m}_n - 2\tilde{m}_{n-1}}{1 - 2\tilde{a}_n - 2\tilde{a}_{n-1}}$$

$$\tilde{m}' = \left[\tilde{m}_1 \quad \tilde{m}_2 \quad ... \quad \tilde{m}_n\right]$$

$$\tilde{a}_{n-1} = c_{n-1} + (0.042981)x - (0.293762)x^2 - (1.752461)x^3 + (5.682633)x^4 - (3.582663)x^5$$

$$\tilde{a}_n = c_n + (0.221157)x - (0.147981)x^2 - (2.07119)x^3 + (4.43685)x^4 - (2.70605)x^5$$

$$c = \frac{\tilde{m}}{\left(\tilde{m}'\hat{m}\right)^{1/2}}$$

$$\mu = -1.5861 - (0.31082)y - (0.083751)y^2 + (0.003915)y^3$$

$$\sigma = \exp\left[-0.4803 - (0.082676)y + (0.0030302)y^2\right]$$

$$y = \ln(n)$$

Se considera que, cuando $F_N(z) \geq 1 - \alpha$, existen razones suficientes para rechazar la hipótesis nula de Normalidad.

10.- Los **tests de Hegazy-Green**, los cuales consideran estadísticos de la forma[171]:

$$T_1 = \frac{1}{n}\sum_{t=1}^{n}\left|\varepsilon_{(t)} - m\left(\hat{\varepsilon}_{(t)}\right)\right|$$

$$T_2 = \frac{1}{n}\sum_{t=1}^{n}\left[\varepsilon_{(t)} - mo\left(\hat{\varepsilon}_{(t)}\right)\right]^2$$

Donde $m\left(\hat{\varepsilon}_{(t)}\right)$ y $mo\left(\hat{\varepsilon}_t\right)$ denotan, respectivamente, el momento absoluto de primer orden y el modo del estadístico de orden $\hat{\varepsilon}_{(t)}$.

A partir de un análisis respecto de las distribuciones de referencia Normal, de Laplace, exponencial y de Cauchy, se puede concluir: **i)** que la potencia de estos dos tests es similar y **ii)** que se obtienen resultados

---

171. Ver Hegazy, Y.A.S.; Green, J.R. (1975).

levemente mejores si los desvíos son calculados con respecto al $mo\left(\hat{\varepsilon}_{(t)}\right)$.

Como se mencionó al tratar los tests en forma individual, si la distribución de referencia contiene coeficientes estimados y los estimadores son asintóticamente eficientes, entonces los estadísticos conservan sus propiedades asintóticas pero, dado que las funciones de referencia poseen límites en-distribución distintos a los correspondientes a los casos en los que los coeficientes son conocidos, pierden la propiedad de ser independientes de la distribución de referencia. De lo expuesto en páginas anteriores se puede concluir que, en los casos de coeficientes desconocidos, las investigaciones estuvieron dirigidas fundamentalmente al diseño de tests para las distribuciones de referencia Normal y exponencial.

11.- El **test de Wilk-Francia**, el cual se basa en el principio que, si la hipótesis nula :

$$H_0: F_{\varepsilon_t}(\varepsilon) = F_0\left(\frac{\varepsilon - \theta_0}{\theta_1}\right)$$

(donde θ_0 y θ_1 denotan las medidas de posición y dispersión) es verdadera, los puntos del gráfico de los estadísticos de orden $\varepsilon_{(t)}$ respecto de la esperanza matemática de los estadísticos de orden normalizados, $m\left(\varepsilon_{(t)}\right)$, deben agruparse alrededor de la recta $\varepsilon_{(t)} = \theta_0 + \theta_1 m\left(\varepsilon_{(t)}\right)$

Los estimadores de los coeficientes θ_0 y θ_1, obtenidos de aplicar cuadrados mínimos ordinarios, son de la forma $\hat{\theta}_0 = \bar{\varepsilon}_{(t)}$ y $\hat{\theta}_1 = \dfrac{\sum_{t=1}^{n} \hat{\varepsilon}_{(t)}\, m\left(\hat{\varepsilon}_{(t)}\right)}{\sum_{t=1}^{n} m^2\left(\hat{\varepsilon}_{(t)}\right)}$. Luego, la correspondiente suma de los cuadrados de los residuos puede ser expresada de la siguiente forma[172]:

$$\sum_{t=1}^{n}\left(\hat{\varepsilon}_t^*\right)^2 = S_t^2 = \sum_{t=1}^{n}\left(\varepsilon_{(t)} - \hat{\varepsilon}_{(t)}\right)^2 = \sum_{t=1}^{n}\left[\varepsilon_{(t)} - \bar{\varepsilon} - \hat{\theta}_1 m\left(\varepsilon_{(t)}\right)\right]^2 = \sum_{t=1}^{n}\left(\varepsilon_{(t)} - \bar{\varepsilon}\right) - \left(\sum_{t=1}^{n} a_t\, \varepsilon_{(t)}\right)^2$$

(donde $a_t = \dfrac{m\left(\varepsilon_{(t)}\right)}{\sqrt{\sum_{t=1}^{n} m^2\left(\varepsilon_{(t)}\right)}}$). A fin de eliminar el efecto escala, si se divide m. a m. la expresión anterior por $\sum_{t=1}^{n}\left(\varepsilon_{(t)} - \bar{\varepsilon}\right)^2$, se obtiene que:

$$\frac{S_t^2}{\sum_{t=1}^{n}\left(\varepsilon_{(t)} - \bar{\varepsilon}\right)^2} = 1 - \frac{\left[\sum_{t=1}^{n} a_t \varepsilon_{(t)}\right]}{\sum_{t=1}^{n}\left(\varepsilon_{(t)} - \bar{\varepsilon}\right)^2} = 1 - W_{(n)}^*$$

donde $W_{(n)}^*$ denota el **estadístico de Wilk-Francia.**

---

[172]. El criterio de optimización conocido como de los cuadrados mínimos ordinarios (CMO) será tratado con detalle en el Cap. 6.

Este estadístico se utiliza habitualmente para decidir acerca de la aproximación Normal, pero puede utilizarse para testear otras distribuciones de referencia. Su distribución asintótica fue analizada por Durbin, J. (1974)[173].

Ejemplo 2.9:

Sea una serie cronológica $\{\hat{\varepsilon}_t\}$ de parámetro diario, obtenida de aplicar al proceso $\{Y_t\}$ un modelo $ARMA(1,1)$ y cuya trayectoria para el período [01-02-1995;20-05-2004] figura en el **GRAFICO 2.17**.

De la observación del correspondiente correlograma (ver **CUADRO 2.13**) se puede concluir que, con una confiabilidad del 95%, $\{\varepsilon_t\}$: $AR(0)$.

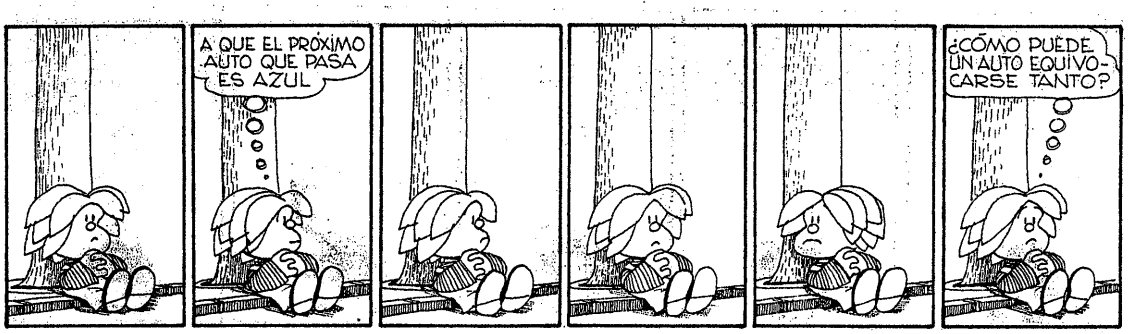
A partir de los datos proporcionados por la mencionada serie cronológica se obtuvieron los siguientes resultados acerca de la distribución de los residuos $\hat{\varepsilon}_t$ (ver **GRAFICO 2.18**): $\hat{As}(\hat{\varepsilon}) = 0.339$ y $\hat{K}(\hat{\varepsilon}) = 8.756$. y, por lo tanto, $JB = 3427.116$. La comparación de este valor con el correspondiente a la distribución $\chi^2_{(2)} = 5.991$ (para un $\alpha = 0.05$), permite concluir que no existen razones suficientes para aceptar la hipótesis de Normalidad.

El diagrama Q-Q (ver **GRAFICO 2.19**) permite concluir que la distribución de los residuos $\{\hat{\varepsilon}_t\}$ posee "colas" más largas que la distribución Normal

---

La condición de homocedasticidad (necesaria para poder garantizar la estacionariedad de segundo orden de un proceso) y los tests desarrollados para detectarla serán analizados al tratar la no-estacionariedad en la varianza, en la Sec. 3.1.

2.15.- La teoría de la predicción de Kolmogorov-Wiener para series cronológicas estacionarias



Sea $\{Y_t\}$ un proceso estocástico estacionario de segundo orden, discreto en el dominio del tiempo que admite una representación $ARMA(p,q)$ de la forma $\Phi(B)Y_t = \Theta(B)\varepsilon_t$ (donde: **i)** las variables Y_t son centradas y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma^2_\varepsilon)$), sea $\{Y_{t-n}, Y_{t-n+1}, ..., Y_t\}$ la serie cronológica formada por el conjunto de sus

---

[173]. Para una análisis de otros tests de Normalidad, ver Baltagi, B.H. (2001), Seier, E. (2002), Thode, H.C. (2002), Nguyen, T.T.; Dinh, K.T. (2003). Para contrastar la hipótesis nula que un proceso $\{\varepsilon_t\}$:*WN – Normal* tiene valor medio nulo, ver Casella, G.; Berger, R.L. (1990).

GRAFICO 2.17: Trayectoria del proceso $\{\hat{\varepsilon}_t\}$ para el período [02-01-1995;20-05-2004]

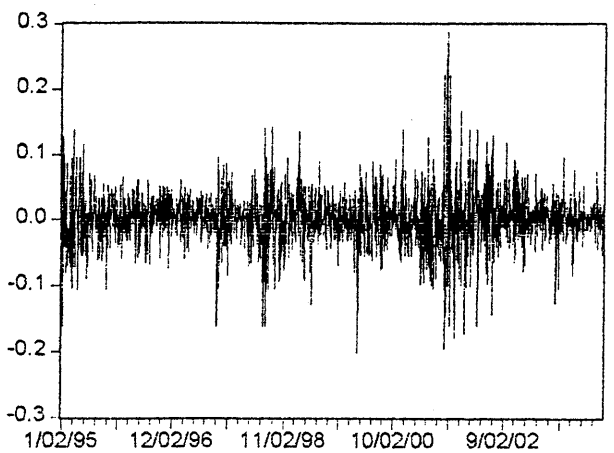
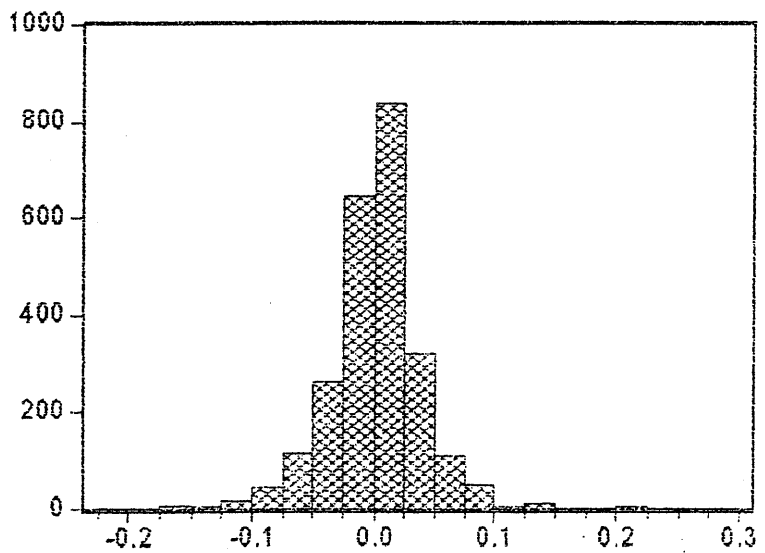
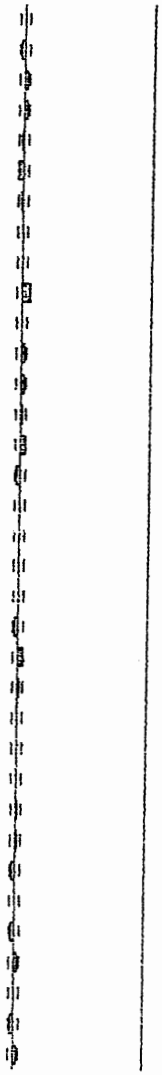


GRAFICO 2.18: Distribución de las observaciones correspondientes al proceso $\{\hat{\varepsilon}_t\}$ para el período [02-01-1995;20-05-2004]



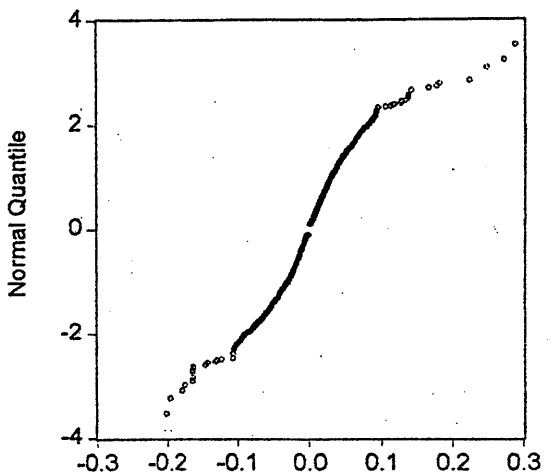
CUADRO 2.13: Correlograma correspondiente al proceso $\{\hat{\varepsilon}_t\}$

Sample: 1/02/1995 5/20/2004
Included observations: 2448



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.002	0.002	0.0088	0.925
		2	-0.021	-0.021	1.1093	0.574
		3	0.029	0.029	3.1056	0.376
		4	0.024	0.024	4.5401	0.338
		5	-0.016	-0.015	5.1487	0.398
		6	-0.039	-0.039	8.9402	0.177
		7	-0.014	-0.015	9.3899	0.226
		8	-0.001	-0.002	9.3906	0.310
		9	-0.008	-0.005	9.5340	0.390
		10	0.068	0.070	20.747	0.023
		11	0.003	0.002	20.766	0.036
		12	0.024	0.026	22.228	0.035
		13	0.026	0.022	23.933	0.032
		14	0.016	0.013	24.568	0.039
		15	0.034	0.035	27.408	0.026
		16	-0.017	-0.014	28.148	0.030
		17	0.006	0.009	28.247	0.042
		18	0.004	0.003	28.282	0.058
		19	0.000	0.003	28.282	0.078
		20	0.003	0.002	28.303	0.102
		21	-0.019	-0.017	29.216	0.109
		22	0.036	0.033	32.378	0.071
		23	0.012	0.008	32.743	0.086
		24	-0.006	-0.005	32.834	0.108
		25	-0.001	-0.008	32.838	0.135
		26	0.005	0.003	32.891	0.165
		27	0.015	0.011	33.442	0.183
		28	0.015	0.017	33.993	0.201
		29	-0.029	-0.028	36.066	0.172
		30	-0.002	-0.004	36.072	0.206
		31	-0.031	-0.032	38.527	0.166
		32	0.022	0.019	39.778	0.162
		33	0.000	0.001	39.778	0.194
		34	-0.018	-0.015	40.611	0.202
		35	0.028	0.026	42.596	0.177

GRAFICO 2.19: Diagrama Q-Q del proceso $\{\hat{\varepsilon}_t\}$ según observaciones correspondientes al período [02-01-1995;20-05-2004]



realizaciones y sea $f^{(t,\ell)}(Y_t, Y_{t-1}, \ldots, Y_{t-n})$ $(\ell \geq 1)$ la **función de predicción** (o **predictor**)[173] del proceso $\{Y_t\}$ basada en el mencionado conjunto de observaciones[174].

Supóngase que se desea predecir el valor $Y_{t+\ell}$ $(\ell \geq 1)$. El problema consiste en la selección de la función de predicción que haga que el valor estimado resulte lo más aproximado que sea posible a $Y_{t+\ell}$. Usualmente, el criterio de aproximación utilizado es el del error medio cuadrático, según el cual la función $f^{(t,\ell)}$ elegida debe ser un estimador insesgado y tal que minimice la distancia L_2, es decir, la **varianza del error de la predicción** para ℓ períodos ($e(\ell,t)$):

$$\sigma^2_{(\ell,t)}(e) = E\left\{\left[Y_{t+\ell} - f^{(t,\ell)}(Y_t, Y_{t-1}, \ldots, Y_{t-n})\right]^2\right\}$$

(debe tenerse en cuenta que el futuro ya no es lo que era).

Sea $f^{(t,\ell)}$ el predictor definido como la esperanza matemática de $Y_{t+\ell}$ condicionada por el conjunto de realizaciones $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$ y sea $f^{(t,\ell)*}$ un predictor distinto de $f^{(t,\ell)}$, tal que $f^{(t,\ell)*} - f^{(t,\ell)} = \eta(\ell,t)$. La varianza del error de predicción asociado a $f^{(t,\ell)*}$, para ℓ períodos ($e*(\ell,t)$) será:

$$\sigma^2_{(\ell,t)}(e*) = E\left\{\left[Y_{t+\ell} - f^{(t,\ell)}(Y_t, Y_{t-1}, \ldots, Y_{t-n})\right]^2\right\} =$$

---

173. Ver Kolmogorov, A.N. (1941), Wiener, N. (1949).

174. Esta función de predicción es equivalente a la función complementaria de la representación truncada analizada en la Sec. 2.12.1.

$$= E\left\{\left[Y_{t+\ell} - f^{(t,\ell)} - \eta(\ell,t)\right]^2\right\} =$$

$$= E\left\{\left[\left(Y_{t+\ell} - f^{(t,\ell)}\right) - \left(f^{(t,\ell)*} - f^{(t,\ell)}\right)\right]^2\right\} = E\left\{\left[\left(Y_{t+\ell} - f^{(t,\ell)}\right) - \eta(\ell,t)\right]^2\right\} =$$

$$= E\left[\left(Y_{t+\ell} - f^{(t,\ell)}\right)^2\right] + \eta^2(\ell,t) - 2\eta(\ell,t)\,E\left(Y_{t+\ell} - f^{(t,\ell)}\right)$$

De acuerdo con la definición dada para $f^{(t,\ell)}$, el tercer sumando del segundo miembro de esta expresión es nulo, de modo que el valor mínimo a asumir por la varianza se verificará cuando $\eta(\ell,t) = 0$. Por lo que se puede concluir que, para que se cumpla la condición de varianza mínima del error de predicción, el predictor debe ser de la forma $f^{(t,\ell)} = E\left(Y_{t+\ell} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right)$. Este resultado es aplicable sólo si se conoce la función de probabilidades conjunta de la variable multidimensional $\left\{Y_{t+\ell}, Y_t, Y_{t-1}, \ldots, Y_{t-n}\right\}$ o si, como en este caso, se cuenta con una representación para $\left\{Y_t\right\}$ que involucre un número finito de coeficientes, a partir de la cual los valores esperados condicionados pueden ser calculadas en forma recursiva [177].

Teniendo en cuenta que:

$$E\left(Y_{t+\ell} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) =$$

$$= \phi_1 E\left(Y_{t+\ell-1} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) + \phi_2 E\left(Y_{t+\ell-2} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) + \ldots +$$

$$+ \phi_p E\left(Y_{t+\ell-p} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) +$$

$$+ E\left(\varepsilon_{t+\ell} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) - \theta_1 E\left(\varepsilon_{t+\ell-1} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) -$$

$$- \ldots - \theta_q E\left(\varepsilon_{t+\ell-q} / Y_y, Y_{t-1}, \ldots, Y_{t-n}\right)$$

y que [178]:

$$E\left(\varepsilon_{t+j} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right) = \begin{cases} 0 & para \quad j > 0 \\ \varepsilon_{t+j} & para \quad j \le 0 \end{cases}$$

se puede concluir que el valor esperado condicionado es lineal en $\left\{Y_t, Y_{t-1}, \ldots, Y_{t-n}\right\}$:

---

[177]. En otros términos, se puede asegurar que $E\left(Y_{t+\ell} / Y_t, \ldots, Y_{t-n}\right)$ es la variable aleatoria en $L^2\left(Y_t, \ldots, Y_{t-n}\right)$ (es decir, tal que $E\left[\left(Y_{t+\ell} / Y_t, \ldots, Y_{t-n}\right)^2\right] < \infty$) más próxima a $Y_{t+\ell}$ en error medio cuadrático, $E\left\{\left[Y_{t+\ell} - E\left(Y_{t+\ell} / Y_t, \ldots, Y_{t-n}\right)\right]^2\right\} = \min_{f^{(t,\ell)} \in L^2\left(Y_t, \ldots, Y_{t-n}\right)} E\left\{\left[Y_{t+\ell} - f^{(t,\ell)}\right]^2\right\}$. Una demostración alternativa de esta propiedad se obtuvo en la Sec. 2.12.1, al tratar la representación truncada de un proceso $ARMA(p,q)$.

[178]. Si el proceso $\left\{\varepsilon_t\right\}$ es un WN-débil para procesos no-lineales que admitan una aproximación lineal, podría suceder que $E\left(Y_{t+\ell} / Y_t, Y_{t-1}, \ldots, Y_{t-n}\right)$ no coincida con el predictor $f^{(t,\ell)}$.

$$f^{(t,\ell)} = a_0 Y_t + a_1 Y_{t-1} + ... + a_n Y_{t-n}$$

Es decir que el problema de la predicción queda reducido a la definición de los métodos analíticos para determinar los coeficientes $a_0, a_1, ..., a_n$ que minimicen $\sigma^2_{(\ell,t)}(e)$.

Supóngase que el proceso $\{Y_t\}$ además de ser estacionario, tenga espectro continuo y función de densidad espectral $\omega(\alpha)$, que satisfaga la condición $\int_{-\infty}^{\infty} \ln[\omega(\alpha)] d\alpha > -\infty$ y que, por lo tanto, pueda ser expresada de la siguiente forma: $\omega(\alpha) = b(e^{-i\alpha}) b(e^{i\alpha})$, donde $b(z) = \sum_{j=0}^{\infty} b_j z^j$ es una función que no posee raíces para $|z| < 1$. Entonces, de acuerdo a lo demostrado en la Sec.2.1.6.1 $\{Y_t\}$ admite una representación de la forma:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \Psi(B)\varepsilon_t \quad (\psi_0 = 1)$$

donde $\{\varepsilon_t\}$ denota un proceso WN-débil con varianza $\sigma^2(\varepsilon_t) = \sigma^2_\varepsilon = 2\pi$.

Suponiendo un conjunto infinito de realizaciones del proceso $\{Y_t\}$ ($\{Y_t, Y_{t-1}, ...\}$) se puede escribir:

$$Y_{t+\ell} = \sum_{j=0}^{\infty} b_j \varepsilon_{t+\ell-j} = \sum_{j=0}^{\ell-1} b_j \varepsilon_{t+\ell-j} + \sum_{j=\ell}^{\infty} b_j \varepsilon_{t+\ell-j} =$$

$$= \sum_{j=0}^{\ell-1} b_j \varepsilon_{t+\ell-j} + \sum_{j=0}^{\infty} a_j Y_{t-j} =$$

$$= e(\ell,t) + a(B) Y_t = e(\ell,t) + f^{(t,\ell)}(Y_t, Y_{t-1}, ...)$$

(donde: **i)** $a(B)$ denota la función generatriz del predictor; **ii)** $e(\ell,t)$ -que involucra a "shocks" futuros- representa la parte no-predecible de Y_t o el error de la predicción $f^{(t,\ell)}$ con origen en el punto t, y **iii)** $f^{(t,\ell)}$ -que involucra a valores presentes y pasados de ε_t - representa la parte predecible de Y_t). Este resultado constituye la llamada **aproximación de Kolmogorov**. Kolmogorov, A.N. (1931) propuso una solución independiente al problema de la predicción lineal para procesos estacionarios, basada en una representación de la varianza de la predicción en el dominio de las frecuencias de la forma $\sigma^2_\ell(\varepsilon) = \int_{-\pi}^{\pi} \left| e^{it\alpha} \Psi(e^{-it\alpha}) - \alpha(e^{-it\alpha} \Psi(e^{-it\alpha})) \right|^2 d\alpha$.

Dado que, para $\ell > 0$, $E(\varepsilon_{t+\ell} / Y_t, Y_{t-1}, ...) = 0$ y, por lo tanto, $E(e(\ell,t)) = 0$, se obtiene que:

$$E(Y_{t+\ell} / Y_t, Y_{t-1}, ...) = f^{(t,\ell)} = \sum_{j=\ell}^{\infty} b_j \varepsilon_{t+\ell-j} = \sum_{j=0}^{\infty} b_{j+\ell} \varepsilon_{t-j} = b^{(\ell)}(B)\varepsilon_t$$

donde $b^{(\ell)}(B) = \sum_{j=0}^{\infty} b_{j+\ell} B^j$. Lo que demuestra que la predicción es insesgada. La varianza del error de predicción

está dada por[179]:

$$\sigma^2_{(\ell,t)}(e) = E\left[\left(Y_{t+\ell} - f^{(t,\ell)}\right)^2\right] = E\left\{\left[\left(\sum_{j=0}^{\ell-1} b_j \varepsilon_{t+\ell-j} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) - \sum_{j=0}^{\infty} b_{j+\ell}\varepsilon_{t-j}\right]^2\right\} =$$

$$= E\left\{\left[\sum_{j=0}^{\ell-1} b_j \varepsilon_{t+\ell-j} + \sum_{j=0}^{\infty}\left(\psi_j - b_{j+\ell}\right)\varepsilon_{t-j}\right]^2\right\} =$$

$$= \sigma^2_{\varepsilon}\left[\sum_{j=0}^{\ell-1} b_j^2 + \sum_{j=0}^{\infty}\left(\psi_j - b_{j+\ell}\right)^2\right] \qquad (\psi_0 = 1)$$

Esta expresión -en la que el primer término es independiente de $\{\psi_j\}$- se minimiza cuando $\psi_j = b_{j+\ell}\ (j = 1,2,...)$, es decir cuando $b_{j+\ell}$ es una función de los coeficientes $\phi_j\ (j = 1,2,...,p)$ y $\theta_j\ (j = 1,2,...,q)$ de la representación $ARMA(p,q)$. En este caso la varianza de la predicción para ℓ períodos asume la forma:

$$\sigma^2_{(\ell,t)}(e) = E\left[\left(Y_{t+\ell} - f^{(t,\ell)}\right)^2\right] = \sigma^2_{\varepsilon}\sum_{j=0}^{\ell-1} b_j^2$$

A partir de estos resultados y suponiendo la distribución Normal de las perturbaciones ε_t, se puede construir el siguiente intervalo de confiabilidad:

$$p\left[f^{(t,\ell)} - z_{\alpha/2}\sigma_{(\ell,t)}(e) \le Y_{t+\ell} \le f^{(t,\ell)} + z_{\alpha/2}\sigma_{(\ell,t)}(e)\right] = 1 - \alpha$$

Supóngase, en particular, que el proceso $\{Y_t\}$ admita una representación $AR(1)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

donde: **i)** el coeficiente ϕ_1 es conocido; **ii)** Y_t denota una variable centrada y **iii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma^2_{\varepsilon}\right)$.

Entonces, utilizando un algoritmo recursivo, se puede escribir:

$$Y_{t+1} = \phi_1 Y_t + \varepsilon_{t+1}$$

$$Y_{t+2} = \phi_1 Y_{t+1} + \varepsilon_{t+2} = \phi_1\left(\phi_1 Y_t + \varepsilon_{t+1}\right) + \varepsilon_{t+2} = \phi_1^2 Y_t + \left(\varepsilon_{t+2} + \phi_1\varepsilon_{t+1}\right) = \phi_1^2 Y_t + e(1,t)$$

---

[179]. A partir de una función $\Psi^*(B)$, que admite un desarrollo de Laurent de la forma $\Psi^*(B) = \sum_{j=-\infty}^{\infty} \psi_j^* B^j$, se pueden definir las funciones $\left[\Psi^*(B)\right]_+ = \sum_{j=0}^{\infty} \psi_j^* B^j$ y $\left[\Psi^*(B)\right]_- = \sum_{j=-\infty}^{-1} \psi_j^* B^j$. Entonces, se puede escribir $b^{(\ell)}(B) = B^{-\ell}\Psi(B)$. Expresión de la cual se obtiene la siguiente forma alternativa para $a(B) = \dfrac{\left[B^{-\ell}\Psi(B)\right]_+}{\Psi(B)}$. Ver Whittle, P. (1962), Yaglom, A.M. (1962).

$$Y_{t+3} = \phi_1 Y_{t+2} + \varepsilon_{t+3} = \phi_1\left(\phi_1^2 Y_t + \varepsilon_{t+2} + \phi_1\varepsilon_{t+1}\right) + \varepsilon_{t+3} = \phi_1^3 Y_t + \left(\varepsilon_{t+3} + \phi_1\varepsilon_{t+2} + \phi_1^2\varepsilon_{t+1}\right) = \phi_1^3 Y_t + e(2,t)$$

.....

$$Y_{t+\ell} = \phi_1^{\ell} Y_t + \sum_{j=0}^{\ell-1}\phi_1^j \varepsilon_{t+\ell-j} = \phi_1^{\ell} Y_t + e(\ell,t) \qquad (\ell = 1,2,...)$$

La función de predicción para $t+1$ será:

$$f^{(t,1)} = E\left(Y_{t+1} / Y_t\right) = \phi_1 Y_t$$

De la misma forma:

$$f^{(t,2)} = \phi_1 E\left(Y_{t+2} / Y_t\right) = \phi_1^2 Y_t = \phi_1 E\left(Y_{t+1} / Y_t\right)$$

.....

$$f^{(t,\ell)} = \phi_1^{\ell} Y_t = \phi_1 E\left(Y_{t+\ell-1} / Y_t\right) \qquad (\ell = 1,2,...)$$

El error de predicción será:

$$e(\ell,t) = Y_{t+\ell} - f^{(t,\ell)} = \sum_{j=0}^{\ell-1}\phi_1^j \varepsilon_{t+\ell-j} \qquad (\ell = 1,2,...)$$

y $\sigma^2_{(\ell,t)}(e) = \left(\sum_{j=0}^{\ell-1}\phi_1^{2j}\right)\sigma_\varepsilon^2$ $(\ell = 1,2,...)$.

Suponiendo una distribución Normal de los ε_t, el estimador de $Y_{t+\ell}$, $f^{(t,\ell)} = \phi_1^{\ell} Y_t$, se distribuye de acuerdo con una función $N\left(Y_{t+\ell}; \sigma^2_{(\ell,t)}(e)\right)$. Obsérvese que, de acuerdo con la propiedad de **retorno al valor esperado**, cuando $\ell \to \infty$, $f^{(t,\ell)}$ converge a la esperanza matemática no-condicionada $E(Y_t) = 0$. En este caso particular de un proceso $AR(1)$, la velocidad de retorno al valor esperado está medida por el coeficiente conocido como **vida media**, definido por $\ell^* = \ln\left(2|\phi_1|\right)^{-1}$. Asimismo, se puede concluir fácilmente que la varianza del error de predicción tiende a la varianza no-condicionada de Y_t:

$$\lim_{\ell\to\infty}\sigma^2_{(\ell,t)}(e) = \left(\sum_{j=0}^{\infty}\phi_1^{2j}\right)\sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\phi_1^2} = \sigma_Y^2$$

Supóngase que el proceso $\{Y_t\}$ admita una representación $MA(q)$ $(q = 1,2,...)$ de la forma:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-q}$$

donde: **i)** los coeficientes θ_j $(j = 1,2,...,q)$ son conocidos; **ii)** Y_t denota una variable centrada y **iii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$. Entonces, será:

$$Y_{t+1} = f^{(t,1)} + e(1,t) = \left(-\theta_1\varepsilon_t - \theta_2\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q+1}\right) + \varepsilon_{t+1}$$

$$Y_{t+2} = f^{(t,2)} + e(2,t) = \left(-\theta_2\varepsilon_t - \theta_3\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q+2}\right) + \left(\varepsilon_{t+2} - \theta_1\varepsilon_{t+1}\right)$$

.....

$$Y_{t+q} = f^{(t,q)} + e(q,t) = -\theta_q\varepsilon_t + \sum_{j=0}^{q-1}\theta_j\varepsilon_{t+q-j} \qquad (\theta_0 = 1)$$

$$Y_{t+\ell} = e(\ell,t) = \varepsilon_{t+\ell} - \theta_1\varepsilon_{t+\ell-1} - \ldots - \theta_q\varepsilon_{t+\ell-q} \qquad (\ell = q+1, q+2, \ldots)$$

La función de predicción será, entonces, de la forma:

$$f^{(t,\ell)} = E\left(Y_{t+\ell} / Y_t\right) = \begin{cases} -\theta_\ell\varepsilon_t - \theta_{\ell+1}\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q+\ell} & (\ell = 1,2,\ldots,q) \\ 0 & (\ell = q+1, q+2, \ldots) \end{cases}$$

y el error de predicción será:

$$e(\ell,t) = \begin{cases} \sum_{j=0}^{\ell-1}\theta_j\varepsilon_{t+\ell-j} & (\theta_0 = 1; \ell = 1,2,\ldots,q) \\ \sum_{j=0}^{q}\theta_j\varepsilon_{t+\ell-j} & (\theta_0 = 1; \ell = q+1, q+2, \ldots) \end{cases}$$

De modo que:

$$\sigma^2_{(\ell,t)}(e) = \begin{cases} \sigma^2_\varepsilon\left(1 + \theta_1^2 + \ldots + \theta^2_{\ell-1}\right) \\ \sigma^2_\varepsilon\left(1 + \theta_1^2 + \ldots + \theta^2_q\right) \end{cases}$$

De los resultados anteriores se puede concluir que: **i)** para $\ell > q$, la función de predicción coincide con la esperanza matemática no-condicionada de $\{Y_t\}$ (en otros términos, que el proceso demora q períodos en retornar al valor esperado); **ii)** para $\ell > q$, la varianza del error de predicción coincide con la varianza no-condicionada de $\{Y_t\}$.

Supóngase ahora que el proceso $\{Y_t\}$ admita una representación $ARMA(1,1)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

donde: **i)** los coeficientes ϕ_1 y θ_1 son conocidos; **ii)** Y_t denota una variable centrada y **iii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma^2_\varepsilon\right)$.

Entonces será:

$$Y_{t+1} = \phi_1 Y_t + \varepsilon_{t+1} - \theta_1 \varepsilon_t = \left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \varepsilon_{t+1}$$

$$Y_{t+2} = \phi_1 Y_{t+1} + \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1} = \phi_1\left[\left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \varepsilon_{t+1}\right] + \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1} =$$

$$= \phi_1\left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \left[\varepsilon_{t+2} + \left(\phi_1 - \theta_1\right)\varepsilon_{t+1}\right]$$

$$Y_{t+3} = \phi_1 Y_{t+2} + \varepsilon_{t+3} - \theta_1 \varepsilon_{t+2} =$$

$$= \phi_1\left\{\phi_1\left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \left[\varepsilon_{t+2} + \left(\phi_1 - \theta_1\right)\varepsilon_{t+1}\right]\right\} + \varepsilon_{t+3} - \theta_1 \varepsilon_{t+2} =$$

$$= \phi_1^2\left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \left[\varepsilon_{t+3} + \left(\phi_1 - \theta_1\right)\varepsilon_{t+2} + \phi_1\left(\phi_1 - \theta_1\right)\varepsilon_{t+1}\right]$$

.....

$$Y_{t+\ell} = \phi_1^{\ell-1}\left(\phi_1 Y_t - \theta_1 \varepsilon_t\right) + \left[\varepsilon_{t+\ell} + \left(\phi_1 - \theta_1\right)\sum_{j=1}^{\ell-1}\phi_1^{j-1}\varepsilon_{t+\ell-j}\right] \qquad (\ell = 1,2,...)$$

La función de predicción será:

$$f^{(t,\ell)} = E\left(Y_{t+\ell} / Y_t\right) = \phi_1^{\ell} Y_t - \phi_1^{\ell-1}\theta_1 \varepsilon_t = \phi_1 E\left(Y_{t+\ell} / Y_t\right) \qquad (\ell = 1,2,...)$$

Luego, el error de predicción será:

$$e(\ell,t) = \varepsilon_{t+\ell} + \left(\phi_1 - \theta_1\right)\sum_{j=1}^{\ell-1}\phi_1^{j-1}\varepsilon_{t+\ell-j} \qquad (\ell = 1,2,...)$$

y $\sigma^2_{(\ell,t)}(e) = \sigma_\varepsilon^2 + \left(\phi_1 - \theta_1\right)^2 \sigma_\varepsilon^2 \sum_{j=1}^{\ell-1}\phi_1^{2(j-1)}$.

Obsérvese que esta varianza tiende a la varianza no-condicionada de $\{Y_t\}$:

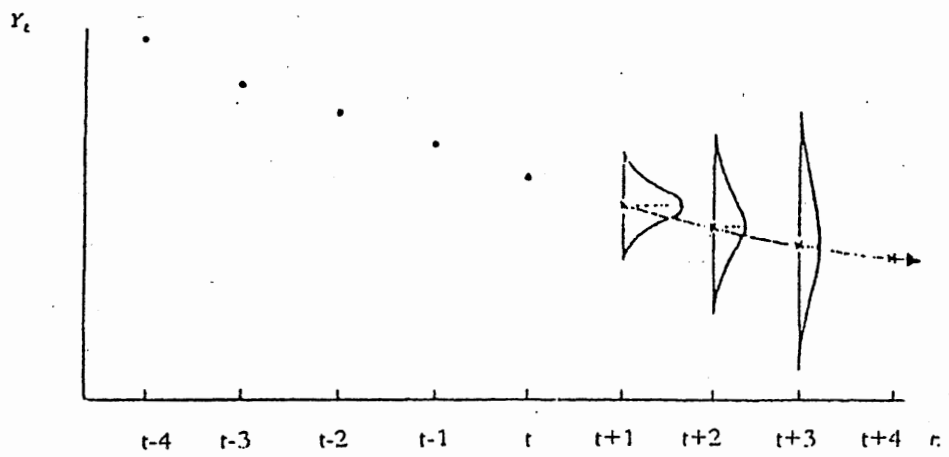
$$\lim_{\ell\to\infty}\sigma^2_{(\ell,t)}(e) = \left[1 + \frac{\left(\phi_1 - \theta_1\right)^2}{1 - \phi_1^2}\right]\sigma_\varepsilon^2 = \frac{\left(1 - \theta_1\right)^2}{1 - \phi_1^2}\sigma_\varepsilon^2 = \sigma_Y^2$$

El error de predicción para el período siguiente al de la última observación ($e(1,t)$) debe ser incorrelacionado con los "shocks" aleatorios $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. El no cumplimiento de esta condición significaría que el error de predicción podría ser predicho -aunque sea en parte- por los "shocks" $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ y, en consecuencia $f^{(t,1)} + e(1,t)$ resultaría mejor predictor de Y_{t+1} que $f^{(t,1)}$.

Aún en el caso en que los errores de predicción óptimos para un período sean incorrelacionados, los errores de predicción para períodos mayores serán, en general, correlacionados. Supónganse, por ejemplo, dos predicciones para ℓ períodos, con origen en los momentos t y $t - j\ (j = 1,2,...)$. Si $j = \ell, \ell+1, \ell+2, \ldots$, los errores de predicción no contienen elementos comunes, pero si $j = 1,2,..., \ell - 1$, existirán valores de ε_t incluidos en ambas predicciones:

$$e(\ell,t) = Y_{t+\ell} - f^{(t,\ell)} = \varepsilon_{t+\ell} + \psi_1 \varepsilon_{t+\ell-1} + ... + \psi_{\ell-1}\varepsilon_{t+1}$$

GRAFICO 2.20.- Distribuciones de probabilidades de las variables $Y_{t+\ell}$ condicionadas por el conjunto de información hasta el período t



$$e(\ell,t-j) = Y_{t+\ell-j} - f^{(t-j,\ell)} = \varepsilon_{t-l+\ell} + \psi_1\varepsilon_{t-l+\ell-1} + ,,, + \psi_{\ell-1}\varepsilon_{t-j+1}$$

De modo que, para $j < \ell$, se verifica que $\gamma\left[e(\ell,t-j),e(\ell,t)\right] = \sigma_\varepsilon^2 \sum_{i=j}^{\ell-1} \psi_i\psi_{i-j}$. Por lo tanto, la correspondiente función de autocorrelaciones es de la forma:

$$\rho\left[e(\ell,t-j),e(\ell,t)\right] = \begin{cases} \dfrac{\sum_{i=j}^{\ell-1}\psi_i\psi_{i-j}}{\sum_{i=0}^{\ell-1}\psi_i^2} & (0 \le j < \ell) \\ 0 & (j \ge \ell) \end{cases}$$

Supónganse, asimismo, dos predicciones para los momentos ℓ y $\ell + j\,(j = 1,2,...)$, realizadas con el mismo origen t. Los errores de predicción serán, respectivamente:

$$e(\ell,t) = Y_{t+\ell} - f^{(t,\ell)} = \varepsilon_{t+\ell} + \psi_1\varepsilon_{t+\ell-1} + ... + \psi_{\ell-1}\varepsilon_{t+1}$$

$$e(\ell+j,t) = Y_{t+\ell+j} - f^{(t,\ell+j)} = \varepsilon_{t+\ell+j} + \psi_1\varepsilon_{t+\ell+j-1} + ... + \psi_j\varepsilon_{t+\ell} + \psi_{j+1}\varepsilon_{t+\ell-1} + ... + \psi_{\ell+j-1}\varepsilon_{t+1}$$

De modo que la función de autocovarianzas queda definida por $\gamma\left[e(\ell,t),e(\ell+j,t)\right] = \sigma_\varepsilon^2 \sum_{j=0}^{\ell-1} b_i b_{j+i}$ y la correspondiente función de autocorrelaciones será de la forma $\rho\left[e(\ell,t),e(\ell+j,t)\right] = \dfrac{\sum_{i=0}^{\ell-1} b_i b_{j+i}}{\sqrt{\left(\sum_{i=0}^{\ell-1} b_i^2\right)\left(\sum_{h=0}^{\ell-j-1} b_h^2\right)}}$. Si se supone que las variables ε_t tienen distribución Normal, entonces, se puede asegurar que:

$$f^{(t,\ell)}: N\left[E\left(Y_{t+\ell} / Y_t, Y_{t-1}, \ldots\right); \sigma_\varepsilon^2 \sum_{j=0}^{\ell-1} b_j^2 \right]$$

Luego, dado un conjunto de información disponible al momento t, se puede asegurar que:

$$p\left[f^{(t,\ell)} - z_{\alpha/2}\sigma_\varepsilon \sqrt{\sum_{j=0}^{\ell-1} b_j^2} \leq Y_{t+\ell} \leq f^{(t,\ell)} + z_{\alpha/2}\sigma_\varepsilon \sqrt{\sum_{j=0}^{\ell-1} b_j^2} \right] = 1 - \alpha$$

(ver **GRAFICO 2.20**).

CAPITULO 3

Los procesos no-estacionarios de parámetro discreto

"El tiempo es un problema para nosotros, un tembloroso y exigente problema, acaso el más vital de la metafísica: la eternidad, un juego o una fatigada esperanza".

(Borges, J.L.: "Historia de la eternidad")

3.1.- No-estacionariedad en la varianza

De acuerdo a lo expresado en la Sec.1.3, los procesos pueden presentar comportamientos no-estacionarios en los momentos de primer orden o en las varianzas o en ambas características a la vez. En el caso de los procesos no-estacionarios en la varianza, la solución en el ámbito de la dinámica lineal consiste en aplicar transformaciones que conviertan al proceso original en un proceso condicionadamente homocedástico.

3.1.1.- Las transformaciones de potencia

Según se vio en los capítulos anteriores, las técnicas relacionadas con la teoría de los modelos lineales que aquí se analizan han sido desarrolladas en base a los siguientes supuestos: **i)** que la estructura de su esperanza matemática condicionada $E\left(Y_t / \Omega\left(Y_{t-1}\right),\Omega\left(Y_{t-2}\right),...\right)$ es aditiva (es decir que las características relevantes del proceso pueden ser descriptas a partir de una combinación lineal de factores de su filtración); **ii)** que la varianza de las variables que componen el proceso es constante (es decir, que la varianza del proceso es independiente del valor esperado de sus variables); **iii)** que las variables Y_t están Normalmente distribuidas.

Para el caso en que una o más de estas condiciones no se verifique[1], existen transformaciones relacionadas con cada una de las propiedades a analizar (ver Tukey, J.W. (1958)). Alternativamente, existen otras transformaciones (ver Box, G.E.P.; Cox, D.R. (1964) y Draper, N.R.; Hunter, W.G. (1969)) que, en principio, intentan actuar sobre las tres propiedades simultáneamente[2] y que se utilizan fundamentalmente para lograr la estabilidad en el tiempo de la varianza condicionada, la simetría y, en

---

[1]. Debe tenerse en cuenta que, como se analizó en la Sec. 2.1.4.2, aún en los casos de presencia de linealidad en la variable original, existen transformaciones que pueden ayudar a aproximar el cumplimiento de las hipótesis de estabilidad de la varianza y de Normalidad de la distribución, propiedades que constituyen el fundamento de las técnicas estadísticas relacionadas con la teoría de los modelos lineales (debe recordarse que estas hipótesis, por ejemplo, son necesarias y suficientes para poder testear, en forma relativamente simple, la significatividad individual de los coeficientes de la representación correspondiente).

[2]. Anscombe, F.J. (1948)(1955)(1961), Elston, R.E. (1948), Moore, P.G.; Tukey, J.W. (1954), Anscombe, F.J.; Tukey, J.W. (1963) desarrollaron, a este respecto, una clase de transformaciones (conocidas como **de potencia**) del tipo $T(Y) = Y^{\lambda}$ y, como se verá inmediatamente, sus trabajos se refieren fundamentalmente a la estimación del coeficiente λ.

lo posible, la Normalidad de la distribución de las variables Y_t.

Sea un proceso $\{Y_t\}$ que admite una descomposición de la forma:

$$Y_t = m(Y_t) + \varepsilon_t^* = \sum_{j=0}^{d} a_j t^j + \varepsilon_t^*$$

en la que $m(Y_t)$ denota una **tendencia determinística** o **estacionaria** y la varianza de las perturbaciones es una función de dicho nivel promedio del proceso[3]:

$$\sigma^2(Y_t) = \sigma^2(\varepsilon_t^*) = h[m(Y_t)]\sigma_\varepsilon^2$$

donde $h(\bullet)$ denota una función conocida[4]. La cuestión radica, entonces, en la definición de una transformación $T(Y_t)$ de la forma:

$$T(Y_t) = \left\{ \int [h(v)]^{-1/2} dv \right\}_{v=Y}$$

tal que $\sigma^2(T(Y_t)) = \sigma_Y^2$.

Tomando en consideración el término de primer orden del desarrollo de Taylor de la definición de la transformación en el entorno de $m(Y_t)$, se obtiene que:

$$T(Y_t) \approx T(m(Y_t)) + [Y_t - m(Y_t)]T'(m(Y_t))$$

(donde $T'(m(Y_t))$ denota el valor de la primera derivada de la función $T(Y_t)$ en el punto $m(Y_t)$).

La varianza de la variable transformada puede ser aproximada entonces de la siguiente forma[5]:

$$\sigma^2[T(Y_t)] \approx \sigma^2\{T(m(Y_t)) + [Y_t - m(Y_t)]T'(m(Y_t))\} = [T'(m(Y_t))]^2 \sigma^2(Y_t) =$$

$$= [T'(m(Y_t))]^2 h[m(Y_t)]\sigma_\varepsilon^2$$

---

[3]. En la Sec.5.2, al analizar las representaciones estáticas de procesos multidimensionales estacionarios, se tratará el problema de la estructura de la heterocedasticidad como función de las variaciones de las variables explicativas del modelo.

[4]. Tradicionalmente se ha considerado a la heterocedasticidad como un fenómeno asociado a las aplicaciones conocidas como de corte transversal y a la presencia de autocorrelación en las perturbaciones como un factor asociado a los fenómenos dinámicos. Fue Engle, R.F. (1982) quien advirtió sobre la posibilidad de existencia de heterocedasticidad en series cronológicas y consideró la posibilidad de una representación autorregresiva de la varianza condicionada de las perturbaciones (condicionada por la información con que se cuenta al momento $t-1$) de la forma $\sigma^2(\varepsilon_t^* / \varepsilon_{t-1}^*,\ldots,\varepsilon_{t-p}^*) = E(\varepsilon_t^{*2} / \varepsilon_{t-1}^*,\ldots,\varepsilon_{t-p}^*) = \phi_0 + \phi_1\varepsilon_{t-1}^{*2}+\ldots+\phi_p\varepsilon_{t-p}^{*2}$. Esta representación de la varianza condicionada surge asociada a un proceso $\{\varepsilon_t^*\}$ que admite una representación de la forma $\varepsilon_t^* = \varepsilon_t[\phi_0 + \phi_1\varepsilon_{t-1}^{*2}+\ldots+\phi_p\varepsilon_{t-p}^{*2}]^{1/2}$ (donde $\{\varepsilon_t\}$:$WN - N(0;1)$ con varianza unitaria), conocida como $ARCH(p)$ (estos procesos serán tratados con detalle en la Parte 3).

[5]. Ver Kendall, M.G.; Stuart, S. (1976).

Luego, para lograr la estabilidad de la varianza, se debe verificar que $T'\left(m\left(Y_t\right)\right) = \dfrac{1}{\sqrt{h\left(m\left(Y_t\right)\right)}}$.

Supóngase, por ejemplo, que $\sqrt{h\left(m\left(Y_t\right)\right)} = m\left(Y_t\right)$. Entonces, se debe verificar que $T'\left(m\left(Y_t\right)\right) =$ $= \dfrac{1}{m\left(Y_t\right)}$, lo cual implica que la transformación debe ser de la forma $T\left(m\left(Y_t\right)\right) = \ln\left(m\left(Y_t\right)\right)$.

De la misma forma, si se verifica que la dependencia de la varianza del nivel del proceso Y_t es de la forma $\sqrt{h\left(m\left(Y_t\right)\right)} = \sqrt{m\left(Y_t\right)}$, entonces la transformación debe ser tal que $T'\left(m\left(Y_t\right)\right) = \left[m\left(Y_t\right)\right]^{-1/2}$, es decir, se debe verificar que $T\left(m\left(Y_t\right)\right) = 2\sqrt{m\left(Y_t\right)}$ [6].

Estos resultados constituyen dos casos particulares de la familia de **transformaciones de potencia de Box-Cox**[7]definida, en general, de la siguiente forma:

$$T^{(\lambda)}\left(Y_t\right) = \begin{cases} \dfrac{Y_t^{\lambda} - 1}{\lambda} & (0 < \lambda \leq 1) \\ \ln\left(Y_t\right) & (\lambda = 0) \\ -Y_t & (\lambda < 0) \end{cases}$$

(donde λ es el coeficiente que caracteriza la transformación). En todos los casos debe verificarse que $Y_t > 0$ (el cambio de signo cuando $\lambda < 0$ es para asegurar que la serie transformada conserve el mismo orden relativo que la serie original)[8].

Por su parte, Tukey, J.W. (1957)(1958) definió otra familia de transformaciones de potencia, alternativa a la de Box-Cox, de la forma:

$$T^{*(\lambda)}\left(Y_t\right) = \left(Y_t + c\right)^{\lambda}$$

(donde c es una constante tal que $c + Y_t > 0$, para $t = 1,2,\ldots$). La función $T^{*(\lambda)}\left(m\left(Y_t\right)\right)$ cumple, entonces, la condición:

---

[6].Igual que con las demás propiedades, en lo que hace a la heterocedasticidad condicionada, resulta imposible definir una transformación que asegure la estabilidad completa de la varianza. Lo máximo que se puede pretender, en términos prácticos, es una transformación tal que $\sigma^2\left[T\left(Y_t\right)\right] = c\left[1 + o\left(k^{-1}\right)\right]$ (donde k denota una constante conocida, lo suficientemente grande como para que k^{-1} sea despreciable).

[7]. Ver Box, G.E.P.; Cox, D.R. (1964).

[8]. Como se verá en la Sec. 5.1, en las representaciones estáticas esta propiedad consiste en considerar que la esperanza matemática condicionada $E\left(Y_t / X_{t1}, X_{t2}, \ldots, X_{tk}\right)$ puede ser representada como una función lineal de los regresores. En estos modelos, a fin de obtener el modelo de regresión más simple que sea posible, es posible transformar tanto a la variable dependiente como a los regresores.

$$\frac{dT^{*(\lambda)}(Y_t)}{d^2T^{*(\lambda)}(Y_t)} = \frac{c + Y_t}{\lambda - 1}$$

En particular, para $\lambda = 0$, se obtiene que $\frac{dT^{*(\lambda)}(Y_t)}{d^2T^{*(\lambda)}(Y_t)} = -(c + Y_t)$. Es decir, que $\frac{d}{dY_t}\left\{\ln\left[T^{*(\lambda)}(Y_t)\right]\right\} =$ $= 0$. Ecuación de la que resulta la transformación $T^{*(\lambda)}(Y_t) = b_0 \ln(c + Y_t) + b_1$. De la misma forma, para $\lambda = \pm\infty$, se obtiene que $\frac{d}{dY_t}\left[\frac{dT^{*(\lambda)}(Y_t)}{d^2T^{*(\lambda)}(Y_t)}\right] = 0$. Por lo tanto, será $\frac{d}{dY_t}T^{*(\lambda)}(Y_y) = b_0 e^{b_1 Y_t}$. De modo que la transformación queda definida de la siguiente forma:

$$T^{*(\lambda)} = \frac{b_0}{b_1} e^{b_1 Y_t} + b_2$$

Otras formas típicas de la **transformación de Tukey** son:

$$T^{*(\lambda)}(Y_t) = \begin{cases} Y_t & para \quad \lambda = 1 \\ Y_t^2 & para \quad \lambda = \frac{1}{2} \\ \frac{1}{Y_t} & para \quad \lambda = -1 \\ \sqrt{Y_t} & para \quad \lambda = 2 \end{cases}$$

En cuanto a la transformación logarítmica, si bien es la más ampliamente aplicada para reducir la asimetría de la distribución (en particular en series económicas), no resulta muy apropiada cuando las variables Y_t son acotadas inferior y superiormente. En particular, si el proceso $\{Y_t\}$ sólo puede asumir valores en el intervalo $[0;1]$, Wallis, K.F. (1987) propuso como alternativa la **transformación logística** $T(Y_t) = \ln\left(\frac{Y_t}{1 - Y_t}\right)$. En general, cuando el dominio de las variables aleatorias Y_t incluye solamente valores no-negativos, es decir cuando poseen una cota inferior pero no una cota superior, se genera una asimetría hacia la derecha en su distribución. En este caso la más adecuada parece ser la utilización de una transformación con $\lambda < 1$, para lograr una retracción en la "cola" de la derecha y una expansión en la "cola" de la izquierda. Para valores de $\lambda > 1$ se obtiene el efecto contrario, es decir, una disminución de la asimetría hacia la izquierda.

La transformación $T(Y_t) = Y_t^{-1}$ es particularmente apropiada en los casos en que la distribución de Y_t es fuertemente asimétrica (por ejemplo, cuando la probabilidad de que $Y_t \leq 0$ es despreciable) y la transformación $T(Y_t) = Y_t^{1/2}$ es particularmente apropiada en los casos en que $Y_t : \wp(\theta)$. Por su parte Anscombe, F.J. (1948), utilizando la **familia de transformaciones de Johnson**[9] demostró que, en este caso de variables distribuidas de acuerdo con una función de Poisson, $T(Y_t) = \sqrt{Y_t + \frac{3}{8}}$ es una transformación óptima para eliminar la heterocedasticidad, obteniéndose los siguientes resultados:

---

[9]. Ver Sec. 2.14.2.

$$\sigma^2\left(T(Y_t)\right) \approx \frac{1}{4}\left(1+\frac{1}{16\theta^2}\right)$$

$$As\left(T(Y_t)\right) \approx -\frac{1}{2\sqrt{\theta}}\left(1+\frac{7}{16\theta}\right)$$

$$K\left(T(Y_t)\right) \approx \frac{1}{\theta}\left(1+\frac{369}{256\,\theta}\right)$$

Fisher, R.A. (1925) demostró que la transformación $T(Y_t) = \sqrt{2Y_t}$ es particularmente útil en el caso en que Y_t tenga una distribución $\chi^2_{(r)}$, obteniéndose una serie transformada $T(Y_t)$ distribuida aproximadamente de acuerdo a una función $N\left(\sqrt{2r-1};1\right)$.

3.1.2.- La estimación del coeficiente λ

Si se conociera la distribución de probabilidades de la variable Y_t, la validez del resultado de aplicar una transformación del tipo de las analizadas en la sección precedente podría ser testeada. Por el contrario, si -como ocurre habitualmente- la información disponible fuera de carácter exclusivamente muestral y no se conociera la distribución de Y_t, ni la definición de la función $h\left(m(Y_t)\right)$, entonces, como se verá inmediatamente, la forma de la transformación debe ser determinada empíricamente a partir de la estimación del coeficiente λ [10].

Para la estimación de los coeficientes λ en las transformaciones dirigidas a la eliminación de la heterocedasticidad se suele utilizar la conocida como **relación rango-mediana**, que se obtiene de dividir la serie cronológica original $\{Y_t, t = 1,2,...,n\}$ en k subseries de n_k observaciones cada una (en general, se recomienda que $4 \le n_k \le 12$) y calcular para cada una de ellas el rango y la mediana[11]. en el **GRAFICO 3.1** figuran algunos comportamientos típicos de esta relación y los correspondientes valores de λ que pueden ser utilizados en la transformación de Box-Cox. Los gráficos **a)** y **b)** corresponden a los valores $\lambda = 1$ y $\lambda = 0$, respectivamente. En el **GRAFICO c)** figuran los comportamientos (de izquierda a derecha) correspondientes a los valores $\lambda = -\frac{1}{2}$, $\lambda = -\frac{1}{3}$ y $\lambda = -\frac{1}{4}$ (en realidad, la finalidad de esta relación rango-mediana no es de terminar el valor de λ en forma precisa, sino distinguir en forma grosera si, por ejemplo, λ puede ser igual a $-\frac{1}{2}, 0, \frac{1}{2}$ ó 1) . De la misma forma, en el **GRAFICO d)** figuran los comportamientos (de izquierda a derecha) correspondientes a los valores $\lambda = \frac{1}{2}$, $\lambda = \frac{1}{3}$ y $\lambda = \frac{1}{4}$.

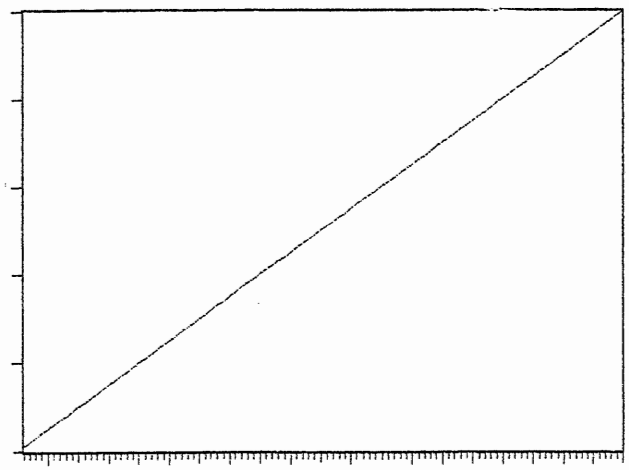
Wilson, E.B.;Hilferty, M.M. (1931) (ver también Garwood, F. (1936), Haldane, J.B.S. (1937), Merrington, M. (1941)) demostraron que la transformación $T(Y_t) = \left(\frac{Y_t}{r}\right)^{1/3}$ es particularmente útil en el caso

---

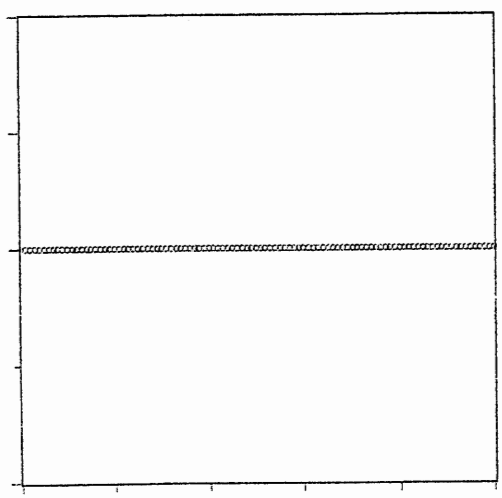
[10] . Ver Fisher, R.A.; Mather, K. (1943), Dolby, J.I. 19(63), Quenouille, M.H. (1965).

[11] . Alternativamente puede calcularse la **relación rango-valor medio**. La relación rango-mediana es preferible a la **relación varianza-valor medio** (en realidad, **ln(varianza)- ln(valor medio)**) propuesta por Bartlett, M.S. (1947)) porque asegura la robustez del estimador.

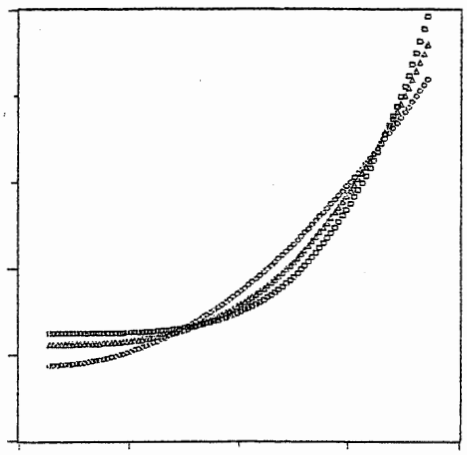
GRAFICO 3.1: Diagrama rango-mediana



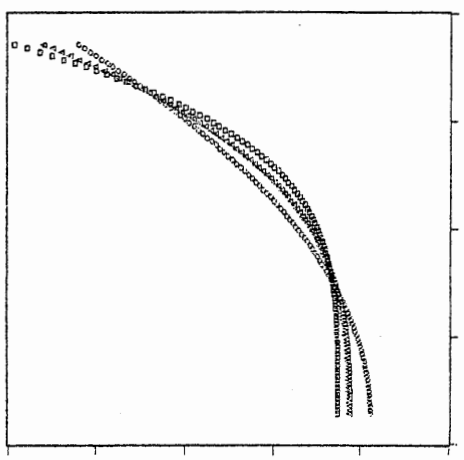
(a)



(b)



(c)



(d)

útil en el caso en que $Y_t : \chi^2_{(r)}$, concluyendo que $T(Y_t)$) se distribuye, aproximadamente, de acuerdo a una función $N\left(1-\frac{2r}{9};\frac{2r}{9}\right)$

Ejemplo 3.1:

Sea un proceso $\{Y_t\}$ de parámetro diario, cuya trayectoria para el período [01-01-2005;10-04-2005] (ver **GRAFICO 3.2**) sugiere la presencia de una tendencia determinística lineal.

A partir de la división de esta serie cronológica en 10 subseries de longitud 10 y del cálculo, para cada una de ellas, del rango y la mediana, se obtuvieron los siguientes resultados:

j	me(Y(j))	R(Y(j))
1	4.781	8.371
2	7.735	8.822
3	13.652	8.599
4	18.021	17.521
5	22.638	13.284
6	28.237	13.275
7	33.539	22.925
8	32.311	27.693
9	41.999	23.400
10	42.992	23.411

De la observación del diagrama de dispersión correspondiente a estos datos (ver **GRAFICO 3.3**) se podría concluir que existe heterocedasticidad y que el valor de λ más adecuado para estabilizar la varianza de $\{Y_t\}$ es $-\frac{1}{2}$.

Realizando un análisis similar a partir de la serie transformada $T(Y_t) = Y_t^{-1/2}$, se obtuvieron los siguientes valores del rango y la mediana para las 10 subseries:

j	me((T(Y$_j$))	R((T(Y$_j$))	j	me((T(Y$_j$))	R((T(Y$_j$))
1	0,457	0,367	6	0,188	0,043
2	0,360	0,211	7	0,173	0,067
3	0,271	0,117	8	0,176	0,071
4	0,236	0,109	9	0,154	0,053
5	0,210	0,066	10	0,153	0,042

GRAFICO 3.2: Trayectoria del proceso $\{Y_t\}$ en el período [01-01-2005;10-04-2005]

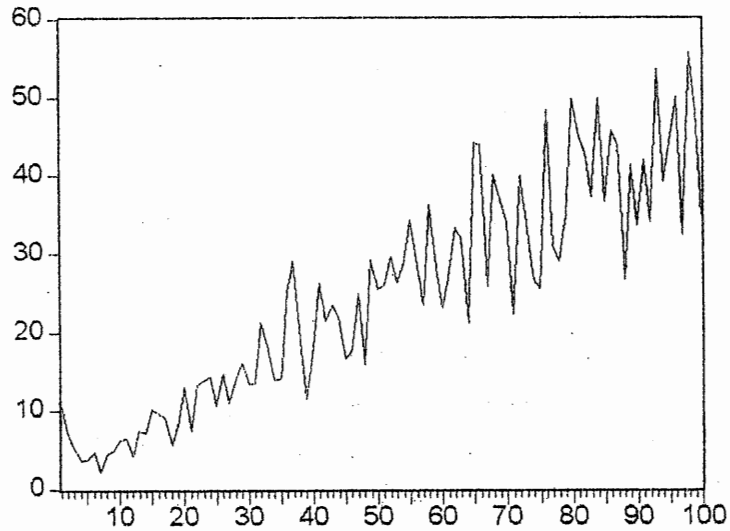


GRAFICO 3.3: Diagrama de dispersión correspondiente a la mediana y el ran de las subseries obtenidas de la serie cronológica $\{Y_t, t=1,2,...,100\}$

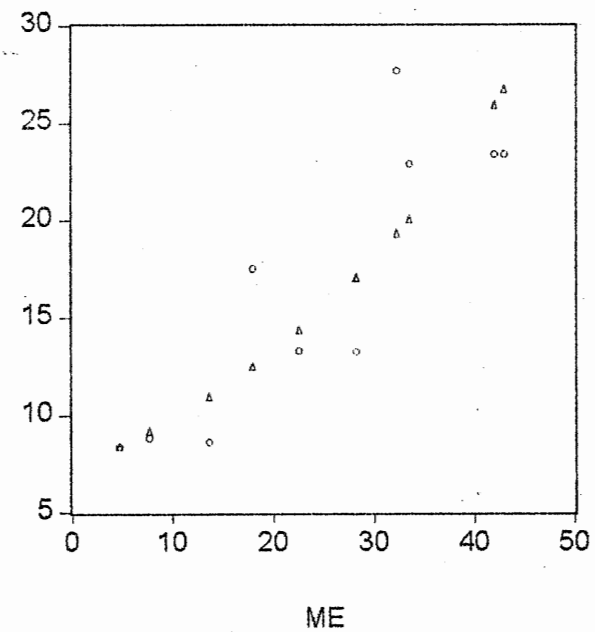
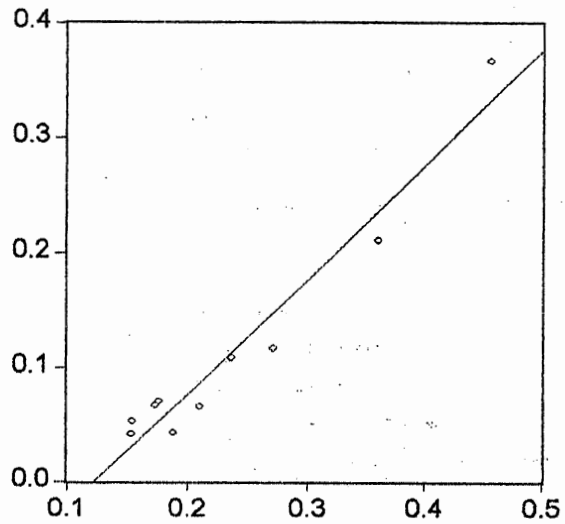


GRAFICO 3.4: Diagrama de dispersión correspondiente a la mediana y el rango de las subseries obtenidas de la serie cronológica $\{T(Y_t),t=1,2,...,100\}$



Del diagrama de dispersión correspondiente a estos datos (ver **GRAFICO 3.4**) se puede concluir que la serie transformada es aproximadamente homocedástica.

3.1.3.- Algunas transformaciones alternativas a las de potencia

Como alternativa a las transformaciones de potencia analizadas precedentemente merecen mencionarse:

i) La **transformación de Granger-Hughes**[12] :

$$T(Y_t) = \frac{Y_t}{\frac{1}{2m+1}\sum_{j=-m}^{m} Y_{t-j}}$$

la cual resulta particularmente apropiada en los casos en que el valor medio y la varianza de Y_t varían con la misma tendencia.

ii) La **transformación arco seno**, $T(Y_t) = arc\ sen(Y_t)$, propuesta por Fisher, R.A. (1922)(1930), la cual resulta particularmente apropiada en el caso que Y_t:$b(n,p)$. Una alternativa de características similares a la de Fisher, para eliminar la heterocedasticidad en una variable binomial, es la **transformación angular**:

$$T(Y_t) = arc\ sen\sqrt{\frac{Y_t + w_1}{n + w_2}}$$

En particular, Anscombe, F.J. (1948) sugirió, para casos como éste, la transformación

---

[12] . Ver Granger, C.W.J.; Hughes, A.D. (1971).

$$T(Y_t) = arc\ sen\sqrt{\frac{Y_t + \frac{3}{8}}{n + \frac{3}{4}}}$$

a partir de la cual se obtienen los siguientes resultados[13]:

$$\sigma^2\left(T(Y_t)\right) = \frac{1}{4} + o\left(\frac{1}{n^2}\right)$$

$$As\left(T(Y_t)\right) \approx \frac{2m(Y_t) - n}{2\sqrt{n\, m(Y_t)\left[n - m(Y_t)\right]}}$$

$$K\left(T(Y_t)\right) \approx \frac{n^2 - 2\, m(Y_t)\left[n - m(Y_t)\right]}{nm(Y_t)\left[n - m(Y_t)\right]}$$

iii) La **transformación hiperbólica**:

$$T(Y_t) = arc\ sh\sqrt{\frac{Y_t}{Y_t - n}} \qquad \left(Y_t = n, n+1, \ldots\right)$$

propuesta por Beall, G. (1942) para eliminar la heterocedasticidad en el caso en que $Y_t : b^{-1}(n, p)$. En particular, Anscombe, F.J. (1948) propuso una variante de esta transformación de la forma[14]:

$$T(Y_t) = arc\ sh\sqrt{\frac{Y_t + \frac{3}{8}}{(Y_t - n) - \frac{3}{4}}} \qquad \left(Y_t = n, n+1, \ldots\right)$$

En la Parte 3 de este trabajo, en la que se considerarán los procesos no-lineales, se analizarán modelos que consideran, explícitamente, casos de heterocedasticidad en los que la función $h(\bullet)$ asume la forma general $h\left(\varepsilon^*_{t-1}, \ldots, \varepsilon^*_{t-p}, Y_t, Y_{t-1}, \ldots, Y_{t-p}, X, \delta\right)$, donde $\varepsilon^*_t = Y_t - f\left(Y_{t-j}, X, \delta\right)$.

3.2.- No-estacionariedad en el valor medio

3.2.1.- El método de la diferenciación sucesiva

La forma típica de no-estacionariedad en el valor medio se caracteriza por la presencia de **tendencia estocástica** en el comportamiento del proceso, que se origina en el no-cumplimiento de las condiciones de estacionariedad referidas a las raíces de la ecuación característica de su componente autorregresiva.

---

[13] . Ver también Blom, G. (1954), Anscombe, F.J. (1956).

[14] . La literatura reconoce 19 transformaciones, de las cuales las que se han comentado aquí son las que aparecen más frecuentemente en las aplicaciones. Entre las menos habituales en las series económicas merecen mencionarse la log-log, la función de distribución, la función de distribución inversa, el desvío equivalente, la probit, la logit, la legit, los "scores" Normal y exponencial y la monótona. Algunas de estas transformaciones son casos particulares de otras y cada una está vinculada con los supuestos de Normalidad y homocedasticidad condicionada (ver Hoyle, M.H. (1973)).

Un caso es aquél en el que las raíces de la ecuación característica de la componente AR son, en módulo, menores que la unidad. En esta circunstancia se dice que el proceso $\{Y_t\}$ es **explosivo**. Otro caso de no-estacionariedad de este tipo es aquél en el que algunas o todas las raíces de la ecuación característica de la componente AR son, en módulo, iguales a la unidad[15].

Sea un proceso $\{Y_t\}$ $(t = 1,2,...)$ que admite una representación de la forma:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_{p+d} Y_{t-p-d} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

$$\left(1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_{p+d} B^{p+d}\right) Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q\right) \varepsilon_t$$

$$\varphi(B) Y_t = \Theta(B) \varepsilon_t$$

(donde $\varphi(B)$ denota un operador autorregresivo no-estacionario, tal que d de sus raíces son iguales a la unidad y las p restantes son, en módulo, mayores que la unidad). El operador AR puede ser, entonces, expresado de la siguiente forma:

$$\varphi(B) = (1 - B)^d \left(1 - G_1^{-1} B\right)\left(1 - G_2^{-1} B\right)...\left(1 - G_p^{-1} B\right) = (1 - B)^d \Phi(B)$$

(donde $\Phi(B)$ denota un operador autorregresivo estacionario de orden p).

El modelo ARMA original puede ser expresado, entonces, de acuerdo con una ecuación en diferencias de la forma:

$$\varphi(B) Y_t = \Phi(B)(1 - B)^d Y_t = \Phi(B)\left(\nabla^d Y_t\right) = \Phi(B) Z_t = \Theta(B) \varepsilon_t$$

(donde $Z_t = \nabla^d Y_t = \sum_{j=0}^{d} (-1)^j \binom{d}{j} B^j Y_t$ denota un proceso estacionario). De modo que, aplicando la transformación inversa se puede escribir [16]:

$$Y_t = \nabla^{-d} Z_t = S^d Z_t$$

Es decir que, el proceso original puede ser expresado como la sumación (integración) d veces de un proceso $ARMA(p,q)$ estacionario. Por esta razón se dice que $\{Y_t\}$ es un proceso integrado de orden d, $I(d)$, (de acuerdo con esta nomenclatura, un proceso $\{Y_t\}$ estacionario de segundo orden puede ser denotado como $\{Y_t\}:I(0)$; es decir, se puede concluir que, si $\{Y_t\}:I(d)$, entonces $\left\{(1 - B)^d Y_t\right\} =$ $= \left\{\nabla^d Y_t\right\} = \{Z_t\}:I(0)$), y la correspondiente familia de representaciones lineales generales que constituye la aproximación discreta a su comportamiento es la de las **autorregresivas integradas de promedios móviles de orden** (p,d,q) ($ARIMA(p,d,q)$).

---

[15]. Box, G.E.P.; Jenkins, G.M. (1994) denominaron a este tipo de no-estacionariedad como **no-estacionariedad homogénea**.

[16]. Debe tenerse en cuenta que si el proceso $\{Z_t\}$ es estacionario de segundo orden con variables centradas, entonces los procesos $\left\{\nabla^d Z_t\right\}$ $(d = 1,2,3,...)$ también serán estacionarios con variables centradas pero, como se verá en forma inmediata, no-invertibles y no-parsimoniosos.

Debe tenerse en cuenta que si, como se mencionó en la Sec. 3.1.1, el nivel del proceso $\{Y_t\}$ ($m(Y_t)$), presenta una tendencia determinística representable por un polinomio de grado d en t, entonces, de acuerdo con la extensión de Cramér del teorema de descomposición de Wold, $\{Y_t\}$ admite una representación de la forma[17]:

$$Y_t = m(Y_t) + \varepsilon_t^* = \sum_{j=0}^{d} a_j t^j + \Psi(B)\varepsilon_t = \sum_{j=0}^{d} a_j t^j + \frac{\Theta(B)}{\Phi(B)}\varepsilon_t$$

(donde: **i)** los a_j denotan coeficientes constantes en el tiempo y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$). Teniendo en cuenta que $E(\varepsilon_t^*) = \Psi(B)E(\varepsilon_t) = 0$, la estructura temporal del proceso será de la forma $E(Y_t) = m(Y_t) = = \sum_{j=0}^{d} a_j t^j$ y $\sigma^2(Y_t) = \sigma_\varepsilon^2$. Entonces diferenciando d veces el proceso $\{Y_t\}$:

$$\nabla^d Y_t = Z_t = \frac{\nabla^d \Theta(B)}{\Phi(B)}\varepsilon_t = \frac{\Theta(B)}{\Phi(B)}\left[(1-B)^d \varepsilon_t\right]$$

se obtiene un proceso $\{Z_t\}$ no-invertible. Luego, se puede concluir que la eliminación de la tendencia estacionaria mediante la diferenciación sucesiva del proceso $\{Y_t\}$ no es aconsejable, ya que genera no-invertibilidad. La solución se obtiene calculando los desvíos de $\{Y_t\}$ respecto de la función $f(t) = \sum_{j=0}^{d} a_j t^j$ estimada por cuadrados mínimos ordinarios[18].

Según se vio en la Sec. 2.12.2, la función de autocorrelaciones de un proceso $ARMA(p,q)$ satisface una ecuación en-diferencias de la forma:

$$\Phi(B)\rho_j(Y) = \rho_j(Y) - \phi_1\rho_{j-1}(Y) - \phi_2\rho_{j-2}(Y) - \ldots - \phi_p\rho_{j-p}(Y) \qquad (j = q+1, q+2, \ldots)$$

cuya solución (suponiendo que las raíces sean todas distintas entre sí), es de la forma:

$$\rho_j(Y) = A_1\left(G_1^{-1}\right)^j + A_2\left(G_2^{-1}\right)^j + \ldots + A_p\left(G_p^{-1}\right)^j \qquad (j = q+1, q+2, \ldots)$$

Esta expresión permite confirmar el decrecimiento exponencial de la función de autocorrelaciones, si se verifica la condición de estacionariedad. Si, por el contrario, alguna de las raíces (por ejemplo, G_i) se aproxima a la unidad: $G_i = 1+\varepsilon$, entonces, para valores grandes de j, la función $\rho_j(Y)$ puede ser aproximada de la siguiente forma: $\rho_j(Y) \approx A_i(1 - j\varepsilon)$. Lo cual demuestra un decrecimiento lineal de la función de autocorrelaciones. Una circunstancia similar se produce si hay más de una raíz próxima a la unidad.

---

[17]. Ver Cramér, H. (1945).

[18]. El método de estimación basado en el criterio de los cuadrados mínimos será tratado con detalle en el Cap. 6 (ver también Landro, A.H. (2002)).

De acuerdo con estas consideraciones, una metodología primaria para detectar la no-estacionariedad de un proceso podría consistir en el análisis de la forma de decrecimiento de su función de autocorrelaciones. Si de la observación de dicha función se concluye que el proceso $\{Y_t\}$ es no-estacionario, se toma en consideración la función de autocorrelaciones de su primera diferencia, $\{\nabla Y_t\}$, si este proceso también resulta no-estacionario, se toma la diferencia segunda, $\{\nabla^2 Y_t\}$, y así sucesivamente, hasta obtener una función de autocorrelaciones exponencialmente decreciente (en las aplicaciones el orden de diferenciación difícilmente debería ser mayor que dos)[19].

A pesar de su aparente simplicidad, la aplicación de este método suele traer aparejado el **problema de la sobre-diferenciación** que puede provocar serios inconvenientes: Sea, por ejemplo, un proceso $\{Y_t\}$ que admite una representación $MA(1)$ invertible de la forma:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 B)\varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$). A su primera diferencia le corresponde, entonces, una representación de la forma:

$$\nabla Y_t = (1 - B)Y_t = (1 - B)(1 - \theta_1 B)\varepsilon_t = \left[1 - (1 + \theta_1)B + \theta_1 B^2\right]\varepsilon_t = (1 - \theta_1^* B - \theta_2^* B^2)\varepsilon_t$$

Nótese que esta representación es más "compleja" que la original, en la medida que involucra a dos coeficientes, θ_1^* y θ_2^*, en vez de uno. Por otra parte, como se mencionó en páginas anteriores, dado que $\theta_1^* + \theta_2^* = 1$, por lo menos una de las raíces de la ecuación característica:

$$\Theta^*(B) = 1 - \theta_1^* B - \theta_2^* B^2 = 0$$

será igual a la unidad, por lo que se puede asegurar que esta representación es no-invertible (es decir, que su correspondiente representación $AR(\infty)$ estacionaria no existe).

Por otra parte, la varianza de Y_t está dada por $\sigma_Y^2 = (1 + \theta_1^2)\sigma_\varepsilon^2$, en tanto que la varianza de ∇Y_t es de la forma $\sigma_{\nabla Y}^2 = \left[1 + (1 + \theta_1)^2 + \theta_1^2\right]\sigma_\varepsilon^2 = 2(1 + \theta_1 + \theta_1^2)\sigma_\varepsilon^2$. Lo que demuestra que la varianza del proceso sobre-diferenciado es mayor que la del proceso original:

$$\sigma_{\nabla Y}^2 - \sigma_Y^2 = (1 + \theta_1)^2 \sigma_\varepsilon^2 > 0$$

Como un corolario de este resultado, se puede concluir que el comportamiento de las varianzas de las series diferenciadas sucesivamente pueden proporcionar un método útil para determinar el orden de diferenciación adecuado: las varianzas muestrales decrecerán hasta que el proceso alcance el estado estacionario (**"variate difference method"**)[20], luego, si se produce sobre-diferenciación tenderán a crecer.

Denotando por $\omega_Z(\alpha)$ a la función de densidad espectral del proceso $\{Z_t\} = \{\nabla^d Y_t\}$:$I(0)$, la

---

[19]. Debe tenerse en cuenta que no siempre las funciones de autocorrelaciones observan exactamente los comportamientos que se han supuesto aquí. A veces esta función decreceen forma más rápida que la lineal, pero no exponencialmente. Como se verá en el Cap. 5, un caso interesante es el del decrecimiento hiperbólico de la función de autocorrelaciones que da origen a representaciones (conocidas como **ARFIMA**) con un orden de diferenciación fraccionario.

[20]. Ver Tintner, G. (1940), Landro, A.H.; González, M.L. (1997).

función de densidad espectral de $\{Y_t\}$ puede ser expresada de la siguiente forma:

$$\omega_Y(\alpha,d) = \frac{1}{\left|1-e^{-i\alpha}\right|^{2d}}\omega_Z(\alpha) = \frac{1}{\left|1-e^{-i\alpha}\right|^{2d}}\left|\Psi\left(e^{-i\alpha}\right)\right|^2\frac{\sigma_\varepsilon^2}{2\pi} = \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{\left[4sen^2\left(\frac{\alpha}{2}\right)\right]^d} \quad (\alpha \neq 0)$$

(Nótese que al diferenciar una vez el proceso $\{Y_t\}$ su función de densidad espectral queda multiplicada por $\left|1-e^{-i\alpha}\right|^2 = \ = 2[1-\cos(\alpha)]$. En general, dado un proceso $\{Y_t\}$ con función de densidad espectral $\omega_Y(\alpha) = c[1-\cos(\alpha)]^{-d}$ (donde c denota una constante positiva), se verifica que $\{\nabla^d Y_t\}$:*WN*).

Si se verifica que $\{Z_t\}$:$I(0)$, entonces la función de densidad espectral en la frecuencia $\alpha = 0$ es acotada, $\lim_{\alpha\to 0}\omega(\alpha) = c > 0$. Si se verifica que $c = 0$, entonces se puede concluir que el proceso está sobrediferenciado[21].

Para $d < 0$, $\omega_Y(0,d) = 0$ y, para $d > 0, \omega_Y(0,d) = \infty$. Para valores de α próximos a cero y $d > 0$, se verifica que:

$$\omega_Y(\alpha) \approx \frac{\sigma_\varepsilon^2}{2\pi}\left[\Psi(1)\right]^2 2\alpha^{-2d}$$

Ejemplo 3.2:

Sea un proceso $\{Y_t\}$ cuya trayectoria -que no observa un comportamiento estacionario- figura en el **GRAFICO 3.5**. Consecuentemente el correlograma (ver **CUADRO 3.1**) observa un decrecimiento lineal que confirma su no-estacionariedad. Por su parte, en el **CUADRO 3.2** figura el correlograma correspondiente al proceso $\{Z_t\} = \{\nabla Y_t\}$.

Del análisis de las funciones de autocorrelaciones y autocorrelaciones parciales se puede concluir que el proceso $\{Z_t\}$ admite una representación $AR(1)$ de la forma:

$$Z_t = \phi_1 Z_{t-1} + \varepsilon_t$$

obteniéndose el siguiente modelo:

$$\hat{Z}_t = (0.4)Z_{t-1}$$

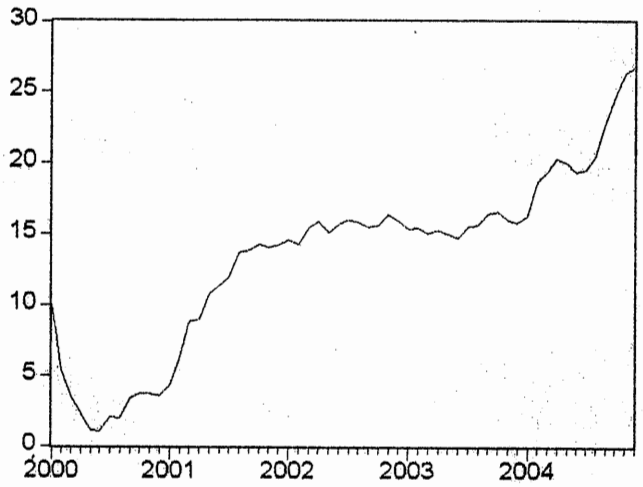
que cumple las condiciones de estacionariedad.

Teniendo en cuenta que $\{Z_t\} = \{\nabla Y_t\}$, se puede escribir:

---

[21]. Obsérvese que esta sobrediferenciación elimina del proceso $\{Y_t\}$ la componente de las bajas frecuencias, la cual es muy importante en las predicciones de largo plazo. Como se verá en el Cap. 5, existe una clase de procesos -muy importante en las aplicaciones económicas- que no son estacionarios, pero en los cuales la primera diferenciación es excesiva. Estos procesos, conocidos como **ARIMA fraccionalmente integrados (ARFIMA)**, se caracterizan por que $0 < d < 1$, lo cual implica la propiedad de atracción por su valor medio, incluso en el intervalo $\frac{1}{2} \leq d < 1$, en el que el proceso es no-estacionario.

GRAFICO 3.5: Trayectoria del proceso $\{Y_t\}$



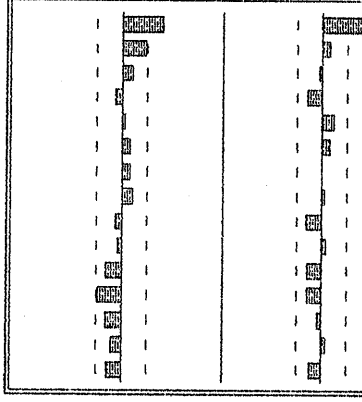
CUADRO 3.1: Correlograma correspondiente al proceso $\{Y_t\}$

Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.944	0.944	56.230	0.000
		2	0.870	-0.200	104.80	0.000
		3	0.793	-0.043	145.83	0.000
		4	0.719	-0.002	180.21	0.000
		5	0.656	0.046	209.34	0.000
		6	0.595	-0.052	233.72	0.000
		7	0.531	-0.064	253.55	0.000
		8	0.460	-0.109	268.66	0.000
		9	0.383	-0.068	279.35	0.000
		10	0.311	0.014	286.57	0.000
		11	0.243	-0.049	291.06	0.000
		12	0.191	0.085	293.90	0.000
		13	0.149	-0.005	295.66	0.000
		14	0.111	-0.006	296.66	0.000
		15	0.076	-0.020	297.13	0.000

CUADRO 3.2: Correlograma correspondiente al proceso $\{Z_t\}$

Sample: 2000:01 2004:12
Included observations: 59



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.410	0.410	10.448	0.001
		2	0.242	0.089	14.152	0.001
		3	0.100	-0.032	14.797	0.002
		4	-0.072	-0.147	15.135	0.004
		5	0.027	0.120	15.184	0.010
		6	0.077	0.089	15.585	0.016
		7	0.075	0.008	15.975	0.025
		8	0.110	0.032	16.835	0.032
		9	-0.072	-0.168	17.208	0.046
		10	-0.049	0.034	17.382	0.066
		11	-0.166	-0.146	19.458	0.053
		12	-0.246	-0.140	24.101	0.020
		13	-0.172	-0.044	26.409	0.015
		14	-0.104	0.047	27.277	0.018
		15	-0.142	-0.126	28.834	0.016

$$\hat{Y}_t - Y_{t-1} = 0.410\left(Y_{t-1} - Y_{t-2}\right)_t$$

$$\hat{Y}_t = (1.410)Y_{t-1} - (0.410)Y_{t-2}$$

y, por lo tanto, se puede concluir que $\{Y_t\}$: $ARIMA(2,1,0)$.

Sea un proceso $\{Y_t\}$ no-estacionario que admite una representación de la forma:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \ldots + \varphi_{p+d} Y_{t-p-d} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

(donde $\{\varepsilon_t\}$: $WN - N(0,\sigma_\varepsilon^2)$).

De acuerdo a lo expresado en la Sec. 2.12.3, el correspondiente coeficiente de determinación podría ser estimado de la siguiente forma:

$$R^2 = 1 - \frac{\sum_{t=p+d+1}^{n} \hat{\varepsilon}_t^2}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2}$$

Pero, dado que en un proceso integrado el valor medio $\overline{Y}$ carece de significado, cualquier modelo que se aplique logrará una explicación apreciablemente mejor de $\{Y_t\}$ y, por lo tanto, este coeficiente asumirá siempre valores muy próximos a la unidad. A fin de resolver esta restricción se aplica una modificación de este coeficiente de la forma

$$R^2 = 1 - \frac{\sum_{t=p+1}^{n} \hat{\xi}_t^2}{\sum_{t=2}^{n} \left(Z_t - \overline{Z}\right)^2}$$

donde $\{Z_t\} = \{\nabla Y_t\}$ y los $\hat{\xi}_t$ denotan los residuos obtenidos de aplicar al proceso $\{Z_t\}$ un modelo de tipo[22]:

$$\hat{Z}_t = \hat{\phi}_0 + \hat{\phi}_1 Z_{t-1} + \hat{\phi}_2 Z_{t-2} + ... + \hat{\phi}_p Z_{t-p}$$

3.2.2.- La representación MA

De la misma forma que en el caso estacionario, el proceso $\{Y_t\}: ARIMA(p,d,q)$ admite una representación MA de orden infinito de la forma:

$$Y_t = \Psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad (\psi_0 = 1)$$

(donde: **i)** Y_t denota una variable centrada; **ii)** $\{\varepsilon_t\}: WN - N(0,\sigma_\varepsilon^2)$ y **iii)** la sucesión $\{\psi_j\}$ no es de cuadrado sumable).

Aplicando m. a m. a esta expresión el operador $\varphi(B)$, se obtiene que:

$$\varphi(B)Y_t = \varphi(B)\Psi(B)\varepsilon_t = \Theta(B)\varepsilon_t$$

de lo que resulta que:

$$\varphi(B)\Psi(B) = \Theta(B)$$

$$\left(1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_{p+d} B^{p+d}\right)\left(1 + \psi_1 B + \psi_2 B^2 + ...\right) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$$

Relación a partir de la cual se obtienen las siguientes fórmulas de recurrencia para los coeficientes ψ_j:

$$\psi_j = \begin{cases} 1 & (j = 0) \\ \varphi_1 - \theta_1 & (j = 1) \\ \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + ... + \varphi_{p+d}\psi_1 - \theta_j & \left(2 \le j \le \min(p+d,q)\right) \\ \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + ... + \varphi_{p+d}\psi_1 & \left(\min(p+d,q) < j \le Max(p+d-1,q)\right) \\ \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + ... + \varphi_{p+d}\psi_{j-p-d} & \left(j > Max(p+d-1,q)\right) \end{cases}$$

Obsérvese que, para $j > Max(p+d,q)$, los coeficientes ψ_j satisfacen la ecuación en-diferencias:

---

22. Ver Sec. 2.10.8.

$$\psi_j - \varphi_1\psi_{j-1} - \ldots - \varphi_{p+d}\psi_{j-p-d} = \varphi(B)\psi_j =$$

$$= \left(1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_{p+d} B^{p+d}\right)\psi_j =$$

$$= \left(1 - \phi_1 B - \ldots - \phi_p B^p\right)(1 - B)^d \psi_j =$$

$$= \Phi(B)(1 - B)^d \psi_j = 0$$

(debe tenerse en cuenta que, en este caso, el operador B afecta al subíndice j) por lo que se puede concluir que, para j suficientemente grande, los coeficientes ψ_j se comportan de acuerdo con una combinación de polinomios, funciones exponenciales amortiguadas y funciones sinusoidales amortiguadas.

Ahora bien, dado que los coeficientes ψ_j no son de cuadrado sumable y, por lo tanto, el sistema no es estable (ver Sec. 2.9.1), esta representación MA de orden infinito de $\{Y_t\}$ no es admisible. Pero, suponiendo que se cuenta con una serie cronológica $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$, se puede hallar una representación truncada finita en términos de una combinación lineal de $n+1$ "shocks" aleatorios, $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-n}$.

Esta forma de representación truncada consiste en considerar a la representación $ARIMA(p,d,q)$, $\varphi(B)Y_t = \Theta(B)\varepsilon_t$ como una ecuación en diferencias cuya solución es de la forma:

$$Y_t = f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1) + e(t-n, n)$$

donde :

i) $f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1)$, que denota la llamada **función complementaria** (y representa la parte de $\{Y_t\}$ conocida al momento $t-n$), es la solución general de la ecuación en-diferencias:

$$\varphi(B) f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1) = 0$$

y, por lo tanto, asumirá la forma:

$$f^{(n)}(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1) = A_1 + A_2 t + \ldots + A_d t^d + C_1\left(G_1^{-1}\right)^n + \ldots + C_p\left(G_p^{-1}\right)^n$$

ii) $e(t-n,n)$, que denota la **integral particular** (y representa la parte de $\{Y_t\}$ no-predecible al momento $t-n$), es cualquier función que satisfaga la ecuación:

$$\varphi(B)e(t-n,n) = \Theta(B)\varepsilon_t$$

y que, por lo tanto, asumirá la forma:

$$e(t-n,n) = \begin{cases} 0 & (t < t-n) \\ \varepsilon_t + \psi_1\varepsilon_{t-1} + \ldots + \psi_n\varepsilon_{t-n+1} & (t \geq t-n) \end{cases}$$

donde los coeficientes ψ_j se calculan a partir de la expresión iterativa que figura en la página precedente.

En consecuencia, la expresión truncada de la representación MA de orden infinito del proceso

$\{Y_t\}$: $ARIMA(p,d,q)$ puede ser escrita de acuerdo con la siguiente descomposición agregativa:

$$Y_t = \sum_{j=0}^{n-1} \psi_j \varepsilon_{t-j} + f^{(n)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right)$$

Obsérvese que, de acuerdo con su definición, se puede escribir:

$$f^{(n)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = f^{(n+n^*)}\left(Y_{t-n-n^*}, Y_{t-n-n^*-1}, \ldots, Y_1\right) + \psi_n \varepsilon_{t-n} + \psi_{n+1} \varepsilon_{t-n-1} + \ldots + \psi_{n+n^*-1} \varepsilon_{t-n-n^*+1}$$

por lo que se puede concluir que la función complementaria varía con el origen $t-n$. Asimismo, se demuestra en forma inmediata que:

$$E\left(Y_t / Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = f^{(n)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right)$$

Sea, en particular, un proceso $\{Y_t\}$ que admite una representación $ARIMA(1,1,1)$ de la forma:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

(donde: i) Y_t denota una variable centrada y ii) $\{\varepsilon_t\}$: $WN - N\left(0, \sigma_\varepsilon^2\right)$), es decir, tal que:

$$\left(1 - \varphi_1 B - \varphi_2 B^2\right) Y_t = \left(1 - \phi_1 B\right)\left(1 - B\right) Y_t = \left(1 - \theta_1 B\right) \varepsilon_t$$

(donde $\varphi_1 = 1 + \phi_1$ y $\varphi_2 = -\phi_1$). La función complementaria, que se obtiene de resolver la ecuación:

$$\left(1 - \varphi_1 B - \varphi_2 B^2\right) f^{(n)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = 0$$

puede ser expresada de la siguiente forma:

$$f^{(n)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = A_1 \left(G_1^{-1}\right)^n + A_2 \left(G_2^{-1}\right)^n = A_1 + A_2 \phi_1^n$$

A partir de las condiciones iniciales Y_{t-n} e $Y_{t-(n+1)}$, se obtienen los valores de A_1 y A_2:

$$\begin{cases} f^{(0)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = A_1 + A_2 = Y_{t-n} \\ f^{(1)}\left(Y_{t-n}, Y_{t-n-1}, \ldots, Y_1\right) = A_1 + A_2 \phi_1 = Y_{t-(n-1)} = \varphi_1 Y_{t-n} + \varphi_2 Y_{t-(n+1)} - \theta_1 \varepsilon_{t-n} \end{cases}$$

De modo que:

$$A_1 = \frac{\phi_1 Y_{t-n} - Y_{t-(n+1)}}{\phi_1 - 1} = \frac{\left(\phi_1 - \varphi_1\right) Y_{t-n} - \varphi_2 Y_{t-(n+1)} + \theta_1 \varepsilon_{t-n}}{\phi_1 - 1} = \frac{Y_{t-n} - \phi_1 Y_{t-(n+1)} - \theta_1 \varepsilon_{t-n}}{1 - \phi_1}$$

$$A_2 = \frac{Y_{t-(n+1)} - Y_{t-n}}{\phi_1 - 1} = \frac{\left(\varphi_1 - 1\right) Y_{t-n} + \varphi_2 Y_{t-(n+1)} - \theta_1 \varepsilon_{t-n}}{\phi_1 - 1} = \frac{-\phi_1 \left(Y_{t-n} - Y_{t-(n+1)}\right) + \theta_1 \varepsilon_{t-n}}{1 - \phi_1}$$

Por su parte, la integral particular se obtiene de resolver la ecuación:

$$\left(1-\varphi_1 B-\varphi_2 B^2\right)e(t-n,n)=\left(1-\phi_1 B\right)e(t-n,n)=\Theta(B)\varepsilon_t$$

y, en consecuencia, puede ser expresada de la siguiente forma:

$$e(t-n,n)=\frac{\Theta(B)}{\left(1-\phi_1 B\right)(1-B)}\varepsilon_t=\sum_{j=0}^{n-1}\left(C_0+C_1\phi_1^j\right)\varepsilon_{t-j}$$

Luego, el proceso $\{Y_t\}$: $ARIMA(1,1,1)$ puede ser representado de la siguiente forma:

$$Y_t=A_1+A_2\phi_1^n+\sum_{j=0}^{n-1}\left(C_0+C_1\phi_1^j\right)\varepsilon_{t-j}$$

3.2.3.- La representación AR

De la misma forma que en el caso estacionario, el proceso $\{Y_t\}$ -que admite una representación $ARIMA(p,d,q)$ de la forma $\varphi(B)Y_t=\Theta(B)\varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN-N(0,\sigma_\varepsilon^2)$)- admite, a su vez, una representación AR de orden infinito de la forma:

$$\Pi(B)Y_t=\left(1-\sum_{j=1}^{\infty}\pi_j B^j\right)Y_t=\varepsilon_t$$

$$Y_t=\sum_{j=1}^{\infty}\pi_j Y_{t-j}+\varepsilon_t$$

(donde ka sucesión $\{\pi_j\}$ no es absolutamente sumable). Luego se puede escribir:

$$\frac{\varphi(B)}{\Theta(B)}Y_t=\left(1-\pi_1 B-\pi_2 B^2-\ldots\right)Y_t=\varepsilon_t$$

Es decir, que:

$$\varphi(B)=\Theta(B)\Pi(B)$$

$$\left(1-\varphi_1 B-\ldots-\varphi_{p+d}B^{p+d}\right)=\left(1-\theta_1 B-\ldots-\theta_q B^q\right)\left(1-\pi_1 B-\pi_2 B^2-\ldots\right)$$

Relación a partir de la cual se obtienen las siguientes fórmulas de recurrencia para los coeficientes π_j:

$$\psi_j=\begin{cases}1 & (j=0)\\ \varphi_1-\theta_1 & (j=1)\\ \varphi_j+\theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\ldots+\theta_{p+d}\pi_1-\theta_j & \left(2\le j\le \min(p+d,q)\right)\\ \varphi_j+\theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\ldots+\theta_{p+d}\pi_1-\theta_j & \left(\min(p+d,q)<j\le Max(p+d-1,q)\right)\\ \theta_1\pi_{j-1}+\theta_2\pi_{j-2}+\ldots+\theta_{p+d}\pi_{j-p-d} & \left(j>Max(p+d-1,q)\right)\end{cases}$$

Obsérvese que, para $j > Max(p+d,q)$, los coeficientes π_j satisfacen la ecuación en-diferencias:

$$\pi_j - \theta_1\pi_{j-1} - ... - \theta_q\pi_{j-q} = \Theta(B)\pi_j = (1 - \theta_1 B - ... - \theta_q B^q)\pi_j = 0$$

(debe tenerse en cuenta que, en este caso, el operador B afecta al subíndice j). De modo que se puede concluir que, para j suficientemente grande, los coeficientes π_j se comportan de acuerdo con una combinación de funciones exponenciales y sinusoidales amortiguadas.

Teniendo en cuenta que, si $d \geq 1$, $\varphi(B) = \Phi(B)(1-B)^d$ es igual a cero para $B = 1$ y, dado que las raíces de la ecuación $\Theta(B) = 0$ son en módulo mayores que la unidad, entonces $\Theta(1) \neq 0$, se puede concluir que:

$$\frac{\varphi(B)}{\Theta(B)} = \frac{\Phi(B)(1-B)^d}{\Theta(B)} = 1 - \sum_{j=1}^{\infty}\pi_j B^j$$

será igual a cero para $B = 1$. Es decir, se verifica que $\sum_{j=1}^{\infty}\pi_j = 1$ y, por lo tanto, que el proceso $\{Y_t\}$ admite una representación como **promedio ponderado** de sus realizaciones pasadas de la forma:

$$Y_t = \sum_{j=1}^{\infty}\pi_j Y_{t-j} + \varepsilon_t$$

3.3.- Los tests de las raíces unitarias

3.3.1.- El caso de los procesos AR(1)

A fin de decidir si un proceso $\{Y_t\}$ puede ser considerado o no estacionario de segundo orden, se han desarrollado procedimientos paramétricos para testear la existencia de raíces unitarias en la ecuación característica de su componente autorregresiva[23].

Sea, en particular, un proceso $\{Y_t\}$ que admite una representación $AR(1)$ de la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\} WN - N(0,\sigma_\varepsilon^2)$). Como se vio en la Sec. 2.10.3, teniendo en cuenta que el orden de autorregresividad es conocido, el coeficiente ϕ_1 puede ser estimado consistentemente utilizando el criterio de optimización de los cuadrados mínimos. En consecuencia, este estimador debería permitir testear la hipótesis nula $H_0: \phi_1 = 1$, con respecto a la hipótesis alternativa $H_1: \phi_1 < 1$ (un test de una cola debido a que los fenómenos explosivos, para los cuales $\phi_1 > 1$, como se mencionó en la Sec. 3.2, no tienen solución en el ámbito de la teoría de los procesos estocásticos) pero, bajo el supuesto de que la hipótesis nula es verdadera, $\{Y_t\}$ admite una representación MA no-

---

[23]. Ver Fuller, W.A. (1976), Dickey, D.A.; Bell, W.R.; Miller, R.B. (1986), Dolado, J.J.; Jenkinson, T.; Sosvilla-Rivero, S. (1990). Como se verá más adelante, a menudo estos tests han sido confundidos con tests para contrastar las hipótesis de $\{Y_t\}: RW$ (los procesos "random walk" serán tratados en la Sec. 4.1).

estacionaria de la forma $Y_t = \sum_{j=0}^{\infty} \varepsilon_{t-j}$, a partir de la cual se puede concluir que $\lim_{t\to\infty} \sigma(Y_t) = \infty$ y que, por lo tanto, los métodos de estimación clásicos y los correspondientes tests de significatividad respecto del coeficiente ϕ_1 resultan inapropiados [24].

Obsérvese que, bajo el supuesto que la hipótesis nula sea verdadera, el estimador mínimo-cuadrático de ϕ_1:

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} Y_{t-1}Y_t}{\sum_{t=2}^{n} Y_{t-1}^2} = \phi_1 + \frac{\sum_{t=2}^{n} Y_{t-1}\varepsilon_t}{\sum_{t=2}^{n} Y_{t-1}^2}$$

que, como se demostró en la Sec. 3.2.1 , está íntimamente relacionado con el estimador $\hat{\rho}_1(Y)$ ($\hat{\rho}_1(Y) = \frac{t-1}{t} < 1$), presenta un sesgo a ser menor que la unidad cuando su verdadero valor sea igual a la unidad. Una aparente solución a este problema consiste en utilizar una forma equivalente -y más conveniente- del test anterior que se obtiene a partir de una expresión de la mencionada representación $AR(1)$ anterior en términos de las diferencias primeras de $\{Y_t\}$ de la forma:

$$Y_t - Y_{t-1} = \nabla Y_t = (\phi_1 - 1)Y_{t-1} + \varepsilon_t$$

De acuerdo con esta expresión, la hipótesis nula $\phi_1 = 1$ se transforma en la hipótesis $H_0{:}\phi_1 - 1 = 0$ que podría ser contrastada respecto de la alternativa $H_1{:}\phi_1 - 1 < 0$, utilizando el estadístico T sobre el coeficiente $(\phi_1 - 1)$. Ahora bien, como se demostró en la Sec. 2.10.6, bajo el supuesto de estacionariedad, es decir, bajo el supuesto que $|\phi_1| < 1$, se verifica que $\sqrt{n-1}(\hat{\phi}_1 - \phi_1) \xrightarrow{d} N(0, 1-\phi_1^2)$, de modo que, si se supone que $\phi_1 = 1$, entonces $\sigma^2\left[\sqrt{n-1}(\hat{\phi}_1 - \phi_1)\right] = 0$ y, por lo tanto, se puede concluir que esta variable no es útil para la costrucción del test [25].

Suponiendo que: **i)** $\phi_1 = 1$ y **ii)** $Y_1 = 0$, la representación $AR(1)$ anterior puede ser expresada de la siguiente forma:

$$Y_t + \varepsilon_t = \sum_{j=0}^{t-2} \varepsilon_{t-j}$$

por lo que se puede concluir que $Y_t{:}N(0,(t-1)\sigma_\varepsilon^2)$. Por otra parte, a partir de la misma expresión, se obtiene que:

$$Y_t^2 = (Y_{t-1} + \varepsilon_t)^2 = Y_{t-1}^2 + 2Y_{t-1}\varepsilon_t + \varepsilon_t^2$$

---

[24]. Ver Sec. 4.1.

[25]. Como se verá en el Cap. 5, existen otros tests de no-estacionariedad (ver Park, J.Y.; Choi, B. (1988), Tanaka, K. (1990), Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. (1992), Saikkonen, P.; Lukkonen, R. (1993), Kahn, J.A.; Ozaki, M. (1997)) que toman como nula la hipótesis de que el proceso es estacionario. Pero es obvio que, para este caso en particular, resulta mucho más conservador tomar a H_0 como la hipótesis de no-estacionariedad.

es decir, que:

$$Y_{t-1}\varepsilon_t = \frac{1}{2}\left(Y_t^2 - Y_{t-1}^2 - \varepsilon_t^2\right)$$

y, teniendo en cuenta que $Y_1 = 0$, resulta que

$$\sum_{t=2}^{n} Y_{t-1}\varepsilon_t = \frac{1}{2}Y_n^2 - \frac{1}{2}\sum_{t=2}^{n}\varepsilon_t^2$$

y, por lo tanto, que:

$$\frac{1}{(n-1)\sigma_\varepsilon^2}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t = \frac{1}{2}\left(\frac{Y_n}{\sigma_\varepsilon\sqrt{n-1}}\right)^2 - \frac{1}{2\sigma_\varepsilon^2}\frac{1}{n-1}\sum_{t=2}^{n}\varepsilon_t^2$$

(donde: **i)** $\left(\frac{Y_n}{\sigma_\varepsilon\sqrt{n-1}}\right)^2 : \chi^2_{(1)}$ y **ii)** $\frac{1}{n-1}\sum_{t=2}^{n}\varepsilon_t^2 \xrightarrow{d} \sigma_\varepsilon^2$). Luego, se puede concluir que:

$$\frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t \xrightarrow{d} \frac{\sigma_\varepsilon^2}{2}\left(\chi^2_{(1)} - 1\right)$$

Por otra parte, teniendo en cuenta que $E\left(Y_t^2\right) = (t-1)\sigma_\varepsilon^2$, se obtiene que:

$$E\left(\sum_{t=2}^{n} Y_{t-1}^2\right) = \sigma_\varepsilon^2\sum_{t=2}^{n}(t-1) = \frac{(n-1)n}{2}\sigma_\varepsilon^2$$

y, por lo tanto, que:

$$\lim_{n\to\infty} E\left(\frac{1}{n^2}\sum_{t=2}^{n} Y_{t-1}^2\right) = \frac{\sigma_\varepsilon^2}{2}\left(1 - \frac{1}{n-1}\right) = \frac{\sigma_\varepsilon^2}{2}$$

De los resultados anteriores se obtiene que la variable:

$$(n-1)\left(\hat{\phi}_1 - 1\right) = \frac{\frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t}{\frac{1}{(n-1)^2}\sum_{t=2}^{n} Y_{t-1}^2}$$

converge en-distribución a un cociente entre una variable $\chi^2_{(1)}$ y una función de densidad no estándar.

Como se verá en la Sec.4.4, a partir de los postulados del teorema central del límite funcional (conocido también como **teorema de Donsker**), se demuestra que la variable $\frac{Y_t}{\sigma_\varepsilon\sqrt{n-1}}$ converge débilmente a un **movimiento Browniano estandarizado** $\{B(t)\}$. Asimismo, se demuestra que $\frac{1}{t}[B(t)]^2 : \chi^2_{(1)}$.

De modo que, a partir de la expresión que figura más arriba, se verifica que[26]:

$$\frac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t \xrightarrow{d} \frac{\sigma_\varepsilon^2}{2}\left\{\left[B(1)\right]^2 - 1\right\}.$$

De la misma forma, se demuestra que (ver Phillips, P.C.B. (1987)) [27]:

$$\frac{1}{(n-1)^2}\sum_{t=2}^{n} Y_{t-1}^2 = \frac{\sigma_\varepsilon^2}{n-1}\sum_{t=2}^{n}\left(\frac{Y_{t-1}}{\sigma_\varepsilon\sqrt{n-1}}\right)^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 \left[B(t)\right]^2 dt$$

Luego, a partir de estos resultados se puede concluir que:

$$(n-1)\left(\hat{\phi}_1 - 1\right) \xrightarrow{d} \frac{\frac{1}{2}\left\{\left[B(1)\right]^2 - 1\right\}}{\int_0^1 \left[B(t)\right]^2 dt}$$

Como se verá en la Sec. 4.5, el numerador de esta expresión es la solución (obtenida a partir de la aproximación que proporciona la suma de Riemann-Stieltjes) de la **integral estocástica de Itô**[28], de modo que se puede escribir:

$$(n-1)\left(\hat{\phi}_1 - 1\right) \xrightarrow{d} \frac{\int_0^1 B(t)\,dB(t)}{\int_0^1 \left[B(t)\right]^2 dt}$$

---

[26]. Ver Banerjee, A.; Dolado, J.J.; Galbraith, J.W.; Hendry, D.F. (1993), Hamilton, J.D. (1994). Esta convergencia en-distribución de un proceso $AR(1)$ de parámetro discreto no-estacionario a un movimiento Browniano, a partir de la aplicación del teorema central del límite funcional y las características del movimiento Browniano, serán tratadas en la Sec. 4.4.

[27]. Como se mencionó en la Sec. 1.4.2, se dice que una sucesión de funciones $\{h_n(y)\}$ **converge débilmente** a la función $h(y)$ si, para todos los puntos de continuidad de $h(y)$, se verifica que $\underset{n\to\infty}{Lim}\, h_n(y) = h(y)$ (la notación Lim es utilizada aquí a fin de distinguir la convergencia débil de la convergencia punto a punto). Aplicando esta definición a las funciones de distribución de probabilidades, se dice que una sucesión de funciones de distribución $\{F_{Y_t}(y)\}$ converge en-distribución a una función $F_Y(y)$ ($F_{Y_t}(y) \xrightarrow{d} F_Y(y)$), si se verifica que $\underset{t\to\infty}{Lim}\, F_{Y_t}(y) = F_Y(y)$. Debe tenerse en cuenta que $F_Y(y)$ puede no ser una función de distribución (ver Lukacs, E. (1970)). La condición necesaria y suficiente para que se produzca esta convergencia débil (que demuestra que la correspondencia biunívoca entre funciones de distribución y funciones características es continua) es que la correspondiente sucesión de funciones características $\{C_{Y_t}(w)\}$ converja a la función característica correspondiente a la función $F_Y(y)$, $C_Y(w)$, para todo w (suponiendo que $C_Y(w)$ es continua en $w = 0$). Asimismo, se demuestra que es condición necesaria y suficiente para la convergencia débil de la sucesión $\{F_{Y_t}(y)\}$ a $F_Y(y)$ que $\lim_{t\to\infty} L\left(F_{Y_t}(y); F_Y(y)\right) = 0$ (donde $L(\bullet;\bullet)$ denota la **medida de distancia de Lévy** entre dos funciones de distribución). A partir de esta definiciones, se dice que una sucesión de variables aleatorias $\{Y_t\}$, con funciones de distribución $F_{Y_t}(y)$, converge **débilmente** o **en-distribución** a una variable Y, con función de distribución $F_Y(y)$, si se verifica que $\underset{t\to\infty}{Lim}\, F_{Y_t}(y) = F_Y(y)$.

[28]. Ver Sec. 4.6.

Obsérvese que: **i)** el denominador es positivo; **ii)** dado que $[B(1)]^2 : \chi^2_{(1)}$ y que $p\left(\chi^2_{(1)} < 1\right) = 0.68$, cuando $n \to \infty$, se verifica que $p\left[(n-1)(\hat{\phi}_1 - 1) < 1\right] \to 0.68$ (es decir, que se debería esperar que en $\frac{2}{3}$ de las observaciones generadas por un proceso que admite una representación de la forma $Y_t = Y_{t-1} + \varepsilon_t$, donde $\{\varepsilon_t\} : WN - N(0, \sigma^2_\varepsilon)$, ϕ_1 sea menor que la unidad) y **iii)** la distribución de la variable $(n-1)(\phi - 1)$ presenta asimetría hacia la izquierda.

Ahora bien, como se mencionó en la Sec. 2.10.3, el estadístico convencional para contrastar la hipótesis nula planteada precedentemente, $H_0 : \phi_1 - 1 = 0$, respecto de la alternativa $H_1 : \phi_1 - 1 < 0$, es de la forma $T = \dfrac{\hat{\phi}_1 - 1}{\hat{\sigma}(\hat{\phi}_1)} = \dfrac{\hat{\phi}_1 - 1}{\dfrac{\hat{\sigma}_\varepsilon}{\sqrt{\sum_{t=2}^{n} Y^2_{t-1}}}}$ (donde $\hat{\sigma}^2_\varepsilon = \dfrac{1}{n-2}\sum_{t=2}^{n}\left(Y_t - \hat{\phi}_1 Y_{t-1}\right)^2$ es el estimador consistente de σ^2_ε, obtenido de aplicar CMO) y, dado que, como ya se demostró, cuando $\phi_1 = 1$ el estadístico T no converge en-distribución a la función Normal, resulta conveniente expresar el estadístico T de la siguiente forma[29]:

$$\tau = \frac{n-1}{\hat{\sigma}_\varepsilon}(\hat{\phi}_1 - 1)\frac{1}{n-1}\sqrt{\sum_{t=2}^{n} Y^2_{t-1}} = \frac{1}{\hat{\sigma}_\varepsilon}\frac{\dfrac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t}{\dfrac{1}{(n-1)^2}\sum_{t=2}^{n} Y^2_{t-1}}\frac{1}{n-1}\sqrt{\sum_{t=2}^{n} Y^2_{t-1}} = \frac{\dfrac{1}{n-1}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t}{\hat{\sigma}_\varepsilon\sqrt{\dfrac{1}{(n-1)^2}\sum_{t=2}^{n} Y^2_{t-1}}}$$

De modo que, de acuerdo con los resultados obtenidos en las páginas anteriores, se puede concluir que[30]:

$$\tau \xrightarrow{d} \frac{\dfrac{\sigma^2_\varepsilon}{2}\left\{[B(1)]^2 - 1\right\}}{\sigma_\varepsilon\sqrt{\sigma^2_\varepsilon \int_0^1 [B(t)]^2\, dt}} = \frac{\dfrac{1}{2}\left\{[B(1)]^2 - 1\right\}}{\sqrt{\int_0^1 [B(t)]^2\, dt}} = \frac{\int_0^1 B(t)\, dt}{\sqrt{\int_0^1 [B(t)]^2\, dt}}$$

Aquí se ha considerado exclusivamente el caso en que el estimador del coeficiente ϕ_1 es positivo, pero debe tenerse en cuenta que, de acuerdo a lo expresado en la Sec. 2.10.6, la condición de

---

29. La denominación τ para denotar a este estadístico (debida a Fuller, W.A. (1996)) es para distinguirlo del estadístico T convencional.

30. Una forma alternativa de este test, basado en la condición de $\hat{\phi}_1$ como estimador superconsistente de $\phi_1 = 1$, se obtiene de dividir la variable $(n-1)(\hat{\phi}_1 - 1)$ por $\sqrt{n-1}$, de lo que resulta $\sqrt{n-1}(\hat{\phi}_1 - 1) = \dfrac{\left(\dfrac{1}{n-1}\right)^{3/2}\sum_{t=2}^{n} Y_{t-1}\varepsilon_t}{\dfrac{1}{(n-1)^2}\sum_{t=2}^{n} Y^2_{t-1}}$. El numerador de esta variable converge en-distribución a $\dfrac{1}{2\sqrt{n-1}}\sigma^2_\varepsilon\left(\chi^2_{(1)} - 1\right)$ y, como $\sigma^2\left(\chi^2_{(1)}\right) = 2$, la varianza de esta variable es de orden $\frac{1}{n}$ y, en consecuencia, se verifica que $\sqrt{n-1}(\hat{\phi}_1 - 1) \xrightarrow{p} 0$ (ver Hamilton, J.D. (1994)). Este resultado permite concluir que esta versión de la variable $(n-1)(\hat{\phi}_1 - 1)$ converge en-distribución a la misma función de probabilidades que la versión comentada en el texto, con la ventaja que puede ser utilizada como estadístico del test sin necesidad de calcular su desvío estándar.

estacionariedad exige que $-1 < \phi_1 < 1$. De modo que, si el estimador de ϕ_1 es próximo a -1, también existe una probabilidad no-nula de no-estacionariedad y, si se denota por $\phi_1^* = \phi_1 - 1$, la condición de estacionariedad equivalente será que $-2 < \phi_1^* < 0$. En este caso habrá que decidir si, con un cierto nivel de confiabilidad, ϕ_1^* puede o no ser considerado mayor que -2[31].

Dickey, D.A.; Fuller, W.A. (1981), basándose en métodos de simulación de Monte Carlo, tabularon esta distribución límite y calcularon los correspondientes valores críticos [32]. Estas técnicas de Monte Carlo fueron aplicadas a partir de la fijación de la hipótesis $\phi_1 = 1$ y un diseño muestral que supone que las perturbaciones ε_t están Normalmente distribuidas[33]. Se generaron, entonces, miles de muestras para Y_t de acuerdo con la representación en diferencias anterior y se consideró el porcentaje de veces en que se rechazaba la hipótesis nula de presencia de raíz unitaria, cuando ésta era verdadera, basándose en el estadístico $T = \dfrac{\hat{\phi}_1 - 1}{\hat{\sigma}\left(\hat{\phi}_1\right)}$.

Asimismo, Dickey, D.A.; Fuller, W.A. (1981) demostraron que la definición de la zona crítica para testear la hipótesis $H_0: \phi_1 - 1 = 0$ depende, además del tamaño de la muestra, de la forma de la ecuación autorregresiva (por ejemplo, si incluye o no término independiente o términos de tendencia determinística) y definieron tres clases de valores críticos (τ, τ_μ y τ_τ) para ser utilizados, respectivamente, en el test según que la ecuación autorregresiva: **i)** no incluya término independiente ni término lineal; **ii)** incluya término independiente, pero no incluya término lineal; **iii)** incluya ambos términos.

Dado que, como se mencionó más arriba, suponiendo que el estimador de ϕ_1 es positivo se trata de un test de una cola, se considerará que existen razones suficientes para rechazar la hipótesis nula de no-estacionariedad si el valor del estadístico $T = \dfrac{\hat{\phi}_1 - 1}{\hat{\sigma}\left(\hat{\phi}_1\right)}$ es menor que el correspondiente valor τ.

---

[31] . Si, hipotéticamente, no se contara con información acerca del signo del estimador de ϕ_1, debe construirse un test de dos "colas".

[32]. Los llamados métodos de Monte Carlo (cuyos antecedentes pueden ser hallados, entre otros, en los trabajos de Erastus Lyman de Forest (ver Stigler, S.M. (1976)), Student (1908), Yule, G.U. (1926) y Orcutt, G.H.; Cochrane, D. (1949) y cuya denominación se debe a Metropolis, N.; Ulam, S. (1949)) fueron diseñados para analizar las propiedades de las distribuciones de variables aleatorias, mediante la utilización de técnicas de simulación de observaciones iid. Contrariamente a lo que ocurre en las aplicaciones en estadística, en las que la simulación del comportamiento de un fenómeno de resultado eventual requiere de la construcción de un modelo, en los métodos de Monte Carlo el objeto de la investigación es el modelo mismo. Todo resultado de un experimento de Monte Carlo es, esencialmente, una estimación del valor de una integral de la forma $m = \dfrac{1}{n}\int_0^1 f_X(x)dx$ la cual, si existe, define el valor esperado de una variable $X: U(0;1)$. De acuerdo con la técnica conocida como **Monte Carlo-cerrada**, a partir de una muestra aleatoria $\{x_1, x_2, ..., x_n\}$ obtenida de una distribución $U(0;1)$, se obtiene el estimador insesgado $\hat{m} = \dfrac{1}{n}\sum_{j=1}^{n} f_X(x_j)$, con varianza $\sigma^2(\hat{m}) = \dfrac{1}{n}\int_0^1 [f_X(x) - m]^2 dx$ (debe tenerse en cuenta que, dado que las propiedades consideradas como más importantes de un estimador son la de insesgado y de varianza mínima, muchas de las investigaciones en el método de Monte Carlo está dirigidas a la reducción de la varianza del estimador). El caso mencionado en el texto se refiere al **test de Monte Carlo** introducido por Barnard, G.A. (1963). Para más detalles sobre las características de este método, ver Hamersley, J.M.; Handscomb, D.C. (1964), Hendry, D.F. (1973), Hendry, D.F.; Harrison, R.W. (1974)(en Griliches, Z.; Intriligator, M.D. (eds.))(1984), Kennedy, W.J.; Gentle, J.E. (1980), Kleijnen, J.P.C. (1974), Mikhail, W.M. (1975), Mizon, G.E.; Hendry, D.F. (1980), Sobol', I.M. (1974).

[33] . Ver Dickey, D.A. (1976), Fuller, W.A. (1976), Dickey, D.A.; Fuller, W.A. (1976), Fuller, W.A. (1976), Said, S.E.; Dickey, D.A. (1984)(1985), Schmidt, P.; Phillips, P.C.B. (1992).

En los casos en que la representación incluya componentes determinísticos, el test para detectar la presencia de raíces unitarias se conoce como **Dickey-Fuller aumentado**. Si el proceso $\{Y_t\}$ es tal que admite una representación de la forma $Y_t = a_0 + \phi_1 Y_{t-1} + \varepsilon_t$ (donde $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$), entonces, de la misma forma que en el caso precedente, se puede concluir que:

$$(n-1)(\hat{\phi}_1 - 1) \xrightarrow{d} \frac{\frac{1}{2}\left\{[B(1)]^2 - 1\right\} - B(1)\int_0^1 B(t)\,dt}{\int_0^1 [B(t)]^2\,dt - \left(\int_0^1 B(t)\,dt\right)^2}$$

y que el correspondiente estadístico del test es tal que:

$$\tau_\mu \xrightarrow{d} \frac{\frac{1}{2}\left\{[B(1)]^2 - 1\right\} - B(1)\int_0^1 B(t)\,dt}{\sqrt{\int_0^1 \left[B(t)^2\,dt - \left(\int_0^1 B(t)\,dt\right)^2\right]}}$$

(obsérvese que el estadístico τ_μ converge, en este caso, a una distribución Normal desplazada del origen).

De la misma forma, si el proceso $\{Y_t\}$ es tal que admite una representación de la forma $Y_t = a_0 + a_1 t + \phi_1 Y_{t-1} + \varepsilon_t$ (donde $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$), entonces se demuestra que el correspondiente estadístico del test es tal que:

$$\tau_\tau \xrightarrow{d} \frac{\frac{1}{2}\left\{[B(1)]^2 - 1\right\} - B(1)\int_0^1 B(t)\,dt + 12\left[\int_0^1 tB(t)\,dt - \frac{1}{2}\right]\left[\int_0^1 B(t)\,dt - \frac{1}{2}B(1)\right]}{\sqrt{\int_0^1 [B(t)]^2\,dt - \left(\int_0^1 B(t)\,dt\right)^2 + 12\left[\int_0^1 B(t)\,dt \int_0^1 tB(t)\,dt - \left(\int_0^1 tB(t)\,dt\right)^2 - 3\left(\int_0^1 B(t)\,dt\right)^2\right]}}$$

Si se supone que la representación de $\{Y_t\}$ puede incluir un término independiente, entonces la ecuación de Dickey-Fuller será de la forma:

$$\nabla Y_t = a_0 + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$$

El valor crítico con respecto al cual contrastar el valor del estadístico T en este caso, será τ_μ, el cual fue generado suponiendo que la verdadera representación de la variable diferenciada es de la forma $\nabla Y_t = (\phi_1 - 1)Y_{t-1} + \varepsilon_t$. Debe tenerse en cuenta que ϕ_1 y τ_μ son invariantes respecto de Y_1, es decir que, cualquiera sea el valor inicial de la serie, la distribución del estadístico τ_μ será la misma (**propiedad de similitud**).

Si se supone que la representación de $\{Y_t\}$ puede estar afectada además por una tendencia estacionaria lineal, entonces será necesario incluir en ésta tantos regresores determinísticos como componentes determinísticos involucre dicha tendencia. En el caso particular de una hipótesis alternativa que suponga la presencia de una tendencia estacionaria lineal, la ecuación de Dickey-Fuller será de la forma:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$$

En este caso, el valor del estadístico T debería ser contrastado con el valor crítico τ_τ [34]. Que existan razones suficientes para rechazar la hipótesis $H_0{:}\phi_1 - 1 = 0$ implica aceptar la hipótesis de que la representación en términos de las primeras diferencias posee sólo tendencia determinística más "shocks" aleatorios (debe tenerse en cuenta que, igual que en el caso precedente, ϕ_1 y τ_τ son invariantes con respecto a Y_1 y a a_0).

De la misma forma supóngase una hipótesis alternativa que incluya una tendencia determinística parabólica de segundo grado, entonces para testear la presencia de una raíz unitaria sería necesaria una representación que incluyera un término cuadrático en dicha tendencia determinística[35]:

$$\nabla Y_t = a_0 + a_1 t + a_2 t^2 + (\phi_1 - 1) Y_{t-1} + \varepsilon_t$$

Como se demostró en las páginas anteriores, el agregado de componentes determinísticos en el modelo utilizado en el test de Dickey y Fuller, más allá de los verdaderamente involucrados en la representación del proceso, aumenta la probabilidad de aceptación de la hipótesis nula de no-estacionariedad, cuando la hipótesis verdadera es la de estacionariedad. Este fenómeno se manifiesta en el comportamiento de los valores críticos: $\tau_\tau < \tau_\mu < \tau$ (al decrecer los valores críticos se vuelve más difícil rechazar la hipótesis nula, aún cuando debería ser rechazada). Asimismo, la omisión de un término independiente y un término lineal significativos, provoca una disminución en el poder del test directamente proporcional a la significatividad de los coeficientes.

Por otra parte, debe tenerse en cuenta que los tests para analizar la significatividad de los coeficientes de los regresores determinísticos están afectados por la presencia de raíces unitarias. A fin dice resolver este problema, Dickey, D.A.; Fuller, W.A. (1981) diseñaron tres estadísticos, Φ_1, Φ_2 y Φ_3 (cuya estructura es similar a la de los estadísticos utilizados en los tests de Wald de significatividad conjunta de los coeficientes de un modelo clásico), para testear en forma conjunta la significatividad de los coeficientes de la representación del proceso $\{\nabla Y_t\}$.

Estos estadísticos están definidos de la siguiente forma:

$$\Phi_i = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR(i)}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR(i)}} : \Im_{r,n-k} \qquad (i = 1,2,3)$$

donde: **i)** $\left(\sum \hat{\varepsilon}_t^2\right)_{CR}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo **con restricciones** de la forma $\nabla \hat{Y}_t = \left(\hat{\phi}_1 - 1\right) Y_{t-1}$; **ii)** $\left(\sum \hat{\varepsilon}_t^2\right)_{SR(1)}$ denota la suma de los cuadrados de los residuos

---

[34]. Obsérvese que, suponiendo que $Y_t = 0$, esta representación puede ser expresada de la siguiente forma $\nabla Y_t = a_0 \sum_{j=2}^{t} \phi_1^{t-j} +$

$+ a_1 \sum_{j=2}^{t} j\phi_1^{t-j} + \sum_{j=2}^{t} \phi_1^j \varepsilon_{t-j}$. Bajo el supuesto que la hipótesis nula $H_0{:}\phi_1 = 1$ es verdadera, se puede escribir $\nabla Y_t = -(a_0 + a_1) +$

$+ \left(a_0 + \frac{a_1}{2}\right) t + \frac{a_1}{2} t^2 + \sum_{j=2}^{t} \varepsilon_{t-j}$ por lo que se puede concluir que, en este caso, $\{\nabla Y_t\}$ está afectada por una tendencia de segundo grado (que es poco habitual en las aplicaciones).

[35]. Este caso no es de presentación frecuente en las aplicaciones, ya que,si $\{Y_t\}$ está expresado en escala logarítmica, esta tendencia implica una tasa de variación siempre creciente (o decreciente).

que resultan de aplicar un modelo **sin restricciones** de la forma $\nabla \hat{Y}_t = \hat{a}_0 + \left(\hat{\phi}_1 - 1\right)Y_{t-1}$; **iii)** $\left(\sum \hat{\varepsilon}_t^2\right)_{SR(2)}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo **sin restricciones** de la forma $\nabla \hat{Y}_t = \hat{a}_0 + \hat{a}_1 t + \left(\hat{\phi}_1 - 1\right)Y_{t-1}$; **iv)** $\left(\sum \hat{\varepsilon}_t^2\right)_{SR(3)}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo **sin restricciones** de la forma $\nabla \hat{Y}_t = \hat{a}_1 t + \left(\hat{\phi}_1 - 1\right)Y_{t-1}$; **v)** r denota el número de restricciones; **vi)** n denota el número de observaciones y **vii)** k denota el número de coeficientes incluidos en la representación sin restricciones.

El estadístico Φ_1 se utiliza para contrastar la hipótesis nula $H_0{:}a_0 = \phi_1 - 1 = 0$; el estadístico Φ_2 se utiliza para contrastar la hipótesis nula $H_0{:}a_0 = a_1 = \phi_1 - 1 = 0$ y el estadístico Φ_3 se utiliza para contrastar la hipótesis nula $H_0{:}a_1 = \phi_1 - 1 = 0$.

Dickey, D.A.; Fuller, W.A. (1981) calcularon y tabularon las distribuciones límite de los estadísticos Φ_i que mejoran la distribución clásica $\Im_{r,n-k}$ del test de Wald.

Dickey-Fuller generaron, además, los estadísticos $\tau_{\beta\tau}$ y $\tau_{\alpha\tau}$ para contrastar, respectivamente, las hipótesis de significatividad del coeficiente correspondiente al término tendencia lineal y del término independiente bajo el supuesto que la hipótesis $\phi_1 - 1 = 0$ es verdadera (es decir, contrastar la hipótesis nula $H_0{:}a_1 = 0/\left(\phi_1 - 1\right) = 0$ con respecto a la alternativa $H_1{:}a_1 \neq 0/\left(\phi_1 - 1\right) = 0$ y la hipótesis nula $H_0{:}a_0 = 0/\left(\phi_1 - 1\right) = 0$ con respecto a la alternativa $H_1{:}a_0 \neq 0/\left(\phi_1 - 1\right) = 0$). Estos estadísticos están definidos de la siguiente forma:

$$\tau_{\beta\tau} = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

(donde: **i)** $\left(\sum \hat{\varepsilon}_t^2\right)_{CR}$ denota la suma de los cuadrados de los residuos correspondientes a un modelo con restricciones de la forma $\nabla Y_t = \hat{a}_0$ y **ii)** $\left(\sum \hat{\varepsilon}_t 2\right)_{SR}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo sin restricciones de la forma $\nabla Y_t = \hat{a}_0 + \hat{a}_1 t + \left(\hat{\phi}_1 - 1\right)Y_{t-1}$) y:

$$\tau_{\alpha\tau} = \frac{\frac{1}{r}\left[\left(\sum_t \nabla Y_t^2\right) - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

(donde: **i)** $\left(\sum \hat{\varepsilon}_t^2\right)_{CR}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo con restricciones de la forma $\nabla Y_t = \hat{a}_1 t$ y **ii)** $\left(\sum \hat{\varepsilon}_t^2\right)_{SR}$ denota la suma de los cuadrados de los residuos que resultan de aplicar un modelo sin restricciones de la forma $\nabla \hat{Y}_t = \hat{a}_0 + \hat{a}_1 t + \left(\hat{\phi}_1 - 1\right)Y_{t-1}$).

Todos estos tests de significatividad se aplican de acuerdo con la siguiente secuencia:

i) Partiendo de la representación más general $\nabla Y_t = a_0 + a_1 t + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$, se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (el correspondiente valor crítico es τ_τ). Si existen razones suficientes para rechazar H_0, la secuencia se detiene.

ii) Si no existen razones suficientes para rechazar la hipótesis H_0 del punto **i)**, entonces, utilizando la misma representación, se contrasta la hipótesis $H_0{:}a_1 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}a_1 \neq 0 / \phi_1 - 1 = 0$ (se utiliza el estadístico $\tau_{\beta\tau}$). La decisión sobre la aceptación o el rechazo de esta hipótesis H_0 puede ser reforzada contrastando la hipótesis $H_0{:}\phi_1 - 1 = a_1 = 0$ (el correspondiente valor crítico es Φ_3).

iii) Si existen razones suficientes para rechazar la hipótesis nula $H_0{:}\phi_1 - 1 = a_1 = 0$ del punto **ii)**, se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (el valor crítico será el que resulte de las distribuciones "t" o Normal)[36]. Si existen razones suficientes para rechazar H_0, la secuencia se detiene.

iv) Si no existen razones suficientes para rechazar la hipótesis H_0 del punto **ii)**, entonces, utilizando la representación $\nabla Y_t = a_0 + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$, se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (el correspondiente valor crítico es τ_μ).

v) Si no existen razones suficientes para rechazar la hipótesis H_0 del punto **iv)**, entonces, utilizando la misma representación, se contrasta la hipótesis $H_0{:}a_0 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}a_0 \neq 0 / \phi_1 - 1 = 0$ (el correspondiente valor crítico es $\tau_{\alpha\tau}$). La decisión sobre el rechazo o no de esta hipótesis H_0 puede ser reforzada contrastando la hipótesis $H_0{:}\phi_1 - 1 = a_0 = 0$ (el correspondiente valor crítico es Φ_1).

vi) Si existen razones suficientes para rechazar la hipótesis H_0 del punto **v)**, se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (el correspondiente valor crítico será el que resulte de las distribuciones "t" o Normal). Si existen razones suficientes para rechazar H_0 se puede concluir que $\{Y_t\}$ no posee raíces unitarias.

vii) Si no existen razones suficientes para rechazar la hipótesis $H_0{:}\phi_1 - 1 = a_0 = 0$ del punto **v)**, entonces se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (el correspondiente valor crítico es τ). Si existen razones suficientes para rechazar H_0 se puede concluir que $\{Y_t\}$ no posee raíces unitarias.

viii) Si no existen razones suficientes para rechazar la hipótesis H_0 del punto **vii)**, entonces, utilizando la representación $\nabla Y_t = (\phi_1 - 1)Y_{t-1} + \varepsilon_t$, se contrasta la hipótesis $H_0{:}\phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}\phi_1 - 1 < 0$ (El correspondiente valor crítico es τ).

Los tests diseñados hasta aquí se basan en el supuesto que $\{\varepsilon_t\}{:}WN - N(0,\sigma_\varepsilon^2)$. Si se verifica que

---

[36].Es decir, si se sabe que el proceso no incluye términos determinísticos, el estadístico τ o bedece a una distribución de Dickey-Fuller, por el contrario, si se sabe que el proceso incluye un término independiente y un término de tendencia determinística, la hipótesis $H_0{:}\phi_1 - 1 = 0$ puede ser contrastada utilizando una distribución Normal.

$\{\varepsilon_t\}$:*WN - estricto* , las distribuciones de los estadísticos serán igualmente válidas como aproximaciones para n suficientemente grande. El caso en que las variables ε_t sean de varianza infinita (habitual en las aplicaciones al mercado financiero) será tratado en el Cap. 4.

Phillips-Perron[37], por su parte, propusieron una generalización de estos tests de raíces unitarias que admite la heterocedasticidad del proceso $\{\varepsilon_t\}$.

Ejemplo 2.3:

Sea un proceso $\{Y_t\}$:$AR(1)$ de parámetero mensual, cuya trayectoria para el período [ENE-2000,DIC-2004] figura en el **GRAFICO 3.6** y cuyo estimador de la función de autocorrelaciones figura en el **CUADRO 3.3**.

Considerando una representación general sin restricciones de la forma:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$$

se obtuvo el siguiente modelo:

$$\nabla Y_t = \underset{(1.184)}{2.523} - \underset{(0.041)}{0.013}t - \underset{(0.068)}{0.091}Y_{t-1} \qquad \left(\sum_{t=2}^{60}\varepsilon_t^2 = 847.424\right)$$

a partir del cual se construyó un test para contrastar la hipótesis H_0:$\phi_1 - 1 = 0$ con respecto a la alternativa H_1:$\phi_1 - 1 < 0$. El correspondiente estadístico $T = \dfrac{\hat{\phi}_1 - 1}{\hat{\sigma}(\hat{\phi}_1)}$ asumió el valor -1.329 el cual, comparado con el valor crítico $\tau_\tau = -3.486$ (para un $\alpha = 0.05$) permite concluir que **no existen** razones suficientes para rechazar H_0.

Siguiendo, entonces, la secuencia comentada precedentemente se construyó un test para contrastar la hipótesis H_0:$a_1 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa H_1:$a_1 \neq 0 / \phi_1 - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\dfrac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\dfrac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

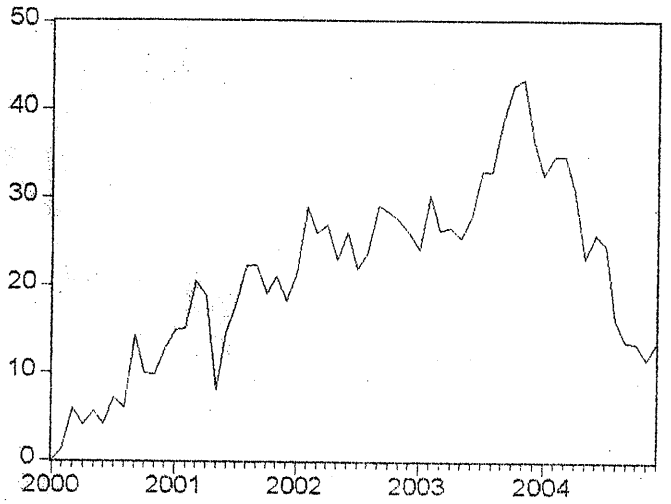
que surge de la comparación de las sumas de los cuadrados de los residuos correspondientes, respectivamente, a la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.518)}{0.226}\left(\sum_{t=2}^{60}\hat{\varepsilon}_t^2 = 917.541\right)$ y del modelo sin restricciones:

$$\nabla Y_t = \underset{(1.059)}{1.787} - \underset{(0.030)}{0.050}t \left(\sum_{t=2}^{60}\hat{\varepsilon}_t^2 = 874.145\right)$$

---

[37] . Ver Phillips, P.C.B.; Perron, P. (1988).

GRAFICO 3.6: Trayectoria del proceso $\{Y_t\}$



CUADRO 3.3: Estimación de la función de autocorrelaciones de $\{Y_t\}$

Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.887	0.887	49.576	0.000
		2	0.789	0.013	89.506	0.000
		3	0.721	0.090	123.46	0.000
		4	0.650	-0.038	151.55	0.000
		5	0.579	-0.028	174.21	0.000
		6	0.496	-0.100	191.16	0.000
		7	0.426	0.002	203.91	0.000
		8	0.329	-0.187	211.66	0.000
		9	0.271	0.121	217.02	0.000
		10	0.222	-0.025	220.69	0.000
		11	0.176	0.034	223.04	0.000
		12	0.152	0.062	224.82	0.000
		13	0.116	-0.044	225.89	0.000
		14	0.089	-0.009	226.52	0.000
		15	0.086	0.112	227.14	0.000

asumió el valor:

$$\Phi = \frac{\frac{1}{2}(917.837 - 874.145)}{\frac{1}{57}(874.145)} = 1.415$$

Este valor, comparado con el valor crítico $\tau_{\beta\tau} \approx 2.8$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar H_0.

De la misma forma, se construyó un test para contrastar la hipótesis $H_0{:}a_0 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa $H_1{:}a_0 \neq 0 / \phi_1 - 1 = 0$. El correspondiente estadístico, que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.015)}{0.006t} \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 917.837 \right)$ y del modelo sin restricciones mencionado más arriba, asumió el valor:

$$\Phi = \frac{\frac{1}{2}(917.837 - 874.145)}{\frac{1}{57}(874.145)} = 1.424$$

Este valor comparado con el valor crítico $\tau_{\alpha\tau} \approx 3.14$ (para un $\alpha = 0.05$) permite concluir que no existen razones suficientes para rechazar H_0.

A fin de confirmar las conclusiones anteriores se construyó el siguiente test para contrastar la hipótesis $H_0{:}a_0 = a_1 = \phi_1 - 1 = 0$. En este caso, el estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2 \right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR} \right]}{\frac{1}{n-k} \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR}}$$

que surge de la comparación de la suma $\sum_{t=2}^{60} (\nabla Y_t)^2 = 933.360$ y la suma de los cuadrados de los residuos obtenidos de la aplicación del modelo general ya mencionado, asumió el valor:

$$\Phi = \frac{\frac{1}{3}(933.360 - 847.424)}{\frac{1}{56}(847.424)} = 1.893$$

Este valor, comparado con el valor crítico $\phi_2 \approx 5.13$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar H_0.

Los resultados anteriores indican (con una significatividad del 5%): **i)** que la ecuación correspondiente al test de Dickey-Fuller no posee término independiente ni tendencia determinística; **ii)** que el proceso $\{Y_t\}$ presenta una raíz unitaria.

Ejemplo 3.3:

Sea un proceso $\{Y_t\}$: $AR(1)$ de parámetro mensual, cuya trayectoria para el período [ENE-2000,DIC-2004] figura en el **GRAFICO 3.7** y cuyo estimador de la función de autocorrelaciones figura en el **CUADRO 3.4.**

Considerando una representación general de la forma:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1 - 1) Y_{t-1} + \varepsilon_t$$

se obtuvo el siguiente modelo:

$$\nabla Y_t = \underset{(0.8123)}{4.2814} - \underset{(0.0207)}{0.0361} t - \underset{(0.0255)}{0.0601} Y_{t-1} \qquad \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 217.6815 \right)$$

a partir del cual se construyó un test para contrastar la hipótesis de no-estacionariedad H_0:$\phi_1 - 1 = 0$ con respecto a la alternativa H_1:$\phi_1 - 1 < 0$. El correspondiente estadístico T asumió el valor -2.3546 el cual, comparado con el valor crítico $\tau_\tau = -3.4862$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis de existencia de una raíz unitaria.

De acuerdo con este decisión y siguiendo la secuencia mencionada en las páginas anteriores, se construyó un test para contrastar la hipótesis H_0:$a_1 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa H_1:$a_1 \neq 0 / \phi_1 - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2 \right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR} \right]}{\frac{1}{n-k} \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.307)}{0.723} \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 322.161 \right)$ y del modelo sin restricciones

$$\nabla Y_t = \underset{(0.554)}{0.2881} - \underset{(0.016)}{0.070} t \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 239.232 \right)$$

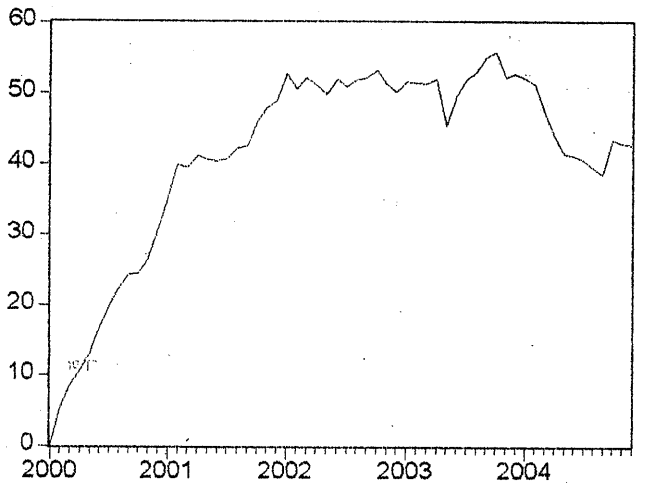
asumió el valor:

$$\Phi = \frac{\frac{1}{2}(322.161 - 239.232)}{\frac{1}{57}(239.232)} = 9.880$$

Este valor, comparado con el valor crítico $\tau_{\beta\tau} \approx 2.81$ (para un $\alpha = 0.05$) permite concluir que **existen** razones suficientes para rechazar H_0.

De la misma forma, se construyó un test para contrastar la hipótesis H_0:$a_0 = 0 / \phi_1 - 1 = 0$ con respecto a la alternativa H_1:$a_0 \neq 0 / \phi_1 - 1 = 0$. El correspondiente estadístico:

GRAFICO 3.7: Trayectoria del proceso $\{Y_t\}$



CUADRO 3.4: Correlograma de $\{Y_t\}$

Date: 08/07/04 Time: 08:...
Sample: 2000:01 2004:12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.909	0.909	52.104	0.000
		2	0.828	0.008	96.046	0.000
		3	0.749	-0.029	132.62	0.000
		4	0.669	-0.049	162.31	0.000
		5	0.593	-0.021	186.11	0.000
		6	0.524	-0.013	204.99	0.000
		7	0.458	-0.020	219.69	0.000
		8	0.399	-0.001	231.09	0.000
		9	0.342	-0.030	239.62	0.000
		10	0.280	-0.071	245.44	0.000
		11	0.216	-0.059	248.97	0.000
		12	0.163	0.014	251.04	0.000
		13	0.120	0.016	252.18	0.000
		14	0.090	0.044	252.84	0.000
		15	0.058	-0.044	253.11	0.000

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.0091)}{0.0018}t \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 352.7842\right)$ y del modelo sin restricciones $\nabla Y_t = \underset{(0.554)}{2.881} - \underset{(0.016)}{0.070}t \left(\sum_{t=2}^{60} \hat{\varepsilon}_t^2 = 239.232\right)$ asumió el valor:

$$\Phi = \frac{\frac{1}{2}(352.784 - 239.232)}{\frac{1}{57}(239.232)} = 13.528$$

Este valor, comparado con el valor crítico $\tau_{\alpha\tau} \approx 3.14$ (para un $\alpha = 0.05$), permite concluir que **existen** razones suficientes para rechazar H_0.

De los resultados anteriores se puede concluir (con una confiabilidad del 95%): **i)** que la ecuación correspondiente al test de Dickey-Fuller posee término independiente y tendencia determinística y **ii)** que el proceso $\{Y_t\}$ presenta una raíz unitaria.

El test de Dickey-Fuller no es adecuado para detectar la presencia de tendencia determinística. Para distinguir entre la presencia de tendencia estocástica o determinística, Durbin-Hausmann[38] propusieron un test similar al de Dickey-Fuller (y uniformemente más potente que éste)[39]que contrasta la hipótesis nula $H_0:\{Y_t\}:I(1)$ con respecto a la hipótesis alternativa $H_1:\{Y_t\}:I(0)+$ *tendencia*. Es decir, dado un proceso que admite una representación de la forma:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1 - 1)Y_{t-1} + \varepsilon_t = a_0 + a_1 t + \phi_1^* Y_{t-1} + \varepsilon_t$$

el test consiste en contrastar la hipótesis $H_0:\phi_1^* = 0$ con respecto a la alternativa $H_1:\phi_1^* < 0$, utilizando un estadístico de la forma $DH = \frac{\left(\hat{\phi}_1^{**} - \hat{\phi}_1^*\right)^2}{\sigma^2\left(\hat{\phi}_1\right)}$, donde $\hat{\phi}_1^{**}$ denota el estimador del coeficiente que acompaña al término $AR(1)$ en la representación $\nabla Y_t = a_0 + \phi_1^{**} Y_{t-1} + \varepsilon_t$. Los límites de la zona crítica son los correspondientes a la tabla de Dickey-Fuller[40].

---

[38] . Ver Choi, I. (1992).

[39] . Para una evidencia de esta propiedad utilizando métodos de Monte Carlo, ver Alexander, C.O.; Rendell, H. (1995).

[40] . Este test fue generalizado por Schmidt-Phillips para el caso de una tendencia polinómica (ver Schmidt, P.; Phillips, P.C.B. (1992)). Cochrane, J.H. (1991) obtuvo resultados analíticos que demuestran que los tests para detectar la presencia de tendencia determinística pueden poseer una potencia arbitrariamente baja.

3.3.2.- El caso de los procesos AR(p)

Sea ahora un proceso $\{Y_t\}$ que admite una representación $AR(p)$ de la forma:

$$\Phi(B)Y_t = \left(1 - \phi_1 B - ... - \phi_p B^p\right)Y_t = \varepsilon_t$$

Sumando y restando términos de la siguiente forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} +$$
$$+ \left[\left(\phi_2 Y_{t-1} - \phi_2 Y_{t-1}\right) + ... + \left(\phi_p Y_{t-1} - \phi_p Y_{t-1}\right)\right] + \left[\left(\phi_3 Y_{t-2} - \phi_3 Y_{t-2}\right) + ... + \left(\phi_p Y_{t-2} - \phi_p Y_{t-2}\right)\right] +$$
$$+ ... + \left(\phi_p Y_{t-(p-1)} - \phi_p Y_{t-(p-1)}\right) + \varepsilon_t$$

la representación original puede ser expresada como:

$$Y_t = \left(\phi_1 + \phi_2 + ... + \phi_p\right)Y_{t-1} - \left(\phi_2 + \phi_3 + ... + \phi_p\right)\left(Y_{t-1} - Y_{t-2}\right) - \left(\phi_3 + \phi_4 + ... + \phi_p\right)\left(Y_{t-2} - Y_{t-3}\right) -$$
$$- ... - \phi_p\left(Y_{t-(p-1)} - Y_{t-p}\right) + \varepsilon_t = \phi_1^* Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

(donde $\phi_j^* = \sum_{i=j}^{p} \phi_i$). Dado que, según lo expresado en la Sec. 2.10.1, si $\sum_{i=1}^{p} \phi_i = 1$, se puede asegurar que el polinomio $\Phi(B)$ posee por lo menos una raíz unitaria, plantear la hipótesis nula de la presencia de una raíz unitaria es equivalente a plantear la hipótesis nula $H_0: \phi_1^* = 1$.

Igual que en el caso anterior, si se supone que el orden p es conocido, ϕ_1^* puede ser estimado en forma consistente utilizando el criterio de optimización de los cuadrados mínimos pero, como se demostró en la Sec. 3.1, este estimador no permite testear la hipótesis nula $\phi_1^* = 1$ con respecto a la alternativa $\phi_1^* < 1$, a partir del estadístico T, por lo que, también en esta situación conviene aplicar el test "t" sobre el coeficiente de la variable Y_{t-1}, en la expresión:

$$\nabla Y_t = \left(\phi_1^* - 1\right)Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

De modo que la hipótesis nula $\phi_1^* = 1$ se transforma en la hipótesis $H_0: \phi_1^* - 1 = 0$ y se contrasta con la hipótesis alternativa $H_1: \phi_1^* - 1 < 0$ mediante la utilización del estadístico [41]:

$$T = \frac{\hat{\phi}_1^* - 1}{\hat{\sigma}\left(\hat{\phi}_1\right)}$$

Si n es suficientemente grande, el contraste de las hipótesis se realiza considerando los mismos valores críticos $\tau, \tau_\mu, \tau_\tau$ (debe tenerse en cuenta que, para muestras chicas los porcentajes de la distribución de

---

[41] . Si el proceso $\{Y_t\}$ estuviera expresado para períodos menores que el año, debería ser testeado respecto de la existencia de raíces unitarias estacionales.

Dickey-Fuller aumentada no son los mismos que en el caso particular del $AR(1)$).

De la misma forma que en el caso del proceso $AR(1)$, la representación más general es aquella en la que el proceso $\{Y_t\}$ no-estacionario posee además un término independiente y una componente correspondiente a una tendencia lineal determinística. Su expresión en términos de las primeras diferencias será, entonces, de la forma[42]:

$$\nabla Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

Debe tenerse en cuenta que, igual que en el caso del $AR(1)$: **i)** si se incluye un término independiente y una componente tendencia, y el proceso posee una raíz unitaria, una constante no-nula y un coeficiente no-nulo para el regresor t entonces, bajo la hipótesis nula de una raíz unitaria, el estadístico τ converge a una distribución del tipo $N(0;1)$ [43] y **ii)** la inclusión de regresores irrelevantes en la ecuación reduce la potencia del test (en particular, los términos autorregresivos que hacen que el número de observaciones utilizables en la estimación disminuya).

Las secuencias de tests desarrollados en la sección precedente rigen también para este caso.

3.3.3.- Los procesos ARMA(p,q)

Said, S.; Dickey, D. (1984)(1985) demostraron que un proceso $\{Y_t\}: ARIMA(p,1,q)$ puede ser aproximado por una representación $ARIMA\left(p^*,1,0\right)$, tal que $p^* \leq n^{1/3}$, lo que permite concluir que, bajo la hipótesis nula de una raíz unitaria, el test es válido aún si el proceso $\{Y_t\}$ tuviera una componente MA. Es decir, si la representación de $\{Y_t\}$ fuera[44]:

$$\nabla Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \Theta(B)\varepsilon_t$$

entonces, podría ser aproximada de la siguiente forma:

$$\nabla Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} - \sum_{j=1}^{p^*} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

---

42 . En 1991, MacKinnon, J.G. (en Engle, R.F.; Granger, C.W.J. (eds.)(1991)) (1996) obtuvo valores críticos para la distribución de Dickey-Fuller para cualquier valor de n y cualquier número de variables en la ecuación de Dickey-Fuller, a partir de un conjunto mayor de simulaciones.

43 . Es decir, como se mencionó en la sección precedente, si se sabe que el proceso no incluye términos determinísticos, el estadístico τ obedece a una distribución de Dickey-Fuller, por el contrario, si se sabe que el proceso incluye un término independiente y un término de tendencia determinística, la hipótesis $H_0: \phi_1^* - 1 = 0$ puede ser testeada utilizando una distribución Normal.

44 . Para procesos expresados en períodos menores que el año, se suele adoptar como valor inicial de p^* la frecuencia estacional más 12 (ver Ng, S.; Perron, P. (1995)). En lo que se refiere a los datos del mercado financiero -que, en general, no están significativamente afectados por variaciones estacionales- se suele utilizar $p^* = \left[n^{1/4}\right]$ (donde $[\bullet]$ denota la parte entera del argumento). Ver Schwert, G.W. (1987), Diebold, F.X.; Nerlove, M. (1990).

De la misma forma que en los precedentes, en este caso y en el caso en que el orden de autorregresividad, que se supone finito, sea desconocido y aproximable por p^*, se aplica la secuencia de tests mencionada en la Sec. 3.3.1 partiendo de la ecuación sin restricciones precedente.

El problema que permanece es el de la selección del orden de autorregresividad adecuado. Esta cuestión es sumamente importante, en la medida que un valor de p^* demasiado grande producirá una disminución del número de grados de libertad (aumento del número de coeficientes a estimar y disminución del tamaño del conjunto de información) que, obviamente genera la reducción en la potencia del test. Por el contrario, un valor de p^* demasiado bajo no permitirá una definición adecuada del proceso residual y generará un sesgo hacia el rechazo de la hipótesis nula cuando ésta es verdadera.

Una solución a este dilema consiste en tomar un orden p^* suficientemente grande y utilizar los resultados de los tests sobre la significatividad individual y conjunta de los coeficientes. Si, por ejemplo, el estadístico T correspondiente al coeficiente que acompaña al término autorregresivo de orden p^* indicara, con un cierto grado de probabilidad, que es no-significativo, se debería a estimar nuevamente el modelo con un orden de autorregresividad $p^* - 1$ y así sucesivamente hasta hallar un coeficiente significativo[45]. Una segunda forma de solución consiste en aplicar los criterios de información mencionados en la Sec. 2.12.4.

A fin de evitar el agregado de términos autorregresivos ∇Y_{t-j} para asegurar el cumplimiento de la condición $\{\varepsilon_t\}$:*WN*, Phillips, P.C.B. (1987)(1987) propuso una forma alternativa de tratamiento del caso $ARMA(p,q)$ (generalizada por Perron, P. (1988))[46], consistente en realizar una modificación no-paramétrica de los estadísticos que considera la presencia de autocorrelación (y, eventualmente, de heterocedasticidad) en el proceso $\{\varepsilon_t\}$ [47], "a posteriori" de la estimación de la representación $Y_t = a_0 + \phi_1 Y_{t-1} + \varepsilon_t$. Supóngase que en esta representación el proceso $\{\varepsilon_t\}$ sea asíntoticamente incorrelacionado, es decir, que cumpla la siguientes condiciones: **i)** que $E(\varepsilon_t) = 0\,(t = 1,2,...)$; **ii)** que sea posible hallar un $\beta > 2$ para el cual $\sup_t E\left(|\varepsilon_t|^{\beta}\right) < \infty$ (como ya se demostró, ésta es condición necesaria y suficiente para asegurar que $\sigma^2(\varepsilon_t)$ existe y que existe, al menos, un momento de orden superior de ε_t, pero no implica que estos momentos sean constantes[48]); **iii)** que $\sigma_{\varepsilon}^{*2} = \lim_{n\to\infty} E\left[\frac{1}{n}\left(\sum_{t=1}^{n}\varepsilon_t\right)^2\right]$ existe y es positiva (ésta es condición necesaria para poder asegurar que la distribución de los ε_t no es degenerada); **iv)** que el proceso $\{\varepsilon_t\}$ cumple la condición α-mixing con coeficientes α_j tales que $\sum_{j=1}^{\infty}\alpha_j^{1-2/\beta} < \infty$ (donde el coeficiente β, que figura tanto en la definición de los momentos de orden superior de ε_t como en la condición α-mixing, indica la existencia de una relación

---

[45]. Una vez determinado el orden de autorregresividad óptimo se debe realizar un análisis sobre las condiciones de WN-Normal del proceso representativo de las perturbaciones $\{\varepsilon_t\}$.

[46]. Ver también Phillips, P.C.B.; Perron, P. (1988)

[47]. Yap, S.F.; Reinsel, G.C. (1995) propusieron tests alternativos para la detección de raíces unitarias en una representación ARMA, basados en cocientes de verosimilitudes.

[48]. Obsérvese que, de acuerdo a lo mencionado más arriba, esta generalización admite la presencia de heterocedasticidad.

entre la dependencia de las perturbaciones ε_t y la posible presencia de "outliers")[49].

A partir de estas condiciones, el único caso en que se verificará la igualdad $\sigma_\varepsilon^{*2} = \sigma_\varepsilon^2 = \lim_{n\to\infty} E\left(\frac{1}{n}\sum_{t=2}^{n}\varepsilon_t^2\right)$ es aquél en que $\{\varepsilon_t\}:WN$. En cualquier otro caso, de acuerdo a la demostración de Phillips, es necesario aplicar a los estadísticos de Dickey-Fuller una corrección no-paramétrica, a fin de poder contrastar la hipótesis $H_0:\phi_1 - 1 = 0$ con respecto a la alternativa $H_1:\phi_1 - 1 < 0$, en la representación $\nabla Y_t = a_0 + (\phi_1 - 1)Y_{t-1} + \varepsilon_t$, cuando la verdadera representación de $\{Y_t\}$ admite perturbaciones que cumplen las condiciones mencionadas más arriba.

Otro test asintóticamente válido puede obtenerse utilizando el **estadístico z de Phillips**[50]:

$$Z(\tau_\mu) = T(\hat{\phi}_1 - 1) - \frac{1}{2}(\hat{\sigma}_\varepsilon^{*2} - \hat{\sigma}_\varepsilon^2)\frac{1}{\frac{1}{n^2}\sum_{t=2}^{n}\left(Y_{t-1} - \overline{Y}_{(n-1)}\right)^2}$$

donde: **i)** $\hat{\sigma}_\varepsilon^{*2} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 + \frac{2}{n}\sum_{j=1}^{p^*}\left(1 - \frac{j}{p^*+1}\right)\sum_{t=j+1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$ y $\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2$ son estimadores consistentes de las varianzas $\hat{\sigma}_\varepsilon^{*2}$ y σ_ε^2, respectivamente (siendo $\frac{w_j}{p^*} = 1 - \frac{j}{p^*+1}$ un conjunto triangular de ponderaciones que aseguran que estos estimadores sean positivos[51]y suponiendo que el orden p^* puede ser aproximado por $\left[n^{1/4}\right]$) y **ii)** $\overline{Y}_{(n-1)} = \frac{1}{n-1}\sum_{t=1}^{n-1}Y_t$.

Una aproximación alternativa puede ser obtenida utilizando el siguiente estadístico:

$$z^*(\tau_\mu) = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^{*2}}\tau_\mu - \frac{n}{2}(\hat{\sigma}_\varepsilon^{*2} - \hat{\sigma}_\varepsilon^2)\frac{1}{\hat{\sigma}_\varepsilon^*\sqrt{\sum_{t=2}^{n}\left(Y_{t-1} - \overline{Y}_{(n-1)}\right)^2}}$$

Bajo el supuesto que la hipótesis nula de no-estacionariedad sea verdadera, los dos estadísticos z mencionados convergen a la misma distribución que $(n-1)(\phi_1 - 1)$ y τ_μ, respectivamente. Por lo tanto, en ambos casos, pueden ser utilizados los mismos valores críticos que en el test de Dickey-Fuller[52].

---

[49]. Debe recordarse que la condición α -mixing" sobre el proceso $\{\varepsilon_t\}:I(0)$, permite asegurar que este proceso es ergódico y que los coeficientes α_j miden la intensidad y la extensión de la dependencia temporal en $\{\varepsilon_t\}$ (ver Sec. 1.4).

[50]. Ver Phillips, P.C.B. (1987).

[51]. Ver Newey, W.K.; West, K.D. (1987).

[52]. De acuerdo con Phillips, P.C.B. (1987), p^* debe aumentar con n, pero a una tasa inferior a $n^{1/4}$. De los trabajos sobre simulación de Monte Carlo debidos a Schwert, G.W. (1987), Phillips, P.C.B.; Perron, P. (1988) y Kim, K.; Schmidt, P. (1990) no surge ninguna regla para la selección de p^*. De la misma forma que en el caso comentado en la página precedente, aquí se ha

Esta aproximación no-paramétrica también es aplicable en el caso de una representación con tendencia determinística lineal de la forma:

$$Y_t = a_0 + a_1 t + \varphi_1 Y_{t-1} + \varepsilon_t$$

donde, igual que en el caso anterior, $\{\varepsilon_t\}$ presenta autocorrelaciones. Un test sintóticamente válido puede ser obtenido utilizando un estadístico de la forma:

$$z(\tau_\tau) = \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^{*2}}\tau_\tau - \frac{\frac{n^3}{2}\left(\hat{\sigma}_\varepsilon^{*2} - \hat{\sigma}_\varepsilon^2\right)}{4\hat{\sigma}_\varepsilon^{*}\sqrt{3\left[\frac{n^2\left(n^2-1\right)}{12}\sum_{t=2}^{n} Y_{t-1}^2 - n\left(\sum_{t=2}^{n} tY_{t-1}\right)^2 + n(n+1)\sum_{t=2}^{n} tY_{t-1}\sum_{t=2}^{n} Y_{t-1} - \frac{n(n+1)(2n+1)}{6}\left(\sum_{t=2}^{n} Y_{t-1}\right)^2\right]}}$$

Otras aproximaciones no-paramétricas que merecen ser mencionadas son: **i)** el método de la variable instrumental de Hall, A. (1989)[53]; **ii)** la aproximación de Bhargava, A. (1986); **iii)** la aproximación que utiliza la varianza de largo plazo de $\{\nabla Y_t\}$, propuesta por Phillips, P.C.B.; Ouliaris, S. (1988)[54]; **iv)** la familia de métodos conocida como de adición de variable[55]; **v)** el test DF-GLS propuesto por Elliot, G.; Rothenberg, T.J.; Stock, J.H. (1996)[56].

3.3.4.- La raíces unitarias múltiples

Los tests mencionados precedentemente fueron desarrollados bajo el supuesto que $\{Y_t\}$ admite, como máximo, una raíz unitaria (es decir, que es un proceso $I(1)$). Si la hipótesis nula de la existencia de una raíz unitaria no es rechazada, entonces, surge la necesidad de testear si el proceso $\{Y_t\}$ admite una segunda raíz unitaria (es decir, si es un proceso $I(2)$). En otros términos, si se acepta la hipótesis de no-estacionariedad de primer orden será necesario testear la no-estacionariedad del proceso $\{\nabla Y_t\}$, si se rechazara esta hipótesis se podría concluir que $\{Y_t\}:I(1)$, si se aceptara esta hipótesis se podría concluir que $\{Y_t\}:I(2)$ y, así sucesivamente.

Lógicamente este procedimiento de testear los órdenes de integración de menor a mayor debería continuar hasta que la hipótesis de no-estacionariedad fuera rechazada. Pero, supóngase que $\{Y_t\}:I(2)$, al aplicar entonces el test de Dickey-Fuller aumentado utilizando la representación:

---

propuesto $n^{1/4}$ como aproximación conveniente en las aplicaciones financieras. Modificaciones a esta aproximación no-paramétrica pueden ser halladas en Perron, P.; Ng, S. (1996).

53. Ver también Pantula, S.G.; Hall, A. (1991).

54. Ver también Stock, J.H. (en Engle, R.F.; McFadde, D.L. (eds.) (1994))

55. Ver Park, J.Y. (en Rhodes, G.F.; Fomby, T.B. (eds.)(1990)).

56. Ver también Stock, J.H. (1994).

$$\nabla Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

la hipótesis alternativa con la que se contrasta la hipótesis de no-estacionariedad es que $\{Y_t\}: I(0)$ cuando lo que realmente se necesita es testear la hipótesis de que $\{\nabla Y_t\}$ es no-estacionario con respecto a la hipótesis de que es estacionario.

De modo que la secuencia de tests debe proceder desde el orden de integración mayor (en la práctica, desde $\{Y_t\}: I(2)$) y reducirlo cada vez hasta que existan razones suficientes para rechazar la hipótesis nula de no-estacionariedad.

En este sentido, Dickey, D.A.; Pantula, S. (1987) propusieron una sucesión de tests según los cuales, si se supone que $\{Y_t\}$ admite, como máximo, dos raíces unitarias, primero se debe testear la hipótesis nula de dos raíces unitarias contra la hipótesis alternativa de una raíz unitaria, utilizando el estadístico T sobre el coeficiente que acompaña a $\phi_1^{**} - 1$ de la variable ∇Y_{t-1}, a partir de la expresión:

$$\nabla^2 Y_t = a_0 + a_1 t + \left(\phi_1^{**} - 1\right) \nabla Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^{**} \nabla^2 Y_{t-j} + \varepsilon_t$$

Si no existen razones suficientes para rechazar la hipótesis nula $H_0: \phi_1^{**} - 1 = 0$, se concluye que $\{Y_t\}: I(2)$. Si, por el contrario, existen razones suficientes para rechazar esta hipótesis, se debe testear la hipótesis nula que $\{Y_t\}$ presenta solamente una raíz unitaria utilizando el estadístico τ sobre los coeficientes ϕ_1^{**} y ϕ_2^{**} correspondientes a las variables ∇Y_{t-1} e Y_{t-1}, respectivamente, a partir del modelo:

$$\nabla Y_t = \hat{a}_0 + \hat{\phi}_1^{**} \nabla Y_{t-1} + \hat{\phi}_2^{**} Y_{t-1} - \sum_{j=1}^{p-1} \hat{\phi}_{j+1}^{*} \nabla Y_{t-j}$$

Rechazar la hipótesis nula $\phi_1^{**} - 1 = 0$ con respecto a la hipótesis alternativa $\phi_1^{**} - 1 < 0$ implica aceptar que $\{Y_t\}$ es estacionario.

Este razonamiento es generalizable al caso en que se supone que el proceso $\{Y_t\}$ admite como máximo d raíces unitarias. En primer lugar se testea la hipótesis nula que el proceso presenta como máximo d raíces unitarias con respecto a la hipótesis alternativa que presenta como máximo $d - 1$ raíces unitarias, utilizando el estadístico T sobre el coeficiente $\phi_1^{(d)*} - 1$ que acompaña a la variable $\nabla^{d-1} Y_{t-1}$ a partir de la expresión:

$$\nabla^d Y_t = \left(\phi_1^{(d)*} - 1\right) \nabla^{d-1} Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^{(d)*} \nabla^d Y_{t-j} + \varepsilon_t$$

Si no existen razones suficientes para rechazar la hipótesis nula $\phi_1^{**} - 1 = 0$, entonces se acepta la hipótesis de que $\{Y_t\}: I(d)$. Si se rechaza la hipótesis nula, entonces se debe testear la hipótesis nula que el proceso presenta como máximo $d - 1$ raíces unitarias con respecto a la alternativa que presenta como máximo $d - 2$ raíces unitarias, utilizando el estadístico T sobre los coeficientes ϕ_1^{**} y ϕ_2^{**} a partir del modelo:

$$\nabla^d Y_t = \hat{\phi}_1^{(d)**}\nabla^{d-1}Y_{t-1} + \hat{\phi}_2^{(d)**}Y_{t-1} - \sum_{j=1}^{p-1}\hat{\phi}_{j+1}^{(d)**}\nabla^d Y_{t-j}$$

Si ϕ_1^{**} y ϕ_2^{**} resultan significativos (ambos negativos), se rechaza la hipótesis de la existencia de $d-1$ raíces unitarias. El procedimiento continúa hasta testear la significatividad de los coeficientes $\phi_1^{(d)**}, \phi_2^{(d)**}, \ldots, \phi_d^{(d)**}$ a partir de la expresión:

$$\nabla^d Y_t = \phi_1^{(d)**}\nabla^{d-1}Y_{t-1} + \phi_2^{(d)**}\nabla^{d-2}Y_{t-1} + \phi_3^{(d)**}\nabla^{d-3}Y_{t-1} + \ldots + \phi_d^{(d)**}Y_{t-1} - \sum_{j=1}^{p-1}\phi_{j+1}^{(d)*}\nabla^d Y_{t-j} + \varepsilon_t$$

Debe tenerse en cuenta: **i)** que la potencia de los tests de raíces unitarias, si bien depende del número de observaciones, no está relacionada en forma directamente proporcional con el número de observaciones contenidas en la serie cronológica (dada una extensión, el aumento en la frecuencia de las observaciones, sólo produce un incremento marginal en la potencia del test)[57]y **ii)** que estos tests no son muy potentes con respecto a hipótesis alternativas relevantes (De Jong, D.N.; Nankervis, J.C.; Savin, N.E.; Whiteman, C.H. (1981) demuestraron que los tests de Dickey-Fuller tienen baja potencia con respecto a alternativas AR con raíces próximas a la unidad y Diebold, F.X.; Rudebusch, G.D. (1990) demuestraron que estos tests son de baja potencia con respecto a alternativas de procesos con orden de diferenciación fraccionario[58]).

Ejemplo 3.5:

Sea un proceso $\{Y_t\}$: $AR(2)$ de parámetro mensual cuya trayectoria para el período [ENE-2001;DIC-2004] figura en el **GRAFICO 3.8**.

Considerando una representación general de la forma:

$$\nabla^2 Y_t = a_0 + a_1 t + \left[\left(\phi_1 + \phi_2\right) - 1\right]\nabla Y_{t-1} - \phi_2\nabla^2 Y_{t-1} + \varepsilon_t =$$

$$= a_0 + a_1 t + \left(\phi_1^* - 1\right)\nabla Y_{t-1} - \phi_2\nabla^2 Y_{t-1} + \varepsilon_t$$

se obtuvo el siguiente modelo:

$$\nabla^2 Y_t = \underset{(0.822)}{0.962} + \underset{(0.030)}{0.009}t - \underset{(0.260)}{1.478}\nabla Y_{t-1} + \underset{(0.157)}{0.072}\nabla^2 Y_{t-1} \qquad \left(AIC = 4.8650, SBC = 5.0256\right)$$

De la aplicación del test t de significatividad individual de los coeficientes se puede concluir en forma inmediata que, con una significatividad del 5%, existen razones suficientes para no rechazar la hipótesis $H_0: \phi_2^{**} = 0$, con respecto a la alternativa $H_1: \phi_2^{**} \neq 0$. De modo que se construyó el modelo alternativo:

$$\nabla^2 Y_t = \underset{(0.822)}{1.120} + \underset{(0.029)}{0.002}t - \underset{(0.145)}{1.379}\nabla Y_{t-1} \qquad \left(AIC = 4.8252, SBC = 4.9444\right)$$

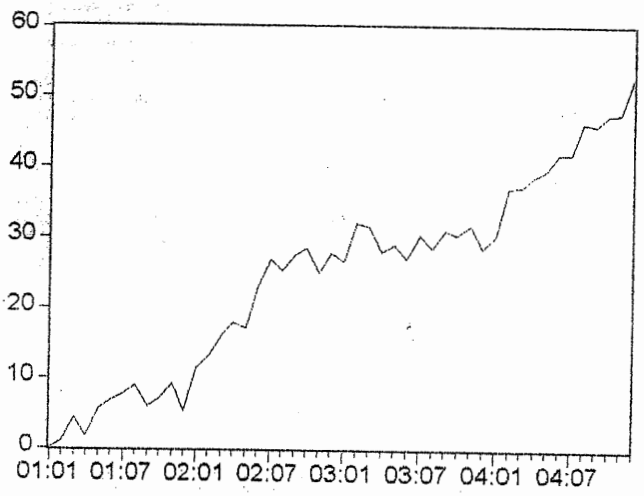
A partir de este modelo se construyó un test para contrastar la hipótesis de no-estacionariedad

---

[57] . Ver Shiller, R.J.; Perron, P. (1985), Perron, P. (1991), Pierse, R.G.; Snell, A.J. (1995), De Jong, D.N.; Nankervis, J.C.; Savin, N.E.; Whiteman, C.H. (1992), Choi, I.; Chung, B.S. (1995).

[58] . Estos procesos, conocidos como ARFIMA, serán analizados en el Cap. 5.

GRAFICO 3.8: Trayectora del proceso $\{Y_t\}$



$H_0: \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: \phi_1^* - 1 < 0$. El correspondiente estadístico T asumió el valor -9.522 el cual, comparado con el valor crítico $\tau_\tau = -3.5112$ (para un $\alpha = 0.05$), permite concluir que existen razones suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{\nabla Y_t\}$. Es decir, este resultado permite concluir (con una confiabilidad del 95%) que $\{\nabla Y_t\}: I(0)$ o, lo que es lo mismo, que el proceso $\{Y_t\}$ posee, como máximo, una raíz unitaria.

A fin de decidir si $\{Y_t\}$ es o no $I(1)$ se consideró la representación general de la forma:

$$\nabla Y_t = a_0 + a_1 t + \left[\left(\phi_1 + \phi_2\right) - 1\right] Y_{t-1} - \phi_2 \nabla Y_{t-1} + \varepsilon_t =$$

$$= a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} - \phi_2 \nabla Y_{t-1} + \varepsilon_t$$

obteniéndose el siguiente modelo:

$$\nabla Y_t = \underset{(0.836)}{1.586} + \underset{(0.086)}{0.155} t - \underset{(0.116)}{0.219} Y_{t-1} - \underset{(0.152)}{0.273} \nabla Y_{t-1} \qquad \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 271.530\right)$$

a partir del cual se construyó un test para contrastar la hipótesis de no estacionariedad $H_0: \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: \phi_1^* - 1 < 0$. El correspondiente estadístico T asumió el valor -1.887 el cual, comparado con el valor crítico $\tau_\tau = -3.5088$ (para un $\alpha = .05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{Y_t\}$.

De acuerdo con esta decisión y siguiendo la secuencia de Perron analizada en la Sec.3.3.1, se construyó un test para contrastar la hipótesis $H_0: a_1 = 0 / \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: a_1 \neq 0 / \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.398)}{1.156} - \underset{(0.143)}{0.379}\nabla Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 294.578\right)$ y de la aplicación del modelo sin restricciones:

$$\nabla Y_t = \underset{(0.847)}{1.119} + \underset{(0.029)}{0.001}t - \underset{(0.145)}{0.379}\nabla Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 294.561\right)$$

asumió el valor:

$$\Phi = \frac{\frac{1}{2}(294.578 - 294.561)}{\frac{1}{43}(294.561)} = 0.001$$

Este valor, comparado con el valor crítico $\tau_{\beta\tau} = 2.81$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar H_0.

De la misma forma, se construyó para contrastar la hipótesis $H_0: a_0 = 0 / \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: a_0 \neq 0 / \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.014)}{0.035}t - \underset{(0.144)}{0.347}\nabla Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 306.502\right)$ y de la aplicación del modelo general sin restricciones, asumió el valor:

$$\Phi = \frac{\frac{1}{2}(306.502 - 271.530)}{\frac{1}{42}(271.530)} = 0.851$$

Este valor, comparado con el valor crítico $\tau_{\alpha\tau} = 3.4$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar H_0.

A fin de confirmar las conclusiones anteriores, se construyó un test a fin de contrastar la hipótesis $H_0: a_0 = a_1 = \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.148)}{-0.262}\nabla Y_{t-1}\left(\sum_{t=2}^{47}\hat{\varepsilon}_t^2 = 351.232\right)$ y del modelo general sin restricciones $\nabla Y_t = \underset{(0.836)}{1.586} + \underset{(0.086)}{0.155}t - \underset{(0.116)}{0.219}Y_{t-1} - \underset{(0.151)}{0.173}\nabla Y_{t-1}\left(\sum_{t=2}^{47}\hat{\varepsilon}_t^2 = 271.530\right)$ asumió el valor:

$$\Phi = \frac{\frac{1}{3}(351.232 - 271.530)}{\frac{1}{42}(271.530)} = 3.996$$

Este valor, comparado con el valor crítico $\Phi \approx 5.13$ (para un $\alpha = 0.05$) permite concluir que **no existen** razones suficientes para rechazar H_0.

De los resultados anteriores se puede concluir (con una confiabilidad del 95%): **i)** que la ecuación correspondiente al test de Dickey-Fuller no posee término independiente ni tendencia determinística y **ii)** que el proceso $\{Y_t\}$ presenta una raíz unitaria.

Ejemplo 3.6:

Sea un proceso $\{Y_t\}$:$AR(2)$ de parámetro mensual cuya trayectoria para el período [ENE-2001;DIC-2004] figura en el **GRAFICO 3.9**.

Considerando la representación general de la forma:

$$\nabla^2 Y_t = a_0 + a_1 t + \left(\phi_1^{**} - 1\right)\nabla Y_{t-1} - \phi_2^{**}\nabla^2 Y_{t-1} + \varepsilon_t$$

(donde $\phi_1^{**} = \phi_1 + \phi_2$ y $\phi_2^{*} = \phi_2$) se obtuvo el siguiente modelo:

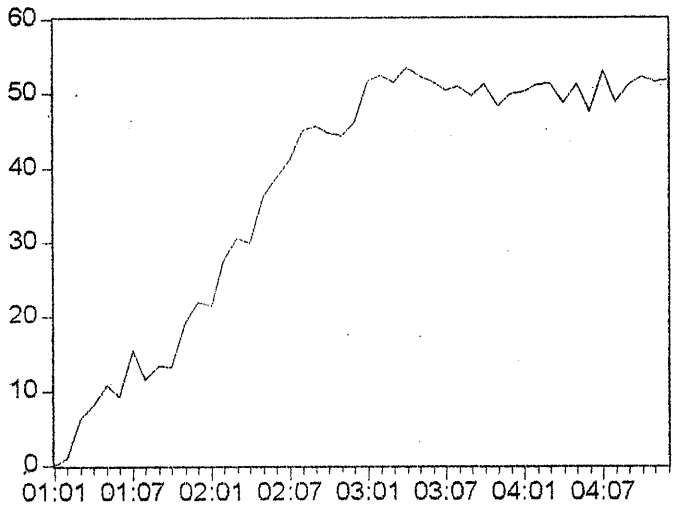
$$\nabla^2 Y_t = \underset{(1.083)}{3.574} - \underset{(0.033)}{0.087}t - \underset{(0.262)}{1.353}\nabla Y_{t-1} - \underset{(0.154)}{0.052}\nabla^2 Y_{t-1} \qquad (AIC = 4.6903, SBC = 4.8509)$$

De la aplicación del test "t" de significatividad individual de los coeficientes, se puede concluir que, con una confiabilidad del 95%, **no existen** razones suficientes para rechazar la hipótesis $H_0: \phi_2^{**} = 0$ con respecto a la alternativa $H_1: \phi_2^{**} \neq 0$. En consecuencia, se construyó el modelo alternativo:

$$\nabla^2 Y_t = \underset{(0.829)}{4.016} - \underset{(0.028)}{0.099}t - \underset{(0.136)}{1.440}\nabla Y_{t-1} \qquad (AIC = 4.6418, BSC = 4.7611)$$

a partir del cual se construyó un test para contrastar la hipótesis de no-estacionariedad $H_0: \phi_1^{**} - 1 = 0$ con respecto a la alternativa $H_1: \phi_1^{**} - 1 < 0$. El correspondiente estadístico T asumió el valor -10.567 el cual, comparado con el valor crítico $\tau_\tau = -3.5088$ (para un $\alpha = 0.05$), permite concluir que **existen** razones

GRAFICO 3.9: Trayectoria del proceso $\{Y_t\}$



suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{\nabla Y_t\}$ ($\{\nabla Y_t\}: I(0)$).

A fin de decidir si $\{Y_t\}$ es o no $I(1)$ se consideró la representación general de la forma:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1^* - 1)Y_{t-1} - \phi_2^* \nabla Y_{t-1} + \varepsilon_t$$

obteniéndose el siguiente modelo:

$$\nabla Y_t = \underset{(0.947)}{4.401} - \underset{(0.060)}{0.054}\, t - \underset{(0.047)}{0.040} Y_{t-1} - \underset{(0.139)}{0.420} \nabla Y_{t-1} \qquad \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 241.078\right)$$

a partir del cual se construyó un test para contrastar la hipótesis de no-estacionariedad del proceso $H_0: \phi_1^* - 1 = 0$, con respecto a la alternativa $H_1: \phi_1^* - 1 < 0$. El correspondiente estadístico T asumió el valor -0.849 el cual, comparado con el valor crítico $\tau_\tau = -3.5088$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{Y_t\}$.

De acuerdo con esta decisión y siguiendo la secuencia de Perron, se construyó un test para contrastar la hipótesis $H_0: a_1 = 0 / \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, repectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.427)}{1.422} - \underset{(0.145)}{0.285} Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 316.070 \right)$ y del modelo sin restricciones $\nabla Y_t = \underset{(0.854)}{4.115} - \underset{(0.028)}{0.099} t - \underset{(0.136)}{0.440} Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 245.213 \right)$ asumió el valor:

$$\Phi = \frac{\frac{1}{2}(316.070 - 245.213)}{\frac{1}{42}(245.213)} = 6.213$$

Este valor, comparado con el valor crítico $\tau_{\beta\tau} = 2.81$ (para un $\alpha = 0.05$), permite concluir que **existen** razones suficientes para rechazar H_0.

De la misma forma, se construyó un test para contrastar la hipótesis $H_0: a_0 = 0 / \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: a_0 \neq 0 / \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2 \right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR} \right]}{\frac{1}{n-k} \left(\sum_t \hat{\varepsilon}_t^2 \right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones $\nabla Y_t = \underset{(0.015)}{0.022} t - \underset{(0.149)}{0.145} \nabla Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 377.724 \right)$ y del modelo sin restricciones $\nabla Y_t = \underset{(0.854)}{4.115} - \underset{(0.028)}{0.099} t - \underset{(0.136)}{0.440} \nabla Y_{t-1} \left(\sum_{t=2}^{47} \hat{\varepsilon}_t^2 = 245.213 \right)$ asumió el valor:

$$\Phi = \frac{\frac{1}{2}(377.724 - 245.213)}{\frac{1}{42}(245.213)} = 11.618$$

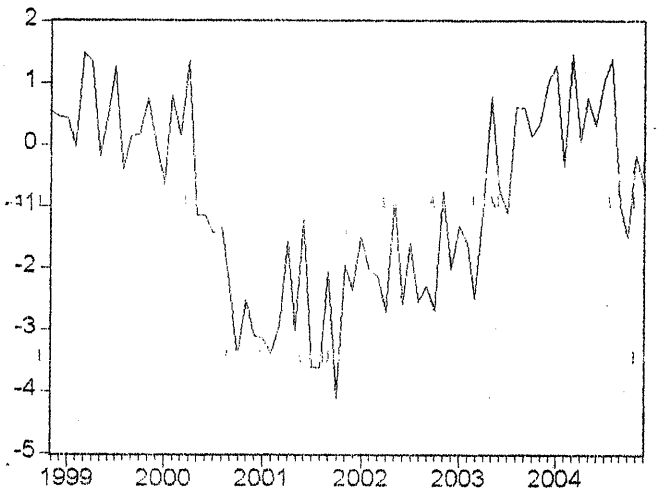
Este valor, comparado con el valor crítico $\tau_{\alpha\tau} = 3.14$ (para un $\alpha = 0.05$), permite concluir que **existen** razones suficientes para rechazar H_0.

De los resultados anteriores se puede concluir (con una confiabilidad del 95%): **i)** que la ecuación correspondiente al test de Dickey-Fuller posee término independiente y tendencia determinística y **ii)** que el proceso $\{Y_t\}$ presenta una raíz unitaria.

Ejemplo 3.7:

Sea un proceso $\{Y_t\}$ de parámetro mensual, cuya trayectoria para el período [NOV-1998;DIC-2004] figura en el **GRAFICO 3.10** y cuyo estimador de la función de autocorrelaciones figura en el **CUADRO 3.5**.

GRAFICO 3.10: Trayectoria del proceso $\{Y_t\}$



CUADRO 3.5: Correlograma correspondiente al proceso $\{Y_t\}$

Sample: 1998:11 2004:12
Included observations: 74

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.712	0.712	39.051	0.000
		2	0.713	0.419	78.804	0.000
		3	0.688	0.233	116.31	0.000
		4	0.662	0.127	151.55	0.000
		5	0.587	-0.064	179.68	0.000
		6	0.499	-0.191	200.27	0.000
		7	0.474	-0.053	219.14	0.000
		8	0.415	-0.038	233.82	0.000
		9	0.321	-0.124	242.74	0.000
		10	0.283	-0.012	249.78	0.000
		11	0.219	-0.039	254.07	0.000
		12	0.188	0.027	257.26	0.000
		13	0.083	-0.117	257.90	0.000
		14	0.049	-0.045	258.13	0.000
		15	-0.036	-0.146	258.25	0.000
		16	-0.097	-0.110	259.17	0.000
		17	-0.142	-0.018	261.15	0.000
		18	-0.210	-0.051	265.57	0.000

A partir de la información proporcionada por la serie cronológica y aplicando MCO a la representación general de la forma:

$$\nabla^2 Y_t = a_0 + a_1 t + \left(\phi_1^{**} - 1\right)\nabla Y_{t-1} - \sum_{j=1}^{p^*} \phi_{j+1}^{**} \nabla^2 Y_{t-j} + \varepsilon_t$$

(de acuerdo con la aproximación propuesta por Said-Dickey, se consideró $p^* = n^{1/3}$), se obtuvo el siguiente modelo[59]:

$$\nabla^2 Y_t = \underset{(0.237)}{-0.143} + \underset{(0.005)}{0.002} t - \underset{(0.500)}{1.960} \nabla Y_{t-1} + \underset{(0.446)}{0.253} \nabla^2 Y_{t-1} - \underset{(0.364)}{0.129} \nabla^2 Y_{t-2} -$$

$$- \underset{(0.255)}{0.246} \nabla^2 Y_{t-3} - \underset{(0.127)}{0.140} \nabla^2 Y_{t-4} \qquad (AIC = 2.7818, SBC = 2.9933)$$

Dado que, de acuerdo con el test "t", el coeficiente que acompaña al término autorregresivo $\nabla^2 Y_{t-4}$, con una confiabilidad del 95%, puede ser considerado nulo, se construyó el siguiente modelo $AR(3)$:

$$\nabla^2 Y_t = \underset{(0.228)}{-0.156} + \underset{(0.005)}{0.003} t - \underset{(0.403)}{2.273} \nabla Y_{t-1} + \underset{(0.377)}{0.567} \nabla^2 Y_{t-1} -$$

$$- \underset{(0.240)}{0.166} \nabla^2 Y_{t-2} - \underset{(0.124)}{0.003} \nabla^2 Y_{t-4} \qquad (AIC = 2.7600, SBC = 2.9399)$$

Dado que, de acuerdo con el test "t", el coeficiente que acompaña al término $\nabla^2 Y_{t-3}$, con una confiabilidad del 95%, puede ser considerado nulo, se construyó el siguiente modelo $AR(2)$:

$$\nabla^2 Y_t = \underset{(0.219)}{-0.134} + \underset{(0.004)}{0.003} t - \underset{(0.292)}{2.274} \nabla Y_{t-1} + \underset{(0.219)}{0.569} \nabla^2 Y_{t-1} +$$

$$+ \underset{(0.115)}{0.169} \nabla^2 Y_{t-2} \qquad (AIC = 2.7231, SBC = 2.8719)$$

Dado que, de acuerdo con el test "t", el coeficiente que acompaña al término $\nabla^2 Y_{t-2}$, con una confiabilidad del 95%, puede ser considerado nulo, se calculó el modelo $AR(1)$:

$$\nabla^2 Y_t = \underset{(0.214)}{-0.108} + \underset{(0.004)}{0.002} t - \underset{(0.190)}{1.944} \nabla Y_{t-1} + \underset{(0.110)}{0.289} \nabla^2 Y_{t-1} \qquad (AIC = 2.7140, SBC = 2.8323)$$

en el cual, de acuerdo con el test "t" (con una confiabilidad del 95%), el coeficiente que acompaña al término $\nabla^2 Y_{t-1}$ puede ser considerado significativo.

A partir de este modelo se construyó un test para contrastar la hipótesis de no-estacionariedad $H_0: \phi_1^{**} - 1 = 0$ con respecto a la alternativa $H_1: \phi_1^{**} - 1 < 0$. El correspondiente estadístico T asumió el valor -10.212 el cual, comparado con el valor crítico $\tau_\tau = -3.4652$ (para un $\alpha = 0.05$), permite concluir que existen razones suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{\nabla Y_t\}$. Es decir, permite concluir (con un nivel de significación del 5%) que $\{\nabla Y_t\}: I(0)$ o, lo que es lo mismo, que el proceso $\{Y_t\}$ posee, como máximo, una raíz unitaria.

---

[59]. Ver Said, S.; Dickey, D. (1985).

A fin de decidir si $\{Y_t\}$ es o no $I(1)$ se consideró la representación general de la forma:

$$\nabla Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right) Y_{t-1} + \sum_{j=1}^{4} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

obteniéndose el siguiente modelo:

$$\nabla Y_t = \underset{(0.228)}{-0.250} + \underset{(0.005)}{0.003} t - \underset{(0.082)}{0.111} t -$$

$$- \underset{(0.135)}{0.636} \nabla Y_{t-1} - \underset{(0.142)}{0.353} \nabla Y_{t-2} - \underset{(0.152)}{0.135} \nabla Y_{t-3} + \underset{(0.128)}{0.021} \nabla Y_{t-4} \qquad (AIC = 2.8124, SBC = 3.0303)$$

Dado que, de acuerdo con el test "t", con una confiabilidad del 95%, el coeficiente correspondiente al término ∇Y_{t-4} puede ser considerado nulo, se calculó el siguiente modelo $AR(3)$:

$$\nabla Y_t = \underset{(0.225)}{-0.254} + \underset{(0.005)}{0.003} t - \underset{(0.080)}{0.110} Y_{t-1} -$$

$$- \underset{(0.131)}{0.641} \nabla Y_{t-1} - \underset{(0.142)}{0.362} \nabla Y_{t-2} - \underset{(0.120)}{0.151} \nabla Y_{t-3} \qquad (AIC = 2.7858, SBC = 3.9726)$$

Dado que, de acuerdo con el test "t", con una confiabilidad del 95%, el coeficiente correspondiente al término ∇Y_{t-3} puede ser considerado nulo, se calculó el siguiente modelo $AR(2)$:

$$\nabla Y_t = \underset{(0.226)}{-0.238} + \underset{(0.005)}{0.003} t - \underset{(0.080)}{0.125} Y_{t-1} -$$

$$- \underset{(0.124)}{0.587} \nabla Y_{t-1} - \underset{(0.116)}{0.259} \nabla Y_{t-2} \qquad \left(AIC = 2.7818, SBC = 2.9375, \sum_{t=2}^{82} \hat{\varepsilon}_t^2 = 64.835\right)$$

De acuerdo con el test "t" (con una confiabilidad del 95%) el coeficiente correspondiente al término $AR(2)$ puede ser considerado significativo.

A partir de este modelo se construyó un test para contastar la hipótesis $H_0: \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: \phi_1^* - 1 < 0$. El correspondiente estadístico T asumió el valor -1.559 el cual, comparado con el valor crítico $\tau_\tau = -3.4704$ (para un $\alpha = 0.05$), permite concluir que no existen razones suficientes para rechazar la hipótesis de presencia de una raíz unitaria en el proceso $\{Y_t\}$.

De acuerdo con esta decisión y siguiendo la secuencia mencionada en la Sec. 3.3.1, se construyó un test para contrastar la hipótesis $H_0: a_1 = 0 / \phi_1^* - 1 = 0$ con respecto a la alternativa $H_1: a_1 \neq 0 / \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones:

$$\nabla Y_t = \underset{(0.052)}{-0.010} - \underset{(0.109)}{0.654}\nabla Y_{t-1} - \underset{(0.110)}{0.288}\nabla Y_{t-2} \qquad \left(\sum_{t=2}^{82}\hat{\varepsilon}_t^2 = 65.018\right)$$

y del modelo general sin restricciones:

$$\nabla Y_t = \underset{(0.111)}{-0.056} + \underset{(0.002)}{0.001}t - \underset{(0.109)}{0.656}\nabla Y_{t-1} - \underset{(0.110)}{0.289}\nabla Y_{t-2} \qquad \left(\sum_{t=2}^{82}\hat{\varepsilon}_t^2 = 64.835\right)$$

asumió el valor:

$$\Phi = \frac{\frac{1}{2}(65.018 - 64.835)}{\frac{1}{77}(64.835)} = 0.109$$

Este valor, comparado con el valor crítico $\tau_{\beta\tau} = 2.79$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis nula.

De la misma forma, se construyó un test para contrastar la hipótesis:

$$H_0: a_0 = 0 \,/\, \phi_1^* - 1 = 0$$

con respecto a la alternativa $H_1: a_0 \neq 0 \,/\, \phi_1^* - 1 = 0$. El correspondiente estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones:

$$\nabla Y_t = \underset{(0.001)}{0.000}t - \underset{(0.109)}{0.653}\nabla Y_{t-1} - \underset{(0.110)}{0.287}\nabla Y_{t-2} \qquad \left(\sum_{t=2}^{82}\hat{\varepsilon}_t^2 = 65.049\right)$$

y del modelo sin restricciones mencionado más arriba, asumió el valor:

$$\Phi = \frac{\frac{1}{2}(65.049 - 64.835)}{\frac{1}{77}(64.835)} = 0.127$$

Este valor, comparado con el valor crítico $\tau_{\alpha\tau} \approx 3.12$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis nula.

A fin de confirmar los resultados anteriores se construyó el siguiente test para contrastar la hipótesis:

$$H_0: a_0 = a_1 = \phi_1^* - 1 = 0$$

En este caso, el estadístico:

$$\Phi = \frac{\frac{1}{r}\left[\left(\sum_t \hat{\varepsilon}_t^2\right)_{CR} - \left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}\right]}{\frac{1}{n-k}\left(\sum_t \hat{\varepsilon}_t^2\right)_{SR}}$$

que surge de la comparación de las sumas de los cuadrados de los residuos obtenidos, respectivamente, de la aplicación del modelo con restricciones:

$$\nabla Y_t = \underset{(0.108)}{-0.653}\nabla Y_{t-1} - \underset{(0.109)}{0.287}\nabla Y_{t-2} \qquad \left(\sum_{t=2}^{82} \hat{\varepsilon}_t^2 = 65.051\right)$$

y del modelo general sin restricciones:

$$\nabla Y_t = \underset{(0.226)}{-0.238} + \underset{(0.005)}{0.003}t - \underset{(0.080)}{0.125}Y_{t-1} - \underset{(0.124)}{0.587}\nabla Y_{t-1} - \underset{(0.116)}{0.259}\nabla Y_{t-2} \qquad \left(\sum_{t=2}^{82} \hat{\varepsilon}_t^2 = 64.835\right)$$

asumió el valor:

$$\Phi = \frac{\frac{1}{3}(65.051 - 64.835)}{\frac{1}{76}(64.835)} = 0.084$$

Este valor, comparado con el valor crítico $\Phi_2 \approx 3.12$ (para un $\alpha = 0.05$), permite concluir que **no existen** razones suficientes para rechazar la hipótesis nula.

De los resultados anteriores se puede concluir (con una confiabilidad del 95%): **i)** que la ecuación correspondiente al test de Dickey-Fuller no posee término independiente ni tendencia determinística y **ii)** que el proceso $\{Y_t\}$ presenta una raíz unitaria.

3.4.- Modificaciones en la estructura temporal

Cuando el proceso experimenta un cambio en su estructura temporal los tests de Dickey-Fuller y Phillips-Perron presentan un sesgo hacia el no-rechazo de la hipótesis de existencia de una raíz unitaria.

A partir de esta característica, Perron, P. (1991) generalizó el criterio de aplicación del test de la raíz unitaria a fin de permitir detectar un posible cambio en la estructura que produce realizaciones atípicas ("outliers") del proceso en un momento $t^* + 1 \left(1 < t^* + 1 < n\right)$ conocido.

En principio, Perron consideró tres casos:

1) Una hipótesis nula de existencia de un salto en el nivel de un proceso con una raíz unitaria (es decir, la hipótesis de que $\{Y_t\}$ es un proceso con una raíz unitaria con un "salto" en su nivel en el período $t = t^* + 1$) $H_0^{(1)}: Y_t = a_0 + Y_{t-1} + b_1(DP)_t + \varepsilon_t$, con respecto a la hipótesis alternativa de una variación en un período t en el término independiente de un proceso con tendencia estacionaria $H_1^{(1)}: Y_t = a_0 + a_1 t + b_2(DL)_t + \varepsilon_t$.

Donde $(DP)_t$ denota una **variable "input" determinística** (o **variable "dummy") de pulso o aditiva,** tal que:

$$(DP)_t = \begin{cases} 1 & si \quad t = t^* + 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y $(DL)_t$ denota una variable **"dummy" de nivel** o innovadora, tal que:

$$(DL)_t = \begin{cases} 1 & si \quad t > t^* \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Se estima, entonces, la ecuación:

$$Y_t = a_0 + a_1 t + \phi_1^* Y_{t-1} + b_2 (DL)_t + \sum_{j=1}^{p} \phi_j^* \nabla Y_{t-j} + \varepsilon_t$$

a partir del modelo resultante se contrasta la hipótesis nula $H_0: \phi_1^* - 1 = 0$, mediante la comparación del estadístico $T = \dfrac{\hat{\phi}_1^* - 1}{\hat{\sigma}(\hat{\phi}_1^* - 1)}$ con el correspondiente valor crítico calculado por Perron, P. (1991). Si el valor de T es mayor que dicho valor crítico, se considerará que existen razones suficientes para rechazar la hipótesis nula de presencia de una raíz unitaria.

2) Una hipótesis nula de existencia de un cambio permanente en la tasa de crecimiento del proceso $H_0^{(2)}: Y_t = a_0 + Y_{t-1} + b_2 (DL)_t + \varepsilon_t$, con respecto a la hipótesis alternativa de existencia de un cambio en la pendiente de la tendencia determinística, $H_1^{(2)}: Y_t = a_0 + a_1 t + b_3 (DN)_t + \varepsilon_t$. Donde,

$$(Dn)_t = \begin{cases} t - t^* & si \quad t > t^* \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Obsérvese que la hipótesis nula supone que:

$$\nabla Y_t = \begin{cases} a_0 + \varepsilon_t & si \quad t \le t^* \\ (a_0 + b_2) + \varepsilon_t & si \quad t > t^* \end{cases}$$

y la hipótesis alternativa supone la presencia de una tendencia segmentada, es decir, un cambio en la pendiente determinística de a_1, para $t \le t^*$, a $(a_1 + b_3)$, para $t > t^*$, sin variaciones imprevistas en el nivel.

3) Una hipótesis nula -que combina las dos anteriores- de existencia de un cambio imprevisto en el nivel y un cambio en la forma de crecimiento:

$$H_0^{(3)}: Y_t = a_0 + Y_{t-1} + b_1 (DP)_t + b_2 (DL)_t + \varepsilon_t$$

con respecto a la hipótesis alternativa que puede ser parametrizada de la siguiente forma:

$$H_1^{(3)} : Y_t = a_0 + a_1 t + b_2 (DL)_t + b_3 (Dn)_t + \varepsilon_t$$

Se estima, entonces, la ecuación:

$$Y_t = a_0 + a_1 t + b_2 (DL)_t + b_3 (Dn)_t + b_4 (DP)_t + \phi_1^* Y_{t-1} + \sum_{t=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

Bajo la hipótesis nula de existencia de una raíz unitaria se debe testear que $\phi_1^* - 1 = a_1 = b_2 = b_3 = 0$ y $b_4 \neq 0$, con respecto a la hipótesis alternativa $\phi_1^* - 1 < 0, a_1, b_2, b_3 \neq 0$ y $b_4 = 0$ [60].

Teniendo en cuenta que uno de los problemas fundamentales en el estudio de los cambios estructurales es la **determinación del momento** en que se supone que éstos se producen, Zivot, E.; Andrews, D.W.K. (1990) demostraron que la selección de t no debería realizarse en forma exógena, consideraron como hipótesis nula, simplemente, que $\{Y_t\}$ es un proceso $I(1)$ sin ningún cambio estructural exógeno, y supusieron como hipótesis alternativa que el proceso $\{Y_t\}$ admite una representación estacionaria en la tendencia, con un cambio estructural en dicha tendencia para un valor de t desconocido. La aproximación propuesta por Zivot-Andrews consiste en seleccionar el punto $t^{(\lambda)} = \frac{t^*}{n}$ que minimice el test de la raíz unitaria calculado a partir de la regresión:

$$Y_t = a_0 + a_1 t + b_2 \left(DL\left(t^{(\lambda)}\right)\right)_t + b_3 \left(DT\left(t^{(\lambda)}\right)\right)_t + \phi_1^* Y_{t-1} + \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \varepsilon_t$$

donde,

$$\left(DL\left(t^{(\lambda)}\right)\right)_t = \begin{cases} 1 & si \quad t > \left[nt^{(\lambda)}\right] \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

$$\left(Dn\left(t^{(\lambda)}\right)\right)_t = \begin{cases} 1 - \left[nt^{(\lambda)}\right] & si \quad t > \left[nt^{(\lambda)}\right] \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y $\frac{2}{n} \leq t^{(\lambda)} \leq \frac{n-1}{n}$.

Un problema que suele presentarse en la práctica se refiere a la posibilidad de distinguir entre modelos integrados y modelos con tendencia segmentada. Balke, N.S.; Fomby, T.B. (1991),(1991) propusieron una aproximación basada en la utilización del **analisis de intervención** para identificar "outliers" causados por "shocks" no-permanentes[61].

---

[60] . Rappoport, P.; Reichlin, I. (1989) generalizaron estos tests a casos con dos cambios de tendencia.

[61] . Ver Mills, T.C. (1990). Los temas relacionado con el análisis de intervención serán tratados en la Parte 2, al estudiar los procesos multidimensionales.

3.5.- Las medidas de persistencia

Sea un proceso estocástico $\{Y_t\}$ que admite una representación $ARIMA(p,1,q)$ de la forma:

$$\Phi(B)\nabla Y_t = \Phi(B)(1-B)Y_t = \Theta(B)\varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$) siendo $\nabla Y_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \Psi(B)\varepsilon_t$ su correspondiente representación MA.

Según lo expresado en la Sección 2.9.1, $\Psi(1) = \sum_{j=0}^{\infty}\psi_j = \frac{\Theta(1)}{\Phi(1)}\ (\psi_0 = 1)$ define la medida de la persistencia de la influencia de un "shock" aleatorio sobre $\{Y_t\}$ [62]. De modo que, si $\{Y_t\}$ es estacionario con una representación MA de la forma $Y_t = \left(\sum_{j=0}^{\infty}\psi_j B^j\right)\varepsilon_t\ (\psi_0 = 1)$, se verificará que $\nabla Y_t = (1-B)\left(\sum_{j=0}^{\infty}\psi_j B^j\right)$. Es decir, que $\Psi(B) = (1-B)\left(\sum_{j=0}^{\infty}\psi_j B^j\right)$ y, por lo tanto, que $\Psi(1) = 0$.

Si, en particular, el proceso $\{\nabla Y_t\}$ admite una representación $AR(1)$ de la forma:

$$\nabla Y_t = \phi_1 \nabla Y_{t-1} + \varepsilon_t$$

la medida de persistencia queda definida como $\Psi(1) = (1-\phi_1)^{-1}$. Luego, se puede concluir que, si $\{Y_t\}$:RW (es decir, si $\phi_1 = 0$), entonces $\Psi(1) = 1$ y que, en la medida que $\phi_1 \to 1$, $\Psi(1) \to \infty$.

Cochrane, J.H. (1991) propuso, por su parte, una definición no-paramétrica de la medida de persistencia a partir de un cociente de varianzas de la forma $V_{(\ell)} = \frac{\sigma_{(\ell)}^2}{\sigma_{(1)}^2}$, donde $\sigma_{(\ell)}^2 = \frac{\sigma^2(Y_t - Y_{t-1})}{\ell} = \frac{\sigma^2(\nabla^{(\ell)}Y_t)}{\ell}$ (siendo $\nabla^{(\ell)} = 1 - B^\ell$). Teniendo en cuenta que:

$$\begin{aligned}
&\nabla^{(\ell)}Y_t = (1-B^\ell)Y_t = Y_t - Y_{t-\ell} = \\
&= Y_t + (Y_{t-1} - Y_{t-1}) + (Y_{t-2} - Y_{t-2}) + \ldots + (Y_{t-(\ell-1)} - Y_{t-(\ell-1)}) - Y_{t-\ell} = \\
&= (Y_t - Y_{t-1}) + (Y_{t-1} - Y_{t-2}) + \ldots + (Y_{t-(\ell-1)} - Y_{t-\ell}) = \\
&= \nabla Y_t + \nabla Y_{t-1} + \ldots + \nabla Y_{t-(\ell-1)} = \\
&= \nabla(1 + B + B^2 + \ldots + B^{\ell-1})Y_t
\end{aligned}$$

si $\{Y_t\}$ es un proceso no-estacionario que admite una representación $AR(1)$ de la forma:

---

[62] . Ver Campbell, J.Y.; Mankiw, N.G. (1987).

$$Y_t = \phi_0 + Y_{t-1} + \varepsilon_t$$

entonces, se verifica que[63]:

$$\sigma^2\left(\nabla^{(\ell)}Y_t\right) = \sigma^2\left(\sum_{j=0}^{\ell-1}\left(Y_{t-j} - Y_{t-j-1}\right)\right) = \sigma^2\left(\sum_{j=0}^{\ell-1}\varepsilon_{t-j}\right) = \ell\sigma_\varepsilon^2$$

Es decir, que la varianza del proceso $\left\{\nabla^{(\ell)}Y_t\right\}$ crece en forma directamente proporcional a ℓ. De modo que $\sigma^2_{(\ell)} = \sigma^2_\varepsilon$ y, por lo tanto, que $V_{(\ell)} = 1$. Por otra parte, si $\left\{Y_t\right\}$ es un proceso estacionario, se verificará que:

$$\sigma^2\left(\nabla^{(\ell)}Y_t\right) = \sigma^2\left(Y_t - Y_{t-\ell}\right) = \sigma^2\left[\Psi(B)\varepsilon_t - \Psi(B)\varepsilon_{t-\ell}\right] =$$

$$= \sigma^2\left(\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} - \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-\ell-j}\right) = \sigma^2\left(\sum_{j=0}^{\ell-1}\psi_j\varepsilon_{t-j} + \sum_{j=\ell}^{\infty}\psi_j\varepsilon_{t-j} - \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-\ell-j}\right) =$$

$$= \sigma^2\left(\sum_{j=0}^{\ell-1}\psi_j\varepsilon_{t-j} + \sum_{j=0}^{\infty}\psi_{j+\ell}\varepsilon_{t-\ell-j} - \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-\ell-j}\right) = \sigma^2\left(\sum_{j=0}^{\ell-1}\psi_j\varepsilon_{t-j} + \sum_{j=0}^{\infty}\left(\psi_{j+\ell} - \psi_j\right)\varepsilon_{t-\ell-j}\right) =$$

$$= \sigma_\varepsilon^2\sum_{j=0}^{\ell-1}\psi_j^2 + \sigma_\varepsilon^2\sum_{j=0}^{\infty}\left(\psi_{j+\ell} - \psi_j\right)^2$$

De modo que, cuando ℓ aumenta indefinidamente se verifica que $\lim_{\ell\to\infty}\sigma^2\left(\nabla^{(\ell)}Y_t\right) = 2\sigma_\varepsilon^2\sum_{j=0}^{\infty}\psi_j^2 = 2\sigma_Y^2$. Es decir, que $\lim_{\ell\to\infty}\sigma^2_{(\ell)} = \lim_{\ell\to\infty}\frac{2\sigma_Y^2}{\ell} = 0$.

Cochrane, J.H. (1991) demuestra que el cociente $V_{(\ell)}$ puede ser expresado como $V_{(\ell)} = 1 + 2\sum_{j=1}^{\ell-1}\frac{\ell-j}{\ell}\rho_j(Y)$ y, dado que $\lim_{\ell\to\infty}\sigma^2_{(\ell)} = \frac{\left(\sum_{j=0}^{\infty}\psi_j\right)^2}{\sum_{j=0}^{\infty}\psi_j^2}\sigma^2_{(1)} = \left[\Psi(1)\right]^2\sigma_\varepsilon^2 = \left(\sum_{j=0}^{\infty}\psi_j\right)^2\sigma_\varepsilon^2\ (\psi_0 = 1)$, el cociente de varianzas en el límite queda definido como $V = \lim_{\ell\to\infty}V_{(\ell)} = 1 + 2\sum_{j=1}^{\infty}\rho_j(Y) = \frac{\sigma_\varepsilon^2}{\sigma^2_{(1)}}\left[\Psi(1)\right]^2$. Luego, teniendo en cuenta que $1 - \frac{\sigma_\varepsilon^2}{\sigma^2_{(1)}} = R^2$ denota la fracción de la varianza de ∇Y_t que puede ser predicha a partir de sus realizaciones pasadas, se puede escribir $\Psi(1) = \sqrt{\frac{V_{(\ell)}}{1-R^2}}$. Lo que demuestra que $V_{(\ell)}$ define el menor valor que puede asumir $\Psi(1)$ y que, cuanto mayor sea la autorregresividad del proceso $\left\{\nabla Y_t\right\}$, mayor será la

---

63. Como se verá en la Sec. 4.1, esta es una forma particular de un proceso "random walk" con término independiente y, en consecuencia, afectado por tendencias determinística y estocástica.

diferencia entre las medidas $\Psi(1)$ y $V_{(\ell)}$.

Un estimador de $\sigma^2_{(\ell)}$, insesgado cuando se aplica a un modelo de la forma $\hat{Y}_t = \hat{\phi}_0 + Y_{t-1}$ puede ser definido de la siguiente forma[64]:

$$\hat{\sigma}^2_{(\ell)} = \frac{n}{\ell(n-\ell)(n-\ell+1)} \sum_{t=\ell}^{n} \left[\nabla^{(\ell)} Y_t - \frac{\ell}{n}(Y_n - Y_1) \right]^2$$

con varianza asintótica $\sigma^2\left(\hat{\sigma}^2_{(\ell)}\right) = \frac{4\ell}{3n}\hat{\sigma}^4_{(\ell)}$. A partir de este resultado se puede construir un estimador del cociente de varianzas de la forma $\hat{V}_{(\ell)} = \frac{\hat{\sigma}^2_{(\ell)}}{\hat{\sigma}^2_{(1)}}$.

Luego, como corolario de los resultados anteriores se puede concluir que el cociente de varianzas de Cochran puede ser utilizado para contrastar la hipótesis nula de que un proceso admite una representación $AR(1)$ no-estacionaria de la forma $Y_t = \phi_0 + Y_{t-1} + \varepsilon_t$ donde $\{\nabla Y_t\} = \{\varepsilon_t\}: WN - N(0, \sigma^2_\varepsilon)$. Es decir, para contrastar la hipótesis nula $H_0: V_{(\ell)} - 1 = 0$.

De acuerdo con Lo, A.W.; McKinlay, A.C. (1988)(1989) el estadístico de este test puede ser expresado de la siguiente forma:

$$M(\ell) = \frac{\hat{\sigma}^2_{(\ell)}}{\hat{\sigma}^2_{(1)}} - 1 = \hat{V}_{(\ell)} - 1 = \frac{n(n-1)}{\ell(n-\ell)(n-\ell+1)} \frac{\sum_{t=\ell}^{n} \left[Y_t - Y_{t-\ell} - \frac{\ell}{n}(Y_n - Y_1) \right]^2}{\sum_{t=1}^{n} \left[Y_t - Y_{t-1} - \frac{1}{n}(Y_n - Y_1) \right]^2}$$

Bajo el supuesto de que la hipótesis nula es verdadera, se verifica que:

$$Z(\ell) = M(\ell)\sqrt{\frac{3n\ell}{2(2\ell-1)(\ell-1)}} \xrightarrow{d} N(0;1)$$

3.6.- La teoría de la predicción en procesos no-estacionarios

Sea $\{Y_t\}$ un proceso estocástico que admite una representación $ARIMA(p,d,q)$ de la forma:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_{p+d} Y_{t-p-d} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}: WN - N(0, \sigma^2_\varepsilon)$), sea $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$ la serie cronológica formada por un conjunto de sus realizaciones y sea:

$$f^{(t,\ell)}(Y_t, Y_{t-n}, \ldots, Y_{t-n}) = E(Y_{t+\ell} / Y_t, Y_{t-1}, \ldots, Y_{t-n})$$

---

[64]. Ver Cochrane, J.H. (1991).

la **función de predicción** (o **predictor**) del proceso $\{Y_t\}$ basada en el mencionado conjunto de observaciones.

Supóngase que se desea predecir el valor $Y_{t+\ell}$ $(\ell \geq 1)$. El problema consiste en la selección de la función de predicción que haga que el valor estimado resulte lo más aproximado que sea posible a $Y_{t+\ell}$. Igual que en el caso de los procesos estacionarios, el criterio de optimización utilizado será el de la minimización de la distancia L_2, es decir, de la minimización del error medio cuadrático de la predicción, según el cual la función $f^{(t,\ell)}$ elegida debe ser tal que minimice la expresión correspondiente a la varianza del error de la predicción para ℓ períodos ($e(\ell,t)$):

$$\sigma^2_{(\ell,t)}(e) = E\left\{\left[Y_{t+\ell} - f^{(t,\ell)}\left(Y_t, Y_{t-1}, \ldots, Y_{t-n}\right)\right]^2\right\} \qquad (\ell \geq 1)$$

Supóngase que el proceso $\{Y_t\}$ tenga espectro evolutivo continuo y función de densidad espectral ($w_t(\alpha)$) que satisfaga la condición $\int_{-\pi}^{\pi} \ln\left[w_t(\alpha)\right] d\alpha > -\infty$. Entonces, de acuerdo a lo demostrado en la Sec.2.9.1, $\{Y_t\}$ admite una representación de la forma:

$$Y_t = \sum_{j=0}^{\infty} b_t(j)\varepsilon_{t-j}$$

donde $\{\varepsilon_t\}$:*WN* débil con varianza $\sigma^2(\varepsilon_t) = \sigma^2_\varepsilon = 2\pi$. Luego, a partir de la serie cronológica $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$, se puede escribir:

$$Y_{t+\ell} = \sum_{j=0}^{\ell-1} b_{t+\ell}(j)\varepsilon_{t+\ell-j} = e(\ell,t) + a_t(B)Y_t = e(\ell,t) + f^{(\ell,t)}\left(Y_t, Y_{t-1}, \ldots, Y_{t-n}\right)$$

De modo que la varianza del error de predicción queda definida por[65]:

$$\sigma^2_{(\ell,t)}(e) = E\left\{\left[Y_{t+\ell} - f^{(t,\ell)}\right]^2\right\} = \sigma^2_\varepsilon \sum_{j=0}^{\ell-1} b_t^2(j)$$

Esta varianza se minimiza cuando los coeficientes $b_t(j)$ son funciones de los coeficientes de la representación $ARIMA(p,d,q)$:

$$f^{(t,\ell)} = E\left(\varphi_1 Y_{t+\ell-1} + \varphi_2 Y_{t+\ell-2} + \ldots + \varphi_{p+d} Y_{t+\ell-p-d} + \varepsilon_{t+\ell} - \theta_1 \varepsilon_{t+\ell-1} - \ldots - \theta_q \varepsilon_{t+\ell-q} \,/\, Y_t, \ldots, Y_{t-n}\right)$$

Supóngase, en particular, un proceso $\{Y_t\}$ con incrementos incorrelacionados que admite una

---

[65] . La utilización de la representación mediante el espectro evolutivo (cuya forma permite determinar en completamente los coeficientes del predictor lineal óptimo aún cuando la estructura del proceso sea completamente desconocida), debido a Abdrabbo, N.A.; Priestley, M.B. (1967), asimila la teoría de la predicción para procesos no-estacionarios con la teoría de Kolmogorov-Wiener para procesos-estacionarios. Otros aportes fundamentales al problema de la predicción para procesos no-estacionarios que merecen ser mencionados son Cramér, H. (1945), Davis, R.C. (1952), Whittle, P. (1963), Parzen, E. (1974).

representación de la forma $Y_t = Y_{t-1} + \varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:*WN - debil*). Entonces, se puede escribir [66]:

$$Y_{t+1} = Y_t + \varepsilon_{t+1}$$

$$Y_{t+2} = Y_{t+1} + \varepsilon_{t+2} = Y_t + (\varepsilon_{t+2} + \varepsilon_{t+1})$$

$$Y_{t+3} = Y_{t+2} + \varepsilon_{t+3} = Y_t + (\varepsilon_{t+3} + \varepsilon_{t+2} + \varepsilon_{t+1})$$

.....

$$Y_{t+\ell} = Y_t + \sum_{j=1}^{\ell} \varepsilon_{t+j} \qquad (\ell \geq 1)$$

De modo que la función de predicción será $f^{(t,\ell)} = E(Y_{t+\ell} / Y_t) = Y_t$ (como se verá en la Sec.4.3, esta expresión se conoce como la propiedad de la martingala) y el error de predicción queda, entonces, definido por:

$$e(\ell,t) = Y_{t+\ell} - f^{(t,\ell)} = \sum_{j=1}^{\ell} \varepsilon_{t+j} \ (\ell \geq 1)$$

Luego, se puede concluir que la varianza del error de predicción crece en forma directamente proporcional a ℓ, $\sigma_{(\ell,t)}(e) = \ell\sigma_\varepsilon^2$.

Supóngase, ahora, un proceso $\{Y_t\}$ que admite una representación $ARIMA(1,1,q)$ de la forma:

$$(1-B)Y_t = (1-\theta_1 B - ... - \theta_q B^q)\varepsilon_t$$

$$Y_t = Y_{t-1} + \varepsilon_t - \theta_1\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-q}$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:*WN - debil*). Entonces, la función de predicción será de la forma:

$$f^{(t,1)} = Y_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ... - \theta_q\varepsilon_{t-(q-1)}$$

$$f^{(2,t)} = Y_t - (\theta_1 + \theta_2)\varepsilon_t - (\theta_2 + \theta_3)\varepsilon_{t-1} - ... - (\theta_{q-1} - \theta_q)\varepsilon_{t-(q-2)} - \theta_q\varepsilon_{t-(q-1)}$$

.....

$$f^{(t,q-1)} = Y_t - (\theta_1 + \theta_2 + ... + \theta_{q-1})\varepsilon_t - (\theta_2 + \theta_3 + ... + \theta_q)\varepsilon_{t-1} - (\theta_3 + \theta_4 + ... + \theta_q)\varepsilon_{t-2} - ... -$$

$$- (\theta_{q-1} + \theta_q)\varepsilon_{t-(q-2)} - \theta_q\varepsilon_{t-(q-1)}$$

$$f^{(t,\ell)} = Y_t - (\theta_1 + \theta_2 + ... + \theta_q)\varepsilon_t - (\theta_2 + \theta_3 + ... + \theta_q)\varepsilon_{t-1} - ... - (\theta_{q-1} + \theta_q)\varepsilon_{t-(q-2)} - \theta_q\varepsilon_{t-(q-1)} \qquad (\ell \geq q)$$

---

[66]. Este proceso, que se conoce como RW de tipo 3 (RW-3) y se caracteriza porque las perturbaciones forman un proceso WN con variables incorrelacionadas (ver Sec. 4.1).

El error de predicción será:

$$e(1,t) = \varepsilon_{t+1}$$

$$e(2,t) = \varepsilon_{t+2} + (1-\theta_1)\varepsilon_{t+1}$$

$$e(3,t) = \varepsilon_{t+3} + (1-\theta_1)\varepsilon_{t+2} + (1-\theta_1-\theta_2)\varepsilon_{t+1}$$

.....

$$e(q-1,t) = \varepsilon_{t+q-1} + (1-\theta_1)\varepsilon_{t+q-2} + ... + (1-\theta_1 - ... - \theta_{q-1})\varepsilon_{t+1}$$

$$e(\ell,t) = \varepsilon_{t+\ell} + (1-\theta_1)\varepsilon_{t+\ell-1} + (1-\theta_1-\theta_2)\varepsilon_{t+\ell-2} + ... + (1-\theta_1 - ... - \theta_q)\sum_{j=1}^{\ell-q}\varepsilon_{t+j} \quad (\ell \geq q)$$

y la varianza del error de predicción será:

$$\sigma^2_{(1,t)}(e) = \sigma^2_\varepsilon$$

$$\sigma^2_{(2,t)}(e) = \left[1 + (1-\theta_1)^2\right]\sigma^2_\varepsilon$$

$$\sigma^2_{(3,t)}(e) = \left[1 + \left[1-\theta_1\right]^2 + (1-\theta_1-\theta_2)^2\right]\sigma^2_\varepsilon$$

.....

$$\sigma^2_{(q-1,t)}(e) = \left[1 + (1-\theta_1)^2 + (1-\theta_1-\theta_2)^2 + ... + (1-\theta_1 - ... - \theta_{q-1})^2\right]\sigma^2_\varepsilon$$

$$\sigma^2_{(\ell,t)}(e) = \left[1 + (1-\theta_1)^2 + (1-\theta_1-\theta_2)^2 + ... + (1-\theta_1 - ... - \theta_q)^2\right]\sigma^2_\varepsilon$$

Supóngase, ahora, un proceso $\{Y_t\}$ que admite una representación $ARIMA(2,2,q)$ de la forma:

$$(1-B)^2 Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-q}$$

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-q}$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:WN - *debil*). Entonces, será:

$$Y_{t+1} = \left(2Y_t - Y_{t-1} - \theta_1\varepsilon_t - \theta_2\varepsilon_{t-1} - ... - \theta_q\varepsilon_{t-(q-1)}\right) + \varepsilon_{t+1}$$

$$Y_{t+2} = 2Y_{t+1} - Y_t + \varepsilon_{t+2} - \theta_1\varepsilon_{t+1} - \theta_2\varepsilon_t - ... - \theta_q\varepsilon_{t-(q-2)} =$$

$$= \left[3Y_t - 2Y_{t-1} - (2\theta_1 + \theta_2)\varepsilon_t - (2\theta_2 + \theta_3)\varepsilon_{t-1} - ... - (2\theta_{q-1} + \theta_q)\varepsilon_{t-(q-2)} - 2\theta_q\varepsilon_{t-(q-1)}\right] +$$

$$+ \left[\varepsilon_{t+2} + (2-\theta_1)\varepsilon_{t+1}\right]$$

.....

$$Y_{t+q-1} = \Big\{ qY_t - (q-1)Y_{t-1} - \big[(q-1)\theta_1 + (q-2)\theta_2 + \ldots + \theta_{q-1}\big]\varepsilon_t -$$

$$- \big[(q-1)\theta_2 + (q-2)\theta_3 + \ldots + \theta_q\big]\varepsilon_{t-1} - \ldots - (q-1)\theta_q\varepsilon_{t-(q-1)} \Big\} +$$

$$+ \Big\{ \varepsilon_{t+(q-1)} + (2-\theta_1)\varepsilon_{t+(q-2)} + \ldots + \big[(q-1) - (q-2)\theta_1 - \ldots - \theta_{q-1}\big]\varepsilon_{t+1} \Big\}$$

$$Y_{t+\ell} = \Big\{ (\ell+1)Y_t - \ell Y_{t-1} - \big[\ell\theta_1 + (\ell-1)\theta_2 + \ldots + (\ell-q+1)\theta_q\big]\varepsilon_t -$$

$$- \big[\ell\theta_2 + (\ell-1)\theta_3 + \ldots + (\ell-q+2)\theta_q\big]\varepsilon_{t-1} - \ldots - \ell\theta_q\varepsilon_{t-(q-1)} \Big\} +$$

$$+ \Big\{ \varepsilon_{t+\ell} + (2-\theta_1)\varepsilon_{t+\ell-1} + (3-2\theta_1-\theta_2)\varepsilon_{t+\ell-2} + \ldots + \big[\ell - (\ell-1)\theta_1 - (\ell-2)\theta_2 - \ldots - \theta_q\big]\varepsilon_{t+1} \Big\} \qquad (\ell \geq q)$$

De modo que la función de predicción será:

$$f^{(t,\ell)} = (\ell+1)Y_t - \ell Y_{t-1} - \sum_{i=1}^{q}\left[\sum_{j=1}^{q}(\ell - j + i)\theta_j\right]\varepsilon_{t-i+1}$$

El error de predicción será:

$$e(\ell,t) = \varepsilon_{t+\ell} + (2-\theta_1)\varepsilon_{t+\ell-1} + (3-2\theta_1-\theta_2)\varepsilon_{t+\ell-2} + \ldots + \big[\ell - (\ell-1)\theta_1 - (\ell-2)\theta_2 - \ldots - \theta_q\big]\varepsilon_{t+1} \qquad (\ell \geq q)$$

y la varianza del error de predicción será:

$$\sigma^2_{(\ell,t)}(e) = \Big\{ 1 + (2-\theta_1)^2 + (3-2\theta_1-\theta_2)^2 + \ldots + \big[\ell - (\ell-1)\theta_1 - (\ell-2)\theta_2 - \ldots - \theta_q\big]^2 \Big\}\sigma_\varepsilon^2 \qquad (\ell \geq q)$$

3.7.- Procesos afectados por factores periódicos

3.7.1.- Las condiciones de estacionariedad e invertibilidad

Se dice que un proceso $\{Y_t\}$ es periódico, con período T, si su comportamiento observa "similitudes" en las realizaciones ocurridas con T **períodos básicos**. En particular, si el período T coincide con el año calendario, ese factor periódico se denomina **estacionalidad**[67].

En estos proceso periódicos el problema a resolver consiste en obtener una aceptable representación parsimoniosa de $\{Y_t\}$. Una solución consiste en eliminar la periodicidad mediante la aplicación de filtros adecuados (en particular, si la influencia periódica es estable en el tiempo, se pueden utilizar **variables "dummy"**), otra (particularmente apropiada en problemas de predicción) consiste en aplicar una diferenciación adecuada.

Debe tenerse en cuenta que, como se mencionó en la Sec. 3.2.1, en la representación:

$$\varphi(B)Y_t = \Theta(B)\varepsilon_t$$

(donde: **i)** $\varphi(B)$ denota un operador AR generalizado; **ii)** Y_t denota una variable centrada y **iii)**

---

[67] . Se denomina **período básico** al período en el cual está expresado el proceso.

$\{\varepsilon_t\}: WN - N(0,\sigma_\varepsilon^2)$) el operador $\varphi(B)$ está definido por una combinación de senos y cosenos combinados, a su vez, con términos polinómicos, originados en las variaciones en el nivel del proceso y en la influencia del factor periódico.

De acuerdo con la definición que figura más arriba, la característica fundamental de los procesos periódicos de período T es que las realizaciones $Y_t, B^T Y_t = Y_{t-T}, B^{2T} Y_t = Y_{t-2T}, \ldots$ son "similares". De modo que un método apropiado para simplificar la representación de $\{Y_t\}$ es aplicar un operador de la forma $\nabla_T Y_t = (1 - B^T) Y_t = Y_t - Y_{t-T}$, conocido como **diferencia periódica de período T.**

Supóngase un proceso $\{Y_t\}$ periódico de parámetro mensual, que presente relaciones significativas entre las variables Y_t correspondientes a los meses sucesivos de un año y entre las variables correspondientes al mismo mes de distintos años. La relación entre las variables Y_t correspondientes a años diferentes puede ser representada de la siguiente forma:

$$\Phi_P^*(B^T) \nabla_T^D Y_t = \Theta_Q^*(B^T) e_t$$

donde $\Phi_P^*(B^T)$ y $\Theta_Q^*(B^T)$ denotan operadores AR (estacionario) y MA (invertible), definidos por polinomios de grado P y Q, respectivamente. De la misma forma, para Y_{t-j}, se puede escribir una representación de la forma:

$$\Phi_P^*(B^T) \nabla_T^D Y_{t-j} = \Theta_Q^*(B^T) e_{t-j} \qquad (j = 1,2,\ldots.)$$

Dada la ya mencionada condición de periódico de período T del proceso $\{Y_t\}$, es de esperar que estos operadores $\Phi_P^*(B^T)$ y $\Theta_Q^*(B^T)$ sean, respectivamente, aproximadamente iguales a los de la representación de $\{Y_t\}$. En general se verifica que las variables $e_t, e_{t-1}, e_{t-2}, \ldots$ obtenidas de aplicar estas representaciones presentan autocorrelaciones significativas y admiten una representación de la forma:

$$\Phi_p(B) \nabla^d e_t = \Theta_q(B) \varepsilon_t$$

donde: **i)** $\Phi_p(B)$ y $\Theta_q(B)$ denotan operadores AR (estacionario) y MA (invertible) de orden p y q, respectivamente y **ii)** $\{\varepsilon_t\}: WN - N(0,\sigma_\varepsilon^2)$.

Sustituyendo esta última expresión en la representación anterior, se obtiene una **representación multiplicativa de orden** $(p,d,q)X(P,D,Q)_T$ de la forma:

$$\Phi_p(B) \Phi_P^*(B^T) \nabla^d \nabla_T^D Y_t = \Theta_q(B) \Theta_Q^*(B^T) \varepsilon_t$$

(una generalización de esta representación puede ser utilizada en el caso en que $\{Y_t\}$ presente varias influencias periódicas)[68].

La condición de estacionariedad implica que todas las raíces de la ecuación característica correspondiente a la parte AR de la representación, $\Phi_p \Phi_P^*(B^T) = 0$, sean en módulo mayores que la unidad

---

[68]. Como se verá en la Sec. 3.7.3, estas representaciones multiplicativas pueden constituir un punto de partida para la construcción de representaciones no-multiplicativas más generales.

y la condición de invertibilidad implica que todas las raíces de la ecuación característica correspondiente a la parte MA de la representación, $\Theta q(B)\Theta_Q^*(B^T) = 0$, sean en módulo mayores que la unidad.

3.7.2.- La función de autocorrelaciones en las representaciones multiplicativas

A continuación se analizan algunas representaciones periódicas que se presentan frecuentemente en las aplicaciones:

1) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,1)\times(0,1,1)_T$ de la forma:

$$\nabla\nabla_T Y_t = (1-B)(1-B^T)Y_t = Z_t = \Theta_1(B)\Theta_1^*(B^T)\varepsilon_t = (1-\theta_1 B)(1-\theta_T^* B^T)\varepsilon_t =$$
$$= \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T^*\varepsilon_{t-T} + \theta_1\theta_T^*\varepsilon_{t-T-1} \qquad (T \geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$).

Multiplicando m. a m. por Z_{t-j} y aplicando m.a m. el operador esperanza matemática, se obtiene la función de autocovarianzas:

$$E(Z_{t-j}Z_t) = E\left[\left(\varepsilon_{t-j} - \theta_1\varepsilon_{t-j-1} - \theta_T^*\varepsilon_{t-j-T} + \theta_1\theta_T^*\varepsilon_{t-j-T-1}\right)\left(\varepsilon_t - \theta_1\varepsilon_{t-1} + \theta_1\theta_T^*\varepsilon_{t-T-1}\right)\right]$$

es decir, resulta que:

$$\gamma_j(Z) = \begin{cases} (1+\theta_1^2)(1+\theta_T^{*2})\sigma_\varepsilon^2 & (j=0) \\ -\theta_1(1+\theta_T^{*2})\sigma_\varepsilon^2 & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ -\theta_T^*(1+\theta_1^2)\sigma_\varepsilon^2 & (j=T) \\ \theta_1\theta_T^*\sigma_\varepsilon^2 & (j=T-1,T+1) \\ 0 & (j=T+2,T+3,\ldots) \end{cases}$$

y, a partir de esta función se obtiene la siguiente función de autocorrelaciones:

$$\rho_j(Z) = \begin{cases} -\dfrac{\theta_1}{1+\theta_1^2} & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ -\dfrac{\theta_T^*}{1+\theta_T^{*2}} & (j=T) \\ \dfrac{\theta_1\theta_T^*}{(1+\theta_1^2)(1+\theta_T^{*2})} = \rho_1(Z)\rho_T(Z) & (j=T-1,T+1) \\ 0 & (j=T+2,T+3,\ldots) \end{cases}$$

(ver **GRAFICO 3.11**)[69].

Sean los procesos $\{Y_t^*\}$: $MA(1)$ y $\{Z_t^*\}$: $MA(T)$. Entonces, de acuerdo a los resultados obtenidos en la Sec. 2.11.4, será:

$$\rho_j(Y^*) = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & (j=1) \\ 0 & (j>1) \end{cases}$$

$$\rho_j(Z^*) = \begin{cases} \dfrac{-\theta_T^*}{1+\theta_T^{*2}} & (j=T) \\ 0 & (j>0, j\neq T) \end{cases}$$

Luego, comparando estos resultados, se obtiene que:

$$\rho_1(Z) = \rho_1(Y^*)$$

$$\rho_T(Z) = \rho_T(Z^*)$$

$$\rho_{T-1}(Z) = \rho_{T+1}(Z) = \rho_1(Y^*)\rho_T(Z^*)$$

De las dos primera igualdades se puede concluir que $\rho_1(Z)$ no está afectada por la influencia del factor MA-estacional $(1-\theta_1^* B^T)$, que $\rho_T(Z)$ no está afectada por el factor MA-regular $(1-\theta_1 B)$ y que, en consecuencia, admitir en este caso una representación multiplicativa implica considerar que las componentes regular y estacional son aproximadamente ortogonales. La tercera igualdad demuestra que los coeficientes $\rho_j(Z)\,(j=T-1, T+1)$ pueden ser considerados como representativos de las interacción de las autocorrelaciones de orden 1 y T (esta circunstancia justifica la denominación de representación multiplicativa).

Obsérvese que: **i)** la función de autocorrelaciones del proceso $\{\nabla\nabla_T Y_t\} = \{Z_t\}$ presenta "picos" (o "valles") en los órdenes $j=1$ y $j=T$ y **ii)** la función de autocorrelaciones de $\{\nabla_T Y_t\}$ decrece a una tasa exponencial (lenta).

2) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,2)\times(0,1,1)_T$ de la forma:

$$\nabla\nabla_T Y_t = (1-B)(1-B^T)Y_t = Z_t = (1-\theta_1 B-\theta_2 B^2)(1-\theta_T^* B^T)\varepsilon_t =$$

$$= \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \theta_T^*\varepsilon_{t-T} + \theta_1\theta_T^*\varepsilon_{t-T-1} + \theta_2\theta_T^*\varepsilon_{t-T-2} \qquad (T\geq 5)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN-N(0,\sigma_\varepsilon^2)$). La función de autocovarianzas será de la forma:

---

[69] . En particular, si $\theta_T^*=1$, entonces la influencia estacional será determinística, por lo que se puede concluir que la **estacionalidad determinística** es un caso particular de la representación multiplicativa

GRAFICO 3.11: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.6\ B)(1-0.7\ B^{12})\varepsilon_t$

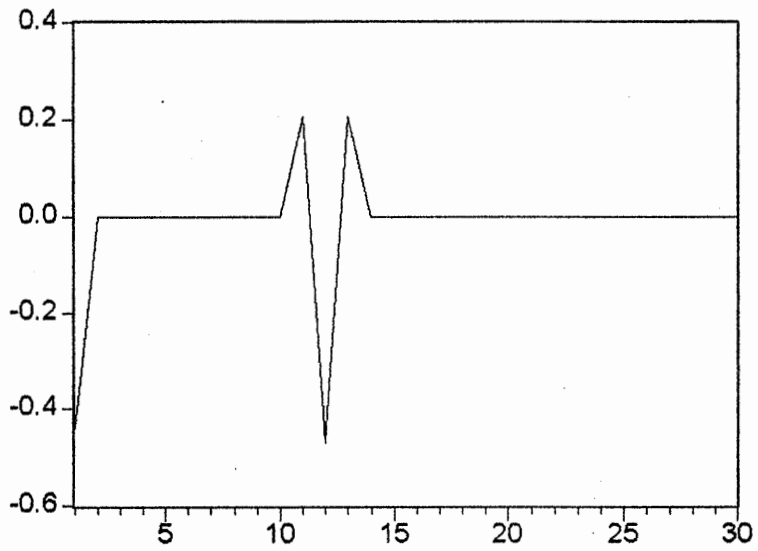
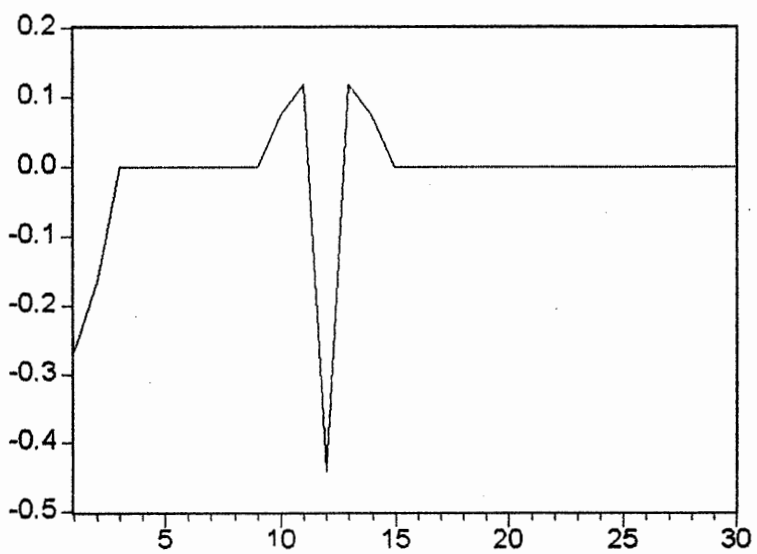


GRAFICO 3.12: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.4\ B-0.6\ B^2)(1-0.6\ B^{12})\varepsilon_t$



$$\gamma_j(Z) = \begin{cases} \left(1+\theta_1^2+\theta_2^2\right)\left(1+\theta_T^{*2}\right)\sigma_\varepsilon^2 & (j=0) \\ -\theta_1\left(1-\theta_2\right)\left(1+\theta_T^{*2}\right)\sigma_\varepsilon^2 & (j=1) \\ -\theta_2\left(1+\theta_T^{*2}\right)\sigma_\varepsilon^2 & (j=2) \\ 0 & (j=3,4,\ldots,T-3) \\ \theta_2\theta_T^*\sigma_\varepsilon^2 & (j=T-2,T+2) \\ \theta_1\theta_T^*\left(1-\theta_2\right)\sigma_\varepsilon^2 & (j=T-1,T+1) \\ \dfrac{-\theta_T^*}{1+\theta_T^{*2}}\sigma_\varepsilon^2 & (j=T) \\ 0 & (j=T+3,T+4,\ldots) \end{cases}$$

y, en consecuencia, la función de autocorrelaciones será:

$$\rho_j(Z) = \begin{cases} \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & (j=1) \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} & (j=2) \\ 0 & (j=3,4,\ldots,T-3) \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_T^*}{1+\theta_T^{*2}} & (j=T-2,T+2) \\ \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_T^*}{1+\theta_T^{*2}} & (j=T-1,T+1) \\ \dfrac{-\theta_T^*}{1+\theta_T^{*2}} & (j=T) \\ 0 & (j=T+3,T+4,\ldots) \end{cases}$$

(ver **GRAFICO 3.12**).

Sean los procesos $\{Y_t^*\}: MA(2)$ y $\{Z_t^*\}: MA(T)$. A partir de los resultados precedentes se obtiene que:

$$\rho_1(Z) = \rho_1\left(Y^*\right)$$

$$\rho_2(Z) = \rho_2\left(Y^*\right)$$

$$\rho_T(Z) = \rho_T\left(Z^*\right)$$

$$\rho_{T-2}(Z) = \rho_{T+2}(Z) = \rho_2\left(Y^*\right)\rho_T\left(Z^*\right)$$

$$\rho_{T-1}(Z) = \rho_{T+1}(Z) = \rho_1(Y^*)\rho_T(Z^*)$$

3) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,1)\times(0,1,2)_T$ de la forma:

$$\nabla\nabla_T Y_t = (1-B)(1-B^T)Y_t = Z_t = (1-\theta_1 B)(1-\theta_T^* B^T - \theta_{2T}^* B^2)\varepsilon_t =$$

$$= \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T^*\varepsilon_{t-T} + \theta_1\theta_T^*\varepsilon_{t-T-1} - \theta_{2T}^*\varepsilon_{t-2T} + \theta_1\theta_{2T}^*\varepsilon_{t-2T-1} \qquad (T \geq 3)$$

(donde: i) Z_t denota una variable centrada y ii) $\{\varepsilon_t\}: WN - N(0,\sigma_\varepsilon^2)$). Multiplicando m. a m. por Z_t y aplicando m. a m. el operador esperanza matemática, se obtiene que:

$$\sigma_Z^2 = (1+\theta_1^2)(1+\theta_T^{*2}+\theta_{2T}^{*2})\sigma_\varepsilon^2$$

La función de autocorrelaciones será de la forma:

$$\rho_j(Z) = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ \dfrac{-\theta_1}{1+\theta_1^2}\dfrac{-\theta_T^*+\theta_T^*\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=T-1,T+1) \\ \dfrac{-\theta_T^*+\theta_T^*\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=T) \\ 0 & (j=T+2,T+3,\ldots,2T-2) \\ \dfrac{-\theta_1}{1+\theta_1^2}\dfrac{-\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=2T-1,2T+1) \\ \dfrac{-\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=2T) \\ 0 & (j=2T+2,2T+3,\ldots) \end{cases}$$

(ver **GRAFICO 3.13**).

Sean los procesos $\{Y_t^*\}: MA(1)$ y $\{Z_t^*\}: MA(2T)$. A partir de los resultados precedentes se obtiene que:

$$\rho_1(Z) = \rho_1(Y^*)$$

$$\rho_T(Z) = \rho_T(Z^*)$$

$$\rho_{T-1}(Z) = \rho_{T+1}(Z) = \rho_1(Y^*)\rho_T(Z^*)$$

$$\rho_{2T-1}(Z) = \rho_{2T+1}(Z) = \rho_1(Y^*)\rho_{2T}(Z^*)$$

$$\rho_{2T}(Z) = \rho_{2T}(Z^*)$$

GRAFICO 3.13: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.6\ B)(1-0.4\ B^{12}-0.2\ B^{24})\varepsilon_t$

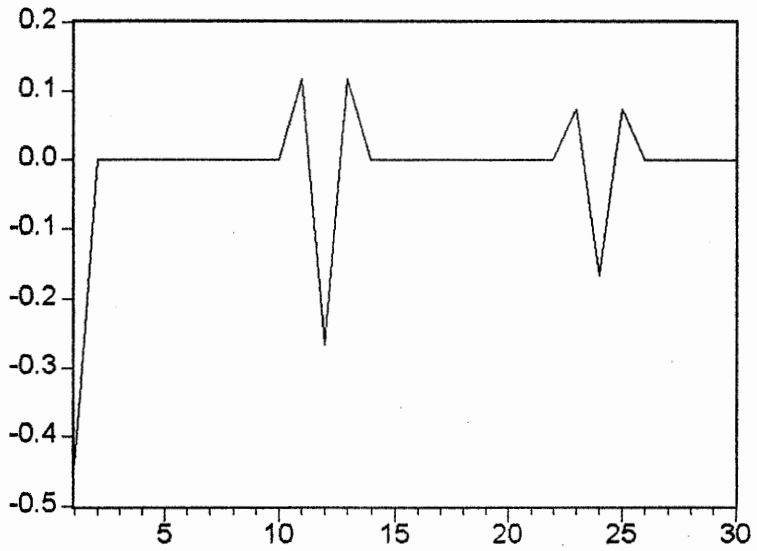
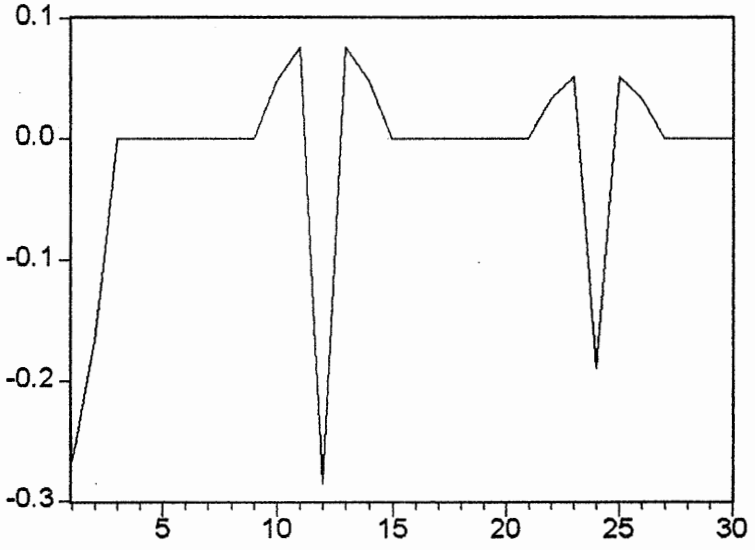


GRAFICO 3.14: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.4\ B-0.2\ B^2)(1-0.5\ B^{12}-0.25\ B^{24})\varepsilon_t$



4) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,2)\times(0,1,2)_T$ de la forma:

$$\nabla\nabla_T Y_t = (1-B)(1-B^T)Y_t = Z_t = (1-\theta_1 B-\theta_2 B^2)(1-\theta_T^* B^T-\theta_{2T}^* B^{2T})\varepsilon_t =$$

$$= \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \theta_T^*\varepsilon_{t-T} + \theta_1\theta_T^*\varepsilon_{t-T-1} + \theta_2\theta_T^*\varepsilon_{t-T-2} -$$

$$- \theta_{2T}^*\varepsilon_{t-2T} + \theta_1\theta_{2T}^*\varepsilon_{t-2T-1} + \theta_2\theta_{2T}^*\varepsilon_{t-2T-2} \qquad (T \geq 5)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN-N(0,\sigma_\varepsilon^2)$). La varianza del proceso $\{Z_t\}$ será de la forma:

$$\sigma_Z^2 = (1+\theta_1^2+\theta_2^2)(1+\theta_T^{*2}+\theta_{2T}^{*2})\sigma_\varepsilon^2$$

y la función de autocorrelaciones asume los siguientes valores:

$$\rho_j(Z) = \begin{cases} \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & (j=1) \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} & (j=2) \\ 0 & (j=3,4,\ldots,T-3) \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_T^*+\theta_T^*\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=T-2,T+2) \\ \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_T^*+\theta_T^*\theta_{2T}^{*2}}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=T-1,T+1) \\ \dfrac{-\theta_T^*+\theta_T^*\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=T) \\ 0 & (j=T+3,T+4,\ldots,2T-3) \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=2T-2,2T+2) \\ \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}\dfrac{-\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^2} & (j=2T-1,2T+1) \\ \dfrac{-\theta_{2T}^*}{1+\theta_T^{*2}+\theta_{2T}^{*2}} & (j=2T) \\ 0 & (j=2T+3,2T+4,\ldots) \end{cases}$$

(ver **GRAFICO 3.14**).

5) Sea, en general, un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,q)\times(0,1,Q)_T$ de la forma:

$$\nabla\nabla_T Y_t = (1-B)(1-B^T)Y_t = Z_t = (1-\theta_1 B-\ldots-\theta_q B^q)(1-\theta_T^* B^T-\ldots-\theta_{QT}^* B^{QT}) \qquad (T \geq q+Q+1)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN-N(0,\sigma_\varepsilon^2)$). La varianza del proceso $\{Z_t\}$ será de la forma:

$$\sigma_Z^2 = (1+\theta_1^2+\ldots+\theta_q^2)(1+\theta_T^{*2}+\ldots+\theta_{QT}^{*2})\sigma_\varepsilon^2$$

y la función de autocorrelaciones asume los siguientes valores:

$$\rho_j(Z) = \begin{cases} \dfrac{-\theta_j+\theta_1\theta_{j+1}+\ldots+\theta_{q-j}\theta_q}{1+\theta_1^2+\ldots+\theta_q^2} & (j=1,2,\ldots,q) \\ 0 & (j=q+1,\ldots,T-q-1) \\ \dfrac{-\theta_q}{1+\theta_1^2+\ldots+\theta_q^2}\dfrac{-\theta_{hT}^*+\theta_T^*\theta_{(h+1)T}^*+\ldots+\theta_{(Q-1)T}^*\theta_{QT}^*}{1+\theta_T^{*2}+\ldots+\theta_{QT}^{*2}} & (j=hT-q,hT+q;h=1,2,\ldots,Q) \\ \dfrac{-\theta_q+\theta_1\theta_q}{1+\theta_1^2+\ldots+\theta_q^2}\dfrac{-\theta_{hT}^*+\theta_T^*\theta_{(h+1)T}^*+\ldots+\theta_{(Q-1)T}^*\theta_{QT}^*}{1+\theta_T^{*2}+\ldots+\theta_{QT}^{*2}} & (j=hT-(q-1),hT+(q-1);h=1,2,\ldots,Q) \\ \dfrac{-\theta_1+\theta_1\theta_2+\ldots+\theta_{q-1}\theta_q}{1+\theta_1^2+\ldots+\theta_q^2}\dfrac{-\theta_{hT}^*+\theta_T^*\theta_{(h+1)T}^*+\ldots+\theta_{(Q-1)T}^*\theta_{QT}^*}{1+\theta_T^{*2}+\ldots+\theta_{QT}^{*2}} & (j=hT-1,hT+1;h=1,2,\ldots,Q) \\ \dfrac{-\theta_{hT}^*+\theta_T^*\theta_{(h-1)T}^*+\ldots+\theta_{(Q-1)T}^*\theta_{QT}^*}{1+\theta_T^{*2}+\ldots+\theta_{QT}^{*2}} & (j=ht;h=1,2,\ldots,Q) \\ 0 & (j=hT+q+1,\ldots,(h+1)T-q-1;h=1,2,\ldots,Q-1) \end{cases}$$

6) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,0)X(1,1,0)_T$ de la forma:

$$\nabla\nabla_T(1-\phi_{1T}^* B^T)Y_t = (1-\phi_{1T}^* B^T)Z_t = Z_t-\phi_{1T}^* Z_{t-T} = \varepsilon_t \qquad (T \geq 2)$$

Es decir, tal que:

$$Z_t = (1-\phi_{1T}^* B^T)^{-1}\varepsilon_t = (1+\phi_{1T}^* B^T+\phi_{1T}^{*2}B^{2T}+\ldots)\varepsilon_t =$$
$$= \varepsilon_t+\phi_{1T}^*\varepsilon_{t-T}+\phi_{1T}^{*2}\varepsilon_{t-2T}+\phi_{1T}^{*3}\varepsilon_{t-3T}+\ldots$$

La función de autocovarianzas del proceso $\{Z_t\}$ será de la forma:

$$\gamma_j(Z) = \begin{cases} (1+\phi_{1T}^{*2}+\phi_{1T}^{*4}+\ldots)\sigma_\varepsilon^2 = \dfrac{\sigma_\varepsilon^2}{1-\phi_{1T}^{*2}} & (j=0) \\ \dfrac{\phi_{1T}^{*h}}{1-\phi_{1T}^{*2}}\sigma_\varepsilon^2 & (j=hT;h=1,2,\ldots) \\ 0 & (\textit{para culaquier otro valor de } j) \end{cases}$$

y, en consecuencia, la función de autocorrelaciones será:

$$\rho_j(Z) = \begin{cases} \phi_{1T}^{*h} & (j = hT; h = 1,2,...) \\ 0 & (\textit{para cualquier otro valor de } j) \end{cases}$$

(ver **GRAFICO 3.15**).

7) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,0)\times(2,1,0)_T$ de la forma:

$$\nabla\nabla_T\left(1-\phi_{1T}^{*}B^T-\phi_{2T}^{*}B^{2T}\right)Y_t = \left(1-\phi_{1T}^{*}B^T-\phi_{2T}^{*}B^{2T}\right)Z_t = Z_t-\phi_{1T}^{*}Z_{t-T}-\phi_{2T}^{*}Z_{t-2T} = \varepsilon_t \qquad (T \geq 2)$$

Es decir, tal que:

$$Z_t = \left(1-\phi_{1T}^{*}B^T-\phi_{2T}^{*}B^{2T}\right)^{-1}\varepsilon_t = \left(1+\psi_{1T}^{*}B^T+\psi_{2T}^{*}B^{2T}+...\right)\varepsilon_t =$$

$$= \varepsilon_t+\psi_{1T}^{*}\varepsilon_{t-T}+\psi_{2T}^{*}\varepsilon_{t-2T}+...$$

donde:

$$\psi_{(2j)T}^{*} = \binom{2j}{0}\phi_{1T}^{*(2j)}+\binom{2j-1}{1}\phi_{1T}^{*(2j-2)}\phi_{2T}^{*}+\binom{2j-2}{2}\phi_{1T}^{*(2j-4)}\phi_{2T}^{*2}+...+\binom{j+1}{j}\phi_{1T}^{*}\phi_{2T}^{*j} \quad (j = 0,1,2,...)$$

$$\psi_{(2j+1)T}^{*} = \binom{2j+1}{0}\phi_{1T}^{*(2j+1)}+\binom{2j}{1}\phi_{1T}^{*(2j-1)}\phi_{2T}^{*}+\binom{2j-1}{2}\phi_{1T}^{*(2j-3)}\phi_{2T}^{*2}+...+\binom{j+1}{j}\phi_{1T}^{*}\phi_{2T}^{*2j} \quad (j = 0,1,2,...)$$

La función de autocovarianzas será de la forma:

$$\gamma_j(Z) = \begin{cases} \left(\sum_{i=0}^{\infty}\psi_{iT}^{*2}\right)\sigma_\varepsilon^2 & (j = 0; \psi_o^{*} = 1) \\ \left(\sum_{i=0}^{\infty}\psi_{iT}^{*}\psi_{(h+i)T}^{*}\right)\sigma_\varepsilon^2 & (j = hT; h = 1,2,...; \psi_0^{*} = 1) \\ 0 & (\textit{para cualquier otro valor de } j) \end{cases}$$

y, en consecuencia, la función de autocorrelaciones será:

$$\rho_j(Z) = \begin{cases} \dfrac{\sum_{i=0}^{\infty}\psi_{iT}^{*}\psi_{(h+i)T}^{*}}{\sum_{i=0}^{\infty}\psi_{iT}^{*2}} & (j = hT; h = 1,2,...; \psi_0^{*} = 1) \\ 0 & (\textit{para cualquier otro valor de } j) \end{cases}$$

(ver **GRAFICO 3.16**).

GRAFICO 3.15: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.7\ B^4)-1\varepsilon_t$

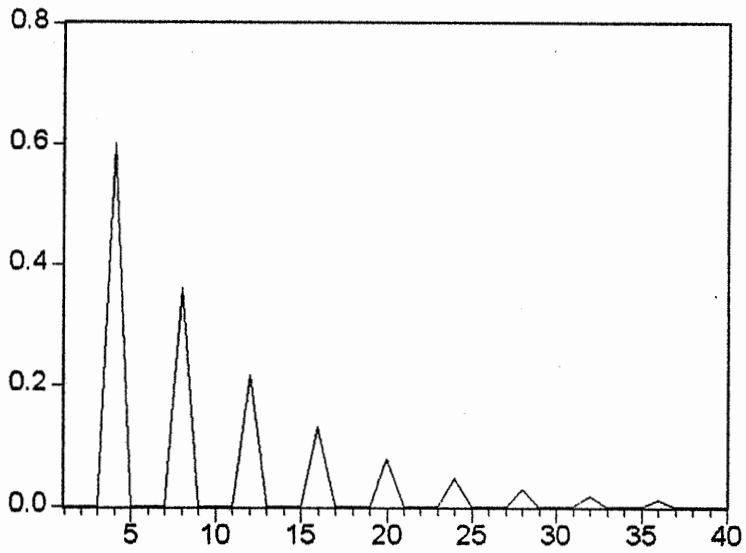


GRAFICO 3.16: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.4\ B^4-0.2\ B^2)-1\varepsilon_t$

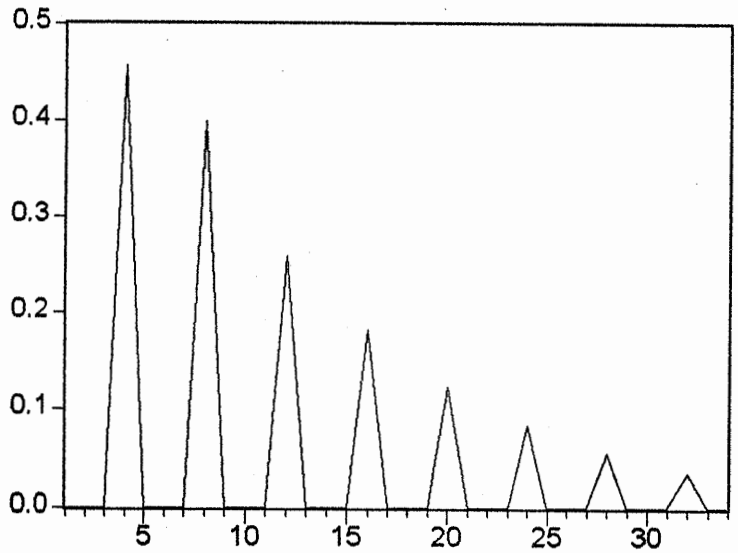
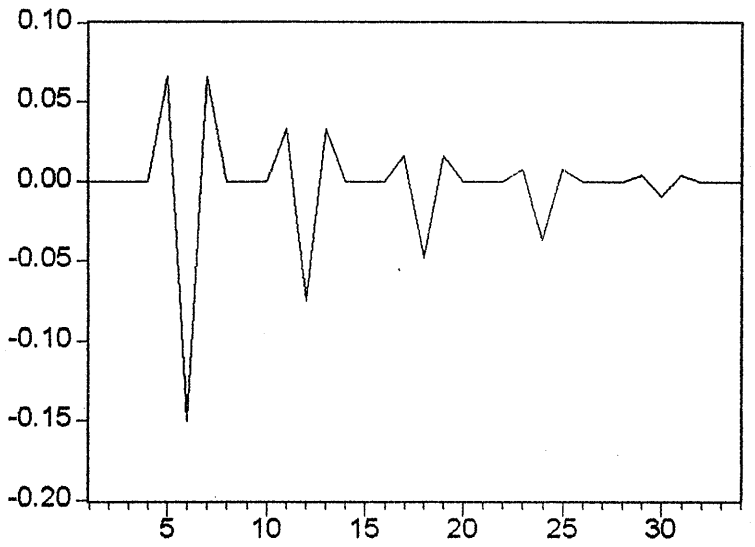


GRAFICO 3.17: Función de autocorrelaciones correspondiente al proceso $Z_t=(1-0.5\ B^6)-1(1-0.6\ B)(1-0.7\ B^6)\varepsilon_t$



8) Sea un proceso $\{Y_t\}$ que admite una representación de orden $(0,1,1)\times(1,1,1)_T$ de la forma:

$$\nabla\nabla_T\left(1-\phi_{1T}^*B^T\right)Y_t=\left(1-\phi_{1T}^*B^T\right)Z_t=Z_t-\phi_{1T}^*Z_{t-T}=\left(1-\theta_1B\right)\left(1-\theta_{1T}^*\right)\varepsilon_t$$

$$Z_t=\phi_{1T}^*Z_{t-T}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_{1T}^*\varepsilon_{t-T}+\theta_1\theta_{1T}^*\varepsilon_{t-T-1} \qquad (T\geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN-N\left(0,\sigma_\varepsilon^2\right)$). La función de autocovarianzas será de la forma:

$$\gamma_j(Z)=\begin{cases}\left(1+\theta_1^2\right)\left[1+\dfrac{\left(\theta_{1T}^*-\phi_{1T}^*\right)}{1-\phi_{1T}^{*2}}\right] & (j=0)\\ 0 & (j=1,2,\ldots,T-2)\\ \theta_1\left[\theta_{1T}^*-\phi_{1T}^*-\dfrac{\phi_{1T}^*\left(\theta_{1T}^*-\phi_{1T}^*\right)^2}{1-\phi_{1T}^{*2}}\right]\sigma_\varepsilon^2 & (j=T-1,T+1)\\ -\left(1+\theta_1^2\right)\left[\theta_{1T}^*-\phi_{1T}^*-\dfrac{\phi_{1T}^*\left(\theta_{1T}^*-\phi_{1T}^*\right)^2}{1-\phi_{1T}^{*2}}\right]\sigma_\varepsilon^2 & (j=T)\\ \phi_{1T}^*\gamma_{j-T}(Z) & (j=T+2,T+3,\ldots)\end{cases}$$

Del comportamiento de estas autocovarianzas se puede concluir fácilmente que la representación

ARMA sobre el proceso $\{\nabla\nabla_T Y_t\}$ puede ser interpretada como una forma especial $ARMA(p+TP, q+TQ)$ (en el **GRAFICO 3.17** figura la correspondiente función de autocorrelaciones).

3.7.3.- La función de autocorrelaciones en las representaciones no-multiplicativas

Las representaciones multiplicativas no siempre proporcionan una explicación adecuada del comportamiento de un proceso periódico. Una forma de representación alternativa es la no-multiplicativa, que consiste en reemplazar el operador

$$\Theta_q(B)\Theta^*_Q(B^T)$$

por un operador más general de la forma $\Theta^*_{q^*}(B)$ y, alternativamente o adicionalmente, el operador AR:

$$\Phi_p(B)\Phi^*_P(B^T)$$

por uno más general de la forma $\Phi^*_{p^*}(B)$. Las representaciones multiplicativas como las analizadas en la sección precedente, pueden constituir un punto de partida para la construcción de representaciones no-multiplicativas más generales.

A continuación se analizan algunas representaciones no-multiplicativas que se presentan frecuentemente en las aplicaciones:

1) Sea un proceso $\{Y_t\}$ que admite una representación de la forma:

$$\nabla\nabla_T Y_t = Z_t = (1 - \theta_1 B - \theta_T B^T)\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T\varepsilon_{t-T} \qquad (T \geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma^2_\varepsilon)$). La función de autocovarianzas será de la forma:

$$\gamma_j(Z) = \begin{cases} (1+\theta_1^2+\theta_T^2)\sigma^2_\varepsilon & (j=0) \\ -\theta_1\sigma^2_\varepsilon & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ \theta_1\theta_T\sigma^2_\varepsilon & (j=T-1) \\ -\theta_T\sigma^2_\varepsilon & (j=T) \\ 0 & (j=T+1,T+2,\ldots) \end{cases}$$

2) Sea un proceso $\{Y_t\}$ que admite una representación de la forma:

$$\nabla\nabla_T Y_t = Z_t = (1 - \theta_1 B - \theta_T B^T - \theta_{T+1} B^{T+1})\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T\varepsilon_{t-T} - \theta_{T+1}\varepsilon_{t-T-1} \qquad (T \geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N(0,\sigma^2_\varepsilon)$). La función de autocovarianzas será de la forma:

$$\gamma_j(Z) = \begin{cases} \left(1+\theta_1^2+\theta_T^2+\theta_{T+1}^2\right)\sigma_\varepsilon^2 & (j=0) \\ \left(-\theta_1+\theta_T\theta_{T+1}\right)\sigma_\varepsilon^2 & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ \theta_1\theta_T\sigma_\varepsilon^2 & (j=T-1) \\ \left(\theta_1\theta_{T+1}-\theta_T\right)\sigma_\varepsilon^2 & (j=T) \\ -\theta_{T+1} & (j=T+1) \\ 0 & (j=T+2,T+3,\ldots) \end{cases}$$

3) Sea un proceso $\{Y_t\}$ que admite una representación de la forma:

$$\nabla\nabla_T\left(1-\phi_T^*B^T\right)Y_t = \left(1-\phi_T^*B^T\right)Z_t = Z_t - \phi_T^*Z_{t-T} =$$
$$= \left(1-\theta_1B-\theta_TB^T\right)\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T\varepsilon_{t-T} \qquad (T \geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$). La función de autocovarianzas será de la forma:

$$\gamma_j(Z) = \begin{cases} \sigma_\varepsilon^2 + \dfrac{\theta_1^2+\left(\theta_T-\phi_T^*\right)^2}{1-\phi_T^{*2}}\sigma_\varepsilon^2 & (j=0) \\ -\theta_1\left[1+\phi_T^*\dfrac{\theta_T-\phi_T^*}{1-\phi_T^{*2}}\right]\sigma_\varepsilon^2 & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ \dfrac{\theta_1\left(\theta_T-\phi_T^*\right)}{1-\phi_T^{*2}}\sigma_\varepsilon^2 & (j=T-1) \\ \dfrac{\phi_T^*\theta_1^2-\left(\theta_T-\phi_T^*\right)\left(1-\phi_T^*\theta_T\right)}{1-\phi_T^{*2}}\sigma_\varepsilon^2 & (j=T) \\ \phi_T^*\gamma_{j-T}(Z) & (j=T+1,T+2,\ldots) \end{cases}$$

4) Sea un proceso $\{Y_t\}$ que admite una representación de la forma:

$$\nabla\nabla_T\left(1-\phi_T^*B^T\right)Y_t = \left(1-\phi_T^*B^T\right)Z_t = Z_t - \phi_T^*Z_{t-T} =$$
$$= \left(1-\theta_1B-\theta_TB^T-\theta_{T+1}B^{T+1}\right)\varepsilon_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T\varepsilon_{t-T} - \theta_{T+1}\varepsilon_{t-T-1} \qquad (T \geq 3)$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$). La función de autocorrelaciones será de la forma:

$$\gamma_j(Z) = \begin{cases} \left[1 + \theta_1^2 + \dfrac{\left(\theta_T - \phi_T^*\right)^2}{1-\phi_T^{*2}} + \dfrac{\left(\theta_{T+1} + \theta_1\phi_T^*\right)^2}{1-\phi_T^{*2}}\right]\sigma_\varepsilon^2 & (j=0) \\ \left[-\theta_1 + \dfrac{\left(\theta_T - \phi_T^*\right)\left(\theta_{T+1} + \theta_1\phi_T^*\right)}{1-\phi_T^{*2}}\right]\sigma_\varepsilon^2 & (j=1) \\ 0 & (j=2,3,\ldots,T-2) \\ \left(\theta_T - \phi_T^*\right)\left(\theta_1 + \phi_T^* \dfrac{\theta_{T+1} - \phi_T^*\theta_1}{1-\phi_T^{*2}}\right)\sigma_\varepsilon^2 & (j=T-1) \\ \left[-\left(\theta_T - \phi_T^*\right)\left(1 - \phi_T^* \dfrac{\theta_T - \phi_T^*}{1-\phi_T^{*2}}\right) + \left(\theta_{T+1} + \theta_1\phi_T^*\right)\left(\theta_1 + \phi_T^* \dfrac{\theta_{T+1} + \theta_1\phi_T^*}{1-\phi_T^{*2}}\right)\right]\sigma_\varepsilon^2 & (j=T) \\ -\left(\theta_{T+1} - \theta_1\phi_T^*\right)\left(1 - \phi_T^* \dfrac{\theta_T - \phi_T^*}{1-\phi_T^{*2}}\right)\sigma_\varepsilon^2 & (j=T+1) \\ \phi_T^*\gamma_j(Z) & (j=T+2,T+3,\ldots) \end{cases}$$

De los resultados anteriores se puede concluir que las representaciones multiplicativas son más parsimoniosas que las no-multiplicativas ya que, si bien involucran el mismo número de coeficientes, las multiplicativas presentan mayor número de coeficientes de autocorrelación significativos.

3.7.4.- La estimación de la función de autocorrelaciones

A fin de decidir acerca de la significatividad de los coeficientes $\rho_j(Z)$, igual que en el caso de los procesos no-estacionales, se puede construir un test que considera una hipótesis nula $H_0{:}\rho_j(Z) = 0$ con respecto a la alternativa $H_1{:}\rho_j(Z) \neq 0$, a partir de la información que proporcionan los correspondientes estimadores muestrales.

Para valores de j suficientemente grandes y, bajo el supuesto que la hipótesis nula es verdadera, se verifica que $\hat{\rho}_j(Z){:}N\left[0,\sigma^2\left(\hat{\rho}(Z)\right)\right]$ donde, de acuerdo con la fórmula de Bartlett mencionada en la Sec. 2.8, la varianza puede ser aproximada de la siguiente forma:

$$\sigma^2\left(\hat{\rho}_j(Z)\right) \approx \frac{1}{n}\left[1 + 2\sum_{i=1}^{j-1}\rho_i^2(Z)\right]$$

3.7.5.- La estimación de los coeficientes

Sea un proceso $\{Y_t\}$ que admite una representación de la forma:

$$\nabla\nabla_T Y_t = Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_T^*\varepsilon_{t-T} + \theta_1\theta_T^*\varepsilon_{t-T-1}$$

(donde: **i)** Z_t denota una variable centrada y **ii)** $\{\varepsilon_t\}{:}WN - N\left(0,\sigma_\varepsilon^2\right)$). Una estimación preliminar de los

GRAFICO 3.18: Trayectoria del proceso $\{Y_t\}$

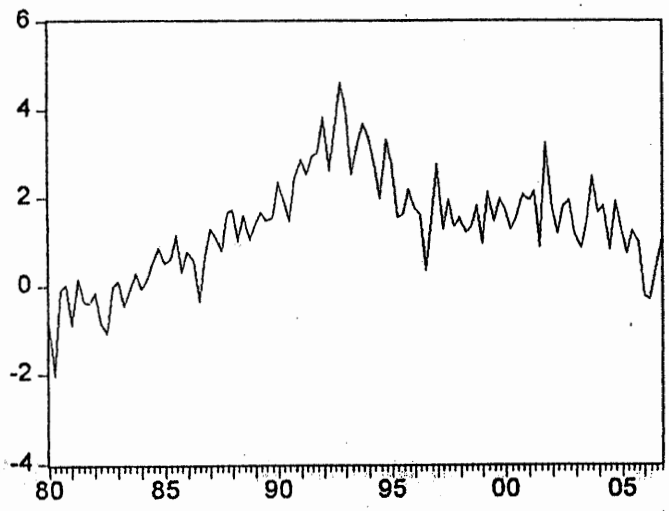


GRAFICO 3.19: Correlograma correspondiente al proceso $\{\nabla Y_t\}$

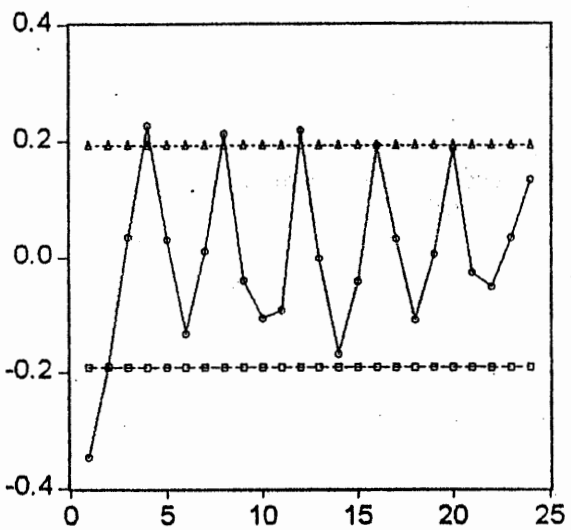


GRAFICO 3.20: Correlograma correspondiente al proceso $\{\nabla^4\nabla Y_t\}=\{Z_t\}$

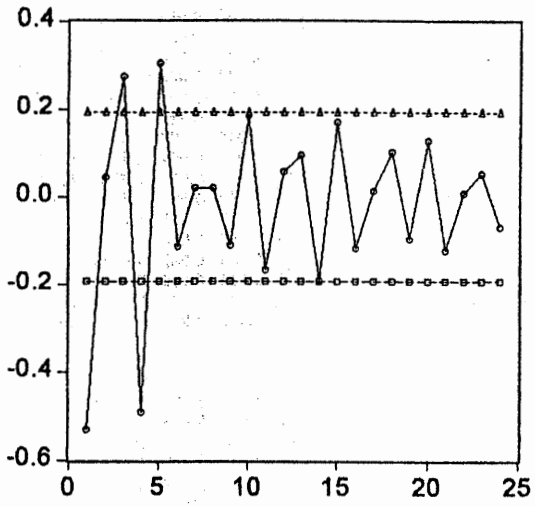
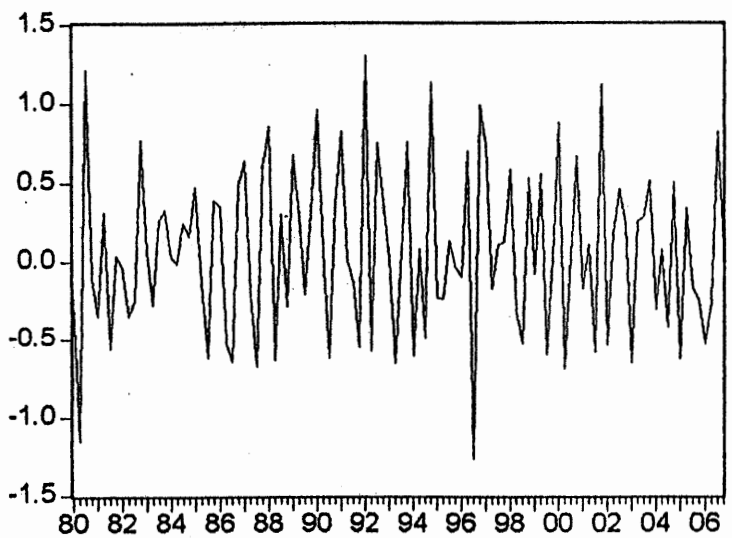
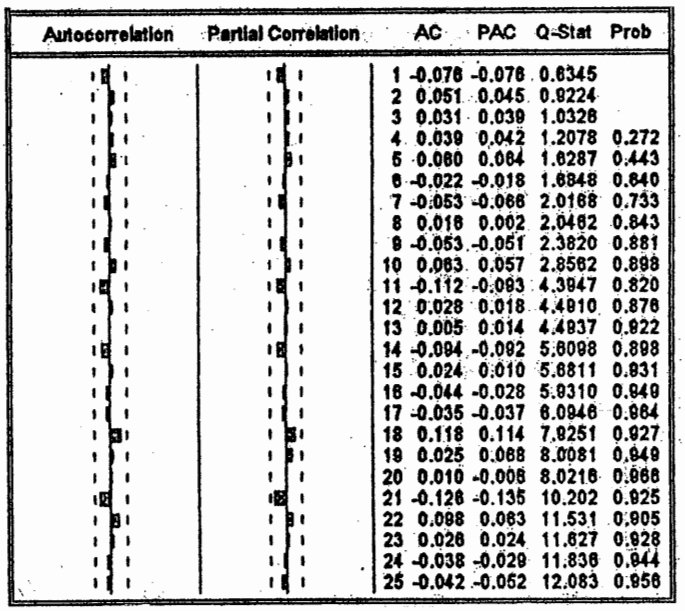


GRAFICO 3.21: Trayectoria del proceso $\{\nabla Y_t\}$



CUADRO 3.6: Correlograma correspondiente a la serie de los residuos que resulta de aplicar el modelo $Z_t=\varepsilon_t-0.667\ \varepsilon_{t-1}-0.846\ \varepsilon_{t-4}+0.475\ \varepsilon_{t-5}$



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.076	-0.076	0.6345	
		2	0.051	0.045	0.9224	
		3	0.031	0.039	1.0328	
		4	0.039	0.042	1.2078	0.272
		5	0.060	0.064	1.6287	0.443
		6	-0.022	-0.018	1.6848	0.640
		7	-0.053	-0.066	2.0168	0.733
		8	0.016	0.002	2.0462	0.843
		9	-0.053	-0.051	2.3820	0.881
		10	0.063	0.057	2.8562	0.898
		11	-0.112	-0.093	4.3947	0.820
		12	0.028	0.018	4.4910	0.876
		13	0.005	0.014	4.4937	0.922
		14	-0.094	-0.092	5.6098	0.898
		15	0.024	0.010	5.6811	0.931
		16	-0.044	-0.028	5.9310	0.949
		17	-0.035	-0.037	6.0946	0.964
		18	0.118	0.114	7.9251	0.927
		19	0.025	0.068	8.0081	0.949
		20	0.010	-0.006	8.0216	0.968
		21	-0.126	-0.135	10.202	0.925
		22	0.098	0.063	11.531	0.905
		23	0.026	0.024	11.627	0.928
		24	-0.038	-0.029	11.836	0.944
		25	-0.042	-0.052	12.083	0.956

coeficientes θ_1 y θ_T^* puede ser obtenida sustituyendo los coeficientes $\rho_j(Z)$ por sus estimadores en las ecuaciones que se obtienen de su función de autocorrelaciones $\hat{\rho}_j(Z)=\dfrac{-\hat{\theta}_1}{1+\hat{\theta}_1^2}$ y $\hat{\rho}_T(Z)=\dfrac{-\hat{\theta}_T^*}{1+\hat{\theta}_T^{*2}}$.

De la misma forma pueden ser obtenidos los estimadores preliminares de los coeficientes de las representaciones que figuran en la sección precedente. Estos estimadores preliminares constituyen los valores iniciales para el cálculo de los estimadores máximo-verosímiles.

Ejemplo 3.8:

Sea un proceso $\{Y_t\}$ de parámetro trimestral, cuya trayectoria para el período [1er.Trim-1992, 4to.Trim-2006] figura en el **GRAFICO 3.18.**

A partir de la información proporcionada por la serie cronológica y aplicando cuadrados mínimos ordinarios a la representación general de la forma:

$$\nabla^d Y_t = a_0 + a_1 t + \left(\phi_1^* - 1\right)\nabla^{d-1} Y_{t-1} + \sum_{j=1}^{p^*} \phi_{j+1}^* \nabla^d Y_{t-j} + \varepsilon_t$$

se obtuvo que, para $d \geq 1$, los procesos $\{\nabla^d Y_t\}: I(0)$ y que, según resulta del modelo:

$$\nabla Y_t = \underset{(0.033)}{-0.031} Y_{t-1} - \underset{(0.099)}{0.645} \nabla Y_{t-1} - \underset{(0.115)}{0.533} \nabla Y_{t-2} - \underset{(0.126)}{0.096} \nabla Y_{t-3} + \underset{(0.117)}{0.178} \nabla Y_{t-4} + \underset{(0.098)}{0.249} \nabla Y_{t-5}$$

el proceso $\{Y_t\}$ presenta una raíz unitaria (obsérvese que el estadístico T correspondiente al coeficiente que acompaña a Y_{t-1} es -0.943 y que el valor crítico que proporciona la tabla de Dickey-Fuller, para $\alpha = 0.05$, es -1.943).

El comportamiento periódico del correlograma correspondiente al proceso $\{\nabla Y_t\}$ (ver **GRAFICO 3.19**) -cuya trayectoria figura en el **GRAFICO 3.21** - parece sugerir que este proceso está afectado por un factor cíclico de período $T = 4$ y que, por lo tanto, debe considerarse el proceso $\{\nabla^4 \nabla Y_t\} = \{Z_t\}$. Del correlograma correspondiente a este proceso (ver **GRAFICO 3.20**) se puede concluir que admite una representación de la forma:

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_4^* \varepsilon_{t-4} + \theta_1 \theta_4^* \varepsilon_{t-5}$$

(donde $\{\varepsilon_t\}$:$WN - N(0, \sigma_\varepsilon^2)$) obteniéndose el siguiente modelo:

$$Z_t = \varepsilon_t - \underset{(0.075)}{0.667}\varepsilon_{t-1} - \underset{(0.060)}{0.846}\varepsilon_{t-4} + \underset{(0.090)}{0.475}\varepsilon_{t-5}$$

En el **CUADRO 3.6** figura el correlograma correspondiente a la serie de los residuos $\{\hat{\varepsilon}_t\}$.

CAPITULO 4

Acerca de la econometría del mercado financiero

"Mirar el río hecho de tiempo y agua
y recordar que el tiempo es otro río,
saber que nos perdemos como el río
y que los rostros pasan como el agua".

(Borges, J.L.: "El hacedor")

4.1.- Los procesos "random walk"

Sea un proceso $\{Y_t\}$ que admite una representación $AR(1)$ de la forma:

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t$$

donde: **i)** Y_t denota una variable centrada; **ii)** $\{\varepsilon_t\}$:WN y **iii)** $\varphi_1 > 1$ (es decir, que, de acuerdo a lo comentado en la Sec. 3.2.1, $\{Y_t\}$ es un proceso explosivo). Entonces, suponiendo que el proceso tiene su estado inicial en $t = 1$, se puede escribir:

$$Y_2 = \varphi_1 Y_1 + \varepsilon_2$$

$$Y_3 = \varphi_1 Y_2 + \varepsilon_3 = \varphi_1\left(\varphi_1 Y_1 + \varepsilon_2\right) + \varepsilon_3 = \varphi_1^2 Y_1 + \left(\varepsilon_3 + \varphi_1 \varepsilon_2\right)$$

$$Y_4 = \varphi_1 Y_3 + \varepsilon_4 = \varphi_1\left(\varphi_1^2 Y_1 + \varepsilon_3 + \varphi_1 \varepsilon_2\right) + \varepsilon_4 = \varphi_1^3 Y_1 + \left(\varepsilon_4 + \varphi_1 \varepsilon_3 + \varphi_1^2 \varepsilon_2\right)$$

.....

Donde, como se vio en la Sec. 2.15, el término $\varphi^{t-1} Y_1$, que puede ser reinterpretado como la esperanza matemática de Y_t calculada en el momento $t = 1$, es una función creciente de t. De la misma forma, la esperanza matemática condicionada de Y_t, calculada en el momento $t = 2$ depende de ε_2 y, en general, las esperanzas matemáticas condicionadas de Y_t, calculadas en los momentos $t = 3,4,5,...,t-1$, dependen de $\varepsilon_2, \varepsilon_3, ..., \varepsilon_{t-2}$. Ahora bien, estos valores esperados condicionados definen la tendencia de Y_t, de modo que se puede concluir que esta tendencia varía en forma aleatoria (**tendencia estocástica**)[1].

La varianza de Y_t está dada por:

---

[1]. Ver Sec. 3.2.1.

$$\sigma^2(Y_t) = \sum_{j=0}^{t-2} \varphi_1^{2j} \sigma_\varepsilon^2 = \frac{\varphi_1^{2(t-1)} - 1}{\varphi_1^2 - 1} \sigma_\varepsilon^2$$

Se puede concluir, entonces, que, en este caso, el proceso $\{Y_t\}$ está afectado por tendencias estocásticas tanto en el valor medio como en la varianza.

De la misma forma, si en la representación $AR(1)$ anterior, se verifica que $\varphi_1 = 1$:

$$Y_t = Y_{t-1} + \varepsilon_t = Y_1 + \sum_{j=0}^{t-2} \varepsilon_{t-j}$$

entonces se dice que $\{Y_t\}$ es un proceso **"random walk" (RW)** (conocido también como **RW-puro** o **RW de segundo orden** según la nomenclatura de Spanos, A. (1999))[2] en el que los "shocks" aleatorios ε_t ejercen un efecto permanente (no-decreciente) sobre $\{Y_t\}$ (el efecto multiplicador de ε_t sobre Y_t, por ejemplo, $\frac{\partial Y_t}{\partial \varepsilon_t}$, es el mismo que el efecto de ε_t sobre todo Y_{t+j}), de modo que, si bien el proceso RW no presenta una tendencia evidente, la acumulación de los ε_t genera variaciones en el valor esperado condicionado de $\{Y_t\}$ que dan origen a la presencia de tendencia estocástica.

De la expresión anterior resulta que $E(Y_t / Y_{t-1}) = Y_{t-1}$; $E(Y_t / Y_1) = Y_1$ y $\sigma^2(Y_t / Y_1) = (t-1)\sigma_\varepsilon^2$ (al proceso $\{Y_t^*\} = \left\{ \frac{1}{\sigma_\varepsilon \sqrt{t-1}} Y_t \right\}$ se lo conoce como proceso **RW-reescalado**). Obsérvese que, cuando t aumenta indefinidamente, se verifica que $\sigma^2(Y_t / Y_1) \to \infty$, por lo que se puede concluir que, cuando $t \to \infty$ el tiempo esperado de recurrencia de $\{Y_t\}$ al valor esperado es infinito.

Se verifica también que:

---

[2]. El RW tiene su origen en un proceso discreto-discreto, conocido como **RW-simple** o **RW de Bernoulli** formado por la agregación de variables aleatorias Z_t iid (que sólo pueden asumir los valores -1, 0 ó 1), cada una de las cuales representa una "transición" de modo que, al cabo de t transiciones, su estado será $Y_t = Y_1 + Z_2 + Z_3 + ... + Z_t = Y_{t-1} + Z_t \ (t = 1,2,...)$. Supóngase, por comodidad, que $Y_1 = 0$ y que las probabilidades de transición simple sean: $p(Z_t = +1) = p$, $p(Z_t = -1) = q$ y $p(Z_t = 0) = 1 - p - q \ (t = 1,2....)$. La función generatriz de momentos será, entonces, de la forma $M_Z(w) = qe^{-w} + (1 - p - q) + pe^{w}$ y sus momentos: $E(Z) = p - q$ y $\sigma^2(Z) = p + q - (p - q)^2$. Luego, la probabilidad de que, al cabo de t transiciones, el proceso se encuentre en el estado Y_t será $p(Y_t = k) = p(Z_1 + Z_2 + ... + Z_t = k) =$ $= \sum \frac{k!}{k_1! k_2! k_3!} p^{k_1} (1 - p - q)^{k_2} q^{k_3}$ $(k_1 + k_2 + k_3 = y; k_1 - k_2 = k)$. Se verificará, entonces, que $M_{Y_t}(w) = \left[M_Z(w) \right]^{t-1}$, $E(Y_t) = \ = (t-1)(p-q) = (t-1)E(Z)$ y $\sigma^2(Y_t) = (t-1)\left[p + (p-q)^2 \right] = (t-1)\sigma^2(Z)$. De estas expresiones se puede concluir que los dos primeros momentos de $\{Y_t\}$ pueden ser descompuestos en: **i)** un factor id (correspondiente al momento respectivo del proceso $\{Z_t\}$:iid y **ii)** un factor heterogéneo que es función de t. Por esta razón se dice que $\{Y_t\}$:*RW* es un **proceso de heterogeneidad separable de segundo orden**. En particular, si $p = q$, queda definido un **RW-simétrico**.

$$\gamma\left(Y_{t-j},Y_t\right) = E\left(Y_{t-j}Y_t\right) = (t-j-1)\sigma_\varepsilon^2 \qquad (j>0)$$

y, en consecuencia, que:

$$\rho\left(Y_{t-j},Y_t\right) = \frac{\gamma\left(Y_{t-j},Y_t\right)}{\sigma\left(Y_{t-j}\right)\sigma\left(Y_t\right)} = \frac{(t-j-1)\sigma_\varepsilon^2}{\sqrt{\left[(t-j-1)\sigma_\varepsilon^2\right]\left[(t-1)\sigma_\varepsilon^2\right]}} = \sqrt{1-\frac{j}{t-1}} \qquad (j\neq 0)$$

Si en esta representación $AR(1)$ de $\{Y_t\}$ se verifica que $\{\varepsilon_t\}:WN-estricto$ (es decir, es un proceso formado por variables aleatorias iid con varianza finita), entonces se dice que $\{Y_t\}$ es un proceso **"random walk" estricto (RW-estricto)** o **"random walk"-1 (RW-1)**[3] y si, en particular, se verifica que $\{\varepsilon_t\}:WN-N\left(0,\sigma_\varepsilon^2\right)$, entonces se dice que $\{Y_t\}$ es un caso especial de RW-1 conocido como **"random walk" Normal (RW-Normal)**[4]. El cumplimiento de esta condición de Normalidad implica que $\{Y_t\}$ puede ser considerado un proceso **Browniano aritmético** aproximado a partir de observaciones obtenidas a intervalos equiespaciados[5]. Si se verifica que $\{\varepsilon_t\}$ es un proceso formado por variables independientes pero no idénticamente distribuidas (es decir que, por ejemplo, pueden ser formado por variables que pueden presentar heterocedasticidad no-condicionada)[6] con varianza finita, se dice que $\{Y_t\}$ es un proceso **"random walk"-2 (RW-2)**[7].

De las definiciones anteriores se puede concluir que la diferencia entre los RW radica en los distintas formas de dependencia entre las variables ε_t. Dadas dos funciones arbitrarias de las variables ε_t y ε_{t+j}, $f(\varepsilon_t)$ y $g\left(\varepsilon_{t+j}\right)$, si se verifica que $\gamma\left[f(\varepsilon_t),g\left(\varepsilon_{t+j}\right)\right]=0\,(t=1,2,...;j\neq 0)$, entonces $\varepsilon_t \perp \varepsilon_{t+j}$ y, por lo tanto, $\{Y_t\}:RW-1$ o $RW-2$. Si la condición necesaria para que se verifique la igualdad anterior es que las funciones $f(\bullet)$ y $g(\bullet)$ sean lineales, entonces sólo se puede asegurar que las variables ε_t son

---

[3]. Según la nomenclatura de Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. (1997).

[4]. Según la nomenclatura de Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. (1997).

[5]. Debe tenerse en cuenta que, si $\{Y_t\}:RW-Normal$, existe una probabilidad no-nula de que $Y_t<0$. Si Y_t representa el precio de un activo al momento t, esta condición implica la violación de la propiedad de **riesgo limitado**. A fin de resolver este inconveniente se suele considerar el proceso $\{Y_t^*\}=\{\ln(Y_t)\}$, que admite una representación de la forma $Y_t^*=Y_{t-1}^*+\varepsilon_t$, donde $\{\varepsilon_t\}:WN-N\left(0,\sigma_\varepsilon^2\right)$ e $\{Y_t\}$ puede ser condsiderado un **movimiento Browniano geométrico**, aproximado a partir de observaciones obtenidas a intervalos equiespaciados (el movimiento Browniano será tratado en la Sec. 4.5).

[6]. Debe tenerse en cuenta que la estructura probabilística de las variables $Z_{(t)}=Z_1+Z_2+...+Z_t$:iid de alguna forma determina la estructura probabilística del RW y, además, que la estructura de los procesos Markovianos ($AR(1)$) no depende de la linealidad de la transformación de las variables Z_t, sino de su recursividad. Es decir que el proceso definido, en general, por la expresión $Y_t=g(Y_{t-1})+Z_t\ (t=1,2,\ldots,n)$ también es un proceso Markoviano..

[7]. Según la nomenclatura de Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. (1997).

linealmente incorrelacionadas con varianza finita (es decir, que $\{\varepsilon_t\}$:*WN - debil*) y, por lo tanto, que $\{Y_t\}$ es un proceso **"random walk" débil (RW-débil)** o **"random walk"-3 (RW-3)**[8]. Luego, se puede concluir que todo proceso RW-1 ó RW-2 es **de incrementos independientes** y todo proceso RW-3 es de **incrementos incorrelacionados**[9].

Ejemplo 4.1:

Sea un proceso $\{Y_t\}$ de parámetro mensual para cuyo análisis se cuenta con una serie cronológica para el período [ENE-2000,DIC-2004], representada en el **GRAFICO 4.1**.

A partir de los datos proporcionados por dicha serie se obtuvo un modelo RW de la forma:

$$Y_t = \underset{(0.0325)}{0.999}\, Y_{t-1} + \hat{\varepsilon}_t$$

Los residuos (cuya representación figura en el **GRAFICO 4.2**) presentan -confirmando lo demostrado en la página precedente- una heterocedasticidad de la forma:

$$\hat{\varepsilon}_t^2 = \underset{(0.0000381)}{0.000247}(t-1)$$

(ver **GRAFICO 4.3**). Por otra parte, los estimadores de las autocorrelaciones del proceso $\{Y_t\}$:

j	$\hat{\rho}$ $(Y_{t-j};Y_t)$	j	$\hat{\rho}$ $(Y_{t-j};Y_t)$
1	0,958	8	0,656
2	0,914	9	0,600
3	0,874	10	0,539
4	0,833	11	0,476
5	0,789	12	0,411
6	0,743	13	0,339
7	0,702	14	0,261
		15	0,191

observan un comportamiento que admite una representación lineal decreciente de la forma:

$$\hat{\rho}\left(Y_{t-j}, Y_t\right) = \underset{(0.016026)}{1.05} - \underset{(0.001763)}{0.054}\, j$$

(ver **GRAFICO 4.4**).

---

[8]. Según la nomenclatura de Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. (1997).

[9]. Como se verá en la próxima sección, si la condición para que se verifique la nulidad de la covarianza es que la restricción de linealidad involucre solamente a la función $g(\bullet)$, entonces se dice que $\{Y_t\}$ es una martingala. Estas condiciones sobre las autocovarianzas son particularmente importante en el ámbito de las aplicaciones al mercado de capitales, ya que casi todos los modelos de equilibrio de valuación de activos se reducen a la definición de conjuntos de condiciones de ortogonalidad.

GRAFICO 4.1: Trayectoria del proceso $\{Y_t\}$

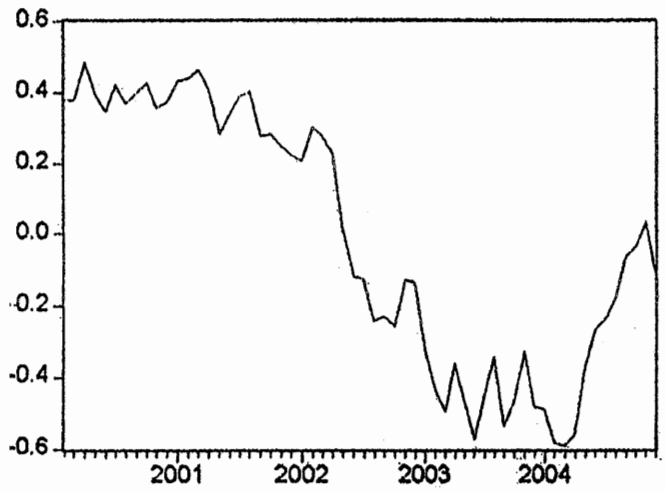


GRAFICO 4.2: Comportamiento de los residuos $\{\hat{\varepsilon}_t\}$

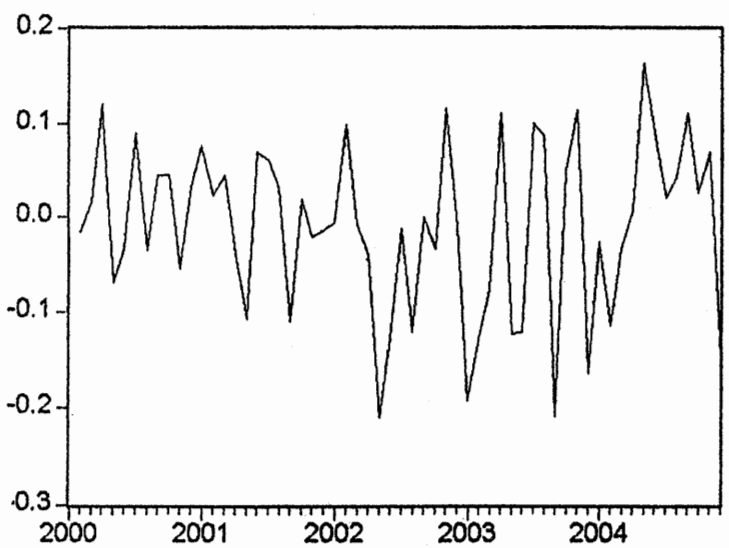


GRAFICO 4.3: Comportamiento de la serie $\{\hat{\varepsilon}_t^2\}$

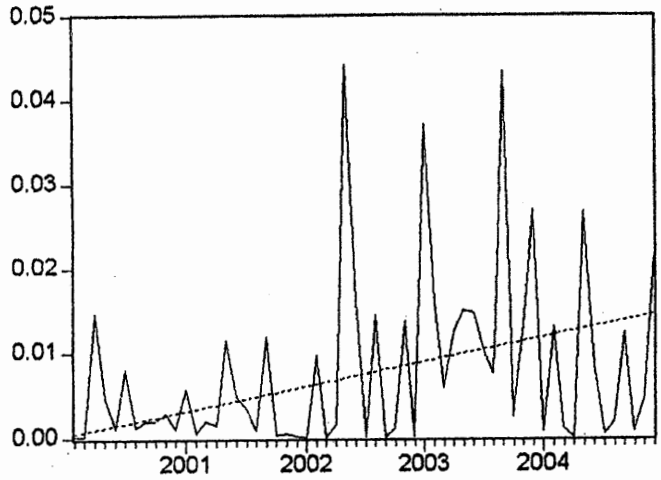
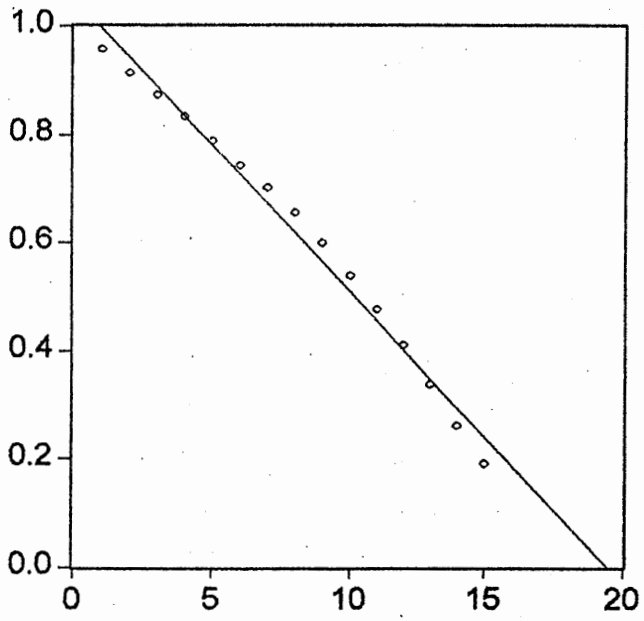


GRAFICO 4.4: Comportamiento del estimador de la función de autocorrelaciones del proceso $\{Y_t\}$



Si a la representación RW se le agrega un término independiente φ_0:

$$Y_t = \varphi_0 + Y_{t-1} + \varepsilon_t$$

queda definido un proceso **"random walk" con tendencia** en el cual $\{\varepsilon_t\}$ está formado, como en el caso anterior, por variables iid o variables independientes o variables incorrelacionadas. Suponiendo, por comodidad, que el proceso asume su estado inicial en $t = 1$, se puede escribir:

$$Y_2 = \varphi_0 + Y_1 + \varepsilon_2$$

$$Y_3 = \varphi_0 + Y_2 + \varepsilon_3 = \varphi_0 + (\varphi_0 + Y_1 + \varepsilon_2) + \varepsilon_3 = Y_1 + 2\varphi_0 + (\varepsilon_2 + \varepsilon_3)$$

.....

$$Y_t = Y_1 + (t-1)\varphi_0 + \sum_{j=0}^{t-2} \varepsilon_{t-j}$$

Expresión de la que resulta que:

$$E(Y_t / Y_1) = Y_1 + (t-1)\varphi_0$$

$$\sigma^2(Y_t / Y_1) = (t-1)\sigma_\varepsilon^2$$

$$\gamma(Y_{t-j}, Y_t / Y_1) = (t-j-1)\sigma_\varepsilon^2$$

y, en consecuencia, que:

$$\rho(Y_{t-j}, Y_t / Y_1) = \frac{\gamma(Y_{t-j}, Y_t / Y_1)}{\sigma(Y_{t-j} / Y_1)\sigma(Y_t / Y_1)} = \sqrt{1 - \frac{j}{t-1}}$$

Luego, se puede concluir que este proceso, que es no-estacionario en el valor medio, presenta una tendencia determinística (creciente o decreciente según que φ_0 sea mayor o menor que cero; es decir que $\{Y_t\}$ tenderá eventualmente a $+\infty$ ó $-\infty$, respectivamente) y una tendencia estocástica generada por la acumulación de los "shocks" aleatorios ε_t.

Debe tenerse en cuenta que: **i)** como se mencionó en la Sec. 3.2, si en este tipo de procesos se calculan los residuos con respecto a la función:

$$f(t) = \varphi_0 + \varphi_1 t$$

se obtiene un proceso sin tendencia determinística, pero no-estacionario por la presencia de tendencia estocástica y **ii)** si se calculan las diferencias primeras se obtiene unproceso con incrementos independientes o incorrelacionados:

$$\nabla Y_t - \nabla Y_{t-1} = (\varphi_0 + \varepsilon_t) - (\varphi_0 + \varepsilon_{t-1}) = \varepsilon_t - \varepsilon_{t-1}$$

y **iii)** si se calculan las diferencias segundas, se obtiene un proceso $MA(1)$ no-invertible con una autocorrelación de primer orden igual a -0.5.

Una expresión más general del RW, conocida como **RW con ruido**, se obtiene a partir de una representación de la forma:

$$Y_t = Z_t + \varepsilon_t^*$$

donde: **i)** $\{\varepsilon_t^*\}$:*WN* y **ii)** $\{Z_t\}$ es un proceso formado por variables centradas, que admite una representación RW de la forma $Z_t = Z_{t-1} + \varepsilon_t$, siendo $\{\varepsilon_t\}$:*WN* y tal que $\{\varepsilon_t\} \perp \{\varepsilon_t^*\}$ (es decir, se verifica que $E(\varepsilon_{t-j}^* \varepsilon_t) = 0$ para $j = 0,\pm 1,\pm 2,\ldots$).

Suponiendo que Z_1 sea conocido, como ya se vio, se puede escribir

$$Z_t = Z_1 + \sum_{j=0}^{t-2} \varepsilon_{t-j}$$

de modo que la representación de $\{Y_t\}$ puede ser expresada de la siguiente forma:

$$Y_t = Z_1 + \sum_{j=0}^{t-2} \varepsilon_{t-j} + \varepsilon_t^* = Y_1 - \varepsilon_1^* + \sum_{j=0}^{t-2} \varepsilon_{t-j} + \varepsilon_t^*$$

lo que permite concluir que: **i)** dado que los "shocks" ε_t ejercen una influencia permanente sobre $\{Y_t\}$, este proceso está afectado por una tendencia estocástica; **ii)** el proceso $\{\varepsilon_t^*\}$ ejerce sólo una influencia temporaria sobre $\{Y_t\}$ (ε_t^* influye sobre el comportamiento de Y_t pero no afecta a Y_{t+j} $(j = 1,2,\ldots)$).

De la expresión anterior se obtiene la esperanza matemática condicionada:

$$E(Y_t / Y_1) = Y_1 - \varepsilon_1^*$$

(formada por una tendencia estocástica y una componente irregular) y la función de autocovarianzas:

$$\gamma(Y_{t-j}, Y_t) = E\left[\left(Y_{t-j} - Y_1 + \varepsilon_1^*\right)\left(Y_t - Y_1 + \varepsilon_1^*\right)\right] =$$

$$= E\left[\left(\sum_{i=0}^{t-j-2} \varepsilon_{t-j-i} + \varepsilon_{t-j}^*\right)\left(\sum_{i=0}^{t-2} \varepsilon_{t-i} + \varepsilon_t^*\right)\right] =$$

$$= E\left(\sum_{i=0}^{t-j-2} \varepsilon_{t-j-i}^2 + \sum\sum_{i \neq h} \varepsilon_{t-i}\varepsilon_{t-h} + \varepsilon_t^* \sum_{i=0}^{t-j-2} \varepsilon_{t-j-i} + \varepsilon_t^* \sum_{i=0}^{t-2} \varepsilon_{t-i}\right) =$$

$$= (t - j - 1)\sigma_\varepsilon^2 \qquad (j = 0,1,2,\ldots)$$

En particular, será:

$$\gamma\left(Y_{t-j},Y_{t-j}\right) = \sigma^2\left(Y_{t-j}\right) = (t-j-1)\sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2$$

$$\gamma\left(Y_t,Y_t\right) = \sigma^2\left(Y_t\right) = (t-1)\sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2$$

De modo que la función de autocorrelaciones será de la forma:

$$\rho\left(Y_{t-j},Y_t\right) = \frac{\gamma\left(Y_{t-j},Y_t\right)}{\sigma\left(Y_{t-j}\right)\sigma\left(Y_t\right)} = \frac{(t-j-1)\sigma_\varepsilon^2}{\sqrt{\left[(t-j-1)\sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2\right]\left[(t-1)\sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2\right]}} =$$

$$= \frac{(t-j-1)\sigma_\varepsilon^2}{\sqrt{(t-j-1)(t-1)\sigma_\varepsilon^4 + (2t-j-2)\sigma_\varepsilon^2\sigma_{\varepsilon^*}^2 + \sigma_{\varepsilon^*}^4}}$$

De los resultados precedentes se puede concluir que: **i)** la presencia del proceso $\left\{\varepsilon_t^*\right\}$ produce una aumento en la varianza de $\left\{Y_t\right\}$, pero no afecta su comportamiento de largo plazo y **ii)** la función de autocorrelaciones de este RW generalizado decrece más rápidamente que la correspondiente al RW puro analizado en las páginas precedentes.

Como una combinación de los RW con tendencia y con ruido, a partir de una representación de la forma:

$$Y_t = Z_t + \varepsilon_t^*$$

donde: **i)** $\left\{\varepsilon_t^*\right\}$:*WN* y **ii)** $\left\{Z_t\right\}$ es un proceso RW con tendencia de la forma:

$$Z_t = \varphi_{Z0} + Z_{t-1} + \varepsilon_t$$

(siendo $\left\{\varepsilon_t\right\}$:*WN* y tal que $\left\{\varepsilon_t\right\} \perp \left\{\varepsilon_t^*\right\}$) se obtiene un **RW con tendencia y con ruido**. Como ya se vio, suponiendo que Z_1 sea conocido, se puede escribir:

$$Z_t = Z_1 + \phi_{Z0}(t-1) + \sum_{j=0}^{t-2}\varepsilon_{t-j}$$

De esta expresión se puede concluir que $\left\{Z_t\right\}$ presenta tendencias determinística y estocástica. Entonces, se puede escribir:

$$Y_t = Z_1 + \varphi_{Z0}(t-1) + \sum_{j=0}^{t-2}\varepsilon_{t-j} + \varepsilon_t^*$$

Suponiendo la condición inicial $Y_1 = Z_1 + \varepsilon_1^*$, resulta la siguiente expresión:

$$Y_t = Y_1 - \varepsilon_1^* + \varphi_{z0}(t-1) + \sum_{J=0}^{T-2}\varepsilon_{t-j} + \varepsilon_t^*$$

Si $\{\varepsilon_t\}$ no es un WN, esta expresión puede ser generalizada de la siguiente forma:

$$Y_t = Y_1 + \varphi_{Z0}(t-1) + \sum_{j=0}^{t-2} \varepsilon_{t-j} + \Theta(B)\varepsilon_t^*$$

Representación conocida como **RW con tendencia general y comportamiento irregular.**

Sean $\{\varepsilon_{Yt}\}, \{\varepsilon_{Zt}\}$ y $\{\varepsilon_{Wt}\}$ tres procesos incorrelacionados entre sí, obtenidos de aplicar, respectivamente, representaciones de la forma:

$$Y_t = Z_t + \varepsilon_{Yt}$$
$$Z_t = Z_{t-1} + W_t + \varepsilon_{Zt}$$
$$W_t = W_{t-1} + \varepsilon_{Wt}$$

que definen un **RW con tendencia lineal local.**

De acuerdo a lo ya expresado, suponiendo que W_1 sea conocido, se puede escribir $W_t =$ $= W_1 + \sum_{j=0}^{t-2} \varepsilon_{W,t-j}$. De modo que resulta que:

$$Z_t = Z_{t-1} + W_1 + \sum_{j=0}^{t-2} \varepsilon_{W,t-j} + \varepsilon_{Zt} =$$

$$= Z_1 + \sum_{j=0}^{t-2} \varepsilon_{Z,t-j} + (t-1)(W_1 + \varepsilon_{W2}) + (t-2)\varepsilon_{W3} + (t-3)\varepsilon_{W4} + ... + \varepsilon_{Wt}$$

Luego, suponiendo la condición inicial $Y_1 = Z_1 + \varepsilon_{Y1}$, resulta la representación:

$$Y_t = Y_1 + (\varepsilon_{Yt} - \varepsilon_{Y1}) + \sum_{j=0}^{t-2} \varepsilon_{Z,t-j} + (t-1)(W_1 + \varepsilon_{W2}) + (t-2)\varepsilon_{W3} + ... + \varepsilon_{Wt}$$

o, en términos más generales:

$$Y_t = Y_1 + \sum_{j=0}^{t-2} \varepsilon_{Z,t-j} + (t-1)(W_1 + \varepsilon_{W2}) + (t-2)\varepsilon_{W3} + ... + \varepsilon_{Wt} + \Theta(B)\varepsilon_{Yt}$$

Se puede concluir, entonces, que el proceso $\{Y_t\}$ presenta: **i)** una tendencia determinística que depende de las realizaciones contemporánea y anteriores del proceso $\{\varepsilon_{Wt}\}$ (obsérvese que, para un período t dado, el coeficiente que representa la tendencia determinística está dado por $W_1 + \varepsilon_{W1} + ... + \varepsilon_{Wt}$ y puede ser positivo o negativo); **ii)** una tendencia estocástica definida por $\sum_{j=0}^{t-2} \varepsilon_{Z,t-j}$ y **iii)** una componente irregular definida por $\Theta(B)\varepsilon_{Yt}$.

Ejemplo 4.2:

Sea un proceso $\{Y_t\}$ de parámetro mensual para cuyo análisis se cuenta con una serie cronológica para el período [ENE-2000,DIC-2004], representada en el **GRAFICO 4.5**.

A partir de los datos proporcionados por dicha serie se obtuvo un modelo de la forma:

$$Y_t = \underset{(0.7061)}{1.6146} + \underset{(0.0177)}{0.999}\, Y_{t-1} + \hat{\varepsilon}_t$$

Obsérvese que, en este caso en el que $\{Y_t\}$ admite una representación RW-con tendencia, la esperanza matemática condicionada por un valor inicial $Y_1 = 0$ observa un comportamiento lineal de la forma:

$$E(Y_t / Y_1) = 1.6146(t-1)$$

(ver **GRAFICO 4.6**). Los residuos que resultan del modelo RW (ver **GRAFICO 4.7**) presentan una heterocedasticidad de la forma:

$$\hat{\varepsilon}_t^2 = 0.2026(t-1)$$

(ver **GRAFICO 4.8**). Los estimadores de las autocorrelaciones del proceso $\{Y_t\}$:

j	$\hat{\rho}$ $(Y_{t-j};Y_t)$	j	$\hat{\rho}$ $(Y_{t-j};Y_t)$
1	0,953	8	0,662
2	0,912	9	0,616
3	0,871	10	0,569
4	0,828	11	0,523
5	0,786	12	0,479
6	0,745	13	0,424
7	0,700	14	0,370
		15	0,310

observan un comportamiento que admite una representación lineal decreciente de la forma:

$$\hat{\rho}(Y_{t-j}, Y_t) = 1.010 - 0.045\, j$$

(ver **GRAFICO 4.9**).

4.2.- Los tests para contrastar las hipótesis de RW

Dado que la hipótesis de RW-1 implica la hipótesis de que $\{\varepsilon_t\}$:*WN – estricto* , sin definir una

GRAFICO 4.5: Trayectoria del proceso $\{Y_t\}$

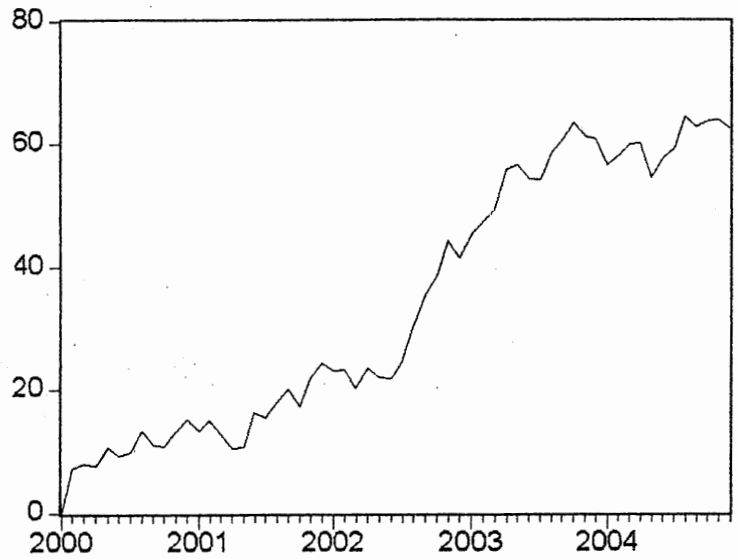


GRAFICO 4.6: Representación del comportamiento lineal de los valores esperado del proceso condicionado $\{Y_t/Y_1=0\}$

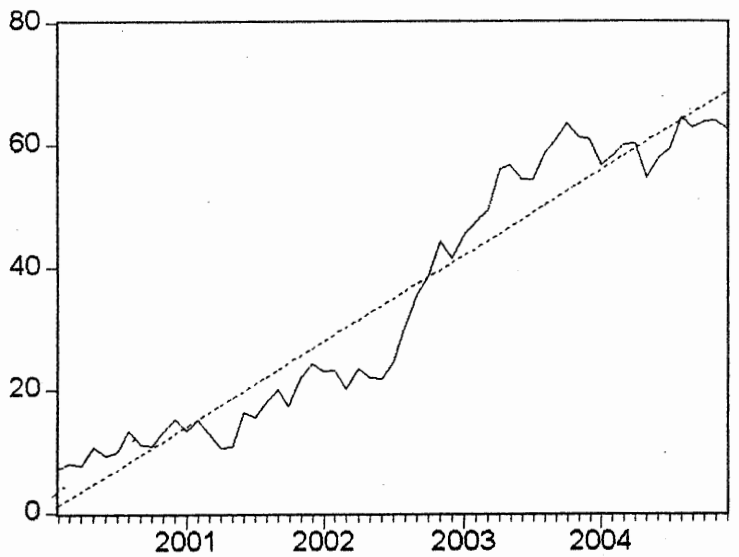


GRAFICO 4.7: Comportamiento de los residuos $\{\hat{\varepsilon}_t\}$ RW-con rendencia

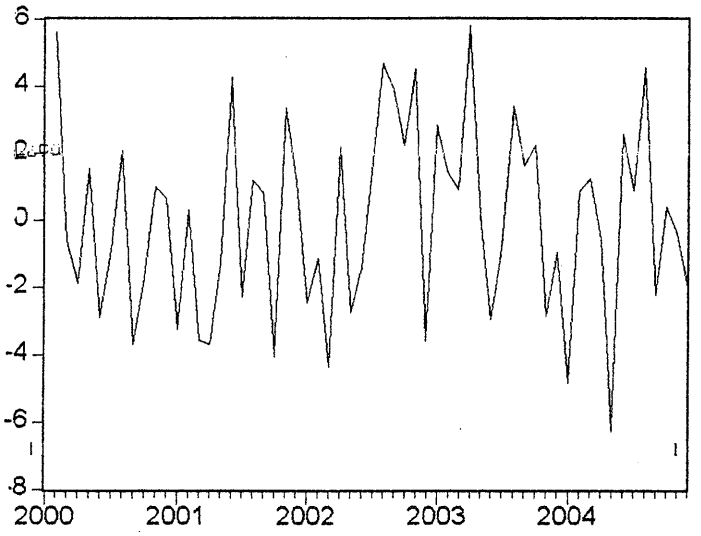


GRAFICO 4.8: Comportamiento de la serie $\{\hat{\varepsilon}_t^2\}$

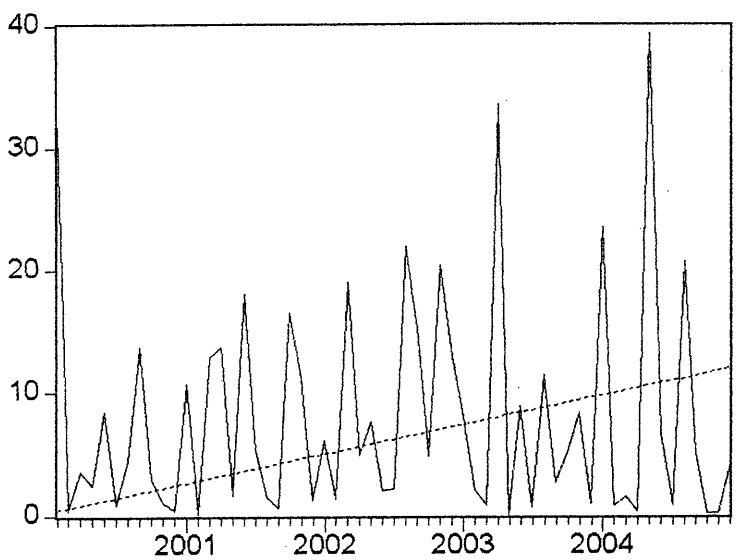
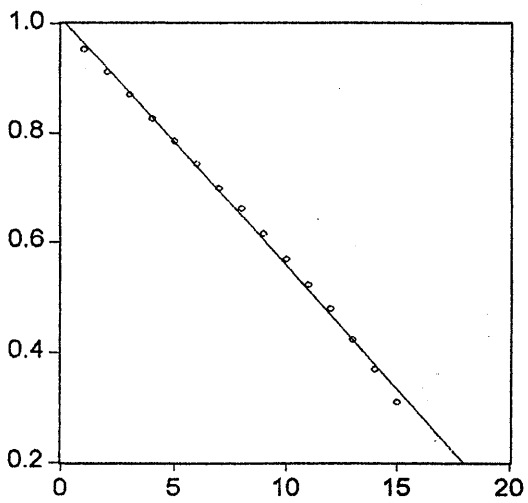


GRAFICO 4.9: Comportamiento del estimador de la función de autocorrelaciones del proceso $\{Y_t\}$



familia específica de distribuciones, los tests desarrollados para decidir si un proceso $\{Y_t\}$ puede ser considerado o no un RW-1 son, en principio, no-paramétricos, robustos respecto a distintas familias de distribuciones e invariantes con respecto a variaciones en las unidades de medida[10]. Otras técnicas basadas en la distribución empírica de los datos -las cuales requieren supuestos más estrictos sobre las distribuciones conjunta y marginal del proceso generador- dieron origen, a su vez, a una serie de tests semi-paramétricos. Una tercera alternativa dio origen a un conjunto de tests paramétricos los cuales, si bien -obviamente-son menos generales, resultan más simples y, a menudo, más potentes que los no-paramétricos. Pero debe tenerse en cuenta que, como contrapartida, adolecen de una debilidad originada en que pequeños desvíos con respecto a la familia de distribuciones seleccionada como hipótesis pueden conducir a diferencias considerables entre las medidas observadas a partir de muestras finitas y las supuestas "a priori".

Testear la hipótesis que $\{Y_t\}:RW-2$, es decir, que el proceso $\{\varepsilon_t\}:WN-estricto$ sin establecer restricciones acerca de la forma de las distribuciones marginales de sus variables, es virtualmente imposible a partir de los métodos de la inferencia clásica. Habitualmente se aplican metodologías empíricas como la de los **filtros**[11] o el **análisis técnico**[12].

Supóngase un proceso $\{Y_t\}:AR(1)$ que admite una representación de la forma:

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}:WN-debil$). Una solución al problema de decidir si $\{Y_t\}$ puede ser considerado o no un RW-3, podría lograrse utilizando los tests de raíz unitaria tratados en el

---

10. Como ocurre, por ejemplo, con el test de correlación de rangos de Spearman o el test τ de correlación de Kendall (ver Randles, R.; Wolfe, D. (1979), Serfling, R. (1980)).

11. Ver Alexander, S. (1961), Fama, E. (1965), Fama, E.; Blume, M. (1966).

12. Ver Edwards, R.; Magee, J. (1966), , Treynor, J.; Ferguson, R. (1985), Murphy, J. (1986), Brown, D.; Jennings, R. (1989), Neftci, S. (1991), Pau, L. (1991), Taylor, M.; Allen, H. (1992), Brock, W.; Lakonishock, J.; Le Baron, B. (1992), Blume, L.; Easley, D.; O'Hara, M. (1994), Le Baron, B. (1996).

capítulo precedente, que contrastan la hipótesis nula $H_0: Y_t = Y_{t-1} + \varepsilon_t$ con respecto a la alternativa $H_1: Y_t = \varphi_1 Y_{t-1} + \varepsilon_t \ (|\varphi_1| < 1)$, es decir, contrastan la hipótesis nula $H_0: \{Y_t\}$ *presenta tendencia estocastica* con respecto a la alternativa $H_1: \{Y_t\}: I(0)$. Pero, debe tenerse en cuenta que en estos tests la hipótesis nula incluye alternativas que no son RW (por ejemplo, representaciones ARFIMA)[13].

Entre los tests para contrastar la hipótesis nula $H_0: \{Y_t\}: RW-1$, cabe mencionar las siguientes propuestas alternativas:

1) El test de Cowles-Jones

Este test considera un proceso $\{Y_t\}: RW - geometrico$ tal que $\{Y_t^*\} = \{\ln(Y_t)\}$ admite una representación RW-1 de la forma[14]:

$$Y_t^* = Y_{t-1}^* + \varepsilon_t$$

(donde Y_t denota una variable centrada y $\{\varepsilon_t\}: WN - binomial$) y un proceso $\{I_t\}$ formado por variables binomiales definidas como[15]:

$$I_t = \begin{cases} 1 & si \quad \varepsilon_t = Y_t^* - Y_{t-1}^* > 0 \\ 0 & si \quad \varepsilon_t = Y_t^* - Y_{t-1}^* \leq 0 \end{cases}$$

con distribución de probabilidades $b(1, p_s)$, donde $p_s = p[(\varepsilon_t < 0)/(\varepsilon_{t-1} > 0)]$ (si se supone, además, que las variables ε_t se distribuyen en forma simétrica, entonces las variables I_t se distribuyen de acuerdo con una función $b\left(1, \frac{1}{2}\right)$).

Definiendo como una **secuencia** a cada par de perturbaciones sucesivas que tengan el mismo signo y como una **reversión** a cada par de perturbaciones sucesivas con signo distinto, dada una muestra de n perturbaciones $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$, el número de secuencias será:

$$f_n^{(s)} = \sum_{t=1}^{n-1} [I_t I_{t+1} + (1 - I_t)(1 - I_{t+1})]$$

---

[13]. Las representaciones ARFIMA, que consideran que el orden de diferenciación puede asumir cualquier valor real en el intervalo (-1,1), serán tratadas en el Cap. 5.

[14]. Este test fue diseñado para contrastar la hipótesis nula que $\{Y_t\}$ es un movimiento Browniano geométrico (es decir que $\{Y^*(t)\} = \{\ln(Y(t))\}: B(t)$), a partir de su aproximación discreta en el dominio del tiempo (ver Cowles, A.; Jones, H. (1937)).

[15]. El proceso $\{I_t\}$ es una cadena de Markov de dos estados (ver Sec. 1.7.1) con probabilidades de transición $p[(\varepsilon_t < 0)/(\varepsilon_{t-1} > 0)] = p_s$ y $p[(\varepsilon_t > 0)/(\varepsilon_{t-1} < 0)] = 1 - p_s$.

y, por lo tanto, el número de reversiones será $f_n^{(r)} = (n-1) - f_n^{(s)}$. Luego, de acuerdo con los postulados de la versión de De Moivre del teorema central del límite ((ver Landro, A.H. (2002)), la variable $f_n^{(s)}$ converge en-distribución a una función Normal con valor esperado np_s y varianza $np_s(1-p_s)$ (donde $1-p_s = p\left[(\varepsilon_t > 0)/(\varepsilon_{t-1} < 0)\right]$).

A fin de contrastar la hipótesis sobre la simetría de la probabilidad p_s, Cowles-Jones propusieron el siguiente estimador consistente:

$$\hat{CJ} = \frac{f_n^{(s)}}{f_n^{(r)}} = \frac{\hat{p}_s}{1-\hat{p}_s}$$

(donde $\hat{p}_s = \frac{f_n^{(s)}}{n-1}$) del parámetro $\frac{p_s}{1-p_s}$ (en particular, si la hipótesis $p_s = \frac{1}{2}$ es verdadera, el estimador de CJ debería ser aproximadamente igual a uno).

Supóngase, ahora, que el proceso $\{Y_t^*\}$ admita una representación de la forma:

$$Y_t^* = \varphi_0 + Y_{t-1}^* + \varepsilon_t$$

(donde $\{\varepsilon_t\}$:*WN – estricto*). Entonces las variables I_t^* se distribuirán de acuerdo con una función $b(1, p_s^*)$, donde: **i)** si el término independiente de la representación es positivo:

$$p_s^* = p(\varepsilon_t^* > 0) = p\left(\ln\left(\frac{Y_t}{Y_{t-1}}\right) > 0\right) = p(\varphi_0 + \varepsilon_t > 0) = p(\varepsilon_t > -\varphi_0) =$$

$$= p\left(\frac{\varepsilon_t}{\sigma_\varepsilon} > -\frac{\varphi_0}{\sigma_\varepsilon}\right) = p\left(u_t > -\frac{\varphi_0}{\sigma_\varepsilon}\right) =$$

$$= 1 - F_u\left(-\frac{\varphi_0}{\sigma_\varepsilon}\right) = F_u\left(\frac{\varphi_0}{\sigma_\varepsilon}\right)$$

(siendo $F_u(\bullet)$ la función de distribución correspondiente a una variable $N(0;1)$) y **ii)** si el término independiente de la representación es negativo:

$$p_s^* = p(\varepsilon_t^* > 0) = p(-\varphi_0 + \varepsilon_t > 0) = p(\varepsilon_t > \varphi_0) = p\left(u_t > \frac{\varphi_0}{\sigma_\varepsilon}\right) = 1 - F_u\left(\frac{\varphi_0}{\sigma_\varepsilon}\right)$$

Es decir que p_s^* será mayor o menor que 1/2 según que φ_0 sea positivo o negativo y el coeficiente CJ, para un proceso RW-1, será de la forma:

$$CJ = \frac{p_s}{1-p_s} = \frac{(p_s^*)^2 + (1-p_s^*)^2}{2p_s^*(1-p_s^*)} \geq 1$$

Igual que en el caso precedente, el estimador:

$$f_n^{(s)} = \sum_{t=1}^{n-1}\left[I_t^* I_{t+1}^* + \left(1 - I_t^*\right)\left(1 - I_{t+1}^*\right)\right] = \sum_{t=1}^{n-1} Z_t$$

(donde $Z_t : b(1, p_s)$, siendo $p_s = \left(p_s^*\right)^2 + \left(1 - p_s^*\right)^2$) converge en-distribución a una función Normal con valor esperado np_s y, dado que las variables Z_t consecutivas no son independientes[16], con varianza:

$$\sigma^2\left(f_n^{(s)}\right) = np_s\left(1 - p_s\right) + 2n\gamma\left(Z_t, Z_{t+1}\right) = np_s\left(1 - p_s\right) + 2\left[\left(p_s^*\right)^3 + \left(1 - p_s^*\right)^3 - \left(p_s^*\right)^2\right]$$

Tomando, entonces, una aproximación de primer orden del desarrollo de Taylor (método delta)[17] del estimador:

$$C\hat{J} = \frac{f_n^{(s)}}{(n-2) - f_n^{(s)}} = \frac{f_n^{(s)}}{f_n^{(r)}}$$

y, teniendo en cuenta la distribución asintóticamente Normal de $f_n^{(s)}$, se obtiene que:

$$C\hat{J} \xrightarrow{d} N\left\{\frac{p_s}{1 - p_s}, \frac{p_s\left(1 - p_s\right) + 2\left[\left(p_s^*\right)^3 + \left(1 - p_s^*\right)^3 - p_s^2\right]}{(n-2)\left(1 - p_s\right)^4}\right\}$$

De modo que, para n suficientemente grande, queda definido el siguiente estadístico para contrastar la hipótesis nula $H_0 : \{Y_t\} : RW - 1$ en el caso binomial:

$$Z = \frac{C\hat{J} - CJ}{\sigma\left(C\hat{J}\right)} : N(0;1)$$

---

16. El proceso $\{I_t\}$ es una cadena de Markov de dos estados (ver Sec. 1.7.1) con probabilidades de transición $p_{1,1}(t-1,t) = \frac{\left(p_s^*\right)^3 + \left(1 - p_s\right)^3}{p_s}$ y $p_{0,0}(t-1,t) = \frac{1}{2}$.

17. De acuerdo con el teorema de Taylor (método Delta), dado un estimador insesgado de un vector θ tal que, bajo el supuesto que la hipótesis nula $H_0 : \theta = \theta_0$ es verdadera, converje en-distribución a una función Normal de la forma $N\left(0, \sigma_\theta^2\right)$ $\left(\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \sigma_\theta^2\right)\right)$. A partir de una aproximación de primer orden de la serie de Taylor para $f\left(\hat{\theta}\right)$ respecto de θ_0, la variable aleatoria definida como una función no-lineal de dicho estimador ($f\left(\hat{\theta}\right)$) es tal que $\sqrt{n}\left[f\left(\hat{\theta}\right) - f\left(\theta_0\right)\right] \xrightarrow{d} N\left(0, \sigma_f^2\right)$, donde $\sigma_f^2 = \frac{\partial f}{\partial \theta'} \sigma_\theta^2 \frac{\partial f}{\partial \theta}$ (debe tenerse en cuenta que los términos de orden superior del desarrollo de Taylor son de orden asintótico menor que $\sqrt{n}$ (ver Sec. 1.7.2.1), por lo que sólo el término lineal influye en la definición de la distribución asintótica de $f\left(\hat{\theta}\right)$).

donde:

$$\sigma^2\left(C\hat{J}\right) = \frac{\hat{p}_s\left(1-\hat{p}_s\right)+2\left[\left(\hat{p}_s^*\right)^3+\left(1-\hat{p}_s^*\right)^3-\hat{p}_s^2\right]}{(n-2)\left(1-\hat{p}_s\right)^4}$$

$$\hat{p}_s^* = \begin{cases} F_u\left(\dfrac{\hat{\varphi}_0}{\hat{\sigma}_\varepsilon}\right) & si \qquad \hat{\varepsilon}_t^* = \hat{\varphi}_0 + \bar{\varepsilon}_t \ y \ \varphi_0 > 0 \\ 1-F_u\left(\dfrac{\hat{\varphi}_0}{\hat{\sigma}_\varepsilon}\right) & si \qquad \hat{\varepsilon}_t^* = -\hat{\varphi}_0 + \hat{\varepsilon}_t \end{cases}$$

y $\hat{p}_s = \left(p_s^*\right)^2 + \left(1-p_s^*\right)^2$.

2) El test de Mood

Este test, propuesto por Mood, A. (1940), se basa en el análisis del número de **rachas**[18] de valores positivos y negativos de $\varepsilon_t^* = \ln\left(\dfrac{Y_t}{Y_{t-1}}\right)$ en una muestra de tamaño n y en la comparación de su distribución con la distribución esperada en el caso que $\left\{Y_t^*\right\}: RW-1$.

Sean: **i)** una sucesión de n variables aleatorias:

$$I_t = \begin{cases} 1 & si \qquad \varepsilon_t = Y_t^* - Y_{t-1}^* > 0 \\ 0 & si \qquad \varepsilon_t = Y_t^* - Y_{t-1}^* \le 0 \end{cases}$$

iid, binomiales y tales que $p\left(I_t=1\right)=p$ y $p\left(I_t=0\right)=1-p$; **ii)** una variable $n_r(i)\,(i=0,1,2,...)$ que representa el número de rachas del resultado $i\,(=0,1)$ en dicha sucesión y **iii)** una variable $n_r = n_r(0)+n_r(1)$ que representa el número total de rachas en la misma sucesión. El valor esperado de la variable n_r es de la forma $E\left(n_r\right) = 2np\left(1-p\right)+p^2+\left(1-p\right)^2$. De modo que (para $n \ge 1$) presenta un valor máximo (igual a $\dfrac{n+1}{2}$) en $p=\dfrac{1}{2}$. La existencia de tendencia en el RW (de cualquier signo) hará que este número esperado de rachas disminuya ($p = F_u\left(\dfrac{\varphi_0}{\sigma_\varepsilon}\right)$ o $p = 1-F_u\left(\dfrac{\varphi_0}{\sigma_\varepsilon}\right)$, según que φ_0 sea positivo o negativo, respectivamente). Es decir, que el número total esperado de rachas se maximiza si $\left\{Y_t\right\}: RW-simple$ (en términos más generales, dada una variable multinomial I_t que puede asumir los valores $\Omega\left(I_t\right)=\left\{0,1,2,...,k\right\}$ con probabilidades $p\left(I_t\right)=p_i\ (i=0,1,2,...,k)$, los momentos de la variable que representa el número de rachas del resultado $i\,(=0,1,2,...,k)$ en una sucesión de n observaciones

---

18. Se define como racha a la sucesión de realizaciones consecutivas de una variable aleatoria que poseen una característica común.

independientes, $n_r(i)$, están definidos por $E[n_r(i)] = np_i(1-p_i)+p_i^2$; $\sigma^2[n_r(i)] = np_i(1-4p_i+p_i^2 - 3p_i^3)+p_i^2(3-8p_i+5p_i^2)$ y $\gamma[n_r(i),n_r(j)] = -np_ip_j(1-2p_i-2p_j+3p_ip_j)-p_ip_j(2p_i+2p_j-5p_ip_j)$).

El estadístico propuesto en este caso para testear la hipótesis nula $H_0:\{Y_t\}:RW-1$ corregido del error de continuidad[19] es de la forma[20]:

$$Z = \frac{n_r + \frac{1}{2} - 2n\hat{p}(1-\hat{p})}{2\sqrt{np(1-\hat{p})[1-3\hat{p}(1-\hat{p})]}} : N(0;1)$$

(donde $\hat{p} = \frac{n_r(1)}{n}$).

Ejemplo 4.3:

Sea un proceso $\{Y_t\}:RW$ de parámetro mensual para cuyo análisis se cuenta con una serie cronológica para el período [ENE-2000,FEB-2004] (ver **GRAFICO 4.10**).

En el **CUADRO 4.1** figura la serie $\{u_t\}$ obtenida de calcular las primeras diferencias de $\{Y_t\}$ y en el **GRAFICO 4.11** figura la representación de su trayectoria.

En la serie $\{u_t\}$ se contabilizan 33 secuencias y 15 reversiones ($f_{50}^{(s)} = 33$ y $f_{50}^{(r)} = 15$) es decir, se obtiene que:

$$\hat{p}_s = \frac{f_{50}^{(s)}}{50-2} = 0{,}6875$$

y $1-p_s = 0.3125$. De modo que:

---

[19]. Ver Mood, A. (1940), Wallis, W.; Roberts, H. (1956), David, F.; Barton, D. (1962), Landro, A.H. (2002).

[20]. En el caso más general de la variable multinomial, los correspondientes estadísticos son de la forma $Z(i) =$

$= \frac{n_r(i) - np_i(1-p_i) - p_i^2}{\sqrt{n[p_i(1-p_i) - 3p_i^2(1-p_i)^2]}} \xrightarrow{d} N(0;1)\ (i=0,1,2,\ldots,k)$; $\lim_{n\to\infty}\gamma[Z(i),Z(j)] = -p_ip_j(1-2p_i-2p_j+3p_ip_j)$

$(i,j=0,1,2,\ldots,k)$ y $Z = \frac{n_r - n\left(1-\sum_{i=1}^{k}p_i^2\right)}{\sqrt{n\left[\sum_{i=1}^{k}p_i^2(1+2p_i) - 3\left(\sum_{i=1}^{k}p_i^2\right)^2\right]}} \xrightarrow{d} N(0;1)$.

$$\hat{CJ} = \frac{0.6875}{0.3125} = 2.2$$

Dado que $\sigma_u^2 = 288.443$, se obtuvieron las probabilidades

$$p_s^* = F_u\left(\frac{\varphi_0}{\hat{\sigma}_u}\right) = F_u\left(\frac{\varphi_0}{16.98362}\right)$$

a partir de las cuales se obtuvieron, a su vez, los estimadores $\hat{p}_s = \left(p_s^*\right)^2 + \left(1 - p_s^*\right)^2$. En el **CUADRO 4.2** figuran los valores asumidos por estas probabilidades y los valores esperados y las varianzas del estimador $\hat{CJ}$ para distintos valores de φ_0, a partir de los cuales se pueden calcular los valores asumidos por el estadístico $Z = \dfrac{2.2 - CJ}{\sigma\left(\hat{CJ}\right)}$ y contrastar la hipótesis que $\{Y_t\}:RW$ para el caso binomial.

De la observación del **CUADRO 4.2** se puede concluir que existen razones suficientes para rechazar la hipótesis de que $\{Y_t\}:WN$ y que, para $\alpha = 0.05$ existen razones suficientes para rechazar la hipótesis $\varphi_0 < 0.9$.

En lo que hace al test de Mood, en la serie cronológica $\{u_t\}$ que figura en el **CUADRO 1.4** se contabilizaron 8 rachas del resultado 0 y 8 rachas del resultado 1. Dado que $\sigma_u^2 = 288.443$, se obtuvieron las probabilidades:

$$\hat{p} = F_u\left(\frac{\varphi_0}{\hat{\sigma}_u}\right) = F_u\left(\frac{\varphi_0}{16.98362}\right)$$

(ver **CUADRO 4.3**) que permiten obtener los valores esperados y las varianzas de las variables **número total de rachas para distintos valores de** φ_0 y, en consecuencia, los valores del estadístico:

$$Z = \frac{\left(16 + \frac{1}{2}\right) - \hat{m}\left(n_r\right)}{\sigma\left(n_r\right)} : N(0:1)$$

que permite contrastar la hipótesis $H_0: \{Y_t\}: RW$ para el caso binomial $n_r : b(49; p)$.

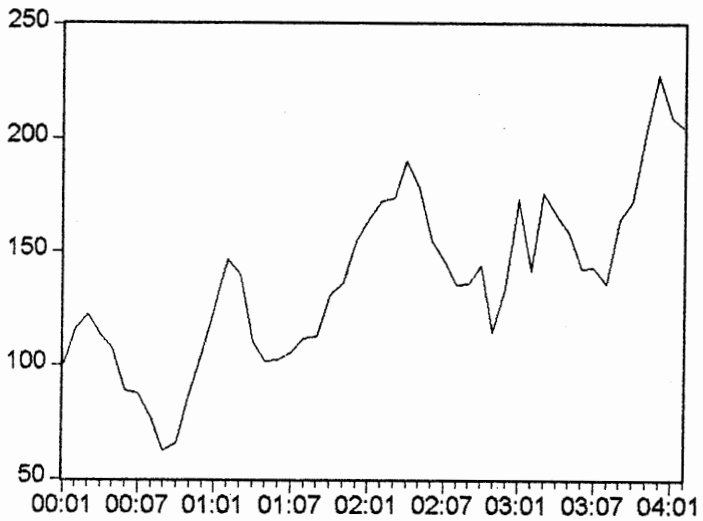
De la observación del **CUADRO 4.3** se puede concluir que hay razones suficientes para rechazar la hipótesis de que $\{Y_t\}:WN - 1$ y que, para un $\alpha = 0.05$, el intervalo de confiabilidad para el número esperado de rachas será $[9;23]$, lo cual implica un intervalo de confiabilidad para φ_0 de la forma $[5;21]$.

---

3) El test Q

Bajo el supuesto que $\{Y_t\}:RW - N\left(0,\sigma_Y^2\right)$ y, en consecuencia, que $\{\varepsilon_t\}:WN - N\left(0,\sigma_\varepsilon^2\right)$ y suponiendo, además, que las variables ε_t tienen momentos de sexto orden proporcionales a σ_ε^6, los

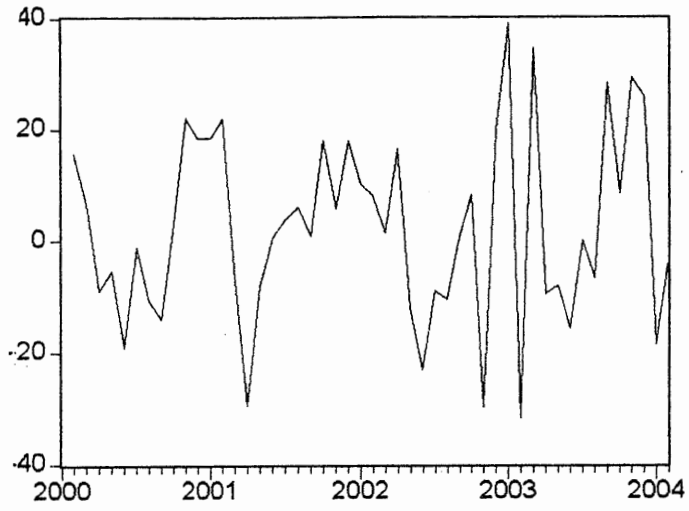
GRAFICO 4.10: Trayectoria del proceso $\{Y_t\}$:RW



CUADRO 4.1: Serie cronológica correspondiente al proceso $\{u_t\}=\{\nabla Y_t\}$

	u		u		u
Ene-00	NA	May-01	-8,225527	Oct-02	8,266698
Feb-00	15,655050	Jun-01	0,430316	Nov-02	-29,652680
Mar-00	6,639223	Jul-01	3,715019	Dic-02	19,813790
Abr-00	-9,121996	Ago-01	5,908094	Ene-03	39,060170
May-00	-5,435373	Sep-01	0,759630	Feb-03	-31,809850
Jun-00	-19,069450	Oct-01	18,060160	Mar-03	34,662480
Jul-00	-1,204232	Nov-01	5,641147	Abr-03	-9,589648
Ago-00	-10,839290	Dic-01	17,872910	May-03	-8,003697
Sep-00	-14,154440	Ene-02	10,097360	Jun-03	-15,819120
Oct-00	3,011134	Feb-02	8,028448	Jul-03	0,095667
Nov-00	22,148830	Mar-02	1,290467	Ago-03	-6,838684
Dic-00	18,461840	Abr-02	16,553330	Sep-03	28,219700
Ene-01	18,513020	May-02	-12,051330	Oct-03	8,358590
Feb-01	21,924690	Jun-02	-23,126360	Nov-03	29,252890
Mar-01	-7,017808	Jul-02	-9,103800	Dic-03	25,920090
Abr-01	-29,538240	Ago-02	-10,613630	Ene-04	-18,547170
		Sep-02	0,474416	Feb-04	-4,120416

GRAFICO 4.11: Trayectoria correspondiente al proceso $\{u_t\}=\{\nabla Y_t\}$



CUADRO 4.2: Valores esperados y desvíos estándar de la variable que representa el estimador de CJ, obtenidos a partir de una serie de 50 datos, con varianza σ_u^2=288.443, calculada como las diferencias de primer orden en un proceso RW-con tendencia

Φo	ρ_s*	ρ_s	CJ	σ (CJ)
0	0,5000	0,5000	1,0000	0,2887
5	0,6140	0,5260	1,1100	0,3357
6	0,6370	0,5380	1,1650	0,3597
7	0,6590	0,5510	1,2270	0,3862
8	0,6810	0,5660	1,3040	0,4211
9	0,7020	0,5810	1,3870	0,2102
10	0,7220	0,5980	1,4880	0,5037
15	0,8110	0,6930	2,2570	0,8837
20	0,8810	0,7900	3,7620	1,7621
30	0,9620	0,9270	12,6990	9,7811

CUADRO 4.3: Valores esperados y desvíos estándar de la variable que representa el estimador de CJ, obtenidos a partir de una serie de 50 datos, con varianza σ_u^2=288.443, calculada como las diferencias de primer orden en un proceso RW-con tendencia

Φo	ρ	E (n_r)	σ (n_r)
0	0,500	25	3,500
1	0,524	24	3,508
2	0,548	24	3,531
3	0,571	24	3,569
4	0,595	24	3,617
5	0,614	23	3,664
10	0,722	20	3,957
15	0,811	15	4,029
20	0,881	10	3,754
21	0,892	9	3,658
22	0,903	9	3,556
23	0,912	8	3,462
24	0,921	7	3,343
25	0,929	6	3,221
30	0,962	4	2,536
35	0,980	2	1,890
40	0,991	1	1,312

correspondientes estimadores muestrales de los coeficientes $\rho_j(\varepsilon)$ serán tales que:

$$E\left(\rho_j(\hat{\varepsilon})\right) = -\frac{n-j}{n(n-1)} + O\left(n^{-2}\right)$$

$$\gamma\left(\rho_i(\hat{\varepsilon}),\rho_j(\hat{\varepsilon})\right) = \begin{cases} O\left(n^{-2}\right) & para \quad i \neq j \\ \frac{n-i}{n^2} + O\left(n^{-2}\right) & para \quad i = j \neq 0 \end{cases}$$

(donde, como ya se mencionó en la Sec. 1.7.2.1, $O\left(n^{-2}\right)$ denota una función con orden asintótico igual a n^{-2}). Debe tenerse en cuenta que estos estimadores están afectados por un sesgo negativo cuyo origen se encuentra en la definición misma de los estimadores de las autocovarianzas como momentos muestrales mixtos centrados respecto de su valor medio muestral $\bar{\varepsilon} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t$: ocurre que la suma de los desvíos respecto de este valor medio debe ser nula, por lo que los desvíos positivos serán seguidos, eventualmente, por desvíos negativos y viceversa y, en consecuencia, el valor esperado de los productos cruzados de estos desvíos será negativo[21]. A fin de solucionar este inconveniente, Fuller, W. (1976) propuso, bajo el supuesto de que la hipótesis nula $H_0: \{Y_t\}: RW-1$ es verdadera, los siguientes estimadores corregidos:

---

21. Debe tenerse en cuenta que, para valores pequeños de n , este sesgo puede ser significativo (por ejemplo, para $n=10$, es de -10%).

$$\rho_j^*(\hat{\varepsilon}) = \rho_j(\hat{\varepsilon}) + \frac{n-j}{(n-1)^2}\left[1-\rho_j^2(\hat{\varepsilon})\right] \qquad \left(j = 1,2,\ldots,\frac{n}{4}\right)$$

que no son insesgados pero son tales que $E\left(\rho_j^*(\hat{\varepsilon})\right) = O\left(n^{-2}\right)$ y, suponiendo que los momentos de sexto orden son uniformemente acotados, se demuestra que son asintóticamente independientes y Normalmente distribuidos, $\rho_j^*(\hat{\varepsilon}) \xrightarrow{d} N\left(0,\frac{n-j}{n^2}\right)$ (como se verá más adelante, Lo, A.; MacKinlay, A.C. (1988), Richardson, M.; Smith, T. (1994) y Romano, J.; Thombs, L. (1996) demostraron que esta convergencia asintótica se verifica en condiciones menos estrictas, lo que les permitió construir tests para probar las hipótesis de $\{Y_t\}: RW-2$ y $\{Y_t\}: RW-3$).

A partir de estos resultados es posible construir un test para contrastar la hipótesis $H_0: \{Y_t\}: RW-N\left(0,\sigma_Y^2\right)$ y, por lo tanto, la hipótesis $H_0: \{\varepsilon_t\}: WN-N\left(0,\sigma_\varepsilon^2\right)$, utilizando un estadístico Q de la forma $Q_{(J)}^{(F)} = n^2\sum_{i=1}^{J}\frac{\rho_i^{*2}(\hat{\varepsilon})}{n-i}$ el cual, bajo el supuesto que H_0 es verdadera, observa un comportamiento asintótico del tipo $Q_{(J)}^{(F)} \xrightarrow{d} \chi^2_{(J-1)}$

Como se expresó en el Cap. 2, este test es potente con respecto a un número muy grande de alternativas lo cual, en el caso del RW, lo coloca en desventaja respecto de otros tests con hipótesis alternativa específica[22].

Ejemplo 4.4:

Sea un proceso $\{Y_t\}: RW$ de parámetro mensual para cuyo análisis se cuenta con una serie cronológica que abarca el período [ENE-2000,MAR-2004] (ver **GRAFICO 4.12**).

En el **GRAFICO 4.13** figura la trayectoria del proceso $\{u_t\}$ obtenido de calcular las primeras diferencias de $\{Y_t\}$ y en el **CUADRO 4.4** figura el correlograma de $\{u_t\}$, que contiene los estimadores de los coeficientes $\rho_j(u)$, los estimadores corregidos según la fórmula de Fuller y los desvíos estándar de estos últimos estimadores.

La observación de este cuadro permite concluir que no existen razones suficientes para rechazar la hipótesis nula $H_0: \{Y_t\}: RW-1$.

---

[22]. Por ejemplo, se puede concluir que la hipótesis nula plantea exclusivamente la ortogonalidad entre ∇Y_t y las combinaciones lineales de $\left(\nabla Y_{t-1}, \nabla Y_{t-2}, \ldots\right)$, no la ortogonalidad entre ∇Y_t y las funciones lineales y no-lineales de dichas diferencias (la experiencia indica que las rentabilidades observan alguna autocorrelación negativa para datos con frecuencia muy alta, ver Goodhart, C.A.E.; Figliouli, L. (1991), Bollerslev, T.; Domowitz, I. (1993), Zhou, B. (1996)). De modo que, si bien el rechazo de H_0 implica el rechazo de la hipótesis de martingala, la aceptación de H_0 implica la aceptación de la hipótesis de RW, pero no implica necesariamente la aceptación de la hipótesis de martingala (∇Y_t podría no ser ortogonal con alguna función no-lineal de $\{\nabla Y_{t-1}, \nabla Y_{t-2}, \ldots\}$). La martingala será tratada en la próxima sección.

GRAFICO 4.12: Trayectoria del proceso $\{Y_t\}$

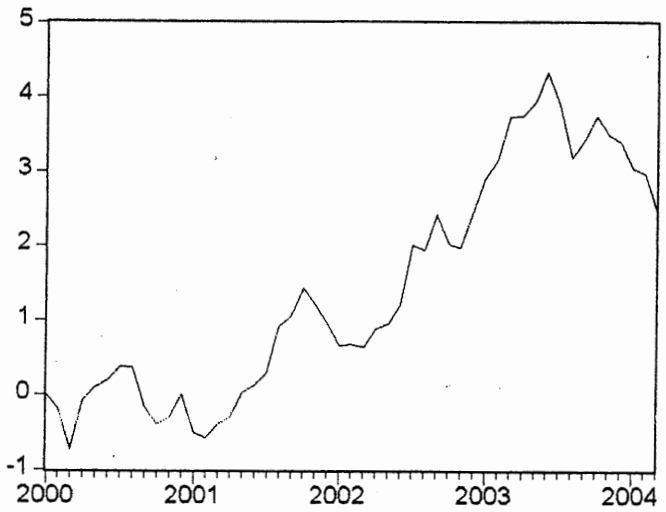
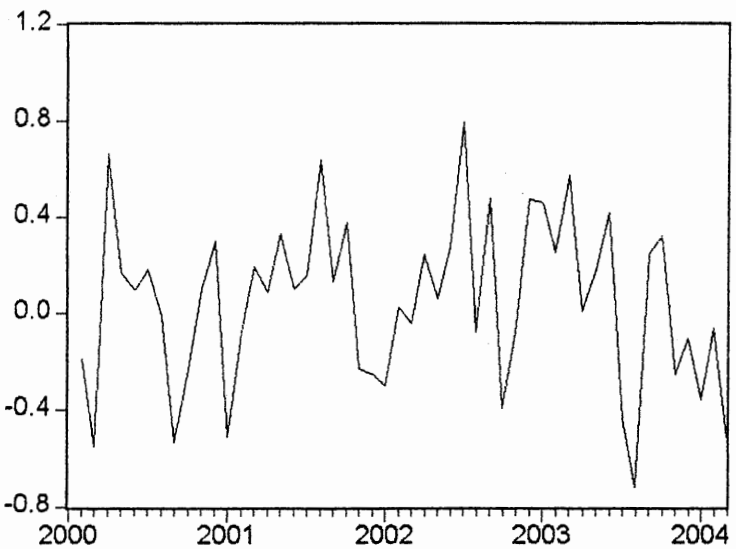


GRAFICO 4.13: Trayectoria del proceso $\{u_t\}$



CUADRO 4.4: Correlograma de $\{u_t\}$

j	$\hat{\rho}_j(\hat{u})$	$\hat{\rho}_j^*(\hat{u})$	$\hat{\sigma}(\hat{\rho}_j^*(\hat{u}))$
1	0,1140	0,1340	0,1440
2	-0,0120	0,0080	0,1386
3	0,0320	0,0520	0,1371
4	-0,0270	-0,0080	0,1356
5	-0,0820	-0,0630	0,1342
6	0,0000	0,0180	0,1327
7	-0,0620	-0,0440	0,1311
8	0,0110	0,0280	0,1311
9	0,0040	0,0210	0,1296
10	0,0450	0,0620	0,1265
11	0,0750	0,0910	0,1249
12	-0,2000	-0,1850	0,1233

4) El test de Lo-MacKinlay para la hipótesis RW-1

Lo, A.; MacKinlay, A.C. (1988), consideran la hipótesis nula:

$$H_0: Y_t - Y_{t-1} = \varphi_0 + \varepsilon_t = m(\nabla Y) + \varepsilon_t$$

(donde $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$). Suponiendo una muestra de tamaño $2n$, $\{Y_1, Y_2, ..., Y_{2n}\}$, sobre el proceso $\{Y_t\}$ se obtienen los siguientes estimadores máximo-verosímiles de los coeficientes $m(\nabla Y)$ y σ_ε^2:

$$\hat{m}(\nabla Y) = \frac{1}{2n-1}\sum_{t=2}^{2n}(Y_t - Y_{t-1}) = \frac{1}{2n-1}(Y_{2n} - Y_1)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{2n-1}\sum_{t=2}^{2n}\left[Y_t - Y_{t-1} - \hat{m}(\nabla Y)\right]^2$$

Bajo el supuesto que la hipótesis nula de comportamiento RW-1 de $\{Y_t\}$ (es decir, la hipótesis que el valor medio y la varianza de los incrementos $\{\nabla Y_t\}$ son funciones lineales de la longitud del intervalo que define dichos incrementos) es verdadera, la varianza σ_ε^2 puede ser estimada a partir de la muestra formada exclusivamente con las observaciones pares $\{Y_2, Y_4, ..., Y_{2n}\}$, obteniéndose que:

$$\left(\hat{\sigma}_\varepsilon^*\right)^2 = \frac{1}{2n-1}\sum_{t=2}^{n}\left[Y_{2t} - Y_{2t-2} - 2\hat{m}(\nabla Y)\right]^2$$

Se demuestra que, cuando n aumenta indefinidamente, los estimadores $\hat{m}(\nabla Y)$, σ_ε^2 y $\left(\hat{\sigma}_\varepsilon^*\right)^2$ convergen casi con certeza a los verdaderos valores de los parámetros (es decir, son fuertemente

consistentes) y que los estimadores $\hat{\sigma}_\varepsilon^2$ convergen en distribución a funciones Normales ((ver Stuart, A.; Ord, K. (1987)):

$$\sqrt{2n}\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right) \xrightarrow{d} N\left(0, 2\sigma_\varepsilon^4\right)$$

$$\sqrt{2n}\left[\left(\hat{\sigma}_\varepsilon^*\right)^2 - \sigma_\varepsilon^2\right] \xrightarrow{d} N\left(0, 4\sigma_\varepsilon^4\right)$$

Teniendo en cuenta este resultado y que, bajo el supuesto de que la hipótesis nula de comportamiento RW-1 de $\{Y_t\}$ es verdadera, se verifica que $\hat{\sigma}_\varepsilon^2$ es un estimador asintóticamente eficiente de σ_ε^2 y, de acuerdo con la condición de Hausman, según la cual la varianza de la diferencia entre un estimador consistente y un estimador asintóticamente eficiente está definida por la diferencia de las varianzas asintóticas[23], se puede asegurar que el estimador de la diferencia de las varianzas definido como $\hat{\sigma}_D^2(2) = \left(\hat{\sigma}_\varepsilon^2\right)^2 - \hat{\sigma}_\varepsilon^2$ es tal que $\sqrt{2n}\hat{\sigma}_D^2(2) \xrightarrow{d} N\left(0, 2\sigma_\varepsilon^4\right)$. Luego, se puede concluir que, en particular, bajo el supuesto de que la hipótesis nula de comportamiento RW-Normal de $\{Y_t\}$ [24], $H_0: \nabla Y_t = m(\nabla Y) + \varepsilon_t$ (donde $\{\varepsilon_t\}: N\left(0, \sigma_\varepsilon^2\right)$), es verdadera, el estadístico a utilizar quedará asintóticamente definido de la siguiente forma: $Z = \dfrac{\hat{\sigma}_D^2(2)}{\sqrt{2\hat{\sigma}_\varepsilon^4}} \xrightarrow{d} N(0;1)$ (el denominador de este estadístico, puede ser construido utilizando cualquier estimador consistente de σ_ε^2). Asimismo, a partir de la convergencia en distribución de la variable definida por la diferencia de varianzas y utilizando la aproximación de primer orden del desarrollo de Taylor, se obtiene la siguiente convergencia en distribución:

$$\sqrt{2n}\left[\frac{\left(\hat{\sigma}_\varepsilon^*\right)^2}{\hat{\sigma}_\varepsilon^2} - 1\right] = \sqrt{2n}\left\{\left[\hat{\sigma}^{**}(2)\right]^2 - 1\right\} \xrightarrow{d} N(0;2)$$

De modo que, bajo el supuesto de que la hipótesis nula $\{Y_t\}: RW - N\left(0, \sigma_Y^2\right)$ es verdadera, el estadístico a utilizar será tal que[25]:

---

[23]. De acuerdo con esta condición, dados dos estimadores ($\hat{\theta}$ y $\hat{\theta}^*$) de un parámetro θ, para que $\hat{\theta}$ sea asintóticamente eficiente debe poseer la propiedad de ser asintóticamente incorrelacionado con la variable aleatoria definida por la diferencia $\left(\hat{\theta}^* - \hat{\theta}\right)$. El no cumplimiento de esta condición implicaría la existencia de alguna combinación lineal de $\hat{\theta}$ y $\left(\hat{\theta}^* - \hat{\theta}\right)$ que es más eficiente que $\hat{\theta}$, lo cual constituiría una contradicción a la supuesta condición de eficiencia. Es decir, si se verifica que $\left[\sigma^{(a)}\left(\hat{\theta}^*\right)\right]^2 = \left[\sigma^{(a)}\left(\hat{\theta} + \hat{\theta}^* - \hat{\theta}\right)\right]^2 + \left[\sigma^{(a)}\left(\hat{\theta}^* - \hat{\theta}\right)\right]^2$ (donde $\left[\sigma^{(a)}(\bullet)\right]^2$ denota la varianza asintótica del estimador), entonces $\left[\sigma^{(a)}\left(\hat{\theta}^* - \hat{\theta}\right)\right]^2 = \left[\sigma^{(a)}\left(\hat{\theta}^*\right)\right]^2 - \left[\sigma^{(a)}\left(\hat{\theta}\right)\right]^2$ (ver Hausman, J. (1978)).

[24]. La condición de Normalidad es conveniente para las demostraciones, pero puede ser sustituida por la más general de incrementos iid con momentos de cuarto orden finitos.

[25]. Debido a su independencia del factor escala, este test es, en general, preferido al construido sobre la diferencia de las variables (ver Cochran, J. (1988), Lo, A.; MacKinlay, A.C. (1988)(1989), Faust, J. (1992)).

$$Z = \frac{\sqrt{2n}\left\{\left[\hat{\sigma}^{**}(2)\right]^2 - 1\right\}}{\sqrt{2}} \xrightarrow{d} N(0,1)$$

Los tests anteriores pueden ser generalizados suponiendo una muestra de tamaño τn (donde τ denota un número entero mayor que uno) sobre $\{Y_t\}$. Se obtienen, entonces, los siguientes estimadores de los coeficientes $m(\nabla Y)$ y σ_ε^2:

$$\hat{m}(\nabla Y) = \frac{1}{\tau n - 1}\sum_{t=2}^{\tau n}(Y_t - Y_{t-1}) = \frac{1}{\tau n - 1}(Y_{\tau n} - Y_1)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{\tau n - 1}\sum_{t=2}^{\tau n}\left[Y_t - Y_{t-1} - \hat{m}(\nabla Y)\right]^2$$

Igual que en el caso anterior, suponiendo que la hipótesis de comportamiento RW-1 de $\{Y_t\}$ sea verdadera, σ_ε^2 puede ser estimado utilizando una muestra reducida, de la siguiente forma:

$$\left(\sigma_\varepsilon^*(\tau)\right)^2 = \frac{1}{\tau n - 1}\sum_{t=2}^{n}\left[Y_{\tau t} - Y_{\tau t-2} - \tau\hat{m}(\nabla Y)\right]^2$$

A partir de estos estimadores se obtiene que

$$\hat{\sigma}_D^2(\tau) = \left(\hat{\sigma}_\varepsilon^*(\tau)\right)^2 - \hat{\sigma}_\varepsilon^2$$

$$\left[\hat{\sigma}^{**}(\tau)\right]^2 = \frac{\left(\hat{\sigma}_\varepsilon^*(\tau)\right)^2}{\hat{\sigma}_\varepsilon^2}$$

y, bajo el supuesto de que la hipótesis de comportamiento RW-1 de $\{Y_t\}$ es verdadera, se verifican las siguientes convergencias en distribución:

$$\sqrt{\tau n}\,\hat{\sigma}_D^2(\tau) \xrightarrow{d} N\left[0, 2(\tau - 1)\sigma_\varepsilon^4\right]$$

$$\sqrt{\tau n}\left\{\left[\hat{\sigma}^{**}(\tau)\right]^2 - 1\right\} \xrightarrow{d} N\left[0, 2(\tau - 1)\right]$$

Existen alternativas que aumentan la eficiencia y la potencia de los tests anteriores, las cuales se obtienen: i) sustituyendo el estimador $\left(\hat{\sigma}_\varepsilon^*\right)^2$ (obtenido a partir de una muestra de n elementos) por el siguiente estimador, obtenido a partir de $(\tau n - \tau)$ términos superpuestos del proceso $\{\varepsilon_t\}$:

$$\left(\hat{\sigma}^{(C)*}(\tau)\right)^2 = \frac{1}{\tau n - 1}\sum_{t=\tau+1}^{\tau n}\left(Y_t^* - Y_{t-1}^* - \tau\hat{m}(\nabla Y)\right)^2$$

y **ii)** utilizando, respectivamente, los siguientes estimadores insesgados de σ_ε^2 y $\left(\sigma_\varepsilon^*(\tau)\right)^2$:

$$\left(\hat{\sigma}_\varepsilon^{**}\right)^2 = \frac{1}{\tau n - 1}\sum_{t=2}^{\tau n}\left[Y_t - Y_{t-1} - \hat{m}(\nabla Y)\right]^2$$

$$\left(\hat{\sigma}_\varepsilon^{(C)*}(\tau)\right)^2 = \frac{1}{n^*}\sum_{t=\tau+1}^{\tau n}\left[Y_t - Y_{t-1} - \tau\hat{m}(\nabla Y)\right]^2 \qquad \left(n^* = \tau(\tau n - \tau)\left(1 - \frac{\tau}{\tau n - 1}\right)\right)$$

En este segundo caso se obtiene el siguiente estimador insesgado de la diferencia de las varianzas de los incrementos:

$$\left(\hat{\sigma}^{**}(\tau)\right)^2 = \left(\hat{\sigma}_\varepsilon^{(C)*}(\tau)\right)^2 - \left(\hat{\sigma}_\varepsilon^{**}\right)^2$$

y un estimador sesgado del cociente de las varianzas de la forma:

$$\left(\hat{\sigma}^{(C)*}(\tau)\right)^2 = \frac{\left(\hat{\sigma}_\varepsilon^{(C)*}(\tau)\right)^2}{\left(\hat{\sigma}_\varepsilon^{**}\right)^2}$$

Este sesgo se origina en la desigualdad de Jensen (ver Jensen, M.; Ruback, R. (1983)). No obstante, debe tenerse en cuenta que, dadas sus propiedades muestrales para n finito, este estimador del cociente de varianzas se aproxima más a su comportamiento asintótico que el estimador $\left(\sigma_\varepsilon^*(\tau)\right)^2$ (ver Lo, A.; MacKinlay, A.C. (1989)). Bajo el supuesto de que la hipótesis nula $H_0: \nabla Y_t = m(\nabla Y) + \varepsilon_t$ (donde $\{\varepsilon_t\}: WN - N\left(0,\sigma_\varepsilon^2\right)$) es verdadera, se puede concluir que estos estimadores convergen en-distribución a la función Normal, de la siguiente forma:

$$\left(\hat{\sigma}_D^{**}(\tau)\right)^2 \xrightarrow{d} N\left[0, \frac{2(2\tau-1)(\tau-1)}{3\tau}\sigma_\varepsilon^4\right]$$

$$\sqrt{\tau n - 1}\left\{\left[\hat{\sigma}^{(C)*}(\tau)\right]^2 - 1\right\} \xrightarrow{d} N\left[0, \frac{2(2\tau-1)(\tau-1)}{3\tau}\sigma_\varepsilon^4\right]$$

donde σ_ε^4 puede ser estimado a partir de la expresión:

$$\left[\left(\hat{\sigma}_\varepsilon^{**}\right)^2\right]^2 = \frac{1}{(\tau n - 2)^2}\left\{\sum_{t=2}^{\tau n}\left[Y_t^* - Y_{t-1}^* - \hat{m}(\nabla Y)\right]\right\}^2$$

Luego, será:

$$\left(\hat{\sigma}_D^{**}(\tau)\right)^2 = \frac{\sqrt{\tau n - 1}}{\left(\hat{\sigma}_\varepsilon^{**}\right)^4}\sqrt{\frac{3\tau}{2(2\tau-1)(\tau-1)}} \xrightarrow{d} N(0;1)$$

$$\sqrt{\tau n - 1}\left\{\left[\hat{\sigma}^{(C)*}(\tau)\right]^2 - 1\right\}\sqrt{\frac{3\tau}{2(2\tau-1)(\tau-1)}} \xrightarrow{d} N(0;1)$$

5) La regla de $\tau^{½}$

La hipótesis sobre la **estabilidad de las volatilidades** -común a los tres tipos de RW- implica la condición de que la varianza de las perturbaciones es una función lineal de la longitud del período que las separa. Es decir, suponiendo que $\{Y_t\}: RW-1$, se debe verificar que:

$$\sigma^2(\varepsilon_{t,\tau}) = \sigma^2(\varepsilon_t + \varepsilon_{t-1} + ... + \varepsilon_{t-(\tau-1)}) = \tau\sigma_\varepsilon^2$$

(donde σ_ε^2 denota la volatilidad para un período unitario)[26] o, lo que es lo mismo, que el desvío estándar de las perturbaciones, para un período de longitud τ, es igual a $\sqrt{\tau}$ multiplicado por el desvío estándar de las perturbaciones para un período unitario (esta condición se conoce como **medida de Wiener** o **regla de $\tau^{½}$**)[27].

Dado un intervalo de tiempo de longitud τ, la **hipótesis varianza-tiempo** correspondiente a un proceso RW-1 puede ser expresada, entonces, como:

$$H_0: \sigma^2(\varepsilon_{t,\tau}) = \tau\sigma_\varepsilon^2$$

siendo $H_1: \sigma^2(\varepsilon_{t,\tau}) \neq \tau\sigma_\varepsilon^2$ la hipótesis alternativa. En particular, para $\tau = 2$, se verifica que:

$$\left[\sigma^{**}(2)\right]^2 = \frac{\sigma^2(\varepsilon_{t,2})}{2\sigma_\varepsilon^2} = \frac{\sigma^2(\varepsilon_t + \varepsilon_{t-1})}{2\sigma_\varepsilon^2} = \frac{2\sigma_\varepsilon^2 + 2\gamma(\varepsilon_{t-1}, \varepsilon_t)}{2\sigma_\varepsilon^2} = 1 + 2\rho_1(\varepsilon)$$

De modo que aceptar la hipótesis varianza-tiempo (es decir, la hipótesis nula $H_0: \{Y_t\}: RW-1$) implica aceptar que $\rho_1(\varepsilon) = 0$ y, por lo tanto, que $\left(\sigma^{**}(2)\right)^2 = 1$. De la misma forma, aceptar que $\left(\sigma^{**}(2)\right)^2 > 1$ (es decir, que la varianza de la suma de las variables ε_{t-1} y ε_t es mayor que la suma de las varianzas para períodos unitarios o, lo que es lo mismo, que las varianzas crecen en forma más rápida que el crecimiento lineal) implica reconocer la existencia de autocorrelación positiva y aceptar que $\left(\sigma^{**}(2)\right)^2 < 1$ (es decir, que la varianza de la suma de las variables ε_{t-1} y ε_t es menor que la suma de las varianzas para períodos unitarios o, lo que es lo mismo, que las varianzas crecen en forma más lenta que el crecimiento lineal) implica reconocer la existencia de autocorrelación negativa en las rentabilidades[28].

---

[26]. Se consideró un proceso RW-1, ya que obviamente, esta propiedad de linealidad es más difícil de demostrar para los procesos RW-2 y RW-3, debido a que las varianzas de los incrementos, en estos casos, pueden ser heterocedásticas.

[27]. Como se verá en la Sec. 4.5, esta **relación tiempo-error** de acuerdo con la nomenclatura de Working, H. (1949), o **función varianza-tiempo** de acuerdo con la nomenclatura de Sprenkle, C.M. (En Cootner, P.H. (1900)) y Poole, W. (1967), se origina en la propiedad del **movimiento Browniano** de ser 0.5-autosimilar. A partir de esta condición se puede concluir en forma inmediata que la volatilidad (es decir, la incertidumbre sobre el comportamiento) de las perturbaciones crece cuando aumenta el valor de τ.

[28]. En la comparación entre las características de las perturbaciones para un período de longitud τ y para un período unitario, debe tenerse en cuenta que la estructura de la kurtosis: **i)** es independiente de la estructura de la volatilidad y **ii)** depende fundamentalmente de los postulados del teorema central del límite. Sea $\{\varepsilon_t\}$ $(t = 1,2,...,n)$ una sucesión de variables aleatorias iid, entonces la kurtosis de la variable $\varepsilon_{t,\tau} = \sum_{t=1}^{\tau} \varepsilon_t$ está definida por $K(\varepsilon_{t,\tau}) = \frac{K(\varepsilon_t)}{\tau}$.

En general, razonando en forma similar, se puede escribir:

$$\left[\sigma^{**}(\tau)\right]^2 = \frac{\sigma^2\left(\varepsilon_{t,\tau}\right)}{\tau\sigma_\varepsilon^2} = 1 + 2\sum_{j=1}^{\tau-1}\left(1-\frac{j}{\tau}\right)\rho_j(\varepsilon)$$

De modo que la aceptación de la hipótesis de comportamiento RW-1 del proceso $\{Y_t\}$ implica la aceptación de la hipótesis $\left(\sigma^{**}(\tau)\right)^2 = 1$ [29].

El diseño de este test encierra una dificultad originada en la no-independencia de los estimadores de las varianzas de la serie para un período unitario ($\hat{\sigma}_\varepsilon^2$) y de la serie para períodos de longitud τ ($\hat{\sigma}^2(t,\tau)$), dada la no-independencia de las series $\{Y_{t+1}^* - Y_t^*\}$ y $\{Y_{t+\tau}^* - Y_t\}$ $(t = 1,2,...,n)$. Una solución a este problema puede ser obtenida mediante la utilización del **"variate difference method"** -debido a G. Tintner[30] - que permite comparar las varianzas correspondientes a dos diferenciaciones sucesivas de una serie, mediante la selección de dos series de observaciones de la forma $\{Y_t\}$ $(t = 1,4,7,10,...)$ y $\{\nabla Y_t\}$ $(t = 2,5,8,11,...)$, de modo de generar muestras independientes[31].

6) El test de Young

Young, W.E. (1971) propuso una solución alternativa al problema de la contrastación de la

---

[29]. En realidad, dado que $\sigma_\varepsilon^2 < \infty$ y que la varianza promedio $\frac{1}{n}\sum_{t=1}^{n}\sigma_\varepsilon^2(t)$ converge a un número positivo finito, aceptar la hipótesis $\left(\sigma^{**}(\tau)\right)^2 = 1$ implica aceptar también las hipótesis de comportamiento $\{Y_t\}: RW-2$ y $\{Y_t\}: RW-3$. Esta relación entre el coeficiente $\left(\sigma^{**}(\tau)\right)^2$ y la función de autocorrelaciones del proceso $\{\varepsilon_t\}$ permite contrastar hipótesis alternativas distintas. Supóngase, por ejemplo, una representación alternativa $AR(1)$ de la forma $\varepsilon_t = \phi_1\varepsilon_{t-1} + \varepsilon_t^*$, entonces, será $\left[\sigma^{**}(\tau)\right]^2 = 1 + 2\sum_{j=1}^{\tau-1}\left(1-\frac{j}{\tau}\right)\varphi_1^j = 1 + \frac{2}{1-\varphi_1}\left[\varphi_1 - \frac{\varphi_1^\tau}{\tau} - \frac{\varphi_1 - \varphi_1^\tau}{\tau(1-\varphi_1)}\right]$.

[30]. Ver Tintner, G. (1940).

[31]. El "variate difference method" -que fue diseñado, originalmente, para la eliminación de la componente tendencia de una serie cronológica- supone que un proceso económico dinámico está formado, aditivamente, por un **comportamiento esperado** -definido por una curva polinómica continua, eliminable por diferenciación- y una **componente aleatoria**, formada por residuos independientes. La varianza del proceso $\{\nabla^k Y_t\}$ $(k = 1,2,...)$ está dada por $\sigma_{\nabla^k Y}^2 = \binom{2k}{k}\sigma_\varepsilon^2$. Si, para un valor dado de k, se verifica que $\sigma_Y^2 = \frac{\sigma_{\nabla^k}^2}{\binom{2k}{k}} = \frac{\sigma_{\nabla^k}^2}{\binom{2k+2}{k+1}}$, se puede concluir que, con la diferenciación de orden k, se ha eliminado la parte componente que define el comportamiento esperado (es decir, se puede suponer que la parte componente tendencia está definida por un polinomio de grado k). El test para la consideración de esta hipótesis está diseñado utilizando el estadístico $\frac{\hat{\sigma}_{\nabla^k}^2}{\binom{2k}{k}} - \frac{\hat{\sigma}_{\nabla^{k+1}}^2}{\binom{2k+2}{k+1}}$. Un inconveniente de este método radica en la pérdida de información originada en la forma en que se realiza la selección de las observaciones.

hipótesis varianza tiempo, basada en una transformación ortogonal aplicada sobre el proceso para el período unitario y sobre el proceso para períodos de longitud τ [32].

El test de Young, en su forma más general, puede ser expresado como sigue: Sean $\{\varepsilon_{1t}\}$ y $\{\varepsilon_{2t}\}$ $(t = 1,2,...,n)$ dos procesos estacionarios Normales y sean los procesos transformados $\{z_t\} = \{\varepsilon_{1t} + \varepsilon_{2t}\sigma^*\}$ y $\{w_t\} = \{\varepsilon_{1t} - \varepsilon_{2t}\sigma^*\}$, donde $\sigma^* = \dfrac{\sigma(\varepsilon_1)}{\sigma(\varepsilon_2)}$. Se verificará, entonces, que:

$$\sigma^2(z) = 2\sigma^2(\varepsilon_1)\left[1 + \rho(\varepsilon_{1t}, \varepsilon_{2t})\right]$$

$$\sigma^2(w) = 2\sigma^2(\varepsilon_1)\left[1 - \rho(\varepsilon_{1t}, \varepsilon_{2t})\right]$$

$$\gamma(z_t, w_t) = E(z_t w_t) - E(z_t)E(w_t) = E(\varepsilon_{1t}^2) - \left[E(\varepsilon_{1t})\right]^2 - \sigma^{*2}\left\{E(\varepsilon_{2t}^2) - \left[E\varepsilon_{2t}\right]^2\right\} = \sigma^2(\varepsilon_1) - \sigma^{*2}\sigma^2(\varepsilon_2)$$

De modo que la igualdad $\sigma^2(\varepsilon_1) = (\sigma^*)^2\sigma^2(\varepsilon_2)$ es condición necesaria y suficiente para que $\gamma(z_t, w_t) = 0$. Con lo que queda demostrado que contrastar las hipótesis $H_0: \sigma^2(\varepsilon_1) = \sigma^{*2}\sigma^2(\varepsilon_2)$ y $H_1: \sigma^2(\varepsilon_1) \neq \sigma^{*2}\sigma^2(\varepsilon_2)$ es equivalente a contrastar las hipótesis $H_0: \rho(z_t, w_t) = 0$ y $H_1: \rho(z_t, w_t) \neq 0$.

Sean, en particular $\varepsilon_t = \ln(Y_{t+1}) - \ln(Y_t)$ $(t = 1,2,...,n)$ y $\varepsilon_t^\tau = \ln(Y_{t+\tau}) - \ln(Y_t)$ $(t = 1,2,...,n-\tau)$, las variables que representan las perturbaciones para el período unitario y para períodos de longitud τ. Para testear la hipótesis $\rho(z_t, w_t) = 0$, se utiliza el estadístico:

$$T = \frac{\sqrt{n-\tau-2}\,\hat{\rho}(z_t, w_t)}{\sqrt{1-\hat{\rho}^2(z_t, w_t)}} : t_{(n-\tau-2)}$$

Este método se caracteriza por la utilización de observaciones superpuestas para la definición de los pares de valores $(\varepsilon_t, \varepsilon_t^{(\tau)})$. Una técnica alternativa consiste en tomar en consideración las τ muestras posibles de las series, que no contengan observaciones superpuestas (por ejemplo, dada una serie unitaria de 60 observaciones y un valor $\tau = 12$, se toman, para las series con períodos de longitud τ, las 11 muestras de tamaño 4 y una de tamaño 5, no-superpuestas)[33]. La ventaja de utilizar observaciones superpuestas radica, fundamentalmente, en el tamaño de la muestra: en este caso el tamaño es $n - \tau$, en tanto que con observaciones no-superpuestas, se reduciría a $\dfrac{n}{\tau}$ ó $\dfrac{n}{\tau} - 1$. Una desventaja del primer método es que las observaciones superpuestas son asimilables a un promedio móvil cuyas ponderaciones, para las observaciones iniciales y finales de la serie unitaria, son menores que las correspondientes a las observaciones intermedias (dada una serie unitaria, $\{\varepsilon_t\}$ $(t = 1,2,...,n)$, en la serie para períodos de

---

[32]. Este test constituye, en realidad, una modificación debida, a Morgan, W.A. (1939-40), de un test anterior construido para considerar la significatividad de la diferencia entre dos varianzas en una muestra obtenida de una población Normal bidimensional.

[33]. Una solución alternativa -utilizada por Sprenkle, C.M. (En Cootner, P.H. (ed.) (1900))- se basa en la utilización de la distribución χ^2 suponiendo que no existe error en la estimación de la varianza de la serie unitaria ($\hat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2$). El inconveniente que afecta a este procedimiento es que requiere una gran cantidad de información, y es aplicable a datos diarios o semanales, pero no a datos mensuales.

longitud τ : los valores ε_1 y ε_n son utilizados sólo una vez, los valores ε_2 y ε_{n-1} son utilizados dos veces,..., los valores $\varepsilon_\tau, \varepsilon_{\tau+1}, ..., \varepsilon_{n-\tau}$ son utilizados τ veces)[34].

Con respecto al análisis de la hipótesis nula $\{Y_t\}$:*RW* - 3 , Lo, A.; McKinlay, A.C. (1988) propusieron el siguiente test. Dado un proceso $\{Y_t\}$ cuyos incrementos admiten una representación de la forma[35]:

$$Y_t - Y_{t-1} = \nabla Y_t = m(\nabla Y_t) + \varepsilon_t$$

se considera la siguiente hipótesis nula compuesta[36]:

$$H_0^* : H_0^{(1)} : El\ proceso\ \{\varepsilon_t\} : WN - debil$$

$$H_0^{(2)} : El\ proceso\ \{\varepsilon_t\}\ es\ asintoticamente\ independiente\ y\ cumple\ las\ condiciones$$

$$\phi - mixing,\ con\ coeficientes\ de\ medida\ \frac{\tau}{2\tau - 1}\ (para\ \tau > 1)\ o\ \alpha - mixing,$$

$$con\ coeficientes\ de\ medida\ \frac{\tau}{\tau - 1}\ (para\ \tau > 1),\ de\ modo\ que,\ para\ todo$$

$$t\ y\ todo\ j \geq 0,\ es\ posible\ hallar\ un\ \delta > 0\ tal\ que\ E\left[\left|\varepsilon_{t-j}\varepsilon_t\right|^{2(\tau+\delta)}\right] < \Delta < \infty.$$

$$H_0^{(3)} : Se\ verifica\ que\ \lim_{m\to\infty} \frac{1}{m} \sum_{t=1}^{m} E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty.$$

$$H_0^{(4)} : Los\ \hat{\rho}_j(\varepsilon)\ son\ asintoticamente\ incorrelacionados.$$

Obsérvese que en la hipótesis $H_0^{(1)}$, queda excluido el caso de las variables ε_t con distribución de probabilidades con varianza infinita (de aparición muy frecuente en las aplicaciones financieras).

---

34. Obviamente, la diferencia entre estos dos métodos será menor cuanto menor sea τ con relación al número de observaciones de la serie para el período unitario. Un inconveniente adicional que parecería acarrear la utilización de observaciones superpuestas es su posibilidad de generar correlaciones negativas entre las series (ver Scott, M.F.G. (1987)(1991)).

35. Este test se basa en los métodos relacionados con las condiciones de heterocedasticidad de White, H. (1980) y White, H.; Domowitz, I. (1984) (ver Lo, A. ;MacKinlay, A.C. (1988)),

36 Obsérvese que la hipótesis nula admite formas generales de heterocedasticidad que incluyen variaciones determinísticas en $\sigma^2(\varepsilon_t)$ (por ejemplo, aquellas debidas a la influencia de factores estacionales) y variaciones de las varianzas condicionadas que admiten representaciones GARCH. Debe tenerse en cuenta que se pueden obtener aproximaciones generales a estas relaciones no-lineales (en realidad, a cualquier tipo de función no-lineal) con un alto grado de confiabilidad, mediante la utilización de **modelos de redes neuronales** (ver White, H. (1988), Fishman, M.B.; Barr, D.S,; Loick, W.J. (1991), Shih, Y.L. (1991), Katz, J.O. (1992), Kean, J. (1992), Swales, G.S.; Yoon, Y. (1992), Wong, F.S. (1992), Baestaens, D.E.; van der Berg, W.M.; Vaudrey, H. (1996), Kuan, C.M.: Liu, T. (1995), Refenes, A.; Abu-Mustafa, Y.; Mood, Y.J.; Wigenda, A. (1996), Bolland, P.J.; Connor, J.T.; Refenes, A.P. (1998)).

En lo que respecta a la hipótesis $H_0^{(2)}$, debe tenerse en cuenta que, como se mencionó en la Sec. 1.5, al tratar las condiciones "mixing", dadas dos variables aleatorias X e Y definidas en los subespacios de probabilidades $\Im$ y $\Re$, respectivamente, las diferencias $\left\| f_{XY}(x,y) - f_X(x) f_Y(y) \right\|$ y $\left\| f(y/x) - f_Y(y) \right\|$ (donde $\|\bullet\|$ denota alguna medida de distancia) pueden ser interpretadas como medidas generales de dependencia entre X e Y. Una forma particular de estas medidas es la denominada **Δ de Hoeffding** (ver Hoeffding, W. (1948)) definida de la siguiente forma:

$$\Delta(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[f_{XY}(x,y) - f_X(x)f_Y(y)\right]^2 f_{XY}(x,y)\,dxdy$$

Alternativas a ésta y basadas en los mismos conceptos, existen otras medidas, entre la que cabe mencionar: **i)** los **coeficientes δ**, que consideran el valor absoluto de la diferencia de una forma análoga al Δ:

$$\delta_1(X,Y) = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|f_{XY}(x,y) - f_X(x)f_Y(y)\right| f_{XY}(x,y)\,dxdy$$

$$\delta_2(X,Y) = 12\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|f_{XY}(x,y) - f_X(x)f_Y(y)\right| f_X(x)f_Y(y)\,dxdy$$

ii) la **diferencia de información** que, como se mencionó en la Sec. 2.10.5.2, está basada en la medida de divergencia entre dos distribuciones de probabilidades o medida de distancia dirigida de Kullback-Leibler (ver Kullback, S.; Leibler, R.A. (1951)), definida por:

$$\mathrm{K}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ln\left(\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right) f_{XY}(x,y)\,dxdy$$

iii) el **cuadrado del coeficiente de contingencia**:

$$\varphi^2(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right] f_{XY}(x,y)\,dxdy - 1$$

y iv) el **coeficiente de rangos de Spearman** (ver Spearman, C. (1904)):

$$S(X,Y) = 3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[2F_x(x)-1\right]\left[2F_Y(y)-1\right] f_{XY}(x,y)\,dxdy$$

(donde la utilización de los factores $\left[2F(\bullet)-1\right]$ se debe a la necesidad de transformar a $F(\bullet)$ en una distribución simétrica con respecto a cero).

Como se mencionó en la Sec. 1.5, las hipótesis $H_0^{(2)}$ y $H_0^{(3)}$ establecen restricciones sobre el grado máximo de dependencia y variabilidad de las variables ε_t aceptables como condición para la vigencia de la ley de los grandes números y del teorema central del límite.

Teniendo en cuenta que, en condiciones bastante generales, se verifica que:

$$\lim_{n\to\infty}\left[\hat{\sigma}^{(C)*}(\tau)\right]^2 = 1+2\sum_{j=1}^{\tau-1}\left(1-\frac{j}{\tau}\right)\hat{\rho}_j(\varepsilon)$$

y que, bajo el supuesto de que la hipótesis H_0^* es verdadera, se verifica que: **i)** cuando n aumenta indefinidamente, $\left[\hat{\sigma}^{(C)*}(\tau)\right]^2 - 1 \xrightarrow{ccc} 0\ (\tau = 2,3,...)$; **ii)** el estimador:

$$\sigma^2\left(\hat{\rho}_j(\varepsilon)\right) = \frac{\tau n \sum_{i=j+1}^{\tau n}\left(Y_i^* - Y_{i-1}^* - \hat{m}(v)\right)^2\left(Y_{i-j}^* - Y_{i-j-1}^* - \hat{m}(v)\right)^2}{\left[\sum_{j=1}^{\tau n}\left(Y_i^* - Y_{i-1}^* - \hat{m}(v)\right)\right]^2}$$

es consistente en la presencia de heterocedasticidad y **iii)** el estimador de la varianza:

$$\hat{\sigma}^2\left\{\left[\hat{\sigma}^{(c)*}(\tau)\right]^2\right\} = 4\sum_{j=1}^{\tau-1}\left(1-\frac{j}{\tau}\right)^2\sigma^2\left(\hat{\rho}_j(\varepsilon)\right)$$

es consistente en la presencia de heterocedasticidad, se puede concluir que la hipótesis H_0^* puede ser contrastada a partir de la información que proporciona el estimador:

$$\frac{\sqrt{\tau n}\left\{\left[\hat{\sigma}^{(c)*}(\tau)\right]^2 - 1\right\}}{\hat{\sigma}\left\{\left[\hat{\sigma}^{(c)*}(\tau)\right]^2\right\}} \xrightarrow{d} N(0;1).$$

4.3.- La martingala

El "random walk" pertenece a una clase más general de procesos, conocida como **martingala** (cuya esencia coincide con el principio de **juego equitativo**[37] en el que las ganancias de los jugadores están valuadas por las esperanzas matemáticas condicionadas)[38], la cual se caracteriza por cumplir las siguientes condiciones: **i)** $E\left(|Y_t|\right) < \infty\ (t = 1,2,...)$ y **ii)** $E\left(Y_t / Y_{t-1}, Y_{t-2}, ...\right) = Y_{t-1}$. Esta última condición de que la mejor predicción para el período t es, simplemente, el estado del proceso en el momento $t-1$ $(t = 2,3,...)$ (conocida como la **propiedad de la martingala**) obviamente también puede ser expresada de la siguiente

---

[37]. "*El principio fundamental de los juegos es simplemente la igualdad de condiciones: de los contrincantes, de los expectadores, del dinero, de la situación, del cubilete y del dado mismo. Para extender el significado de este principio usted debe partir de dicha igualdad, si es a favor de su oponente usted es un necio, si es a su favor usted es injusto*" (Cardano, G.: "Liber de ludo aleæ"). Ver Landro, A.H. (2002).

[38]. El proceso denominado martingala (introducido en la literatura probabilística por Ville, J. (1939), pero apreciado, en cuanto a su utilidad en las aplicaciones, recién después de la publicación de los trabajos de Doob, J.L. (1953)) es asimilable a un proceso que representa la fortuna de un jugador en un estado dado del juego y que, en virtud del criterio del juego equitativo, el valor esperado de dicha fortuna al cabo de una nueva partida es igual al monto presente de su fortuna y no está afectodo por su historia pasada. La denominación "martingala" se relaciona con la ciudad de Martigues (Provence) cuyos habitantes eran famosos por ejecutar un sistema de juego que consistía en duplicar la apuesta precedente cada vez que el jugador pierde, retornando a la apuesta inicial después de cada triunfo (ver Weaver, W. (1982), Gingerenzer, G.; Swijtink, Z.; Porter, T.; Daston, L.; Beatty, J.; Kruger, I. (1989)).

forma[39]:

$$E\left(Y_t - Y_{t-1} / Y_{t-1}, Y_{t-2}, \ldots\right) = E\left(\nabla Y_t / Y_{t-1}, Y_{t-2}, \ldots\right) = E\left(\varepsilon_t / Y_{t-1}, Y_{t-2}, \ldots\right) = 0$$

Al proceso

$$\left\{\nabla Y_t\right\} = \left\{Y_t - Y_{t-1}\right\} = \left\{\varepsilon_t\right\} \qquad (t = 2,3,\ldots)$$

se lo conoce como **diferencia de una matingala con respecto al conjunto de información** $\left\{Y_{t-1}, Y_{t-2}, \ldots\right\}$ y, entonces, se puede concluir que el error medio cuadrático mínimo de la predicción de un incremento futuro de una martingala es nulo[40].

Una extensión interesante del teorema central del límite permite demostrar que, dado un proceso $\left\{\nabla Y_t\right\} = \left\{\varepsilon_t\right\}$ diferencia de una martingala tal que: **i)** $\bar{\varepsilon}_{(n)} = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t$; **ii)** $E\left(\varepsilon_t^2\right) = \sigma^2\left(\varepsilon_t\right)$; **iii)** $\lim_{n\to\infty}\sum_{t=1}^{n}\sigma^2\left(\varepsilon_t\right) = \sigma_\varepsilon^2$; **iv)** $E\left(\left|\varepsilon_t\right|^s\right) < \infty \ (s = 3,4,\ldots; t = 1,2,\ldots)$ y **v)** $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 \xrightarrow{p} \sigma_\varepsilon^2$, la sucesión $\sqrt{n}\, Y_{(n)} \xrightarrow{d} N\left(0,\sigma^2\right)$ (ver White, H. (1984)).

Supóngase que el proceso $\left\{Y_t\right\}$, con respecto a la filtración $\left\{Y_{t-1}, Y_{t-2}, \ldots\right\}$, represente el valor de un juego al momento t . Entonces el proceso $\left\{\nabla Y_t\right\} = \left\{Y_t - Y_{t-1}\right\}$ puede ser interpretado como representativo de la ganancia neta del juego para una apuesta unitaria, entre los momentos $t-1$ y t . Luego, la mejor predicción de la ganancia neta ∇Y_t será:

$$E\left(\nabla Y_t / Y_{t-1}, Y_{t-2}, \ldots\right) = E\left(Y_t / Y_{t-1}, Y_{t-2}, \ldots\right) - Y_{t-1}$$

y, como se mencionó más arriba, si $\left\{Y_t\right\}$ es una martingala con respecto al conjunto de información $\left\{Y_{t-1}, Y_{t-2}, \ldots\right\}$, el segundo miembro de esta expresión es nulo (obsérvese que la condición que establece que la mejor predicción de las ganancias netas futuras, para apuestas unitarias es igual a cero, es la que caracteriza a un juego equitativo). Como corolario de esta propiedad se obtiene inmediatamente que:

$$E\left(Y_t / Y_{t-1}, Y_{t-2}, \ldots\right) = E\left[E\left(Y_{t+1} / Y_t, Y_{t-1}, \ldots\right) / Y_{t-1}, Y_{t-2}, \ldots\right] =$$

$$= E\left(Y_{t+1} / Y_{t-1}, Y_{t-2}, \ldots\right) =$$

$$= E\left[E\left(Y_{t+2} / Y_{t+1}, Y_t, \ldots\right) / Y_{t-1}, Y_{t-2}, \ldots\right] =$$

---

39. El RW y los procesos con incrementos independientes (ver Sec. 1.3) integran la familia de los procesos $AR(1)$. Contrariamente a lo que ocurre con estos dos procesos, la martingala considera una cota superior en los momentos de primer orden.

40. Una extensión de esta definición de martingala se obtiene considerando la siguiente restricción sobre los momentos de orden superior ($E\left(\left|Y_t\right|^s\right) < \infty$, para algún $s > 1$) que permite definir una **martingala de orden s**. En particular, para $s = 2$, se obtiene una **martingala de segundo orden** (o **martingala de cuadrado integrable**).

$$= E\left(Y_{t+2} / Y_{t-1}, Y_{t-2}, \ldots\right) = \ldots = E\left(Y_{t+k} / Y_{t-1}, Y_{t-2}, \ldots\right)$$

Es decir que, en una martingala, la esperanza matemática condicionada es constante[41].

Debe tenerse en cuenta que la definición de la martingala se concentra fundamentalmente en la esperanza matemática y no en su distribución de probabilidades. Esta condición determina la diferencia fundamental entre las hipótesis de martingala y de RW: en la primera (igual que en el RW-3), los incrementos $\nabla Y_t = \varepsilon_t$ son incorrelacionados pero no necesariamente estacionarios (un proceso RW-3 es una martingala con varianza constante y finita, es decir, es una martingala estacionaria)[42], por lo que admite que el precio de un activo puede ser representado utilizando un modelo basado en las autocorrelaciones positivas entre sus varianzas condicionadas[43].

En términos generales, se dice que el proceso $\{Y_t\}$ es una martingala con respecto a la filtración $\Omega_t(Y)$ si se cumplen las siguientes condiciones:

$$E\left(|Y_t|\right) < \infty$$

$$E\left(Y_t / \Omega_{t-1}(Y), \Omega_{t-2}(Y), \ldots, \Omega_0(Y)\right) = Y_{t-1}$$

donde $\Omega_t(Y) = \left\{\{Y_t\}, \{Z_t\}, \{W_t\}, \ldots\right\}$ $(t = 1,2,\ldots)$ denota una sucesión creciente, es decir, tal que $\Omega_1(Y) \subset \Omega_2(Y) \subset \Omega_3(Y) \subset \ldots$ (debe tenerse en cuenta que $\Omega_t(Y)$ constituye la estructura causal de $\{Y_t\}$ y que, por lo tanto, el proceso $\{Y_t\}$ puede ser considerado como una función de $\{\Omega_t(Y)\}$).

De la misma forma, se verifica que $E\left(Y_{t+s} / \Omega_0(Y), \Omega_1(Y), \ldots, \Omega_t(Y)\right) = Y_t$ $(s = 0,1,2,\ldots,t)$. Lo cual implica que la información disponible al momento t coincide con el estado del proceso al momento t [44]. Como corolario de esta condición se puede concluir en forma inmediata que todo incremento, $Y_{t+s} - Y_t$, es independiente de $\Omega_t(Y)$, es decir que el conjunto de información $\Omega_t(Y)$ no sirve para predecir el valor Y_{t+s} (en las aplicaciones referidas al comportamiento de los precios de un activo esta propiedad conduce a la hipótesis de eficiencia del mercado, la condición conocida como "risk neutrality" implica la propiedad

---

41. El no cumplimiento de esta condición permite decidir que un proceso $\{Y_t\}$ no es una martingala. Debe tenerse en cuenta que toda martingala es un proceso con valor esperado constante, pero que no todo proceso con esperanza matemática constante es una martingala; es decir que el cumplimiento de esta condición no permite asegurar que un proceso sea una martingala.

42. Todo proceso definido por una suma de variables aleatorias centradas independientes entre sí es una martingala (para un tratamiento más completo del concepto de martingala lineal, ver Nelson, P.I. (1970), McQueen, J.B. (1973)).

43. Como se verá más adelante, la martingala (en su versión continua en el dominio del tiempo), que implica ausencia de arbitraje (ver Sec. 4.7.1), resulta particularmente adecuada para representar el comportamiento de las rentabilidades de un activo y conduce a la fórmula de Black-Scholes-Merton (ver Black, F.; Scholes, M. (1973), Merton, R.C. (1973)) para la valuación de una opción sobre el mismo (por sus trabajos pioneros en la teoría de valuación de opciones, M. Scholes y R.C. Merton fueron galardonados, en 1977, con el premio Nobel de economía, F. Black había fallecido en 1995).

44. En la interpretación de la martingala en términos de la teoría de los juegos de azar, la variable Y_t , que representa la fortuna del jugador en el momento t del juego, depende además de la sucesión de fortunas pasadas $\left(Y_0, Y_1, \ldots, Y_t\right)$, de cualquier otra información adicional (por ejemplo, resultados de partidas en las cuales el jugador no efectuó apuestas).

de la martingala) [45].

Se dice que el proceso $\{Y_t\}$ es una **martingala inversa** (o **martingala "backwards"**) con respecto al conjunto de información $\Omega_t(Y)$ si se verifica que: **i)** $E(|Y_t| < \infty)$ y **ii)** $E(Y_t / \Omega_{t+1}(Y), \Omega_{t+2}(Y), ..., \Omega_{t+n}(Y)) = Y_{t+n}$, donde la sucesión $\{\Omega_t(Y)\}$ es decreciente (es decir, tal que $\Omega_{t+1}(Y) \supset \Omega_{t+2}(Y) \supset ... \supset \Omega_{t+n}(Y)$). De esta definición se puede concluir que la condición necesaria y suficiente para que el proceso $\{Y_t\}$ sea una martingala inversa con respecto al conjunto de información $\Omega_t(Y)$ es que $\{Y_{t-i+1}\}$ $(i = 1,2,...,n)$ sea una martingala. Esto demuestra que la teoría relacionada con las martingalas inversas es simétrica con respecto a la teoría de las martingalas finitas. Debe tenerse en cuenta que esta simetría no es generalizable en forma automática a la aplicación de los teoremas en el límite.

Sea $\{\nabla Y_t\} = \{Y_t - Y_{t-1}\} = \{\varepsilon_t\}$ un proceso diferencia de una martingala con respecto a su filtración natural $\{Y_{t-1}, Y_{t-2}, ...\}$, tal que $E(\varepsilon_t^2) < \infty$ (es decir, con varianza finita) y sea un proceso $\{C_t\}$ $(t = 1,2,...)$ predecible con respecto a $\{Y_{t-1}, Y_{t-2}, ...\}$ (es decir que su estado al momento t , C_t , se supone completamente conocido en el momento $t - 1$) y tal que $E(C_t^2) < \infty$, entonces, de acuerdo con la desigualdad de Cauchy-Schwarz, se verifica que el proceso $\{Z_t^*\} = \{[C \bullet \nabla Y]_t\} = \{[C \bullet \varepsilon]_t\} = \sum_{\tau=1}^{t} C_\tau \varepsilon_\tau$ es tal que[46]:

$$E(|Z_t^*|) \leq \sum_{\tau=1}^{n} E(|C_\tau \varepsilon_\tau|) \leq \sum_{\tau=1}^{n} \sqrt{E(C_t^2) E(\varepsilon_\tau^2)} < \infty$$

Si, además, se tiene en cuenta que $\{Z_t^*\}$ y $\{C_t\}$ son procesos predecibles con respecto al conjunto de información $\{\nabla Y_t, \nabla Y_{t-1}, ...\} = \{\varepsilon_t, \varepsilon_{t-1}, ...\}$, se puede concluir que:

$$E(Z_t^* - Z_{t-1}^* / \nabla Y_{t-1}, \nabla Y_{t-2}, ...) = E(\nabla Z_t^* / \nabla Y_{t-1}, \nabla Y_{t-2}, ...) =$$

$$= E(C_t \nabla Y_t / \nabla Y_{t-1}, \nabla Y_{t-2}, ...) =$$

$$= C_t E(\nabla Y_t / \nabla Y_{t-1}, \nabla Y_{t-2}, ...) = 0$$

Es decir, que el proceso transformado $\{\nabla Z_t^*\}$ es una diferencia de una martingala y $\{Z_t\}$ es una

---

[45]. La primera propuesta de relación entre el supuesto de eficiencia del mercado de capitales y las martingalas se debe a Samuelson, P.A. (1965). Este vínculo permite concluir que la aparente oposición entre los modelos fundamentalista (que supone un comportamiento completamente sistemático de los precios de los activos) y de eficiencia del mercado de capitales (que supone un comportamiento completamente aleatorio es ficticia). La propuesta de Samuelson implica que, si el modelo fundamentalista (que considera que los precios de un activo están dados por los flujos de caja descontados) es correcto, los precios futuros de un activo son impredecibles, igual que en la hipótesis de la martingala.

[46]. Ver Landro, A.H. (2002).

martingala con respecto al conjunto de información $\{Y_{t-1}, Y_{t-2}, ...\}$. Como se verá en la próxima sección, la transformación C_t para martingalas continuas en el dominio del tiempo es la **integral estocástica de Itô**.

Considerando nuevamente la interpretación de la martingala en términos de la teoría de los juegos de azar, sean: **i)** $\{\nabla Y_t\}$ un proceso que representa las ganancias netas de un juego para una apuesta unitaria y **ii)** $\{C_t\}$ un proceso que representa la apuesta en el $t-esimo$ juego, predecible con respecto al conjunto de información $\{Y_{t-1}, Y_{t-2}, ...\}$. La transformación $C_t(\nabla Y_t)$ representará, entonces, la ganancia neta del $t-esimo$ juego y $[C \bullet (\nabla Y)]_n = \sum_{t=1}^{n} C_t(\nabla Y_t)$ representará el total de las ganancias netas hasta el $n-esimo$ juego. Como el proceso $\{[C \bullet (\nabla Y)]_t\} = \{[C \bullet \varepsilon]_t\}$ es predecible con respecto al conjunto de información $\{Y_{t-1}, Y_{t-2}, ...\}$, se verifica que la condición $E(C_t(\nabla Y_t) / Y_{t-1}, Y_{t-2}, ...) = 0$ es la que caracteriza a un juego equitativo[47].

A partir de los resultados anteriores se pueden deducir las siguientes propiedades:

1) Dada una martingala de segundo orden, de acuerdo con el criterio de Kolmogorov, es posible hallar una sucesión $\{c_t\}$ de constantes positivas ($c_t \uparrow \infty$) tal que $\sum_{t=1}^{\infty} \frac{1}{c_t^2} E(\varepsilon_t^2) < \infty$. Lo cual implica que $\{\varepsilon_t\}$ se comporta como una sucesión de variables incorrelacionadas y, en consecuencia, que $\frac{1}{c_n} \sum_{t=1}^{n} \varepsilon_t \xrightarrow{ccc} 0$. Es decir que cumple con la ley fuerte de los grandes números [48]:

$$p\left[\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\right) = 0\right] = 1$$

2) Sea una martingala de primer orden tal que su diferencia $\{\nabla Y_t\} = \{\varepsilon_t\}$ sea de varianza finita. De

---

[47]. Igual que en el caso de las martingalas, la extensión de las restricciones sobre los momentos de orden superior, $E(|\nabla Y_t|^s) = E(|\varepsilon_t|^s) < \infty \ (s = 2,3,...)$, permite definir procesos diferencia de martingalas de orden s. En particular, para $s = 2$, se verificará que $E(|\nabla Y_t|^2) < \infty$ y que $\gamma(\nabla Y_t, \nabla Y_{t+j}) = \gamma(\varepsilon_t, \varepsilon_{t+j}) = E(\varepsilon_t \varepsilon_{t+j}) = 0 \ (j = 1,2,...; t = 1,2,...)$.

[48]. Esta propiedad es una consecuencia del teorema de convergencia de la martingala, el cual será tratado en las próximas páginas (ver Feller, W. (1971), Loeve, M. (1977)).

$$\lim_{n\to\infty}\left\{\frac{1}{\sigma_{(n)}^2}\sum_{t=1}^{n}E\left[\varepsilon_t^2 I_{\left(|\varepsilon|>\varepsilon\sigma_{(n)}\right)}\right]\right\} = 0$$

(donde: $\varepsilon > 0$ y $\sigma_{(n)}^2 = \sigma^2\left(\sum_{t=1}^{n}\varepsilon_t\right)$) y **iii)** $E\left\{E\left[\varepsilon_t^2 / \Omega_{t-1}(Y),\Omega_{t-2}(Y),...\right]\right\} = \sigma_t^2$ $(t = 1,2,...)$, cumple los postulados del teorema central del límite[50]:

$$\lim_{n\to\infty} p\left[\frac{1}{\sigma_{(n)}}\sum_{t=1}^{n}\varepsilon_t \le u\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{y^2/2}dy = F_N(u)\ .$$

6) Sea una martingala de cuadrado integrable tal que su diferencia sea un proceso estacionario de segundo orden, ergódico. Entonces, se verifica que $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t = \sigma_\varepsilon\varsigma(n)\sqrt{\frac{2}{n}\ln\left[\ln(n)\right]}$ (donde $\limsup_{n\to\infty}\varsigma(n) = 1$ y $\liminf_{n\to\infty}\varsigma(n) = -1$). Este resultado (debido a Stout, W.F. (1970)(1974)) contiene como caso particular a la ley clásica del logaritmo iterado para sumas de variables aleatorias iid (debida a Hartman, P.; Winter, A. (1941)).

Si se verifica que $E\left(Y_t - Y_{t-1} / Y_{t-1},Y_{t-2},...\right) > 0$, es decir, si $E\left(Y_t / Y_{t-1},Y_{t-2},...\right) > Y_{t-1}$, entonces, se dice que el proceso $\left\{Y_t\right\}$ es una **submartingala** y, si $E\left(Y_t - Y_{t-1} / Y_{t-1},Y_{t-2},...\right) < 0$, entonces se dice que el proceso $\left\{Y_t\right\}$ es una **supermartingala.**

En términos más generales, se dice que el proceso $\left\{Y_t\right\}$ es una **supermartingala con respecto al conjunto de información** $\left\{\Omega_t(Y)\right\}$, si se verifica que:

$$E\left(Y_t^{(-)}\right) < \infty \qquad \left(Y_t^{(-)} = \min\left(Y_t,0\right)\right)$$

$$E\left(Y_{t+1} / \Omega_0(Y),\Omega_1(Y),...,\Omega_t(Y)\right) \le Y_t$$

De la misma forma, si se verifica que:

---

[50]. Los primeros resultados relacionados con esta generalización del teorema central del límite (debidos a Bernstein, S. (1928) y Lévy, P.P. (1935)(1935)(1937)) suponían la condición de que la variable condicionada (introducida por Lévy) fuera tal que $V_t^2 = \sum_{\tau=1}^{t}E\left[\varepsilon_\tau^2 / \Omega_{\tau-1}(Y),\Omega_{\tau-2}(Y),...\right] =$ *cte.* Doob, J.L. (1953) desarrolló una demostración del resultado de Lévy, P.P. (1937), utilizando el concepto de función característica. Billingsley, P. (1961) e Ibragimov, I.A. (1963) obtuvieron, en forma independiente, una demostración del teorema central del límite para martingalas con diferencias estacionarias y ergódicas (es decir, con diferencias con varianza condicionada asintóticamente constante, $\frac{1}{\sigma_n^2}\sum_{t=1}^{n}E\left(\varepsilon_t^2 / \Omega_{t-1}(Y),\Omega_{t-2}(Y),...\right) = \frac{V_n^2}{E\left(V_n^2\right)} = \frac{V_n^2}{E\left(Y_n^2\right)} \xrightarrow{p} 1$).

Posteriores gneralizaciones fueron obtenidas por Rosén, P. (1967)(1967)(1967), Dvoretzky, A. (1969)(1971)(1972), Loynes, R.M. (1969)(1970), Bergström, H. (1970), Brown, B.M. (1971), Gänsler, P.; Strobel, J.; Stute, W. (1978).

$$E\left(Y_t^{(+)}\right) < \infty \qquad \left(Y_t^{(-)} = \max(0, Y_t)\right)$$

$$E\left(Y_{t+1} / \Omega_0(Y), \Omega_1(Y), ..., \Omega_t(Y)\right) \geq Y_t$$

se dice que $\{Y_t\}$ es una **submartingala con respecto al conjunto de información** $\{\Omega_t(Y)\}$.

Como corolario de estas definiciones se puede concluir que: **i)** la condición necesaria y suficiente para que $\{Y_t\}$ sea una supermartingala con respecto a la filtración $\{\Omega_t(Y)\}$ es que $\{-Y_t\}$ sea una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$; **ii)** la condición necesaria y suficiente para que $\{Y_t\}$ sea una martingala con respecto a la filtación $\{\Omega_t(Y)\}$ es que $\{Y_t\}$ sea, simultáneamente, una submartingala y una supermartingala con respecto a la filtración $\{\Omega_t(Y)\}$.

Algunas propiedades básicas de las submartingalas particularmente útiles en las aplicaciones, son las siguientes:

1) Sean: **i)** $\{Y_t\}$ una submartingala con respecto a la filtración $\Omega_t(Y)$ y **ii)** una función f no-decreciente y convexa y tal que $E\left(\left|f\left(\sum_{t=1}^{n} \nabla Y_t\right)\right|\right) = E\left(\left|f\left(\sum_{t=1}^{n} \varepsilon_t\right)\right|\right) < \infty \ (n = 1,2,...)$, entonces el proceso $\left\{f\left(\sum_{t=1}^{n} \varepsilon_t\right)\right\}$ es una martingala[51].

2) Sea $\{Y_t\}$ una submartingala con respecto a la filtración $\Omega_t(Y)$, entonces el proceso $\left\{Max\left(\sum_{t=1}^{n} \varepsilon_t ; 0\right)\right\}$ también es una submartingala con respecto a la filtración $\Omega_t(Y)$.

3) Sea $\{Y_t\}$ una martingala con respecto a la filtración $\Omega_t(Y)$, tal que $E\left(\left|\sum_{t=1}^{n} \nabla Y_t\right|^P\right) = E\left(\left|\sum_{t=1}^{n} \varepsilon_t\right|^P\right) < \infty$ $(n = 1,2,...; p \geq 1)$, entonces el proceso $\left\{\left|\sum_{t=1}^{n} \varepsilon_t\right|\right\}$ es una submartingala con respecto a la filtración $\Omega_t(Y)$.

Con un razonamiento similar al empleado en las martingalas de parámetro discreto, se dice que

---

[51]. Si el proceso $\{Y_t\}$ es una martingala con respecto al conjunto de información $\Omega_t(Y)$, no es condición necesaria para el cumplimiento de esta propiedad que la función f sea no-decreciente.

un proceso $\{Y(t)\}$ $(t \geq 1)$ es una **martingala continua en el dominio del tiempo con respecto a su filtración natural** $\{Y(\tau)\}$ $(\tau < t)$, si se verifica que:

$$E\left[Y(t_n) / Y(t_m), Y(t_\ell), \ldots\right] = Y(t_m)$$

y, en general, se dice que $\{Y(t)\}$ es una **martingala continua en el dominio del tiempo con respecto a la filtración** $\Omega(t)$, si se verifica que:

$$E\left[Y(t_n) / \Omega\left(Y(t_m)\right), \Omega\left(Y(t_\ell)\right), \ldots\right] = Y(t_m)$$

Sea $\{Y(t); t \geq 0\}$ un proceso estocástico de parámetro continuo: Si se verifica que: **i)** $E\left[Y^{(+)}(t)\right] < \infty$ $\left(Y^{(+)} = \max(0, Y_t)\right)$ y **ii)** $E\left[Y(t_n) / \Omega\left(Y(t_m)\right), \Omega\left(Y(t_\ell)\right), \ldots\right] \geq Y(t_m)$ $(t_n \geq t_m \geq t_\ell \geq \ldots \geq 0)$, se dice que $\{Y(t)\}$ es una **submartingala continua en el dominio del tiempo con respecto a la filtración** $\{\Omega(Y(t))\}$.

De la misma forma, se puede concluir que: **i)** $\{Y(t)\}$ es una supermartingala continua en el dominio del tiempo con respecto a la filtración $\{\Omega(Y(t))\}$ si se verifica que $\{-Y(t)\}$ es una **submartingala continua en el dominio del tiempocon respecto a la filtración** $\{\Omega(Y(t))\}$ y **ii)** es una **martingala continua en el dominio del tiempo con respecto a la filtración** $\{\Omega(Y(t))\}$, si es a la vez una supermartingala y una submartingala con respecto a dicha filtración.

El fundamento de la teoría límite de las martingalas se desarrolló a partir de la obra de Bernstein, S. (1927)(1939)(1940)(1941) y Lévy, P.P. (1935)(1935)(1937), como una generalización de los teoremas límite para sumas de variables aleatorias independientes. Esta tendencia cambió radicalmente con la publicación del **teorema de convergencia de la martingala**, debido a Doob, J.L. (1953)[52], según el cual, dada una submartingala con respecto al conjunto de información $\Omega_t(Y)$ ($\{Y_t / \Omega_t(Y)\}$), entonces existe una variable Y tal que: **i)** $Y_t \xrightarrow{ccc} Y$ y **ii)** $E(Y) \leq \liminf_{t \to \infty} E(|Y_t|) < \infty$. Asimismo, si la submartingala $\{Y_t / \Omega_t(Y)\}$ tiene varianza finita, entonces $Y_t \xrightarrow{m_2} Y$ (debe tenerse en cuenta que éste es un teorema de existencia que no permite obtener ninguna conclusión acerca de las características de la variable límite, excepto que posee momentos de primero y segundo orden finitos).

4.4.- La mixingala

La martingala pertenece a una clase más general de procesos (propuesta por Andrews, D.W.K.

---

[52]. Esta obra ejerció gran influencia sobre los desarrollos relacionados con los procesos estocásticos en los siguientes 30 años. En los últimos años del siglo XX se produjo un resurgimiento de la teoría límite de la martingala como una generalización de los resultados referidos a la suma de variables independientes.

(1988)) conocida como **mixingala**. Dado un proceso $\{Y_t\}$ con $E(Y_t) = 0\ (t = 1,2,...)$, se dice que es una **$L^1$-mixingala con respecto a una filtración** $\Omega_{(t)}(Y) = \{\Omega_t(Y),\Omega_{t-1}(Y),...,\Omega_1(Y)\}$, si se verifica que:

$$E\left(\left|E\left[Y_t / \Omega_{C,(t-j)}(Y)\right]\right|\right) \leq c_t\xi_j \qquad (t = 1,2,...;j = 0,1,...)$$

(donde $\{c_t\}\ (t = 1,2,...)$ y $\{\xi_j\}\ (j = 0,1,2,...)$ denotan sendas sucesiones de constantes determinísticas tales que: **i)** $c_t \geq 0\ (t = 1,2,...)$; **ii)** $\xi_j \geq 0\ (j = 0,1,2,...)$ y **iii)** $\lim_{j\to\infty}\xi_j = 0$). En otros términos, se puede decir que todo proceso $\{Y_t\}$ con valor esperado nulo, tal que su predictor $E\left(Y_t / \Omega_{C(t-j)}(Y)\right)$ converge en valor absoluto a su valor esperado no-condicionado ($E(Y_t) = 0$), es una L^1-mixingala.

A partir de la interpretación de la mixingala como una martingala asintótica que, como el proceso diferencia de una martingala, satisface el teorema de convergencia[53], Andrews, D.W.K. (1988) desarrolló una ley de los grandes números para L^1-mixingalas, a partir de la condición de un proceso de ser **uniformemente integrable.**

Se dice que un proceso $\{Y_t\}$ es uniformemente integrable: **i)** si, para todo $\varepsilon > 0$ y arbitrariamente pequeño, existe una constante c tal que $E\left[|Y_t|\delta_{(|Y_t|\geq c)}\right] < \varepsilon\ (t = 1,2,...)$, donde:

$$\delta_{(|Y_t|\geq c)} = \begin{cases} 1 & si \qquad |Y_t| \geq c \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y **ii)** si existe un $s > 0$ y un $K < \infty$ tales que $E\left(|Y_t|^s\right) < K\ (t = 1,2,...)$. Por otra parte, si existe un $s > 0$ y un $K < \infty$ tales que $E\left(|Y_t|^s\right) < K\ (t = 1,2,...)$ y una sucesión $\{h_j\}\ (j = 0,\pm 1,\pm 2,...)$ tal que $\sum_{j=-\infty}^{\infty}|h_j| < \infty$, entonces se puede asegurar que el proceso $Z_t = \sum_{j=-\infty}^{\infty}h_jY_{t-j}$ es uniformemente integrable.

A partir de estos resultados, se demuestra que: **i)** si $\{Y_t\}$ es uniformemente integrable y **ii)** si existe una sucesión tal que $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}c_t < \infty$, entonces se verifica que $\overline{Y}_{(n)} = \frac{1}{n}\sum_{t=1}^{n}Y_t \xrightarrow{p} 0$ y, por lo tanto, que $\{Y_t\}$ es L^1-mixingala. Obsérvese que, como se comentó en la Sec. 1.5, esta condición, necesaria para asegurar la ergodicidad de $\{Y_t\}$, es alternativa a la condición $\sum_{j=0}^{\infty}|\gamma_j(Y)| < \infty$ (ver McLeish, D.L. (1977)) y debe su denominación como **condición de la mixingala** a la combinación de los conceptos de

---

[53]. Ver Sec. 4.3.

martingala y de condiciones "mixing" de dependencia en el dominio del tiempo de sucesiones de variables aleatorias.

Como casos particulares de las condiciones de la mixingala:

1) Sea $\{\nabla Y_t\}$ un **proceso diferencia de una mixingala**. Entonces, para $c_t = E(|Y_t|)$ y:

$$\xi_j = \begin{cases} 1 & para \quad j = 0 \\ 0 & para \quad j = 1,2,... \end{cases}$$

se puede concluir que $\{Y_t\}$ es una L^1-mixingala con respeto a la filtración natural $\{\Omega_t(Y)\} = \{Y_t, Y_{t-1},..., Y_1\}$, es decir, se puede concluir que $E\left(\left|E\left(Y_t / \Omega_{C,t-j}(Y)\right)\right|\right) \leq E(|Y_t|)\ (t = 1,2,...)$. Si, además, se verifica que $E\left(|Y_t|^s\right) < K$ (para algún $s > 1$ y $K < \infty$), entonces se puede asegurar que el proceso $\{Y_t\}$ es uniformemente integrable y puede ser interpretado como una L^1-mixingala con $c_t = K$, de modo que $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} c_t = K < \infty$ y, por lo tanto, se puede concluir que $\overline{Y}_{(n)} = \frac{1}{n}\sum_{t=1}^{n} Y_t \xrightarrow{p} 0$.

2) Sea un proceso $\{Y_t\}$ (tal que $E(Y_t) = 0\ (t = 1,2,...)$) que admite una representación de la forma $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, donde: **i)** $\{\varepsilon_t\}$ denota un proceso diferencia de una martingala con $E(|\varepsilon_t|) < K$ (para $t = 1,2,...$ y $K < \infty$) y **ii)** la sucesión de ponderaciones $\{\psi_j\}$ es tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Entonces se verifica que:

$$E\left\{\left|E\left(Y_t / \varepsilon_{t-j}, \varepsilon_{t-j-1},...\right)\right|\right\} = E\left(\left|\sum_{h=j}^{\infty} \psi_h \varepsilon_{t-h}\right|\right) \leq E\left(\sum_{h=j}^{\infty} |\psi_h \varepsilon_{t-h}|\right)$$

y, dado que: **i)** la sucesión $\{\psi_j\}$ es absolutamente sumable y **ii)** $E(|\varepsilon_{t-j}|) < K$, se puede escribir:

$$E\left(\sum_{h=j}^{\infty} |\psi_h \varepsilon_{t-h}|\right) = \sum_{h=j}^{\infty} |\psi_h| E(\varepsilon t - h) \leq K \sum_{h=j}^{\infty} |\psi_h|$$

Luego, para: **i)** $c_t = K$ y **ii)** $\xi_j = \sum_{h=j}^{\infty} |\psi_h| < \infty$, se verifica que $\lim_{j\to\infty} \xi_j = 0$. De modo que se puede concluir que $\{Y_t\}$:L^1-mixingala. Si además se verifica que $E\left(|\varepsilon_t|^s\right) < K < \infty\ (s = 2,3,...)$, se puede

asegurar que $\{Y_t\}$ es un proceso uniformemente integrable y que $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}c_t = K < \infty$. De modo que se puede concluir que $\overline{Y}_{(n)} \xrightarrow{p} 0$.

4.5.- El movimiento Browniano

Sea un proceso $\{Y_t\}: RW-1$ que admite una representación de la forma:

$$Y_t = Y_{t-1} + \varepsilon_t$$

donde el proceso $\{\varepsilon_t\} = \{\nabla Y_t\}$ está formado por variables que pueden asumir sólo dos valores: η, con probabilidad p y $-\eta$, con probabilidad $q = 1-p$.

Dividiendo el intervalo $[t-1,t]$ en k subintervalos de longitud $h = \frac{1}{k}$, queda definido un proceso $Y_k(t) = Y_{[t/h]} = Y_{[kt/T]}\ (t \in [0,T])$ (donde $[x]$ denota el valor del mayor entero menor o igual que x) que es continuo en el dominio del tiempo, constante excepto en los puntos $t = jh\ (j = 1,2,...,k)$ y que, por lo tanto (dado que sólo varía en puntos discretos del tiempo), puede ser considerado como equivalente al proceso $\{Y_t\}$ de parámetro discreto definido más arriba, con medidas $E[Y_k(t)] = k(p-q)\eta$ y $\sigma^2[Y_k(t)] = 4kpq\eta^2$.

Para $k = 1$:

$$E(Y_1(t)) = -\eta q + \eta p = (p-q)\eta, m_2(Y_1(t)) = \eta^2 q + \eta^2 p = (p+q)\eta^2$$

y, por lo tanto:

$$\sigma^2(Y_1(t)) = (p+q)\eta^2 - (p-q)^2\eta^2 = 4pq\eta^2$$

De modo que la variable $Y_k(t)$ puede ser considerada como la suma de k variables $Y_1(t)$ iid.

De estas relaciones se puede concluir en forma inmediata que, cuando k aumenta indefinidamente (es decir cuando $h \to 0$), $\lim_{k\to\infty} E[Y_k(t)] = \lim_{k\to\infty} \sigma^2[Y_k(t)] = \infty$. Luego, si se quiere obtener, en el límite, un proceso continuo en el dominio del tiempo, no-degenerado, sus momentos deben mantenerse finitos y si, en particular, se quiere obtener un proceso equivalente al RW-1 definido al comienzo de esta sección, en el límite, estos momentos deben ser lineales con respecto a t (ver Sec. 4.1):

$$\lim_{k\to\infty} k(p-q)\eta = \lim_{h\to 0}\frac{t}{h}(p-q)\eta = \varphi_0 t$$

$$\lim_{k\to\infty} 4kpq\eta^2 = \lim_{h\to 0}\frac{t}{h}4pq\eta^2 = t\sigma_\varepsilon^2$$

Es decir, se debe verificar que $p = \frac{1}{2}\left(1 + \frac{\varphi_0\sqrt{h}}{\sigma_\varepsilon}\right)$; $q = 1 - p = \frac{1}{2}\left(1 - \frac{\varphi_0\sqrt{h}}{\sigma_\varepsilon}\right)$ y $\eta = \sigma_\varepsilon\sqrt{h}$.

De acuerdo con los postulados del **teorema central del límite funcional**[54], estas condiciones implican que, cuando $k \to \infty$: **i)** la longitud de cada subintervalo -obviamente- tiende a cero, pero a una tasa $\sqrt{h} = \frac{1}{\sqrt{k}}$; **ii)** las probabilidades p y q convergen a $\frac{1}{2}$ también a una tasa $\sqrt{h}$, de modo que se puede escribir $p = q = \frac{1}{2} + O\left(\sqrt{h}\right)$ y $\eta = O(h)$ (donde $O\left(\sqrt{h}\right)$ denota una función de orden asintótico igual a $\sqrt{h}$) y **iii)** el proceso definido por $Y_k^*(t) = \frac{Y_{(k)} - km(Y)}{\sqrt{k}\sigma_Y}\left(t = \frac{i}{k}; i = 0,1,2,\ldots,k; Y_{(0)} = 0\right)$ (donde $Y_{(k)} = Y_1 + Y_2 + \ldots + Y_k$) converge en distribución a una función $N(0;1)$:

$$\lim_{k\to\infty} f_{Y_{(k)}^*(t_1),\ldots,Y_{(k)}^*(t_n)}\left(y_{(k)}^*(t_1),\ldots,y_{(k)}^*(t_n)\right) = f^{(N)}_{Y^*(t_1),\ldots,Y^*(t_n)}\left(y^*(t_1),\ldots,y^*(t_n)\right) =$$

$$= f^{(N)}_{Y^*(t_1)}\left(y^*(t_1)\right)\prod_{j=2}^{n} f_{Y^*(t_j)-Y^*(t_{j-1})}\left(y^*(t_j) - y^*(t_{j-1})\right) =$$

$$= \frac{\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[\frac{y^{*2}(t_1)}{t_1} + \frac{\left(y^*(t_2) - y^*(t_1)\right)^2}{t_2 - t_1} + \ldots + \frac{\left(y^*(t_n) - y^*(t_{n-1})\right)^2}{t_n - t_{n-1}}\right]\right\}}{\sqrt{\left(2\pi\sigma_\varepsilon^2\right)^{n+1} t_1(t_2 - t_1)\ldots(t_n - t_{n-1})}} \qquad \left(0 \le t_1 < t_2 < \ldots < t_n \le T\right)$$

donde:

$$f_{Y^*(t_j)-Y^*(t_{j-1})}\left(y^*(t_j) - y^*(t_{j-1})\right) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi(t_j - t_{j-1})}}\exp\left\{-\frac{\left[y^*(t_j) - y^*(t_{j-1})\right]^2}{2\sigma_\varepsilon^2(t_j - t_{j-1})}\right\} .$$

Es decir, cuando $k \to \infty$ (o, lo que es lo mismo, cuando $h \to 0$) y $\left\{\nabla Y_k^*(t)\right\} \to 0$, el proceso $\left\{Y_k^*(t)\right\}$ (definido a partir de las sumas parciales) posee las siguientes propiedades: **i)** $Y_k^*(0) = 0$; **ii)** el proceso $\left\{Y_k^*\left(\frac{i_j}{k}\right) - Y_k^*\left(\frac{i_{j+1}}{k}\right)\right\}$, para todos los enteros tales que $0 \le i_1 \le i_2 \le \ldots \le i_m \le n$, es de incrementos

---

54. Este fascinante teorema constituye una extensión del teorema central del límite. Su primera expresión (conocida como **principio de invariancia**) se basa en el desarrollo de Erdös, P.; Kac, M. (1946)(1947) según el cual la solución al problema del límite de funciones de sumas puede ser obtenida flexibilizando el límite cuando las sumas están distribuidas de alguna forma particular. La primera forma general del principio de invariancia se debe a Donsker, M. (1951)(1952) y Prokhorov, Y.V. (1956). Como se mencionó en la Sec. 4.3, a partir de un trabajo de Rosén, B. (1967), Billingsley, P. (1968) extendió este principio a martingalas con incrementos estacionarios y ergódicos (los principios de invariancia de las martingalas fueron estudiados con gran detalle por Drogin, R. (1972), Rootzén, H. (1977)(1980), Gänssler, P.; Häusler, E. (1980), Hall, P.; Heyde, C.C. (1980)). Para más detalles sobre las generalizaciones posteriores al ámbito de la teoría de las martingalas de los alcances de las sumas de variables aleatorias independientes, ver Breiman, L. (1968), Csörgö, M. (1968), Brown, B.M.; Eagleson, .K. (1971), Dhrymes, P.J. (1998), Phillips, P.C.B. (1987), Stock, J.H.; Watson, M.W. (1988).

independientes y **iii)** para todo $0 \le i \le n$, se verifica que $\left\{Y_k^*\left(\frac{i}{k}\right)\right\}: N\left(0, \frac{i}{k}\right)$. Estas condiciones caracterizan al proceso conocido como **movimiento Browniano aritmético o de Wiener**[55] (que se denota habitualmente como $\{B(t)\}$), de modo que se puede escribir $Y_k^*(t) \xrightarrow{d} B(t)$ en el intervalo $[0;1]$. Sus momentos condicionados son de la forma[56]:

$$E\left[Y^*(t) / Y^*(t_0)\right] = Y^*(t_0)$$

$$\sigma^2\left[Y^*(t) / Y^*(t_0)\right] = (t - t_0)\sigma_\varepsilon^2$$

y las covarianzas:

$$\gamma\left[Y^*(t_1), Y^*(t_2)\right] = \gamma\left[Y^*(t_1), Y^*(t_2) - Y^*(t_1) + Y^*(t_1)\right] =$$

$$= \gamma\left[Y^*(t_1), Y^*(t_2) - Y^*(t_1)\right] + \gamma\left[Y^*(t_1), Y^*(t_1)\right] = \sigma^2\left[Y^*(t_1)\right] = t_1\sigma_\varepsilon^2$$

En general el proceso $\{Y(t)\}$ obtenido como límite de un proceso RW con tendencia, puede ser

---

55. En 1827, el británico Robert Brown definió este proceso como representativo del comportamiento de las partículas de polen suspendidas en un fluido. Hasta ese momento se creía que el comportamiento aparentemente aleatorio de las partículas se debía a la acción de moléculas vivas. En 1870, I. Carbonelle sugirió que el movimiento era causado por el impacto con las partículas vecinas. Brown demostró que ese comportamiento dinámico estaba motivado por el bombardeo sufrido por cada partícula de millones de moléculas del fluido, debido a la acción de la difusión térmica. Posteriormente, Einstein, A. (1905), basándose en los principios de la mecánica estadística que había desarrollado en sus memorias de (1902)(1903) y (1904), propuso el primer intento sistemático para la estimación de la función de densidad de probabilidades de la variable representativa del comportamiento de una partícula, a partir de una simple ecuación diferencial de la forma $\frac{\partial f(y,t)}{\partial t} = = \delta\left(\frac{\partial^2 f(y,t)}{\partial^2 y}\right)$, donde: **i)** $\delta = \lim_{\Delta t \to 0}\left[\frac{(\Delta y)^2}{2\Delta t}\right]$ denota el coeficiente de difusión y representa la velocidad promedio del desplazamiento medio cuadrático y **ii)** $f_{Y(t)}(y,t) = \frac{n}{\sqrt{4\delta\pi}} \frac{\exp\left(-y^2 / 4\delta t\right)}{\sqrt{t}}$.

Por su parte, von Smoluchowski, M. (1906) desarrolló en forma independiente una conceptualización teórica de carácter netamente probabilístico, basándose en los siguientes supuestos simplificadores: **i)** la velocidad de las partículas es constante; **ii)** cada colisión produce una pequeña variación aleatoria de dirección en la trayectoria de las partículas y **iii)** los ángulos de esas variaciones se suponen que son de la misma medida. Suponiendo la condición inicial $Y(0) = 0$, la solución de esta ecuación (para valores de t suficientemente grandes) demuestra que $Y(t): N(0; 2\delta t)$. A comienzos de la década de 1920 Wiener, N. (1923)(1924) logró la primera expresión matemáticamente formal de este proceso y demostró la posibilidad de asignar probabilidades en el modelo de Einstein. Pero, hasta la publicación de Kolmogorov, A.N. (1933), no se pudo lograr una teoría de la probabilidad consistente que comprendiera a este modelo (ver Jerison, D.; Singer, I.; Stroock, D. (eds.) (1996)). Los trabajos de Wiener fueron continuados por Lévy, P. (1931)(1934)(1934)(1935)(1935) (1937) (para una revisión histórica del desarrollo del movimiento Browniano, ver Haas-Lorentz, G.L. (1913), Brush, S. (1976)). Bachelier, L. (1900) (introduciendo el concepto de lo que posteriormente se conoció como **dependencia Markoviana**) se basó en este proceso RW-continuo para desarrollar el primer modelo estocástico para estudiar el comportamiento de los precios de un activo. Pero sus resultados (que figuran en Bachelier, L. (1912) y que no fueron aceptados por matemáticos como H. Poincaré), permanecieron ignorados hasta su mención -en términos poco favorecedores- por Kolmogorov, A.N. (1931) y su redescubrimiento por Samuelson en la biblioteca de la Universidad de Harvard. Posteriormente -en general- todos los desarrollos dirigidos a aplicaciones financieras consideraron a los precios (y, por lo tanto, a las rentabilidades) como procesos continuos en el dominio del tiempo (con notables excepciones como la teoría del equilibrio de Rubinstein, M. (1976) y la fórmula de valuación de opciones de Cox, J.; Ross, S.; Rubinstein, M. (1979)). Los ejemplos más representativos de la adopción del paradigma del continuo para explicar el comportamiento de las rentabilidades de un activo, son los trabajos de Samuelson, P.A. (1965)(1965), el desarrollo de R. Merton (ver Merton, R. (1969)(1972)(1973)(1973)(1990)) y la fórmula de valuación de opciones de Black-Scholes (ver Black, F.; Scholes, M. (1972)(1973)) (debe tenerse en cuenta que fue recién después de los trabajos de Merton, Black y Scholes que esta teoría alcanzó niveles avanzados).

56. Para un mayor detalle acerca de la convergencia de un proceso discreto en el dominio del tiempo a procesos continuos en el tiempo, ver Cox, J.C.; Ross, S.A.; Rubinstein, M. (1979), Willinger, W.; Taqqu, M. (1991), Duffie, D.; Protter, P. (1992), Amin, K.; Khanna, A. (1994).

representado de la siguiente forma[57]:

$$Y(t) = t\varphi_0 + \sigma_\varepsilon B(t)$$

y es conocido como **movimiento Browniano con tendencia lineal** (donde $t\varphi_0$ define la tendencia determinística del proceso)[58]. Sus momentos condicionados son de la forma:

$$E\left[Y^*(t)/Y^*(t_0)\right] = Y^*(t_0) + \varphi_0(t - t_0)$$

$$\sigma^2\left[Y^*(t)/Y^*(t_0)\right] = (t - t_0)\sigma_\varepsilon^2$$

e, igual que en el caso precedente, las covarianzas son lineales con respecto al tiempo y están definidas por $\gamma\left[Y(t_1), Y(t_2)\right] = \sigma^2\left[Y(t_1)\right] = t_1\sigma_\varepsilon^2 \quad (t_1 < t_2)$.

Como corolario de esta última definición, es posible hallar en la literatura la siguiente expresión para la función de autocorrelaciones $\rho\left[Y(t_1, Y(t_2))\right] = \frac{1}{\sqrt{t_1 t_2}}$. A partir de este resultado se puede concluir que, para valores suficientemente grandes de t_1 e independientemente de la distancia $(t_1 - t_2)$, se verifica que $\lim_{t_1 \to \infty} \rho\left[Y(t_1), Y(t_2)\right] = 1$. Pero debe tenerse en cuenta que, en este caso el hecho que $t_1 \to \infty$ no permite obtener ninguna conclusión conceptualmente interesante (obsérvese que t_1 no está relacionado con la longitud de la serie cronológica). Una expresión alternativa podría ser $\rho\left[Y(t_1), Y(t_2)\right] = \sqrt{\frac{\min(t_1, t_2)}{\max(t_1, t_2)}}$.

Este movimiento Browniano cumple las siguientes condiciones:

i) Sean t_1 y t_2 dos puntos en el tiempo tales que $0 \le t_1 < t_2$, entonces se verifica que $B(t_2) - B(t_1) : N\left[(t_2 - t_1)\varphi_0, (t_2 - t_1)\right]$, es decir, que $B(t_2) - B(t_1) \overset{d}{=} B(t_2 - t_1)$.

ii) Como corolario de la definición de la función de autocovarianzas, para $0 \le t_1 < t_2 < t_3 < t_4$, se puede

---

[57]. A partir del principio de que toda función periódica, con período 2π, admite una representación en series de Fourier y teniendo en cuenta que las trayectorias del movimiento Browniano son funciones continuas, se puede concluir que estas trayectorias admiten una representación como un desarrollo en serie de funciones trigonométricas. Ahora bien, las trayectorias son funciones aleatorias y Normales, de modo que los coeficientes de la serie de Fourier deben ser variables aleatorias Normales. De las infinitas representaciones posibles de un movimiento Browniano merecen ser mencionadas la **de Paley-Wiener** y la **de Lévy**. Ambas pueden ser consideradas como casos particulares de la **representación de Lévy-Ciesielski**. Ciesielski, Z. (1965) demostró que el movimiento Browniano, en el intervalo [0,1], puede ser representado de la siguiente forma $B(t,w) = \sum_{j=1}^{\infty} f_j(w) \int_0^t \varphi_j(x)dx \quad (t \in [0;1])$. Donde $f_j(\bullet) : N(0;1)$ y $\{\varphi_j\}$ denota un sistema completo de funciones normales independientes entre sí, en el intervalo $[0;1]$.

[58]. El planteo probabilístico del movimiento Browniano desarrollado por von Smoluchowski puede ser considerado como una solución al problema, conocido como de los "vuelos aleatorios", propuesto por Rayleigh en 1880 y recreado por Pearson, K. (1905) (para una análisis más detallado sobre el origen de la teoría de los vuelos aleatorios, ver Dutka, J. (1985)), que consistía en la determinación de la probabilidad de que la distancia recorrida por el proceso sobre un plano en n pasos de igual longitud esté incluida en un intervalo dado, suponiendo que el ángulo formado por la trayectoria al cabo de cada paso fuera arbitrario. Debe tenerse en cuenta que el movimiento Browniano de von Smoluchowski, el "random walk" de Pearson y el "random flight" de Rayleigh eran procesos de Markov continuos en el dominio de las variables y discretos en el dominio del tiempo.

escribir:

$$E\left\{\left[Y(t_2) - Y(t_1)\right]\left[Y(t_4) - Y(t_3)\right]\right\} =$$

$$= \left\{\gamma\left[Y(t_2), Y(t_4)\right] - \gamma\left[Y(t_2), Y(t_3)\right]\right\} - \left\{\gamma\left[Y(t_1), Y(t_4)\right] - \gamma\left[Y(t_1), Y(t_3)\right]\right\} =$$

$$= (t_2 - t_2) - (t_1 - t_1) = 0$$

lo que demuestra que los incrementos de $\{B(t)\}$ para los intervalos $[t_1, t_2]$ y $[t_3, t_4]$ $(0 \leq t_1 < t_2 \leq t_3 < t_4)$ son independientes entre sí[59], de modo que la variable $B(t_2)$ no se ve afectada por las realizaciones de $B(t)$, para $t < t_1$. Esta propiedad, conocida como **condición de Markov** o de autorregresividad de primer orden (como se verá en la Sec. 4.7.1, la condición de Markov supone la eficiencia débil del mercado) implica la relación $B(t_1 + t_2) = B(t_1) + \left[B(t_1 + t_2) - B(t_1)\right]$ según la cual el estado del proceso al momento $(t_1 + t_2)$ estará dado por el estado del proceso al momento t_1 ($B(t_1)$) más una variable Normal ($B(t_1 + t_2) - B(t_1)$) independiente[60].

iii) Sus trayectorias son continuas (es decir, no presentan "saltos") casi con certeza, con respecto a una función de probabilidades dada.

iv) Es un proceso Normal. **v)** $\left\{\sqrt{c}B\left(\frac{t}{c}\right)\right\} \overset{d}{=} \{B(t)\}$.

Como corolario de estas propiedades (que son mucho más fuertes que la condición que supone que las variables $Y(t)$ son Normales con varianza $t\sigma_\varepsilon^2$ $(t \in [0, T])$) se puede concluir que: **i)** todo proceso con incrementos independientes, tal que sus trayectorias sean continuas con probabilidad igual a uno, es Normal (ver Breiman, L. (1968)); **ii)** para $0 \leq t_1 < t_2$, el proceso $\{Y(t)\}$ es tal que $Y(t_2) - Y(t_1) : N\left[(t_2 - t_1)\varphi_0, (t_2 - t_1)\sigma_\varepsilon^2\right]$.

En particular, si $\varphi_0 = 0$ y $\sigma_\varepsilon^2 = 1$, entonces se dice que $\{Y_t\}$ es un **movimiento Browniano estandarizado** para el cual se verifica que: **i)** $E\left[B(t_2) / B(t_1)\right] = B(t_1)$ $(t_1 \leq t_2)$ (o, lo que es lo mismo, que $E\left[B(t_2) - B(t_1) / B(t_1)\right] = 0$ $(t_1 \leq t_2)$); **ii)** $\sigma^2\left[B(t_2) / B(t_1)\right] = (t_2 - t_1)$ $(t_1 \leq t_2)$ y **iii)** $\gamma\left[B(t_1), B(t_2)\right] = t_1$ $(t_1 \leq t_2)$.

Sea un vector estocástico de orden n, $\left[B(t_1) \quad B(t_2) \quad \ldots \quad B(t_n)\right]$. Teniendo en cuenta que su función generatriz de momentos es de la forma:

---

59. Luego se puede concluir que, como todo proceso no-constante con incrementos independientes estacionarios, los procesos movimiento Browniano no son estacionarios.

60. Esta propiedad fue planteada por Bachelier, L. (1900).

$$M_{B(t_1),B(t_2),\ldots,B(t_n)}(w_1,w_2,\ldots,w_n) =$$

$$= E\left\{\exp\left[B(t_n)w_n + B(t_{n-1})w_{n-1}+\ldots+B(t_1)w_1\right]\right\} =$$

$$= E\left(\exp\left\{\left[B(t_n) - B(t_{n-1})\right]w_n + \left[B(t_{n-1}) - B(t_{n-2})\right](w_{n-1} + w_n) - B(t_{n-2}) + \right.\right.$$

$$\left.\left. +\ldots+ B(t_1)(w_1 + w_2+\ldots+ w_n)\right\}\right) =$$

$$= E\left(\exp\left\{\left[B(t_n) - B(t_{n-1})\right]w_n\right\}\right)E\left(\exp\left\{\left[B(t_{n-1}) - B(t_{n-2})\right](w_{n-1} + w_n)\right\}\right)\ldots$$

$$\ldots E\left(\exp\left\{B(t_1)(w_1 + w_2+\ldots+ w_n)\right\}\right) =$$

$$= \exp\left[\frac{1}{2}w_n^2(t_n - t_{n-1})\right]\exp\left[\frac{1}{2}(w_{n-1} + w_n)^2(t_{n-1} - t_{n-2})\right]\ldots\exp\left[\frac{1}{2}(w_1 + w_2+\ldots+ w_n)^2 t_1\right] =$$

$$= \exp\left[\frac{1}{2}(w_1 + w_2+\ldots+ w_n)^2 t_1 + \frac{1}{2}(w_2 + w_3+\ldots+ w_n)^2(t_2 - t_1)+\ldots+\frac{1}{2}(w_{n-1} + w_n)^2(t_{n-1} - t_{n-2}) + \right.$$

$$\left. + \frac{1}{2}w_n^2(t_n - t_{n-1})\right]$$

queda demostrado que las variables $B(t_1), B(t_2),\ldots, B(t_n)$ se distribuyen en forma conjunta de acuerdo con una función Normal n-dimensional con valores esperados nulos y matriz de varianzas y covarianzas de la forma:

$$\begin{bmatrix} E\left[B^2(t_1)\right] & E\left[B(t_1)B(t_2)\right] & \ldots & E\left[B(t_1)B(t_n)\right] \\ E\left[B(t_2)B(t_1)\right] & E\left[B^2(t_2)\right] & \ldots & E\left[B(t_2)B(t_n)\right] \\ \ldots & \ldots & \ldots & \ldots \\ E\left[B(t_n)B(t_1)\right] & E\left[B(t_n)B(t_2)\right] & \ldots & E\left[B^2(t_n)\right] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & \ldots & t_2 \\ \ldots & \ldots & \ldots & \ldots \\ t_1 & t_2 & \ldots & t_n \end{bmatrix}$$

Como corolario de estos resultados, se puede concluir que, para $0 \le t_1 < t_2$, $Y(t_2) - Y(t_1)$: $N(0, t_2 - t_1)$. Es decir, que $f_{Y(t_2)-Y(t_1)}\left(y(t_2) - y(t_1)\right) = \dfrac{1}{\sqrt{2\pi(t_2 - t_1)}}\exp\left\{-\dfrac{y(t_2) - y(t_1)}{2(t_2 - t_1)}\right\}$.

El proceso definido por:

$$Y(t) = \int_0^t B(\tau)d\tau \qquad (0 \le t < \infty)$$

(donde $\{B(t)\}$ es un movimiento Browniano estandarizado) constituye el **movimiento Browniano integrado**, cuyos momentos son de la forma[61]:

---

[61]. Ver Parzen, E. (1962).

$$m\big(Y(t)\big) = \int_0^t B(\tau)d\tau = 0$$

$$m_2\big(Y(t)\big) = \int_0^t\int_0^t E\big[B(s)B(\tau)\big]ds\,d\tau = \int_0^t\left[\int_0^t 2\tau\,d\tau\right]ds = \frac{t^3}{3}$$

$$\gamma\big[Y(t_1),Y(t_2)\big] = m_2\big(Y(t_1)\big) + (t_2 - t_1)E\big[Y(t_1)B(t_1)\big] = s^2\left(\frac{t_2}{2} - \frac{t_1}{6}\right) \qquad (0 \le t_1 < t_2)$$

El proceso definido por:

$$Z(t) = B(t) - t\,B(1)$$

constituye el llamado **puente Browniano** o **movimiento Browniano restringido estandarizado**[62], cuyos momentos son de la forma:

$$m\big[Z(t)\big] = 0$$

$$\gamma\big[Z(t_1),Z(t_2)\big] = \gamma\Big\{\big[Z(t_2) - t_2Z(1)\big]\big[Z(t_1) - t_1Z(1)\big]\Big\} =$$

$$= \gamma\big[Z(t_1),Z(t_2)\big] - t_1\gamma\big[Z(t_2),Z(1)\big] - t_2\gamma\big[Z(t_1),Z(1)\big] + t_1t_2\gamma\big[Z(1),Z(1)\big] =$$

$$= t_1 - t_1t_2 - t_1t_2 + t_1t_2 = t_1 - t_1t_2 \qquad (t_1 \le t_2\,;t_1,t_2 \in [0;1])$$

Asimismo, el proceso definido por:

$$Y(t) = e^{\varphi_0 t + \sigma_\varepsilon B(t)} \qquad (t \ge 0)$$

constituye el llamado **movimiento Browniano exponencial con tendencia** o **movimiento Browniano geométrico**[63], cuyos momentos son de la forma[64]:

---

62. Este proceso surge como límite de una sucesión de variables aleatorias independientes, todas con distribución uniforme del tipo $U(0;1)$ y, como se vio en la Sec. 2.14.2, constituye el fundamento de muchos tests de adecuacidad de la representación de una distribución de observaciones.

63. Este proceso constituye el punto de partida que conduce a la fórmula de Black-Scholes-Merton. De la misma forma que en los casos anteriores, si $\varphi_0 = 0$ y $\sigma_\varepsilon^2 = 1$, queda definido el **movimiento Browniano geométrico estandarizado**. Esta denominación se basa en que, a partir del proceso $\{Y(t)\} = \{\exp[B(t)]\}$ $(0 \le t < \infty)$ (donde $\{B(t)\}$ denota un movimiento Browniano estandarizado) puede definirse un proceso $\{Z(t)\}$ iid de la forma $\{Z(t)\} = \left\{\dfrac{Y(t)}{Y(t-1)}\right\}$ $(0 \le t < \infty)$ tal que, si $Y(0) = 1$, constituye la media geométrica de $Y(t)$ (ver Yor, M. (2001)).

64. Debe tenerse en cuenta que, si bien el movimiento Browniano geométrico viola la condición de eficiencia del mercado y hasta admite arbitrajes, no afecta en forma significativa los resultados de la fórmula de Black-Scholes-Merton.

$$m\left[Y(t)\right] = e^{\varphi_0 t}\, E\left(e^{\sigma_\varepsilon B(t)}\right) = \exp\left[\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)t\right]$$

$$\sigma^2\left[Y(t)\right] = m_2\left[Y(t)\right] - m^2\left[Y(t)\right] = \exp\left[2\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)t\right]\left(e^{\sigma_\varepsilon^2 t} - 1\right)$$

$$\gamma\left[Y(t_1), Y(t_2)\right] = E\left[Y(t_1)Y(t_2)\right] - E\left[Y(t_1)\right]E\left[Y(t_2)\right] =$$

$$= \exp\left[\varphi_0(t_1 + t_2)\right]E\left\{\exp\left[\sigma_\varepsilon\left(B(t_1) - B(t_2)\right)\right]\right\} - \exp\left[\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)(t_1 + t_2)\right] =$$

$$= \exp\left[\varphi_0(t_1 + t_2)\right]E\left\{\exp\left[\sigma_\varepsilon\left(B(t_1) - B(t_2)\right) + 2B(t_2)\right]\right\} - \exp\left[\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)(t_1 + t_2)\right] =$$

$$= \exp\left[\varphi_0(t_1 + t_2)\right]E\left\{\exp\left[\sigma_\varepsilon\left(B(t_1) - B(t_2)\right)\right]\right\}E\left[\exp\left(2\sigma_\varepsilon B(t_2)\right)\right] - \exp\left[\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)(t_1 + t_2)\right] =$$

$$= \exp\left[\left(\varphi_0 + \frac{\sigma_\varepsilon^2}{2}\right)(t_1 + t_2)\right]\left(e^{\sigma_\varepsilon^2 t_2} - 1\right)$$

De lo expuesto en esta sección se puede concluir que muchos procesos de aparición habitual en las aplicaciones pueden ser aproximados por un movimiento Browniano y sus propiedades asintóticas pueden ser estudiadas incluyéndolos en la teoría del movimiento Browniano. Esta metodología, conocida como la **inclusión de Skorohod** para sumas de variables aleatorias independientes (ver Skorohod, A.V. (1965)), fue generalizada posteriormente a las martingalas (ver Dambis, K.E. (1965), Dubins, L.E. (1968), Dubins, L.E.; Schwartz, G. (1965), Monroe, I. (1972)(1978), Strassen, V. (1967)).

Un argumento muy importante en el cálculo estocástico es el denominado **cambio de medida** sobre el espacio de las trayectorias del movimiento Browniano, que consiste en la introducción de una nueva medida de probabilidad mediante una **función de densidad** que, en general, no constituye una función de densidad de probabilidades. En otros términos, consiste en una modificación de las probabilidades asignadas a cada trayectoria y, por lo tanto, es equivalente a un cambio en su tendencia[65].

En general, dadas dos medidas de probabilidad, P y Q, si existe una función no-negativa $g(\bullet)$ tal que $Q(A) = \int_A g_1(w)dP(w)$, se dice que g_1 es la **densidad de** Q **con respecto a** P (de acuerdo con el teorema de Radon-Nikodym se demuestra, además, que Q es absolutamente continua con respecto a P). De la misma forma, si existe una función no-negativa $g_2(\bullet)$ tal que:

$$P(A) = \int_A g_2(w)dQ(w)$$

---

[65]. El movimiento Browniano (descubierto en el microscopio en 1827 y explicado, en base a elementos de mecánica estadística clásica) y la radioactividad (descubierta en 1896) son dos fenómenos específicos que, a partir de un desarrollo exclusivamente experimental, jugaron un papel muy importante en el desarrollo de la teoría de la probabilidad.

se dice que g_2 es la densidad de P con respecto a Q. Si se verifica que P es absolutamente continua con respecto a Q y Q es absolutamente continua con respecto a P, se dice que P y Q son **medidas de probabilidad equivalentes**[66].

Sea, entonces, un movimiento Browniano con tendencia de la forma:

$$B^*(t) = B(t) + ct$$

(donde c es una constante). De acuerdo con el **teorema de Girasinov**[67], se demuestra que el proceso:

$$K(t) = \exp\left[-cB(t) - \frac{1}{2}c^2 t\right] \qquad (t \in [0;T])$$

es una martingala con respecto al conjunto de información $\{B(\tau)\}$ $(\tau \le t)$, bajo una medida de probabilidad $P = p(B(t) \le x)$. Ahora bien, bajo la medida de probabilidad P, $B(t)$ no satisface las condiciones de movimiento Browniano estandarizado, pero a partir de la relación:

$$Q(A) = \int_A K(t,w)\,dP(w)$$

es posible hallar una medida de probabilidad alternativa Q, equivalente a P, tal que se puede asegurar que, bajo esta medida Q, el proceso $\{B^*(t)\}$ es un movimiento Browniano estandarizado. Se dice, entonces, que Q es una **medida de martingala equivalente**.

El movimiento Browniano pertenece a una clase más general de procesos conocidos como **autosimilares en el dominio del tiempo.**

Se dice que un proceso $\{Y(t)\}$ es **auto-similar con respecto a un parámetro H** (o **H-autosimilar**) cuando se verifica la siguiente condición con respecto a la distribución de probabilidades conjunta para cualquier sucesión finita de sus variables aleatorias:

---

[66]. Sean Y_1 e Y_2 dos variables aleatorias con distribuciones de probabilidades Normales del tipo $N(m_1,\sigma_1^2)$ y $N(m_2,\sigma_2^2)$, respectivamente y sean las funciones $g_1(y) = \frac{f_{Y_1}(y)}{f_{Y_2}(y)}$ y $g_2(y) = \frac{f_{Y_2}(y)}{f_{Y_1}(y)}$. Obviamente se verificará que $F_{Y_1}(y) = \int_{-\infty}^{y} g_1(x) f_{Y_2}(x)\,dx$ y que $F_{Y_2}(y) = \int_{-\infty}^{y} g_1(x) f_{Y_1}(x)\,dx$. Entonces $g_1(\bullet)$ es la función de densidad de $F_{Y_1}(\bullet)$ con respecto a $F_{Y_2}(\bullet)$ y $g_2(\bullet)$ es la función de densidad de $F_{Y_2}(\bullet)$ con respecto a $F_{Y_1}(\bullet)$. Obsérvese que $g_1(\bullet)$ y $g_2(\bullet)$ son funciones positivas, pero tales que $\int_{-\infty}^{\infty} g_i(x)\,dx \neq 1\ (i = 1,2)$ de modo que no son funciones de densidad de probabilidades.

[67]. Denominado también en la literatura como **fórmula de Cameron-Martin** (ver Girasinov, I.V. (1969), Karatzas, I.; Shreve, S.E. (1988), Kallenberg, O. (1997)).

$$f_{Y(t_1),Y(t_2),\ldots,Y(t_n)}\left(y(t_1),y(t_2),\ldots,y(t_n)\right) = \tau^{-H} f_{Y(\tau t_1),Y(\tau t_2),\ldots,Y(\tau t_n)}\left(y(\tau t_1),y(\tau t_2),\ldots,y(\tau t_n)\right)$$

$$\left\{Y(t_1),Y(t_2),\ldots,Y(t_n)\right\} \overset{d}{=} \left\{\tau^{-H}Y(\tau t_1),\tau^{-H}Y(\tau t_2),\ldots,\tau^{-H}Y(\tau t_n)\right\}$$

(para $t_i \geq 0; i = 1,2,\ldots,n; n \geq 1; \tau > 0; H > 0$), es decir, cuando se verifica que, para escalas distintas, la estructura del comportamiento del proceso es la misma.

Debe tenerse en cuenta que la condición de autosimilitud no implica que las trayectorias observadas deban ser todas iguales. En realidad, contrariamente a lo que ocurre con algunos procesos determinísticos que presentan un tipo de **autosimilitud estricta** justificable sólo matemáticamente, los fenómenos fácticos presenta una clase de **autosimilitud cualitativa**, conocida como **autosimilitud estadística**, según la cual, a cada escala, el proceso se asemeja estadísticamente a las otras escalas -espaciales o temporales-, pero no es idéntico (si bien sus partes poseen la forma de la totalidad, presentan variaciones)[68].

Todo conjunto fractal posee la propiedad de autosimilitud en el espacio. Esta propiedad de invariancia a escala implica que si, mediante la aplicación de una transformación dada (denominada **transformación de similitud**, que se caracteriza porque el factor escala es el mismo para todas las dimensiones del sistema), se somete al conjunto a un proceso de fraccionamiento en pequeñas partes, cada una de estas permanecerá similar al conjunto original en sus características espaciales. La implicación inversa no se verifica necesariamente, es decir, que un proceso sea autosimilar no significa que deba ser fractal. Esta simetrías recursivas que generan la autosimilitud -y las complejas interacciones entre los diferentes niveles del proceso- determinan la base de la interpretación sistémica que, en oposición a la interpretación clásica -que supone el comportamiento del proceso como resultado de una interacción agregada de sus unidades individuales-, constituye la hipótesis fundamental de la teoría de los sistemas dinámicos, la cual permite la representación del comportamiento de un proceso mediante una pocas ecuaciones iterativas.

Se demuestra[69]: **i)** que la función de autocovarianzas de un proceso autosimilar, con respecto al cual se cuenta con un conjunto de observaciones realizadas a intervalos regulares (discretos) de tiempo, es de la forma [70]:

$$\gamma_j(Y) = \frac{c\gamma_0(Y)}{2}\left(|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right) \qquad \left(c > 0; \frac{1}{2} < H < 1\right)$$

y que, para valores suficientemente grandes de j, puede ser aproximada de la siguiente forma: $\gamma_j(Y) \approx cH(2H-1)|j|^{2H-1}$ y **ii)** que todo proceso $\{Y(t)\}$ autosimilar es no-diferenciable para todo $t \in T$.

---

[68]. Existen otros procesos que no poseen la propiedad de autosimilitid, pero presentan un tipo de **autoafinidad**. La **transformación de autoafinidad** se caracteriza porque los factores escala para las distintas dimensiones del sistema pueden ser diferentes entre sí. De la misma forma que en el caso anterior, es posible definir una **autoafinidad estadística** según la cual el proceso original se asemeja estadísticamente al proceso sometido a una **transformación de afinidad** (para más detalles sobre las transformaciones de similitud y de afinidad y sus aplicaciones, ver Mandelbrot, B.B. (1983),(1985),(1986), Voss, R.F. (En Pynn, R.; Skjeltrop, A. (eds.)(1985)), Earnshaw, R.A. (1985)).

[69]. Ver Kolmogorov, A.N. (1940), Mandelbrot, B.B.; van Ness, J.W. (1968).

[70]. A los procesos que poseen esta función de autocovarianzas también se los suele denominar como **WN con orden diferenciación fraccionario**- $N\left(0;\sigma_\varepsilon^2\right)$ (**FNN**) (ver Mandelbrot, B.B.; van Ness, J.W. (1968), Mandelbrot, B.B.; Wallis, J. (1968)) y serán tratados en la Sec. 4.2. De acuerdo con la nomenclatura de Mandelbrot, B.B. (1972)(1975), se dice que un proceso $\{Y_t\}$ es fractal si se verifica que $|\nabla Y(t)| = \left(\frac{\nabla t}{c}\right)^H$ (donde ∇t denota un intervalo de tiempo y c es una constante) y se dice que $\{Y(t)\}$ es un proces fractal auto-similar cuando el coeficiente H es el mismo para cualquier ∇t.

En particular, para $0 < H < 1$ y todo $\tau \in T$, se verifica que $\{Y(t)\}$ es no-diferenciable con respecto a t con probabilidad uno, es decir que:

$$\lim_{n\to\infty} Max\, P\left(\frac{|Y(t_n) - Y(\tau)|}{t_n - \tau} > y\right) = 1 \qquad (\tau < t_n, t_n \downarrow \tau)$$

o, lo que es lo mismo, que $\lim_{t\downarrow\tau} Max \frac{|Y(t) - Y(\tau)|}{t - \tau} = \infty$ [71].

El movimiento Browniano es 0.5-autosimilar, es decir, es tal que[72]:

$$\{B(t_1), B(t_2), ..., B(t_n)\} \overset{d}{=} \{\tau^{-1/2} B(\tau t_1), \tau^{-1/2} B(\tau t_2), ..., \tau^{-1/2} B(\tau t_n)\}$$

$(\tau > 0; t_i \geq 0; i = 1,2,...,n; n \geq 1)$. De modo que se puede concluir que el proceso $\{B(t)\}$ es un proceso fractal, que es no-diferenciable y que, en consecuencia, para su definición no es posible utilizar los métodos habituales de integración[73].

Desde un punto de vista más intuitivo se puede arribar a la misma conclusión si se tiene en cuenta: **i)** que debido a que los incrementos de $\{B(t)\}$ son independientes y que su variación para cualquier intervalo finito $[0,T]$ es no-acotada, es decir, que $\underset{A}{Max} \sum_{i=1}^{n} |B(t_i, w) - B(t_{i-1}, w)| = \infty$, (donde el operador Max afecta a todas las particiones posibles del dominio del tiempo $A: t_0 < t_1 < ... < t_n = T$)[74] y **ii)** que el límite $\lim_{h\to 0}[B(t+h) - B(t)]$ no existe y, por lo tanto, cuando $h \to 0$, la variable $\frac{B(t+h) - B(t)}{h}$ no converge en

---

[71]. Esta propiedad poco intuitiva, que implica reconocer que la partícula de polen que dio origen a todo este desarrollo se mueve a una velocidad infinita, produjo en su momento una cierta conmoción teórica.

[72]. Que el movimiento Browniano sea 0.5-autosimilar implica que, en las condiciones mencionadas, la distancia recorrida por una partícula es igual a la raíz cuadrada de la longitud del período de medición (propiedad conocida como regla de τ a la ½, $R = \sqrt{\tau}$)

[73]. Un proceso fractal íntimamente vinculado con el movimiento Browniano es el conocido como $ARFIMA(p,d,q)$ (denominación resumida de la representación $ARMA$ con orden de integración fraccionario), que admite una representación de la forma $\varphi(B)Y_t = \Phi(B)(1-B)^d Y_t = \Phi(B)(\nabla^d Y_t) = \Theta(B)\varepsilon_t$. Donde: I) Y_t denota una variable centrada; ii) $\{\varepsilon_t\}: WN - N(0, \sigma_\varepsilon^2)$ y iii) d puede asumir cualquier valor real en el intervalo $(-1;1)$. Para $0 < d < \frac{1}{2}$, es de memoria larga y para $-\frac{1}{2} < d < 1$, es de memoria corta (estos procesos, introducidos por B.Mandelbrot y J.B. van Ness para formalizar los resultados obtenidos por H.E. Hurst, serán tratados en el Cap. 5).

[74]. Se dice que una función $f(t)$ es de p-variación acotada, si se verifica que $\sup_{\tau} \sum_{j=1}^{n} |f(t_i) - f(t_{i-1})|^p < \infty$ (donde: **i)** $p > 0$ y **ii)** el operador "supremum" es sobre todas las particiones τ del intervalo $[0;1]$). Debe tenerse en cuenta que la condición $x = \sup_n(x_n)$ implica que: **i)** $x \geq x_n$ para todo n y **ii)** para todo $\varepsilon > 0$, existe un valor j tal que $x - \varepsilon < x_j$.

distribución a ninguna variable[75].

Debe tenerse en cuenta que, no obstante todas estas consideraciones -acerca de que, excepto para los elementos de un subconjunto $\Omega^* \subset \Omega$ tal que $p(\Omega^*) = 0$, la derivada $\frac{dB(t)}{dt}$ no existe- para representar al proceso $\{B(t)\}$ suele utilizarse la siguiente notación $dB(t) = \lim_{h\to 0}[B(t+h) - B(t)]$.

4.6.- El proceso de Itô

Supóngase ahora que la tendencia y la volatilidad de $\{Y(t)\}$ sean funciones de $Y(t)$, utilizando la notación de las secciones precedentes, se puede escribir la siguiente **ecuación diferencial estocástica**[76]:

$$dY(t) = m(t, Y(t))\,dt + \sigma_\varepsilon(t, Y(t))\,dB(t)$$

donde las funciones determinísticas $m(t, Y(t))$ y $\sigma_\varepsilon(t, Y(t))$ pueden ser interpretadas, respectivamente, como la tendencia o la medida de la influencia de la variación de la variable tiempo (dt) y la medida de la influencia de la variación del movimiento Browniano ($dB(t)$)[77], en la variación de $\{Y(t)\}$ [78].

Como se mencionó en la Sec. 1.7.2.1, la calificación de **estocástica** respecto de esta ecuación diferencial es para distinguirla de las ecuaciones diferenciales ordinarias (determinísticas), que se caracterizan: **i)** porque sus soluciones son funciones; **ii)** porque para obtener una solución única es necesario conocer la condición inicial $Y(0) = Y_0$ y, si esta condición coincide con el valor inicial de $Y(t)$, entonces $\{Y(t)\}$ queda completamente determinado para $t > 0$; **iii)** porque integrando m. a m. se obtiene

---

[75]. En la Sec. 1.3 se definió a un ruido blanco como un proceso formado por variables aleatorias independientes o incorrelacionadas linealmente, con valor esperado nulo y varianza constante. En las aplicaciones a la física el proceso ruido blanco se define como una derivada (obtenida por métodos que no pertenecen a la diferenciación ordinaria en el sentido del análisis) del movimiento Browniano. Como el ruido blanco es físicamente inimaginable, se define el proceso **ruido coloreado** como un proceso Normal de la forma $\{Y(t)\} = \left\{\frac{B(t+h) - B(t)}{h}\right\}$ $(t \geq 0, h > 0)$, cuyos momentos están definidos por $m(Y(t)) = 0$, $\sigma^2(Y(t)) = \frac{1}{n}$ y $\gamma(Y(t_1), Y(t_2)) = \frac{1}{h^2}[(t_1 + h) - \min(t_1 + h, t_2)]$ $(t_1 < t_2)$. Obsérvese que: **i)** si $t_2 - t_1 \geq h$, entonces $\gamma(Y(t_1), Y(t_2)) = 0$, es decir que $Y(t_1) \perp Y(t_2)$; **ii)** si $t_2 - t_1 < h$, entonces $\gamma(Y(t_1), Y(t_2)) = \frac{h - (t_2 - t_1)}{h^2}$; **iii)** dado que la varianza de $\{Y(t)\}$ es inversamente proporcional a h , su variabilidad aumenta cuando h disminuye y **iv)** como $\gamma(Y(t_1), Y(t_2))$ es una función de $t_2 - t_1$, se puede concluir que $\{Y(t)\} : I(0)$.

[76]. Conocida también como **ecuación de difusión .**

[77]. Conocida también como **función de difusión.**

[78]. Como se verá inmediatamente, para un valor $Y(0)$ dado, esta ecuación tiene una solución única en la clase de los procesos continuos en valor medio cuadrático en el sentido que, dada una solución alternativa $Y^*(t)$, se verificará que $p[Y(t) - Y^*(0) = 0] = 1$ y, además, satisface una ecuación estocástica en-diferencias con perturbaciones incorrelacionadas.

una **ecuación integral** (no una ecuación integral estocástica). La aleatoriedad de esta ecuación se origina en la aleatoriedad inherente a la condición inicial[79] y en la presencia del movimiento Browniano. Su solución, si existe, será un proceso estocástico. Debe tenerse en cuenta que su solución no requiere de cálculo estocástico, se pueden utilizar métodos clásicos del análisis, adaptando la solución a una realización particular de la condición inicial.

Dado que, como se comentó en las páginas precedentes, el proceso $\{B(t)\}$ no posee trayectorias diferenciables, parece conveniente interpretar esta ecuación como una **ecuación integral estocástica**. Esta teoría de las integrales estocásticas se basa en la condición que, dadas dos funciones $f(t)$ y $g(t)$ tales que: **i)** no poseen discontinuidades en el mismo punto $t \in [0;1]$ y **ii)** $f(t)$ es una función de p -variación acotada ($p > 0$) y $g(t)$ es una función de q -variación acotada ($q > 0$), tal que $p^{-1} + q^{-1} > 1$,la integral:

$$\int_0^1 f(t)dg(t)$$

existe en el sentido de las integrales de Riemann-Stieltjes (ver Rozanov, Y.A. (1967)). En particular, si $f(t)$ es una función determinística o la trayectoria de un proceso estocástico adaptado con respecto a la filtración $\Omega(Y(t))$, diferenciable con derivada continua en el intervalo $[0;1]$, entonces la integral de Riemann-Stieltjes:

$$\int_0^1 f(t)dB(t,w)$$

existe para toda trayectoria del movimiento Browniano $B(t,w)$ (debe tenerse en cuenta que, dado que: **i)** $\{B(t,w)\}$ es un proceso con incrementos independientes (ver Sec. 4.5) y **ii)** $\{f(t)\}$ es $\Omega(Y(t))$ - medible, entonces $\{f(t)\}$ es independiente de los incrementos futuros de $B(t,w)$).

Como se mencionó en la Sec. 1.3, un proceso de este tipo es un proceso $\Omega(Y(t))$ – *medible* para todo $t \geq 0$. En otros términos, un proceso tal que el conjunto de información con que cuenta el observador al momento t es suficiente para evaluar al proceso $\{f(t)\}$ al momento t (o, lo que es lo mismo, tal que la filtración $\Omega(Y(t))$ es suficiente para que $f(t)$ deje de ser una variable aleatoria).

De la ecuación diferencial estocástica definida en la página precedente, se puede concluir que su solución está relacionada, en principio, con una integral del tipo $I(B(w)) = \int_0^1 B(t,w)dB(t,w)$ (conocida como **integral de Itô**)[80] que no tiene solución, para cada trayectoria, en el sentido de Riemann-Stieltjes.

---

[79]. Dada una ecuación diferencial ordinaria de la forma $dY(t) = m(t,Y(t))dt \; (Y(0) = Y_0)$, la correspondiente versión estocástica más inmediata se presenta cuando la condición inicial está afectada por aleatoriedad: $Y(0,w) = Y_0(w) \; (w \in \Omega(Y))$ (obsérvese, entonces, que las ecuaciones diferenciales estocásticas pueden ser consideradas como ecuaciones diferenciales ordinarias con condiciones iniciales perturbadas).

[80]. El concepto de integral estocástica fue introducido por Doeblin, W. (1940), Itô, K. (1942)(1944). Por su parte, Doob, J.L. (1953) vinculó este concepto con la teoría de martingalas. Para un análisis más avanzado, ver Øksendahl, B. (1985), Ikeda, N.; Watanabe, S. (1989), Chung, K.L.; Williams, R.J. (1990), Karatzas, I.;Shreve, S.E. (1998).

A partir de la aproximación que proporciona la **suma de Riemann-Stieltjes**[81]:

$$S_n = \sum_{j=1}^{n} B\left(t_{j-1},w\right)\left[B\left(t_j w\right) - B\left(t_{j-1},w\right)\right] = \sum_{j=1}^{n} B\left(t_{j-1},w\right)\nabla_{(j)}B(t,w)$$

(donde τ_n denota una partición del intervalo $[0;1]$ de la forma $0 = t_0 < t_1 <...< t_{n-1} < t_n = 1$) se obtiene la siguiente solución como límite en-valor medio cuadrático[82]:

$$\int_0^t B(\tau,w)dB(\tau,w) = \frac{1}{2}\left[B^2(t,w) - t\right]$$

De la misma forma se obtiene la siguiente expresión para la variación cuadrática del movimiento Browniano: $d\left[B(t,w)\right]^2 = \left[B(t+dt,w) - B(t,w)\right]^2 = dt$ o, en términos de su expresión integral, $\int_0^t \left[dB(\tau,w)\right]^2 = \int_0^t d\tau = t$.

Debe tenerse en cuenta que: **i)** esta igualdad, que expresa la convergencia en L^2 , se verifica casi con certeza, es decir que pueden existir trayectorias para las cuales esta ecuación no se verifica y que el conjunto de dichas trayectorias es de probabilidad nula, en otros términos, el conjunto de trayectorias para las cuales se verifica la igualdad es de probabilidad 1[83]; **ii)** esta circunstancia no permite asegurar que la

---

[81]. Supóngase que $\left\{B(t,w)\right\}$ denote el proceso que representa el comportamiento del precio de un activo al momento t . Sean: **i)** $t_0,t_1,t_2,...,t_n$ los momentos de comercialización del activo y **ii)** $f(t_0),f(t_1),...,f(t_n)$ las posiciones adoptadas en los momentos t_j $(j = 0,1,2,...,n)$ respecto del activo. Entonces la aproximación a la integral de Itô del proceso $\left\{f(t)\right\}$, $I(t) = \sum_{j=1}^{n} f\left(t_{j-1}\right)\nabla_{(j)}B(t,w)$, representa la ganancia obtenida "a posteriori" de la operación realizada almomento t .

[82]. Debe tenerse en cuenta que utilizando métodos de integración determinística, se obtiene que $\int_0^t f(\tau)df(\tau) = \frac{1}{2}\left[f^2(t) - f^2(0)\right]$.

[83]. Suponiendo un valor de n suficientemente grando, sea $t_j = \frac{jt}{n}$ $(j = 0,1,2,...,n)$. Entonces, será $t_{j+1} - t_j = \frac{t}{n}$ $(j = 0,1,2,...,n)$ y $\left[B\left(t_{j+1}\right) - B\left(t_j\right)\right]^2 = t\frac{Y_{j+1}^2}{n}$. Teniendo en cuenta que las variables $Y_1,Y_2,...,Y_n$ son iid, de acuerdo la ley de los grandes números se puede concluir que cuando n aumenta indefinidamente, $\sum_{j=0}^{n-1}\frac{Y_{j+1}^2}{n} \xrightarrow{p} E\left(Y_{j+1}^2\right)$. Dado que $E\left(Y_{j+1}^2\right) = 1$, se demuestra que $\lim_{n\to\infty}\sum_{j=0}^{n-1}\left[B\left(t_{j+1}\right) - B\left(t_j\right)\right]^2 = t$. Es necesario señalar que, si bien cada uno de los términos de esta suma puede ser bastante diferente de su valor esperado, $t_{j+1} - t_j = \frac{t}{n}$, cuando n aumenta indefinidamente, el promedio de las diferencias converge a cero.

integral $\int_0^t B^2(\tau,w)dB(\tau,w)$ sea una martingala y **iii)** la condición necesaria para poder asegurar la condición de martingala es que $E\left(\int_0^t\left[dB(\tau,w)\right]^2\right) < \infty$.

La variación cuadrática -que puede ser interpretada como una medida de riesgo y que constituye el origen de la volatilidad de los precios de un activo que observan un comportamiento representable por un movimiento Browniano- difiere de la varianza del proceso. Debe tenerse en cuenta que la variación cuadrática se calcula para cada trayectoria y, en consecuencia, el resultado es una función de las trayectorias, en tanto que la varianza de $I\left(B(t)\right)$ es un promedio sobre todas las trayectorias posibles (realizadas y no-realizadas) de la variación cuadrática.

Supóngase, ahora, una función $C\left(t,B(\tau)\right)$ $(\tau \le t)$ (es decir, una función predecible a partir de un movimiento Browniano en el intervalo $[0;T]$) tal que $\int_0^T E\left[C\left(s,B(\tau)\right)\right]ds < \infty$, queda definida, entonces, una **integral de Itô de C(.,.)** de la forma $I(t,C) = \int_0^t C\left(\tau,B(\tau)\right)dB(\tau)$, que es una martingala con respecto a su filtración natural (es decir, con respecto al conjunto de información $\left\{Y_{t-1},Y_{t-2},...\right\}$) y respecto de la cual se verifica que: **i)** $E\left[I(t,C)\right] = 0$ (teniendo en cuenta que la integral de Itô puede ser interpretada como la ganancia después de la operación en la martingala $B(t)$ y como, de acuerdo con lo demostrado en la Sec. 4.3, una martingala no posee tendencia -ni hacia "arriba", ni hacia "abajo"-, el valor esperado de la función $I(t,C)$, entendida como un proceso continuo y $\Omega\left(Y(t)\right)$ - *medible* en su límite superior de integración, t , tampoco posee tendencia); **ii)** posee la **propiedad de isometría**: $E\left\{\left[I(t,C)\right]^2\right\} = \int_0^t E\left\{\left[C\left(\tau,B(\tau)\right)\right]^2\right\}d\tau \left(t \in [0;T]\right)$; **iii)** posee la **propiedad de linealidad**: $\int_0^t\left[c_1C^{(1)}\left(\tau,B(\tau)\right) + c_2C^{(2)}\left(\tau,B(\tau)\right)\right]dB(\tau) = c_1I\left(t,C^{(1)}\right) + c_2I\left(t,C^{(2)}\right)\left(t \in [0;T]\right)$ (donde c_1 y c_2 son constantes) y **iv)** posee la **propiedad de linealidad sobre intervalos adyacentes**: $I(T,C) =$

$$= \int_0^t C\left(\tau,B(\tau)\right)dB(\tau) + \int_t^T C\left(\tau,B(\tau)\right)dB(\tau)\,(0 \le t \le T)\ .$$

Si, en particular, $C(.,.)$ es un **proceso simple** $\left\{C(t)\right\}\left(t \in [0,T]\right)$ (es decir, un proceso tal que, dadas: **i)** una partición $\tau_n : 0 = t_0 < t_1 <...< t_{n-1} < t_n = T$ y **ii)** una sucesión de variables aleatorias $Z_i\ (i = 1,2,...,n)$ predecible a partir de una filtración $\Omega\left(t_{i-1}\right)$ $(i = 1,2,...,n)$ y tal que $E\left(Z_i^2\right) < \infty$), entonces se verifica que: **i)** $C(t) = Z_n$ y **ii)** para $t_{i-1} \le t < t_i$ $(i = 1,2,...,n)$ y $C(t) = Z_i$, la función $I(t,C)$ admite una representación en términos de la suma de Riemann-Stieltjes.

La solución de la integral estocástica de Itô toma en consideración una modificación de la **regla de encadenamiento de la diferenciación** que, en su versión clásica, puede ser expresada de la siguiente forma $\left\{f\left[g(\tau)\right]\right\}' = f'\left[g(\tau)\right]g'(\tau)$, donde $f(\bullet)$ y $g(\bullet)$ denotan funciones diferenciables (esta expresión diferencial puede ser interpretada como el término de primer orden del desarrollo de Taylor

$f[g(t)+dg(t)]-f[g(t)]=f'[g(t)]dg(t)+\frac{1}{2}f''[g(t)][dg(t)]^2+\ldots$)[84]. Integrando m. a m. esta expresión, se obtiene que:

$$f[g(t)]-f[g(0)]=\int_0^t f'[g(\tau)]g'(\tau)d\tau=\int_0^t f'[g(\tau)]dg(\tau)\ .$$

Supóngase que $f(\bullet)$ sea una función diferenciable de segundo orden, si se reemplaza la función $g(\bullet)$ por un movimiento Browniano, $B(t,w)$, entonces la expresión precedente asume la siguiente forma:

$$f[B(t,w)]+dB(t,w)-f[B(t,w)]=f'[B(t,w)]dB(t,w)+\frac{1}{2}f''[B(t,w)][dB(t,w)]^2+\ldots$$

Integrando m.a m. esta expresión y considerando los términos de primero, segundo y tercer orden, se obtiene que[85]:

$$\int_0^t df[B(t,w)]=f[B(t,w)-B(s,w)]\approx\int_0^t f'[B(\tau,w)]dB(\tau,w)+\frac{1}{2}\int_0^t f''[B(\tau,w)]d\tau$$

(donde: **i)** el primer sumando del segundo miembro es la integral de Itô de $f'[B(t,w)]$ y **ii)** el segundo sumando puede ser interpretado como la integral de Riemann de $f''[B(t,w)]$). Esta expresión define el llamado **lema de Itô-Doeblin**[86] o **fórmula de Itô-Doeblin**[87].

Considerando una función $f[t,B(t,w)]$ con derivadas parciales continuas de por lo menos segundo orden, se obtienen la siguientes generalizaciones de la fórmula anterior, conocidas habitualmente como las **extensiones I y II del lema de Itô-Doeblin**[88]:

I) $f[t,B(t,w)]-f[s,B(s,w)]=$

---

[84]. Debe tenerse en cuenta que, en condiciones bastante generales, los términos de orden mayor o igual que dos son despreciables.

[85]. De acuerdo a lo expresado en las páginas anteriores, $d\{[B(t,w)]^2\}=dt$. De lo que se puede concluir que, contrariamente a lo que ocurre en el caso determinístico, la contribución del término de segundo orden del desarrollo de Taylor, no es despreciable. Esta constituye precisamente la "modificación" de la regla clásica de encadenamiento mencionada más arriba.

[86]. A pesar del olvido generalizado de la literatura, debe tenerse en cuenta que Wofgang Doeblin (un soldado alemán con destino en el frente francés) envió a la Academie des Sciences, en Febrero de 1940, un trabajo (que permaneció archivado y olvidado hasta Mayo de 2000) que contenía una definición de la integral estocástica y de la fórmula de cambio de variable, ligeramente diferente (y anterior en 4 años) de la demostración de Itô (ver Doeblin, W. (2000)). Doeblin falleció poco tiempo después de la presentación de su trabajo (para un comentario histórico detallado, ver Bru, B. (2000) y Bru, B.; Yor, M. (2002) y, para un análisis teorico del mismo, ver Yor, M. (2000)).

[87]. Distintas versiones del lema de Itô-Doeblin y sus demostraciones pueden ser halladas en Øksendahl, B. (1985), Karatzas, I.; Shreve, S.E. (1988), Ikeda, N.; Watanabe, S. (1989), Chung, K.L.; Williams, R.J. (1990), Dothan, M.U. (1990), Protter, P. (1992), Duffie, G. (1992), Hull, J. (2004).

[88]. Como se verá en las secciones siguientes el lema de Itô-Doeblin y sus extensiones constituyen herramientas fundamentales en la valuación de derivados financieros.

$$= \int_0^t \left\{ \frac{\partial}{\partial \tau} f\left[\tau, B(\tau,w)\right] + \frac{1}{2} \frac{\partial^2}{\partial B(\tau,w)} f\left[\tau, B(\tau,w)\right] \right\} \partial\tau + \int_0^t \frac{\partial}{\partial B(\tau,w)} f\left[\tau, B(\tau,w)\right] dB(\tau,w) \qquad (s < \tau)$$

$$II) \; f\left[t, B(t,w)\right] - f\left[s, B(s,w)\right] =$$

$$= \int_0^t \left\{ \frac{\partial}{\partial \tau} f\left[\tau, B(\tau,w)\right] + \frac{1}{2} \left[A^{(2)}(\tau)\right]^2 \frac{\partial^2}{\partial B(\tau,w)^2} f\left[\tau, B(\tau,w)\right] \right\} d\tau + \int_0^t \frac{\partial}{\partial B(\tau,w)} f\left[\tau, B(\tau,w)\right] dB(\tau,w)$$

donde $dB(\tau,w) = A^{(1)}(\tau)d\tau + A^{(2)}(\tau)dB(\tau,w)$, siendo $A^{(1)}(\bullet)$ y $A^{(2)}(\bullet)$ funciones no predecibles con respecto a un movimiento Browniano[89].

Luego, considerando dos funciones, $m\left(t,Y(t)\right)$ y $\sigma\left(t,Y(t)\right)$, no predecibles a partir de un movimiento Browniano, se obtiene la siguiente generalización de la fórmula de Itô-Doeblin:

$$Y(t) = Y(0) + \int_0^t m\left(\tau,Y(\tau)\right)d\tau + \int_0^t \sigma_\varepsilon\left(\tau,Y(\tau)\right)dB(\tau,w) \qquad (0 \le t \le T)$$

(donde: **i)** el primer sumando del segundo miembro es una integral de Riemann y **ii)** el segundo sumando es una integral de Itô, que es una martingala). La cual es la expresión integral de la ecuación diferencial estocástica correspondiente a un movimiento Browniano con tendencia, mencionada en las páginas anteriores, conocida como la **ecuación diferencial estocástica de Itô**, en la que $B\left(t,Y(t)\right)$ define el proceso directriz de dicha ecuación[90]. Todo proceso que admite una representación de este tipo se denomina **proceso de Itô** y se demuestra que, si el proceso $\left\{Y(t)\right\}$ satisface esta ecuación, entonces los procesos $m\left(t,Y(t)\right)$ y $\sigma\left(t,Y(t)\right)$ quedan definidos en forma unívoca y, viceversa, dadas la condición inicial y las funciones $m\left(t,Y(t)\right)$ y $\sigma\left(t,Y(t)\right)$, es posible hallar un proceso $\left\{B(t)\right\}$ que satisfaga dicha ecuación. En particular: **i)** si $m\left(t,Y(t)\right) = 0$ y $\sigma\left(t,Y(t)\right) = 1$, queda definido un **proceso de difusión; ii)**

---

[89]. Generalizando las definiciones precedentes, sea $\left\{Y(t)\right\} = \left\{Y^{(1)}(t), Y^{(2)}(t), \ldots, Y^{(k)}(t)\right\}$ un conjunto de procesos de Itô y sea $f(\bullet,\bullet,\ldots,\bullet)$ una transformación no-lineal, entonces $f\left(t, Y^{(1)}(t), \ldots, Y^{(k)}(t)\right)$ será un proceso de Itô que satisface la ecuación diferencial estocástica $df\left(t,Y(t)\right) = df\left(t, Y^{(0)}(t), \ldots, Y^{(k)}(t)\right) = \frac{\partial f}{\partial t}dt + \sum_{h=1}^{k} \frac{\partial f}{\partial Y^{(h)}(t)} +$ $+ \frac{1}{2} \sum_{h=1}^{k} \sum_{i=1}^{k} \frac{\partial f}{\partial Y^{(h)}(t) Y^{(i)}(t)} dY^{(h)}(t) dY^{(i)}(t)$. Por otra parte, sean $\left\{Y^{(i)}(t)\right\}$ e $\left\{Y^{(j)}(t)\right\}$ dos procesos de Itô, entonces se verificará que $d\left(Y^{(i)}(t)Y^{(j)}(t)\right) = Y^{(i)}(t)dY^{(i)}(t) + Y^{(j)}(t)dY^{(j)}(t) + dY^{(i)}(t)dY^{(j)}(t)$.

[90]. Ver Itô, K. (1951). Obviamente, es posible reemplazar el movimiento Browniano por otros procesos directrices, pero esta extensión requiere la definición de otras integrales. Chung, K.L.; Williams, R.J. (1990) y Protter, P. (1992) introdujeron las integrales estocásticas con respecto a la familia de las semimartingalas (que contiene al movimiento Browniano como caso particular), lo cual constituyó una generalización sumamente útil para representar procesos con salto (particularmente adecuados para representar oscilaciones fuertes en el comportamiento del mercado de capitales).

si las funciones coeficientes son lineales respecto de $Y(t)$, $m(t,Y(t)) = a_0 + a_1 Y(t)$ y $\sigma(t,Y(t)) = c_0 + c_1 Y(t)$, queda definida la **ecuación diferencial estocástica de Itô; iii)** si los coeficientes a_0, a_1, c_0 y c_1 son funciones continuas del tiempo, queda definida la **ecuación estocástica lineal general**[91]:

$$Y(t) = Y(0) + \int_0^t \left[a_0(\tau) + a_1(\tau)Y(\tau)\right]d\tau + \int_0^t \left[c_0(\tau) + c_1(\tau)Y(\tau)\right]dB(\tau,w) \qquad \left(t \in [0,T]\right)$$

iv) si en la ecuación anterior se verifica que $c_1(t) = 0$, queda definida la **ecuación lineal con perturbaciones aditivas:**

$$Y(t) = Y(0) + \int_0^t \left[a_0(\tau) + a_1(\tau)Y(\tau)\right]d\tau + \int_0^t c_0(\tau)dB(\tau,w) \qquad \left(t \in [0,T]\right)$$

v) si se verifica que $a_0(t) = c_0(t) = 0$ para todo t, queda definida la **ecuación lineal homogénea con perturbaciones multiplicativas:**

$$Y(t) = Y(0) + \int_0^t a_1(\tau)Y(\tau)d\tau + \int_0^t c_1(\tau)Y(\tau)dB(\tau,w) \qquad \left(t \in [0,t]\right)$$

(obsérvese que en esta ecuación el proceso $\{Y(t)\}$ y los incrementos del movimiento Browniano, $\{B(t,w)\}$, figuran como factores)[92].

---

[91]. Obsérvese que, en este caso general, aplicando m.a m. el operador esperanza matemática, se obtiene que $E[Y(t)] = E[Y(0)] + \int_0^t \{a_0(\tau) + a_1(\tau)E[Y(\tau)]\}d\tau$, expresión que corresponde a la ecuación diferencial general lineal de la forma $\frac{d}{dt}E[Y(t)] = a_0(t) + a_1(t)E[Y(t)]$. A partir de la extensión-II del lema de Itô, se obtiene que $Y^2(t) = Y^2(0) + \int_0^t \{2Y(\tau)[a_0(\tau) + a_1(\tau)Y(\tau)] + [c_0(\tau) + c_1(\tau)Y(\tau)]^2\}d\tau + \int_0^t 2Y(\tau)[c_0(\tau) + c_1(\tau)Y(\tau)]dB(\tau,w)$. Aplicando, entonces, m.a m. el operador esperanza matemática, se obtiene que $E[Y^2(t)] = E[Y^2(0)] + \int_0^t \{[2Y(\tau) + c_1^2(\tau)]E[Y^2(\tau)] + 2[a_0(\tau) + c_0(\tau)c_1(\tau)]E[Y(\tau)] + c_1^2(\tau)\}d\tau$. Esta es la expresión integral de la siguiente ecuación diferencial general lineal para $E[Y(t)]$ $\frac{d}{dt}E[Y(t)] = [2a_1(t) + c_1^2(t)]E[Y(t)] + 2[a_0(t) + c_0(t)c_1(t)]E[Y(t)] + c_1^2(t)$ (debe tenerse en cuenta que $E[Y(t)]$ es conocida).

[92]. Como se verá inmediatamente la aplicación más importante de esta ecuación homogénea, cuya solución es de la forma $Y(t) = Y(0)\exp\left\{\int_0^t \left[a_1(\tau) - \frac{c_1^2(\tau)}{2}\right]d\tau + \int_0^t c_1(\tau)dB(\tau,w)\right\}$ $(\tau \in [0;T])$ a la econometría del mercado financiero, es el **movimiento Browniano geométrico**. Debe tenerse en cuenta que el proceso $\{Z(t)\} = \{\exp[B(t)]\}$ es tal que $Z(t) = Z'(t) = Z''(t)$, de modo que,

La ecuación de Itô admite una solución "débil" y una solución "fuerte"[93]. La **solución débil** se basa en el supuesto que la condición inicial, $Y(0)$ y las funciones $m(t,Y(t))$ y $\sigma_v(t,Y(t))$ están dadas y consiste en hallar el proceso $B(t,w)$ que satisfaga la ecuación. Esta solución es suficiente para estimar la esperanza matemática, la varianza y la función de autocovarianzas de $\{Y(t)\}$. La **solución fuerte**, basada en la trayectoria del movimiento Browniano directriz, consiste en un proceso $\{Y(t)\}$ que es función de dicha trayectoria y de las funciones que definen a los coeficientes $m(t,Y(t))$ y $\sigma_v(t,Y(t))$. De modo que, si se sustituye un movimiento Browniano por otro, se obtienen soluciones diferentes caracterizadas por la misma relación funcional pero referidas a distintos movimientos Brownianos [94]

Se demuestra que las siguientes condiciones: **i)** que la condición inicial, $Y(0)$, sea tal que $E(Y^2(0)) < \infty$ e independiente de $B(t,w)$; **ii)** que los coeficientes $m(t,Y(t))$ y $\sigma_v(t,Y(t))$ sean funciones continuas y tales que satisfagan una **condición de Lipschitz** con respecto a la variable $Y(t)$, es decir, que las funciones coeficientes sean tales que $|m(t,y_1) - m(t,y_2)| + |\sigma_\varepsilon(t,y_1) - \sigma_\varepsilon(t,y_2)| \le k|y_1 - y_2|$, son

---

para $t_1 < t_2$, la correspondiente expresión del lema de Itô-Doeblin será de la forma $Z(t_2) - Z(t_1) = e^{B(t_2)} - e^{B(t_1)} = \int_{t_1}^{t_2} e^{B(\tau)}\, dB(\tau) + \frac{1}{2}\int_{t_1}^{t_2} e^{B(\tau)}\, d\tau$, en la que la integral de Riemann que figura en el segundo miembro es positiva. Luego, se puede concluir que $\{Z(t)\}$ no es un proceso de Itô exponencial.

[93]. Una forma de resolución de la ecuación diferencial estocástica de Itô es utilizando el denominado **cálculo de Stratonovich**. El punto de partida de este método es la **integral estocástica de Stratonovich** (ver Stratonovich, R.L. (1966)) $\int_0^T f[B(\tau,w)] \circ dB(\tau,w)$ (donde $f(\bullet)$ denota una función diferenciable de segundo orden en el intervalo $[0,T]$), cuya vinculación con la integral de Itô puede ser expresada de la siguiente forma $\int_0^T f[B(\tau,w)] \circ dB(\tau,w) = \int_0^T f[B(\tau,w)]\, dB(\tau,w) + \frac{1}{2}\int_0^T f'[B(\tau,w)]\, d\tau$. Debe tenerse en cuenta que la ventaja en la utilización de la integral de Stratonovich radica en que satsisface la regla de encadenamiento de la diferenciación, es decir, se verifica que $\int_0^T g'[B(\tau,w)] \circ dB(\tau,w) = g[B(T,w) - g[B(0),w]]$ (esta propiedad no significa que la integral de Stratonovich pueda ser considerada una integral clásica en el sentido de Riemann, indica solamente que las reglas de encadenamiento tienen la misma estructura). La ecuación de Stratonovich equivalente a la ecuación de Itô que figura en el texto (en el sentido que, suponiendo que exista, tiene la misma solución fuerte) es de la forma $Y(t) = Y(0) + \int_0^t m^*(\tau,Y(\tau))\, d\tau + \int_0^t \sigma_v(\tau,Y(\tau)) \circ dB(\tau,w)$ (donde $m^*(t,y) = m(t,y) - \frac{1}{2}\sigma_v(t,y)\frac{\partial}{\partial y}\sigma_v(t,y)$). La representación del proceso mediante una ecuación diferencial estocástica de Stratonovich es más conveniente si las perturbaciones $v(t)$ observan un comportamiento "suficientemente regular" (condición que podría asimilarse a la diferenciabilidad), de modo que se pueda asegurar la validez de las reglas clásicas del cálculo.

[94]. Debe tenerse en cuenta que existen ecuaciones diferenciales de Itô que admiten solamente soluciones débiles (ver Shung, K.L.; Williams, R.J. (1990)). En realidad, las ecuaciones diferenciales estocásticas que admiten una solución explícita constituyen una excepción y, en consecuencia, en la mayoría de los casos se requiere una solución numérica que permita: **i)** obtener un conjunto de trayectorias muestrales (a este conjunto se lo suele mencionar como un **escenario**) y **ii)** obtener una aproximación razonable a las características de la correspondiente distribución de probabilidades (ver Kloeden, P.E.; Platen, E. (1992), Kloeden, P.E.; Platen, E.; Schurz, H. (1994), Mikosch, T. (2006),).

suficientes para asegurar que la solución fuerte de la ecuación de Itô es única[95].

Los procesos de difusión pueden ser estimados: **i)** utilizando una aproximación por quasi-máxima verosimilitud, a partir del supuesto que $\{dB(t)\}$ está Normalmente distribuido[96]; **ii)** utilizando métodos de los momentos[97]; **iii)** utilizando métodos no-paramétricos[98]; **iv)** utilizando métodos semiparamétricos y de reproyección[99]: **v)** utilizando métodos de simulación de Monte Carlo en cadenas de Markov[100].

4.7.- Procesos estocásticos y mercado de capitales

4.7.1.- Las hipótesis clásicas sobre el comportamiento del mercado de capitales

Los **supuestos fundamentales** sobre los que se desarrolló **la teoría de los modelos estocásticos aplicada al mercado de capitales** fueron los siguientes[101]:

i) Que las rentabilidades se comportan de acuerdo con un **proceso estocástico.**

ii) Que las variables aleatorias que integran dicho proceso se generan cuando se produce una **modificación en el conjunto de información** con que cuenta el mercado.

iii) Que la incorporación de nueva información al mercado se verifica **en forma continua** y que dicha información **se incorpora en forma instantánea a los precios**[102]. Es decir, que el proceso estocástico $\{v_t\}$ representativo del comportamiento de las rentabilidades es **continuo en el dominio del tiempo.**

---

95. Ver Øksendahl, B. (1985), Kloeden, P.E.; Platen, E. (1992).

96. Ver Kessler, M. (1997).

97. Ver Conley, T.G.; Hansen, L.P.; Luttmer, E.G.J.; Scheikman, J.A. (1997).

98. Ver Ait-Sahalia, Y. (1996)(2002).

99. Ver Gallant, A.R.; Long, J.R. (1997), Gallant, A.R.; Tauchen, G. (1997).

100. Ver Eraker, B. (2001), Elerian, O.; Chib, S.; Shephard, N. (2001).

101. Estos supuestos -que se refieren al comportamiento de las rentabilidades, ya que los procesos que representan la evolución de los precios de un activo son no-estacionarios- constituyen, de acuerdo con Cont, A. (2001), los "stylized facts".

102. Esta condición implica suponer que el activo en cuestión no admite arbitrajes. Se dice que, en el momento t, un activo admite arbitrajes si se verifica que: **i)** su valor, al momento t, es no-positivo; **ii)** tiene probabilidad nula de asumir valores negativos en el futuro y **iii)** existe una probabilidad no-nula de asumir valores positivos en el futuro. Si se supone una tasa de interés libre de riesgo, r, se puede considerar el factor e^{-rt} para calcular los precios futuros descontados al momento actual. Entonces, para asegurar la ausencia de arbitrajes, se debe suponer que el proceso $\left\{Y_t e^{-rt}\right\}$ es una martingala que, por lo tanto, cumple la condición $d\left(Y(t)e^{-rt}\right) = e^{-rt}\,dY(t) + Y(t)\,d\left(e^{-rt}\right) = \left(\varphi_0 - r\right)Y(t)e^{-rt}\,dt + \sigma_\varepsilon Y(t)e^{-rt}\,dB(t)$ (debe tenerse en cuenta que la condición de ausencia de arbitraje es similar, pero no igual, a la de mercado eficiente). Ver Delbaen, F.; Schachermayer, W. (2006)

iv) Que las rentabilidades de un activo son **estocásticamente independientes**[103].

v) Que las variables aleatorias que representan el comportamiento de las rentabilidades **tienen todas la misma distribución de probabilidades, con varianza finita**. Es decir, que el proceso representativo del comportamiento de los precios de un activo es un **RW-1** y que su volatilidad puede ser estimada a partir de la observación de sus realizaciones anteriores[104].

vi) Que las rentabilidades tienen una **se distribuyen de acuerdo con una función Normal**. Si a la hipótesis de RW-1 del comportamiento de los precios de un activo se agrega, además, el supuesto de que las rentabilidades de un activo constituyen la agregación de un número grande de variables aleatorias independientes con varianza finita -que representan la opinión (lineal) de los operadores ante un incremento de la información-, en virtud de los postulados del teorema central del límite, se puede concluir que las rentabilidades están Normalmente distribuidas con función de densidad (ver Roberts, H.V. (1959), en Cootner, P.H. (1964))[105]:

$$f\big(v(t,\tau)\big) = \frac{1}{\sqrt{2\pi\left(\tau\sigma_v^2\right)}}\exp\left\{-\frac{1}{2}\left[\frac{v(t,\tau)-m}{\sqrt{\tau}\sigma_v}\right]^2\right\}$$

(donde σ_v^2 denota la volatilidad para un período unitario) para un intervalo de tiempo de longitud τ (debe recordarse que, como se mencionó en la Sec. 4.2, la hipótesis de estabilidad de las volatilidades implica que $\sigma^2\big[v(\tau)\big] = \tau\sigma^2$).

Con respecto a la **rentabilidad**, entendida como **variación de los precios** de un activo (suponiendo que el mismo no paga dividendos), históricamente la literatura sobre este particular menciona tres tipos de definiciones:

$$v^*(t) = Y(t) - Y(t-1)$$

$$v^{**}(t) = \frac{Y(t)-Y(t-1)}{Y(t-1)}$$

$$v(t) = \ln\left(\frac{Y(t)}{Y(t-1)}\right) = \ln\left(1+\frac{\nabla Y(t)}{Y(t-1)}\right) = \nabla\ln\big(Y(t)\big) = \ln\big(Y(t)\big) - \ln\big(Y(t-1)\big) = Y^*(t) - Y^*(t-1)$$

(donde $Y(t)$ denota la variable que representa el precio de una activo al momento t).

La primera definición presenta dos inconvenientes fundamentales: **i)** la dependencia de $v^*(t)$ de la unidad en la que están medidos los precios, con la consiguiente dificultad en las comparaciones de las

---

103. Un supuesto implícito en estas hipótesis (la **condición de liquidez**) considera que la conducta de los operadores no afecta el comportamiento de los precios del activo (es decir, que el monto de las transacciones no afecta el comportamiento del mercado).

104. La hipótesis RW fue propuesta por Kendall, M.G. (1953)(en Cootner, P.H. (1964)) y continuada por Granger, C,W,J,; Morgenstern, O. (1963)(en Cootner, P.H. (1964)), utilizando análisis espectral. Estos trabajos pueden considerarse una consecuencia de la publicación de Bachelier, L. (1900).

105. Como se verá eb la Sec. 4.7.7.2, la debilidad de esta condición radica en considerar que la distribución Normal es el único límite al cual converge una suma de variables aleatorias estandarizadas. En realidad, existe una exrtensión del teorema central del límite para cada miembro de la llamada familia de distribuciones de Lévy-.Khinchin (la cual incluye, además de las funciones Normal, de Poisson y de Cauchy, el conjunto de las distribuciones infinitamente divisibles). Ver Hoffmann; Jorgensen, J. (1994).

rentabilidades correspondientes a distintos activos, y **ii)** la proporcionalidad de las varianzas de dichas rentabilidades con respecto al nivel de precios[106]. Por esta razón los autores han preferido, en general, utilizar cualquiera de las dos expresiones restantes, $v^{**}(t)$ (conocida como **tasa de rentabilidad simple**) y $v(t)$ (llamada **tasa de rentabilidad compuesta**), la cual supone **capitalización continua**.

Si se tiene en cuenta que el valor de una unidad invertida en $\frac{1}{Y(t)}$ unidades de un activo, al precio $Y(t)$, al momento $(t+\tau)$ será igual a:

$$\frac{Y(t+\tau)}{Y(t)} = 1 + v^{**}(t,\tau) = e^{v(t,\tau)} = 1 + v(t,\tau) + \frac{1}{2!}v^2(t,\tau) + \frac{1}{3!}v^3(t,\tau) + \ldots$$

se puede concluir en forma inmediata que $v^{**}(t,\tau)$ y $v(t,\tau)$ definen **tasas de rentabilidad** para un período de longitud τ. En particular, para $v(t,\tau)$, será $e^{v(t,\tau)} = \lim_{n\to\infty}\left[1 + \frac{v(t,\tau)}{n}\right]^n$.

La justificación de la preferencia por la **rentabilidad continua** se basa, fundamentalmente, en su posibilidad de generalización de los resultados obtenidos en forma discreta al caso continuo en el tiempo y en la posibilidad de expresar rentabilidades sobre cualquier período como una función de las rentabilidades unitarias[107]:

$$e^{v(t,\tau)+v(t+\tau,\tau)} = \left[1 + v^{**}(t,\tau)\right]\left[1 + v^{**}(t+\tau,\tau)\right]$$

---

[106]. Obsérvese que, si bien para contrastar la hipótesis nula que la rentabilidad es un proceso $WN - N\left(0,\sigma_v^2\right)$, este test es aplicable al proceso $\left\{v_t^*\right\} = \left\{Y_t - Y_{t-1}\right\} = \left\{\nabla Y_t\right\}$, no resulta útil para el análisis del proceso $\left\{v_t^{**}\right\} = \left\{\frac{Y_t - Y_{t-1}}{Y_{t-1}}\right\} = \left\{\frac{\nabla Y_t}{Y_{t-1}}\right\}$, ya que la varianza $\sigma^2\left(\frac{\nabla Y_t}{Y_{t-1}}\right) = \frac{1}{Y_{t-1}^2}\sigma_v^2$ está afectada por heterocedasticidad condicionada. Obviamente tampoco resulta útil para el análisis del proceso $\left\{v_t\right\} = \left\{\ln\frac{Y_t}{Y_{t-1}}\right\}$.

[107]. A partir de estas definiciones se puede concluir que el proceso $\left\{v^{**}(t,\tau)\right\}$: WN – log *Normal* con parámetros $E\left[v^{**}(t)\right] =$ $= \exp\left\{\varphi_0 + \frac{\sigma_v}{2} - 1\right\} = m_{**}$ y $\sigma^2\left[v^{**}(t)\right] = \exp\left(2\varphi_0 + \sigma_v^2\right)\left(e^{\sigma_v^2} - 1\right) = \sigma_{**}^2$. Alternativamente se puede escribir $\left\{v(t)\right\}$: $WN -$ $- N\left\{\ln\left[\frac{m_{**}+1}{\sqrt{1+\left(\frac{\sigma_{**}^2}{m_{**}+1}\right)^2}}\right]; \ln\left(1+\left(\frac{\sigma_{**}^2}{m_{**}+1}\right)\right)^2\right\}$. Suponiendo que v_t^{**} esté Normalmente distribuida, dado que $v_{t,k\tau}^{**}\ (k = 1,2,\ldots)$ está definida por un producto de tasas unitarias v_t no se distribuye de acuerdo con una función Normal (para más detalles sobre la distribución log-Normal, ver Sec. 4.7.6.1).

Las rentabilidades sobre dos períodos están dadas por[108]:

$$v(t,2\tau) = \ln\left[\frac{Y(t+2\tau)}{Y(t)}\right] = \ln\big(Y(t+2\tau)\big) - \ln\big(Y(t)\big) =$$

$$= \Big[\ln\big(Y(t+2\tau)\big) - \ln\big(Y(t)\big)\Big] + \Big[\ln\big(Y(t+\tau)\big) - \ln\big(Y(t)\big)\Big] =$$

$$= v(t,\tau) + v(t+\tau,\tau)$$

De la misma forma, para $k \geq 2$, será $v(t,k\tau) = v(t,\tau) + v(t+\tau,\tau) + ... + v\big(t+(k-1)\tau,\tau\big)$.

Asimismo:

$$v^{**}(t,2\tau) = \frac{Y(t+2\tau) - Y(t)}{Y(t)} = v^{**}(t+\tau,\tau)\frac{Y(t+\tau)}{Y(t)} + v^{**}(t,\tau) =$$

$$= v^{**}(t+\tau,\tau) + v^{**}(t,\tau) + v^{**}(t+\tau,\tau)v^{**}(t,\tau)$$

Luego, en general, se puede escribir [109]:

$$1 + v^{*}(t+k\tau) = \frac{Y(t+k\tau)}{Y(t)} = \frac{Y(t+k\tau)}{Y\big(t+(k-1)\tau\big)}\frac{Y\big(t+(k-1)\tau\big)}{Y\big(t+(k-2)\tau\big)}\cdots\frac{Y(t+\tau)}{Y(t)} =$$

$$= \Big[1 + v^{*}(t,\tau)\Big]\Big[1 + v^{*}(t+\tau,\tau)\Big]...\Big[1 + v^{*}\big(t+(k-1)\tau,\tau\big)\Big] =$$

$$= \prod_{j=1}^{k-1}\Big[1 + v^{*}(t+j\tau,\tau)\Big]$$

Por otra parte, debe tenerse en cuenta que, a fin de resolver el problema de la violación de la **propiedad de sesgo limitado,** se suele considerar como representativo del comportamiento de los precios de un activo a un proceso de la forma $\{Y^{*}(t)\} = \{\ln(Y(t))\}$. Por lo tanto se puede concluir que la definición más adecuada será la correspondiente a la rentabilidad compuesta. Si a esto se agrega el supuesto que $\{v(t)\}: WN - N\big(0,\sigma_v^2\big)$, resulta que $\{Y^{*}(t)\} : RW - Normal$ y que, en consecuencia, $\{Y(t)\}$ es un proceso

---

[108]. De las expresiones que siguen en el texto se obtiene fácilmente que $\left|v(t,\tau) - v^{**}(t,\tau)\right| \leq 0.1$. Esta desigualdad, unida a las comprobaciones empíricas, permite concluir que ambas definiciones de rentabilidad proporcionan resultados parecidos. Como se verá en la Sec. 4.7.6, a partir de esta propiedad de linealidad, se han desarrollado tests sobre las posibilidades de predicción del comportamiento de las rentabilidades (ver Campbell. J.; Mankiw, N.G. (1987), Cochrane J. (1988), Lo, A.; MacKinlay, A.C. (1988), Poterba, J.; Summers, L. (1988), Richardson, M.; Stock, J. (1989), Faust, J. (1992), Richardson, M. (1993)).

[109]. Debe tenerse en cuenta que esta propiedad de la rentabilidad continua no es aplicable a una cartera formada por n activos, $v^{(c)}(t+\tau,\tau)$: $v^{(c)}(t+\tau,\tau) = \ln\left(\sum_{i=1}^{n}\alpha_i e^{v_i(t+\tau,\tau)}\right) \neq \sum_{i=1}^{n}\alpha_i v_i(t+\tau,\tau)$, donde: i) α_i denota la ponderación correspondiente al i-ésimo activo en la cartera (proporción invertida en el i-ésimo activo); ii) $\sum_{i=1}^{n}\alpha_i = 1$ y iii) $v_i(t+\tau,\tau)$ denota la rentabilidad del i-ésimo activo.

log-Normal, de modo que las variables $Y(t)$ no son simétricas, $As\big(Y(t)\big) = \left[e^{\sigma^2(Y(t))} + 2\right]\sqrt{e^{\sigma^2(Y(t))} - 1}$ (si las variables $Y(t)$ fueran simétricas se verificaría que: **i)** $me(Y) = e^{m(Y)}$ y **ii)** $E\big(Y(t)\big) = \exp\left[m\big(Y(t)\big) + \frac{\sigma^2\big(Y(t)\big)}{2}\right]$, que $E\big(Y(t)\big) > me\big(Y(t)\big)$ implica que la distribución log-Normal está afectada por una asimetría hacia la derecha o asimetría positiva y que esta asimetría será mayor cuanto mayor sea el valor de $\sigma^2\big(Y(t)\big)$). A este respecto, debe tenerse en cuenta que la función $Y(t) = e^{Y^*(t)}$ es no-negativa y que, de acuerdo a lo comentado en la Sec. 3.1.1, cuando las observaciones presentan una cota inferior y no presentan cota superior, se origina una asimetría hacia la derecha cuya corrección se logra aplicando la transformación de Box-Cox con $\lambda < 1$.

4.7.2.- Los modelos clásicos representativos de los precios de un activo

Sea un proceso $\{Y_t^*\} = \{\ln(Y_t)\}$ (donde Y_t denota el precio de un activo al momento t), que admite una representación de la forma:

$$Y_t^* = \varphi_0 + Y_{t-1}^* + v_t$$

donde: **i)** φ_0 denota la variación esperada del precio del activo y **ii)** $\{v_t\} = \{\ln(1 + v_t^{**})\}: WN - N(0, \sigma_v^2)$ denota el proceso de las rentabilidades compuestas-continuas que admite una representación de Bachelier (ver Bachelier, L. (1900) (en Cootner, P.H. (ed.) (1964))[110].

Como se demostró en la Sec. 4.5, este proceso RW admite una representación continua-continua en términos de un movimiento Browniano con tendencia lineal de la forma:

$$dY^*(t) = \frac{dY(t)}{Y(t)} = \varphi_0\, dt + \sigma_v\, dB(t)$$

(Y, en consecuencia, el proceso $\{Y(t)\} = \{e^{Y^*(t)}\}$ es un movimiento Browniano geométrico). La aplicación de la extensión-I del lema de Itô-Doeblin:

$$Y(t) = Y(0) + \int_0^t \left\{\frac{\partial}{\partial \tau} f\big[\tau, B(\tau)\big] + \frac{1}{2}\frac{\partial^2}{\partial B(\tau)} f\big[\tau, B(\tau)\big]\right\} d\tau + \int_0^t \frac{\partial}{\partial B(\tau)} f\big[\tau, B(\tau)\big] dB(\tau)$$

permite concluir que el proceso $\{Y(t)\}$ satisface una ecuación diferencial estocástica de la forma:

---

[110]. Se trata de la propuesta de L.Bachelier "Théorie de la speéculation" (1900), en la que H. Poincaré actuó como uno de los jurados. Su "Calcul des probabilités" (1912) está dedicado a la probabilidad continua y a las vinculaciones del proceso representativo del comportamiento de las rentabilidades de un activo con el movimiento Browniano. Sus principio fundamental se basaba en el supuesto de equilibrio en el comportamiento de este proceso afectado por infinitas causas, equivalente a la hipótesis que en la actualidad se conoce como "eficiencia del mercado" ("*El conjunto de operadores no puede esperar ni un alza ni una caída del mercado y, para cada precio, existen tantos compradores como vendedores (...) La esperanza matemática de un operador es nula*", lo cual generaba la independencia de sus variables aleatorias. Como ya se mencionó en la Sec. 4.5, Kolmogorov, A.N. (1931) criticó duramente la matemática de Bachelier y en su revisión del estado del arte en la teoría de la probabilidad de 1935, la tildó de incomprensible. No obstante debe tenerse en cuenta que sus trabajos contienen las propiedades básicas de los proceso de Markov y definiciones de tipos especiales de procesos como los hoy denominados procesos estocásticos.

$$Y(t) = Y(0) + \varphi_0^* \int_0^t f\left[\tau, B(\tau)\right] d\tau + \sigma_v \int_0^t f\left[\tau, B(\tau)\right] dB(\tau)$$

(obsérvese que, de acuerdo con la segunda integral del segundo miembro, los procesos $\{Y(t)\}$ y $\{B(t)\}$ están vinculados en forma multiplicativa). Si a esta condición se le agrega la de continuidad de las trayectorias del movimiento Browniano, queda definido el siguiente sistema de ecuaciones en-derivadas parciales:

$$\begin{cases} \varphi_0^* f(t,x) = \dfrac{\partial}{\partial t} f(t,x) + \dfrac{1}{2} \dfrac{\partial^2}{\partial x^2} f(t,x) \\ \sigma_v f(t,x) = \dfrac{\partial}{\partial x} f(t,x) \end{cases}$$

De la segunda ecuación se obtiene que $\sigma_v^2 f(t,x) = \dfrac{\partial^2}{\partial x^2} f(t,x)$, de modo que este sistema puede ser escrito de la siguiente forma:

$$\begin{cases} \left(\varphi_0^* - \dfrac{1}{2}\sigma_v^2\right) f(t,x) = \dfrac{\partial}{\partial t} f(t,x) \\ \sigma_v f(t,x) = \dfrac{\partial}{\partial x} f(t,x) \end{cases}$$

Expresando la función $f(t,x)$ como un producto de funciones de la forma $f(t,x) = g(t)h(x)$, entonces el sistema asume la forma:

$$\begin{cases} \left(\varphi_0^* - \dfrac{\sigma_v^2}{2}\right) g(t) = g'(t) \\ \sigma_v h(x) = h'(x) \end{cases}$$

y sus ecuaciones pueden ser resueltas en forma independiente:

$$g(t) = g(0) \exp\left[\varphi_0^* - \frac{\sigma_v^2}{2}\right] t$$

$$h(x) = h(0) e^{\sigma_v x}$$

Es decir, se obtiene la solución general:

$$f(t,x) = g(0)h(0) \exp\left[\left(\varphi_0^* - \frac{\sigma_v^2}{2}\right) + \sigma_v x\right]$$

Teniendo en cuenta que $Y(0) = f\left[0, B(0)\right] = f(0,0) = g(0)h(0)$, se obtiene la siguiente solución particular

para el movimiento Browniano geométrico[111]:

$$Y(t) = f\left[t, B(t)\right] = Y(0)\exp\left\{\left[\varphi_0^* - \frac{\sigma_v^2}{2}\right]t + \sigma_v B(t)\right\} = Y(0)\exp\left[\varphi_0 t + \sigma_v B(t)\right] \qquad \left(t \in [0,T]\right)$$

donde $\varphi_0 = \varphi_0^* - \frac{\sigma_v^2}{2}$. Luego se puede concluir que toda variación en el logaritmo del precio del activo entre los puntos t^* y $t\left(t^* > t\right)$ define una variable aleatoria con distribución Normal de la forma $N\left[\left(\varphi_0^* - \frac{\sigma_v^2}{2}\right)\left(t^* - t\right); \sigma_v^2\left(t^{*-t}\right)\right]$.

Bajo el supuesto que las rentabilidades son linealmente incorrelacionadas, se pueden obtener estimadores consistentes de los coeficientes φ_0^* y σ_v^2 de la forma $\bar{v} = \frac{1}{n}\sum_{t=1}^{n} v_t$; $\hat{\sigma}_v^2 = \frac{1}{t^* - t}\frac{1}{n-1}\sum_{t=1}^{n}\left(v_t - \bar{v}\right)^2$ y $\hat{\varphi}_0^* = \frac{\bar{v}}{t^* - t} + \frac{\hat{\sigma}_v^*}{2}$ (como se verá en las secciones siguientes, si las rentabilidades presentan autocorrelaciones significativas o los precios no admiten una representación mediante un movimiento Browniano, estos métodos de estimación deben ser oportunamente modificados).

Ahora bien, considerando el proceso:

$$B^*(t) = B(t) + \frac{\varphi_0 t}{\sigma_v} \qquad \left(t \in [0,T]\right)$$

la ecuación diferencial estocástica lineal anterior puede ser expresada de la siguiente forma:

$$\frac{dY(t)}{Y(t)} = \sigma_v\, d\left[\frac{\varphi_0 t}{\sigma_v} + \sigma_v B(t)\right] = \sigma_v\, dB^*(t) \qquad \left(t \in [0,T]\right)$$

Luego, de acuerdo con el teorema de Girasinov, el proceso $\left\{B(t)\right\}$ (que no presenta término independiente) es una martingala exponencial bajo la medida Q (siendo $Q(A) = \int_A K(t,w)\,dP(w)$ y $K(t) = \exp\left[-cB(t) - \frac{1}{2}c^2 t\right]$), pero no es una martingala bajo la medida P, cuya representación es de la forma[112]:

$$Y(t) = Y(0)\exp\left[\left[-\frac{\sigma_v^2 t}{2} + \sigma_b B(t)\right]\right] \qquad \left(t \in [0,T]\right)$$

---

111. Del desarrollo anterior se puede concluir que, como ya se mencionó en las páginas precedentes, en algunos casos, la ecuación diferencial estocástica de Itô puede ser resuelta como una ecuación en derivadas parciales determinística.

112. Ver Sec. 4.5.

(obsérvese que el cambio de medida permite la eliminación del término representativo de la tendencia en la ecuación diferencial estocástica)[113].

Una aplicación interesante del lema de Itô-Doeblin es la derivación de la fórmula de Black-Scholes-Merton mediante la utilización del cambio de medida en un movimiento Browniano. Denotando por $G(Y(t),t)$ el precio del derivado, se puede escribir $dG(Y(t),t) = \left[\frac{\partial}{\partial Y(t)}G(Y(t),t)\varphi_0^* Y(t) + \frac{\partial}{\partial t}G(Y(t),t) + \frac{1}{2}\frac{\partial^2}{\partial Y^2(t)}G(Y(t),t)\sigma_v^2 Y^2(t)\right]dt + \frac{\partial}{\partial Y(t)}G(Y(t),t)\sigma_v Y(t)dB(t)$. A partir de esta ecuación se obtiene la ecuación diferencial de Black-Scholes-Merton para valuar el precio de la opción $\frac{\partial}{\partial t}G(Y(t),t) + rY(t)\frac{\partial}{\partial Y(t)}G(Y(t),t) + \frac{1}{2}\sigma_v^2 Y^2(t)\frac{\partial^2}{\partial Y^2(t)}G(Y(t),t) = rG(Y(t),t)$, donde r denota la tasa de interés libre de riesgo (ver Harrison, M.J.. Pliska, S.R. (1981)).

Un resultado interesante con relación a esta martingala exponencial se refiere a la probabilidad de **arribo por primera vez a un estado** ω. Sea $T_\omega = \min\left[t \prec 0, B(t) = \omega\right]$ la variable que representa el tiempo a transcurrir por el proceso hasta su arribo al estado ω. Se demuestra[114] que la función de densidad de esta variable es de la forma

$$f_{T_\omega}(t) = \frac{|\omega|}{t\sqrt{2\pi t}}e^{-\omega^2/2t} \qquad (\omega > 0, t \geq 0)$$

y su valor esperado está definido por $E(T_\omega) = e^{-\sigma_v \omega}$ $(\sigma_v > 0, \omega > 0)$.

A partir de este resultado es posible considerar la variable que representa el número de trayectorias que arriban al estado ω en el momento t o antes del momento t (es decir, el número de trayectorias para las cuales $T_\omega \leq t$). Existen dos tipos de trayectorias de esta clase: **i)** aquellas que arriban al estado ω en un momento anterior a t y, al momento t, se encuentran en un estado inferior a ω y **ii)** aquellas que arriban al estado ω en un momento anterior a t y, al momento t, se encuentran en un estado superior a ω (también existen trayectorias que, al momento t, se encuentran en el estado ω, pero su probabilidad para procesos como el tratado aquí es nula). Para cada trayectoria del primer tipo es posible construir (intercambiando los movimientos hacia arriba y hacia abajo a partir del momento T_ω) una **trayectoria reflejada** que se encuentra en el estado $2\omega - \omega$ al momento t. En consecuencia, queda definida la siguiente **ecuación reflexiva**:

$$P\left[T_\omega \leq t, B(t) \leq \omega_0\right] = p\left[B(t) \geq 2\omega - \omega\right] \qquad (\omega_0 \leq \omega, \omega > 0)$$

Sea, por otra parte, un proceso $B^{(M)}(t) = \underset{0\leq\tau\leq t}{Max}\, B(\tau)$. Para $\omega > 0$, la condición $T_\omega \leq t$ es necesaria y suficiente para que se verifique que $B^{(M)}(t) \geq \omega$. Luego, la ecuación reflexiva puede ser

---

113. De lo que se puede concluir que, de acuerdo a lo expresado en la Sec. 4.1, φ_0^* representa la tasa de rentabilidad esperada del activo.

114. Ver Shreve, S.E. (2001).

expresada de la siguiente forma [115]:

$$p\left[B^{(M)}(t) \geq \omega, B(t) \leq \omega_0\right] = p\left[B(t) \geq 2\omega - \omega_0\right] \qquad (\omega_0 \geq \omega, \omega > 0)$$

y, a partir de esta expresión se demuestra que la función de densidad conjunta de la variable bidimensional $\left[B^{(M)}(t), B(t)\right]$ es de la forma:

$$f_{B^{(M)}(t),B(t)}(\omega,\omega_0) = \frac{2(2\omega - \omega_0)}{t\sqrt{2\pi t}} \exp\left[-\frac{(2\omega - \omega_0)^2}{2t}\right] \qquad (\omega_0 \leq \omega, \omega > 0)$$

y que la función de densidad de la variable condicionada $\left[B^{(M)}(t) = \omega \,/\, B(t) = \omega_0\right]$ está definida por:

$$\left[B^{(M)}(t) = \omega \,/\, B(t) = \omega_0\right] = \frac{2(2\omega - \omega_0)}{t} \exp\left[-\frac{2\omega(\omega - \omega_0)}{t}\right] \qquad (\omega_0 \leq \omega, \omega > 0)$$

Sean $B^{(1)}(t)$ y $B^{(2)}(t)$ dos movimientos Brownianos independientes, entonces el proceso bidimensional:

$$B^{(2d)}(t) = B^{(1)}(t) + B^{(2)}(t) = \frac{1}{\sqrt{\sigma_{(1)v}^2 + \sigma_{(2)v}^2}}\left[\sigma_{(1)v} B^{(1)}(t) - \sigma_{(2)v} B^{(2)}(t)\right]$$

es un movimiento Browniano.

Sea $\{Y(t)\}$ un proceso que representa el valor de una cartera formada por dos activos y que admite una representación en términos de una ecuación integral estocástica de Itô con un movimiento Browniano $B^{(2)}(t)$ como proceso directriz, de la forma:

$$Y(t) = Y(0) + \varphi_1^* \int_0^t Y(\tau)d\tau + \sigma_{(1)v} \int_0^t Y(\tau)dB^{(1)}(\tau) + \sigma_{(2)v} \int_0^t Y(\tau)dB^{(2)}(\tau) =$$

$$= \varphi_1^* \int_0^t Y(\tau)d\tau + \int_0^t Y(\tau)d\left[\sigma_{(1)v} B^{(1)}(\tau) + \sigma_{(2)v} B^{(2)}(\tau)\right] =$$

$$= \varphi_1^* \int_0^t Y(\tau)d\tau + \sqrt{\sigma_{(1)v}^2 + \sigma_{(2)v}^2} \int_0^t Y(\tau)dB^{(2)}(\tau)$$

(donde $\sigma_{(1)v}^2$ y $\sigma_{(2)v}^2$ denotan las volatilidades de cada uno de los activos). De una forma similar a la

---

115. Este proceso es utilizado en la valuación de las denominadas "barrier options". Para un análisis de la aplicación de las propiedades del movimiento Browniano a la valuación de opciones exóticas, ver Borodin, A.; Salminen, P. (1996).

expuesta en páginas anteriores, se obtiene, entonces, la siguiente solución[116]:

$$Y(t) = Y(0)\exp\left\{\left[\varphi_1^* - \frac{\sigma_{(1)v}^2 + \sigma_{(2)v}^2}{2}\right]t + \sqrt{\sigma_{(1)v}^2 + \sigma_{(2)v}^2}\,B^{(2)}(t)\right\} =$$

$$= Y(0)\exp\left\{\left[\varphi_1^* - \frac{\sigma_{(1)v}^2 + \sigma_{(2)v}^2}{2}\right]t + \left[\sigma_{(1)v}B^{(1)}(t) + \sigma_{(2)v}B^{(2)}(t)\right]\right\}$$

4.7.3.- El comportamiento de la rentabilidades como un proceso fractal

Todos los resultados anteriores permiten concluir, en primer lugar, que las hipótesis clásicas conducen al supuesto que el mercado de capitales es eficiente[117] y que los precios de los activos obedecen a un proceso Browniano geométrico (donde φ_0 coincide con la tasa de interés libre de riesgo).

Ahora bien, en términos de procesos estocásticos unidimensionales la hipótesis clásica que considera el comportamiento RW-log-Normal del precio de un activo (que implica que toda la información sobre el comportamiento pasado del mismo está resumida en su precio al momento t), supone solamente la condición de **eficiencia débil** del mercado (conocida también como autorregresividad débil de primer orden o **condición de Markov débil**)[118] la cual, como ya se mencionó en la Sec. 1.3, en la medida que no posee como condición necesaria la independencia de orden superior, no implica necesariamente una representación RW-1[119].

---

[116]. Estos resultados pueden ser generalizados a cualquier conjunto finito de movimientos Brownianos $B^{(i)}(t)\ (i = 1,2,\ldots,k)$ que no tienen por qué ser independientes.

[117]. Teniendo en cuenta que la condición de eficiencia del mercado implica solamente que las rentabilidades poseen distribuciones condicionadas por el conjunto de información disponible ($f\left(Y_t / \Omega_t(Y)\right)$) idénticas e independientes entre sí.

[118]. De acuerdo con Fama, E.F. (1970) se puede calificar como eficiente a todo mercado "*...en el cual los precios reflejan 'totalmente' la información disponible*". En forma más completa, Malkiel, B. (en Newman, P.; Milgate, M.; Eatwell, J. (eds.)(1992)) caracteriza a un mercado de capitales como eficiente cuando "*refleja en forma total y correcta toda información relevante para la determinación de los precios de los activos*" y agrega: "*Se dice que el mercado es formalmente eficiente con respecto a un conjunto de información dado...si los precios de los activos no se ven afectados cuando dicho conjunto de información es puesto en conocimiento de todos los participantes del mercado. Además, la eficiencia con respecto a un conjunto de información...implica la imposibilidad de obtener beneficios económicos operando en base a dicho conjunto de información*" y se podría agregar, cuando las futuras variaciones de los precios pueden solamente resultar "novedades" impredecibles, de modo que la mejor predicción de los precios es simplemente el precio de hoy. Teniendo en cuenta que, como se mencionó en la página precedente, la condición de eficiencia del mercado implica solamente que las rentabilidades poseen poseen distribuciones condicionadas por el conjunto de información disponible ($f\left(Y_t / \Omega_t(Y)\right)$) idénticas e independientes entre sí. Roberts, H. (1967), a partir de la caracterización de tres tipos de conjuntos de información, define las siguientes condiciones de eficiencia del mercado: **i) débil**, la cual implica que no es posible predecir los precios futuros a partir, exclusivamente, de la serie cronológica formada por los precios pasados; **ii) semi-fuerte**, la cual implica que no es posible predecir precios futuros a partir de los precios pasados ni de cualquier otro tipo de información "de dominio público", y **iii) fuerte**, la cual implica que no es posible predecir los precios futuros a partir de los datos mencionados en el punto anterior, ni de cualquier otro tipo de "inside information" (obsérvese que,las definiciones de eficiencia del mercado pueden ser asimiladas a las definiciones de filtraciones, natural y aumentadas, analizadas en el Cap. 1, paradójicamente, una mayor eficiencia implica una mayor aleatoriedad en el proceso de las variaciones de los precios, la eficiencia fuerte implica aleatoriedad pura). Para una revisión histórica sobre la hipótesis de eficiencia del mercado, ver LeRoy, S. (1989) y, para una revisión sobre la bibliografía vinculada con las hipótesis de eficiencia del mercado, ver Cuthbertson, K. (1996).

[119]. Debe tenerse en cuenta que, dado el método de cálculo de los precios de un activo y de los índices de rentabilidad, pueden presentarse autocorrelaciones significativas en las series cronológicas representativas de las rentabilidades (particularmente cuando el período en que está expresada la serie es pequeño). El análisis de las rentabilidades de un activo a partir de la hipótesis de comportamiento RW-1 de los precios implica no considerar restricciones asociadas a teorías económicas o financieras (la teoría de los mercados de capitales eficientes se basa en la aplicación del principio Ricardiano de la ventaja comparativa, que constituye el fundamento de la teoría del equilibrio competitivo).

Esta relación de implicación solamente en el sentido "si RW, entonces eficiencia" permitiría concluir que, en realidad, no fue la hipótesis de eficiencia del mercado la que condujo al supuesto de RW sino exactamente el caso inverso, que la aceptación de la hipótesis de eficiencia fue necesaria para justificar el supuesto -erróneo- de que los precios de los activos se comportaban de acuerdo a un proceso RW-Normal. Si bien este esquema es absolutamente injustificable desde el punto de vista epistemológico, es necesario reconocer que la adopción del supuesto RW-Normal permitió la introducción de la teoría de la probabilidad y del análisis estadístico en el estudio de la evolución de los precios de los activos.

En segundo lugar, se puede concluir que la evolución de la rentabilidad de un activo es un fenómeno dinámico que posee dimensión fractal y una trayectoria compleja que, de acuerdo con los supuestos clásicos mencionados en la sección precedente es aleatoria pura[120], pero que podría ser caótica.

Con respecto a la propiedad de fractalidad, debe tenerse en cuenta que, desde un punto de vista conceptual, no existe ninguna vinculación entre los dominios de los sistemas dinámicos y de la geometría fractal: la dinámica se refiere al estudio de los "objetos" en movimiento y los fractales son "objetos geométricos" que definen imágenes estáticas. No obstante ocurre que muchas regiones caóticas correspondientes a sistemas dinámicos son fractales. Como se verá en la **Parte III**, un atractor caótico puede ser definido como una trayectoria en un espacio de fase que permanece en un conjunto fractal. A partir de esta relación, los fractales pueden ser definidos como instrumentos matemáticos que permiten describir, en términos de estructuras simples, los comportamientos complejos de los sistemas dinámicos caóticos.

Un modo primario de formalización del concepto de caos está relacionado con la noción de sensibilidad del proceso a sus condiciones iniciales[121]. Como se expresó en la Sec. 1.1, un conocimiento más detallado acerca del comportamiento de los sistemas dinámicos (en particular, de los sistemas de comportamiento caótico) puede ser obtenido a partir del estudio de sus conjuntos atractores.

4.7.4.- El proceso de Uhlenbeck-Ornstein

Como una variación del proceso tratado en la sección precedente, sea un proceso $\{Y(t)\}$ que admite una representación $AR(1)$ de la forma:

$$Y_t = \varphi_1 Y_{t-1} + v_t = \left(\varphi_1^* + 1\right) Y_{t-1} + v_t$$

o, alternativamente:

---

120. Por lo menos en términos de procesos unidimensionales.

121. Una definición matemáticamente más completa (debida a Devaney, R.I. (1989) indica que un proceso caótico posee las siguientes propiedades: **i)** sensibilidad a las condiciones iniciales; **ii)** transitividad topológica y **iii)** densidad de sus puntos periódicos. Supóngase que la dinámica del sistema esté definida por el ya mencionado modelo evolutivo simple $Y_t = f_\varphi\left(Y_{t-1}\right)$. La condición necesaria, pero no suficiente, para la generación de un comportamiento caótico es que $f_\varphi(\bullet)$ sea una función no-lineal (las representaciones no-lineales serán tratadas en la Parte III). Se dice que f_φ presenta sensibilidad a las condiciones iniciales, si realizaciones del proceso correspondientes a condiciones iniciales distintas aunque muy próximas entre sí, divergen. En términos formales, si existe un $\delta > 0$ tal que, para todo entorno W de Y, es posible hallar un $Y^* \in W$ y un $n \geq 0$, tal que se verifique que $\left\| f_\varphi^{(n)}(Y) - f_\varphi^{(n)}(Y^*) \right\| = \underset{n \geq 0}{Max} \left| f_\varphi^{(n)}(Y) - f_\varphi^{(n)}(Y^*) \right| > \delta$. Esta condición implica necesariamente una disminución exponencial de la precisión de las potenciales predicciones sobre el comportamiento del proceso, a medida que n crece. Esta característica se conoce, en la nomenclatura de P. Diaconis, como el **fenómeno de la multiplicación de la aleatoriedad**. La teoría del caos modificó las expectativas respecto de la proporcionalidad entre causa y efecto. En los paradigmas clásicos se supone que una "causa pequeña" debe estar asociada a un "efecto pequeño". En un sistema complejo, pequeñas variaciones en las fórmulas iterativas o en las condiciones iniciales pueden provocar diferencias de gran magnitud en el estado final.

$$\nabla Y_t = Y_t - Y_{t-1} = \varphi_1^* Y_{t-1} + v_t$$

(donde $\{v_t\} = \{\sigma_v(B_t - B_{t-1})\}: WN - N(0,\sigma_v^2)$). La correspondiente representación continua en el dominio del tiempo será, entonces, de la forma:

$$dY(t) = \varphi_1^* Y(t)dt + \sigma_v\, dB(t)$$

La aplicación de la extensión-I del lema de Itô-Doeblin permite concluir que el proceso $\{Y(t)\}$ satisface una ecuación diferencial estocástica de la forma (conocida como **ecuación de Langevin**)[122]:

$$Y(t) = Y(0) + \varphi_1^* \int_0^t Y[\tau, B(\tau)]d\tau + \sigma_v \int_0^t dB(\tau) \qquad (t \in [0,T])$$

A los efectos de la resolución de esta ecuación es conveniente utilizar una transformación de la forma $Z(t) = e^{-\varphi_1^* t} Y(t)$, para la cual se verifica que $Z(0) = Y(0)$ (esta transformación y la condición de Normalidad hacen que el proceso $\{Z(t)\}$ sea estrictamente estacionario). Aplicando la extensión-II del lema de Itô-Doeblin en la que: **i)** $f(t,x) = e^{-\varphi_1^* t} x$; **ii)** $\frac{\partial}{\partial t} f(t,x) = -\varphi_1^* f(t,x)$; **iii)** $\frac{\partial}{\partial x} f(t,x) = e^{-\varphi_1^* t}$; **iv)** $\frac{\partial^2}{\partial x^2} f(t,x) = 0$; **v)** $A^{(1)} = \varphi_1^* Y(t)$ y **vi)** $A^{(2)} = \sigma_v$, se obtiene que:

$$Z(t) = Z(0) + \int_0^t \left\{ \frac{\partial}{\partial t} f[\tau, Y(\tau)] + \varphi_1^* Y(t) \frac{\partial}{\partial Y} f[\tau, Y(\tau)] + \frac{1}{2}\sigma_v^2 \frac{\partial^2}{\partial Y^2} f[\tau, Y(\tau)] \right\} + \int_0^t \sigma_v f[\tau, Y(\tau)] dB(\tau) =$$

$$= \int_0^t \left[-\varphi_1^* Z(\tau) + \varphi_1^* Z(\tau) + 0 \right] d\tau + \int_0^t \sigma_v e^{-\varphi_1^* t}\, dB(\tau) =$$

$$= \int_0^t \sigma_v e^{-\varphi_1^* t}\, dB(\tau)$$

De modo que la solución de la ecuación de Langevin será de la forma:

$$Y(t) = e^{\varphi_1^* t} Y(0) + \sigma_v e^{\varphi_1^* t} \int_0^t e^{-\varphi_1^* t}\, dB(\tau)$$

Dado que la ecuación de Langevin es una ecuación diferencial estocástica lineal, de acuerdo a lo demostrado en la Sec. 4.6, se puede concluir que esta solución es única.

En particular, si la condición inicial $Y(0)$ es constante, queda definido un proceso Normal,

---

122. Esta ecuación fue propuesta (muchos años antes del desarrollo de la teoría general de las ecuaciones diferenciales estocásticas) por Langevin, P. (1908) para representar la velocidad de una partícula Browniana. Obsérvese que, contrariamente a lo que ocurre con el proceso analizado en la sección precedente, los procesos $\{Y(t)\}$ y $\{B(t)\}$ no están relacionados directamente en la integral de Itô de esta ecuación diferencial.

conocido como **proceso de Uhlenbeck-Ornstein**[123].

Suponiendo (por razones de mayor simplicidad en las demostraciones) que $Y(0) = 0$, se obtiene que $E[Y(t)] = 0$,

$$\sigma^2[Y(t)] = \frac{\sigma_v^2}{2\varphi_1^*}\left(e^{2\varphi_1^* t} - 1\right)$$

y

$$\gamma[Y(t_1), Y(t_2)] = \frac{\sigma_v^2}{2\varphi_1^*}\left[e^{\varphi_1^*(t_2+t_1)} - e^{\varphi_1^*(t_2-t_1)}\right] \qquad (t_1 < t_2)$$

(ver Dana, R.A.; Jeanblanc, M. (2003), Prigent, J.L. (2003)).

4.7.5.- Acerca de la estructura temporal de la tasa de interés

De acuerdo con el **modelo de Vasicek**, la tasa de interés instantánea ($r(t)$) admite una representación de la forma[124]:

$$dr(t) = c[m(r) - r(t)]dt + \sigma\, dB(t) \qquad (t \in [0,T]; c > 0; m(r) > 0; \sigma > 0)$$

o, en términos de una ecuación diferencial estocástica lineal de la forma:

$$r(t) = r(0) + c\int_0^t [m(r) - r(\tau)]d\tau + \sigma\int_0^t dB(\tau) \qquad (t \in [0,T]; c > 0; m(r) > 0; \sigma > 0)$$

donde: **i)** $m(r)$ denota el valor alrededor del cual el proceso $r(t)$ evoluciona por fluctuaciones; **ii)** c denota la velocidad de reversión a $m(r)$ de $r(t)$; **iii)** σ denota la volatilidad, es decir, la medida de las fluctuaciones de $r(t)$ con respecto a $m(r)$. La solución general de esta ecuación es de la forma:

$$r(t) = r(0)e^{-ct} + m(r)\left(1 - e^{-ct}\right) + \sigma e^{-ct}\int_0^t e^{ct}\, dB(t)$$

Haciendo:

$$f(t,x) = r(0)e^{-ct} + m(r)\left(1 - e^{-ct}\right) + \sigma e^{-ct}\int_0^t e^{c\tau}\, dB(\tau)$$

---

123. Este proceso, debido a Uhlenbeck, G.E.; Ornstein, L.S. (1930) y basado en la versión de P. Langevin de la segunda ley del movimiento de Newton en (ver Breiman, L. (1968), Bhattacharya, R.N.; Waymire, E.C. (1990)), constituye un modelo más adecuado para el estudio del movimiento de partículas suspendidas en un líquido. Dado que sus trayectorias son diferenciables, como se mencionó en la página precedente, proporciona una medida de la velocidad de la partícula.

124. Ver Vasicek, O. (1977).

y subdividiendo el segundo miembro de la expresión anterior en una función no-aleatoria de las variables t y x y un proceso de Itô de la forma $X(t) = \int_0^t e^{c\tau}\, dB(\tau)$, se obtiene que:

$$\frac{\partial}{\partial t} f(t,x) = -c e^{-ct} r(0) + c m(r) e^{-ct} - \sigma c e^{-ct} x = c\left[m(r) - f(t,x)\right]$$

$$\frac{\partial}{\partial x} f(t,x) = \sigma e^{-ct}$$

$$\frac{\partial^2}{\partial x^2} f(t,x) = 0$$

$$dX(t) = e^{ct}\, dB(t)$$

Teniendo en cuenta que $f\left[0, X(0)\right] = r(0)$, se puede concluir que $f\left[t, X(t)\right]$ satisface la ecuación diferencial estocástica de Vasicek para todo $t \geq 0$.

Si se supone que $r(0)$ es una constante, entonces se puede concluir que $\{r(t)\}$ es un proceso Normal y, si $m(r) = 0$, queda definido un proceso de Uhlenbeck-Ornstein. Además, se obtiene en forma inmediata que $E\left[r(t)\right] = r(0)e^{-ct} + m(r)\left(1 - e^{-ct}\right)$ y $\sigma^2\left[r(t)\right] = \frac{\sigma^2}{2c}\left(1 - e^{-2ct}\right)$. A partir de estos resultados se concluye fácilmente que, cuando $t \to \infty$:

$$r(t) \xrightarrow{d} N\left[m(r); \frac{\sigma^2}{2c}\right]$$

Obsérvese que: **i)** si $r(t) = m(r)$, el término representativo de la tendencia es nulo; **ii)** si $r(t) > m(r)$, dicho término es negativo; **iii)** si $r(t) < m(r)$, dicho término es positivo (en estos dos últimos casos se puede decir que el término representativo de la tendencia "empuja" a $r(t)$ hacia $m(r)$, es decir, de acuerdo con este modelo, la tasa de interés es un fenómeno con reversión al valor medio); **iv)** si $r(0) = m(r)$, entonces $E\left[r(t)\right] = m(r)\ (t \geq 0)$ y **v)** si $r(0) \neq m(r)$, entonces $\lim_{t \to \infty} E\left[r(t)\right] = m(r)$.

Este modelo presenta el inconveniente de admitir una probabilidad no-nula de que $r(t)$ sea negativa. A fin de resolver este problema, Cox-Ingersoll-Ross[125] propusieron un modelo de la forma:

$$dr(t) = c\left[m(r) - r(t)\right]dt + \sigma\sqrt{r(t)}\, dB(t) \qquad \left(t \in [0,T]; c > 0; m(r) > 0; \sigma > 0\right)$$

que no admite una solución única y según el cual, como en el caso anterior, el proceso $r(t)$ posee la propiedad de reversión al valor medio.

A partir de la función $f(t,x) = xe^{ct}$ y de la fórmula de Itô-Doeblin, se obtiene que:

---

[125]. Ver Cox, J.C.; Ingersoll, J.E.; Ross, S. (1985).

$$d\left[e^{ct}r(t)\right] = df\left[t,r(t)\right] =$$

$$= c\,e^{ct}r(t)dt + e^{ct}c\left[m(r) - r(t)\right]dt + e^{ct}\sigma\sqrt{r(t)}\,dB(t) =$$

$$= cm(r)e^{ct}\,dt + \sigma\,e^{ct}\sqrt{r(t)}\,dB(t)$$

Integrando m. a m. esta expresión resulta que:

$$e^{ct}r(t) = r(0) + cm(r)\int_0^t e^{c\tau}\,d\tau + \sigma\int_0^t e^{c\tau}\,d\tau + \sigma\int_0^t e^{c\tau}\sqrt{r(\tau)}\,dB(\tau) =$$

$$= r(0) + c\left(e^{ct} - 1\right) + \sigma\int_0^t e^{c\tau}\sqrt{r(\tau)}\,dB(\tau)$$

De modo que, teniendo en cuenta que el valor esperado de una integral de Itô es nulo, igual que en el modelo de Vasicek, se obtiene que:

$$E\left[r(t)\right] = e^{-ct}r(0) + m(r)\left(1 - e^{-ct}\right)$$

Asimismo, haciendo $X(t) = e^{ct}r(t)$, se obtiene que:

$$dX(t) = cm(r)e^{ct}\,dt + \sigma\,e^{ct}\sqrt{r(t)}\,dB(t) = cm(r)e^{ct}\,dt + \sigma\,e^{ct/2}\sqrt{X(t)}\,dB(t)$$

y que $E\left[X(t)\right] = r(0) + m(r)\left(e^{ct} - 1\right)$.

De acuerdo con la fórmula de Itô-Doeblin, se obtiene que:

$$d\left[X^2(t)\right] = 2cm(r)e^{ct}X(t)dt + 2\sigma\,e^{ct/2}\left[X(t)\right]^{3/2}dB(t) + \sigma^2\,e^{ct}X(t)dt$$

Integrando m. a m. esta expresión, resulta que:

$$X^2(t) = X^2(0) + \left[2cm(r) + \sigma^2\right]\int_0^t e^{c\tau}X(\tau)d\tau + 2\sigma\int_0^t e^{c\tau/2}X^{3/2}(\tau)dB(\tau)$$

y, en consecuencia, que:

$$E\left[X^2(t)\right] = X^2(0) + \left[2cm(r) + \sigma^2\right]\int_0^t e^{c\tau}E\left[X(\tau)\right]d\tau =$$

$$= r^2(0) + \left[2cm(r) + \sigma^2\right]\int_0^t e^{c\tau}\left[r(0) + m(r)\left(e^{c\tau} - 1\right)\right]d\tau =$$

$$= r^2(0) + \frac{2cm(r) + \sigma^2}{c}\left[r(0) - m(r)\right]\left(e^{ct} - 1\right) + \frac{2cm(r) + \sigma^2}{c}m(r)\left(e^{ct} - 1\right)$$

De modo que:

$$E\left[r^2(t)\right] = e^{-2ct}E\left[X^2(t)\right] =$$

$$= e^{-2ct}r^2(0) + \frac{2cm(r)+\sigma^2}{c}\left[r(0)+m(r)\right]\left(e^{-ct}-e^{-2ct}\right) + \frac{m(r)\left[2cm(r)+\sigma^2\right]}{2c}\left(1-e^{-2ct}\right)$$

y,

$$\sigma^2\left[r(t)\right] = \frac{\sigma^2}{c}r(0)e^{-ct}\left(1-e^{-ct}\right) + \frac{m(r)\sigma^2}{2c}\left(1-2e^{-ct}+e^{-2ct}\right)$$

De esta última expresión, se obtiene que:

$$\lim_{t\to\infty}\sigma^2\left[r(t)\right] = \frac{m(r)\sigma^2}{2c}.$$

Para un análisis más detallado de la estructura temporal de la tasa de interés, ver Ho, T.; Lee, S. (1986), Chan, K.; Karoly, A.; Longstaff, F.; Sanders, A. (1992), Chen, R.; Scott, L. (1992)(1993) (1995), Duffie, D.; Kan, R. (1994), Chen, L. (1996), Brace, A.; Gątarek, D.; Musiela, M. (1997), Brigo, D.; Mercurio, F. (2001).

4.7.6.- La estimación de la volatilidad

4.7.6.1.- La volatilidad histórica

Con respecto a las volatilidades, debe tenerse en cuenta que sólo pueden ser estimadas en el contexto de un proceso que represente el comportamiento de las rentabilidades[126] y que las estimaciones de los precios de opciones basadas en las **estimaciones históricas** de las volatilidades resultan sistemáticamente sesgadas[127]. Lo cual sugiere que el proceso representativo de las rentabilidades es no-estacionario y que, en consecuencia, contrariamente a las recomendaciones de la teoría clásica, la extensión del período de observación no mejora la calidad de la estimación[128].

Por otra parte, dado que, como se vio, la fórmula de Black-Scholes-Merton es no-lineal respecto de la varianza de las variaciones del precio del activo: **i)** aún una estimación insesgada de la volatilidad puede no producir una estimación insesgada del precio de la opción; **ii)** la relación entre la volatilidad y

---

126. Como se comentó en la Sec. 4.7.1, suponer que las rentabilidades se rigen por un movimiento Browniano implica suponer que tienen volatilidad constante. Como se verá más adelante, en forma menos restrictiva se suelen utilizar alternativamente otros tipos de procesos de difusión o modelos "con salto", en los que la volatilidad varía aleatoriamente (por ejemplo, puede comportarse de acuerdo con un proceso propio de difusión o con salto). Habitualmente la definición del proceso representativo de las volatilidades y, en particular, su relación con el proceso representativo de los precios al momento t, se suele lograr a partir de la utilización de una **superficie de volatilidades** ($\sigma\left(S(t),t\right)$).

127. Ver Galai, D. (1972), Black, F.; Scholes, M. (1973), Boyle, P.P.;Ananthanarayanan, A.L. (1977), Parkinson, M. (1980), Ball, C.A.; Torous, W.N. (1984), Butler, J.S.; Schachter, B. (1986), French, K.R.; Roll, R. (1986), Marsh, T.A.; Rosenfeld, E.R. (1986), Skinner, D.J. (1989), Rogers, L.C.G.; Sachtell, S.E. (1991), Chesney, M. (1993).

128. Como se verá en la Parte III, la representación de la volatilidad utilizada habitualmente es la conocida como **proceso autorregresivo con heterocedasticidad condicionada generalizado (GARCH)** (o, en particular, dado que la volatilidad de un activo habitualmente reacciona en forma diferente según se produzca un aumento significativo o una caída significativa en el precio, una representación **GARCH-exponencial (EGARCH)**). Debe tenerse en cuenta que, cuando se habla de procesos con volatilidad constante, se hace referencia a la **varianza no-condicionada** de las rentabilidades. Esta volatilidad constante tiene significado sólo si el proceso representativo de las rentabilidades es estacionario de segundo orden. La heterocedasticidad condicionada se refiere al comportamiento de las **volatilidades condicionadas** (que representan los valores de volatilidad como función de t) que tienden a aumentar en la medida que aumenta la frecuencia de las observaciones (ver Baillie, R.T.; Bollerslev, T. (1990), Drost, F.C.; Nijman, T.E. (1993), Ghose, D.; Kroner, K.F. (1995), Andersen, T.G.; Bollerslev, T. (1996), Zhou, B. (1996), Taylor, S.J.; Xu, X: (1997)).

el precio de mercado de la opción sólo puede ser obtenida utilizando métodos numéricos[129]; **iii)** un aumento en la volatilidad genera un incremento en la incertidumbre (es decir, una disminución en la confiabilidad) en la estimación del precio de la opción[130].

Dado que v_t^2 puede ser considerada como una aproximación a la volatilidad en el momento t, un estimador puntual de la volatilidad de un activo puede ser obtenido utilizando **promedios simples** de las observaciones históricas:

$$\sigma_v^2 = \frac{1}{n}\sum_{t=1}^{n} v_t^2$$

Debe tenerse en cuenta que no es necesario aplicar la corrección sobre el número de grados de libertad, ya que se ha supuesto que $E(v_t) = 0$. En general, si $E(v_t) \neq 0$, se obtienen los estimadores: **i)** Máximo-verosímil bajo el supuesto de Normalidad $\hat{\sigma}_v^{(MV)2} = \frac{1}{n}\sum_{t=1}^{n}(v_t - \bar{v})^2$ y **ii)** insesgado, $\hat{\sigma}_v^2 = \frac{1}{n-1}\sum_{t=1}^{n}(v_t - \bar{v})^2$ (de acuerdo con la desigualdad de Jensen, $\hat{\sigma}_v$ es un estimador sesgado de σ_v y que este sesgo puede ser corregido bajo el supuesto de Normalidad)[131].

Si se supone que las rentabilidades obedecen a un proceso $WN - N(0,\sigma_v^2)$, entonces se verificará que:

$$\sigma^2(\sigma_v^2) = \frac{1}{n^2}\sum_{t=1}^{n}\sigma^2(v_t) = \frac{1}{n^2}\sum_{t=1}^{n}\left\{E(v_t^2) - \left[E(v_t^2)\right]^2\right\} = \frac{1}{n^2}\left[3n\sigma_v^4 - n\sigma_v^4\right] = \frac{2}{n}\sigma_v^4$$

(obsérvese que esta varianza de las rentabilidades disminuye en la medida que se incrementa el número de observaciones, no obstante debe tenerse en cuenta que aumentar el período de observación implica aumentar el riesgo de introducir rentabilidades excepcionales) y que[132]:

$$\frac{n\hat{\sigma}_v^2}{\sigma_v^2} : \chi_{(n)}^2$$

De lo que resulta un intervalo de confiabilidad de la forma:

$$p\left[\frac{n\hat{\sigma}_v^2}{\chi_{(n),\alpha/2}^2} \leq \sigma_v^2 \leq \frac{n\sigma_v^2}{\chi_{(n),(1-\alpha)/2}^2}\right] = 1 - \alpha$$

Una forma alternativa de estimar la volatilidad es utilizando **promedios móviles**

---

129. El método utilizado habitualmente es el ya mencionado de Newton-Raphson (ver Sec. 2.11.3). Por otra parte, se debe tener en cuenta que no todos los procesos representativos del comportamiento de los precios de un activo admiten una relación funcional con las correspondientes volatilidades.

130. Debe tenerse en cuenta que si no existiera incertidumbre, no sería posible hallar el valor intrínseco de la opción.

131. Ver Cox, J.C.; Rubinsterin, M. (1985).

132. Como se verá en la Sec. 4.7.6.2, la confiabilidad de las estimaciones depende, en gran medida, del período seleccionado y del método de valuación empleado (ver Diebold, F.X.; Mariano, R.S. (1995)).

exponencialmente ponderados (EWMA)[133].

Dada la serie cronológica $\{v_t\}$ $(t = 1,2,...,n)$ de las rentabilidades de un activo, es posible calcular un $EWMA(n+1)$ de la forma [134]:

$$EWMA(n+1) = \frac{v_t^2 + \lambda v_{t-1}^2 + \lambda^2 v_{t-2}^2 + ... + \lambda^n v_{t-n}^2}{1 + \lambda + \lambda^2 + ... + \lambda^n} \qquad (0 < \lambda < 1)$$

a partir del cual se obtiene el siguiente estimador de la volatilidad[135]:

$$\hat{\sigma}^2(v_{t+1}) = \lim_{n\to\infty} EWMA(n+1) = (1-\lambda)\sum_{j=1}^{\infty} \lambda^{j-1} v_{t-j} = (1-\lambda)v_t^2 + \lambda\hat{\sigma}^2(v_t)$$

En esta expresión el primer sumando representa la **intensidad de reacción** de la volatilidad ante la presencia de "shocks" en el sistema y el segundo sumando representa la **persistencia** de un "shock" representativo de la volatilidad[136]. Habitualmente en las aplicaciones se recomienda utilizar valores de λ comprendidos en el intervalo $[0.75;0.98]$. Valores pequeños de λ para estimaciones de corto plazo y valores altos para estimaciones de largo plazo.

---

133. **"Exponentially weighted moving average".**

134. Suponiendo que $v_t : N(0, \sigma^2(v_t))$, se verifica que $E(|v_t|) = \sqrt{\frac{2}{\pi}}\sigma(v_t)$. Luego se puede concluir que, en este caso, $|v_t| = \sqrt{\frac{2}{\pi}}$ puede ser considerada una aproximación a $\sigma(v_t)$. Debe tenerse en cuenta que éste no es un estimador apropiado de la volatilidad condicionada.

135. Debe recordarse que, dadas: **i)** una variable aleatoria X y **ii)** una variable aleatoria Y definida como una función monótona diferenciable de X $(Y = Y(X))$, sus funciones de densidad están relacionadas de la siguiente forma: $f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x)$. De modo que, si $Y = \sqrt{X}$, entonces $\left|\frac{dx}{dy}\right| = 2y$. Luego, dado que en la literatura suele aparecer la volatilidad definida como el desvío estándar de las rentabilidades futuras ($\sigma(v_t)$), debe tenerse en cuenta que las funciones densidad de la varianza y del desvío estándar de las rentabilidades de un activo están relacionadas de la siguiente forma: $f(\sigma(v_t)) = 2\sigma(v_t)\, g(\sigma^2(v_t))$. Por otra parte, dada una función $f(x)$, el desarrollo de Taylor con respecto al valor esperado de X $(E(X) = m)$ será de la forma $E[f(X)] = f(m) + \frac{1}{2}\sigma^2(X) f''(m)$. En forma similar, se puede escribir $E[f(x)] = f^2(m) + \sigma^2(x)\left\{[f'(m)]^2 - f(m)f''(m)\right\}$. De modo que, despreciando los términos de orden superior, será $\sigma^2[f(X)] = E[f^2(X)] - \{E[f(X)]\}^2 \approx [f'(m)]^2\sigma^2(X)$. Por lo tanto, resulta que $\sigma^2(\hat{\sigma}(v_t)) = \left[\frac{2}{n}\sigma^4(v_t)\right]\left[\frac{1}{2\sigma(v_t)}\right]^2 = \frac{1}{2n}\sigma^2(v_t)$.

136. Obsérvese que un $\lambda \approx 1$ estaría suponiendo un proceso representativo de las volatilidades caracterizado por una escasa reacción respecto de variaciones producidas en el mercado en el pasado inmediato y una alta persistencia. Viceversa, un $\lambda \approx 0$ implica una volatilidad con gran reacción que es asimilada rápidamente por el proceso.

De este estimador de la volatilidad se obtiene que:

$$\sigma^2\left[\hat{\sigma}^2\left(v_{t+1}\right)\right] = (1-\lambda)^2\sigma^2\left(v_t^2\right) + \lambda^2\sigma^2\left[\sigma^2\left(v_t\right)\right]$$

y, como se supone que la volatilidad es constante, se puede escribir:

$$\sigma^2\left[\sigma^2\left(v_{t+1}\right)\right] = \frac{(1-\lambda)^2}{1-\lambda^2}\sigma^2\left(v_t^2\right) = \frac{1-\lambda}{1+\lambda}\sigma_v^2$$

Las estimaciones obtenidas de utilizar EWMA sólo son válidas en el caso que los precios obedezcan a un proceso RW-1. Como se verá en la Parte III, las rentabilidades son incorrelacionadas linealmente, pero observan autocorrelaciones de orden superior (en particular, autocorrelaciones de segundo orden)[137], que indican un comportamiento de las volatilidades en el que se suceden períodos con rentabilidades próximas a cero con períodos con rentabilidades apartadas de cero en cualquiera de los dos sentidos[138]. Como se mencionó en páginas anteriores, este comportamiento de las rentabilidades se conoce como **autorregresivo con heterocedasticidad condicionada (ARCH)**[139].

Obviamente, cuando el mercado es estable, las diferencias entre los resultados obtenidos de aplicar los distintos métodos de estimación de la volatilidad comentados precedentemente es pequeña y es posible construir estimaciones a partir de alguna combinación de los minmos[140].

Además debe tenerse en cuenta que la volatilidad está más afectada por las rentabilidades negativas que por las positivas (la experiencia indica que, en promedio, el impacto de una rentabilidad negativa es el doble que el de una rentabilidad positiva)[141]

Entre los distintos métodos para medir la confiabilidad de los estimadores de la volatilidad de un activo merecen ser mencionados[142]:

1.- El que utiliza **la verosimilitud de la rentabilidad, dado el estimador de la volatilidad.**

---

137. Para un análisis de relaciones no-lineales en las rentabilidades en particular, en datos de frecuencia diaria o mayor y, en consecuencia, para un estudio de la posibilidad de predicción con cierto grado de confiabilidad utilizando modelos no-lineales, ver Lo, A.W.; MacKinley, A.C. (1988), Brock, W.A.; Hseih, D.A.; LeBaron, B. (1990), Goodhart, C.A.E.; Figliouli, L. (1991), LeBaron, B. (1995), Abhyankar, A.; Copeland, L.S.; Wong, W. (1997), de Lima, P.J.F. (1998), Drunat, J.; Dufrenot, G.; Mathieu, L. (1998), Zhou, B. (1998). Recientemente se ha desarrollado una nueva generación de modelos para datos de alta frecuencia, basados en el supuesto que el comportamiento de las rentabilidades obedece a sistemas caóticos de baja dimensión.

138. Como se verá en el Cap. 5, también se observan autocorrelaciones significativas en las series correspondientes a los valores absolutos de las rentabilidades, lo cual dá origen a los ya mencionadas representaciones AR con orden de integración fraccionario (ARFIMA).

139. Este tipo de procesos, que serán tratados en la Parte III, fue introducido por Benoit Mandelbrot (ver Mandelbrot, B. (1963)). La implementación de los modelos correspondientes se debe a Engle, R.F. (1982).

140. Ver Granger, C.W.J.; Ramanathan, R. (1984), Diebold, F.X.; López, J.A. (en Maddala, C.S.; Rao, C.R. (eds.)(1996)).

141. Campbell, J.Y.; Hentschel, L. (1992) atribuyeron este comportamiento al mayor impacto que ejercen las malas noticias respecto de las buenas. Ver Black, F. (1976), Christie, A.A. (1982)

142. Ver Cumby, R.; Figlewski, S.; Hasbrouk, J. (1993), West, K.D.; Cho, D. (1995), Brailsford, T.J.; Faff, R.W.; (1996), Alexander, C.O.; Leigh, C. (1997), Figlewski, S. (1997), Andersen, T.G.; Bollerslev, T. (1998).

Sean $\hat{\sigma}^2\left(v_{t+1}^{(A)}\right)$, $\hat{\sigma}^2\left(v_{t+2}^{(A)}\right)$,..., $\hat{\sigma}^2\left(v_{t+\ell}^{(A)}\right)$ y $\hat{\sigma}^2\left(v_{t+1}^{(B)}\right)$, $\hat{\sigma}^2\left(v_{t+2}^{(B)}\right)$,..., $\hat{\sigma}^2\left(v_{t+\ell}^{(B)}\right)$ sendas sucesiones de estimadores de la volatilidad de un activo generadas de la aplicación de dos modelos A y B, a partir de una serie cronológica $\{v_t, v_{t-1},..., v_{t-n}\}$ y sean $L(A / v_t, v_{t-1},..., v_{t-n})$ y $L(B / v_t, v_{t-1},..., v_{t-n})$ las verosimilitudes "post-muestra" del conjunto de variables aleatorias $\{v_t, v_{t-1},..., v_{t-n}\}$ dados los estimadores de las volatilidades obtenidos, respectivamente, de aplicar los modelos A y B[143]. Se puede construir, entonces, un "backtest"[144] basado en el cociente de verosimilitudes, para contrastar la hipótesis nula H_0: *El modelo adecuado es A*, con respecto a la alternativa H_1: *El modelo adecuado es B*, utilizando un estadístico de la forma:

$$-2\ln(LR) = 2\left[\ell(B / v_t, v_{t-1},..., v_{t-n}) - \ell(A / v_t, v_{t-1},..., v_{t-n})\right]$$

(donde $\ell(\bullet)$ denota el logaritmo de la función de verosimilitud $L(\bullet)$) que converge en-distribución a una función $\chi^2_{(k)}$ (donde k denota el número de coeficientes incluidos en los modelos A y B). Debe tenerse en cuenta que, habitualmente, este test supone que las rentabilidades están Normalmente distribuidas pero, en general, puede ser definido a partir de cualquier función de verosimilitud.

2.- El que utiliza el **error medio cuadrático entre la predicción de la volatilidad y las realizaciones de dicho proceso.**

Con respecto a este criterio, debe tenerse en cuenta que el error medio cuadrático es particularmente apropiado en predicciones de valores esperados y no de volatilidades y que, por lo tanto, lo correcto sería aplicar un test F[145].

3.- El que utiliza una **regresión lineal de los cuadrados de las rentabilidades respecto de la predicción de la varianza** del tipo $v_t^2 = a_0 + a_1\sigma^2(v_t) + \varepsilon_t$. Si la volatilidad estuviera correctamente especificada, entonces se debería verificar que $a_0 = 0$ y $a_1 = 1$. Ahora bien, dado que la variable explicativa no es observable (sus realizaciones son valores estimados) y, en consecuencia, que $\sigma^2(v_t) = \hat{\sigma}^2(v_t) + u_t$, la representación anterior puede ser escrita, en función del regresor aleatorio $\hat{\sigma}^2(v_t)$, de la siguiente forma:

---

143. Los datos históricos se dividen en dos partes: **i)** aquellos que corresponden al período muestral, utilizado para construir el modelo y **ii)** aquellos que no corresponden al período muestral (datos "post-muestra"), que se utilizan para evaluar las predicciones.

144. El cual considera un conjunto de medidas de "post-muestra" basadas en la comparación de valores observados y valores predichos por el modelo. Entre estas medidas merecen ser mencionadas: **i)** el **cociente de información** (definido como el valor medio del error de predicción dividido por el desvío estándar de dicho error de predicción); **ii)** el **error medio cuadrático** (definido como el valor medio de los cuadrados de los errores de predicción); **iii)** el **error medio absoluto** (definido como el valor medio de los valores absolutos de los errores de predicción); **iv)** el **error medio cuadrático estandarizado** (definido como el error medio cuadrático dividido el desvío estándar de los errores de predicción); **v)** la **verosimilitud de la predicción** (definida como el producto de las verosimilitudes de cada valor predicho, suponiendo algún tipo de funciones de densidad de los mismos) y **vi)** el **coeficiente de correlación post-muestra** (definido como el coeficiente de correlación entre las predicciones y los valores observados, suponiendo que el proceso $\{v_t\}$ es estacionario).

145. Ver Makridakis, S. (1993). Debe tenerse en cuenta que: **i)** este criterio es equivalente a maximizar la función de verosimilitud cuando las rentabilidades están Normalmente distribuidas con varianza constante y **ii)** la varianza es la esperanza matemática de una variable aleatoria elevada al cuadrado y que, por lo tanto, tendrá una distribución χ^2, de modo que la función de verosimilitud es diferente y no incluye cualquier suma de errores cuadráticos.

$$v_t^2 = a_0 + a_1\hat{\sigma}^2(v_t) + (\varepsilon_t - a_1 u_t) = a_0 + a_1\hat{\sigma}^2(v_t) + \eta_t$$

donde:

$$\gamma\left[\hat{\sigma}^2(v_t), \eta_t\right] = \gamma\left[\sigma^2(v_t) + u_t, \varepsilon_t - a_1 u_t\right] =$$

$$= \gamma\left[\sigma^2(v_t), \varepsilon_t\right] + \gamma(u_t, \varepsilon_t) - \gamma\left[\sigma^2(v_t), a_1 u_t\right] - \gamma(u_t, a_1 u_t) =$$

$$= -a_1\sigma^2(u_t) < 0$$

De modo que se puede concluir que el estimador de a_1 obtenido de aplicar CMO, tendrá un sesgo negativo[146].

El valor del coeficiente de determinación $R^2_{v_t^2/\hat{\sigma}^2(v_t)}$ que se obtiene de aplicar este modelo puede ser utilizado como medida de su calidad (la parte de la variación del proceso $\{v_t^2\}$ explicada por las predicciones obtenidas para $\sigma^2(v_t)$), pero, debe tenerse en cuenta que posee una cota superior que depende del proceso seleccionado para representar el comportamiento de las rentabilidades[147].

4.7.6.2.- La volatilidad implícita

Una forma alternativa a la basada en las estimaciones históricas de la volatilidad, es a partir de la **volatilidad implícita**, definida como la varianza de la rentabilidad de un activo tal que, cuando es introducida como dato en la fórmula de valuación de una opción, se obtiene como resultado el precio de mercado de dicha opción. Lo que permite concluir que esta forma de estimación depende de los supuestos sobre los que se basa el modelo econométrico que da origen a la fórmula de valuación de la opción (que pueden no verificarse en cada aplicación en particular) y, en consecuencia, del proceso utilizado para representar el comportamiento de la rentabilidad del mismo[148].

Si se supone que las rentabilidades se comportan de acuerdo con un movimiento Browniano geométrico y que, en consecuencia, la fórmula a aplicar para la valuación de la opción es la de Black-Scholes-Merton[149]:

---

146. Debe tenerse en cuenta que, para poder asegurar que el estimador obtenido de aplicar CMO es insesgado, se debe verificar que $\gamma\left[\hat{\sigma}^2(v_t), \eta_t\right] = 0$.

147. Por ejemplo, si el proceso $\{v_t\}$ admite una representación GARCH(1,1) simétrica, el R^2 obtenido de aplicar la regresión mencionada en el texto queda definido de la siguiente forma acotada superiormente: $R^2 = \dfrac{a_0^2}{1 - a_1^2 - 2a_0a_1}$ (ver Andersen, T.G.; Bollerslev, T. (1998)). Los valores de R^2 próximos a cero indican, en este caso, un error de especificación de la representación.

148. Debe tenerse en cuenta que la fórmula de valuación de una opción está definida como una esperanza matemática a partir de una distribución de probabilidades dada.

149. Este modelo fue propuesto por Samuelson, P.A. (1965). Su diferencia con el modelo de Bachelier es análoga a la diferencia entre las teorías del interés simple y compuesto. En el corto plazo ambos modelos producen resultados similares. En el largo plazo las diferencias son muy significativas. Una generalización del modelo de Bachelier para lograr una mejor aproximación al modelo de Black-Scholes.Merton y una extensión de esta aproximación más allá de dicho modelo, pueden ser obtenidas a partir del desarrollo

$$C(K,S(t),\sigma,\tau,r) = S(t)F_U(u_1) - Ke^{-rt}F_U\left(u_1 - \sigma\sqrt{\tau}\right)$$

(donde: **i)** $S(t)$ denota el precio del activo al momento t ; **ii)** K es el precio "strike"; **iii)** $\tau = T - t$ denota el período hasta la ejecución de la opción; **iv)** r denota la tasa de interés libre de riesgo; **v)** $F_U(u)$ denota el valor de la función de distribución de una función $N(0;1)$ para la abcisa u y **vi)** $u_1 = \frac{1}{\sigma\sqrt{\tau}}\ln\left(\frac{S(t)}{Ke^{-r\tau}}\right) + \frac{\sigma\sqrt{\tau}}{2}$)[150], la volatilidad implícita será la que se obtiene de aplicar una ecuación de la forma $\sigma_{(i)}^2 = f\left[C(t),S(t),K,r,\tau\right]$.

Debe tenerse en cuenta que el único elemento no observable en la fórmula de valuación de opciones es la volatilidad y que la única variable es el precio del activo y que no existe ninguna fórmula de inversión que permita calcular la volatilidad implícita como una función del precio de la opción. En particular, para opciones "at the money", cuando $S(t) = K$, Bachelier, L. (1900) obtuvo la relación $C^{(B)}(K,S(t),\sigma,r,\tau) = \sigma^{(B)}\sqrt{\frac{\tau}{2\pi}}$ la cual permite, dado un precio $C = C^{(B)}$ para una opción "at the money", con madurez τ , calcular el coeficiente de nerviosismo del activo, $H = \frac{\sigma^{(B)}}{\sqrt{2\pi}}$ y, en consecuencia, su volatilidad implícita. Para valores dados de $\sigma(t) > 0$, $\tau > 0$ y $S(t) = K$ ("at the money"), sean $\sigma^{(BSM)}(t) = \sigma$ y $\sigma^{(B)} = S(t)\sigma$ (donde $\sigma^{(BS)}$ y $\sigma^{(B)}$ denotan, respectivamente, las volatilidades según los modelos de Black-Scholes-Merton y de Bachelier). Entonces se verifica que $0 \le C^{(B)}(t) - C^{(BSM)}(t) \le \frac{S(t)}{12\sqrt{2\pi}}\sigma^2\tau^{3/2} = O\left(\left(\sigma\sqrt{\tau}\right)^3\right)$ (donde $C^{(B)}$ y $C^{(BSM)}$ denotan los precios de una opción europea obtenidos de aplicar los modelos de Bachelier y de Black-Scholes-Merton, respectivamente). La diferencia relativa será de la forma $\frac{C^{(B)}(t) - C^{(BSM)}(t)}{C^{(B)}(t)} \le \frac{\tau}{12}\sigma^2$. Inversamente, para un valor fijo $0 < \sigma^{(BSM)} < S(t)$ del precio de una opción "at the money", se obtiene que $0 \le \sigma^{(BSM)} - \frac{\sigma^{(B)}}{S(t)} \le \frac{\tau}{12}\left(\sigma^{(BSM)}\right)^3$. A partir de estos resultados, se puede concluir que, cuando $\sigma\sqrt{\tau} << 1$ (un caso que se presenta frecuentemente en las aplicaciones), la fórmula de Bachelier proporciona una aproximación satisfactoria a la solución de Black-Scholes-Merton (y resulta más fácil de calcular)[151].

---

caótico o de los teoremas de Strook-Taylor (ver Nualart, D. (1995), Malliavin, P. (1997)(2005), Malliavin, P.; Thalmaier, A. (2006)).

[150]. Existen tres aproximaciones que permiten obtener la fórmula de Black-Scholes-Merton: **i)** una (analizada en la Sec. 4.5) basada en la aproximación al movimiento Browniano a partir de un proceso discreto en el dominio del tiempo que cumple la condición de no-arbitraje; **ii)** otra basada en la derivación de una ecuación en derivadas parciales y **iii)** una tercera basada sobre el concepto de la **medida de una martingala**.

[151]. Ver Schachermayer, W.; Teichmann, J. (en Bernard, P. (ed.) (2003))(2008).

A partir de la fórmula para valuar el precio de una opción, se puede calcular la derivada de dicho precio con respecto a cualquiera de los elementos de la fórmula[152]. Suponiendo volatilidad constante, puede obtenerse el **coeficiente delta**:

$$\Delta\left(K,S(t),\sigma,\tau,r\right) = \frac{\partial}{\partial S(t)} C\left(K,S(t),\sigma(t),\tau,r\right) = F_U\left(u_1\right)$$

que mide la sensibilidad del precio de la opción con respecto a las variaciones en los precios del activo ($S(t)$). De la misma forma se obtiene el **coeficiente Gamma** de la opción:

$$\Gamma\left(K,S(t),\sigma,\tau,r\right) = \frac{\partial^2}{\partial S^2(t)} C\left(K,S(t),\sigma,\tau,r\right) = \frac{\partial}{\partial S(t)} \Delta\left(K,S(t),\sigma,\tau,r\right) = \frac{1}{\sigma S(t)\sqrt{\tau}} f_U\left(u_1\right)$$

que mide la sensibilidad del coeficiente Δ con respecto a variaciones en el precio del activo.

Si se supone que la volatilidad no es constante, sino una función determinística de $S(t)$, entonces será:

$$\Delta\left(K,S(t),\sigma,\tau,r\right) = \frac{\partial}{\partial S(t)} C\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right) + \frac{\partial}{\partial \sigma\left(S(t)\right)} C\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right) \frac{\partial}{\partial S(t)} \sigma\left(S(t)\right) =$$

$$= \Delta\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right) + \Lambda\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right) \frac{\partial}{\partial S(t)} \sigma\left(S(t)\right)$$

donde:

$$\Lambda\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right) = \frac{\partial}{\partial \sigma\left(S(t)\right)} C\left(K,S(t),\sigma\left(S(t)\right),\tau,r\right)$$

define el **coeficiente Vega** de la opción[153], el cual mide la sensibilidad del precio de la opción con respecto a variaciones en la volatilidad implícita y, en consecuencia, proporciona una medida de la incertidumbre en la estimación de dicho precio[154]. Dado que las funciones $F_U(u)$ y $f_U(u)$ son siempre positivas, se puede concluir que los coeficientes Δ, Γ y Λ son siempre positivos y, dado que su segunda derivada es estrictamente positiva, se puede asegurar que el precio de la opción es una función convexa del precio del activo. Asimismo, se puede asegurar que si dos volatilidades proporcionan el mismo precio de la opción, entonces son iguales[155].

Completando la definición dada al comienzo de esta sección, se puede decir que, bajo un supuesto de volatilidad constante, la volatilidad implícita es la volatilidad del activo "implícita" en el precio de

---

152. Dado que a los coeficientes definidos por estas derivadas se los denota por letras del alfabeto griego (Δ, Γ y Λ), en el metalenguaje del mundo de la finanzas se los denomina habitualmente como **los griegos**.

153. Dado que la notación Vega no corresponde a ninguna letra del alfabeto griego, los puristas lo denominan como el **coeficiente Kappa** pero, paradójicamente se lo representa por la letra Lambda.

154. En particular, si la valuación de la opción se obtiene de aplicar la fórmula de Black-Scholes-Merton, se verificará que $\Delta\left(K,S(t),\sigma,\tau,r\right) = F_U(u)$, $\Gamma\left(K,S(t),\sigma,\tau,r\right) = \frac{1}{S(t)\sigma\sqrt{\tau}} F_U(u)$ y $\Lambda\left(K,S(t),\sigma,\tau,r\right) = S(t)\tau F_U(u)$.

155. Ver Manaster, S.; Koheler, G. (1982), Brenner, M.; Subramanyan, M.G. (1988), Corrado, C.J.; Miller, T.W. (1996).

mercado de una opción de acuerdo con un proceso particular de la representación de su rentabilidad. El problema que se plantea en las aplicaciones es que, si bien -obviamente- existe un único verdadero valor de volatilidad para cada proceso representativo del comportamiento del precio de un activo, se presentan muchas volatilidades diferentes implícitas en los precios de mercado de las opciones. Esta circunstancia es atribuible a que el proceso que se supone que rige el comportamiento de los precios del activo, no coincide con el comportamiento del mercado del mismo y, por lo tanto, debe ser atribuida a algún tipo de error muestral[156].

La experiencia indica que las estimaciones obtenidas a partir de la volatilidad implícita de un activo es mayor que la correspondiente estimación obtenida a partir de las volatilidades históricas. De acuerdo con su definición, la volatilidad implícita en el precio de mercado de una opción debería ser interpretada como la **volatilidad promedio** del proceso que rige el comportamiento de los precios de un activo. De modo que, si el proceso seleccionado para representar el comportamiento del precio de la opción fuera el adecuado y si las expectativas de los operadores fueran evaluadas correctamente, entonces las diferencias entre las volatilidades histórica e implícita serán atribuibles a fallas en la estimación estadística. Por el contrario, si la estimación fuera adecuada, entonces las diferencias serían atribuibles a una selección inadecuada del proceso representativo del comportamiento de las rentabilidades[157].

4.7.7.- Las distribuciones representativas del comportamiento de las rentabilidades

4.7.7.1.- las distribuciones Normales

La hipótesis clásica de Normalidad en la distribución de las rentabilidades -de acuerdo con los postulados del teorema central del límite- implica el supuesto que los precios de los activos están formados por la agregación de "shocks" aleatorios representativos de los factores que impulsan a los operadores a realizar estimaciones sobre el comportamiento de dichos precios y que estas estimaciones individuales constituyen variables aleatorias con varianza finita, independientes entre sí.

Ahora bien, de la observación sobre el comportamiento de los activos se puede concluir: **i)** que las rentabilidades suelen presentar cierto grado de asimetría (habitualmente negativa, lo cual implica una mayor probabilidad de caídas significativas que de aumentos significativos en los precios del activo)[158]; **ii)** que las rentabilidades muestran kurtosis mayor que tres (debe tenerse en cuenta que la kurtosis influye sobre la asimetría); **iii)** que las frecuencias contenidas en las "colas" de la distribución son más significativas que las que se deberían esperar en una función Normal (efecto conocido como **comportamiento "fat tailed"**, que permite concluir que existen probabilidades marcadamente no-nulas de que las rentabilidades, asuman valores alejados de cero); **iv)** que, además, dichas frecuencias son **uniformemente** más significativas que las de la Normal (para desvíos de la variable con respecto a su valor medio superiores a 4σ, se tiene la misma frecuencia que para desvíos superiores a 2σ)[159]; **v)** que esta significatividad de las "colas" se mantiene en el tiempo, es decir que se observa un comportamiento similar para las rentabilidades correspondientes a distintos períodos[160].

---

156. Si el modelo de mercado fuera correcto, el diagrama de dispersión construido considerando los precios "strike" de las opciones sobre un activo en el eje de abcisas y las volatilidades implícitas en el de ordenadas debería asumir la forma de una recta paralela al eje horizontal, pero la experiencia indica que el resultado es una curva conocida como **curva sonrisa** (**"smile"**), lo cual evidencia las imperfecciones de la fórmula de valuación de Black-Scholes-Merton.

157. Debe tenerse en cuenta que no existe ningún modelo que represente perfectamente el comportamiento de la rentabilidad de un activo.

158. Circunstancia que, obviamente, influye sobre la selección de una cartera óptima (ver Fama, E. (1965), Arditti, F.D. (1971), Simkowitz, M.A.; Beedles, W.L. (1978), Singleton, J.C.; Wingender, J. (1986)).

159. Debe tenerse en cuenta que la asimetría y el comportamiento "fat tailed" de las rentabilidades persiste aún después de depurarlas de la heterocedasticidad condicionada, lo cual permite concluir que la distribución condicionada de las rentabilidades tampoco es Normal (ver Mandelbrot, B.B. (1963), Fama, E. (1963), Blattberg, R.; Gonedes, N. (1974)).

160. Ver Kon, S. (1984), Badrinath, S.G.; Chatterjee, S. (1988), Mitnik, S.; Rachev, S.T. (1993).

Un intento de solución al comportamiento singular de la asimetría y la kurtosis, debido a Badrinath, S.G.; Chatterjee, S. (1988)(1991), se basa en la representación de las rentabilidades mediante las transformaciones no-lineales de la función Normal definidas por las distribuciones[161]: **i)** $v_g = A + \frac{B}{g}\left(e^{ug} - 1\right)\left(u: N(0;1)\right)$ (esta transformación, conocida como **función** g, implica suponer que la variable $v(t)$ se distribuye de acuerdo a una función log-Normal no-centrada, acotada por g^{-1}); **ii)** $v_h = A + Bue^{hu^2/2}\left(u: N(0;1)\right)$ (esta transformación, conocida como **función h**, es tal que, si $h > 0$, presentará colas más levantadas que las de la Normal) y **iii)** $(g) \times (h)$ (que se obtiene de multiplicar las dos funciones anteriores)[162].

Otro intento de solución, particularmente apropiado en activos cuyo comportamiento se caracteriza por presentar "saltos" en las rentabilidades, a menudo precediendo a períodos de alta volatilidad, consiste en considerar la **mezcla de distribuciones Normales**[163]. Sean n variables Normales

---

161. Ver Tukey, J.W. (1977) y Hoaglin, D.C. (en Hoaglin, D.C.; Mosteller,F.; Tukey, J.W. (eds.) (1985)), Mills, T.C. (1995),.

162. Se dice que una variable X se distribuye de acuerdo con una función log-Normal ($X: LN\left(\theta, m, \sigma^2\right)$) si se verifica que la variable $Y = \log(X - \theta)$ (donde θ denota un parámetro real que caracteriza la posición de la variable) observa una distribución del tipo $N\left(m, \sigma^2\right)$, con función de densidad de la forma $f_X(x) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}}$ $\exp\left\{-\frac{1}{2}\left[\frac{\log(x-\theta)-m}{\sigma}\right]^2\right\} = \frac{1}{(x-\theta)\sigma} f_N\left(\frac{\log(x-\theta)-m}{\sigma}\right)$ (donde: **i)** $\sigma^2 > 0$; **ii)** $-\infty \le m \le \infty$ y **iii)** $f_N(\bullet)$ denotala función de densidad correspondiente a una variable $N(0;1)$). La correspondiente función de distribución es de la forma $F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\log(x-\theta)-m}{\sigma}} e^{-u^2/2}\, du$. Sus momentos están definido por $m_s(X) = \sum_{j=0}^{s} \frac{s}{j} \theta^{s-j} e^{jm} e^{j^2\sigma^2/2}$ $(s = 1,2,...)$ y, en consecuencia, sus medidas son de la forma $\sigma^2(X) = e^{2m} e^{\sigma^2}\left(e^{\sigma^2} - 1\right)$; $As(X) = \left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1}$ y $K(X) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$.

163. Se denominan **distribuciones Normales compuestas** aquellas funciones Normales en las que los coeficientes m o σ^2 (uno de los dos o ambos) son variables aleatorias. En particular, cuando los dominios de la variables m o σ^2 están formados por un número finito (habitualmente pequeño) de valores, entonces quedan definidas **distribuciones Normales mezcladas** (las cuales pueden ser consideradas como distribuciones que combinan los efectos que ejercen subpoblaciones homogéneas en una población heterogénea). En términos más generales, dado un conjunto finito de variables componentes $(X_1, X_2, ..., X_n)$ con funciones de densidad $f_{X_t}(x, \theta_t)$ $(t = 1,2,...,n)$, se obtiene una función de densidad mezclada de la forma $f_{X_{(n)}}(x, a, \theta) = \sum_{t=1}^{n} a_t f_{X_t}(x, \theta_t)$ $\left(a_t \ge 0; t = 1,2,...,n; \sum_{t=1}^{n} a_t = 1\right)$. Considerada θ_t como una variable aleatoria en el dominio $\{\theta_1, \theta_2, ..., \theta_n\}$, es posible interpretar a la sucesión $\{a_t\}$ como una función de probabilidades discreta (denominada **distribución latente**) formada por las probabilidades de que la i -ésima observación $(i = 1,2,...,m)$ provenga de la t -ésima $(t = 1,2,...,n)$ subpoblación componente (el conjunto $\{\theta_1, \theta_2, ..., \theta_n\}$ está formado por los denominados **puntos de soporte**): $f_\theta(\theta_t) = a_t$ $(t = 1,2,...,n)$. Aunque pueda parecer poco intuitivo con respecto a las aplicaciones aquí planteadas, la distribución latente puede ser continua.

$v_t : N\left(m_t, \sigma_t^2\right)\ (t = 1,2,...,n)$ representativas de las rentabilidades de un activo. La correspondiente **distribución Normal mezclada** de las **variables componentes** v_t, será de la forma:

$$f_{v_{(n)}}(a,v) = a_1 f_{v_1}(v) + a_2 f_{v_2}(v) + ... + a_n f_{v_n}(v) =$$

$$= \sum_{t=1}^{n} \frac{a_t}{\sigma_t \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{v - m_t}{\sigma_t} \right)^2 \right] \qquad \left(a_1 + a_2 + ... + a_n = 1; a_t \geq 0; t = 1,2,...,n\right)$$

Sus momentos absolutos quedan definidos de la siguiente forma[164]:

$$m_s\left(v_{(n)}\right) = \sum_{t=1}^{n} E\left(v_{(n)}^s / m_t, \sigma_t^2\right) = \sum_{t=1}^{n} E\left[\left(m_t + \sigma_t\right)^s\right] = \sum_{j=0}^{[s/2]} \binom{s}{2j} E\left(U^{2j}\right) \sum_{t=1}^{n} m_t^{s-2j} \sigma_t^{2j}$$

(donde $U : N(0,1)$ y, por lo tanto, $E\left(U^{2j+1}\right) = 0$, para $j = 0,1,2,...$). De modo que $\sigma^2\left(v_{(n)}\right) = \sum_{t=1}^{n} a_t \sigma_t^2$ y

$K\left(v_{(n)}\right) = 3\left[\dfrac{\sum_{t=1}^{n} a_t \sigma_t^4}{\left(\sum_{t=1}^{n} a_t \sigma_t^2 \right)^2} - 1 \right]$. Se puede concluir entonces que, si las variables componentes son tales que $m_1 = m_2 = ... = m_n$, se verificará que

$$\sum_{t=1}^{n} a_t \sigma_t^4 > \left(\sum_{t=1}^{n} a_t \sigma_t^2 \right)^2$$

es decir, que la distribución mezclada será leptokúrtica y con "colas" más levantadas que una función Normal con la misma varianza[165] y con varianza condicionada por la variable componente específica de la mezcla de la cual proviene la observación (debe tenerse en cuenta que una varianza condicionada variable sugiere distribuciones marginales con "colas" levantadas, aún en el caso en el que las distribuciones condicionadas sean Normales con "colas" relativamente no-significativas).

En particular, para $n = 2$, se verifica que: **i)** si $m_1 = m_2$ o $\left(m_1 - m_2\right)^2 < \dfrac{27\sigma_1^2\sigma_2^2}{4\left(\sigma_1^2 + \sigma_2^2\right)}$, la distribución mezclada es unimodal; **ii)** si $\left(m_1 - m_2\right)^2 > \dfrac{8\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, es posible hallar por lo menos un valor de a_1 ($a_2 = 1 - a_1$) para el cual la distribución mezclada es bimodal; **iii)** para cualquier conjunto de

---

164. En particular, si $a_i \approx 1$ (y, por lo tanto, $a_j \approx 0$ para $j \neq i$), a la distribución mezclada se la denomina **contaminada** y ha sido utilizada como modelo en tests para rechazar "outliers" en muestras obtenidas de poblaciones Normales (ver Tukey, J.W. (1949)).

165. Debe tenerse en cuenta que la kurtosis es especialmente sensible a la significatividad de las "colas". La leptokurtosis es sinónimo de "colas pesadas". No obstante, en este caso, se demuestra que las "colas" decrecen en forma exponencial (no polinomial).

valores m_1, m_2, σ_1^2 y σ_2^2 es posible hallar por lo menos un valor de a_1 para el cual la distribución mezclada es unimodal[166].

Por otra parte: **i)** si $m_1 = m_2$, entonces la distribución mezclada es simétrica para cualquier valor de a_1; **ii)** si $\sigma_1^2 = \sigma_2^2$ y $a_1 = a_2 = \frac{1}{2}$, entonces la distribución mezclada es simétrica (aún si $m_1 \neq m_2$) con kurtosis negativa y "colas" achatadas; **iii)** si $m_1 = m_2$ y $\sigma_1^2 \neq \sigma_2^2$, entonces la distribución mezclada es simétrica con kurtosis positiva y "colas" levantadas; **iv)** si $m_1 \neq m_2$ y $\sigma_1^2 \neq \sigma_2^2$, entonces la distribución mezclada tendrá asimetría positiva, kurtosis negativa y "colas" levantadas.

Utilizando el método de los momentos los estimadores de los coeficientes $a_1, m_1, m_2, \sigma_1^2$ y σ_2^2 se obtienen de la resolución del siguiente sistema de ecuaciones[167]:

$$m_1\left(v_{(M)}\right) = a_1 m_1 + \left(1 - a_1\right) m_2\left(v_{(M)}\right)$$

$$m_2\left(v_{(M)}\right) = a_1\left(m_1^2 + \sigma_1^2\right) + \left(1 - a_1\right)\left(m_2^2\left(v_{(M)} + \sigma_2^2\right)\right)$$

$$m_3\left(v_{(M)}\right) = a_1\left(m_1^3\left(v_{(M)}\right) + 3m_1\left(v_{(M)}\right)\sigma_1^2\right) + \left(1 - a_1\right)\left(m_2^3\left(v_{(M)}\right) + 3m_2\left(v_{(M)}\right)\sigma_2^2\right)$$

$$m_4\left(v_{(M)}\right) = a_1\left(m_1^4\left(v_{(M)}\right) + 6m_1^2\left(v_{(M)}\right)\sigma_1^2 + 3\sigma_1^4\right) + \left(1 - a_1\right)\left(m_2^4\left(v_{(M)}\right) + 6m_2^2\left(v_{(M)}\right)\sigma_2^2 + 3\sigma_2^4\right)$$

$$m_5\left(v_{(M)}\right) = a_1\left(m_1^5\left(v_{(M)}\right) + 10m_1^3\left(v_{(M)}\right)\sigma_1^2 + 15m_1\left(v_{(M)}\right)\sigma_1^4\right) +$$

$$+ \left(1 - a_1\right)\left(m_2^5\left(v_{(M)}\right) + 10m_2^3\left(v_{(M)}\right)\sigma_2^2 + 15m_2\left(v_{(M)}\right)\sigma_2^4\right)$$

Debe tenerse en cuenta que la estimación de estos coeficientes es equivalente a la definición de la distribución latente[168].

4.7.7.2.- Las distribuciones no-Normales

Como alternativas a la función Normal para representar el comportamiento de las rentabilidades

---

166. Las distribuciones marginales de un proceso que admite una representación GARCH (esta representación será estudiada en la Parte III) son Normales mezcladas, formadas por infinitas variables componentes y con varianzas condicionadas con distribución continua.

167. Esta solución fue planteada originalmente por Pearson, K. (1894) y tratada sucesivamente por De Helguero, F. (1905), Charlier, C.V.L.; Wicksell, S.D. (1924), Molenaar, W. (1965), Cohen, A.C. (1967). En particular, si las variables componentes son del tipo $v_1: N\left(0, b_1\sigma_1^2\right)$ y $v_2: N\left(0, b_2\sigma_2^2\right)$, donde $b_1 > 1$ y $b_2 < 1$ y tales que $a_1 b_1^2 + \left(1 - a_1\right)b_2^2 = 1$, Hull, J.; White, A. (1998) propusieron un método de estimación basado en la relación entre las frecuencias teóricas y empíricas cuartil por cuartil. Métodos gráficos de estimación fueron desarrollados por Harding, (1949), Weichselbergber, K. (1961), Molenaar, W. (1965), Taylor, B.J.R. (1965).

168. De los resultados anteriores se puede concluir que las ventajas de la utilización de distribuciones Normales mezcladas radica en que: **i)** conservan la posibilidad de su tratamiento como función Normal; **ii)** poseen momentos de orden superior y **iii)** tienen capacidad para representar kurtosis altas.

afectadas por asimetrías, kurtosis altas y "colas" significativas, se suelen utilizar[169]:

1.- La **distribución hiperbólica** (que se caracteriza porque el logaritmo de la distribución de la variable tiene la forma de una hipérbola)[170] cuya función de densidad para una variable estandarizada del tipo $v_t^* = \frac{v_t - \lambda}{\delta}$ (donde λ y δ denotan coeficientes característicos, respectivamente, de la posición y del factor escala de v_t) es de la forma:

$$f_{v_t^*}(v) = \frac{B_1(\eta)}{2\sqrt{1+\alpha^2}} \exp\left[-\eta\left(\sqrt{1+\alpha^2}\sqrt{1+v^2} - \alpha v\right)\right] \qquad \left(\eta > 0; -\infty \le \alpha \le \infty; -\infty \le v \le \infty\right)$$

donde: **i)** α denota un coeficiente que caracteriza la kurtosis; **ii)** η denota un coeficiente característico de la asimetría (si $\eta = 0$ la función es simétrica) y **iii)** $B_1(\eta) = \sum_{j=0}^{\infty} \frac{(-1)^j \left(\frac{n}{2}\right)^{2j}}{j!} \Gamma(j+1)$ es una función de Bessel estandarizada[171].

Los coeficientes de esta distribución pueden ser estimados utilizando el método de la máxima verosimilitud[172].

Si se supone que las rentabilidades de un activo se distribuyen Normalmente con valor esperado nulo y varianza estocástica con distribución Normal-inversa, entonces se puede concluir que admiten una distribución hiperbólica. Por otra parte, suponer que las rentabilidades de un activo obedecen a una distribución hiperbólica implica suponer que el correspondiente proceso representativo del precio de una opción sobre el mismo es el de **movimiento hiperbólico de Lévy** , que evoluciona sólo a "saltos".

2.- La **distribución t con α grados de libertad**[173] cuya función de densidad para una variable estandarizada del tipo $v_t^* = \frac{v_t - E(v_t)}{\sigma(v_t)}$ es de la forma:

$$f_{v_t^*}(v) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi\alpha}\,\Gamma\left(\frac{\alpha}{2}\right)\left(1+\frac{v^2}{\alpha}\right)^{(\alpha+1)/2}} \qquad \left(-\infty \le v \le \infty; \alpha = 1,2,\ldots\right)$$

---

169. Ver Mandelbrot, B. (1963), Fama, E. (1963).

170. Obsérvese que el logaritmo de la distribución Normal es representable por una parábola. Luego se puede concluir que las distribuciones Normal y de Laplace son límites de la función hiperbólica.

171. Ver Barndorff-Nielsen, O.; Blaesild, P. (1980) (en Taillie, C.; Patil, G.P.;Baldessari, B.A. (eds.)(1981)), Eberlein, E.; Keller, U. , Eberlein, E. (en Alexander , C. (2001)).

172. Ver Eberlein, E.; Keller, U. (1995).

173. Esta variable fue estudiada en primer término por W.S. Gosset (quien publicó sus trabajos bajo el seudónimo de "the Student" (1908)). Fisher, R.A. (1925)(1925) la definió formalmente como el cociente entre dos variables aleatorias independientes, una $N(0,1)$ y la raíz cuadrada de una variable χ^2 dividida por su número de grados de libertad (ver Landro, A.H. (2002)).

y su función característica es de la forma:

$$C_{v_t}(w) = \frac{\left(\frac{|w|}{2\sqrt{\alpha}}\right)^{\alpha/2}}{\pi\Gamma\left(\frac{\alpha}{2}\right)}\Gamma_{(\alpha/2)}\left(\frac{|w|}{\sqrt{\alpha}}\right)$$

donde $\Gamma_{(\alpha/2)}(y) = \int_0^{\alpha/2} x^{y-1}e^{-x}\,dx \;(y > 0)$ denota una función Gamma-incompleta.

De las expresiones anteriores se puede concluir que esta distribución [174]: **i)** es simétrica con respecto a un máximo en $v = 0$; **ii)** es proporcional a $\frac{1}{\left(1+\frac{v^2}{\alpha}\right)^{(\alpha+1)/2}}$, de modo que, para valores de $|v|$ suficientemente grandes, esta expresión puede ser aproximada por $|v|^{-(\alpha+1)}$, lo que demuestra que esta distribución incluye , en el centro y en las "colas", una densidad de probabilidades mayor que la función $N(0;1)$.

Los momentos absolutos de la variable v_t están definidos por:

$$m_{2s}(v_t) = \frac{1.3.5...(2s-1)}{(\alpha-2)(\alpha-4)...(\alpha-2s)} \qquad \left(s = 1,2,...,\frac{\alpha-1}{2}\right)$$

$$m_{2s+1}(v_t) = 0 \qquad (s = 1,2,...)$$

y sus momentos centrados son de la forma $\mu_2(v_t) = \frac{\alpha}{\alpha-2}\,(\alpha > 2)$; $\mu_3(v_t) = 0$ y $\mu_4(v_t) = \frac{3\alpha^2}{(\alpha-2)(\alpha-4)}\;(\alpha > 4)$. Luego, se puede concluir que la distribución t es leptkúrtica, $K(v_t) = \frac{3(\alpha-2)}{\alpha-4}\,(\alpha > 4)$. Obsérvese que el exceso de kurtosis respecto de la función Normal es igual a $\frac{6}{\alpha-4}$ y que la kurtosis y el peso de las "colas" (polinomiales) crece cuando α disminuye (cuando $\alpha \le 4$ la kurtosis es infinita).

3.- La **distribución logística**, cuya función de densidad para una variable del tipo $v_t^* = \frac{v_t - \lambda}{\delta}$ $(\delta > 0; -\infty \le \lambda \le \infty)$ es de la forma:

---

[174]. Esta distribución, como la de Laplace, la de Pareto, la Gamma y la Beta, pertenecen a la familia de las distribuciones de Pearson (ver Pearson, K. (1895)) las cuales quedan definidas a partir de la ecuación diferencial $\frac{d}{dv}f(v) = \frac{v-\theta_0}{\theta_1+\theta_2 v+\theta_3 v^2}f(v)\;(-\infty \le v \le \infty)$ donde $\theta_0, \theta_1, \theta_2$ y θ_3 denotan parámetros desconocidos.

$$f_{v_t^*}(v) = \frac{e^{-v}}{\left(1+e^{-v}\right)^2} = \frac{1}{4}\sec^2\left(\frac{v}{2}\right) \qquad (-\infty \le v \le \infty)$$

y su función generatriz de momentos es de la forma[175]:

$$M_{v_t^*}(w) = e^{\lambda w} B(1-\delta w, 1+\delta w) = (\pi\delta w)e^{\lambda w}\operatorname{cosec}(\pi\delta w) \qquad (|w| < \delta^{-1})$$

donde:

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx \quad (p,q>0)$$

denota una función Beta-completa.

De las expresiones anteriores se puede concluir que los momentos absolutos de la variable v_t están definidos por $m_s(v_t) = \sum_{j=0}^{s}\binom{s}{j}\lambda^{s-j}\delta^j g(j)$ (debe recordarse que $\sum_{i=1}^{s} i^{-s} = \varsigma(s)\,(s>1)$ denota una función Zeta de Riemann), donde[176]:

$$g(j) = \begin{cases} 1 & si\ j = 0 \\ 0 & si\ j\ es\ impar \\ 2s!\left(1-\frac{1}{2^{s-j}}\right)\sum_{i=1}^{s} i^{-s} & si\ s>1\ y\ j\ es\ par \end{cases}$$

A partir de estos momentos se obtienen las siguientes medidas: $E(v_t) = \lambda$; $\sigma^2(v_t) = \lambda^2 + \frac{\pi^2\delta^2}{3}$; $As(v_t) = 0$ y $K(v_t) = 4.2$.

Los estimadores máximo-verosímiles de los coeficientes λ y $\sigma(v_t)$ se obtienen de resolver (en forma iterativa y utilizando como valores iniciales el valor medio muestral para λ y la varianza muestral para $\sigma^2(v_t)$, que coinciden con los estimadores obtenidos por el método de los momentos) el siguiente sistema de ecuaciones[177]:

---

[175]. La correspondiente función de distribución de esta variable, $F_{v_t^*}(v) = \frac{1}{1+e^{-v}}$, fue habitualmente utilizada desde fines del siglo XIX, para representar funciones de crecimiento (fundamentalmente en aplicaciones demográficas y económicas) en la que v representa la variable tiempo (ver Verhulst, P.F. (1845), Perks, W.F. (1932). Pearl, R. (1940), Fisk, P.R. (1961)).

[176]. Esta función fue obtenida por Gumbel, E.J. (1944) como distribución límite del promedio entre los valores máximo y mínimo de un conjunto de n observaciones (cuando $n \to \infty$). Posteriormente, Gumbel, E.J.; Keeney, R.D. (1950) demostraron que la logística es la distribución límite de un múltiplo del promedio entre los valores extremos de un conjunto de n observaciones.

[177]. Plackett, R.L. (1958) y Tiku, M.L. (1968) propusieron algunos métodos para la linealización de estas ecuaciones.

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=1}^{n}\left\{1+\exp\left[-\frac{\pi}{\hat{\sigma}(v_t)\sqrt{3}}\left(v_t-\hat{\lambda}\right)\right]\right\}^{-1}=\frac{1}{2} \\ \dfrac{1}{n}\displaystyle\sum_{t=1}^{n}\frac{\left(v_t-\hat{\lambda}\right)\left\{1-\exp\left[\frac{\pi}{\hat{\sigma}(v_t)\sqrt{3}}\left(v_t-\hat{\lambda}\right)\right]\right\}}{\hat{\sigma}(v_t)\left\{1+\exp\left[\frac{\pi}{\hat{\sigma}(v_t)\sqrt{3}}\left(v_t-\lambda\right)\right]\right\}}=\frac{\sqrt{3}}{\pi} \end{cases}$$

Las varianzas de estos estimadores, para valores de n suficientemente grandes, pueden ser aproximadas de la siguiente forma $\sigma^2\left(\hat{\lambda}\right)\approx\dfrac{9\sigma^2(v_t)}{\pi^2 n}$ y $\sigma^2\left(\hat{\sigma}^2(v_t)\right)\approx\dfrac{9\sigma^2(v_t)}{\left(3+\pi^2\right)^n}$.

4.- La **distribución de Laplace**[178], cuya función de densidad es de la forma:

$$f_{v_t}(v)=\frac{1}{2\varphi}e^{-|v-\lambda|/\varphi}\qquad\left(\varphi>0;-\infty\le\lambda\le\infty;\lambda\le v\le\infty\right)$$

(esta distribución se conoce, en particular, como la **primera ley de Laplace**)[179]. Se puede concluir en forma inmediata que: **i)** posee un máximo en $v=\lambda$ y **ii)** dado que la función de densidad es proporcional a $e^{-|v-\lambda|/\varphi}$, las "colas" son más significativas que las de la distribución Normal (en general, se verifica que toda función de densidad proporcional a $\exp\left(\left|\dfrac{v-\lambda}{\varphi}\right|^{\alpha}\right)$, para $0<\alpha<2$, presentará "colas" más significativas que las de la distribución Normal).

La correspondiente función de distribución es de la forma:

$$F_{v_t}(v)=\begin{cases}\dfrac{1}{2}e^{-(\lambda-v)/\varphi} & (v\le\lambda)\\ 1-\dfrac{1}{2}e^{-(v-\lambda)/\varphi} & (v\ge\lambda)\end{cases}$$

Su función generatriz de momentos está definida por $M_{v_t}(w)=\dfrac{1}{1-w^2}$. A partir de la cual se

---

178. Conocida también como **distribución doble-exponencial** o **exponencial de dos "colas"** (ver Greenwood, J.A.; Olkin, I.; Savage, I.R. (1962)) o **exponencial bilateral** (ver Feller, W. (1966)). De acuerdo con algunos autores (ver Stigler, S.M. (1980)), la denominación como distribución de Laplace no es sino un "reconocimiento eponímico" (algo así como un premio consuelo) en desagravio a la histórica omisión sobre la pseudo paternidad de la distribución Normal compartida con Gauss (ver Landro, A.H. (2002).

179. Fue propuesta por Laplace, P.S. (1774) como la distribución tal que el máximo de la función de verosimilitud se obtiene cuando $\lambda=me(v)$ en un conjunto de un número impar de variables aleatorias iid. Posteriormente, Laplace reemplazó la mediana por la media aritmética como valor que maximiza la función de verosimilitud y concluyó que la distribución correspondiente era la Normal (denominada también como **segunda ley de Laplace**).

obtienen los momentos:

$$E(v_t) = \lambda$$

$$\mu_s(v_t) = \begin{cases} 0 & para \quad s = 1,3,5,... \\ s! & para \quad s = 2,4,6,... \end{cases}$$

Debe tenerse en cuenta que, dado que el valor esperado y la varianza de v_t son finitos, no se puede asegurar que esta distribución tenga "colas" realmente significativas (en el sentido que, como se verá inmediatamente, posee la distribución de Pareto, la cual es proporcional a $|v_t|^{-(\alpha+1)}$).

Dadas n variables aleatorias observables mutuamente independientes, $v_1, v_2, ..., v_n$, todas con distribución de probabilidades de Laplace, la función de verosimilitud será de la forma:

$$\ell = -n\log(2\varphi) - \frac{1}{\varphi}\sum_{t=1}^{n}|v_t - \lambda|$$

De modo que el estimador máximo-verosímil de λ será aquél que minimice la expresión $\sum_{t=1}^{n}|v_t - \lambda|$.

Si n es impar $\hat{\lambda}^{MV} = me(v_t)$ [180]. Si n es par, el estimador de λ puede ser calculado como el promedio aritmético de los dos valores centrales de la sucesión $\{v_t\}$ ordenada de menor a mayor, $v_{(1)}, v_{(2)}, ..., v_{(n/2)}, v_{((n+1)/2)}, ..., v_{(n)}$, es decir, $\hat{\lambda}^{MV} = \frac{v_{(n/2)} + v_{((n+2)/2)}}{2}$.

Debe tenerse en cuenta que, si bien $m\hat{e}(v_t)$ es un estimador insesgado de λ, no es un estimador de varianza mínima (se demuestra que, para valores de n relativamente pequeños, es posible hallar estimadores insesgados con varianza menor que $m\hat{e}(v_t)$).

Si se supone que λ es conocido, el estimador máximo-verosímil de φ es de la forma $\hat{\varphi}^{(MV)} = \frac{1}{n}\sum_{t=1}^{n}|v_t - \lambda|$ y tal que su distribución de probabilidades está definida por $f(\hat{\varphi}^{(MV)}) = \frac{\varphi}{2n}\chi^2_{(2n)}$. Por el contrario, si λ es desconocido, el estimador máximo-verosímil de φ será $\hat{\varphi}^{(MV)} = \frac{1}{n}\sum_{t=1}^{n}|v_t - \hat{\lambda}^{MV}|$ (para un análisis de la distribución de probabilidades de este estimador, ver Karst, O.J.; Polowy, H. (1963))[181].

---

[180]. Ver Keynes, J.M. (1911).

[181]. Para una comparación de las varianzas de los estimadores insesgados de λ calculados como $m\hat{e}(v_t)$, $\hat{m}(v_t) = \frac{1}{n}\sum_{t=1}^{n}v_t$ y $semi - rango(v_t) = \frac{1}{2}(v_{(1)} - v_{(n)})$ (ver Sarhan, A.E. (1954)). Por su parte, Chu, J.T.; Hotelling, H. (1955) demostraron que, para $n \geq 7$, $\sigma^2(m\hat{e}(v_t)) < \sigma^2(\hat{m}(v_t))$. Ver tambien, Govindarajulu, Z, (1966).

Otra alternativa no-Normal para representar el comportamiento de las rentabilidades es a partir de las **distribuciones de valores extremos** que consideran exclusivamente la distribución de las rentabilidades excepcionalmente altas y de las pérdidas excepcionales [182].

Sean: **i)** una serie cronológica formada por rentabilidades observadas de una activo, $\{v_t\}$ $(t = 1,2,...,n)$, a partir de la cual se construyen k muestras no-superpuestas de tamaño n_k $(n = kn_k)$, $v_1, v_2, ..., v_{n_k}$ (donde, a fin de asegurar el comportamiento WN-estricto de las rentabilidades y la estabilidad de la estructura de comportamiento del activo, se supone que las v_t son realizaciones de las correspondientes n_k variables aleatorias iid); **ii)** $v_{t,n_k}^{(Max)} = Max\left(v_1, v_2, ..., v_{n_k}\right)$ y $v_{t,n_k}^{(\min)} = \min\left(v_1, v_2, ..., v_{n_k}\right)$ y **iii)** $v_{t,n_k}^{(Max)*}$ y $v_{t,n_k}^{(\min)*}$ las correspondientes variables estandarizadas que se obtienen de restar de las variables originales un coeficiente λ que representa su posición y de dividir por un coeficiente δ que representa su factor escala, $v_{t,n_k}^{(Max)*} = \dfrac{v_{t,n_k}^{(Max)} - \lambda}{\delta}$ y $v_{t,n_k}^{(\min)*} = \dfrac{v_{t,n_k}^{(\min)} - \lambda}{\delta}$. En el límite, cuando $n_k \to \infty$, esta variables, que representan las rentabilidades extremas estandarizadas, $v_t^{(Max)*} = \lim_{n_k \to \infty} v_{t,n_k}^{(Max)*}$ y $v_t^{(\min)*} = \lim_{n_k \to \infty} v_{t,n_k}^{(\min)*}$, admiten tres tipos de distribuciones definidos de la siguiente forma [183]:

$$Tipo\ 1: F_{v_t^*}(v) = \exp\left(-e^{-v}\right)$$

$$Tipo\ 2: F_{v_t^*}(v) = e^{-v^{-\alpha}} \qquad (v \geq \alpha)$$

$$Tipo\ 3: F_{v_t^*}(v) = e^{-v^{\alpha}} \qquad (v \leq \alpha)$$

(obviamente las características de la variable $v_t^{(\min)*}$ pueden ser obtenidas a partir de las características de la variable $v_t^{(Max)*}$ y viceversa, mediante un simple cambio de signo).

Las distribuciones de tipo **2** y **3** pueden ser transformadas en una distribución de **tipo 1** utilizando, respectivamente, las siguientes transformaciones: $y = \log(v - \alpha)$ y $y = -\log(\alpha - v)$.

Estas funciones dan origen a las siguientes familias de distribuciones[184]:

1) La distribución de Gumbel:

$$f_{v_t^{(Max)}}(v) = e^{-v} e^{-e^{-v}} \qquad (-\infty \leq v \leq \infty)$$

---

182. El análisis de los valores extremos ha sido extendido a procesos estocásticos estacionarios formados por variables Normales (ver Berman, S.M. (1964) y Leadbetter, M.R.; Lindgreen, G.; Rootzén, H. (1983)).

183. Obviamente la selección del valor n_k depende del período de observación que abarca la serie cronológica. En series correspondientes a precios de activos, que habitualmente son de parámetro diario, algunos autores recomiendan considerar $n_k = 21$ (valor que corresponde aproximadamente al número de días hábiles en un mes) o $n_k = 63$ (valor que corresponde aproximadamente al número de días hábiles en un trimestre).

184. Gnedenko, B.V. (1943) determinó las condiciones necesarias y suficientes que deben satisfacer las funciones de distribución para poder ser consideradas asociadas con cada uno de estos tres tipos de distribuciones en el límite. Para un análisis más detallado sobre las distribuciones de valores extremos, ver Hareer, H.L. (1978).

cuya función generatriz de momentos es de la forma $M_{v(Max)}(w) = \Gamma(1-w)\,(w<1)$.

De estas expresiones se puede concluir, en forma inmediata, que esta distribución: **i)** posee un máximo en $v = 0$; **ii)** decrece exponencialmente hacia las colas; **iii)** su valor esperado y su varianza están dados por $E\left(v_t^{(Max)}\right) = \lambda + (0.57721)\delta$ y $\sigma^2\left(v_t^{(Max)}\right) = \dfrac{\delta^2\pi^2}{6}$ (donde $\gamma = 0.57721$ define la **constante de Euler**); **iv)** posee asimetría positiva[185]:

$$As\left(v_t^{(Max)*}\right) = -\left(\frac{\sqrt{6}}{\pi}\right)^3\left[\frac{d^3}{du^3}\log\left(\Gamma(u)\right)\right]_{u=1} = 1.1395$$

(donde la expresión $\psi^{(s)}(u) = \dfrac{d^{s+1}}{du^{s+1}}\log\left(\Gamma(u)\right)\,(u>0)$ definde la **función poli-Gamma**) y **iv)** posee alta kurtosis, $K\left(v_t^{(Max)*}\right) = 5.4$.

Los estimadores máximo-versímiles de los coeficientes λ y δ pueden ser obtenidos resolviendo el siguiente sistema de ecuaciones[186]:

$$\begin{cases} \hat{\lambda} = -\hat{\delta}\log\left(\dfrac{1}{n_k}\displaystyle\sum_{t=1}^{n_k} e^{-v_t^{(Max)}/\delta}\right) \\ \hat{\delta} = \dfrac{1}{n_k}\displaystyle\sum_{t=1}^{n_k} v_t^{(Max)} - \dfrac{\displaystyle\sum_{t=1}^{n_k} v_t^{(Max)}\, e^{-v_t^{(Max)}/\delta}}{\displaystyle\sum_{t=1}^{n_k} e^{-v_t^{(Max)}/\delta}} \end{cases}$$

Estos estimadores son asintóticamente insesgados[187] y con varianzas mínimas definidas por $\lim_{n_k\to\infty}\hat{\sigma}^2\left(\hat{\lambda}\right) = \dfrac{(1.10867)\delta^2}{n_k}$ y $\lim_{n_k\to\infty}\sigma^2\left(\hat{\delta}\right) = \dfrac{(0.60793)\delta^2}{n_k}$ y asintóticamente Normales. Además se verifica que $\lim_{n_k\to\infty}\rho\left(\hat{\lambda},\hat{\delta}\right) = 0.313$.

De la misma forma, se demuestra que la variable que representa las pérdidas máximas posee una

---

185. Esta función fue estudiada por Frechét, M. (1927) y Fisher, R.; Tipett, L.H.C. (1928) como distribución límite, bajo ciertas condiciones, de los valores máximos de conjuntos de n observaciones (cuando $n\to\infty$). Por su parte, Talacko, J. (1956) demostró que la variable $v^{**} = \displaystyle\sum_{t=1}^{n}\frac{v_t}{t}$ (donde las variables v_t son independientes con distribución de Gumbel-Max) converge en-distribución a una variable logística, cuando $n\to\infty$. De acuerdo con los teoremas asintóticos de Gnedenko, B.V. (1943) se demuestra que esta distribución (conocida también como **distribución log-Weibull**) puede ser obtenida a partir de las distribuciones Normal, log-Normal, logística y Gamma (ver Landro, A.H. (2002)).

186. Ver Embrecht, S.P.; Klüppelberg, G.; Mikosch, T. (1997).

187. Para obtener una simplificación del estimador $\hat{\delta}$, ver Kimball, B.F. (1956).

distribución de probabilidades de la forma[188]:

$$f_{v_t^{(\min)}}(v) = e^{v} e^{-e^{v}} \qquad (-\infty \le v \le \infty)$$

con medidas $E\left(v_t^{(\min)}\right) = \lambda - (0.57721)\delta$; $\sigma^2\left(v_t^{(\min)}\right) = \sigma^2\left(v_t^{(Max)}\right)$; $As\left(v_t^{(\min)}\right) = -As\left(v_t^{(Max)}\right)$ y $K\left(v_t^{(\min)}\right) = K\left(v_t^{(Max)}\right)$.

Los estimadores máximo-verosímiles de los coeficientes λ y δ pueden ser obtenidos de la solución del siguiente sistema de ecuaciones:

$$\begin{cases} \hat{\lambda} = -\hat{\delta}\log\left(\dfrac{1}{n_k}\displaystyle\sum_{t=1}^{n_k} e^{-v_t^{(\min)}/\delta}\right) \\ \hat{\delta} = \dfrac{1}{n_k}\displaystyle\sum_{t=1}^{n_k} v_t^{(\min)} - \dfrac{\displaystyle\sum_{t=1}^{n_k} v_t^{(\min)} e^{-v_t^{(\min)}/\delta}}{\displaystyle\sum_{t=1}^{n_k} e^{-v_t^{(\min)}/\delta}} \end{cases}$$

Obviamente estos estimadores poseen las mismas propiedades que los de las rentabilidades máximas.

2) La distribución de Weibull[189]:

$$f_{v_t^{(Max)}}(v) = \alpha(-v)^{\alpha-1} e^{-(-v)^{\alpha}} \qquad (\alpha > 0; -\infty \le v \le 0)$$

cuyas medidas están definidas de la siguiente forma[190]:

$$E\left(v_t^{(Max)}\right) = \lambda - \delta\Gamma\left(1 + \frac{1}{\alpha}\right)$$

---

188. Esta función fue estudiada por Frechét, M. (1927) y Fisher, R.; Tippett, L.H.C. (1928). Obsérvese que las funciones de distribución de probabilidades de las variables máxima y mínima están relacionadas de la siguiente forma: $F_{v_t^{(Max)*}}(v) = 1 - F_{v_t^{(\min)*}}(-v)$.

189. Gnedenko, B. (1943) demostró que esta función representa la distribución de una variable v_t correspondiente a los valores máximos de conjuntos de observaciones obtenidas de universos caracterizados por una densidad acotada superiormente ($v_t \le \lambda$) y tal que, para ciertos valores positivos de α y δ, se verifica que $\lim_{v_t \to \lambda}\left[1 - F_{v_t}(v)\right] = \left(\dfrac{\lambda - v}{\delta}\right)^{\alpha}$. Por esta razón, en la literatura rusa a esta distribución se la conoce también como de Weibull-Gnedenko (ver Mann, N.R. (1968), Landro, A.H. (2002)).

190. Dada una variable v_t con distribución de este tipo, con coeficientes de posición y de escala λ y δ, respectivamente, la variable $v_t^{**} = -\alpha\log\left(-\dfrac{v_t - \lambda}{\delta}\right) = -\alpha\log\left(-v_t^{*}\right)$ tiene distribución de Gumberl-Max con coeficientes $\lambda = 0$ y $\delta = 1$.

$$\sigma^2\left(v_t^{(Max)}\right) = \delta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right]$$

$$As\left(v_t^{(Max)}\right) = -\frac{\Gamma\left(1+\frac{3}{\alpha}\right) - 3\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right) + 2\Gamma^3\left(1+\frac{1}{\alpha}\right)}{\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right]^2}$$

$$K\left(v_t^{(Max)}\right) = \frac{\Gamma\left(1+\frac{4}{\alpha}\right) - 4\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma\left(1+\frac{1}{\alpha}\right) + 6\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma^2\left(1+\frac{1}{\alpha}\right) - 3\Gamma^4\left(1+\frac{1}{\alpha}\right)}{\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right]^2}$$

Los estimadores máximo-verosímiles (eficientes y consistentes) de los coeficientes λ, δ y α se obtienen de resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \hat{\delta} = \left\{\frac{1}{n_k}\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(Max)}\right]^{\hat{\alpha}}\right\}^{1/\hat{\alpha}} & \left(\hat{\lambda} > Max\left(v_1, v_2, \ldots, v_{n_k}\right)\right) \\ \frac{n_k}{\hat{\alpha}} + \sum_{t=1}^{n_k}\log\left[\hat{\lambda} - v_t^{(Max)}\right] - \frac{n_k}{\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(Max)}\right]^{\hat{\alpha}}}\sum_{t=1}^{n_k}\left\{\left[\hat{\lambda} - v_t^{(Max)}\right]^{\hat{\alpha}}\log\left[\hat{\lambda} - v_t^{(Max)}\right]\right\} = 0 \\ \left(\hat{\alpha} - 1\right)\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(Max)}\right]^{-1} = \frac{n_k\alpha}{\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(Max)}\right]^{\hat{\alpha}}}\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(Max)}\right]^{\hat{\alpha}-1} \end{cases}$$

Asimismo, utilizando el método de los momentos, los correspondientes estimadores se obtienen de la resolución del siguiente sistema de ecuaciones:

$$\begin{cases} \hat{As}\left(v_t^{(Max)}\right) = \frac{\Gamma\left(1+\frac{3}{\hat{\alpha}^*}\right) - 3\Gamma\left(1+\frac{1}{2\hat{\alpha}^*}\right)\Gamma\left(1+\frac{1}{\hat{\alpha}^*}\right) + 2\Gamma^3\left(1+\frac{1}{\hat{\alpha}^*}\right)}{\left[\Gamma\left(1+\frac{2}{\hat{\alpha}^*}\right) - \Gamma^2\left(1+\frac{1}{\hat{\alpha}^*}\right)\right]^{3/2}} \\ \hat{\delta}^* = \frac{\hat{\sigma}\left(v_t^{(Max)}\right)}{\sqrt{\Gamma\left(1+\frac{2}{\hat{\alpha}^*}\right) - \Gamma^2\left(1+\frac{1}{\hat{\alpha}^*}\right)}} \\ \hat{\lambda}^* = \bar{v} + \hat{\delta}^*\Gamma\left(1+\frac{1}{\hat{\alpha}^*}\right) \end{cases}$$

De la misma forma, se demuestra que la variable que representa los quebrantos máximos (es decir, las rentabilidades mínimas) posee una distribución de probabilidades de la forma[191]:

$$f_{v_t^{(\min)}}(v) = \alpha v^{\alpha-1} e^{-v^{\alpha}} \qquad (\alpha > 0; 0 \le v \le \infty)$$

con medidas $E\left(v_t^{(\min)}\right) = \lambda + \delta\Gamma\left(1+\frac{1}{\alpha}\right)$;

$$\sigma^2\left(v_t^{(\min)}\right) = \sigma^2\left(v_t^{(Max)}\right)$$

y

$$As\left(v_t^{(\min)}\right) = -As\left(v_t^{(Max)}\right)$$

y la misma kurtosis que la distribución de las rentabilidades máximas[192].

Los estimadores máximo-verosímiles de los coeficientes λ, δ y α se obtienen de resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \hat{\delta} = \left\{ \frac{1}{n_k} \sum_{t=1}^{n_k} \left[\hat{\lambda} - v_t^{(\min)}\right]^{\hat{\alpha}} \right\}^{1/\hat{\alpha}} & \left(\hat{\lambda} < \min\left(v_1, v_2, \ldots, v_{n_k}\right)\right) \\ \frac{n_k}{\hat{\alpha}} + \sum_{t=1}^{n_k} \log\left[\hat{\lambda} - v_t^{(\min)}\right] - \frac{n_k}{\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(\min)}\right]^{\hat{\alpha}}} \sum_{t=1}^{n_k} \left\{ \left[\hat{\lambda} - v_t^{(\min)}\right]^{\hat{\alpha}} \log\left[\hat{\lambda} - v_t^{(\min)}\right] \right\} = 0 \\ \left(\hat{\alpha} - 1\right) \sum_{t=1}^{n_k} \left[\hat{\lambda} - v_t^{(\min)}\right]^{-1} = \frac{n_k \alpha}{\sum_{t=1}^{n_k}\left[\hat{\lambda} - v_t^{(\min)}\right]^{\hat{\alpha}}} \sum_{t=1}^{n_k} \left[\hat{\lambda} - v_t^{(\min)}\right]^{\hat{\alpha}-1} \end{cases}$$

De acuerdo con Aroian, L.A. (1965) y Miller, D.W. (1966), este estimador de α está afectado por un sesgo que depende del valor de n_k y no del valor de α. Asimismo, se demuestra que, para $\alpha > 2$, los estimadores máximo-verosímiles de α y δ convergen en-distribución a una función Normal.

Utilizando el método de los momentos los correspondientes estimadores se obtienen de la resolución

---

191. De las dos distribuciones de Weibull aquí mencionadas, la de rentabilidades mínimas es la de utilización más frecuente en las aplicaciones.

192. En la resolución de este sistema es conveniente calcular los estimadores de λ y β a partir de las dos últimas ecuaciones y sustituir estos resultados en la primera ecuación, para obtener el estimador de δ.

del siguiente sistema de ecuaciones[193]:

$$\begin{cases} A\hat{s}\left(v_t^{(\min)}\right) = \dfrac{\Gamma\left(1+\dfrac{3}{\hat{\alpha}^*}\right) - 3\Gamma\left(1+\dfrac{1}{2\hat{\alpha}^*}\right)\Gamma\left(1+\dfrac{1}{\hat{\alpha}^*}\right) + 2\Gamma^3\left(1+\dfrac{1}{\hat{\alpha}^*}\right)}{\left[\Gamma\left(1+\dfrac{2}{\hat{\alpha}^*}\right) - \Gamma^2\left(1+\dfrac{1}{\hat{\alpha}^*}\right)\right]^{3/2}} \\ \hat{\delta}^* = \dfrac{\hat{\sigma}\left(v_t^{(\min)}\right)}{\sqrt{\Gamma\left(1+\dfrac{2}{\hat{\alpha}^*}\right) - \Gamma^2\left(1+\dfrac{1}{\hat{\alpha}^*}\right)}} \\ \hat{\lambda}^* = \overline{v} - \hat{\delta}^*\Gamma\left(1+\dfrac{1}{\hat{\alpha}^*}\right) \end{cases}$$

3) La distribución de Fréchet[194]:

$$f_{v_t^{(Max)}}(v) = \alpha v^{-\alpha-1}e^{-v^{-\alpha}} \qquad (\alpha > 0; 0 \le v \le \partial)$$

(obsérvese que la "cola" de la distribución decrece a una velocidad menor que en el caso de la distribución de Weibull)[195].

Cuya esperanza matemática es de la forma $E\left(v_t^{(Max)}\right) = \lambda + \delta\Gamma\left(1-\dfrac{1}{\alpha}\right)$ $(\alpha > 1)$ y sus momentos absolutos, en general, están definidos por $m_s\left(v_t^{(Max)}\right) = \sum_{j=0}^{s}\binom{s}{j}\lambda^{s-j}\delta^j\Gamma\left(1-\dfrac{j}{\alpha}\right)$ $(s < \alpha)$. A partir de estas expresiones se obtienen las siguientes medidas:

$$\sigma^2\left(v_t^{(Max)}\right) = \hat{\delta}\left[\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma^2\left(1-\frac{1}{\alpha}\right)\right] \qquad (\alpha > 2)$$

---

193. Para el caso en que el verdadero valor de λ sea conocido existen métodos alternativos de estimación debidos a Gumbel, E.J. (1958), Menon, M.V. (1963), Miller, I.; Freund, J.E. (1965), Dubey, S.D. (1966)(1966)(1967). Weibull, W. (1967), Vogt, H. (1968), Bain, L.J.; Antle, C.E. (1967).

194. Propuesta por Frechét, M. (1927). Se demuestra (ver Gnedenko, B. (1943)) que esta distribución de Frechét-Max es el límite hacia el cual converge una variable representativa del valor máximo de un conjunto de n observaciones obtenidas de un universo con dominio no-acotado superiormente ($v_t \ge \lambda$) y tal que, para ciertos valores positivos de K y A, se verifique la relación $\lim_{v_t \to \infty} v_t^K\left[1 - F_{v_t}(v)\right] = A$ (donde $F_{v_t}(v)$ denota la función de distribución de probabilidades de la variable v_t). Dado que estas condiciones también se verifican para la variable de Cauchy, frecuentemente esta distribución de Frechét-Max es denominada equivocadamente como distribución de Cauchy (ver Landro, A.H. (2002)).

195. Dada una variable v_t^* con distribución de Frechét-Max, se verifica que la variable $v_t^{**} = \alpha\log\left(v_t^*\right)$ se distribuye de acuerdo con una función de Gumbel-Max.

$$\sigma^2\left(v_t^{(Max)}\right) = \hat{\delta}\left[\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma^2\left(1-\frac{1}{\alpha}\right)\right] \qquad (\alpha > 2)$$

$$As\left(v_t^{(Max)}\right) = \frac{\Gamma\left(1-\frac{3}{\alpha}\right) - 3\Gamma\left(1-\frac{2}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right) + 2\Gamma^2\left(1-\frac{1}{\alpha}\right)}{\left[\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma^2\left(1-\frac{1}{\alpha}\right)\right]^{3/2}} \qquad (\alpha > 3)$$

$$K\left(v_t^{(Max)}\right) = \frac{\Gamma\left(1-\frac{4}{\alpha}\right) - 4\Gamma\left(1-\frac{3}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right) + 6\Gamma\left(1-\frac{2}{\alpha}\right)\Gamma^2\left(1-\frac{1}{\alpha}\right) - 3\Gamma^4\left(1-\frac{1}{\alpha}\right)}{\left[\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma^2\left(1-\frac{1}{\alpha}\right)\right]^2} \qquad (\alpha > 4)$$

Los estimadores máximo-verosímiles de los coeficientes λ, δ y α se obtienen de resolver el siguiente sistema de ecuaciones[196]:

$$\begin{cases} \hat{\delta} = \left\{ \dfrac{n_k}{\sum_{t=1}^{n_k}\left[v_t^{(Max)} - \hat{\lambda}\right]^{\hat{\alpha}}} \right\}^{1/\hat{\alpha}} & \left(\hat{\lambda} \geq Max\left(v_1, v_2, \ldots, v_{n_k}\right)\right) \\ \dfrac{n_k}{\hat{\alpha}} - \sum_{t=1}^{n_k} \log\left[v_t^{(Max)} - \hat{\lambda}\right] + \dfrac{n_k}{\sum_{t=1}^{n_k}\left[v_t^{(Max)} - \hat{\lambda}\right]^{-\hat{\alpha}}} \sum_{t=1}^{n_k} \dfrac{\log\left[v_t^{(Max)} - \hat{\lambda}\right]}{\left[v_t^{(Max)} - \hat{\lambda}\right]^{\hat{\alpha}}} = 0 & \\ \left(\hat{\alpha}+1\right)\sum_{t=1}^{n_k}\left[v_t^{(Max)} - \hat{\lambda}\right]^{-1} = \dfrac{n_k\hat{\alpha}}{\sum_{t=1}^{n_k}\left[v_t^{(Max)} - \hat{\lambda}\right]^{-\hat{\alpha}}} \sum_{t=1}^{n_k}\left[v_t^{(Max)} - \hat{\lambda}\right]^{-\hat{\alpha}-1} & \end{cases}$$

La matriz de varianzas y covarianzas de estos coeficientes está formada por los siguientes elementos

$$\hat{\sigma}^2\left(\hat{\lambda}\right) = \frac{\left(\hat{\alpha}+1\right)^2}{n_k\hat{\delta}^2}\Gamma\left(1+\frac{2}{\hat{\alpha}}\right), \quad \sigma^2\left(\hat{\delta}\right) = \sigma^2\left(\hat{\alpha}\right)\frac{\hat{\alpha}}{n_k\hat{\delta}}, \quad \gamma\left(\hat{\lambda},\hat{\delta}\right) = \frac{\hat{\alpha}\left(\hat{\alpha}+1\right)}{n_k\hat{\delta}^2}\Gamma\left(1+\frac{1}{\hat{\alpha}}\right), \quad \gamma\left(\hat{\lambda},\hat{\alpha}\right) = \frac{\hat{\alpha}+1}{n_k\hat{\delta}}\cdot$$

$$\cdot\left[\Gamma\left(1+\frac{1}{\hat{\alpha}}\right) + \Gamma\left(1+\frac{1}{\hat{\alpha}}\right)\right] \text{ y } \gamma\left(\hat{\delta},\hat{\alpha}\right) = \frac{0.422785}{n_k\hat{\delta}}.$$

Utilizando el método de los momentos, los estimadores se obtienen de resolver el siguiente sistema

---

196. De acuerdo con Tiago de Oliveira, J. (1972) el proceso de estimación considera obtener, mediante un método iterativo, los estimadores $\hat{\lambda}$ y $\hat{\alpha}$ a partir de las dos últimas ecuaciones del sistema, y sustituir estos valores en la primera ecuación, para obtener el estimador $\hat{\delta}$. Los estimadores máximo-verosímiles son consistentes pero, dado que el conjunto de observaciones (formado exclusivamente por los valores extremos) es, en general, pequeño, se sugiere utilizar el estimador de Hill (ver Embrechts, P.; Klüppelberg, G.; Mikosch, T. (1997)).

de ecuaciones[197]:

$$\begin{cases} \hat{\delta} = \dfrac{\hat{\sigma}\left(v_t^{(Max)}\right)}{\sqrt{\Gamma\left(1-\dfrac{2}{\hat{\alpha}}\right) - \Gamma^2\left(1-\dfrac{1}{\hat{\alpha}}\right)}} \\ \overline{v}_t^{(Max)} = \overline{v} - \hat{\delta}\,\Gamma\left(1-\dfrac{1}{\hat{\alpha}}\right) \end{cases}$$

De la misma forma, se demuestra que la variable que representa los quebrantos mínimos posee una distribución de probabilidades de la forma[198]:

$$f_{v_t^{(\min)}}(v) = \alpha(-v)^{-\alpha-1} e^{-(-v)^{-\alpha}} \qquad (\alpha > 0; -\infty \le v \le 0)$$

con momentos absolutos definidos de la siguiente forma:

$$m_s\left(v_t^{(\min)}\right) = \sum_{j=0}^{s} \binom{s}{j} \lambda^{s-j} \delta^j \Gamma\left(1-\frac{j}{\alpha}\right) \qquad (s < \alpha)$$

a partir de los cuales se obtienen las siguientes medidas:

$$E\left(v_t^{(\min)}\right) = \lambda - \delta\Gamma\left(1-\frac{1}{\alpha}\right) \quad (\alpha > 1)$$

$$\sigma^2\left(v_t^{(\min)}\right) = \sigma^2\left(v_t^{(Max)}\right)$$

y $As\left(v_t^{(\min)}\right) = -As\left(v_t^{(Max)}\right)$ y la misma kurtosis que la distribución de las rentabilidades máximas[199].

Los estimadores máximo-verosímiles de los coeficientes λ, δ y α se obtienen de resolver el siguiente sistema de ecuaciones[200]:

---

197. Debe tenerse en cuenta que, como la variable de Fréchet no posee momentos de orden mayor que α, este método de estimación sólo puede ser utilizado en el supuesto (poco frecuente en las aplicaciones al mercado de capitales) que $\alpha > 2$. Para un análisis de métodos de estimación alternativos, ver Harter, H.L.; Moore, A.H. (1968).

198. Dada una variable $v_t^* = \dfrac{v_t - \lambda}{\delta}$ con distribución de Fréchet-min, se verifica que la variable $v_t^{**} = -\alpha \log\left(-v_t^*\right)$ tiene una distribución de Gumbel con $\lambda = 0$ y $\delta = 1$.

199. Estas distribuciones resultan particularmente adecuadas para el caso en que el proceso representativo de las volatilidades admitan una representación GARCH o sus variables se distribuyan de acuerdo con una función "t" o estable de Pareto (esta distribución será tratada a continuación). Por esta razón suele ser utilizada para representar el comportamiento de las pérdidas máximas en una cartera de inversión.

200. De acuerdo con Tiago de Oliveira, J. (1972), el proceso de estimación considera obtener, mediante un método iterativo, los estimadores $\hat{\lambda}$ y $\hat{\alpha}$ a partir de las dos últimas ecuaciones del sistema y sustituir esto valores en la primera ecuación, para obtener el estimador $\hat{\delta}$.

$$\begin{cases} \hat{\delta} = \left\{ \dfrac{n_k}{\sum_{t=1}^{n_k} \left[v_t^{(\min)} - \hat{\lambda} \right]^{\hat{\alpha}}} \right\}^{1/\hat{\alpha}} & \left(\hat{\lambda} \leq \min\left(v_1, v_2, \ldots, v_{n_k} \right) \right) \\ \dfrac{n_k}{\hat{\alpha}} - \sum_{t=1}^{n_k} \log\left[v_t^{(\min)} - \hat{\lambda} \right] + \dfrac{n_k}{\sum_{t=1}^{n_k} \left[v_t^{(\min)} - \hat{\lambda} \right]^{-\hat{\alpha}}} \sum_{t=1}^{n_k} \dfrac{\log\left[v_t^{(\min)} - \hat{\lambda} \right]}{\left[v_t^{(\min)} - \hat{\lambda} \right]^{\hat{\alpha}}} = 0 & \\ \left(\hat{\alpha} + 1 \right) \sum_{t=1}^{n_k} \left[v_t^{(\min)} - \hat{\lambda} \right]^{-1} = \dfrac{n_k \hat{\alpha}}{\sum_{t=1}^{n_k} \left[v_t^{(\min)} - \hat{\lambda} \right]^{-\hat{\alpha}}} \sum_{t=1}^{n_k} \left[v_t^{(\min)} - \hat{\lambda} \right]^{-\hat{\alpha}-1} & \end{cases}$$

Utilizando el método de los momentos, los estimadores (para $\alpha > 2$) se obtienen de resolver el siguiente sistema de ecuaciones[201]:

$$\begin{cases} \hat{\delta} = \dfrac{\hat{\sigma}\left(v_t^{(\min)} \right)}{\sqrt{\Gamma\left(1 - \dfrac{2}{\hat{\alpha}^*} \right) - \Gamma^2\left(1 - \dfrac{1}{\hat{\alpha}^*} \right)}} \\ \overline{v}_t^{(\min)} = \overline{v} - \hat{\delta}\Gamma\left(1 - \dfrac{1}{\hat{\alpha}^*} \right) \end{cases}$$

Una forma alternativa de estimación no-paramétrica del coeficiente $\alpha^* = \dfrac{1}{\alpha}$, para las distribuciones de valores extremos mencionadas precedentemente, propuesta por Pickands, J. (1975), puede ser obtenida a partir de los estadísticos de orden, $v_{(1)}, v_{(2)}, \ldots, v_{(n)}$ (es decir, sin necesidad de subdividir la serie original) mediante una expresión de la forma $\hat{\alpha}^{*(P)}(q) = -\dfrac{1}{\ln(2)} \ln\left[\dfrac{-v_{(q)} + v_{(2q)}}{-v_{(2q)} + v_{(4q)}} \right]$ (donde q denota un entero positivo). Se demuestra[202] que: **i)** si q crece a un ritmo adecuado con respecto a n, entonces este estimador es consistente y **ii)** la variable

$$\hat{\alpha}^{**(P)}(q) = \sqrt{q}\left[\hat{\alpha}^{*(P)}(q) - \alpha^* \right] \xrightarrow{d} N\left\{ 0; \frac{\alpha^{*2}\left(2^{-2\alpha^*-1} + 1 \right)}{\left[2\left(2^{-\alpha^*} - 1 \right) \ln(2) \right]^2} \right\}$$

---

201. Para un análisis de métodos de estimación alternativos, ver Harter, H.L.; Moore, A.H. (1968).

202. Ver Dekker, A.L.M.; De Haan, L. (1989).

En particular, para la distribución de Fréchet, y utilizando una metodología similar a la del estimador de Pickands, Hill, B.M. (1975) propuso un estimador para el coeficiente α^* de la forma:

$$\hat{\alpha}^{*(H)}(q) = -\frac{1}{q}\sum_{j=1}^{q}\left[\ln\left(-v_{(j)}\right) - \ln\left(v_{(q+1)}\right)\right]$$

(la selección del valor q difiere de la propuesta por Pickands) el cual resulta más eficiente que el anterior. Se demuestra[203] que la variable:

$$\hat{\alpha}^{**(H)}(q) = \sqrt{q}\left[\hat{\alpha}^{*(H)}(q) - \alpha^*\right] \xrightarrow{d} N\left(0;\alpha^{*2}\right)$$

4.- La **distribución de Pareto-Lévy**[204], cuya función de densidad es de la forma[205]:

$$f_{v_t^*}(v) = \frac{\alpha\delta^\alpha}{v^{\alpha+1}} \qquad (\delta \le v \le \infty ; \delta > 0)$$

donde: **i)** $v_t^* = v_t - U$ (siendo U el valor del "umbral" a partir del cual se definen las "colas")[206]; **ii)** δ denota un factor que caracteriza la escala y representa el menor quebranto significativo y **iii)** α es el exponente característico, también denominado **índice de la cola.**

En su versión original esta distribución puede ser expresada de la siguiente forma:

$$p(v_t \ge v) = \left(\frac{\delta}{v}\right)^\alpha \qquad (\delta > 0; \alpha > 0; v \ge \delta)$$

(donde δ representa el ingreso máximo de la población en estudio). De esta igualdad se obtiene inmediatamente que $F_{v_t}(v) = 1 - \left(\frac{\delta}{v}\right)^\alpha \ (\delta > 0; \alpha > 0; v \ge \delta)$ y, por lo tanto, que esta función constituye una forma especial de la distribución de Pearson-Tipo VI[207].

---

203. Ver Goldie, C.M.; Smith, R.L. (1987).

204. Esta función fue propuesta por Pareto, V. (1897) para representar la distribución del ingreso en una población (es decir, originalmente no fue pensada como una distribución de valores extremos) y, en su versión original puede ser escrita de la siguiente forma: $N = Ax^{-\alpha}$ (donde: **i)** N denota el número de personas de la población con ingresos mayores o iguales a x; **ii)** A es un parámetro; **iii)** α es un coeficiente que caracteriza la forma de la función y es conocido como la **constante de Pareto**; **iv)** A y α son independientes de x, pero dependen de la población). Pareto propuso esta ley como "*universal e inevitable*" e independiente de las condiciones políticas, económicas y sociales de la población, pero fue objeto de muchas críticas (ver Pigou, A.C. (1932)). Posteriormente se conocieron muchos intentos de aplicación en la explicación de distintos fenómenos fácticos (ver, por ejemplo, Steindl, J. (1965), Hagstroem, K.G. (1960)). En particular, Mandelbrot, B.B. (1960)(1963)(1963) propuso esta función como una aproximación a la distribución de las rentabilidades de una activo, a partir de la eliminación de la restricción de que las estimaciones individuales de los operadores constituyen variables aleatorias con varianza finita.

205. Ver Landro, A.H. (2002).

206. Los resultados para un umbral positivo o negativo son simétricos.

207. Ver Landro, A.H. (2002).

Esta distribución se conoce como **de Pareto de primera clase**[208] y de acuerdo con la nomenclatura de Samorodnitsky-Taqqu[209] se la denomina α-estable simétrica [210] (en términos compactos **v:SαS**)[211]. En particular: **i)** si $\alpha = 2$, entonces $v: N(0{:}2)$; **ii)** si $\alpha = 1$, entonces la variable v se distribuye de acuerdo con

---

[208]. Pareto, V, (1897) propuso otras dos formas de distribución representativa de la distribución de las rentas: **i)** la conocida como **de segunda clase** (también conocida como **distribución de Lomax**), con función de distribución definida por $F_{v_t}(v) = 1 - \frac{\delta_2}{(v+c)^{\alpha}}$ (la cual también es una forma especial de la distribución de Pearson-tipo VI) y **ii)** la denominada **de tercera clase**, con función de distribución definida por $F_{v_t}(v) = 1 - \frac{\delta_1 e^{-\beta v}}{(v+c)^{\alpha}}$. Por su parte, Gourieroux, C.; Jasiak, J. (2001) propusieron una forma de distribución de Pareto con función de distribución $F_{v_t}(v) = \begin{cases} \frac{1}{2}\frac{1}{(1-v)^{\alpha}} & (v<0) \\ 1 - \frac{1}{2}\frac{1}{(1+v)^{\alpha}} & (v \geq 0) \end{cases}$ y, en consecuencia, función de densidad de la forma $f_{v_t}(v) = \frac{\alpha}{2(1+|v|)^{\alpha+1}} = \frac{\alpha}{2\sigma\left(1+\frac{|v-m|}{\sigma}\right)^{\alpha+1}}$.

[209]. Ver Samorodnitsky, G.; Taqqqu, M.S. (1994).

[210]. Hasta 1930 los estudios sobre las distribuciones límite se refirieron exclusivamente al análisis de las condiciones mínimas para poder asegurar la convergencia a la distribución Normal. Lévy, P. (1925)(1930)(1931)(1934)(1934)(1935)(1937) demostró que la condición para que una distribución pueda ser considerada límite de la distribución de una suma parcial de variables aleatorias es la de ser **estable**. Dada una sucesión de variables aleatorias $\{v_t\}$ $(t = 1,2,...)$ iid, si existen dos constantes, $a_n > 0$ y b_n para las cuales se verifica que $F_{v_t}(v) = \lim_{n\to\infty} p\left(\frac{\sum_{(n)} - b_n}{a_n} \leq v\right) = p(v_t \leq v)$ $(t = 1,2,...,n)$ donde $\sum_{(n)} = \sum_{t=1}^{n} v_t$ se dice que la distribución $F_V(v)$ es estable (esta familia de distribuciones incluye a la Normal y a la de Cauchy como casos particulares). Estas varaibles poseen la siguiente propiedad: dadas dos variables, v_1 y v_2 , iid, distribuidas de acuerdo con una función estable y dadas las constantes c_1 y c_2 , entonces es posible hallar una tercera constante c tal que $c_1 v_1 + c_2 v_2 = cv$, donde v es una variable con la misma distribución estable que v_1 y v_2 (como corolario de esta propiedad, se puede concluir que $C_{v_1}(c_1, w) C_{v_2}(c_2, w) = C_v(c, w)$, donde $C_v(w)$ denota la función característica de la variable v). A partir de un esquema de razonamiento similar queda definida la familia de las **distribuciones Max-estables**. Sea una sucesión de variables $\{v_t\}$ $(t = 1,2,...)$ iid y sea $v_{Max}(n) = Max\{v_1, v_2, ..., v_n\}$. Si existen dos constantes $a_n > 0$ y b_n tales que: $\lim_{n\to\infty} p\left[\frac{v_{Max}(n) - b_n}{a_n} \leq v\right] = F_{v_t}(v) = F(v)$ se dice que la distribución $F(v)$ es Max-estable (ver Gnedenko, B.V. (1943), Galambos, J. (1995)). De acuerdo con Galambos, J.(1995) y utilizando su notación, estas distribuciones se pueden clasificar en tres tipo s: 1) $G_{1,\alpha}(v) = \begin{cases} \exp\left[-v^{-\alpha}\right] & [v>0] \\ \exp(0) & (v \leq 0) \end{cases}$; 2) $G_{2,\alpha}(v) = \begin{cases} \exp\left[-(-v)^{\alpha}\right] & (v<0) \\ \exp(1) & (v>0) \end{cases}$; 3) $G_{3,\alpha}(v) = \exp\left(-e^{-v}\right)$ $(-\infty < v < \infty)$. Una extensión de la familia de distribuciones estables para suma de variables independientes pero no necesariamente id, es la familia de las variables infinitamente divisibles (ver Moran, P.A.P. (1968), Galambos, J. (1995)).

[211]. "Symmetric α-stable".

una función de Cauchy[212].

Sus momentos absolutos son de la forma $m_s(v) = \frac{\alpha\delta^s}{\alpha - s}(s < \alpha)$. A partir de los cuales se obtienen las medidas: $E(v_t) = \frac{\alpha\delta}{\alpha - 1}(\alpha > 1)$, $mo(v_t) = \delta$, $me(v_t) = \delta\alpha^{1/\alpha}$, $\sigma^2(v_t^*) = \frac{\alpha\delta^2}{(\alpha - 2)(\alpha - 1)}$ $(\alpha > 2)$, $As(v_t^*) = \frac{2(\alpha + 1)\sqrt{\alpha - 2}}{(\alpha - 3)\sqrt{\alpha}}(\alpha > 3)$ y $K(v_t^*) = \frac{3(3\alpha^2 + \alpha + 2)(\alpha - 2)}{\alpha(\alpha - 4)(\alpha - 3)}$ $(\alpha > 4)$.

Obsérvese que: **i)** si $\alpha = 1$, entonces $E(v)$ no existe; **ii)** si $\alpha > 1, E(v) < \infty$; **iii)** si $1 < \alpha < 2, \sigma^2(v) = \infty$; **iv)** si $\alpha < 2, \sigma^2(v) < \infty$. En particular, si v puede asumir solamente valores no-negativos, se verifica que: **i)** si $\alpha = 1, E(v) = \infty$; **ii)** si $\alpha > 1, E(v) < \infty$; **iii)** si $\alpha \leq 2, \sigma^2(v) = \infty$; **iv)** si $\alpha > 2, \sigma^2(v) < \infty$.

Dado que la variable v no posee momentos de orden mayor o igual que α $E(|v|^s) = \infty$ $(s \geq \alpha)$, se puede concluir que, de acuerdo con la experiencia, en las aplicaciones al mercado de capitales, esta distribución presenta kurtosis infinita, lo cual implica que su estimador debe crecer a medida que aumenta el tamaño de la muestra (ver Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C. (1997))[213].

Si $\sigma^2(v) = \infty$, la distribución es de **variación regular en el infinito**:

$$\lim_{\tau \to \infty} \frac{1 - F(\tau, v)}{1 - F(\tau)} = v^{-\alpha}$$

---

212. Ver Landro, A.H. (2002).

213. Como se vio, esta distribución está caracterizada por dos parámetros (d y α). En el caso más general, en el que la simetría es con respecto a un valor distinto de cero, la distribución Paretiana estable requiere para su caracterización de dos parámetrica adicionales. Lévy, P. (1925) obtuvo la siguiente definición del logaritmo de la función característica de la variable estable generalizada (dada la dificultad para expresar la distribución de las variables estables mediante su función de densidad, habitualmente se las define a partir de su función característica)

$$\ln(C_v(w)) = \begin{cases} i\delta w - d|w|^\alpha \left\{1 - i\beta\left[signo(w)\, tg\left(\alpha\frac{\pi}{2}\right)\right]\right\} (-\infty \leq \delta \leq \infty; 0 \leq d < \infty; 0 < \alpha \leq 2; \alpha \neq 1; -1 \leq \beta \leq 1) \\ i\delta w - d|w|^\alpha \left\{1 + i\beta\left[signo(w)\right]\frac{2}{\pi}\ln(|w|)\right\} (-\infty \leq \delta \leq \infty; 0 \leq d < \infty; \alpha = 1; -1 \leq \beta \leq 1) \end{cases}$$

. Donde:

i) $signo(w) = w/|w|$, si $w \neq 0$ e igual a cero si $w = 0$; **ii)** δ es el **parámetro de posición**; **iii)** σ es el **parámetro de escala**; **iv)** α es el **exponente característico** y **v)** β es el **índice de asimetría**. Si $\beta = 0$, la distribución es simétrica con respecto a δ; que β sea positivo o negativo significa que la distribución es asimétrica hacia la derecha o hacia la izquierda, respectivamente; si $\alpha = 2$ entonces $v: N(\delta, 2d)$; si $1 < \alpha < 2$, entonces $E(v) = \delta$; si se verifica que $\beta = 0$ y $\alpha = 1$, se obtiene la **distribución de Cauchy** con función de densidad de la forma $f(v) = \frac{1}{\pi}\frac{d}{d^2 + (v - d)^2}$ $(-\infty \leq v \leq \infty)$ y, si se verifica que $\beta = 1, \delta = 0, d = 1$ y $\alpha = \frac{1}{2}$, se obtiene la distribución de Bernoulli. A partir de estos resultados se puede concluir en forma inmediata que esta generalización complica la definición y no mejora la explicación del comportamiento de las colas de la distribución (ver Fama, E. (1965), Mandelbrot, B.; Taylor, H. (1967), Samuelson, P. (1967)(1976), Granger, C.W.J.; Morgenstgern, O. (1970), Fama, E.; Roll, R. (1971), Officer, R. (1972), Blattberg, R.; Gonedes, N. (1974), Hsu, D.R.; Miller, R.; Wichern, D. (1974), Fielitz, B. (1976), Hagerman, R. (1978), Simkowitz, M.; Beedles, W. (1980), Fielitz, B.; Rozell, J. (1983), Tucker, A. (1992), Landro, A.H. (2002)).

condición necesaria y suficiente para que observe un comportamiento "fat tailed"[214]:

$$p(v \geq v^*) = p(v \leq -v^*) = (\sigma^*)^{\alpha}(v^*)^{-\alpha} \qquad (v^* \to \infty, 0 < \alpha < 2)$$

(ver Feller, W. (1966), Koedijk, K.G.; Schafgans, M.M.A.; de Vries, C.G. (1990)) donde σ^* denota una constante denominada habitualmente como la **dispersión de** v (obsérvese que esta propiedad implica: **i)** un decrecimiento potencial hacia las colas, no exponencial como en el caso de la distribución Normal y **ii)** que las "colas" serán más significativas cuanto menor sea α).

Del logaritmo de la función de verosimilitud para un conjunto de observaciones $(v_1, v_2, ..., v_n)$, $L = \prod_{t=1}^{n_k} \frac{\alpha \delta^{\alpha}}{v_t^{\alpha+1}}$, se obtiene el siguiente estimador del coeficiente α: $\hat{\alpha} = \frac{n_k}{\sum_{t=1}^{n_k} \log\left(\frac{v_t}{\hat{\delta}}\right)}$. Debe tenerse en cuenta que, dado que la función $\log(L) = \ell$ no es acotada con respecto a δ, no se puede obtener en la forma habitual una segunda ecuación de la forma $\frac{\partial \ell}{\partial \delta} = 0$. Luego, dado que δ es la cota inferior de la variable v, la función $\log(L)$ debe ser maximizada sujeta a la restricción $\hat{\delta} \leq \min(v_1, v_2, ..., v_{n_k})$, obteniéndose el siguiente estimador del coeficiente δ: $\hat{\delta} = \min(v_1, v_2, ..., v_{n_k})$ (debe tenerse en cuenta que este método de la máxima verosimilitud es aplicable cuando el número de observaciones es suficientemente grande y los valores de rentabilidad que superan un umbral, en cualquiera de los dos sentidos, tienen baja frecuencia)[215].

Suponiendo que δ es conocido, el estimador máximo-verosímil de $me(v)$ es de la forma $\hat{me}(v) = \delta 2^{\log(\hat{mg}(v)/\delta)}$. Donde $\hat{mg}(v)$ denota el estimador máximo-verosímil del promedio geométrico de v, $\hat{mg}(v) = \delta e^{1/\alpha}$ (ver Muniruzzaman, A.N.M. (1957)).

El estimador $\hat{\alpha}$ tiene una distribución de probabilidades de la forma:

---

[214]. Debe tenerse en cuenta que una función $f(x)$ es de **variación regular en el infinito** con índice λ, si se verifica que $\lim_{t \to \infty} \frac{f(tx)}{f(t)} = x^{\lambda}$ $(para\ x > 0)$ (es decir, si se verifica que $f(x)$ es asintóticamente una función potencial). De la misma forma, se dice que $f(x)$ es de **variación lenta en el infinito** si se verifica que $\lim_{t \to \infty} \frac{f(tx)}{f(t)} = cte$. Un ejemplo típico de esta última clase de funciones es $f(x) = \log(x)$. Ver Resnick, J. (1987).

[215]. Para la variable estable generalizada se puede escribir $\lim_{v^* \to \infty} p(v > v^*) = p\sigma^*(v^*)^{-\alpha}$ y $\lim_{v^* \to \infty} p(v \leq v^*) = q\sigma^*(v^*)^{-\alpha} = (1-p)\sigma^*(v^*)^{-\alpha}$ (donde $0 < \alpha < 2$ y $0 \leq p \leq 1$). Las condiciones necesarias y suficientes para que se verifiquen las igualdades anteriores son las siguientes $p = \frac{1+\beta}{2}$; $\sigma^* = \begin{cases} \frac{d}{\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)} & (\alpha \neq 1) \\ \frac{2d}{\pi} & (\alpha = 1) \end{cases}$.

$$f_{\hat{\alpha}}(v) = \frac{\alpha^{n-1} n^{n-1}}{v^n \Gamma(n-1)} \qquad (v \geq 0)$$

que es asintóticamente Normal (obsérvese que $\frac{2\pi\alpha}{\hat{\alpha}} : \chi^2_{(2n_k)}$). Además, se verifica que $E(\hat{\alpha}) = \frac{n_k \alpha}{n_k - 2}$ y que $\sigma^2(\hat{\alpha}) = \frac{n_k^2 \alpha^2}{(n_k - 2)(n_k - 3)}$ y se demuestra que son estocásticamente independientes (ver Malik, H.J. (1960)).

Utilizando el método de los momentos, para $\alpha > 1$, se obtienen los siguientes estimadores: $\hat{\alpha}^{(m)} = \frac{n_k \bar{v}_t - v_1^*}{n_k (\bar{v}_t - v_1^*)}$ y $\hat{\delta} = \frac{(\hat{\alpha}^{(m)} n_k - 1)}{\hat{\alpha}^{(m)} n_k}$, donde v_1^* denota el estadístico de primer orden, es decir, el primer valor de la sucesión de observaciones $v_t^* \; (t = 1,2,...,n_k)$ ordenadas de menor a mayor[216].

A partir de la expresión $\log[1 - F_{v_t}(v)] = \alpha \log(\delta) - \alpha \log(v)$, considerando a $\log[1 - F_{v_t}(v)]$ como variable dependiente, se obtienen los siguientes estimadores mínimo-cuadráticos de los coeficientes de la distribución:

$$\hat{\alpha}^{(CM)} = \frac{-n_k \sum_{t=1}^{n_k} \log(v_t) \log[1 - F_{v_t}(v)] + \left[\sum_{t=1}^{n_k} \log(v_t)\right]\left[\sum_{t=1}^{n_k} \log(1 - F_{v_t}(v))\right]}{n_k \sum_{t=1}^{n_k} [\log(v_t)]^2 - \left[\sum_{t=1}^{n_k} \log(v_t)\right]^2}$$

$$\hat{\delta}^{(CM)} = \exp\left\{\frac{1}{n_k}\left(\frac{\sum_{t=1}^{n_k} \log[1 - F_{v_t}(v)]}{\hat{\alpha}^{(CM)}} + \sum_{t=1}^{n_k} \log(v_t)\right)\right\}$$

Se demuestra que los tres tipos de estimadores analizados son consistentes[217].

En la Sec. 2.14.2 se analizaron tests para decidir acerca de la adecuacidad de una función de probabilidades cualquiera considerada como de referencia, para representar la distribución de un conjunto de observaciones. Con relación al diagrama Q-Q, debería esperarse que, con respecto al comportamiento de los activos, los puntos del diagrama observaran un ordenamiento no-lineal caracterizado por un comportamiento creciente, por encima de la recta $Q(p) = Q_0(p)$ (donde $Q(p)$ y $Q_0(p)$ denotan, respectivamente, los cuantiles de orden p de la distribución de las observaciones v_t y de la correspondiente distribución Normal) en el extremo superior de la distribución y descendente (por debajo de la recta $Q(p) = Q_0(p)$) en el extremo inferior (es decir que los cuantiles de menor orden son menores que los de la función Normal y, viceversa, los de mayor orden superiores a los de la Normal).

---

216. Ver Quandt, R.E. (1966).

217. Para un análisis detallado de los problemas de estimación en la función de Pareto, ver Hagstroem, K.G. (1960), Malik, H.J. (1960), Quandt, R.E. (1966), Scala, C. (1967)(1967), Koutrouvelis, I.A. (1980). Para un tratamiento del problema de estimación para valores extremos a partir de un umbral, ver Pickands, J. (1975), Bernardo, J.M.; Smith, A.F.M. (1994).

4.7.7.3.- Los métodos de predicción en procesos autorregresivos con varianza infinita

La caracterización de una distribución como estable se basa en la propiedad de **invariancia con respecto a la adición de variables aleatorias**, la cual expresa que, en principio, la suma de un número finito de variables estables dá origen a una variable estable $v_t^{(k)} = \sum_{j=0}^{k} v_{t-j}$ con el mismo exponente característico. De modo que suponer que las rentabilidades diarias poseen distribución estable implica que las rentabilidades semanales, mensuales, etc. (consideradas como sumas de rentabilidades diarias) también están distribuidas de acuerdo con una función estable con el mismo coeficiente α.

Un análisis de los procesos estables en el dominio del tiempo permite demostrar[218] que, dado un proceso $\{v_t\}$ formado por variables aleatorias independientes todas con distribución estable con varianza infinita, la condición suficiente para que el proceso $\{Z_t\}$ admita una representación MA de la forma:

$$Z_t = \sum_{j=0}^{\infty} \psi_j v_{t-j} = \left(\sum_{j=0}^{\infty} \psi_j B^j \right) v_t = \Psi(B) v_t = \frac{\Theta(B)}{\Phi(B)} v_y$$

(donde $\Theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ y $\Phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$) es que $\sum_{j=0}^{\infty} |\psi_j|^{\alpha} < \infty$ $(0 < \alpha < 2)$. Como corolario de esta demostración y, de acuerdo a lo expresado más arriba, se verificará que $\lim_{v^* \to \infty} p\left(\left| \sum_{j=0}^{\infty} \psi_j v_{t-j} \right| > v^* \right) = \frac{1}{(v^*)^{\alpha}} \left(\sum_{j=0}^{\infty} |\psi_j|^{\alpha} \right) \sigma^{**}$, donde $\sigma^{**} = \left(\sum_{j=0}^{\infty} |\psi_j|^{\alpha} \right) \sigma^{*}$. Es decir que el proceso autorregresivo que se obtiene a partir de una representación MA de un proceso estable, también es estable[219]. Igual que en el caso de los procesos $ARMA(p,q)$ (ver Sec.2.1.12), si las raíces de la ecuación característica $\Theta(B) = 0$ son, en módulo, mayores que la unidad, entonces $\{Z_t\}$ también será invertible.

Debe tenerse en cuenta que, si bien el proceso $\{Z_t\}$ es estrictamente estacionario, dado que $E(|Z_t|^2) = \infty$, no es estacionario de segundo orden. No obstante, se puede definir una función similar a la función de autocorrelaciones de la forma:

$$\rho_j^*(Z) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+j}}{\sum_{i=0}^{\infty} \psi_i^2} \qquad (j = 1,2,...)$$

(obsérvese que, si $\{Z_t\}$:WN - *debil* y, por lo tanto, tiene varianza finita, esta función coincide con la función de autocorrelaciones), para la cual se puede utilizar el siguiente estimador:

---

218. Ver Mandelbrot, B.B. (1963)(1963), Feller, W. (1966), Brockwell, P.J.; Davis, R.A. (1987).

219. Ver Cline, D.B.H. (1983), Cline, D.B.H.; Brockwell, P.J. (1985)

$$\hat{\rho}_j^*(Z) = \begin{cases} \dfrac{\sum_{t=1}^{n-j} Z_t Z_{t+j}}{\sum_{t=1}^{n} Z_t^2} & \left(\alpha > 1; j = 1,2,\ldots,\dfrac{n}{4}\right) \\ \dfrac{\sum_{t=1}^{n-j}\left(Z_t - \overline{Z}\right)\left(Z_{t+j} - \overline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2} & \left(\alpha > 1; \overline{Z} = \dfrac{1}{n}\sum_{t=1}^{n} Z_t ; j = 1,2,\ldots,\dfrac{n}{4}\right) \end{cases}$$

que, de acuerdo con los postulados del teorema de Davis-Resnick[220], posee la propiedad de ser consistente:

$$\hat{\rho}_j^*(Z) \xrightarrow{p} \rho_j(Z)$$

o, más específicamente, tal que $\hat{\rho}_j^*(Z) - \rho_j^*(Z) = O\left[\left(\dfrac{n}{\ln(n)}\right)^{-1/\alpha}\right] (\delta > \alpha)$ (obsérvese que esta tasa de convergencia es mayor que la tasa $O\left(n^{-1/2}\right)$, correspondiente a la diferencia $\hat{\rho}_j(Z) - \rho_j(Z)$ de las representaciones con varianza finita (ver Sec. 2.8)). El teorema de Davis-Resnick demuestra también que la distribución asintótica de este estimador de $\rho_j^*(Z)$ puede ser obtenida a partir de la variable:

$$\lim_{n\to\infty}\left[\frac{n}{\ln(n)}\right]^{1/\alpha}\left[\hat{\rho}_j^*(Z) - \rho_j^*(Z)\right] = \left(\sum_{i=1}^{\infty}\left|\rho_{i+j}(Z) + \rho_{i-j}(Z) - 2\rho_i(Z)\rho_j(Z)\right|^{\alpha}\right)\frac{U}{V} \qquad (V \geq 0)$$

donde U y V son variables aleatorias independientes con funciones características:

$$C_U(w) = \begin{cases} \exp\left[-\Gamma(1-\alpha)\cos\left(\dfrac{\alpha\pi}{2}\right)|u|^{\alpha}\right] & (\alpha \neq 1) \\ \exp\left(\dfrac{\pi}{2}|u|\right) & (\alpha = 1) \end{cases}$$

$$C_V(w) = \exp\left\{-\Gamma\left(1-\frac{\alpha}{2}\right)\cos\left(\frac{\pi}{4}\alpha\right)|u|^{\alpha/2}\left[1 - i\ signo(u)\ tg\left(\frac{\pi}{4}\alpha\right)\right]\right\}$$

Excepto en el caso en que $\alpha = 1$, para el cual se verifica que: **i)** la variable U tiene distribución de probabilidades de Cauchy de la forma $f_U(u) = \dfrac{1}{2}\left(\dfrac{n^2}{4} + u^2\right)^{-1} \left(-\infty \leq u \leq \infty\right)$; **ii)** la variable V tiene una distribución estable no-negativa con función de densidad de la forma $f_V(v) = \dfrac{2}{2}v^{-1/2}e^{-n/4v} \left(v \geq 0\right)$ (ver

---

220. Ver Davis, R.A.; Resnick, S.I. (1986).

Feller, W. (1971)) y, en consecuencia, **iii)** la función de distribución de la variable $\frac{U}{V}$ es de la forma:

$$F_{U/V}(x) = \int_0^\infty F_U(xy) f_V(y) dy = \frac{1}{\sqrt{2}} \int_0^\infty \frac{arco\ tg(xu)}{(\pi u)^{3/2}} e^{-1/2u} du$$

la distribución conjunta de la variable bidimensional (U,V) no puede ser expresada en términos de su función de densidad (obsérvese que, en el caso que $\alpha = 1$, la distribución de probabilidades de la variable $\frac{U}{V}$ coincide con la distribución conjunta de una variable bidimensional definida por el producto de dos variables independientes: una variable de Cauchy estandarizada con función de densidad de la forma $\frac{1}{\pi(1+x^2)}$ y una variable $\chi^2_{(1)}$).

Supóngase, en particular, que el comportamiento del proceso $\{Z_t\}$ pueda ser aproximado por una representación $MA(q)$ de la forma:

$$Z_t = v_t - \theta_1 v_{t-1} - ... - \theta_q v_{t-q}$$

donde $\{v_t\}$ es un proceso Paretiano estable con $0 < \alpha < 2$. Entonces se verifica que $\rho_j^*(Z) = 0$ para $|j| > q$. El estimador asintótico de esta función será de la forma[221]:

$$\lim_{n\to\infty} \left(\frac{n}{\ln(n)}\right)^{1/\alpha} \left[\hat{\rho}_j^*(Z) - \rho_j^*(Z)\right] = \left(1 + 2\sum_{j=1}^{q} \left|\rho_j^*(Z)\right|^{1/\alpha} \frac{U}{V}\right) \qquad (j > q)$$

Sea, ahora, un proceso $\{Z_t\}$ estrictamente estacionario que admite una representación de la forma $Z_t = \Psi(B)v_t$ y supóngase que $\{v_t\}$ sea un proceso Paretiano estable simétrico con $0 < \alpha < 2$ que representa las rentabilidades de un activo. El predictor lineal de Z_{t+1}, a partir de una serie cronológica $\{Z_1, Z_2, ..., Z_t\}$, será de la forma $f^{(1,t)} = \varphi_1 Z_t + \varphi_2 Z_{t-1} + ... + \varphi_t Z_1$. Ahora bien, dado que, de acuerdo a lo demostrado más arriba, $\{Z_t\}$ es un proceso de varianza infinita no se puede considerar a la varianza del error de predicción, $E\left[\left(Z_{t+1} - f^{(1,t)}\right)^2\right]$, como un criterio para la selección del mejor predictor. Teniendo en cuenta la relación $Z_t = \sum_{j=0}^{\infty} \psi_j v_{t-j}$, se puede escribir:

$$f^{(1,t)} = \sum_{j=0}^{\infty} \left(\varphi_1 \psi_j + \varphi_2 \psi_{j-1} + ... + \varphi_t \psi_{j-(t-1)}\right) v_{t-j}$$

lo cual permite concluir que el error de predicción es simétrico estable y está dado por:

---

221. En particular, si $q = 0$, entonces el segundo miembro de esta expresión será igual a $\frac{U}{V}$.

$$Z_{t+1} - f^{(1,t)} = v_{t+1} + \sum_{j=0}^{\infty}\left(\psi_{j+1} - \varphi_1\psi_j - \varphi_2\psi_{j-1} - \ldots - \varphi_t\psi_{j-(t-1)}\right)v_{t-j}$$

con un desvío estándar:

$$\sigma\left(Z_{t+1} - f^{(1,t)}\right) = \left(1 + \sum_{j=0}^{\infty}\left|\psi_{j+1} - \varphi_1\psi_j - \ldots - \varphi_t\psi_{j-(t-1)}\right|^{\alpha}\right)\sigma^*$$

(donde, como se mencionó en páginas anteriores, σ^* denota una constante positiva, habitualmente denominada como la dispersión de v_t) y que minimizar esta dispersión implica minimizar el factor escala y, por lo tanto, definir un criterio de optimización alternativo que consiste en minimizar:

$$p\left(\left|Z_{t+1} - f^{(1,t)}\right| > \varepsilon\right) \qquad (\varepsilon > 0)$$

Debe tenerse en cuenta que, contrariamente a lo que ocurre cuando $\alpha = 2$, para $0 < \alpha < 2$, el predictor obtenido de aplicar este criterio, en general, no es único.

Este criterio de predicción **de dispersión mínima** también es utilizado en modelos de regresión asociados con sucesiones estables. En general se demuestra que, si el proceso $\{v_t\}$ satisface las (ya mencionadas) condiciones de las variables estables generalizadas ($\lim_{v^* \to -\infty} p\left(v > v^*\right) = p\sigma^*\left(v^*\right)^{-\alpha}$ y:

$$\lim_{v^* \to \infty} p\left(v \le v^*\right) = q\sigma^*\left(v^*\right)^{-\alpha} = (1-p)\sigma^*\left(v^*\right)^{-\alpha}$$

donde $0 < \alpha < 2$ y $0 \le p \le 1$), entonces el criterio de dispersión mínima, minimiza las probabilidades de las "colas" de la distribución del error de predicción[222].

Supóngase, por ejemplo, un proceso $\{Z_t\}$ cuyo comportamiento puede ser aproximado por una representación $MA(1)$ de la forma:

$$Z_t = v_t - \theta_1 v_{t-1} \qquad (|\theta_1| < 1)$$

Suponiendo que $\{v_t\}$:*WN – debil* y que se cuente con una serie cronológica de longitud t , $\{Z_1, Z_2, \ldots, Z_t\}$, de acuerdo a lo expresado en la Sec. 2.11, el predictor óptimo (de error medio cuadrático mínimo) de Z_{t+1} será de la forma:

$$f^{(1,t)} = \varphi_1 Z_t + \varphi_2 Z_{t-1} + \ldots + \varphi_t Z_1$$

Teniendo en cuenta que los coeficientes $\varphi_j \ (j = 1,2,\ldots,t)$ satisfacen las ecuaciones en diferencias:

---

222. Ver Blattberg, R.; Sargent, T. (1971)). Para el análisis de la estimación de $Z_{t+\ell}$ $(\ell > 1)$ ver Cline, D.B.H.; Brockwell, P.J. (1985).

$$-\theta_1\varphi_{j-1}+\left(1+\theta_1^2\right)\varphi_j-\theta_1\varphi_{j+1}=0 \qquad (j=2,3,\ldots,t-1)$$

$$\left(1+\theta_1^2\right)\varphi_1-\theta_1\varphi_2=-\theta_1$$

$$\left(1+\theta_1^2\right)\varphi_t-\theta_1\varphi_{t-1}=0$$

este predictor óptimo puede ser expresado de la siguiente forma:

$$f^{(1,t)}=\frac{1}{1-\theta_1^{2(t+1)}}\sum_{j=1}^{t}\left[\left(-\theta_1\right)^j-\left(-\theta_1\right)^{2(t+1)-j}\right]Z_{t+1-j}$$

De modo que resulta:

$$Z_{t+1}-f^{(1,t)}=v_{t+1}-\frac{\left(-\theta_1\right)^{t+1}\left(1-\theta_1^2\right)}{1-\theta_1^{2(t+1)}}\sum_{j=0}^{t}\left(-\theta_1\right)^j v_j$$

Supóngase, ahora, que en esta representación $MA(1)$, $\{v_t\}$ sea un proceso formado por variables iid con distribución de Cauchy, el predictor de desvío estándar mínimo de Z_{t+1} será de la forma $f^{(1,t)}=-\sum_{j=1}^{t}\left(-\theta_1\right)^j Z_{t+1-j}$ y el desvío estándar del error de predicción será $\sigma\left(Z_{t+1}-f^{(1,t)*}\right)=\frac{2}{\pi}\left(1+\left|\theta_1\right|^{t+1}\right)$. Obsérvese que, si este predictor es aplicado al proceso de Cauchy, el desvío estándar del error de predicción será tal que $\sigma\left(Z_{t+1}-f^{(1,t)*}\right)=\frac{2}{\pi}\left(1+\left|\theta_1\right|^{t+1}\frac{1+\left|\theta_1\right|}{1+\left|\theta_1\right|^{t+1}}\right)>\sigma\left(Z_{t+1}-f^{(1,t)}\right)$. De modo que se puede concluir que, en el caso particular de una representación $MA(1)$ formada por una combinación lineal de variables de Cauchy, el predictor definido para un proceso $\{Z_t\}$ estacionario de segundo orden no es óptimo (no produce una varianza de error de predicción mínima).

Aquí se han considerado solamente los métodos de predicción lineal, pero debe tenerse en cuenta que la condición de convergencia absoluta de los coeficientes ψ_j ($\sum_{j=0}^{\infty}\left|\psi_j\right|^{\alpha}<\infty$) permite mejorar la predicción de los procesos autorregresivos con varianza infinita a partir de la utilización de predictores no-lineales. En la Parte III de este trabajo se tratarán los modelos no-lineales, en particular los **modelos con umbral** y los **modelos GARCH**, útiles para explicar el comportamiento autorregresivo de las varianzas condicionadas, es decir de las volatilidades de los activos.

Supóngase que, de acuerdo con las hipótesis clásicas, el proceso que representa a las rentabilidades pueda ser aproximado mediante una representación de la forma:

$$v_t=m+\sigma_t U_t$$

Donde: **i)** $\{U_t\}=\left\{\frac{v_t-m}{\sigma_t}\right\}:WN-N(0;1)$; **ii)** $\{\sigma_t^2\}$ es un proceso formado por variables aleatorias positivas que definen las varianzas condicionadas de las variables v_t y que admite una representación $ARCH(1)$ de

la forma $\sigma_t^2 = \sigma^2(v_t / \sigma_t) = \varphi_0 + \varphi_1 v_{t-1}^2$ (los procesos ARCH serán tratados en la Parte III) y **iii)** $U_t \perp v_t$ $(t = 1,2,...)$.

La función de autocovarianzas del proceso $\{v_t\}$ será, entonces, de la forma:

$$\gamma(v_{t-j}, v_t) = E\left[(v_{t-j} - m)(v_t - m)\right] = E\left[(\sigma_{t-j}U_{t-j})(\sigma_t U_t)\right] = E\left[(\sigma_{t-j}\sigma_t)U_{t-j}U_t\right] = 0 \qquad (j \neq 0)$$

En particular, para $j = 0$, será:

$$\sigma^2(v_t) = E\left[(v_t - m)^2\right] = E(\sigma_t^2 U_t^2) = E(\sigma_t^2)E(U_t^2) = E(\sigma_t^2)$$

por lo que se puede concluir que $\{v_t\}$: $WN - debil$. Entonces se puede escribir $v_t^2 = \varphi_0 U_t^2 + \varphi_1 U_t^2 v_{t-1}^2$. Luego, la función de autocovarianzas del proceso $\left\{(v_t - m)^2\right\} = \left\{\sigma_t^2 U_t^2\right\} = \{w_t\}$ será de la forma:

$$\gamma(w_{t-j}, w_t) = E\left\{\left[w_{t-j} - E(w_t)\right]\left[w_t - E(w_t)\right]\right\} = E(w_{t-j}w_t) - \left[E(w_t)\right]^2 =$$

$$= E(\sigma_{t-j}^2\sigma_t^2)E(U_{t-j}^2U_t^2) - \left[E(\sigma_t^2)\right]^2 =$$

$$= E(\sigma_{t-j}^2\sigma_t^2) - \left[E(\sigma_t^2)\right]^2 \qquad (j = 0,\pm 1,\pm 2,...)$$

y, por lo tanto, la función de autocorrelaciones queda definida como:

$$\rho(w_{t-j}, w_t) = \frac{E(\sigma_{t-j}^2\sigma_t^2) - \left[E(\sigma_t^2)\right]^2}{E(\sigma_t^4) - \left[E(\sigma_t^2)\right]^2} \qquad (j = 0,\pm 1,\pm 2,...)$$

Lo que demuestra que, si bien el proceso $\{v_t\}$ es un $WN - debil$, contrariamente a lo supuesto por las hipótesis clásicas de independencia de las variables v_t, el proceso $\left\{(v_t - m)^2\right\}$ presenta un comportamiento autorregresivo[223] (es decir que, tal como se comentó en la Sec. 4.3, el proceso $\{Z_t\}$ que, en este caso representa la evolución de los precios, no se comporta como un RW, sino como una martingala).

De acuerdo a lo expresado en las páginas precedentes, habitualmente, el proceso $\{v_t\}$: $ARCH(1)$ tiene varianza infinita con un exponente característico (β) definido por la ecuación $\Gamma\left(\frac{\beta+1}{2}\right) = \sqrt{\frac{\pi}{(2\varphi_1)^\beta}}$, donde $\varphi_1 < 2e^{\nu} = 3.56856$, siendo $\nu = 0.578069$ la llamada constante de Euler (obsérvese que el coeficiente β que, como se vio, determina el orden máximo de los momentos finitos, puede ser definido para

---

[223]. A partir de una demostración similar se puede concluir que el proceso $\{|v_t - m|\}$ también presenta autocorrelaciones no-nulas. Esta caso será analizado con más detalle en el Cap. 5, al tratar los procesos ARFIMA.

distribuciones distintas de la estable y no tiene por qué coincidir con el exponente α). Se demuestra que: **i)** para $\varphi_1 = 1$ se verifica que $\beta = 2$; **ii)** para $\varphi_1 = 0$ se verifica que $\beta = \infty$ (por lo que se puede concluir que en estos dos casos las colas de la distribución se comportan en forma similar a las de la función Normal); **iii)** para $\varphi_1 = 2e^{\nu}$ se verifica que $\beta = 0.00279$ (por lo que se puede concluir que, en lo que hace al comportamiento de las colas, los procesos estable y ARCH coinciden parcialmente)[224]; **iv)** para $0 < \varphi_1 < 1$ se verifica que $\beta > 2$ (por lo que se puede concluir que, si $\beta \geq 2$, $\{v_t\}$ es débilmente estacionario de segundo orden y, por lo tanto, con varianza finita, de modo que no tiene distribución estable pero tampoco es necesariamente Normal: puede tener una distribución $t_{(\beta)}$ y puede tener "colas" levantadas); **v)** para $\beta < 2$, la varianza del proceso es infinita y puede ser representado por una distribución estable con exponente característico $\alpha = \beta$ y **vi)** para $\beta < 0.00279$, el proceso $\{v_t\}$ no admite representación ARCH (ver de Haan, L.; Resnick, S.I.; Rootzén, H.; de Vries, C.G. (1989)).

En conclusión, para toda distribución que observe un comportamiento "fat tailed" y que posea la propiedad de variación regular en el infinito, la distribución de las colas converge en distribución a la función estable, de modo que se puede escribir[225]:

$$p\left(v > v^*\right) = \left(\sigma^*\right)^{\beta}\left(v^*\right)^{-\beta}\left[1 + \beta_D\left(v^*\right)\right] \qquad \left(v^* > 0\right)$$

$$p\left(v < -v^*\right) = \left(\sigma^*\right)^{\beta}\left(v^*\right)^{-\beta}\left[1 - \beta_I\left(v^*\right)\right] \qquad \left(v^* > 0\right)$$

donde $\lim_{v\to\infty} \beta_D(v) = \lim_{v\to\infty} \beta_I(v) = 0$.

A partir de los estadísticos de orden (ascendente) $v_{(1)} \leq v_{(2)} \leq ... \leq v_{(n)}$ obtenidos de la serie cronológica $\{v_t\}$ $(t = 1,2,...,n)$, se define el siguiente estimador de β (ver de Haan, L.; Resnick, S.I. (1980), Groenendijk, P.A.: Lucas, A.; de Vries, C.G. (1995)):

$$\hat{\beta} = \frac{\tau}{\sum_{j=1}^{\tau}\left[\log\left(v_{(n-j+1)}\right) - \log\left(v_{(n-\tau)}\right)\right]}$$

donde $\lim_{n\to\infty} \tau(n) = \infty$. Obviamente, para la utilización de este estimador en las aplicaciones, es necesario fijar un valor finito para $\tau(n)$. Una solución a este problema, propuesta por Koedijk, K.J.; Schafgans, M.M.A.; de Vries, C.G. (1990) (basada en métodos de simulación de Monte Carlo) consiste en seleccionar el valor de τ que minimice el error medio cuadrático del estimador de β . Por su parte, Phillips, P.C.B.; McFarland, J.W.; McMahon, P.C. (1996), a partir de los trabajos de Hall, P.; Welsh, A.H. (1985), propusieron una solución asintótica[226] según la cual la minimización del error medio cuadrático del límite de la distribución

---

224. Ver Groenedijk, P.A.; Lucas, A.; de Vries, C.G. (1995).

225. Ver Loretan, M.; Phillips, P.C.B. (1994).

226. Ver Hall, P. (1982).

del estimador $\hat{\beta}$ se obtiene seleccionando $\tau(n) = \lambda n^{2/3}$, donde el coeficiente λ puede ser estimado en forma iterativa de acuerdo con la expresión $\hat{\lambda} = \left| \frac{\hat{\beta}_1}{\sqrt{2}} \frac{n\left(\hat{\beta}_1 - \hat{\beta}_2\right)}{\tau_2} \right|^{2/3}$, siendo $\hat{\beta}_1$ y $\hat{\beta}_2$ los estimadores iniciales de β, obtenidos a partir de los valores $\tau_1 = n^{\alpha_1}$ y $\tau_2 = n^{\alpha_2}$ $\left(0 < \alpha_1 < \frac{2}{3} < \alpha_2 < 1\right)$. Phillips, P.C.B.; McFarland, J.W.; McMahon, P.C. (1996) recomiendan, en particular, que $\alpha_1 = 0.06$ y $\alpha_2 = 0.9$ y Loretan, M.; Phillips, P.C.B. (1994) recomiendan que $\tau \leq \frac{n}{10}$ (debe tenerse en cuenta que estos estimadores están relacionados con la cola superior de la distribución de v_t, para los estimadores relacionados con la cola inferior basta con multiplicar por -1 los mencionados estadísticos de orden).

De acuerdo a las condiciones analizadas más arriba, con respecto al coeficiente β, hay dos problemas que resultan particularmente interesantes:

i) Decidir si la hipótesis $H_0: \beta < 2$ es verdadera, con respecto a la alternativa $H_1: \beta \geq 2$ (debe tenerse en cuenta que, como se mencionó en la página precedente, el valor $\beta = 2$ divide el dominio de existencia de $\{v_t\}$ en dos clases: con varianza infinita ($\beta < 2$) y con varianza finita ($\beta \geq 2$)). El estadístico a utilizar será $\left(\hat{\beta} - \beta\right)\sqrt{\tau} \xrightarrow{d} N\left(0, \beta^2\right)$.

ii) Decidir acerca de la constancia del índice β. Es decir, dados dos índices ($\beta_{(1)}$ y $\beta_{(2)}$), decidir si la hipótesis $H_0: \beta_{(1)} = \beta_{(2)} = \beta$ es verdadera, a partir de la información que proporcionan los correspondientes estimadores, obtenidos de dos muestras independientes. Una solución a este problema se obtiene utilizando el estadístico $\left(\frac{\beta_{(1)}}{\hat{\beta}_{(1)}} - 1\right)^2 \tau_1 + \left(\frac{\beta_{(2)}}{\hat{\beta}_{(2)}} - 1\right)^2 \tau_2 \xrightarrow{d} \chi^2_{(2)}$. Los límites (superior e inferior) que definen la zona de aceptación de la hipótesis nula se obtienen de la resolución de la ecuación:

$$\left(\frac{\beta_{(1)}}{\hat{\beta}_{(1)}} - 1\right)^2 \tau_1 + \left(\frac{\beta_{(2)}}{\hat{\beta}_{(2)}} - 1\right)^2 \tau_2 - \chi^2_{(2)}(\alpha) = 0$$

Alternativamente, Loretan, M.; Phillips, P.C.B. (1994) considerando $\hat{\tau}_\beta = \hat{\beta}_{(1)} - \hat{\beta}_{(2)}$ propusieron el estadístico $V_\beta = \frac{\hat{\tau}_\beta}{\frac{\hat{\beta}^2_{(1)}}{\tau_1} + \frac{\hat{\beta}^2_{(2)}}{\tau_2}} \xrightarrow{d} \chi^2_{(1)}$. A partir del resultado obtenido para el estimador $\hat{\beta}$, queda definido el siguiente estimador del coeficiente σ (ver Hill, B.M. (1975), Hall, P. (1982)): $\hat{\sigma}^* = \frac{\tau}{n} v^{\hat{\beta}}_{(n-\tau)}$, a partir del cual se demuestra que el estadístico[227]:

---

[227]. Ver Hall, P. (1982).

$$\frac{\sqrt{\tau}\left(\hat{\sigma}_{\tau}^{*}-\sigma^{*}\right)}{\log\left(\frac{n}{\tau}\right)} \xrightarrow{d} N\left(0,\sigma^{*2}\right)$$

Igual que en el caso del coeficiente β, un problema importante en las aplicaciones es el referido a la constancia de σ^{*}. Es decir, dados dos índices $\sigma_{(1)}^{*}$ y $\sigma_{(2)}^{*}$, decidir si la hipótesis $H_0:\sigma_{(1)}^{*} = \sigma_{(2)}^{*} = \sigma^{*}$ es verdadera, a partir de la información que proporcionan los estimadores obtenidos de dos muestras independientes. Loretan, M.; Phillips, P.C.B. (1994) propusieron el estadístico:

$$V_{\sigma} = \frac{\hat{\eta}_{\sigma^{*}}^{2}}{\frac{\hat{\beta}_{(1)}^{2}}{\tau_1}+\frac{\hat{\beta}_{(2)}^{2}}{\tau_2}} \xrightarrow{d} \chi_{(1)}^{2}$$

(donde $\hat{\eta}_{\sigma^{*}} = \hat{\sigma}_{(1)}^{*} - \hat{\sigma}_{(2)}^{*}$).

4.7.8.- Las distribuciones Normales compuestas

4.7.8.1.-El modelo de Clark

El hecho que, de acuerdo a lo expresado en la sección precedente: **i)** las distribuciones estables que representan el comportamiento "fat tailed" de las rentabilidades ($\alpha < 2$) no posean momentos de orden superior a α, lo cual implica que los estimadores de la varianza, la asimetría y la kurtosis no converjen a los verdaderos valores de los parámetros cuando el número de observaciones aumenta indefinidamente y **ii)** la propiedad de invariancia con respecto a la adición de variables aleatorias implique que las rentabilidades semanales, mensuales, etc. también sean estables, cuando en las aplicaciones las rentabilidades obtenidas de la agregación de rentabilidades diarias observan un comportamiento no-Normal mucho más débil que el de las variables sumando, hizo que surgieran propuestas alternativas a la utilización de distribuciones estables, que mantuvieron las hipótesis de Normalidad e independencia de las variables que representan a las rentabilidades.

Considerando, además, que la no-estabilidad de las volatilidades se asoció a la hipótesis de que la información, en realidad, no se genera sólo en forma regular y continua, sino que aparece también en forma de "paquetes" en puntos discontinuos del tiempo generando aumentos en el nivel de actividad[228] y, consecuentemente, en las varianzas condicionadas de las rentabilidades[229], estas soluciones alternativas se basaron, en general en modelos construidos a partir de distribuciones Normales compuestas en las que la varianza es, a su vez, una variable aleatoria[230]. De modo que la distribución de las rentabilidades queda definida como una combinación de funciones normales, algunas con varianzas condicionadas pequeñas -que concentran la densidad de la probabilidad alrededor del valor central- y otras con varianzas

---

228. Esta relación positiva entre volatilidad y volumen operado dio origen a dos clases de modelos para observaciones de alta frecuencia: una en la que el dominio del tiempo está determinado por el volumen operado (ver Giot, P. (2001), Ghysels, E.; Jasiak, J. (1995)(1995), Zhou, B. (1996), Hysels, E.; Gourieroux, C.; Jasiak, J. (1998)) y otra en la que el dominio del tiempo está definido por el intervalo entre transacciones o entre arribos de información al mercado o entre variaciones en el precio que superen un cierto umbral (**modelos autorregresivos de duración condicionada (ADC)**) propuestos por Engle, R.F.; Russell, J.R. (1997)(1998)).

229. Debe recordarse en cuenta que, como ya se mencionó en la Sec. 4.3, dado un proceso no-lineal estacionario $\{Y_t\}$, su varianza no-condicionada $\sigma^{2}(Y_t) = \sigma_Y^{2}$ es constante, pero su varianza condicionada, $\sigma^{2}(Y_t / Y_{t-1}, Y_{t-2}, \ldots)$, es una función de sus realizaciones pasadas y puede variar.

230. A estas distribuciones se las conoce también como **contaminadas** (ver Tukey, W. (1957)).

condicionadas grandes -que determinan la densidad de probabilidad en las colas. El resultado es una distribución que observa un comportamiento "fat tailed" con varianza y momentos de orden superior finitos lo cual, de acuerdo con los postulados del teorema central del límite, permite concluir que la representación de las rentabilidades obtenidas de la agregación de las rentabilidades diarias se aproximaría más a la Normal que las rentabilidades sumando.

Sea $\{Y(t)\}$ el proceso que representa el comportamiento del precio de un activo y sea $T(\tau)$ la variable que representa el incremento en la cantidad de información disponible para un observador[231], ocurrida en un intervalo de tiempo τ. Supóngase que $\{T(\tau)\}$ $(\tau \geq 0)$ sea un proceso formado por variables aleatorias independientes que pueden asumir solamente valores reales no-negativos, con momentos de segundo orden finitos, y tal que $E[T(\tau)] = T_m$.

De acuerdo con el modelo de Clark, P.K. (1973), el proceso representativo del comportamiento de la rentabilidad de un activo, para dicho observador, puede ser expresado de la siguiente forma:

$$v(t,T(\tau)) = Y^*(t) - Y^*(t-\tau) = \Delta_\tau Y^*(t) = \mu T(\tau) + \sigma B(T(\tau))$$

Esto implica suponer que las variaciones de precios que el observador considera que se habrán de producir en el activo en un intervalo de longitud τ, están afectadas por las variaciones en la cantidad de su información en dicho intervalo. Se dice, entonces, que el proceso estocástico $\{v(t,\tau)\}$ está **subordinado** al proceso $\{B(t)\}$, siendo $\{T(\tau)\}$ el **proceso directriz.** Teniendo en cuenta que el proceso definido por los incrementos del movimiento Browniano $\{B(T(\tau))\}$ es Normal, se puede concluir que (dado que cada rentabilidad está distribuida en forma diferente, el proceso $\{v(t,\tau)\}$ constituye una mezcla de distribuciones) $v(t,\tau): N[mT(\tau); \sigma^2 T(\tau)]$ y se obtiene que:

$$E[v(t,\tau)] = 0$$

$$\sigma^2\{v(t,\tau)/[T(\tau) = T']\} = \sigma^2 T'$$

$$\sigma^2[v(t,\tau)] = E[\sigma^2 T(\tau)] = \sigma^2 T_m$$

$$E\{[v(t,\tau)]^4 / [T(\tau) = T']\} = 3T'\sigma^4$$

$$\mu_4[v(t,\tau)] = \{[v(t,\tau)]^4\} = 3\sigma^4 E[3T(\tau)\sigma^4] = 3\sigma^4 E\{[T(\tau)]^2\} = 3\sigma^4[T_m + \sigma^2(T(\tau))]$$

De modo que la kurtosis de $v(t,\tau)$ es una función que varía con la varianza de $T(\tau)$:

$$K[v(t,\tau)] = \frac{\mu_4[v(t,\tau)]}{\sigma^4[v(t,\tau)]} = \frac{3\sigma^4[T_m^2 + \sigma^2(T(\tau))]}{T_m^2\sigma^4} = 3\frac{[E(T(\tau))]^2 + \sigma^2(T(\tau))}{[E(T(\tau))]^2}$$

lo cual demuestra que la introducción de un proceso directriz, no importa cómo sea éste, aumenta la kurtosis de las variables $v(t,\tau) = \Delta_\tau Y^*(\tau)$ y, por lo tanto, mejora su representación.

---

231. Operativamente, esta variable es asimilada al volumen operado.

De acuerdo a lo expresado precedentemente, es posible construir un test basado en el siguiente principio: Si la rentabilidad de un activo no estuviera relacionada con su volumen operado, entonces las variables $T(\tau)$ y $\left[v(t,T(\tau))\right]^2$, obviamente, deberían ser incorrelacionadas. Si, por el contrario, $\{T(\tau)\}$ constituyera un proceso directriz, la relación entre dichas variables debería ser lineal, con $\sigma^2(v(t,\tau))$ como coeficiente de regresión. Un análisis de la relación entre el volumen operado de un activo y la correspondiente varianza de sus rentabilidades sugiere la existencia de vínculos funcionales del tipo:

$$\sigma^2[v(t,\tau)] = a_0 + a_1 T$$

$$\ln[\sigma^2(v(t,\tau))] = f(T)$$

(debe tenerse en cuenta que los resultados obtenidos de la observación muestran que la relación entre estas variables admite, habitualmente, una representación no-lineal doble logarítmica, $\ln[\sigma^2(v(t,\tau))] = f(\ln(T))$).

En particular, Clark, P.K (1973) propuso un modelo que supone que las variables $v(t,\tau)$ se distribuyen de acuerdo con una función Normal y que $T(\tau)$ tiene una distribución log-Normal[232] tal que a la variable $\ln[T(\tau)]$ le corresponde una distribución del tipo $N(m_{LT},\sigma^2_{LT})$:

$$f_{LN}(T(\tau)) = \frac{1}{\sigma_{LT}\sqrt{2\pi T(\tau)}} \exp\left[-\frac{\ln(T(\tau)) - m_{LT}}{2\sigma^2_{LT}}\right]$$

La función de densidad de $v(t,\tau)$, condicionada a la realización T' del proceso $\{T(\tau)\}$, es una función de la forma $N(0,\sigma^2 T')$:

$$f_N[v(t,\tau)/T'] = \frac{1}{\sigma\sqrt{2\pi T'}} \exp\left[-\frac{u^2}{2\sigma^2 T'}\right] \qquad (-\infty \le u \le \infty)$$

Se demuestra que[233]:

$$m_T = m(T(\tau)) = \exp\left(m_{LT} + \frac{\sigma^2_{LT}}{2}\right)$$

$$\sigma^2_T = \sigma^2(T(\tau)) = e^{2m_{LT}+\sigma^2_{LT}}\left(e^{\sigma^2_{LT}} - 1\right)$$

Luego, teniendo en cuenta que las variables $v^*(t,\tau)$: $N(0,\sigma^2_{v^*})$, entonces las variables $v'(t,\tau) = v^*(T(\tau))$ tendrán distribuciones de probabilidades log-Normales compuestas del tipo

---

232. Ver Sec. 4.7.6.1.

233. Ver Aitchison, J.; Brown, J.A.O. (1957).

$N\left[0,T(\tau)\sigma^2\right]_{\underset{T(\tau)}{\wedge}}$ log- $N\left(m_{LT},\sigma_{LT}^2\right)$, con valor esperado nulo y varianza:

$$\sigma^2\left[v\left(t,T(\tau)\right)\right] = \sigma_v^2 \exp\left(m_{LT} + \frac{\sigma_{LT}^2}{2}\right)$$

Es decir, será:

$$f\left[v^*\left(t,T(\tau)\right)\right] = \int_0^\infty f_N\left[v(t,\tau)/T(\tau)\right] f_{LT}\left[T(\tau)\right] d\tau =$$

$$= \frac{1}{2\pi\sigma_{LN}^2\sigma_{v^*}^2}\int_0^\infty \exp\left\{-\frac{1}{2T(\tau)}\left[\frac{v^*(\tau)}{\sigma_{v^*}}\right]\right\}\left[T(\tau)\right]^{3/2}\exp\left[-\frac{1}{2}\frac{\ln\left[T(\tau)-m_{LT}\right]}{\sigma_{LT}^2}\right] dT(\tau)$$

(debe tenerse en cuenta que no es posible obtener una forma explícita de esta función, lo cual hace dificultosa su aplicación).

4.7.8.2.-El modelo de Praetz-Blattberg-Gonedes

Supóngase, ahora, de acuerdo con el modelo propuesto por Praetz, P. (1972), Blattberg, R.C.; Gonedes, W.J. (1974), que la variable $T^*(\tau) = \left[T(\tau)\right]^{-1} = \left[\sigma^2\left(v(\tau)\right)\right]^{-1}$ se distribuye de acuerdo a una función de densidad Gama-2 ($f_\Gamma\left[T(\tau)\right]$) con $\alpha = c^2\tau$, r grados de libertad y momentos $E\left[T^*(\tau)\right] =$ $= \frac{1}{c^2\tau}$ y $\sigma^2\left[T^*(\tau)\right] = \frac{2}{r\left(c^2\tau\right)^2}$, entonces, las variables $v(\tau) = v^*\left(T(\tau)\right)$ se distribuirán de acuerdo con funciones compuestas de la forma $N\left[0,T(\tau)\sigma^2\right]_{\underset{T^*(\tau)}{\wedge}} \Gamma\left(c^2\tau,r\right)$, es decir, se distribuirán de acuerdo con una función t de Student con parámetro $\left[\sigma^2c^2\tau\right]^{-1}$ y r grados de libertad de la forma (ver Landro, A.H. (2002))[234]:

$$f\left[v(\tau)\right] = \int_0^\infty f_N\left[v(t,\tau)/T(\tau)\right] f_T\left[T(\tau)\right] dT(\tau) = \frac{r^{r/2}}{\sigma c\sqrt{\tau}\, B\left(\frac{1}{2},\frac{r}{2}\right)}\left\{r + \frac{\left[v(t,\tau)-m_v\right]^2}{\sigma^2 c\tau}\right\}^{-(r+1)/2}$$

Donde $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ denota, en este caso, una distribución Beta. Sus momentos serán $E\left[v(\tau)\right] = m_v$ $\left(para\ r > 1\right)$ y $\sigma^2\left[v(\tau)\right] = \left(\sigma^2c^2\tau\right)\frac{r}{r-2}\ (r > 2)$. En general, se verifica que todos los momentos de orden menor que r son finitos. Para $r = 1$, la distribución t se transforma en la función de Cauchy, y cuando $r \to \infty$ la función t converge en-distribución a la función Normal (si $\sigma^2\left(v'(\tau)\right)$ define una variable estrictamente positiva estable, con $\alpha < 1$, entonces, $v(\tau)$ se distribuye de acuerdo con una función de

---

234. Ver Raiffa, H.; Schlaifer, R. (1961).

probabilidades estrictamente positiva estable con $\alpha < 2$) [235].

4.7.9.- Los modelos para el estudio de las rentabilidades que utilizan procesos con salto

4.7.9.1.- El proceso de difusión con salto

Como una solución alternativa al caso tratado en la sección precedente que supone, además de una información que ingresa en forma continua (en puntos aleatorios del tiempo) la incorporación de paquetes de información -que también son asumidos en forma instantánea por los operadores, se suelen utilizar procesos que admiten una probabilidad no-nula de que se produzcan variaciones de magnitud extraordinaria (**"jumps"** o **saltos**) en intervalos de tiempo arbitrariamente pequeños[236]. Es decir, procesos que consideran que las variaciones en el precio del activo están formadas por dos componentes: una que evoluciona **por fluctuaciones**[237], con distribución Normal, representada por un proceso **movimiento Browniano geométrico**, y otra, debida al arribo de un conjunto importante de nueva información, que produce sobre los precios impactos de magnitud cierta, $(Y-1)$, y está representada por un **proceso de Poisson** ($\wp(t)$).

Supóngase, por ejemplo, un proceso representativo del comportamiento de la rentabilidad de un activo de la forma [238]:

$$\frac{dY(t)}{Y(t)} = \varphi_0\, dt + \sigma_V\, dB(t) + (Z-1)d\wp(\beta t)$$

donde: **i)** la ocurrencia del "salto" define una variable aleatoria (que puede asumir valores 1 ó 0) con distribución $\wp(\beta t)$; **ii)** β denota la tasa instantánea de la ocurrencia del "salto" (ver Sec. 1.6.2.2):

$$\beta = \lim_{\Delta t \to 0} \frac{1 - p\left\{\dfrac{Y(t+\Delta t) - Y(t)}{Y(t)} = \varphi_0 \Delta t + \sigma_v\left[B(t+\Delta t) - B(t)\right]\right\}}{\Delta t}$$

y **iii)** $dB(t) \perp d\wp(\beta t)$.

Es decir que, con probabilidad $\beta\Delta t$, la rentabilidad del activo en un intervalo $[t, t+\Delta t)$ será $\dfrac{Y(t+\Delta t)-Y(t)}{Y(t)} = \varphi_0\Delta t + \sigma_v \Delta B(t) + (Z-1)$ y, con probabilidad $1-\beta\Delta t$, dicha rentabilidad será $\dfrac{Y(t+\Delta t)-Y(t)}{Y(t)} = \varphi_0\Delta t + \sigma_v \Delta B(t)$ (debe tenerse en cuenta que el proceso $\{\wp(t)\}$ supone solamente la posibilidad de que se produzca un número finito de "saltos" para todo intervalo de tiempo finito).

Supóngase ahora, que el proceso que describe la evolución de la rentabilidad admite una

---

235. Ver Sec. 4.7.5.

236. Formalmente, este proceso "con salto" sigue trayectorias que presentan discontinuidades con probabilidad igual a uno (ver Sec. 1.7.2), en tanto, que (como se vio en la Sec. 4.5) el proceso de difusión sigue trayectorias continuas con probabilidad igual a uno.

237. Merton, R.C. (1976) las denomina **"vibraciones normales"** en los precios.

238. Debe tenerse en cuenta que la asimetría y la kurtosis se ven afectadas por la presencia de "saltos" significativos.

representación de la forma:

$$\frac{dY(t)}{Y(t)} = \varphi_0\, dt + \sigma_v\, dB(t) + d\left[\sum_{j=1}^{N}\left(Z_j - 1\right)\right]$$

donde: **i)** Z_j denota una variable aleatoria que representa la medida del j-ésimo "salto" expresada como proporción de $Y(t)$; **ii)** $\{B(t)\} \perp \{Z_j\}$ y **iii)** las variables $Z_j^* = \ln(Z_j)$ son idénticamente distribuidas con autocorrelaciones significativas y con distribución de probabilidades conjunta de la forma[239]:

$$f\left[\ln(Z_1), \ln(Z_2), ..., \ln(Z_N)\right] = f\left(Z_1^*, Z_2^*, ..., Z_N^*\right) = \int_{-\infty}^{\infty} f\left(Z_1^*, Z_2^*, ..., Z_N^*, W\right) dW$$

Para $N \geq 1$ la función de probabilidades condicionada será de la forma:

$$p\left[v(t+\Delta t) / N\right] = \int_{-\infty-\infty}^{\infty\ \infty} \int \ldots \int_{-\infty}^{\infty} p\left(Z_1^*, Z_2^*, ..., Z_N^*, W\right) dZ_1^* dZ_2^* \ldots dZ_N^* dW =$$

$$= \int_{-\infty-\infty}^{\infty\ \infty} \int \ldots \int_{-\infty}^{\infty} p\left[v(t+\Delta t) - \left(\varphi_0^* - \frac{\sigma^2}{2}\right)\Delta t - \sigma W\sqrt{\Delta t} - \sum_{i=2}^{n} Z_i^*, Z_2^*, ..., Z_n^*, W \right] dZ_1^* dZ_2^* \ldots dZ_n^* dW$$

(donde $-\infty < v(t+\Delta t) < \infty$). A partir de esta función, se demuestra, además, que las medidas de las variables condicionadas son [240]:

$$E\left\{\left[v(t+\Delta t)\right] / N\right\} = \left(\varphi_0^* - \frac{\sigma_v^2}{2}\right)\Delta t + \sigma_v\sqrt{\Delta t}\, E\left[W(t)\right] + \sum_{j=1}^{N} E\left[\ln(Z_j)\right] = \left(\varphi_0^* - \frac{\sigma_v^2}{2}\right)\Delta t + NE^*$$

(donde $E^* = E\left[\ln(Z_j)\right] = E(Z_j^*)$) y,

---

239. Debe tenerse en cuenta que dZ denota el incremento de un movimiento Browniano con valor medio nulo y varianza unitaria, y $d\wp$ denota un proceso de "salto" (ver Sec. 1.6.2.4). En términos generales, el proceso $\{Y(t)\}$ puede ser definido de la siguiente forma agregativa: $Y(t) = Y(0) + I(t) + R(t) + J(t)$, donde: **i)** $Y(0)$ denota una condición inicial no-aleatoria; **ii)** $I(t) = \int_0^t \Gamma(\tau) dB(\tau)$ (siendo $\Gamma(t)$ un proceso adaptado respecto de un movimiento Browniano $\{B(t)\}$) denota la parte de $\{Y(t)\}$ debida a una integral de Itô; **iii)** $R(t) = \int_0^t \Theta(\tau) d\tau$ denota la parte de $\{Y(t)\}$ debida a una integral de Riemann y **iv)** $J(t)$ denota un proceso de "salto". De modo que la parte continua de $\{Y(t)\}$ puede ser definida por la siguiente fórmula de Itô-Doeblin: $Y^{(c)}(t) = Y^{(c)}(0) + \int_0^t \Gamma(\tau) dB(\tau) + \int_0^t \Theta(\tau) d\tau$ (ver Shreve, S.E. (2004)).

240. Nótese que la presencia de autocorrelación no afecta al valor esperado condicionado.

$$\sigma^2\left(v(t+\Delta t)/N\right) = \Delta t\,\sigma^2 + \sigma^2\left\{\left[\sum_{j=1}^{N}\ln\left(Z_j\right)\right]/N\right\} =$$

$$= \Delta t\,\sigma^2 + N\sigma^{*2} + 2\sum_{i=1}^{N-1}(N-i)\gamma\left[\ln\left(Z_{j-i}\right),\ln\left(Z_j\right)\right] =$$

$$= \Delta t\,\sigma^2 + N\sigma^{*2} + 2\sigma^{*2}\sum_{i=1}^{N-1}(N-i)\rho_i^*$$

(donde: **i)** $\sigma^{*2} = \sigma^2\left[\ln\left(Z_j\right)\right] = \sigma^2\left(Z_j^*\right)$ y **ii)** $\rho_i^* = \rho\left(Z_{j-i},Z_j\right)\left(j = 1,2,\ldots,N;i > 0\right)$). El primer sumando del segundo miembro de esta expresión ($\tau\sigma^2$) constituye la contribución del proceso de difusión a la variación total del proceso que representa la evolución de la rentabilidad del activo en caso de que ocurra un número N de saltos, y los sumandos segundo y tercero ($N\sigma^{*2} + 2\sigma^{*2}\sum_{i=1}^{N-1}(N-i)\rho_i^*$) representan la contribución del proceso de salto[241].

Por otra parte, la función de probabilidades:

$$p\left[v(t+\Delta t)\right] = \sum_{N=0}^{\infty} p\left[v(t+\Delta t)/N\right]p(N)$$

(donde $p(N)$ denota la probabilidad de que ocurran N saltos en un período $[t,t+\tau)$), permite analizar la variable rentabilidad sobre un intervalo de longitud fija (τ) en el que el número de transacciones (N) es variable. Luego, será:

$$E\left[v(t+\Delta t)\right] = \left(\varphi_0^* - \frac{\sigma^2}{2}\right)\Delta t + E^*\,E(N)$$

$$\sigma^2\left[v(t+\Delta t)\right] = \sum_{N=0}^{\infty}\sigma^2\left[v(t+\Delta t)/N\right]p(N) + E^{*2}\,\sigma^2(N)$$

Se concluye fácilmente que la componente del valor esperado y de la varianza debida al proceso de difusión, depende de la longitud del intervalo (τ), y la componente debida al proceso de salto depende del número de transacciones (N) (nótese que las definiciones de valor esperado y varianza requieren la definición previa de la distribución de probabilidades $p(N)$).

4.7.9.2.- El modelo de Kou

Supóngase ahora que el proceso que describe la rentabilidad del activo admita una representación de la forma:

---

241. Nótese que la varianza condicionada es una función del número de saltos en el intervalo de longitud Γ, pero no de Γ, y no requiere la definición de la distribución de probabilidades de la variable $Y(t)$.

$$\frac{dY(t)}{Y(t)} = \varphi_0\,dt + \sigma_v\,dB(t) + d\left[\sum_{j=1}^{N(t)}\left(Z_j - 1\right)\right]$$

donde: **i)** $\left(Z_j - 1\right)$ representa la medida del j-ésimo "salto" expresada como una proporción de $Y(t)$; **ii)** $\{N(t)\}$ es un proceso de Poisson con parámetro βt, que representa el número de "saltos" a ocurrir al momento t y **iii)** β denota la tasa instantánea de ocurrencia de un "salto":

$$\beta = \lim_{\Delta t \to 0} \frac{1 - p\left[N(t+\Delta t) - N(t) = 0\right]}{\Delta t}$$

(debe tenerse en cuenta que, como se mencionó en la Sec. 1.3, un proceso de Poisson es un proceso de Markov homogéneo, continuo hacia la derecha, con límite por la izquierda, es decir un proceso CADLAG).

Kou, S.G. (2002) propuso para este caso un modelo -particularmente apropiado para representar variables leptokúrticas con asimetría con respecto a cero, que definen un mercado incompleto en el cual los métodos de "hedging" habituales no son aplicables-, donde: **i)** $\{Z_j\}\,(j = 1,2,\ldots,N(t))$ denota una sucesión de variables aleatorias iid e independientes de $B(t)$ y $N(t)$, que pueden asumir solamente valores no-negativos y tales que las variables $Z^* = \ln(Z)$ se distribuyen de acuerdo con una función de Laplace de la forma $f_{Z^*}(z^*) = \frac{1}{2\varphi} e^{-|z^* - \lambda|/\varphi}\ (0 < \varphi < 1; \lambda \le z^* \le \infty)$ (ver Sec. 4.7.5) y **ii)** $\{B(t)\} \perp \{N(t)\}$.

A partir de estos supuestos se obtiene la siguiente solución a la ecuación diferencial anterior:

$$Y(t) = Y(0)\exp\left\{\left[\varphi_0^* - (0.5)\sigma_v^2\right]t + \sigma_v B(t)\right\}\prod_{j=1}^{N(t)} Z_j$$

La versión de esta ecuación para un intervalo Δt arbitrariamente pequeño asume la forma:

$$\frac{Y(t+\Delta t) - Y(t)}{Y(t)} = v(t) = \exp\left\{\left(\varphi_0^* - \frac{\sigma_v^2}{2}\right)\Delta t + \sigma_v\left[B(t+\Delta t) - B(t)\right] + \sum_{j=N(t)+1}^{N(t+\Delta t)} Z_j^*\right\} - 1 \qquad \left(Z_j^* = \ln\left(Z_j\right)\right)$$

y permite utilizar las aproximaciones $e^{f(t)} \approx 1 + f(x) + \frac{1}{2}f^2(x)$ y $\left[\Delta B(t)\right]^2 \approx \Delta t$. De modo que se puede escribir:

$$\frac{Y(t+\Delta t) - Y(t)}{Y(t)} = v(t) \approx \left(\varphi_0^* - \frac{\sigma_v^2}{2}\right)\Delta t + \sigma_v \Delta B(t) + \sum_{j=N(t)+1}^{N(t+\Delta t)} Z_j^* + \frac{\sigma_v^2}{2}\left[\Delta B(t)\right]^2 =$$

$$\approx \varphi_0^* \Delta t + \sigma_v\,\Delta B(t)\sqrt{\Delta t} + \sum_{j=N(t)+1}^{N(t+\Delta t)} Z_j^*$$

(donde, como se vio en la Sec. 4.5, $\Delta B(t) = B(t+\Delta t) - B(t)$: $N(0;1)$). Teniendo en cuenta que $\{N(t)\}$:$\wp(\beta t)$ y que, por lo tanto, la probabilidad de que se produzca más de un "salto" en un intervalo de

longitud Δt (arbitrariamente pequeño) tiende a cero, se verifica que $p\left(\sum_{j=N(t)+1}^{N(t+\Delta t)} Z_j^* = Z_{n(t)+1}\right) = \beta\Delta t$ y que $p\left(\sum_{j=N(t)+1}^{N(t+\Delta t)} Z_j^* = 0\right) = 1-\beta\Delta t$. De modo que la ecuación anterior, simbólicamente, puede ser aproximada de la siguiente forma:

$$\frac{Y(t+\Delta t)-Y(t)}{Y(t)} = v(t) \approx \varphi_0^*\Delta t + \sigma_v\Delta t\sqrt{\Delta t} + b(1,\beta)Z^*$$

(donde: **i)** $b(1,\beta)$ denota una variable binomial tal que $p(b=1)=\beta\Delta t$ y $p(b=0)=1-\beta\Delta t$ y **ii)** Z^* es la variable con distribución de probabilidades doble-exponencial definida en la página precedente). Denotando por W a la variable aleatoria definida por el segundo miembro de esta expresión y teniendo en cuenta que $\Delta B(t)\perp Z^*$, se obtiene la siguiente definición de su función de densidad:

$$f_W(w) = \frac{\beta\Delta t}{2\varphi}e^{\sigma_v^2\Delta t/(2\varphi^2)}\left[e^{-w^*/\varphi}F_{(N)}\left(\frac{w^*-\sigma_v^2\Delta t}{\sigma_v\varphi\sqrt{\Delta t}}\right) + e^{w^*/\varphi}F_{(N)}\left(\frac{w^*\varphi+\sigma_v^2\Delta t}{\sigma_v\varphi\sqrt{\Delta t}}\right)\right] +$$

$$+ (1-\beta t)\frac{1}{\sigma_v\sqrt{\Delta t}}f_{(N)}\left(\frac{w-\varphi_0^*\Delta t}{\sigma_v\sqrt{\Delta t}}\right)$$

(donde: **i)** $w^* = w-\varphi_0\Delta t-\lambda$; **ii)** $F_{(N)}(\bullet)$ denota la función de distribución de una variable $N(0;1)$ y **iii)** $f_{(N)}(\bullet)$ denota la función de densidad de una variable $N(0;1)$). A partir de esta distribución de probabilidades, se obtiene que $E(W) = \varphi\Delta t + \lambda\beta\Delta t$ y $\sigma^2(W) = \sigma_v^2 + \Delta t + \beta\Delta t\left[2\varphi_0^2 + \lambda^2(1-\beta\Delta t)\right]$.

4.7.9.3.- El modelo de Press

Press, S.J. (1967) propuso un modelo similar al precedente, en el que los "saltos" $Z_j^* = \ln(Z_j)$ constituyen una sucesión de variables aleatorias iid con distribución Normal del tipo $N(m_{Z^*},\sigma_{Z^*}^2)$. La distribución de probabilidades de la variable $v(t)$ será, entonces, de la forma:

$$f\left(\frac{dY(t)}{Y(t)}\right) = f(v(t)) = \sum_{j=0}^{\infty}\frac{\beta^j e^{-\beta}}{j!}\frac{\exp\left\{-\frac{1}{2}\frac{[v(t)-j\varphi_0]^2}{j\sigma_{Z^*}^2+\sigma_v^2}\right\}}{\sqrt{2\pi(9\sigma_v^2)+j\sigma_{Z^*}^2}}$$

La asimetría de esta función está dada por $As(v(t)) = \frac{\beta\varphi_0(\varphi_0^2+3\sigma_v^2)}{\left[\sigma_{Z^*}^2+\beta(\varphi_0^2+\sigma_v^2)\right]^{3/2}}$. Obsérvese: **i)** que el signo de este coeficiente es el mismo que el de β ; **ii)** que la condición necesaria y suficiente para que $As(v(t)) = 0$

es que $\beta = 0$ y **iii)** que, a su vez, la condición necesaria y suficiente para que $\beta = 0$ es que $E\left(v(t)\right) = 0$ (es decir que esta distribución es simétrica si, y sólo si, el valor esperado de la variable $v(t)$ es nulo).

El logaritmo de la función característica de la variable $v(t)$ es de la forma:

$$\ln\left[C_{v(t)}(w)\right] = -\frac{w^2\sigma_v^2}{2} + \beta\left[\exp\left(iw\varphi_0 - \frac{w^2\sigma_v^2}{2}\right) - 1\right]$$

A partir de esta expresión se puede concluir que el grado de apartamiento de la distribución de la variable $v(t)$ de la función Normal está dado por la magnitud de los coeficientes de los términos de grado superior a 2 en el desarrollo en serie de potencias de w (si β es aproximadamente igual a 0, la distribución es aproximadamente Normal; para valores grandes de β, los términos de orden superior en w, producen un apartamiento de la Normal).

Los momentos de la variable $v(t)$ son de la forma $E\left[v(t)\right] = \varphi_0\beta$ y $\sigma^2\left[v(t)\right] = \sigma_{Z^*}^2 + \beta\left(\varphi_0^2 + \sigma_v^2\right)$

4.7.9.4.- El modelo de Oldfield-Rogalski-Jarrow

En el contexto de los procesos de difusión con "salto" como los analizados en las secciones precedentes, Oldfield, G.S.; Rogalski, R.J.; Jarrow, R.A. (1977) propusieron un modelo que supone que los intervalos de tiempo entre el i -ésimo y el $(i+1)$ -ésimo ($i = 0,1,2,...$) "salto" ($T_{i,i+1}$) son variables aleatorias independientes, todas con distribución de probabilidades Gama de la forma:

$$f_{T_{i,i+1}}(t) = \frac{\alpha(\alpha t)^{r-1}}{\Gamma(r)}e^{\alpha t} \qquad (t \geq 0; r > 0; \alpha > 0)$$

con medidas $E(T) = \frac{r}{\alpha}$ y $\sigma^2(T) = \frac{r}{\alpha^2}$ (donde α denota el factor escala y r denota un parámetro que caracteriza la forma de la distribución). Se demuestra, entonces, que la probabilidad de que ocurran N saltos en el período $[t, t+\tau)$ queda definida por:

$$p(N) = p\left(T_{0,1} + T_{1,2} + ... + T_{N-1,N} \leq \tau\right) - p\left(T_{0,1} + T_{1,2} + ... + T_{N,N+1} \leq \tau\right) =$$

$$= p\left(T^{(N)} \leq \tau\right) - p\left(T^{(N+1)} \leq \tau\right) =$$

$$= \frac{\alpha}{\Gamma(rN)}\int_0^{\tau}\left(\alpha T^{(N)}\right)^{rN-1}e^{-\alpha T^{(N)}}dT^{(N)} - \frac{\alpha}{\Gamma(rN+r)}\int_0^{\tau}\left(\alpha T^{(N+1)}\right)^{rN+r-1}e^{-\alpha T^{(N+1)}}dT^{(N+1)} =$$

$$= \frac{\Gamma^*(rN, \alpha\tau)}{\Gamma(rN)} - \frac{\Gamma^*(rN+r, \alpha\tau)}{\Gamma(rN+r)} \qquad (N \geq 1)$$

(donde $\Gamma^*(x,y) = \int_0^y e^{-u}u^{x-1}du$ denota una función **Gama incompleta**). Sustituyendo esta función de

probabilidades en la distribución de la variable $v(t,\Delta t) = \dfrac{Y(t+\Delta t) - Y(t)}{Y(t)}$, se obtiene como caso particular:

$$p\left[v(t+\Delta t)\right] = \left[1 - \frac{\Gamma^*(r,\alpha\Delta t)}{\Gamma(r)}\right] p\left[v(t,\Delta t)/(N=0)\right] + \sum_{N=1}^{\infty}\left[\frac{\Gamma^*(rN,\alpha\Delta t)}{\Gamma(rN)} - \frac{\Gamma^*(rN+r,\alpha\Delta t)}{\Gamma(rN+r)}\right] p\left[v(t,\Delta t)/N\right]$$

De modo que:

$$E\left[v(t,\Delta t)\right] = \left(\varphi_0 - \frac{\sigma_v^2}{2}\right)\Delta t + E^* \sum_{N=1}^{\infty} N\left[\frac{\Gamma^*(rN,\alpha\Delta t)}{\Gamma(rN)} - \frac{\Gamma^*(rN+r,\alpha\Delta t)}{\Gamma(rN+r)}\right]$$

$$\sigma^2\left[v(t,\Delta t)\right] = \Delta t\,\sigma_v^2 + E(N)\sigma^{*2} + E^{*2}\sigma^2(N) + 2\sigma^{*2}\sum_{N=2}^{\infty}\sum_{i=1}^{N-1}(N-i)\rho_i^*\left[\frac{\Gamma^*(rN,\alpha\Delta t)}{\Gamma(rN)} - \frac{\Gamma^*(rN+r,\alpha\Delta t)}{\Gamma(rN+r)}\right]$$

donde: **i)** $E^* = E\left(Z_j^*\right) = E\left[\ln\left(Z_j\right)\right]$; **ii)** $\sigma^{*2} = \sigma^2\left(Z_j^*\right) = \sigma^2\left[\ln\left(Z_j\right)\right]$ y **iii)** $\rho_i^* = \rho\left(Z_{j-i}^*, Z_j^*\right)\left(j = 1,2,\ldots,N(t);\ i > 0\right)$.

4.7.9.5.- El modelo de Merton

Este modelo[242] puede ser considerado como un caso particular del modelo con incrementos autorregresivos anterior, en el que: **i)** el proceso de "salto" no presenta autocorrelaciones significativas; **ii)** la variable $T_{i,i+1}$ se distribuye de acuerdo con una función de probabilidades exponencial y **iii)** la variable $N(t)$ se distribuye de acuerdo con una función de probabilidades de Poisson de la forma $p\left[N(t) = n\right] = \dfrac{(\alpha t)^n}{n!} e^{-\alpha t}\ \left(n = 0,1,2,\ldots;t > 0\right)$. Luego, los valores característicos de las variables condicionadas $\left(v(t,\Delta t)/N\right)$, serán $E\left[v(t,\Delta t)/N\right] = \left(\varphi_0 - \dfrac{\sigma^2}{2}\right)\Delta t + NE^*$ y $\sigma^2\left[v(t,\Delta t)/N\right] = \Delta t\,\sigma^2 + N\sigma^{*2}$ (donde $E^* = E\left(Z_j^*\right) = E\left[\ln\left(Z_j\right)\right]$, $\sigma^{*2} = \sigma^2\left(Z_j^*\right) = \sigma^2\left[\ln\left(Z_j\right)\right]$ y $\rho_i^* = \rho\left(Z_{j-i}^*, Z_j^*\right)\ \left(j = 1,2,\ldots,N(t);i > 0\right)$). Nótese que el valor esperado condicionado es igual al correspondiente al modelo con-incrementos autorregresivos analizado en la Sec. 4.1. En lo que hace a la varianza condicionada, la diferencia con dicho modelo radica en la contribución del término $2\sigma^2\sum_{i=1}^{N-1}(N-i)\rho_i^*$ (donde $\rho_i^* = \rho\left(Z_{j-i}^*, Z_j^*\right)$ $\left(j = 1,2,\ldots,N(t);i > 0\right)$)correspondiente al proceso de "salto".

Supóngase que la variable Z_j que representa la medida del j-ésimo "salto" pueda asumir solamente dos valores: $Z^{(+)} > 1$ y $Z^{(-)} < 1$, con probabilidades $\pi^{(+)}$ y $\pi^{(-)}$, respectivamente. Dejando de lado (a fin de simplificar) a la componente tendencia, si se supone que el precio de un activo es el resultado de la agregación de unidades individuales estocásticamente independientes entre sí, y que la probabilidad de que el evento "incorporación al mercado de un paquete importante de información" afecte a una cualquiera de dichas unidades en un intervalo de longitud Δt, está dada por $\lambda\Delta t$ (donde λ representa el número medio de arribos de paquetes de información por unidad de tiempo), entonces, $\pi^{(+)}$ puede ser

---

242. Ver Merton, R.C. (1976).

interpretada como la probabilidad instantánea de que "nazcan" (se incorporen) $\left(Z^{(+)}-1\right)$ unidades en el sistema, y $\pi^{(-)}$ como la probabilidad instantánea de que "fallezcan" $\left(1-Z^{(-)}\right)$ unidades. De esta forma, el comportamiento del valor del activo puede ser representado por un **proceso de nacimiento-muerte lineal** para el cual la correspondiente ecuación "backward" de Kolmogorov es de la forma:

$$-\frac{\partial}{\partial\tau}p_{x,y}(\tau,t)=-\lambda x\,p_{x,y}(\tau,t)+\lambda x\pi^{(+)}p_{x+\Delta x,y}(\tau,t)+\lambda x\pi^{(-)}p_{x-\Delta x,y}(\tau,t)$$

(donde $\Delta x=Y^{(+)}-1$ y $-\Delta x=Y^{(-)}-1$).

Definiendo las intensidades de ocurrencia de la siguiente forma: $\lambda=\frac{\sigma^2}{(\Delta x)^2}$; $\lambda\pi^{(+)}=$ $=\frac{1}{2}\left[\left(\frac{\sigma}{\Delta x}\right)^2+\frac{\mu}{\Delta x}\right]$ y $\lambda\pi^{(-)}=\frac{1}{2}\left[\left(\frac{\sigma}{\Delta x}\right)^2-\frac{\mu}{\Delta x}\right]$ y haciendo $\Delta x\to 0$, se obtiene la siguiente ecuación "backward":

$$-\frac{\partial}{\partial\tau}p_{x,y}(\tau,t)=$$

$$=-\lambda x\,p_{x,y}(\tau,t)+\lambda x\pi^{(+)}\left[p_{x,y}(\tau,t)+\frac{\partial}{\partial x}p_{x,y}(\tau,t)\Delta x+\frac{1}{2}\frac{\partial^2}{\partial x^2}p_{x,y}(\tau,t)(\Delta x)^2\right]+$$

$$+\lambda x\pi^{(-)}\left[p_{x,y}(\tau,t)-\frac{\partial}{\partial x}p_{x,y}(\tau,t)\Delta x+\frac{1}{2}\frac{\partial^2}{\partial x^2}p_{x,y}(\tau,t)(\Delta x)^2\right]=$$

$$=\mu x\frac{\partial}{\partial x}p_{x,y}(\tau,t)+\frac{1}{2}\sigma^2x\frac{\partial^2}{\partial x^2}p_{x,y}(\tau,t)$$

correspondiente al **proceso de difusión** que queda definido cuando en el proceso de nacimiento-muerte se hace $Z^{(+)}\to 1$, $Z^{(-)}\to 1$ y $\lambda\to\infty$. Este proceso de difusión, se obtiene de suponer un mercado en el que las variaciones en los estados del proceso son de pequeña magnitud, y en el que la varianza de las variaciones de los precios del activo crecen con su precio, pero más lentamente que en el proceso de difusión, de modo que la varianza de la tasa de rentabilidad resulta decreciente:

$$dY(t)=\varphi_0Y(t)dt+\sigma_v\sqrt{Y(t)}\,dB(t)$$

El valor medio instantáneo de este proceso de difusión y su varianza son $\varphi_0Y(t)$ y $\sigma_v^2Y(t)$, respectivamente, donde φ_0 y σ_v^2 se obtienen de resolver las ecuaciones $E\left[dY(t)\right]=\left[\varphi_0+\lambda(Z-1)\right]Y(\tau)d\tau$ y $\sigma^2\left[dY(t)\right]=\lambda E\left[(Z-1)^2\right]Y(\tau)d\tau$.

Asimismo, si se supone que el valor del activo crece endógenamente a una tasa φ_0 y que los saltos exógenos de magnitud $(Z-1)$ ocurren con una intensidad constante λ, se obtiene que el proceso se comporta de acuerdo con una ecuación "backward" de la forma:

$$-\frac{\partial}{\partial\tau}p_{x,t}(\tau,t) = -\lambda\, p_{x,y}(\tau,t) + \lambda\pi^{(+)}p_{x+\Delta x,y}(\tau,t) + \lambda\pi^{(-)}p_{x-\Delta x,y}(\tau,t) + \mu x\frac{\partial}{\partial x}p_{x,y}(\tau,t)$$

Definiendo las intensidades de ocurrencia como $\lambda\pi^{(+)} = \lambda\pi^{(-)} = \frac{1}{2}\left(\frac{\sigma}{\Delta x}\right)^2$ se demuestra que, como en el caso precedente, el sistema de ecuaciones anterior converge al sistema[243]:

$$-\frac{\partial}{\partial\tau}p_{x,y}(\tau,t) = \mu x p_{x,y}(\tau,t) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}p_{x,y}(\tau,t)$$

correspondiente al proceso $dY(t) = \varphi_0 Y(t)dt + \sigma_v\, dB(t)$. De modo que:

$$E\left[dY(t)\right] = \left\{\varphi_0 Y(t) + \lambda\left[\pi^{(+)}\left(Z^{(+)} - 1\right) + \pi^{(-)}\left(Z^{(-)} - 1\right)\right]\right\}dt$$

$$\sigma^2\left[dY(t)\right] = \lambda\left[\pi^{(+)}\left(Z^{(+)} - 1\right)^2 + \pi^{(-)}\left(Z^{(-)} - 1\right)^2\right]dt$$

(debe tenerse en cuenta que $Y(t)$ representa el precio del activo y que, por lo tanto, a los procesos anteriores debe agregarse una **"barrera absorbente"** en $Y(t) = 0$).

4.7.10.- Los modelos para el estudio de las volatilidades

Otra razón por la cual el proceso de difusión resulta inadecuado para representar la evolución del precio de un activo radica en el comportamiento estocástico de sus volatilidades. De modo que una cuestión a resolver consiste en la definición del proceso estocástico que describe el comportamiento de $\sigma(t)$. Una primera propuesta consiste en suponer que la volatilidad admite una representación de la forma:

$$d\sigma(t) = f\left[\sigma(t),t\right]dt + g\left[\sigma(t),t\right]dB^*(t)$$

(donde $\left\{B^*(t)\right\}$ denota un movimiento Browniano estandarizado, correlacionado con $\left\{\sigma(t)\right\}$). La solución (única en condiciones bastante generales) de este sistema de ecuaciones diferenciales estocásticas no-lineales es un **proceso de difusión bidimensional**[244].

Un caso particularmente interesante se presenta cuando la rentabilidad admite una representación de la forma:

$$dY(t) = m\left[Y(t),t\right]dt + h\left[Y(t),t\right]dB(t)$$

(donde: **i)** $Y(t)$ representa el precio del activo al momento t y **ii)** $h(\bullet)$ denota una función estrictamente creciente (o decreciente)). De acuerdo con la fórmula de Itô-Doeblin[245], será:

---

243. Ver Sec. 1.7.3.

244. Ver Ikeda, N.; Watanabe, S. (1981), Karatzas, I.; Shreve, S. (1988).

245. Ver Sec. 4.6.

$$d\sigma(t) = h'\big(Y(t)\big)\,dY(t) + \frac{1}{2}h''\big(Y(t)\big)\,dY(t)$$

entonces, se puede escribir:

$$d\sigma(t) = \left\{h'\Big[h^*\big(\sigma(t)\big)\Big] + \frac{1}{2}h''\Big[h^*\big(\sigma(t)\big)\Big]\sigma^2(t)\right\}dt + h'\Big[h^*\big(\sigma(t)\big)\Big]\sigma(t)dB(t)$$

(donde $h^*(\bullet)$ denota la inversa de $h(\bullet)$). De lo que se puede concluir que el proceso $\{\sigma(t)\}$ admite una representación que constituye un caso particular de la ecuación general expresada en la página precedente, en el que $B(t) = B^*(t)$.

Una forma especial de esta representación, debida a J.C.Cox[246] se basa en la hipótesis de que la varianza condicionada $\sigma^2(t) = \sigma^2\big(v(t)/\sigma(t)\big)$ varía en forma inversamente proporcional al precio del activo y que, en consecuencia, la función $h^*(\bullet)$ puede ser definida como $h(x) = \sigma x^{\delta}$ $(x > 0)$ (donde $\sigma > 0$ y $0 \le \delta < 1$ son constantes).

Otro caso particular de la misma expresión general considera la siguiente representación para la volatilidad[247]:

$$d\sigma(t) = a\big(\sigma(t)\big)dt + \theta\sigma(t)dB^*(t)$$

(donde $\{B^*(t)\}$ denota un movimiento Browniano estandarizado correlacionado con $\{Z(t)\}$).

A partir de esta expresión, Wiggins,J.B. (1987) propuso la representación de la volatilidad mediante un proceso (recurrente al valor medio) de Uhlenbeck-Ornstein de la forma[248]:

$$d\ln\big(\sigma(t)\big) = \kappa\Big[\xi - \ln\big(\sigma(t)\big)\Big]dt + \theta\sigma(t)dB^*(t)$$

donde κ, ξ y θ son constantes (ver Scott, L.O. (1987)(1991), Chesney, M.; Scott, L.O. (1989)).

Otras formas de representación de la volatilidad, habituales en las aplicaciones, son las siguientes:

$$d\big(\sigma(t)\big) = \kappa\big[\xi - \sigma(t)\big]dt + \theta dB^*(t)$$

$$d\big(\sigma(t)\big) = \kappa\sigma(t)\big[\xi - \sigma(t)\big]dt + \theta\sigma(t)dB^*(t)$$

$$d\big(\sigma(t)\big) = \kappa\sigma(t)dt + \theta\sigma(t)dB^*(t)$$

---

246. Ver Cox, J.C.(1975), Beckers, S. (1980).

247. Obsérvese que, de acuerdo con esta hipótesis, se verifica que $\frac{Y(t)}{\sigma(t)}h'\big(Y(t)\big) = \delta$, por lo que a esta representación se la conoce también como el **proceso de difusión con elasticidad de la varianza constante**.

248. Ver Sec. 4.7.4.

$$d(\sigma(t)) = \frac{\xi - \kappa\sigma^2(t)}{\sigma(t)}dt + \theta dB^*(t)$$

$$d(\sigma^2(t)) = \kappa(\xi - \sigma^2(t))dt + \theta\sigma(t)dB^*(t)$$

Por su parte, Grünbichler, A.; Longsstaff, F.A. (1996) consideraron una representación de la forma:

$$d\sigma(t) = \kappa[\xi - \sigma(t)]dt + \theta\sqrt{\sigma(t)}\,dB^*(t)$$

donde el índice de volatilidad está calculado en base a volatilidades implícitas de una cartera de opciones[249].

Supóngase un activo cuya rentabilidad admita una representación en términos de un movimiento Browniano con tendencia de la forma:

$$v(t) = \varphi_0 dt + \sigma_v\, dB(t)$$

(ver Sec. 4.5). La aproximación discreta en el dominio del tiempo de esta ecuación puede ser expresada de la siguiente forma $v_t = \varphi_0 + \sigma_t U_t$ (donde $\{U_t\} = \left\{\frac{v_t - \varphi_0}{\sigma_t}\right\}$ denota un proceso estandarizado que habitualmente se supone un $WN - estricto$ o, en particular, un $WN - N(0;1)$ independiente de $\sigma(t)$ ver Sec. 1.3). Teniendo en cuenta que las variables σ_t son positivas y habitualmente con asimetría hacia la derecha, parecería apropiado asignarles una distribución log-Normal con una representación AR de la forma[250] $\ln(\sigma_t^2) = Z_t = \varphi_{Z0} + \varphi_{Z1}Z_{t-1} + \varepsilon_{Zt}$ (donde: **i)** $\{\varepsilon_{Zt}\}: WN - N(0,\sigma_{\varepsilon Z}^2)$; **ii)** $\{\varepsilon_{Zt}\} \perp \{U_t\}$; **iii)** $Z_t: N(m_Z,\sigma_Z^2) = N(\varphi_{Z0},\sigma_{\varepsilon Z}^2)$; **iv)** $\varphi_{Z0} = (1 - \varphi_{Z1})m_Z$ y **v)** $\{Z_t\}$ denota el proceso que representa el flujo de información asimétrica que ingresa al mercado)[251]. Entonces, a partir de estos supuestos se puede escribir $v_t - m = U_t\sigma_t = U_t e^{Z_t/2}$. Como $\{U_t\}$ siempre es estacionario, si se supone que $\{Z_t\}: I(0)$ (es decir, que $|\varphi_{Z1}| < 1$), entonces se puede asegurar que $\{v_t\}$ es débilmente estacionario y, teniendo en cuenta las propiedades de las variables log-Normales, se puede concluir que todos los momentos centrados de orden impar de v_t y los momentos absolutos de orden impar de $w_t = (v_t - m)^2$ son nulos y los correspondiente momentos de orden par son de la forma:

$$\mu_s(v_t) = E\left[(v_t - m)^s\right] = E\left(w_t^{s/2}\right) = m_{s/2}(w_t) =$$

$$= E\left(U_t^s e^{sZ_t/2}\right) = E\left(U_t^s\right)E\left(e^{Z_t\frac{s}{2}}\right) = E\left(U_t^2\right)M_{Z_t}\left(\frac{s}{2}\right) =$$

---

249. Ver Buff, R. (2002).

250. Conocida como **representación autorregresiva de la varianza (ARV)** (ver Wiggins, J.B. (1987), Chesney, M.; Scott, L.O. (1989), Melino, A.; Turnbull, S.M. (1990), Duffie, D.; Singleton, K. (1993)).

251. Ver Clark, P.K. (1973), Tauchen, G.E.; Pitts, M. (1983).

$$= \frac{s!}{2^{s/2}\left(\frac{s}{2}\right)!}\exp\left\{\left[m(Z)+\frac{1}{2}\sigma_Z^2\right]\frac{s}{2}\right\} \qquad (s = 2,4,6,...)$$

(donde $M_Z(\bullet)$ denota la función generatriz de momentos de la variable Z_t y, de acuerdo con la representación AR supuesta más arriba, $m(Z) = E(Z_t) = \frac{\varphi_{Z0}}{1-\varphi_{Z1}}$ y $\sigma_Z^2 = \sigma^2(Z_t) = \frac{\sigma_{Z\varepsilon}^2}{1-\varphi_{Z1}}$). A partir de estos resultados se obtiene que:

$$K(w_t) = \frac{E(w_t^2)}{\left[E(w_t)\right]^2} = \frac{E\left[(v_t - m)^4\right]}{\left\{E\left[(v_t - m)^2\right]\right\}^2}$$

Lo que demuestra que, en concordancia con la experiencia, esta representación supone que las variables v_t son leptokúrticas y que las "colas" de su distribución son más levantadas que en la Normal.

Asimismo, la función de autocovarianzas del proceso $\{w_t\}$ es de la forma:

$$E(w_{t-j}, w_t) = E\left[\left(U_{t-j}^2 e^{Z_{t-j}}\right)\left(U_t^2 e^{Z_t}\right)\right] = E\left(U_{t-j}^2 U_t^2\right)E\left(e^{Z_{t-j}} e^{Z_t}\right) = E\left(e^{Z_{t-j}} e^{Z_t}\right) = E\left(e^{Z_{t-j}+Z_t}\right)$$

Luego, de acuerdo con las propiedades de la distribución log-Normal, se puede escribir:

$$E(w_{t-j}w_t) = \exp\left\{\left(m_Z + \sigma_Z^2\right) + \left[m_Z + \rho_j(Z)\sigma_Z^2\right]\right\} = \exp\left[2m_Z + \left(1+\varphi_{Z1}^j\right)\sigma_Z^2\right]$$

y, por lo tanto:

$$\gamma(w_{t-j}, w_t) = E(w_{t-j}w_t) - E(w_{t-j})E(w_t) =$$

$$= \exp\left[2m_Z + \left(1+\varphi_{Z1}^j\right)\sigma_Z^2\right] - \exp\left(m_Z + \frac{\sigma_Z^2}{2}\right)\exp\left(m_Z + \frac{\sigma_Z^2}{2}\right) =$$

$$= \exp\left(2m_Z + \sigma_Z^2\right)\left[\exp\left(\sigma_Z^2\varphi_{Z1}^j\right) - 1\right] \qquad (j = 0,\pm1,\pm2,...)$$

De modo que la función de autocorrelaciones queda, en este caso, definida de la siguiente forma:

$$\rho(w_{t-j}, w_t) = \frac{\exp\left(\varphi_{Z1}^j\sigma_Z^2\right) - 1}{e^{\sigma_Z^2} - 1} \qquad (j = 0,\pm1,\pm2,...)$$

Teniendo en cuenta, entonces, que: **i)** $w_t = (v_t - m)^2 = \sigma_t^2 U_t^2$ y **ii)** $Z_t = \ln\left(\sigma_t^2\right)$, se puede escribir $\ln(w_t) = \ln\left(\sigma_Z^2\right) + \ln\left(U_t^2\right) = Z_t + \ln\left(U_t^2\right)$ y, como $(1-\varphi_{Z1}B)(Z_t - m_Z) = \varepsilon_{Zt}$, será:

$$\ln(w_t) = w_t^* = m_Z + \frac{\varepsilon_{Zt}}{1-\varphi_{Z1}} + \ln\left(U_t^2\right)$$

Lo que demuestra que el proceso $\{w_t^*\}$ admite una representación $ARMA(1,1)$ con perturbaciones no-Normales[252].

Debe tenerse en cuenta que si, como se supuso, $U_t: N(0;1)$, entonces la variable $\ln(U_t^2)$ tiene una distribución con una cola muy larga hacia la izquierda (generada por los logaritmos de los valores muy pequeños de U_t^2), con medidas $E[\ln(U_t^2)] = -1.27$ y $\sigma^2[\ln(U_t^2)] = 4.93$.

A partir de los estadísticos de orden (ascendente), $v_{(1)} \leq v_{(2)} \leq ... \leq v_{(n/2)} \leq ... \leq v_{(n)}$, obtenidos de la serie cronológica $\{v_t, t = 1,2,...,n\}$, su mediana puede ser definida como $me(v) = v_{([n+1]/2)}$ (donde $\left[\frac{n+1}{2}\right]$ denota la parte entera de $\frac{n+1}{2}$). Luego, si la variable v_t es simétrica, se verificará que:

$$v_{(n+1-t)} - me(v) = me(v) - v_{(t)} \qquad \left(t < \frac{n+1}{2}\right)$$

Es decir, que la variable v_t puede ser considerada simétrica si el diagrama de dispersión de los puntos $(v_{(n-t)}, v_{(t)})$ ocupan el mismo lugar geométrico que una recta de pendiente -1.

Como se mencionó en la Sec. 4.7.6.1, otra aproximación a la representación de la volatilidad como proceso estocástico (que será tratada en la **Parte III**, referida a los modelos no-lineales) está dada por los modelos autorregresivos generales para procesos con heterocedasticidad condicionada (GARCH).

---

[252]. Dado que w_t puede ser nula, se puede emplear la transformación de Koopman-Harvey-Doornik-Shephard $\ln(w(t)) = \ln[w(t) + (0.02)\sigma^2(w)] - \frac{(0.02)\sigma^2(w)}{w_t + (0.02)\sigma^2(w)}$ (ver Koopman, S.J.; Harvey, A.C.; Doornik, J.A.; Shephard, N. (1995)).

CAPITULO 5

Una introducción a los procesos de parámetro discreto con orden de integración fraccionario

"En un día del hombre están los días
del tiempo, desde aquél inconcebible
día inicial del tiempo, en que un terrible
Dios prefijó los días y agonías
hasta aquél otro en que el ubicuo río
del tiempo terrenal torne a su fuente,
que es lo eterno, y se apague en el presente".

(Borges, J.L.: "Elogio de la sombra")

5.1.- Acerca de la memoria de los procesos

Sea un proceso $\{Y_t\}:I(d)$ que admite una representación de la forma:

$$\varphi(B)Y_t = \Theta(B)\varepsilon_t$$

$$\Phi(B)(1-B)^d Y_t = \Phi(B)\left(\nabla^d Y_t\right) = \Theta(B)\varepsilon_t$$

(donde **i)** Y_t denota una variable centrada; **ii)** $\{\varepsilon_t\}:WN-N\left(0.\sigma_\varepsilon^2\right)$; **iii)** $\Phi(B)$ denota un operador $AR(p)$ estacionario y **iv)** $\Theta(B)$ denota un operador $MA(q)$) en la que, contrariamente a lo considerado en las representaciones ARIMA analizadas en el Cap. 3, d puede asumir cualquier valor real comprendido en el intervalo $(-1;1)(d \neq 0)$.

Este tipo de procesos, conocido como $ARFIMA(p,d,q)$, para $0 < d < \frac{1}{2}$, es **de memoria larga** (es estacionario, aunque presenta autocorrelaciones significativas entre variables Y_t separadas por períodos muy largos) y, para $-\frac{1}{2} < d < 0$, es de memoria corta.

Los procesos ARFIMA (**"autorregressive fractionally integrated moving average"**) fueron introducidos por Mandelbrot, B.; Van Ness, J.W. (1968) para formalizar los resultados obtenidos empíricamente por H.E. Hurst en su "Long-term storage capacity of reservoirs" (ver Sec. 5.4). Admitir que d puede asumir valores no-enteros coloca a estas representaciones como un nexo entre las representaciones $ARIMA(p,1,q)$ y $ARMA(p,q)$ (ver Bailli, R.I. (1996)).

Como se mencionó en la Sec. 1.2, a partir de una interpretación constructivista, la memoria del proceso puede ser caracterizada a partir del comportamiento de sus sumas parciales $S_t = \sum_{j=1}^{t} Y_j \quad (t = 1,2,...)$ donde, como se expresó más arriba, las variables Y_t son centradas. De acuerdo con Lee, D.; Schmidt, P. (1996), la condición necesaria para que $\{Y_t\}$ pueda ser considerado un proceso de memoria corta es que $0 < \sigma^2 = \lim_{t\to\infty} E\left(\frac{1}{t}S_t^2\right) < \infty$ (es decir, que $S_t = O_p\left(\sqrt{t}\right)$), y la condición suficiente es que $\frac{1}{\sigma\sqrt{t}} S_{[kt]} \xrightarrow{d} B(k) \, (0 \le k \le 1)$ (donde $[kt]$ denota la parte entera de kt y $B(k)$ es un movimiento Browniano)[1]. Esta condición se conoce como el **principio de invariancia[2] para la convergencia de sumas parciales a un movimiento Browniano)[3]**.

5.2.- Los procesos WN con orden de diferenciación fraccionario

5.2.1.- Definición

Sea, en particular, un proceso $\{Y_t\}$ que admite una representación $ARFIMA(0,d,0)$ de la forma:

$$(1 - B)^d Y_t = \nabla^d Y_t = \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$). A estos procesos habitualmente se los denomina **WN con orden de diferenciación fraccionario (FWN)[4]** o **fraccionalmente integrados**, de acuerdo con la nomenclatura de Granger-Joyeux-Hosking (ver Granger, C.W.J.; Joyeux, R. (1980), Hosking, J.R.M. (1981))[5].

---

[1]. La definición de Lee-Schmidt sobre la memoria corta supone implícitamente: **i)** la existencia de momentos en la distribución de las variables Y_t (obsérvese que un proceso formado por variables independientes con distribución de Cauchy no puede ser considerado de memoria corta); **ii)** el cumplimiento de las restricciones sobre la persistencia de la dependencia sobre las variables aleatorias del proceso. Por su parte, Lo, A.W. (1991) define como de memoria corta a todo proceso que posea momentos absolutos de orden $s > 2$ y cumpla la condición α-mixing con coeficientes α_j tales que $\sum_{j=1}^{\infty} \alpha_j^{1-2/\alpha} < \infty$ (condiciones conocidas como **supuesto de Phillips**, ver Phillips, P.C.B. (1987)).

[2]. Ver Sec. 2.5.

[3]. Obsérvese que, de acuerdo con estas condiciones, un proceso de memoria corta no tiene por qué ser necesariamente estacionario de segundo orden. En particular, si $\{Y_t\}$ fuera estacionario de segundo orden, la varianza de largo plazo, σ^2, será proporcional a la función de densidad espectral para $\alpha = 0$ que, como se expresó en la Sec. 2.5, no es nula ni infinita.

[4]. **"Fractionally white noise"**.

[5]. Como se verá en la Sec. 5.5, esta condición sobre el proceso $\{\varepsilon_t\}$, debida a Lee, P.; Schmidt, P. (1996), está dirigida a asegurar la consistencia de los tests de estacionariedad con respecto a alternativas de orden de diferenciación fraccionario. En realidad, se demuestra que las condiciones necesarias para la consistencia de estos tests son menos restrictivas: **i)** Sowell, F. (1990) propone la condición que $\{\varepsilon_t\}$:$WN - estricto$, con variables que posean valor esperado nulo y el momento

Teniendo en cuenta que, para $d > -1$, la función $f(B) = \nabla^d = (1-B)^d$ admite un desarrollo binomial de la forma:

$$f(B) = f(0) + \frac{1}{1!}\frac{d}{dB}f(B)\Big|_{B=0}B + \frac{1}{2!}\frac{d^2}{dB^2}f(B)\Big|_{B=0}B^2 + \ldots =$$

$$= \sum_{j=0}^{\infty}\frac{d(d-1)\ldots\left[d-(j-1)\right]}{j!}(-B)^j =$$

$$= \sum_{j=0}^{\infty}\frac{\left\{d(d-1)\ldots\left[d-(j-1)\right]\right\}\left\{(d-j)\left[d-(j+1)\right]\ldots 2.1\right\}}{j!\left\{(d-j)\left[d-(j+1)\right]\ldots 2.1\right\}}(-B)^j$$

la representación AR de estos procesos puede ser expresada de la siguiente forma:

$$\sum_{j=0}^{\infty}\binom{d}{j}(-B)^j Y_t = \sum_{j=0}^{\infty}\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}(-B)^j Y_t = \sum_{j=0}^{\infty}\pi_j Y_{t-j} = \Pi(B)Y_t = \varepsilon_t$$

donde,

$$\Gamma(x) = \begin{cases} \int_0^{\infty} u^{x-1}e^{-u}\,du & (x > 0) \\ \infty & (x = 0) \\ \dfrac{\Gamma(x+1)}{x} & (x < 1) \end{cases}$$

De modo que, aplicando la fórmula de De Moivre-Stirling[6], para valores de j suficientemente grandes, los coeficientes[7]:

$$\pi_j = \pi_{j-1}\frac{j-1-d}{j} = \prod_{i=0}^{j}\frac{i-1-d}{i}\ (j = 1,2,\ldots)$$

pueden ser aproximados de la siguiente forma: $\pi_j \approx \dfrac{j^{-d-1}}{\Gamma(-d)}$. En particular, para $d > -\dfrac{1}{2}$ se verifica

---

absoluto de orden s finito, para algún $s \geq Max\left[\dfrac{(4.8)d}{1+2d}\right]$; **ii)** Taqqu, M. (1975) y Lo, A.W. (1991) proponen la condición que $\{\varepsilon_t\}: I(0)$ con variables Normales, pero no iid.

[6]. Según la cual $\lim_{x\to\infty}\Gamma(x) = \sqrt{2\pi}e^{-x+1}(x-1)^{x-1/2}$.

[7]. Obsérvese que, según que d sea positivo o negativo, los π_j serán negativos o positivos.

que la sucesión de coeficientes $\{\pi_j\}$ es absolutamente sumable $\left(\sum_{j=0}^{\infty}|\pi_j| < \infty\right)$, de modo que se puede asegurar que el proceso $\{Y_t\}$ es invertible y que, por lo tanto, admite una representación MA de orden infinito de la forma:

$$Y_t = (1-B)^{-d}\varepsilon_t = \nabla^{-d}\varepsilon_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \left\{\sum_{j=0}^{\infty}\frac{d(d+1)...[d+(j-1)]}{j!}B^j\right\}\varepsilon_t =$$

$$= \left\{\sum_{j=0}^{\infty}\frac{[d+(j+1)]...(d+1)d(d-1)...2.1}{j![d(d-1)...2.1]}B^j\right\}\varepsilon_t =$$

$$= \left[\sum_{j=0}^{\infty}\frac{\Gamma(d+j)}{\Gamma(j+1)\Gamma(d)}B^j\right]\varepsilon_t$$

Aplicando, entonces, la fórmula de De Moivre-Stirling, para valores de j suficientemente grandes, los coeficientes:

$$\psi_j = \frac{d(d+1)...[d+(j-1)]}{j!} =$$

$$= \frac{d}{1}\frac{d+1}{2}...\frac{d+(j-3)}{j-2}\frac{d+(j-2)}{j-1}\frac{d+(j-1)}{j} =$$

$$= \frac{j-(j-1)+d-1}{j-(j-1)}\frac{j-(j-2)+d-1}{j(j-2)}...\frac{j-2+d-1}{j-2}\frac{j-1+d-1}{j-1}\frac{j+d-1}{j} =$$

$$= \left(1+\frac{d-1}{j-(j-1)}\right)\left(1+\frac{d-1}{j-(j-2)}\right)...\left(1+\frac{d-1}{j-2}\right)\left(1+\frac{d-1}{j-1}\right)\left(1+\frac{d-1}{j}\right) =$$

$$= \prod_{i=0}^{j}\left(1+\frac{d-1}{j}\right) = \psi_{j-1}\left(1+\frac{d-1}{j}\right)$$

pueden ser aproximados de la siguiente forma $\psi_j \approx \left(1+\frac{1}{j}\right)^{d-1} = \frac{j^{d-1}}{\Gamma(d)}$ (obsérvese que, según que d sea positivo o negativo, los ψ_j serán positivos o negativos)[8]. De esta expresión se puede concluir que, para $d < \frac{1}{2}$, la sucesión de coeficientes $\{\psi_j\}$ es de cuadrado sumable y, en consecuencia, se verifica que:

$$\sigma^2(Y) = \sigma^2\left(\varepsilon_t + \frac{1}{\Gamma(d)}\sum_{j=1}^{\infty}j^{d-1}\varepsilon_{t-j}\right) = \left\{1+\frac{1}{[\Gamma(d)]^2}\sum_{j=1}^{\infty}j^{2(d-1)}\right\}\sigma_\varepsilon^2 < \infty$$

---

[8]. Ver Granger, C.W.J. (1980), Granger, C.W.J.; Joyeux, R. (1980), Hosking, J.R.M. (1981).

y, por lo tanto, que el proceso $\{Y_t\}$ es débilmente estacionario (es decir que, para $|d| < \frac{1}{2}$, $\{Y_t\}$ es estacionario e invertible).

Obsérvese que: **i)** para $|d| > \frac{1}{2}$, $\sigma^2(Y)$ es infinita y, por lo tanto, el proceso es no-estacionario; **ii)** para $d = \frac{1}{2}$, es invertible y **iii)** para $d = -\frac{1}{2}$, es no-invertible.

A partir de los resultados obtenidos con relación a las representaciones AR y MA de $\{Y_t\}$ se puede concluir que las sucesiones $\{\psi_j\}$ y $\{|\pi_j|\}$ observan una convergencia más lenta que la exponencial y, por lo tanto, que el proceso $\{Y_t\}$ (no importa que se apliquen o no diferencias sucesivas) no admitirá una representación $ARMA(p,q)$ satisfactoria para valores de j suficientemente grandes (la característica particular de los procesos con memoria larga, conocidos también como **sensible a las condiciones iniciales**, es la posibilidad de admitir representaciones marcadamente diferentes de las representaciones ARMA, en los cuales los coeficientes observan, de acuerdo con los postulados del teorema de Wold, un decrecimiento exponencial) .

5.2.2.- Las funciones de autocorrelaciones y de autocorrelaciones parciales

Como se ha visto hasta aquí, se dice que un proceso $\{Y_t\}:I(0)$ es de **memoria corta**, debido a que su función de autocorrelaciones -que es geométricamente acotada- observa un comportamiento del tipo $|\rho_j(Y)| \leq cr^{-j}$ $(c > 0; 0 < r < 1; j = 1,2,...)$ y se denominan de **memoria larga** aquellos procesos estacionarios que se caracterizan porque su función de autocorrelaciones observa un comportamiento de la forma $\lim_{j\to\infty} \rho_j(Y) \approx c\, j^{2d-1}$ $\left(c > 0; d < \frac{1}{2}\right)$ (esta condición, debida a Brockwell, P.J.; Davies, R.A. (1991) y Beran, J. (1994) es equivalente a la siguiente condición sobre la función de densidad espectral: $\lim_{\alpha\to\infty} w_t(\alpha) = c^*|\alpha|^{2d-1}$ $\left(c^* > 0; d < \frac{1}{2}\right)$). Esta condición de persistencia de largo plazo de la función de autocorrelaciones -que se conoce como el "**fenómeno de Hurst**" y que puede ser interpretada como una tendencia a que los valores grandes de Y_t sean seguidos por valores grandes del mismo signo, de modo que la estructura del proceso asume la forma de una sucesión de factores cíclicos (que puede incluir ciclos de período tan largo como la longitud de la serie observada)- hace que estos procesos sean particularmente útiles para realizar predicciones de largo plazo[9]. En términos más generales, se puede decir[10] que un proceso es de memoria larga si su función de autocorrelaciones, para valores de j suficientemente grandes, puede ser aproximada de la siguiente forma:

$$\rho_j(Y) \approx G(j)\, j^{2H-2} \quad (H > 0)$$

---

[9]. Parzen, E. (1982) consideró como procesos de memoria larga a aquellos que admiten una representación de la forma $\varphi(B)\Phi(B)Y_t = \Theta(B)\varepsilon_t$ (donde: **i)** $\varphi(B)$ denota un operador AR con raíces menores que la unidad; **ii)** $\Phi(B)$ denota un operador AR estacionario; **iii)** Y_t denota una variable centrada y **iv)** $\{\varepsilon_t\}:WN$) y los denominó **ARARMA**. Más adelante, en esta misma sección, se analizarán métodos para detectar la presencia y, eventualmente, medir la memoria de largo plazo del proceso.

[10]. Ver Resnick, J. (1987), Helson, J.; Sarason, Y. (1967).

(donde $G(j)$ es una función de variación lenta en el infinito (ver Sec. 5.4) y H denota el **coeficiente de Hurst**), en la que no se cumplen las condiciones "mixing" (ver Sec. 1.5)[11].

De acuerdo a lo expresado en la Sec. 2.5, la función de autocovarianzas del proceso $\{Y_t\}$ puede ser expresada de la siguiente forma[12]:

$$\gamma_j(Y) = \int_{-\pi}^{\pi} e^{i\alpha j}\omega(\alpha)d\alpha = \int_{-\pi}^{\pi}\omega(\alpha)\cos(\alpha j)d\alpha = c\int_{-\pi}^{\pi}2^{-d}\cos(\alpha j)\left(sen\left(\frac{\alpha}{2}\right)\right)^{-2d}d\alpha =$$

$$= \frac{\sigma_\varepsilon^2}{2\pi}sen(\pi d)\frac{\Gamma(j+d)\Gamma(1-2d)}{\Gamma(j+1-d)} = \sigma_\varepsilon^2\frac{\Gamma(j+d)\Gamma(1-2d)}{\Gamma(j+1-d)\Gamma(1-d)\Gamma(d)} =$$

$$= (-1)^j\frac{\Gamma(1-2d)}{\Gamma(1-j-d)\Gamma(j+1-d)}\sigma_\varepsilon^2 \qquad \left(j = 0,1,2,\ldots;-1 < d < \frac{1}{2};d \neq 0\right)$$

De modo que la función de autocorrelaciones queda definida por:

$$\rho_j(Y) = \frac{\Gamma(1-d)}{\Gamma(d)}\frac{\Gamma(j+d)}{\Gamma(j+1-d)} = \prod_{i=1}^{j}\frac{i-1+d}{i-d}\ (j = 1,2,\ldots)$$

Es decir, que $\rho_1(Y) = \frac{d}{1-d}$ y, para valores de j suficientemente grandes y $d < \frac{1}{2}\ (d \neq 0)$, puede ser aproximada de la siguiente forma: $\rho_j(Y) \approx \frac{\Gamma(1-d)}{\Gamma(d)}j^{2d-1}$ (debe tenerse en cuenta que, de acuerdo con la fórmula de Sheppard, para valores de j suficientemente grandes, $\frac{\Gamma(j+a)}{\Gamma(j+b)} \approx \approx j^{a-b}$).

Obsérvese que: **i)** para $0 < d < \frac{1}{2}$ la función de autocorrelaciones es positiva para todo $j > 0$, experimenta un decrecimiento hiperbólico cuya velocidad depende de d (j^{2d-1}, más lento que el de la función de autocorrelaciones correspondiente a un proceso ARMA) y no es absolutamente sumable[13]

---

[11]. Diseñado por Harold Edwin Hurst para estudiar el comportamiento del río Nilo (ver Hurst, H.E. (1951), Hurst, H.E.; Black, R.P.; Sinaika, Y.M. (1965)). B.B. Mandelbrot denotó a este coeficiente con la letra H como reconocimiento a su descubridor y lo utilizó para distinguir entre series cronológicas aleatorias y series cronológicas fractales en las aplicaciones económicas (ver Mandelbrot, B.B. (1974) (1975)).

[12]. Debe tenerse en cuenta que $\int_0^{\pi}\cos(\alpha j)(sen(\alpha))^{k-1}d\alpha = \frac{\pi\cos\left(\frac{j\pi}{2}\right)\Gamma(k+1)2^{1-k}}{k\Gamma\left(\frac{k+j+1}{2}\right)\Gamma\left(\frac{k-j+1}{2}\right)}$ (ver Gradszteyn, I.S.; Ryzhik, M. (1980)).

[13]. En general, los procesos de memoria larga se caracterizan porque la función de autocorrelaciones de la serie original observa un comportamiento no-estacionario y la correspondiente al proceso diferenciado se comporta como la función de autocorrelaciones correspondiente a un proceso sobre-diferenciado.

y **ii)** para $-\frac{1}{2} < d < 0$ los coeficientes de autocorrelación para $j = 1,2,...$ son negativos y absolutamente sumables y, por lo tanto, $\{Y_t\}$ es un proceso de memoria corta. Si bien el segundo caso es menos interesante que el primero, suele resultar de utilidad en la representación de las primeras diferencias de los procesos $ARFIMA(0,d,0)$ no-estacionarios, con $\frac{1}{2} < d < 1$. Estos procesos son particularmente apropiados para representar el comportamiento de procesos no-estacionarios, pero con reversión al valor medio (que no admiten representaciones con raíces unitarias).

Luego, para $0 < d < \frac{1}{2}$ la condición $\sum_{j=-\infty}^{\infty} |\rho_j(Y)| = \infty$ es necesaria y suficiente para que $\{Y_t\}$ pueda ser considerado de memoria larga[14]. Lo cual, de acuerdo a lo expresado en la Sec. 2.5, implica que, en ese caso, la función de densidad espectral es no-acotada para frecuencias bajas y, por lo tanto que $\lim_{\alpha \to 0} w_t(\alpha) = \infty$. Esta propiedad se conoce como "**comportamiento espectral típico de una variable económica**" de acuerdo con la nomenclatura de Granger, C.W.J. (1966) o "**efecto José**" de acuerdo con la nomenclatura de Mandelbrot, B.B.; Wallis, J. (1968)[15].

De acuerdo con el algoritmo de Durbin-Levinson mencionado en la Sec. 2.10.4, se obtienen los coeficientes:

$$\phi_{pj} = -\binom{p}{j} \frac{\Gamma(j-d)\Gamma(p-d-j+1)}{\Gamma(-d)\Gamma(p-d+1)} \qquad (j = 1,2,...,p)$$

de modo que la función de autocorrelaciones parciales queda definida de la siguiente forma:

$$\phi_{jj}(Y) = -\frac{\Gamma(j-d)\Gamma(1-d)}{\Gamma(-d)\Gamma(j-d+1)} \qquad (j = 1,2,...)$$

Para valores de j suficientemente grandes, estos coeficientes pueden ser aproximados de la siguiente forma $\phi_{jj}(Y) \approx \frac{d}{j-d}$, lo que demuestra que la función de autocorrelaciones parciales observa un comportamiento decreciente del tipo j^{-1}, independiente de d.

---

[14]. Ver McLeod, A.J.; Hipel, K.W. (1978). Como corolario de esta condición se puede asegurar que, en consecuencia, la sucesión de coeficientes de la representación MA de orden infinito serán de cuadrado sumable pero no absolutamente sumables. Algunos autores distinguen entre: **i) procesos de memoria intermedia** (o **antipersistentes**) que se presentan cuando $-\frac{1}{2} < d < 0$ y, por lo tanto $\sum_{j=-\infty}^{\infty} |\rho_j(Y)| < \infty$ y **ii)** estos **procesos de memoria larga**, que se presentan cuando $0 < d < \frac{1}{2}$, es decir, cuando la función de autocorrelaciones no es absolutamente sumable. Como se verá más adelante, existe una propiedad más general que la de memoria larga, es la de **memoria extendida**, la cual se verifica cuando el predictor mínimo-cuadrático (óptimo) $f^{(t,\ell)} = f(Y_{t+\ell} / Y_t, Y_{t-1}, ..., Y_{t-n})$ no converge a una constante para todo $t \geq 0$, cuando $\ell \to \infty$. La diferencia entre estas dos propiedades radica en que la de memoria larga toma en consideración sólo las características lineales del proceso, mientras que la de memoria extendida puede incluir predicciones no-lineales óptimas (ver Mandelbrot, B.B. (1972), Granger, C.W.J. (1980)).

[15]. Debe tenerse en cuenta que la suma de procesos $AR(1)$ independientes con coeficiente autorregresivo aleatorio -por ejemplo, con distribución Beta- define un proceso ARFIMA y que muchos procesos de la economía surgen de la agregación de otros procesos, por lo que su aparición en las aplicaciones es muy frecuente (ver Haubrich, J.G.; Lo, A.W. (1989)).

5.2.3.- La función de densidad espectral

A partir de la expresión correspondiente a la función de densidad espectral del proceso $\{\varepsilon_t\}: wn-debil$, $w_\varepsilon(\alpha)=\frac{\sigma_\varepsilon^2}{2\pi}$, queda definida la función de densidad espectral de $\{Y_t\}$:

$$\omega(\alpha)=\left|1-e^{-i\alpha}\right|^{-2d}\omega_\varepsilon(\alpha)=\left[2(1-\cos(\alpha))\right]^{-d}\frac{\sigma_\varepsilon^2}{2\pi}=\left[4\,sen^2\left(\frac{\alpha}{2}\right)\right]^{-d}\frac{\sigma_\varepsilon^2}{2\pi}$$

Para , $\alpha\to 0$ esta función puede ser aproximada de la siguiente forma: $\omega_Y(\alpha)\approx\frac{\sigma_\varepsilon^2}{2\pi}\alpha^{-2d}$ (dado un proceso $\{Y_t\}: I(d)$ con función de densidad espectral de la forma:

$$\omega_Y(\alpha)=\left|1-e^{-i\alpha}\right|^{-2d}\omega_Z(\alpha)=\quad\left(4\,sen^2\left(\frac{\alpha}{2}\right)\right)^{-d}\omega_Z(\alpha)\,(\alpha\neq 0)$$

(donde $Z_t=\nabla^d Y_t$), con d fraccionario, es posible hallar una representación $ARIMA(p,d,q)$, con d entero, cuya estructura presente una muy buena aproximación a $\omega_Y(\alpha)$ para cualquier frecuencia α , pero no para $\alpha=0$) .

De las expresiones anteriores se puede concluir que: **i)** que como se mencionó en la sección precedente, para $0<d<\frac{1}{2}$, la función $w_Y(\alpha)$ es decreciente, no-acotada en el origen y, por lo tanto, que $\{Y_t\}$ es de memoria larga pero que, para el proceso $\{\nabla Y_t\}$, la función de densidad espectral es nula en el origen, por lo que se puede concluir que está sobrediferenciado; **ii)** que para $d<0$, $w_Y(\alpha)$ se anula en el origen, observa una dominancia de las frecuencias altas y, por lo tanto, que $\{Y_t\}$ es antipersistente (en el origen la función de densidad espectral de la suma parcial del proceso (S_t) es infinita, ver Sec. 5.1).

5.3.- Los procesos ARFIMA(p,d,q)

5.3.1.- Definición

Retornando a la expresión general mencionada en la Sec. 5.1, sea un proceso $\{Y_t\}: I(d)$ que admite una representación $ARIMA(p,d,q)$ de la forma:

$$\varphi(B)Y_t=\Phi(B)(1-B)^d Y_t=\nabla^d Y_t=\Theta(B)\varepsilon_t\qquad\qquad(-1<d<1;d\neq 0)$$

De acuerdo con Campbell, J.Y.; Mankiw, N.G. (1987), a partir de la primera diferencia de

$\{Y_t\}$ se obtiene la siguiente representación MA[16]:

$$(1-B)Y_t = (1-B)^{1-d}\frac{\Theta(B)}{\Phi(B)}\varepsilon_t = H(d-1,1,1;B)\frac{\Theta(B)}{\Phi(B)}\varepsilon_t = \Psi^*(B)\varepsilon_t = \left(1+\sum_{j=1}^{\infty}\psi_j^* B^j\right)\varepsilon_t$$

(donde $H(\bullet)$ denota la función hipergeométrica $H(a,b,c;e) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\sum_{j=1}^{\infty}\frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)\Gamma(j+1)}e^j$).

Teniendo en cuenta que, para $d > 0$, $H(d-1,1,1;B) = 0$, entonces se puede concluir que $\Psi^*(1) =$ $= H(d-1,1,1;1)\frac{\Theta(1)}{\Phi(1)} = 0$ (ver Gradszteyn, I.S.; Ryzhic, I.M. (1980)). Dado que, de acuerdo a lo expresado en la Sec. 2.9.1, la medida de la influencia de un "shock" introducido en el proceso en el momento t sobre la variable Y_{t+j} estará dada por $1+\sum_{i=1}^{j}\psi_i^*$ y que, para un proceso con reversión al valor medio, se verifica que $\Psi(1) = 0$, entonces de las expresiones que figuran más arriba se puede concluir que todo proceso $\{Y_t\}: I(d)\,(d < 1)$ posee la propiedad de reversión al valor medio. Esta propiedad se verificará aún en el caso en que $\frac{1}{2} < d < 1$, para el cual $\{Y_t\}$ no es estacionario de segundo orden. En otros términos, podría decirse que una representación ARFIMA no-estacionaria es "un poco menos" no-estacionaria que una representación ARIMA (debe tenerse en cuenta que, para $d = 1$, se verifica que $\Psi^*(1) = \frac{\Theta(1)}{\Phi(1)} \neq 0$, es decir, que no hay reversión al valor medio).

5.3.2.- La función de autocovarianzas

La función de autocovarianzas puede ser expresada de la siguiente forma[17]:

$$\gamma_j(Y) = \sigma_\varepsilon^2\sum_{h=1}^{p}\xi_h\sum_{i=0}^{q}\sum_{k=0}^{q}\theta_i\theta_k\, f(d,d,p+i-k-j;G_j)$$

donde: G_j denota la j -ésima raíz de la ecuación característica $\Phi(B) = 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p = 0$

$$\xi_h = \frac{1}{G_h\prod_{i=1}^{p}[1-\rho_i(Y)\rho_h(Y)]\prod_{\substack{j=1\\(j\neq h)}}^{p}[\rho_i(Y)-\rho_j(Y)]}$$

---

16. Esta representación generalizada permite describir simultáneamente las características de largo plazo (mediante el coeficiente d) y las de corto plazo (a partir de los coeficientes de la componente ARMA). Dado que, para observaciones muy distantes en el tiempo, la influencia de los coeficientes de la componente ARMA se vuelven despreciables, en el largo plazo el comportamiento de este proceso es asimilable al del $ARFIMA(0,d,0)$.

17. Ver Sowell, F.B. (1986)(1992), Chung, C.F. (1994).

$$f(a,b,c;e) = g(a,b,c)\left[e^{2e}H(b+c,1,1-a+c;e) + H(a-c,1,1-b-c;e)\right]$$

y,

$$g(a,b,c) = \frac{\Gamma(1-a-b)\Gamma(b+c)}{\Gamma(1-a+c)\Gamma(1-b)\Gamma(b)}$$

De acuerdo con Chung, C.F. (1994), esta función de autocovarianzas puede ser expresada alternativamente mediante la siguiente fórmula de cálculo:

$$\gamma_j(Y) = \sum_{i=-q}^{q} a_i \sum_{h=1}^{p} \theta_h f(j-p-i;\psi_h)$$

donde:

$$a_i = \frac{1}{\psi_i \prod_{j=1}^{p}(1-\psi_i\psi_j)\prod_{\substack{h=1 \\ h\neq i}}^{p}(\psi_i-\psi_h)}$$

$$\psi_i = \sum_{j=0}^{q-|i|} \theta_j\theta_{j+|i|}$$

y,

$$f(j-p-i,\psi_h) = \psi_i^{2p}\sum_{k=0}^{\infty}\psi_j^k\gamma_{j-p-i-k}(W) + \sum_{\ell=1}^{\infty}\psi_j^{\ell}\gamma_{j-p-i+\ell}(W)$$

siendo $\{W_t\}$ un proceso que admite una representación $ARFIMA(0,d,0)$.

Se demuestra que: **i)** si $d < \frac{1}{2}$ y la raíces de la ecuación característica de la componente AR ($\Phi(B) = 0$) son todas, en módulo, mayores que la unidad, entonces $\{Y_t\}$ es estacionario de segundo orden y **ii)** si $d > -\frac{1}{2}$ y las raíces de la ecuación característica de la componente MA ($\Theta(B) = 0$) son todas, en módulo, mayores que la unidad, entonces $\{Y_t\}$ es invertible (ver Hosking, J.R.M. (1981)).

5.3.3.- La función de densidad espectral

Sea un proceso $\{Y_t\}$ que admite una representación AR de orden infinito de la forma:

$$\sum_{j=0}^{\infty}\pi_j Y_{t-j} = \Pi(B)Y_t = \varepsilon_t$$

Su función de densidad espectral será, entonces, de la forma:

$$\omega(\alpha) = \frac{c}{\Pi\left(e^{-i\alpha}\right)\Pi\left(e^{i\alpha}\right)}$$

(donde, de acuerdo a lo expresado en la Sec. 2.5, $\Pi\left(e^{-i\alpha}\right) = \left(1 - e^{-i\alpha}\right)^d$). A partir de esta expresión se puede concluir que: **i)** para $d < 0$, $w(0) = 0$; **ii)** para $d > 0$, $w(0) = \infty$; **iii)** para $d > 0$ y, para valores de α próximos a cero, $w(\alpha) = \alpha^{-2d}$ y **iv)** para $d > \frac{1}{2}$, $\sigma^2(Y) = \infty$.

Alternativamente a la definición anterior, la función de densidad espectral puede ser expresada de la siguiente forma[18]:

$$\omega(\alpha) = \frac{\sigma_Y^2}{2\pi}\frac{\left|\Theta\left(e^{-i\alpha}\right)\right|^2}{\left|\Phi\left(e^{-i\alpha}\right)\right|^2}\frac{1}{\left|1 - e^{-i\alpha}\right|^{2d}} = \frac{\sigma_Y^2}{2\pi}\frac{\left|\Theta\left(e^{-i\alpha}\right)\right|^2}{\left|\Phi\left(e^{-i\alpha}\right)\right|^2}\frac{1}{\left[2\left|1 - \cos(\alpha)\right|\right]^{2d}}$$

De modo que, para frecuencias bajas, se puede utilizar la siguiente aproximación[19]:

$$\lim_{\alpha \to 0}\omega(\alpha) = \frac{\sigma_Y^2}{2\pi}\left[\frac{\Theta(1)}{\Phi(1)}\right]^2 \alpha^{-2d}$$

5.3.4.- La estimación de los coeficientes

Existen dos métodos de estimación de las representaciones de memoria larga: **i)** uno **en dos etapas**, en el que se estima en primer lugar el parámetro d mediante la expresión $\hat{d} = \frac{\hat{\rho}_1(Y)}{1 + \hat{\rho}_1(Y)}$ y, en segundo lugar, los coeficientes de la representación ARMA y **ii)** otro **en una etapa** en el que se estiman en forma conjunta todos los coeficientes de la representación, utilizando métodos de máxima verosimilitud.

Entre los incluidos en esta segunda clase se pueden distinguir: **i)** los **métodos paramétricos**, que requieren información sobre toda la estructura de correlaciones (de corto y de largo plazo) o sobre la densidad espectral para todas las frecuencias, lo cual permite caracterizar completamente el comportamiento de la función de autocorrelaciones (no solamente la parte que corresponde a su decrecimiento hiperbólico); **ii)** los **métodos semiparamétricos**, en los que no se especifica un modelo completo de la estructura de autocorrelaciones (o del espectro), sino su comportamiento para valores grandes de j (o frecuencias bajas del espectro).

Sea un proceso $\{Y_t\}$ que admite una representación $ARFIMA(0,d,0)$ de la forma $(1 - B)^d Y_t = \varepsilon_t$ (donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0,\sigma_\varepsilon^2\right)$) se demuestra que un estimador paramétrico consistente y asintóticamente Normal del vector $\Theta = \left[\sigma_\varepsilon^2 \quad d\right]$ se obtiene de

---

[18]. Ver Chung, C.F. (1994).

[19]. Resumiendo las propiedades mencionadas en páginas anteriores se puede concluir que una condición necesaria para que se verifique que $\lim_{\alpha \to 0}\omega(\alpha) = \infty$, es que la sucesión $\{\psi_j\}$ sea divergente.

maximizar la función de verosimilitud exacta[20]:

$$L(\Theta,Y) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\left|\Gamma(\Theta)\right|\right) - \frac{1}{2}Y'\left(\Gamma(\Theta)\right)^{-1}Y$$

donde: **i)** Y denota el vector de las observaciones; **ii)** $\Gamma(\theta)$ denota la matriz de varianzas y autocovarianzas de $\{Y_t\}$ (que puede ser calculada utilizando el algoritmo de Sowell, F. (1992)) y **iii)** $|\Gamma(\theta)|$ denota el determinante de dicha matriz.

Generalizando estos resultados se demuestra que, dado un proceso $\{Y_t\} : ARFIMA(p,d,q)$, siendo $0 < d < \frac{1}{2}$, el estimador máximo verosímil del vector Θ también es consistente, asintóticamente Normal y asintóticamente eficiente[21].

A fin de evitar los problemas de cálculo que trae aparejada la maximización de la función de verosimilitud exacta (debe tenerse en cuenta que este método requiere la utilización de métodos numéricos que incluyen el cálculo de la inversa de la matriz $\Gamma(\theta)$ en cada iteración y que, si $d = \frac{1}{2}$, dicha matriz es quasi-singular y, por lo tanto, su inversa puede ser muy inestable ante variaciones pequeñas en las covarianzas), Fox, R.; Taqqu, M.S. (1986) propusieron los siguientes **estimadores máximo-verosímiles aproximados de Whittle**[22] que se obtienen de maximizar la **función de Whittle**[23]:

$$L_W(\Theta,Y) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\ln\left[\omega(\alpha,\Theta)\right] + \frac{I_n^{(p)}(\alpha)}{\omega(\alpha,\Theta)}\right\}d\alpha$$

donde: **i)** $w(\alpha,\theta)$ es la función de densidad espectral y **ii)** $I_n^{(p)}(\alpha)$ denota el período grama, definido de la siguiente forma:

$$I_n^{(p)}(\alpha) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)e^{it\alpha}\right|^2 .$$

En particular, si $\{Y_t\}: I(0) - N\left(m(Y),\sigma_Y^2\right)$, con función de densidad espectral representable de la siguiente forma: $w(\alpha,\theta) = \theta^* f\left(\alpha,\theta^{**}\right)$ (donde $\theta = \left[\theta^* \quad \theta^{**}\right]$ y $\int_{-\pi}^{\pi}\ln\left[f\left(\alpha,\theta^{**}\right)\right]d\alpha = 0$), se demuestra que los estimadores máximo-verosímiles de Whittle de los coeficientes θ^* y θ^{**} son consistentes y que el estimador máximo-verosímil de Whittle del coeficiente θ^{**} es, además, asintóticamente Normal (ver

---

20. Ver Yajima, Y. (1985).

21. Ver Dahlhauss, R. (1989).

22. Ver Whittle, P. (1962).

23. El subíndice W en la función $L_W(\bullet)$ hace referencia a P. Whittle.

Fox, R.; Taqqu, M.S. (1986))[24].

Asimismo, si $\{Y_t\}: I(0) - N(m(Y), \sigma_Y^2)$ con función de densidad espectral arbitraria, se demuestra que el estimador de Whittle del vector θ es consistente, asintóticamente Normal y, bajo ciertas condiciones, asintóticamente eficiente[25].

Según se vio, los procesos $\{Y_t\}: ARFIMA(p,d,q)$ cumplen las condiciones mencionadas precedentemente, por lo tanto se puede concluir que el estimador máximo-verosímil de Whittle de d es consistente, asintóticamente Normal y asintóticamente eficiente.

Si bien, como se mencionó en la página precedente, estos estimadores de Whittle permiten evitar el cálculo de la inversa de la matriz de varianzas y covarianzas de Y_t en cada iteración de los métodos numéricos necesarios en la maximización de la función de verosimilitud exacta mediante su sustitución por transformadas de Fourier del vector Y, la integral que figura en la definición de la función $L_W(\Theta, Y)$ anterior debe resolverse por métodos numéricos y esta solución debe repetirse en cada iteración. A fin de solucionar este inconveniente se suele sustituir dicha integral por una suma finita de funciones de Fourier de la forma:

$$L_W^{(D)}(\Theta, Y) = \frac{1}{2\pi} \sum_{j=1}^{n-1} \left\{ \ln\left[f(\alpha_j, \Theta) \right] + \frac{I_n^{(p)}(\alpha_j)}{f(\alpha_j, \Theta)} \right\}$$

(donde $\alpha_j = \frac{2\pi j}{n}$ denota la j-ésima frecuencia de Fourier). De la maximización de esta función se obtiene la **versión discreta del estimador máximo-verosímil de Whittle en el dominio de las variables** ($\Theta_W^{(D)}$), que posee las mismas propiedades asintóticas que el estimador obtenido de la expresión original de Fox-Taqqu. En particular, si la función de densidad espectral de $\{Y_t\}$ -igual que en el caso anterior- admite una representación de la forma $w(\alpha, \theta) = \theta^* f(\alpha, \theta^{**})$, donde $\theta = \left[\theta^* \quad \theta^{**}\right]$, se logra una simplificación metodológica importante, ya que de la minimización de la función:

$$Q(\Theta^{**}, Y) = \sum_{j=1}^{n-1} \frac{I_n^{(p)}(\alpha_j)}{f(\alpha_j, \Theta)}$$

se obtiene el estimador de θ^{**} y el estimador de θ^* se obtiene a partir de la expresión:

$$\hat{\Theta}^* = \frac{1}{n} \sum_{j=1}^{n-1} \frac{I_n(\alpha_j)}{f(\alpha_j, \hat{\Theta}^*)}$$

Por otra parte, teniendo en cuenta que, si $\{Y_t\}: ARFIMA(0,d,0)$, su función de densidad

---

[24]. Si $\{Y_t\}$ es un proceso de memoria larga no necesariamente Normal, que admite una representación lineal y cuya función de densidad espectral admite una representación de la forma mencionada, entonces es posible desarrollar una versión del teorema central del límite para el estimador de Whittle del coeficiente Θ^* (ver Giriatis, L.; Surgalis, D. (1990)).

[25]. Ver Dahlhaus, R. (1989).

espectral admite una representación de la forma $w(\alpha,\theta) = \theta^* f(\alpha,d)$ (donde $\theta^* = \frac{\sigma_\varepsilon^2}{2\pi}$), se puede concluir que $\hat{\sigma}_\varepsilon^2 = 2\pi\hat{\theta}^*$. Asimismo, suponiendo un conjunto de información finito, si $m(Y)$ es conocido el estimador máximo-verosímil exacto del vector θ es más eficiente, pero si $m(Y)$ es desconocido, es preferible el estimador $\theta_W^{(D)}$ (especialmente en el caso de conjuntos de información suficientemente grandes)[26].

El problema de los estimadores paramétricos radica en que, ante una especificación equivocada de la función de densidad espectral, se vuelven sesgados e inconsistentes[27].

Como se vio en la Sec. 3.2.1, la función de densidad espectral de un proceso $\{Y_t\}:I(d)$ puede ser expresada de la siguiente forma:

$$\omega_Y(\alpha) = \left|1 - e^{-i\alpha}\right|^{-2d} \omega_Z(\alpha) = \left[4\,sen^2\left(\frac{\alpha}{2}\right)\right]^{-d} \omega_Z(\alpha)$$

(donde $Z_t = \nabla^d Y_t = (1-B)^d Y_t$). Luego, se puede escribir:

$$\log\left[I_n^{(p)}(\alpha_j)\right] \approx \log\left[\omega_Z(\alpha)\right] - d\log\left[4\,sen^2\frac{\alpha}{2}\right]$$

(donde $w_Z(\alpha) = \frac{\sigma_\varepsilon^2}{2\pi}\frac{\left|\Phi\left(e^{-i\alpha}\right)\right|^2}{\left|\Theta\left(e^{-i\alpha}\right)\right|^2}$ denota la función de densidad espectral de un proceso estacionario de segundo orden $\{Z_t\} = \{\nabla^d Y_t\} = (1-B)^d Y_t$ que admite una representación $ARMA(p,q)$).

Luego, a partir de los valores de $w_Y(\alpha)$ para las frecuencias de Fourier y aplicando m. a m. el operador logaritmo, se obtiene la siguiente expresión aproximada debida a Geweke-Porter Hudak[28]:

$$\log\left[I_n^{(p)}(\alpha_j)\right] \approx \left\{\log\left[\omega_Z(0) - \gamma\right] - d\log\left[4\,sen^2\frac{\alpha_j}{2}\right]\right\} + \varepsilon_j^*$$

(donde: **i)** $\alpha_j = \frac{2\pi j}{n}$ denota la j-ésima frecuencia de Fourier; **ii)** $I_n(\alpha_j)$ denota la j-ésima frecuencia del periodograma, **iii)** γ es la constante de Euler; **iv)** las perturbaciones $\varepsilon_j^* = \ln\frac{I_n(\alpha_j)}{w(\alpha_j)}$ son

---

[26]. Ver Cheung, Y.W.;Diebold, F.X. (1994).

[27]. Ver Robinson, P.M. (1995).

[28]. Ver Geweke, J.; Porter-Hudak, S. (1983).

asintóticamente iid con valor esperado nulo y varianza $\frac{\pi^2}{6}$ y v) la tasa de convergencia en-distribución, $k = g(n)$,es un entero menor que n y la función $g(\bullet)$ tiende a infinito a una tasa menor que n) .

Esta aproximación permite estimar el coeficiente d por cuadrados mínimos ordinarios, como la pendiente de un modelo lineal en el dominio espectral[29].

Para $d < 0$, este estimador de Geweke-Porter Hudak (**GPH)** es consistente y asintóticamente Normal. Dado que el coeficiente de regresión de este modelo es $-d$, se podría testear si un proceso es de memoria larga contrastando la hipótesis nula H_0:$d = 0$ (H_0:*el proceso es de memoria corta*) con respecto a la alternativa H_1:$d \neq 0$, mediante un test "t" de significatividad individual. Pero debe tenerse en cuenta que, si $\{Z_t\}$ admite una representación $AR(1)$ o $MA(1)$ con φ_1 o θ_1 cercano a la zona de no-estacionariedad, el estimador GPH de d , en muestras finitas, es muy sesgado e ineficiente (ver Agiakloglou, C.; Newbold, P.; Wohar, M. (1993)), y el test presenta un sesgo importante hacia el rechazo de H_0 . Una conclusión similar se obtiene utilizando el método por simulación de Cheung, Y.W. (1993): cuando el proceso $\{Z_t\}$ admite una representación $AR(1)$ con un valor de ϕ_1 próximo a la zona de no-estacionariedad, el test presenta un sesgo hacia la aceptación de hipótesis alternativas del tipo H_1:$d > 0$.

Por su parte, dado un proceso $\{Y_t\}$:$N\left(m(Y),\sigma_Y^2\right)$ con función de densidad espectral tal que, para las bajas frecuencias, se verifica que $\lim_{\alpha\to 0} w_Y(\alpha) \approx c|\alpha|^{-2d}$ sin especificar, como en el caso anterior, la forma de la función $w_Y(\alpha)$, para $-\frac{1}{2} < d < \frac{1}{2}$, se puede obtener un estimador $\hat{d}$ consistente, asintóticamente Normal (con una tasa de convergencia en-distribución igual a $\sqrt{k}$, menor que $\sqrt{n}$, que es la tasa de convergencia típica de los estimadores Normales)[30] y más eficiente que el estimador GPH, a partir de un modelo de la forma[31]:

$$\log\left[I_n(\alpha_j)\right] \approx \log(c) - 2d\log\left(\left|\alpha_j\right|\right) + \log\left(\frac{I_n(\alpha_j)}{\omega(\alpha_j)}\right) \qquad (j = 1,2,...,k)$$

Alternativamente a este estimador obtenido de aplicar cuadrados mínimos ordinarios en el dominio espectral, Robinson, P.M. (1995) propuso, para procesos cuya función de densidad espectral, igual que en el caso precedente, para las bajas frecuencias es tal que $\lim_{\alpha\to 0} w_Y(\alpha) \approx c|\alpha|^{-2d}$, un estimador Normal semiparamétrico -más eficiente que el anterior, que no supone la Normalidad de Y_t y que permite una forma objetiva del número k de ordenadas a promediar- que se obtiene de maximizar una versión de la aproximación discreta de Whittle a la ya mencionada función de verosimilitud Normal:

---

[29]. Debe tenerse en cuenta que la selección del valor k constituye un problema importante en la aplicación de este método de estimación. Habitualmente se considera $k = \sqrt{n}$.

[30]. Esta condición de Normalidad de $\{Y_t\}$ es sumamente restrictiva (ver Velasco, C. (1998)(1998)).

[31]. Ver Robinson, P.M. (1995).

$$L_W^{(D)}(\Theta, Y) = \frac{1}{2n}\sum_{j=1}^{n-1}\left\{\log\left[f(\alpha_j,\Theta)\right] + \frac{I_n(\alpha_j)}{f(\alpha_j,\Theta)}\right\}$$

con la suma solamente en el entorno de la frecuencia cero. Se demuestra (ver Robinson, P.M. (1995)) que, para $-\frac{1}{2} < d < \frac{1}{2}$, este estimador es consistente y asintóticamente Normal (con una tasa de convergencia en-distribución inferior a $\sqrt{n}$). Velasco, C. (1999) generalizó estos resultados al caso no-estacionario ($d \geq \frac{1}{2}$).

5.4.- El coeficiente de Hurst

Teniendo en cuenta que detectar la presencia y, eventualmente, medir la memoria de largo plazo de un proceso a partir del estimador de la función de autocorrelaciones es sumamente difícil[32], se suele utilizar un método alternativo basado en el **rango sobre el desvío estándar** -también conocido como **rango reescalado** (o **estandarizado**) o **coeficiente R/S-** el cual se funda en una propiedad del movimiento Browniano, según la cual la distancia recorrida por el proceso es igual a la raíz cuadrada del tiempo de medición (llamada **ley de T a la ½**), $R = \sqrt{T}$ (ver Sec. 4.2).

Hurst, H.E. (1951) propuso la siguiente generalización de esta fórmula:

$$(R/S)_n = (an)^H$$

donde H denota el ya mencionado coeficiente de Hurst, a es una constante y los valores del estadístico $(R/S)_n$ -definido como el rango de sumas parciales de desvíos de la serie cronológica $\{Y_t\}$ $(t = 1,2,...,n)$ respecto de su valor medio corregido de su efecto escala por su coeficiente de dispersión- se calculan de la siguiente forma:

$$(R/S)_n = \frac{1}{\hat{\sigma}_n^{(MV)}(Y)}\left|\underset{1\le j\le n}{Max}\sum_{t=1}^{j}\left(Y_t - \overline{Y}_{(n)}\right) - \min_{1\le j\le n}\sum_{t=1}^{j}\left(Y_t - \overline{Y}_{(n)}\right)\right| =$$

$$= \underset{1\le j\le n}{Max}\sum_{t=1}^{j}u_t - \min_{1\le j\le n}\sum_{t=1}^{j}u_t = \underset{1\le j\le n}{Max}\left(u_{jt}^*\right) - \min_{1\le j\le n}\left(u_{jt}^*\right) \qquad (j = 1,2,...,n)$$

(donde: **i)** $\overline{Y}_{(n)} = \frac{1}{n}\sum_{t=1}^{N} Y_t$; **ii)** $\hat{\sigma}_n^{(MV)2}(Y) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \overline{Y}_{(n)}\right)^2$ denota el **factor de Normalización**, definido como el estimador máximo verosímil de la varianza de Y_t y **iii)** $u_j^* = \sum_{t=1}^{j} u_t$, siendo u_t una variable estandarizada)[33].

---

[32]. Ver Newbold, P.; Agiakloglou, Ch. (1993).

[33]. Que los valores de $(R/S)_n$ sean estandarizados obviamente facilita la comparación de fenómenos referidos a períodos diferentes.

Como la suma de las n variables u_t $\left(u_{nt}^*\right)$ es nula, entonces se verificará que: **i)** $Max\left(u_{jt}^*\right)$ $\left(j = 1,2,...,n\right)$, que representa el máximo de las sumas parciales de los j primeros desvíos de Y_t con respecto a $\overline{Y}_{(n)}$, será siempre mayor o igual que cero y **ii)** $\min\left(u_{jt}^*\right)\left(j = 1,2,...,n\right)$ será siempre menor o igual que cero, de modo que $\left(R/S\right)_n \geq 0$.

Se demuestra que $\underset{n\to\infty}{p\lim}\left[\frac{1}{n^H}\left(R/S\right)_n\right] = cte.$ (ver Hurst, H.E. (1951), Mandelbrot, B.B.; Wallis, J. (1968), Mandelbrot, B.B.; Taqqu, M.S. (1979), Taqqu, M.S. (1975)(1976), Lo, A.W. (1991)). De modo que el coeficiente H de Hurst puede ser estimado a partir de la expresión $\frac{\log\left(R/S\right)_n}{\log(n)}$. Un valor de $H = \frac{1}{2}$ indica que $\left\{Y_t\right\}$ puede ser considerado como de memoria corta, si $H > \frac{1}{2}$, se puede asegurar que el proceso es **persistente** lo cual implica la presencia de memoria larga, si se verifica que $0 \leq H < \frac{1}{2}$, se dice que el proceso es **anti-persistente**, circunstancia que indica la presencia de un mayor número de cambios de curvatura que en el "random walk", lo cual implica una autocorrelación negativa de los incrementos: una realización con tendencia creciente tiene más probabilidad de estar seguida por una realización con tendencia decreciente que por una con tendencia decreciente ("**pink noise**")[34].

5.5.- Los tests para contrastar la hipótesis ARFIMA

Una forma alternativa de estimación del coeficiente H se obtiene de considerarlo como pendiente del modelo de regresión lineal con $\log\left[\left(R/S\right)_n\right]$ y $\log(t)$ como variables dependiente e independiente, respectivamente. Con este fin se divide la serie cronológica original $\left\{Y_t\ ; t = 1,2,...,n\right\}$ en k períodos (denotados por $\tau_1,\tau_2,...,\tau_k$) de longitud $n_k \geq 10$ (se demuestra que valores más pequeños de n producen estimaciones inestables), tal que $kn_k = n$ (es decir que, n_k deber ser tal que $\frac{n}{n_k}$ sea un número

---

[34]. Como se mencionó en la Sec. 4,4, el coeficiente H fue definido para minimizar los costos de construcción de reservorios en una represa, que permitieran mantener un comportamiento relativamente constante en el caudal del río Nilo. De acuerdo con la nomenclatura empleada en el texto, Y_t denota la medida del caudal de agua en el momento t, antes de la represa; $\overline{Y}_{(n)}$ denota el caudal deseado después de la represa y $\left(R/S\right)_n$ representa la capacidad del reservorio que debería asegurar el caudal deseado, dado un cierto comportamiento del río para $t = 1,2,...,n$. Sea, por ejemplo la siguiente serie cronológica correspondiente a 10 mediciones sobre el caudal del río para $t = 1,2,...,10$, respectivamente: $\left\{Y_t^{(1)}\right\} = \left\{10,8,7,3,5,9,10,6,4,2\right\}$ y sea $\overline{Y}_{(10)}^{(1)} = 6.4$ el caudal constante deseado. La capacidad del reservorio debería ser: $\left(R/S\right)_{10}^{(1)} = 7.2 - 0.0 = 7.2$. Sea, ahora, otra serie cronológica correspondiente también a 10 mediciones sobre el caudal del río: $\left\{Y_t^{(2)}\right\} = \left\{10,10,9,8,7,6,5,4,3,2\right\}$. En este caso la capacidad del reservorio debería ser: $\left(R/S\right)_{10}^{(2)} = 12.0 - 0.0 = 12.0$. Obsérvese: **i)** que el estimador de $\sigma_{10}^2(Y)$ es común a ambas series e igual a 7.44, razon por la cual no se utilizaron variables estandarizadas; y **ii)** que, como se verá inmediatamente, el aumento de la persistencia en comportamiento de $\left\{Y_t\right\}$ produce un aumento del coeficiente $\left(R/S\right)$.

entero), quedando definidas k sub-series del tipo $\left\{Y_{1,\tau_i}, Y_{2,\tau_i}, ..., Y_{n_k,\tau_i}\right\}\left(i = 1,2,...,k\right)$. A partir de estas sub-series, para cada sub-período τ_i $\left(i = 1,2,...,k\right)$, se calcula: **i)** el promedio $\overline{Y}_{\tau_i} = \frac{1}{n_k}\sum_{j=1}^{n_k} Y_{j,\tau_i}$ $\left(i = 1,2,...,k\right)$

ii) la serie cronológica de los desvíos acumulados $Y^*_{j,\tau_i} = \sum_{h=1}^{j}\left(Y_{h,\tau_i} - \overline{Y}_{\tau_i}\right)\left(i = 1,2,...,k\right)$ y **iii)** los valores del rango de las sumas parciales de los desvíos de cada una de las sub-series cronológicas con respecto a su valor medio, corregidos del efecto escala por su desvío estándar:

$$\left(R/S\right)_{\tau_i} = \frac{1}{\hat{\sigma}_{\tau_i}}\left[Max\left(Y^*_{j,\tau_i}\right) - \min\left(Y^*_{j,\tau_i}\right)\right] \qquad \left(i = 1,2,...,k; j = 1,2,...,n_k\right)$$

(donde $\sigma^2_{\tau_i} = \quad = \frac{1}{n_k}\sum_{j=1}^{n_k}\left(Y_{j,\tau_i} - \overline{Y}_{\tau_i}\right)^2$).

Igual que en el caso anterior, como la suma de los n_k desvíos ($Y^*_{n_k,\tau_i}$) es igual a cero, entonces $Max\left(Y^*_{j,\tau_i}\right)$ $\left(i = 1,2,...,k; j = 1,2,...,n_k\right)$ será siempre mayor o igual que cero y $\min\left(Y^*_{j,\tau_i}\right)$ será siempre menor o igual que cero, de modo que $\left(R/S\right)_{\tau_i} \geq 0\left(i = 1,2,...,k\right)$.

Dado que el estadístico $\left(R/S\right)_n$ puede presentar deficiencias ante la existencia de dependencia de corto plazo combinada con una componente correspondiente a la presencia de memoria larga[35] (lo cual implica que, en un test con hipótesis nula de no-existencia de memoria larga, la no aceptación de H_0 no implica necesariamente la existencia de dependencia de largo plazo, ya que puede deberse a la existencia de una dependencia de corto plazo)[36]. A fin de resolver esta restricción, Haubrich, J.G.; Lo, A.W. (1989) y Lo, A.W. (1991) propusieron una modificación del factor de Normalización que incorpora la dependencia de corto plazo y genera un **rango reescalado modificado**, de la forma[37]:

$$\left(R/S\right)_{q,n} = \frac{1}{\hat{\sigma}_q(Y)}\left[\underset{1\leq j\leq n}{Max}\left(Y^*_{jt}\right) - \underset{1\leq j\leq n}{\min}\left(Y^*_{jt}\right)\right] = \frac{\hat{\sigma}_n(Y)}{\hat{\sigma}_q(Y)}\left(R/S\right)_n$$

Donde $Y^*_{jt} = \sum_{t=1}^{j}\left(Y_t - \overline{Y}\right)\left(j = 1,2,...,n\right)$ y:

---

35. Por ejemplo. si $\left\{Y_t\right\}: AR(1)$, el estimador del valor esperado es sesgado.

36. Como se mencionó en la Sec. 3.5, Cochrane, J.H. (1988) propuso un test alternativo para la no-estacionariedad basado en el cociente de varianzas. Para un análisis de la superioridad del test de rango reescalado sobre el de Cochrane, ver Mandelbrot, B.B. (1972).

37. Este rango reescalado modificado, que es robusto respecto de muchas formas de heterocedasticidad y dependencia débil, permite la construcción de un test, con hipótesis nula de existencia de memoria corta consistente con respecto a alternativas $I(d)$.

$$\hat{\sigma}_q^2(Y) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \overline{Y}_{(n)}\right)^2 + \frac{2}{n}\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\sum_{i=q+1}^{n}\left(Y_i - \overline{Y}_{(n)}\right)\left(Y_{i-j} - \overline{Y}_{(n)}\right) =$$

$$= \hat{\sigma}_n^2(Y) + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_j(Y) =$$

$$= \hat{\sigma}_n^2(Y)\left[1 + \frac{2}{n}\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\rho}_j(Y)\right] \qquad (q < n)$$

(obsérvese que este rango reescalado utiliza una variable residual Y_{jt}^* que considera exclusivamente la presencia de término independiente en la ecuación de la tendencia determinística)[38].

Si $\{Y_t\}$ es un proceso de memoria corta, el estimador de $\sigma_q^2(Y)$ incluirá en su definición, no sólo la suma de las varianzas de los términos individuales, sino las correspondientes autocorrelaciones hasta el "lag" q, ponderadas. Dichas ponderaciones $\left(1 - \frac{j}{q+1}\right)$ (correspondientes a la llamada **ventana de Bartlett**) aseguran que el estimador (consistente) $\hat{\sigma}_q^2(Y)$ sea positivo y que, para $q = 0$, $(R/S)_{q,n} = (R/S)_n$ (ver Newey, W.; West, K. (1987)).

Bajo el supuesto que $\{Y_t\}$: $RW - 1$, el estimador $n^{-1/2}(R/S)_{q,n}$ converge a una variable V representativa del rango de un puente Browniano estandarizado sobre un intervalo unitario[39], cuya función de distribución de probabilidades es de la forma:

$$F_V(v) = \sum_{j=1}^{\infty}\left(1 - 4v^2j^2\right)e^{-2v^2j^2}$$

con valor esperado $\sqrt{\frac{\pi}{2}} = 1.253$ y varianza $\sigma^2(V) = \frac{n(n-3)}{6} = 0.07398$ (ver Kennedy, D. (1976), Siddiqui, M. (1976)). En el **CUADRO 5.1** figuran valores de los cuantiles de la variable V. Obsérvese: **i)** que la función de densidad de V tiene asimetría positiva y **ii)** que la mayor concentración de probabilidades se verifica en el intervalo [0.75;2][40].

Bajo el supuesto que $\{Y_t\}$: $I(0) - AR(1)$, el estimador de $(R/S)_{q,n}$ estandarizado converge a la variable $\sqrt{\frac{1+\phi_1}{1-\phi_1}}$.

---

38. Una generalización inmediata se obtiene de considerar $Y_{jt}^* = \sum_{t=1}^{j}\left(Y_t - \hat{a}_0 - \hat{a}_1 t\right)$.

39. Esta distribución límite es invariante con respecto a muchas formas de dependencia de corto plazo, pero sensible a la presencia de dependencia de largo plazo (ver Ross, S.M. (1983), Lo, A.W. (1991)).

40. Ver Lo, A.W. (1991).

CUADRO 5.1: Cuantiles correspondientes a la variable V

$F_V(v)$	v
0,005	0,721
0,025	0,809
0,050	0,861
0,100	0,927
0,200	1,018
0,300	1,090
0,400	1,157
0,500	1,223
0,543	$\sqrt{\pi}/2$
0,600	1,294
0,700	1,374
0,800	1,473
0,900	1,620
0,950	1,747
0,975	1,862
0,995	2,098

La dificultad mayor en la construcción de este estadístico consiste en la determinación de d y en la distinción entre dependencias de corto y largo plazo. De acuerdo a lo propuesto por Andrews, D. (1991), Lo, A.W. (1991), $q = [n_k]$ (donde $[\bullet]$ denota el mayor entero menor o igual que el argumento) y $n_k = \left(\frac{3n}{2}\right)^{1/3}\left[\frac{2\hat{\rho}_1(Y)}{1-\hat{\rho}_1^2(Y)}\right]^{2/3}$ denota una aproximación a la medida de la dependencia entre las variables Y_t. Debe tenerse en cuenta que: **i)** el valor de q obtenido por este método es óptimo sólo en el caso en que $\{Y_t\}: AR(1)$; **ii)** de acuerdo con este procedimiento, las ponderaciones $1-\frac{j}{q+1}$ propuestas por Newey-West deberían ser reemplazadas por $1-\left|\frac{j}{n}\right|$, propuestas por Andrews.

Por otra parte, si las variables Y_t son "fat-tailed", entonces la distribución de probabilidades del estimador $(R/S)_{q,n}$ presenta un desplazamiento hacia la izquierda con respecto a la distribución asintótica $F_V(v)$. Esto implica que la probabilidad de rechazo de la hipótesis nula de no existencia de memoria larga, con respecto a la alternativa de que el proceso es persistente ($d < 0$), es decir, la probabilidad de rechazo de H_0 por la cola de la izquierda, es mayor que la indicada por la distribución asintótica. Por el contrario, la probabilidad de rechazo de la hipótesis nula con respecto a la alternativa de que el proceso es persistente de memoria larga ($d > 0$), es decir, la probabilidad de rechazo de H_0 por la cola de la derecha, es menor que la indicada por la distribución asintótica[41].

---

[41]. Ver Hiemstra, C.; Jones, J.D. (1997).

Robinson, P.M. (1991) propuso un test LM para contrastar la hipótesis nula $H_0:\{Y_t\} - WN$ con respecto a una familia muy general de alternativas que incluye a los procesos $ARFIMA(0,d,0)$, utilizando un estadístico de la forma:

$$\Lambda = \frac{6n}{\pi^2}\left(\sum_{j=1}^{n-1}\frac{\hat{\rho}_j(Y)}{j}\right)^2$$

el cual, bajo el supuesto que H_0 es verdadera, converge en-distribución a una función $\chi^2_{(1)}$.

Este test permite contrastar la hipótesis de WN: **i)** con respecto a la **alternativa de memoria larga** ($H_1:d > 0$), rechazando H_0 cuando el estadístico Λ es significativamente positivo con relación al límite de la zona crítica que proporciona la distribución $N(0;1)$; **ii)** con respecto a la **alternativa de sobrediferenciación** ($H_1:d < 0$), rechazando H_0 cuando el estadístico Λ es significativamente negativo con relación al límite de la zona crítica que proporciona la distribución $N(0;1)$; **iii)** con respecto a la alternativa $H_1:d \neq 0$, rechazando H_0 cuando el estadístico Λ toma valores significativamente grandes con relación al límite de la zona crítica que proporciona la distribución $\chi^2_{(1)}$ (en Robinson, P.M. (1991) se puede hallar una variante de este test robusta con respecto a la heterocedasticidad condicionada de $\{\varepsilon_t\} = Y_t - \delta_0 X_t$ y, en Sims, C.A. (en Robinson, P.M. (1994)), figura otra variante de este test robusta con respecto a la presencia de autocorrelación débil en H_0).

El test de Dickey-Fuller aumentado y los tests no-paramétricos con $H_0: d = 1$, analizados en la Sec. 3.3, son consistentes contra hipótesis alternativas con d fraccionario (ver Diebold, F.X.; Rudebusch, G.D. (1991), Hassler, U.; Wolters, J. 1(994))[42].

Wu, P. (1992) propuso otro test (basado en el test de King, M.L.; Hillier, G. (1985)) para contrastar la hipótesis nula $H_0:d = 0$ con respecto a las alternativas $d > 0$ ó $d < 0$ en un FWN utilizando, respectivamente, los estadísticos $\Lambda^* = -2\sum_{j=1}^{n-1}\frac{\hat{\rho}_j(Y)}{j}$ y $-\Lambda^*$ (para una generalización de este test para procesos ARFIMA, ver Wu, P. (1992), Agiakloglou, Ch.; Newbold, P. (1994)).

El **test de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)** considera la hipótesis nula de estacionariedad, $H_0:d = 0$ con respecto a alternativas de procesos integrados de memoria corta en las primeras diferencias. Este test (que es consistente con respecto a hipótesis alternativas de d fraccionarios en el intervalo $-\frac{1}{2} < d < \frac{1}{2}$ y $d \neq 0$) considera la hipótesis nula de presencia de tendencia determinística, es decir, la hipótesis nula que el proceso representativo de las perturbaciones de un fenómeno dinámico con respecto a una tendencia determinística, es de memoria corta[43].

---

[42]. Bajo el supuesto que la hipótesis nula $H_0:\{Y_t\}:ARFIMA(p,d,q)$ no-estacionario con $\frac{1}{2} < d < \frac{3}{2}$ es verdadera, Sowell, F. (1990) obtuvo una distribución asintótica del estadístico de Dickey-Fuller cuando el proceso definido por las primeras diferencias es $I(d)$ y demostró que sólo posee una distribución límite bien definida cuando $d = 1$ (para las críticas a esta solución, ver Diebold, F.X.; Rudebush, G.D. (1991)).

[43]. Ver Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. (1992).

Sean:

1.- Un proceso $\{Y_t\}$ cuya estructura temporal admite una representación compuesta por la suma de una tendencia determinística y un RW, de la forma:

$$S(Y_t) = a_0 + a_1 t + Z_t$$

donde: **i)** $Z_t = S(Y_t) - a_0 - a_1 t$, que denota el proceso definido por las perturbaciones de $\{Y_t\}$ con respecto a la tendencia determinística lineal, se supone que admite una representación de la forma $Z_t = Z_t^* + e_t$, siendo $\{Z_t\}$ un proceso RW-1 de la forma $Z_t^* = Z_{t-1}^* + v_t \ (Z_0^* = 0)$ (donde $\{v_t\}$:*WN - estricto* (en particular, si se supone que: **i)** $\{Z_t^*\}$:*RW - Normal* y **ii)** $\{e_t\}$:$WN - N(0,\sigma_e^2)$, este test constituye una continuación del test de Nabeya, S.; Tanaka, K. (1988)) y **ii)** $\{e_t\}$ es un proceso de memoria corta que satisface las condiciones de Lee-Schmidt[44] (obsérvese que, de acuerdo a lo mencionado más arriba, el test KPSS considera la hipótesis nula de estacionariedad $H_0:\sigma_v^2$, la cual implica la hipótesis que $Z_t = e_t$ es un proceso de memoria corta).

2.- El proceso $\hat{\varepsilon}_t^{(TD)} = Y_t - \hat{a}_0 - \hat{a}_1 t$, formado por los residuos del proceso $\{Y_t\}$ a partir de la aplicación del modelo lineal representativo de la tendencia determinística.

3.- Las sumas parciales del proceso residual definido en el punto precedente, $S_t^{(TD)} = \sum_{j=1}^{t} \hat{\varepsilon}_j^{(TD)} \ (t = 1,2,...,n)$.

4.- El estimador de Newey, W.K.; West, K.D. (1987) de la varianza de largo plazo de las perturbaciones $\varepsilon^{(TD)}$:

$$\hat{\sigma}^2_{\varepsilon^{(TD)}}(q) = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{\varepsilon}_t^{(TD)}\right)^2 + \frac{2}{n}\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\sum_{i=j+1}^{n}\hat{\varepsilon}_i^{(TD)}\hat{\varepsilon}_{i-j}^{(TD)} =$$

$$= \hat{\gamma}_0\left(\varepsilon^{(TD)}\right) + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_j\left(\varepsilon^{(TD)}\right) =$$

$$= \hat{\gamma}_0\left(\varepsilon^{(TD)}\right)\left[1 + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\rho}_j\left(\varepsilon^{(TD)}\right)\right] \qquad (q < n)$$

(donde, como se mencionó en la sección precedente, el término $\left[1 - \frac{j}{q-1}\right]$ debe ser tal que garantice que el estimador sea no-negativo y q puede ser estimado como el mayor entero menor o igual a

---

[44]. Ver Sec. 5.2.1.

$n_k = \left(\frac{3n}{2}\right)^{1/2}\left[\frac{2\rho_1\left(\hat{\varepsilon}^{TD}\right)}{1-\rho_1^2\left(\hat{\varepsilon}^{TD}\right)}\right]^{2/3}$). Debe tenerse en cuenta que, bajo el supuesto de que la hipótesis nula de estacionariedad $H_0{:}\sigma_v^2 = 0$ es verdadera (es decir, bajo el supuesto de que no existen razones suficientes para rechazar la hipótesis $H_0{:}el\ proceso\ \{Z_t\} = \{e_t\}\ es\ de\ memoria\ corta$), la condición necesaria para que $\left(\sigma_e^{(TD)}\right)^2$ sea un estimador consistente de $\left(\sigma_e^{LP}\right)^2 = \lim_{n\to\infty} E\left(e_t^2\right)$, que denota la varianza de largo plazo de las perturbaciones e_t, es que $q \to \infty$ y $n \to \infty$ (ver Andrews, D.W.K. (1991)).

El estadístico propuesto por el test KPSS para contrastar la hipótesis nula de estacionariedad, suponiendo una tendencia lineal de la forma $a_0 + a_1 t$, es:

$$\eta_\tau = \frac{1}{n^2\hat{\sigma}_\varepsilon^2(q)}\sum_{t=1}^{n}\left(S_t^{TD}\right)^2$$

Si $a_1 = 0$, el estadístico propuesto es η_μ en el cual $\hat{\varepsilon}_t^{TD} = Y_t - \overline{Y}$ (obsérvese que, en este caso, el estadístico η_μ es similar al rango reescalado modificado de Haubrich-Lo mencionado en páginas anteriores).

Bajo el supuesto que la hipótesis nula $H_0{:}el\ proceso\{Z_t\} = \{e_t\}\ es\ de\ memoria\ corta$ es verdadera, se demuestra que:

$$\frac{1}{n^2}\sum_{t=1}^{n}\left(S_t^{TD}\right)^2 \xrightarrow{\ d\ } \left(\sigma_e^{LP}\right)^2\int_0^1 V_2^2(x)\,dx$$

(donde $V_2(x) = B(x) - x(3x-2)B(1) - 6x(x-1)\int_0^1 B(x)\,dx$ denota un puente Browniano de segundo nivel de la forma)[45]. En consecuencia, se verificará que:

$$\eta_\tau \xrightarrow{\ d\ } \int_0^1 V_1(x)\,dx = \int_0^1\left[B(x) - xB(1)\right]^2 dx$$

(donde $V_1(x)$ denota un puente Browniano estacionario).

Si se supone que la hipótesis alternativa a la hipótesis nula $H_0{:}el\ proceso\{Z_t\}\ es\ de\ memoria\ corta$ es $H_1{:}el\ proceso\{\nabla Z_t\}\ es\ de\ memoria\ corta$, se demuestra que, para los estadísticos η_τ y η_μ, se verifica que[46]:

---

[45]. De acuerdo con a nomenclatura de MacNeill, I. (1978), V_2 es equivalente a B_1 (ver Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. (1992)).

[46]. Ver Park, J.Y.; Phillips, P.C.B. (1988).

$$\frac{1}{n}\eta \xrightarrow{d} \frac{\int_0^1 \left[\int_0^w B^*(x)dx\right]^2 dw}{\int_0^1 \left[B^*(x)\right]^2 dx}$$

(donde $B^*(x)$ denota un movimiento Browniano estandarizado) y que ambos estadísticos son consistentes con respecto a hipótesis alternativas de procesos $I(d)$: **i)** para tests de dos colas, η_τ y η_μ son consistentes con respecto a procesos con $-\frac{1}{2} < d < \frac{1}{2}$ y $d \neq 0$; **ii)** para tests de cola superior, η_μ es consistente con respecto a procesos con $0 < d < \frac{1}{2}$; **iii)** para tests de cola inferior, η_μ es consistente con respecto a procesos con $-\frac{1}{2} < d < 0$ (ver Lee, D.; Schmidt, P. (1996)).

Lee, H.S.; Amsler, C. (1997) generalizaron los resultados anteriores a procesos $ARFIMA(p,d,q)$ no-estacionarios con $\frac{1}{2} < d < 1$ y demostraron que el test KPSS no puede distinguir en forma consistente los procesos fraccionarios de aquellos con raíz unitaria.

Otro tipo de tests sobre la presencia de orden de diferenciación fraccionario se basa en la aplicación al proceso $\{Y_t\}$ de un modelo $ARMA(p,q)$ de la forma $\hat{\Phi}(B)Y_t = \hat{\Theta}(B)\hat{\varepsilon}_t$ y en la construcción de un test LM para contrastar la hipótesis nula H_0:$d = 0$ (es decir, H_0:$\{Y_t\}$:$ARMA(p,q)$) con respecto a la alternativa H_1:$\{Y_t\}$:$ARFIMA(p,d,q)$, a partir de la utilización del estadístico T para decidir acerca de la significatividad del coeficiente δ en la representación:

$$\varepsilon_t = \beta_1 W_{t-1} + \beta_2 W_{t-2} + ... + \beta_p W_{t-p} + \gamma_1 Z_{t-1} + \gamma_2 Z_{t-2} + ... + \gamma_q Z_{t-q} + \delta K_t(m) + \varepsilon_t^*$$

donde $\hat{\Theta}(B)W_t = Y_t$, $\hat{\Theta}(B)Z_t = \hat{\varepsilon}_t$ y $K_t(m) = \sum_{j=1}^{n} j^{-1}\hat{\varepsilon}_{t-j}$ (ver Agiakloglou, C.; Newbold, P. (1994)).

Debe tenerse en cuenta que los estimadores de la función de autocorrelaciones de un procesoF WN son sesgados, por lo que es muy difícil detectar comportamientos que sugieran memoria larga a partir de los estimadores de los $\rho_j(\varepsilon)$ de series no muy largas (ver Newbold, P.; Agiakloglou, C. (1993)).

Ejemplo 5.1:

Sea un proceso $\{Y_t\}$ que representa la rentabilidad diaria de un activo, para cuyo análisis se cuenta con una serie cronológica correspondiente al período [02-01-1995;20-05-2004] y cuya trayectoria figura en el **GRAFICO 5.1**.

En el **GRAFICO 5.2** figuran las observaciones diarias correspondientes a los valores absolutos de dichas rentabilidades. De la observación de este gráfico se puede concluir que los valores absolutos de las rentabilidades grandes tienen mayor probabilidad de ser seguidas por valores absolutos de rentabilidad pequeñas que de rentabilidades grandes.

En los cuadros 5.2 y **5.3** figuran los correlogramas de los procesos $\{Y_t\}$ e $\{|Y_t|\}$. Obsérvese que, con un 95% de confiabilidad, se puede concluir que la función $\rho_j(Y)\,(j = 1,2,...)$ asume valores significativos solamente para $j = 1$ (aunque no lo suficientemente significativo como para proporcionar un grado de predictibilidad aceptable). Contrariamente, con un 95% de confiabilidad, se puede concluir: **i)** que los coeficientes $\rho_j(|Y|)$ son significativos (positivos) para todos los valores de j y **ii)** que decrecen rápidamente para los primeros 35 valores de j y después observan un decrecimiento lento.

Esta forma de comportamiento parecería indicar que la función de autocorrelaciones de $\{|Y_t|\}$ no observa el comportamiento exponencialmente decreciente, típico de un proceso estacionario. En el **GRAFICO 5.3** figuran los estimadores de la función de autocorrelaciones correspondiente al proceso $\{|Y_t|\}$ y la curva obtenida de aplicar a dichos datos el modelo:

$$\hat{\rho}_j^*(Y) = (0.0165)(0.9793)^j$$

(obsérvese que la curva ajustada decrece lentamente). Del análisis de la serie cronológica $\{|Y_t|\}$ $(t = 1,2,...2448)$ se obtuvieron los siguientes resultados: $|\overline{Y}_{2448}| = 0.027631$ y $\sigma^2_{2448} = 0.0007815$, a partir de los cuales se definió el proceso estandarizado $\{u_t\} = \left\{\dfrac{|Y_t| - |\overline{Y}_{(2448)}|}{\hat{\sigma}_{2448}(|Y|)}\right\} = \left\{\dfrac{|Y_t| - 0.027631}{0.027956}\right\}$ y las variables acumuladas $u_t^* = \sum_{t=1}^{j} u_t\,(j = 1,2,...,2448)$, obteniéndose los siguientes resultados: $\underset{1\le j\le 2448}{Max}(u_j^*) = 135.1044$ y $\underset{1\le j\le 2448}{\min}(u_j^*) = -133.5570$ y, en consecuencia, el siguiente coeficiente de rango reescalado:

$$(R/S)_{2448} = \underset{1\le j\le 2448}{Max}(u_j^*) - \underset{1\le j\le 2448}{\min}(u_j^*) = 268.6614$$

Estos resultados permitieron calcular el siguiente estimador del coeficiente de Hurst:

$$H = \frac{\log[(R/S)_{2448}]}{\log(2448)} = 0.7168$$

Que este coeficiente sea mayor que $\frac{1}{2}$ implica que el proceso $\{|Y_t|\}$ es persistente y que, por lo tanto, puede ser considerado de memoria larga.

Teniendo en cuenta que, de acuerdo al correlograma que figura en el **CUADRO 5.3**, $\hat{\rho}_1(|Y|) = 0.311$, se puede concluir que el proceso $\{|Y_t|\}$ admite una representación ARFIMA con orden de diferenciación estimado $\hat{d} = \dfrac{\hat{\rho}_1(|Y|)}{1 + \hat{\rho}_1(|Y|)} = 0.2372$. Dado que $|d| < \frac{1}{2}$, se puede concluir que el proceso $\{|Y|\}$ es estacionario e invertible y que su función de autocorrelaciones decrece en forma

GRAFICO 5.1: Trayectoria del proceso {Yt}

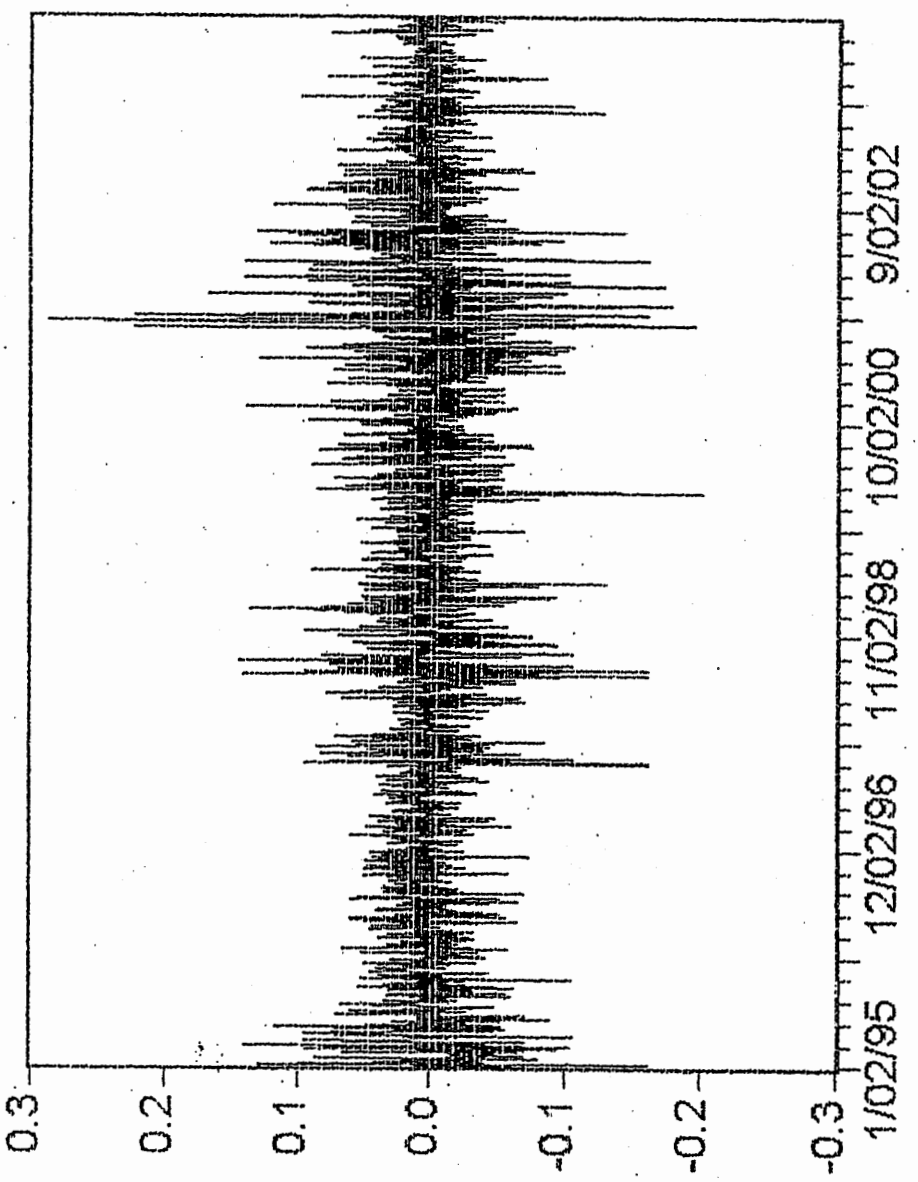
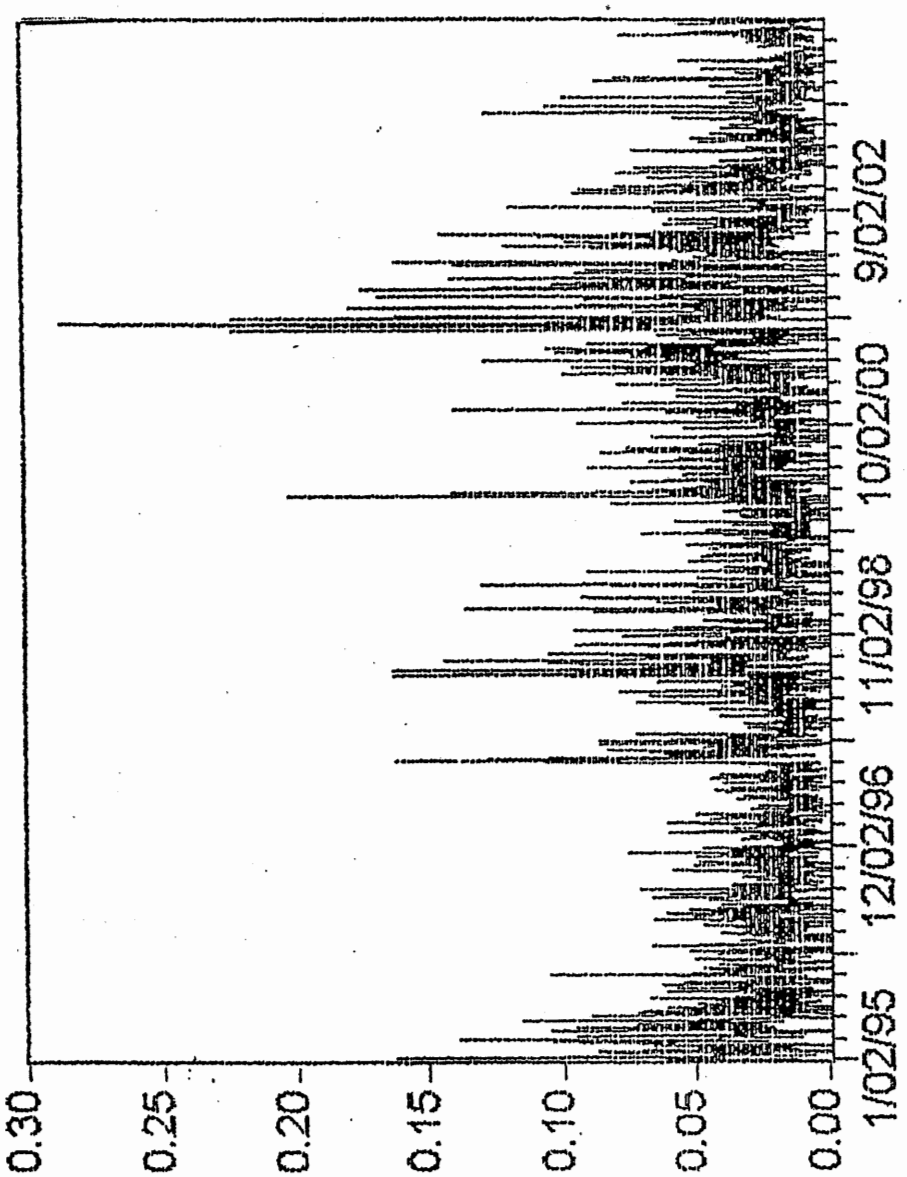


GRAFICO 5.2: Trayectoria del proceso $\{|Y_t|\}$



CUADRO 5.2: Correlograma correspondiente al proceso $\{Y_t\}$

Sample: 1/02/1995 5/20/2004
Included observations: 2449

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.161	0.161	63.695	0.000
		2	0.010	-0.016	63.955	0.000
		3	0.033	0.035	66.621	0.000
		4	0.026	0.016	68.284	0.000
		5	-0.018	-0.024	68.952	0.000
		6	-0.044	-0.039	73.719	0.000
		7	-0.021	-0.010	74.831	0.000
		8	-0.004	0.001	74.865	0.000
		9	0.003	0.007	74.890	0.000
		10	0.067	0.070	86.064	0.000
		11	0.019	-0.004	86.903	0.000
		12	0.032	0.029	89.369	0.000
		13	0.034	0.019	92.166	0.000
		14	0.026	0.015	93.871	0.000
		15	0.034	0.028	96.777	0.000
		16	-0.010	-0.018	97.032	0.000
		17	0.005	0.011	97.095	0.000
		18	0.005	0.002	97.149	0.000
		19	0.001	0.003	97.151	0.000
		20	0.000	-0.002	97.151	0.000
		21	-0.013	-0.011	97.548	0.000
		22	0.034	0.036	100.49	0.000
		23	0.017	0.002	101.19	0.000
		24	-0.003	-0.008	101.22	0.000
		25	0.000	-0.005	101.22	0.000
		26	0.007	0.006	101.35	0.000
		27	0.013	0.013	102.14	0.000
		28	0.013	0.010	102.55	0.000
		29	-0.027	-0.031	104.33	0.000
		30	-0.010	-0.004	104.58	0.000
		31	-0.028	-0.027	106.58	0.000
		32	0.017	0.023	107.30	0.000
		33	0.000	-0.005	107.30	0.000
		34	-0.014	-0.011	107.78	0.000
		35	0.023	0.024	109.04	0.000
		36	-0.017	-0.030	109.73	0.000
		37	-0.048	-0.048	115.53	0.000
		38	-0.081	-0.070	131.70	0.000
		39	-0.007	0.020	131.83	0.000
		40	0.025	0.026	133.39	0.000
		41	0.024	0.028	134.78	0.000
		42	0.012	0.003	135.12	0.000
		43	-0.014	-0.023	135.60	0.000
		44	0.021	0.022	136.75	0.000
		45	0.004	-0.011	136.80	0.000
		46	0.000	0.010	136.80	0.000
		47	-0.011	-0.006	137.10	0.000
		48	0.013	0.030	137.54	0.000
		49	-0.009	-0.015	137.75	0.000
		50	0.012	0.017	138.12	0.000

CUADRO 5.3: Correlograma correspondiente al proceso $\{|Y_t|\}$

Sample: 1/02/1995 5/20/2004
Included observations: 2448

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.311	0.311	236.79	0.000
		2	0.284	0.208	434.90	0.000
		3	0.282	0.170	629.70	0.000
		4	0.237	0.092	767.33	0.000
		5	0.253	0.111	923.80	0.000
		6	0.199	0.034	1021.0	0.000
		7	0.215	0.069	1134.9	0.000
		8	0.232	0.085	1267.4	0.000
		9	0.185	0.021	1351.7	0.000
		10	0.186	0.029	1436.9	0.000
		11	0.165	0.011	1504.0	0.000
		12	0.193	0.057	1596.1	0.000
		13	0.176	0.027	1672.3	0.000
		14	0.192	0.057	1763.6	0.000
		15	0.196	0.047	1857.8	0.000
		16	0.189	0.037	1945.8	0.000
		17	0.151	-0.015	2001.8	0.000
		18	0.157	0.012	2062.8	0.000
		19	0.134	-0.014	2107.1	0.000
		20	0.149	0.018	2161.7	0.000
		21	0.150	0.022	2217.6	0.000
		22	0.143	0.014	2268.5	0.000
		23	0.120	-0.017	2303.9	0.000
		24	0.140	0.023	2352.4	0.000
		25	0.122	0.002	2389.4	0.000
		26	0.103	-0.018	2415.5	0.000
		27	0.140	0.039	2463.9	0.000
		28	0.113	-0.005	2495.6	0.000
		29	0.105	-0.009	2523.2	0.000
		30	0.086	-0.029	2541.4	0.000
		31	0.088	-0.003	2560.6	0.000
		32	0.126	0.042	2600.2	0.000
		33	0.100	0.009	2624.9	0.000
		34	0.109	0.018	2654.3	0.000
		35	0.070	-0.036	2666.7	0.000
		36	0.089	0.003	2686.5	0.000
		37	0.105	0.024	2713.9	0.000
		38	0.108	0.036	2743.0	0.000
		39	0.083	-0.014	2760.2	0.000
		40	0.108	0.027	2789.6	0.000
		41	0.114	0.026	2821.7	0.000
		42	0.067	-0.037	2832.7	0.000
		43	0.101	0.027	2858.2	0.000
		44	0.102	0.025	2884.3	0.000
		45	0.109	0.027	2914.0	0.000
		46	0.099	0.003	2938.3	0.000
		47	0.068	-0.025	2950.0	0.000
		48	0.081	-0.012	2966.3	0.000
		49	0.080	0.002	2982.3	0.000
		50	0.103	0.041	3008.9	0.000

GRAFICO 5.3: Estimador de la función de autocorrelaciones del proceso $\{|Y_t|\}$ y de la función $\rho_j^*=(0.0165)(0.9793)^j$

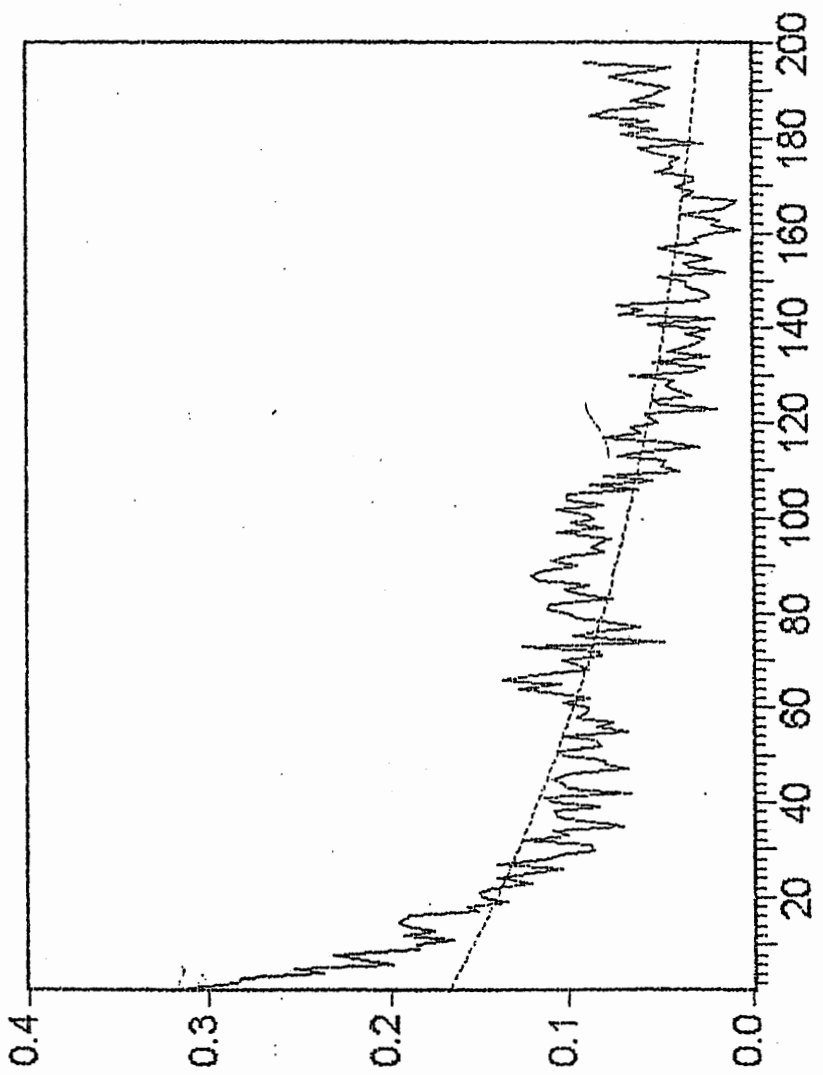
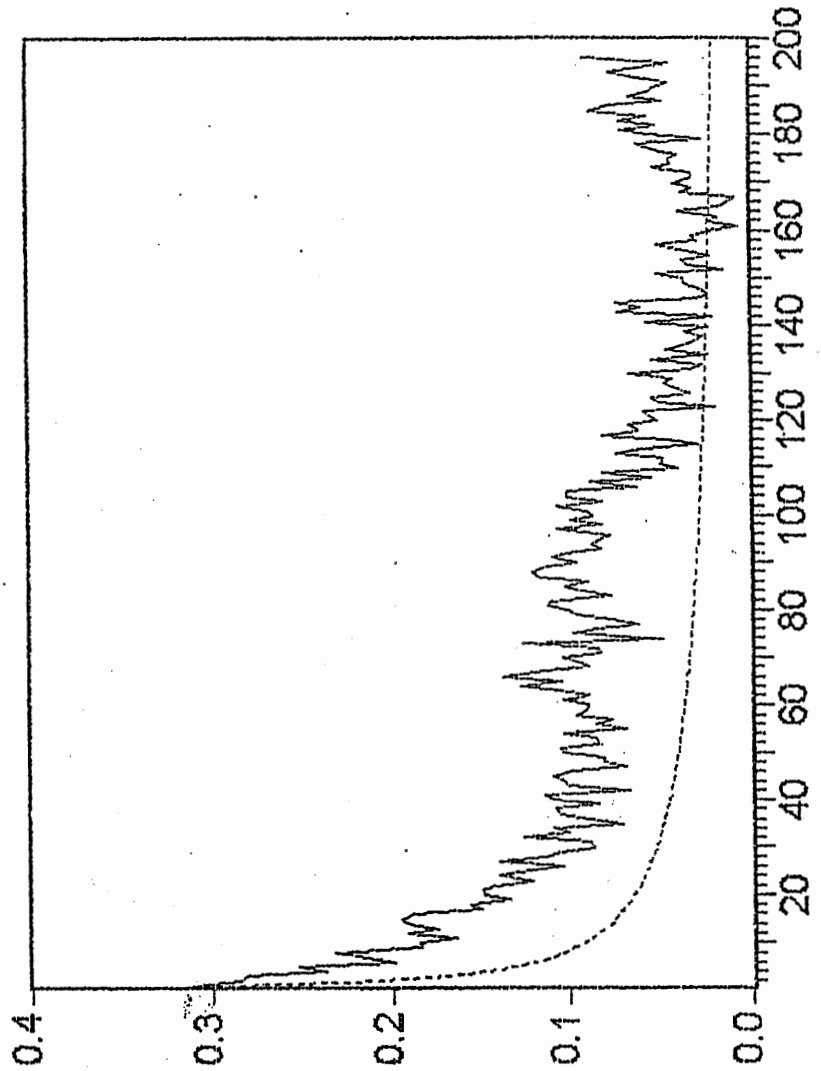


GRAFICO 5.4: Estimador de la función de autocorrelaciones del proceso $\{|Y_t|\}$ y de la función $\varrho_j^*(|Y|)=(0.315)j^{-0.526}$



aproximadamente hiperbólica de acuerdo con una función de la forma:

$$\hat{\rho}_j^*(|Y|) \approx \frac{\Gamma(1-\hat{d})}{\Gamma(\hat{d})} j^{2\hat{d}-1} = (0.311) j^{-0.526}$$

En el **GRAFICO 5.4** figuran el estimador de la función de autocorrelaciones y la trayectoria de esta función hiperbólica.

Por otra parte, el estimador de la varianza modificada $\sigma_\eta^2(|Y|)$ asumió el valor:

$$\hat{\sigma}_q^2(|Y|) = \hat{\sigma}_n^2(|Y|)\left[1 + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\rho}_j(|Y|)\right] = (0.0007815)\left[1 + 2\sum_{j=1}^{12}\left(1 - \frac{j}{13}\right)\hat{\rho}_j(|Y|)\right] =$$

$$= (0.0007815)[1 + 2(1.499)] = 0.003124$$

De la aplicación de la fórmula de Andrews se obtuvo que:

$$\log(n_k) = \frac{1}{3}\log\left(\frac{3n}{2}\right) + \frac{2}{3}\log\left(\frac{2\hat{\rho}_1(|Y|)}{1-\hat{\rho}_1^2(|Y|)}\right) =$$

$$= \frac{1}{3}\log\left[\frac{2(2448)}{2}\right] + \frac{2}{3}\log\left[\frac{2(0.311)}{1-(0.311)^2}\right] = 1.0804$$

es decir, que $n_k = 12.03$ y, por lo tanto, que $q = [n_k] = 12$.

Utilizando estos resultados se obtuvo el siguiente valor del rango reescalado modificado:

$$(R/S)_{12/2448} = \frac{\hat{\sigma}_{12}(|Y|)}{\hat{\sigma}_{2448}(|Y|)}(R/S)_{2448} = \frac{(0.027956)(268.6614)}{0.055897} = 134.3668$$

y el estadístico $n^{-1/2}(R/S)_{q,n}$ asumió el valor:

$$\frac{(R/S)_{12/2448}}{\sqrt{2448}} = 2.7154$$

Este valor, comparado con el valor crítico 1.862 (ver **CUADRO 5.1**) para un $\alpha = 0.05$, permite concluir que existen razones suficientes para rechazar la hipótesis H_0: *RW – estricto* por la cola de la derecha, es decir, con respecto a la alternativa H_1: $\{|Y_t|\}$ *es un proceso persistente de memoria larga*.

A partir de la ya mencionada serie cronológica sobre el proceso $\{|Y_t|\}$ se obtuvo el siguiente modelo representativo de su tendencia determinística:

$$|\hat{Y}_t^{TD}| = \hat{a}_0 + \hat{a}_1 t = \underset{(0.001131)}{0.026331} + \underset{(0.0000008)}{(0.000001)}t$$

Los correspondientes estadísticos T, $T(\hat{a}_0) = 23.2829$ y $T(\hat{a}_1) = 1.3261$, permiten concluir (con un 95% de confiabilidad) que $a_1 = 0$.

Considerando los residuos de la aplicación de este modelo $\hat{\varepsilon}_t^{TD} = |Y_t| - |\hat{Y}_t^{(TD)}| = |Y_t| - \hat{a}_0 - \hat{a}_1 t$ se calcularon los valores de las sumas parciales $S_t^{TD} = \sum_{j=1}^{t} \hat{\varepsilon}_j^{TD} = S_{t-1}^{TD} + \hat{\varepsilon}_t^{TD} \left(t = 1,2,...,n; S_0^{TD} = 0\right)$, obteniéndose el valor $\sum_{t=1}^{2448} \left(S_t^{TD}\right)^2 = 13631.4603$.

De los resultados anteriores se obtuvo el siguiente valor del estadístico η_μ:

$$\eta_\mu = \frac{1}{(2448)^2 \hat{\sigma}^2_{\varepsilon^{TD}}(12)} \sum_{t=1}^{2448} \left(S_t^{TD}\right)^2 = \frac{13631.4603}{(2448)^2 (0.003124)} = 0.728$$

(donde $q = [n_k] = [12.03] = 12$). De modo que se puede concluir que, con un 95% de confiabilidad y teniendo en cuenta los valores críticos, 0.809 y 1.862 (ver **CUADRO 5.4**), existen razones suficientes para rechazar la hipótesis nula $H_0: d = 0$ (es decir, $H_0: \{|Y_t|\}$ *es un proceso de memoria corta*) con respecto a la alternativa $H_1: |d| < \frac{1}{2} (d \neq 0)$ (es decir, $H_1: \{|Y_t|\}$ *es un proceso de memoria larga*).

5.6.- Los procesos GARMA

Una interesante generalización de los procesos ARFIMA es la relacionada con el **proceso ARMA de Gegenbauer (GARMA)**[47] cuya representación es de la forma[48]:

$$\Phi(B)\left(1 - 2\xi B + B^2\right)^d Y_t = \Theta(B)\varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$: *WN – debil*). Se puede concluir, entonces, que: **i)** si $|\xi| < 1$ y $d < \frac{1}{2}$, el proceso es estacionario y **ii)** si $|\xi| < 1$ y $d > -\frac{1}{2}$, el proceso es invertible (ver Gray, H.L.; Zhang, N.F.; Woodward, W.A. (1989)).

La forma MA de esta representación ARMA es de la forma:

$$Y_t = \frac{\Theta(B)}{\Phi(B)}\left(1 - 2\xi B + B^2\right)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

---

47. Denominación debida a Gray, H.L.; Zhang, N.F.; Woodward, W.A. (1989).

48. Este proceso fue introducido por Hosking, J.R.M. (1981). En particular, si d es un entero positivo, queda definido un proceso no-estacionario con d pares de raíces complejas conjugadas y constituye una generalización de los procesos AR no-estacionarios con raíces unitarias reales (ver Ahtola, J.; Tiao, G.C. (1987)(1987)). Para una teoría general de los procesos AR, ver Chan, N.H.; Wei, C.Z. (1988).

donde los coeficientes de la función de respuesta a un impulso (conocidos como los **polinomios de Gegenbauer**) pueden ser aproximados de la siguiente forma $\psi_j \approx \cos\left[j\beta + d\left(\beta - \frac{\pi}{2}\right)\right] j^{d-1} \left(j \to \infty\right)$.

Para $j \to \infty$ la función de autocovarianzas puede ser aproximada de la siguiente forma $\gamma_j(Y) \approx \cos(j\beta) j^{2d-1}$ (donde $\beta = \cos^{-1}(\xi)$ denota la llamada **frecuencia G** y define el comportamiento armónico que caracteriza la memoria larga del proceso). Es decir, asume la forma de una función sinusoidal hiperbólicamente amortiguada[49].

La función de densidad espectral está definida por:

$$\omega(\alpha) = \frac{\sigma_Y^2}{2\pi} \frac{\left|\Theta\left(e^{-i\alpha}\right)\right|^2}{\left|\Phi\left(e^{-i\alpha}\right)\right|^2} \frac{1}{\left[2\left|\cos(\beta) - \cos(\alpha)\right|\right]^{2d}}$$

De modo que, para $\alpha \to \beta$, se verifica que $w(\alpha) \to \frac{1}{\left|\alpha^2 - \beta^2\right|^{2d}}$. Lo que demuestra que, para $\alpha \to \beta$, $\omega(\alpha)$ es no-acotada.

Un caso particular del GARMA es el **proceso puro de Gegenbauer** cuya representación es de la forma:

$$\left(1 - 2\xi B + B^2\right)^d Y_t = \varepsilon_t$$

(donde: **i)** Y_t denota una variable centrada y **ii)** $\{\varepsilon_t\}$:$WN - N\left(0, \sigma_\varepsilon^2\right)$). En este caso: **i)** si $|\xi| < 1$ y $0 < d < \frac{1}{2}$ o si $|\xi| = 1$ y $d < \frac{1}{4}$, el proceso es estacionario de segundo orden; **ii)** si $|\xi| < 1$ y $d > \frac{1}{2}$, el proceso es invertible y **iii)** si $\xi = 1$ el proceso se reduce a un $ARFIMA(0,d,0)$.

La forma MA de esta representación es de la forma:

$$Y_t = \left(1 - 2\xi B + B^2\right)^{-d} \varepsilon_t = \left(\sum_{j=0}^{\infty} \psi_j B^j\right) \varepsilon_t$$

Para $|B| \leq 1$ y $|\xi| \leq 1$, los coeficientes ψ_j pueden ser calculados de la siguiente forma:

$$\psi_j = \sum_{h=0}^{[j/2]} (-1)^h (2\xi)^{j-2h} \frac{\Gamma(d + j - h)}{\Gamma(d)\Gamma(h+1)\Gamma(j - 2h + 1)}$$

(donde $\left[\frac{j}{2}\right]$ denota la parte entera de $\frac{j}{2}$), o utilizando la siguiente fórmula recursiva:

---

[49]. Esta propiedad distingue a los procesos GARMA de los procesos ARFIMA en general.

$$\psi_j = 2\xi\left(\frac{d-1}{j}+1\right)\psi_{j-1} - \left[\frac{2(d-1)}{j}+1\right]\psi_{j-2} \quad (\psi_0 = 1)$$

debida a Gradshteyn, I.S.; Ryzhik, I.M. (1980). Para $j \to \infty$, estos coeficientes de Gegenbauer

pueden ser aproximados de la siguiente forma[50]:

$$\psi_j \approx \frac{1}{\Gamma(d)}\left(\frac{2}{j}\right)^{1-d} \frac{\cos\left[j\beta + d\left(\beta - \frac{\pi}{2}\right)\right]}{sen^d(\beta)}$$

(donde $\beta = \cos^{-1}(\xi)$). Obsérvese que, para $d < 1$, cuando $j \to \infty$, los coeficientes ψ_j decrecen en forma hiperbólica.

La función de densidad espectral es de la forma:

$$\omega(\alpha) = \frac{\sigma_\varepsilon^2}{2\pi}\left(\left|\cos(\alpha) - \cos(\beta)\right|\right)^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi}\left|4\,sen\left(\frac{\alpha+\beta}{2}\right)sen\left(\frac{\alpha-\beta}{2}\right)\right|^{-2d} \qquad (0 \le \alpha \le \pi)$$

(donde $\beta = \cos^{-1}(\xi)$). De modo que, cuando $\alpha \to \beta$, la función $\omega(\alpha)$ puede ser aproximada de la siguiente forma: $\omega(\alpha) \approx \left|\alpha^2 - \beta^2\right|^{-2d}$. Obsérvese que, para $0 < d < \frac{1}{2}$, cuando $\alpha \to \beta$, la función $\omega(\alpha)$ es no-acotada, por lo que se puede concluir que el proceso de Gegenbauer posee un ciclo persistente en la frecuencia β.

A partir de esta función se obtiene la siguiente función de autocovarianzas[51]:

$$\gamma_j(Y) = \frac{\Gamma(1-2d)}{2\sqrt{\pi}}\left[2\,sen(\beta)\right]^{\frac{1}{2}-2d}\left[P_{j-\frac{1}{2}}^{2d-\frac{1}{2}}(\xi) + (-1)P_{j-\frac{1}{2}}^{2d-\frac{1}{2}}(-\xi)\right]\sigma_\varepsilon^2$$

donde $\beta = \cos^{-1}(\xi)$ y $P_a^b(x)$ denota una función de Legendre asociada, que puede ser calculada utilizando la siguiente expresión recursiva:

$$P_a^b(x) = \frac{2a-1}{a-b}xP_{a-1}^b(x) - \frac{a+b-1}{a-b}P_{a-2}^b(x)$$

con los siguientes valores iniciales:

$$P_{-1/2}^{2d-1/2}(\xi) = \left(\frac{1+\xi}{1-\xi}\right)^{d-1/4}\frac{1}{\Gamma\left(\frac{3}{2}-2d\right)}H\left(\frac{1}{2},\frac{1}{2},\frac{3}{2}-2d,\frac{1-\xi}{2}\right)$$

---

50. Ver también Rainville, E.D. (1960).

51. Ver Gradshteyn, I.S.; Ryzhik, I.M. (1980).

y,

$$P_{1/2}^{2d-1/2}(\xi) = \left(\frac{1+\xi}{1-\xi}\right)^{d-1/4} \cdot \frac{1}{\Gamma\left(\frac{3}{2}-2d\right)} H\left(-\frac{1}{2},\frac{3}{2},\frac{3}{2}-2d,\frac{1-\xi}{2}\right)$$

siendo $H(a,b,c,x) = \sum_{h=1}^{\infty} \frac{\Gamma(c)\Gamma(a+h)\Gamma(b+h)}{\Gamma(a)\Gamma(b)\Gamma(c+h)\Gamma(h+1)} x^h$ una función hipergeométrica.

Esta función de autocovarianzas puede ser aproximada de la siguiente forma[52]:

$$\gamma_j(Y) = \frac{2^{d-1}}{\pi} \frac{\Gamma(1-2d)\Gamma(j+2d)}{\Gamma(j+1)} \frac{sen(d\pi)\cos(j\beta)}{sen^{2d}(\beta)} \left[1+O(1/j)\right]\sigma_\varepsilon^2$$

A partir de esta expresión y utilizando la fórmula de Stirling, para $j \to \infty, |\xi| < 1$ y $0 < d < \frac{1}{2}$, se obtiene la siguiente aproximación para la función de autocorrelaciones: $\rho_j(Y) \approx C\cos(j\beta) j^{2d-1}$, donde: **i)** C es una constante que puede depender de d y de β, pero no depende de j y **ii)** $\beta = \cos^{-1}(\xi)$ (ver Chung, C.F. (1996)).

Si se verifica que $\xi = 1$, entonces el proceso de Gegenbauer se reduce a un proceso $ARFIMA(0,d,0)$ y, por lo tanto, para $0 < d < \frac{1}{2}$, la correspondiente función de autocorrelaciones será de la forma:

$$\rho_j(Y) \approx \frac{\Gamma(1-d)\Gamma(j+d)}{\Gamma(d)\Gamma(j-d+1)}$$

Expresión que, para $j \to \infty$, puede ser aproximada de la siguiente forma $\lim_{j\to\infty} \rho_j(Y) \approx j^{2d-1}$ (ver Gray, H.L.; Zhang, N.F.; Woodward, W.A. (1989))

De la misma forma, si $\xi = 1$ y $0 < d < \frac{1}{2}$, se verifica que:

$$\rho_j(Y) = (-1)^j \frac{\Gamma(1-d)\Gamma(j+d)}{\Gamma(d)\Gamma(j-d+1)}$$

Expresión que, para $j \to \infty$, puede ser aproximada de la siguiente forma: $\lim_{j\to\infty} \rho_j(y) \approx (-1)^j j^{2d-1}$.

5.7.- El proceso ARFIMA y el movimiento Browniano

Avram, F.; Taqqu, M.S. (1987) desarrollaron un teorema central del límite para procesos con orden de integración fraccionario: Sea $\{Y_t\}: FWM$ que admite una representación de la forma:

---

52. Ver Gradshteyn, I.S.; Ryzhik, I.M. (1980).

$$(1-B)^d Y_t = \varepsilon_t$$

(donde Y_t denota una variable centrada y $\{\varepsilon_t\}$:$WN - N(0,\sigma_\varepsilon^2)$). Se demuestra que $\frac{S_{(kn)}}{\sigma(S_{kn})} \xrightarrow{d} B_d(k)$, donde $S_{[kn]} = \sum_{t=1}^{[kn]} Y_t$ (siendo $[kn]$ la parte entera de kn) y $B_d(k)$ es un movimiento Browniano fraccionario definido de la siguiente forma:

$$B_d(k) = \frac{1}{\Gamma(d+1)} \int_0^k (k-x)^d \, dB(x) \, (0 < k < 1)$$

(siendo $B(x)$ un movimiento Browniano estandarizado)[53], cuya función de autocovarianzas es de la forma[54]:

$$E\left(\left|B_d(k_1) - B_d(k_2)\right|^2\right) = |k_1 - k_2|^{2d}$$

la cual, por lo tanto, puede ser aproximada de la siguiente forma: $\gamma_j(B_d(k)) \approx |j|^{2(d-1)}$ (por lo que se puede concluir que, para valores de j suficientemente grandes, la función γ_j observa un decrecimiento hiperbólico) y donde $d = H - \frac{1}{2}$.

---

[53]. Obsérvese que la función $S_{[kn]}$ posee, además, las siguientes propiedades: **i)** $S_{[kn]} = O_p\left(n^{d+1/2}\right)$; **ii)** $\sigma^2\left(S_{[kn]}\right) = t^{2d+1}G(n)$ (donde G denota una función de variación lenta en el infinito).

[54]. Un proceso FWN continuo en el dominio del tiempo puede ser considerado, entonces, como la derivada de un movimiento Browniano fraccionario.

Epifonema

"¿Qué me impide soñar que alguna vez,
descifré la sabiduría
y dibujé con aplicada mano los símbolos?"

(J.L. Borges: "Elogio de la sombra")

Bibliografía

- Aalen, O.: "*Statistical inference for a family of counting processes*". University of California Press, 1975.
- Aalen, O.: "Nonparametric inference for a family of counting processes". Annals of Statistics, vol. 6, 1978.
- Abdrabbo, N.A.; Priestley, M.B.: "On the prediction of non-stationary processes". JRSS, Serie B, vol. 29, 1967.
- Abhyankar, A.; Copeland, L.S.; Wong, W.: "Uncovering nonlinear structure in real-time stock-market indexes: TheSP500,the Dax, the Nikkei 225 and the FTSE-100". Journal ob Business an Economic Statistics, vol. 15, 1997.
- Abrhamse, A.P.J.: "*On the theory and application of the general linear model*". Rotterdam University Press, 1969.
- Abrhamse, A.P.J.; Louter, A.S.: "On the new test for autocorrelation in least squares regression".Biometrika, vol. 58, 1971.
- Abril, J.C.: "*Modelos para el análisis de las series de tiempo*". Ediciones Cooperativas, 2004.
- Agiakloglou, C.; Newbold, P.: "Lagrange multiplier tests for fractional difference". Journal of Time Series analysis, vol. 15, 1994.
- Agiakloglou, C.; Newbold, P.; Wohar, M.: "Bias in an estimator of the fractional difference paramenter". Journal of Time Series Analysis, vol. 14, 1993.
- Ahtola, J.; Tiao, G.C.: "Distributions of least squares estimators of autoregressive parameters for a process with complex roots on the unit circle". Journal of Time Series Analysis, vol. 8, 1987.
- Ahtola, J.; Tiao, G.C.: "A note on the asymptotic inference in autoregressive models with roots on the unit circle". Journal of Time Series Analysis, vol. 8, 1987.
- Aitchison, J.; Brown, J.A.O.: "*The lognormal distribution*". Cambridge University Press, 1957.
- Ait-Sahalia, Y.: "Testing continuous-time models for the spot interest rate". Review of Financial Studies, vol. 9, 1996.
- Ait-Sahalia, Y.: "Maximum likelihood estimation of discretely sampled diffusions: A closed approach". Econometrica, vol. 70, 2002
- Akaike, H.: "Fitting autoregressive models for prediction". Annals Institute Statistical Mathematics, vol. 21, 1969.
- Akaike, H.: "Power spectrum estimation through autoregressive model fitting". Annals Institute STatistical Mathematics, vol. 21, 1969.
- Akaike, H.: "A fundamental relation between predictor identification and power spectrum estimation". Annals Institute Statistical Mathematics, vol. 22, 1970.
-Akaike, H.: "Statistical predictor identification". Annals Institute Statistical Mathematics, vol. 22, 1970.
- Akaike, H.: "A new look at the statistical model identification". IEEE Transactions on Automatic Control, AC-19, 1974.
- Akaike, H.: "A Bayesian analysis of the minimum AIC procedure". Annals Institute Statistical Mathematics, vol. 30, 1978.
- Aldrich, J.: "Autonomy". Oxford Economic Papers, nº 41, 1989.
- Alexander, C.O.: "*Mastering risk*". Finacial Times-Prentice Hall, 2001.
- Alexander, C.O.; Leigh, C.: "On the covariance matrices used en VaR models". Journal of Derivatives, vol. 4, 1997.
- Alexander, C.O.; Rendell, H.: "Data generation processes of spatial series: Analysis of ephemeral channel form". Geographical Analysis, vol. 27, 1995.
- Alexander, S.: "Price movements in speculative markets: Trends or random walks". Industrial Management Review, vol.2 ,1961.
- Amemiya, T.: "Regression analysis when the variance of the dependent variable is proportional to the square of its expectation". JASA, vol. 68, 1973.
- Amemiya, T.: "A note on heteroscedastic model". Journal of Econometrics, vol. 6, 1977.
- Amin, K.; Khanna, A.: "Convergence of american option values from discrete-to-continuous-time financial models". Mathematical Finance, vol. 4, 1994.
- Amstrong, A.H.: "*An introduction to ancient philosophy*". Methuen, 1957.
- Andersen, T.G.; Bollerslev, T.: "DM-dollar volatility: Introductory activity patterns macroeconomic announcements and longer runs dependence". Working Paper nº 217, Kellog Graduate School of Management, Northwestern University, 1996.
- Andersen, T.G.; Bollerslev, T.: "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts". International Economic Review, vol. 39, 1998.
- Anderson, T.W.: "*The statistical analysis of time series*". Wiley, 1971.

- Anderson, T.W.; Walker, A.M.: "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process". Annals Mathematical Statistics, vol. 53, 1964.
- Andrews, D.W.K.: "Laws of large numbers for dependent non-identically distributed random variables". Econometric Theory, vol. 4, 1988.
- Andrews, D.W.K.: "Herteroskedasticity and autocorrelation consistent covariance matrix estimation". Econometrica, vol. 59, 1991.
- Anscombe, F.J.: "The transformation of Poisson, binomial and negative binomial data". Biometrika, vol. 35, 1948.
- Anscombe, F.J.: "Contribution to the discussion of the paper by Box, G.E.P.; Anderson, S.E.L.". JRSS, Serie B, vol. 17, 1955.
- Anscombe, F.J.: "Examination of residuals". 4th. Berkeley Symposium of Mathematical Statistics and Probability, University of California Press, 1961.
- Anscombe, F.J.; Tukey, J.W.: "The examination and analysis of residuals". Technometrics, vol. 5, 1963.
- Ansley, C.F.; Newbold, P.: "Finite sample properties of estimators for autoregressive moving average models". Journal of Econometrics, vol. 13, 1980.
- Arditti, F.D.: "Another look of a mutual fund performance". Journal of Financial and Quantitative Analysis, vol. 10, 1971.
- Arley, N.: "On the theory of stochastic processes and their applications". Wiley, 1948.
- Aroian, L.A.: "Some properties of the conditional Weibull distribution". Transactions 19th. Technical Conference of the American Society for Qality Control, 1965.
- Ash, R.B.; Gardner, M.F.: "*Topics in stochastic processes*". Academic Press, 1975.
- Atlan, H.: "*Entre le cristal et la fumée*". Seuil, 1979.
- Avram, F.; Taqqu, M.S.: "Non central limit theorems and Appell polynomials". Annals of Probability, vol. 15, 1987.
- Azzalini, A.: "Statistica inference: Based on the likelihood". Chapman & Hall, 1996.
- Bachelier, L.B.J.A.: "*Theory of speculation*". Gauthier-Villars, 1900.
- Badrinath, S.G.; Chatterjee, S.: "On measuring skewness and elongation in common stock distributions: The case of the market index". Journal of Business, vol. 61, 1988.
- Badrinath, S.G.; Chatterjee, S.: "A data-analytic lock at skewness and elongation in common stock return distributions". Journal Business and Economic Statistics, vol. 9, 1991.
- Baestaens, D.E.; van der Berg, W.M.; Vaudrey, H.: "Money market headline new flashes, effective news and the DEM/USD swap rate: An introductory analysis in operational time". 3rd. Forecasting Financial Market Conference, London, 1996.
- Bai, I.; Ng, S.N.: "Testing skewness, kurtosis and Normality for time series". Journal of Business and Economic Statistics, vol. 23, 2005.
- Baillie, R.T.; Bollerslev, T.G.: "Intra-day and inter-market volatility in foreign exchange rates". Review of Economics Studies, vol. 58, 1990.
- Bain, L.J.; Antle, C.E.: "Estimation of parameters in the Weibull distribution". Technometrics, vol. 9, 1967.
- Balke, N.S.; Fomby, T.B.: "Shifting trends, segmented trends an infrequent permanent shocks". Journal of Monetary Economics, vol. 28, 1991.
- Balke, N.S.; Fomby, T.B.: "Infrequent permanent shocks and the finite sample performance of unit roots tests". Economic Letters, vol. 36, 1991.
- Ball, C.A.; Torous, W.N.: "The maximum likelihood estimation of security price volatility: Theory, evidence and application on option pricing". Journal of Business, vol. 57, 1984.
- Baltagi, B.H.: "*A companion to theoretical econometrics*". Blackwell, 2001.
- Bancroft, T.A.: "On biases in estimation due to the use of preliminary tests of significance". Annals of Mathematical Statistics, vol. 15, 1944.
- Barndorff-Nielsen, O.; Blaesild, P.: "Global maxima and likelihood in inear models". Res. n° 57, Dept. Theoretical Statistics, Aarhus University, 1980.
- Barndorff-Nielsen, O.; Cox, D.R.: "*Asymptotic techniques for use in statistics*". Chapman & Hall, 1996.
- Banerjee, A.; Dolado, J.; Galbraith, J.W.; Hendry, D.F.: "*Co-integration, error-correction and the economic analysis of non-stationary data*". Oxford University Press, 1993.
- Barnard, G.A.: "New methods in quality control". JRSS, Serie B, vol. 25, 1963.
- Bartels, R.; Bornholt, G.; Hanslow, K.: "The polynomial trend model with autocorrelated residuals". Communications in Statistics, vol. 11, 1982.
- Bartels, R.; Goodhew, J.: "The robustness of the Durbin-Watson test". Review Economics and Statistics, vol. 63, 1968.
- Bartlett, M.S.: "The use of transformations". Biometrics, vol. 3, 1947.
- Bartlett, M.S.: "Some evolutionary stochastic processes". JRSS, Serie B, vol. 11, 1949.
- Basawa, I.V.: "Estimation of the utocorrelation coefficient in simple Markov chains". Biometrika, vol. 59, 1972.
- Basawa, I.V.; Prakasa Rao, B.L.S.: "*Convergence of probability measures*". Wiley, 1968.

- Basawa, I.V.; Prakasa Rao, B.L.S.: "*Statisitcal inference for stochastic processes*". Academic Press, 1980.
- Basawa, I.V.; Scott, D.J.: "Efficient tests for stochastic processes". Sankhya, Serie A, vol. 39, 1977.
- Bates, G.E.; Neyman, J.: "Contribution to the theory of accident processes, I: An optimistic model of the correlation between light and severe accidents". University of California Publications on Statistics, vol. 1, 1952.
- Battaglia, F.: "Inverse autocovariances and a measure of linear determinism for a stationary process". Journal of Time Series Analysis, vol. 4, 1983.
- Beall, G.: "The transformation of data from entomological field experiments so that the analysis of variance becomes applicable". Biometrika, vol. 29, 1942.
- Beckenback, E.F. (ed.): "*Modern mathematics for engineers*". McGraw-Hill, 1956.
- Beckers, S.: "The constant elasticity of variance model and its implications for option pricing". Journal of Finance, vol. 35, 1980.
- Bellman, R.; Cooke, K.L.: "*Differential-difference equations*". Princeton University Press, 1963.
- Bera, A.K.; Jarque, C.M.: "Efficient tests for Normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence". Economic Letters, vol. 7, 1981.
- Beran, J.: "*Statistics fon long-memory processes*". Chapman & Hall, 1994.
- Berenblutt, I.I.; Webb, G.I.: "A new test for autocorrelated errors en the linear regression model". JRSS, Serie B, vol. 35, 1973.
- Bergström, H.: "Comparison method for distribution functions of sums of independent random variables". Theory of Probability and Applications, vol. 15, 1970.
- Berman, S.M.: "Limiting theorems for the maximum term in stationary squences". Annals of Mathematical Statistics, vol. 35, 1964.
- Bernard, P. (ed.): "Lecture notes in mathematics 1816. Lectures on probability theory and statistics, Saint-Flour Summer School". Springer-Verlag, 2003.
- Bernard, P.: "How close are the optio priceng formulas of Bachelier and Black-Scholes?". Mathematical Finance, vol. 18, 2008.
- Bernardo, J.M.; Smith, A.F.M.: "*Bayesian theory*". Wiley, 1994.
- Bernoulli, J.: "*Ars conjectandi*". Thurnisiorum, Basilea, 1713.
- Bernstein, P.: "*Capital ideas: The improbable origins of modern Wall Street*". Free Press, 1992.
- Bernstein, S.: "Sur l'extension du théorème limité du calcul des probabilités aux sommes des quantités dépendantes". Mathematische Annalen, vol. 97, 1927.
- Bernstein, S.: "Quelques remarques sùr le théorème limite de Liapunoff". Dokl. Akad. Nauk. SSSR, Serie Matemática, vol. 24, 1939.
- Bernstein, S.: "New applications of almost independent quantities". Izv. Akad. Nauk. SSSR, Serie Matemática, vol. 4, 1940.
- Bernstein, S,: "On sums of dependent quantities having a joint regression wich is almost zero". Dokl. Akad. Nauk. SSSR,Serie Matemática, vol. 32, 1941.
- Beśka, M.: "Poisson distribution as a limit law for sums of dependent random variables". Preprint nº 6/78, Institute of Mathematics Nicholas Copernicus, Univ. Torum, 1978.
- Bhansali, R.J.; Downham, D.Y.: "Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion". Biometrika, vol. 64, 1977.
- Bharucha-Reid, A.T.: "On stochastic processes". Biometrics, vol. 9, 1953.
- Bharucha-Reid, A.T.: "*Elements of the theory of Markov processes and their applications*". McGraw-Hill, 1960.
- Bhattacharya, R.N.; Waymire, E.C.: "*Stochastic processes with applications*". Wiley, 1990.
- Bickel, P.J.; Ritov, Y.; Stoker, T.M.: "Tailor-made tests for goodness-of-fit to semiparametric hypotheses". JRSS, vol. 48, 1986.
- Billingsley, P.: "Statistical methods for Markov chains". Annals of Mathematical Statistics, vol. 32, 1961.
- Billingsley, P.: "*Statistical inference for Markov processes*". University of Chicago Press, 1961.
- Billingsley, P.: "The Lindeberg-Lévy theorem for martingales". Proc. American Mathematical Society, vol. 12, 1961.
- Billingsley, P.: "*Convergence of probability measures*". Wiley, 1968.
- Billingsley, P.: "*Probability and measure*". Wiley, 1986.
- Birch, J.B.; Martin, R.D.: "Confidence intervals for robust estimates of the first order autoregressive parameter". Journal of Time Series Analysis, vol. 2, 1981.
- Birkhoff, G.D.: "Proof of the ergodic theorem". Proc. National Academy of Sciences, USA, vol. 17, 1931.
- Black, F.: "Studies in stock price volatility changes". Proceedings of the 1976 Meeting of the Business ans Economic Statistics Section. American Statistical Association, 1976.
- Black, F.; Scholes, M.: "The valuation of option contacts and a test of market efficiency". Journal of Finance, vol. 27, 1972.
- Black, F.; Scholes, M.: "The pricing of options an corporate liabilities". Journal Political Economy, May/Jun, 1973.

- Blanchard, O.; Quah, D.: "The dynamic effects of aggregate demand and supply disturbances". American Economic Review, vol. 79, 1989.
- Blattberg, R.C.: "Evaluation of the power of the Durbin-Watson statistic for non-first order serial correlation alternatives". Review Economic and Statistics, vol. 55, 1973.
- Blattberg, R.C.; Gonedes, W.J.: "A comparison of the stable and Student distributions as statistical models for stock prices". Journal of Business, vol. 47, 1974.
- Blattberg, R.C.; Sargent, P.: "Regression with non-Gaussian stable distributions: Some sampling results". Econometric, vol. 39, 1971.
- Bloomfield, P.: "***Fourier analysis of time series: An introduction***". Wiley, 1976.
- Blume, L.; Easley, D.; O'Hara, M.: "Market statistics and technical analysis: The role of volume". Journal of Finance, vol. 49, 1994.
- Bodkin, R.G.; Klein, L.R.; Marwah, K. (eds.): "*A history of macroeconometric model building*". Elgar, 1991.
- Bollerslev, T.; Domowitz, I.: "Trading patterns and prices in the interbank foreign exchange market". Journal of Finance, vol. 48, 1993.
- Boltzmann, L.: "Ueber die mecbanische Bedeutung des sweiten Hauptsatzes der Wärmetheorie". Wissenschaftliche Abhandlungen, vol.1, 1866.
- Boltzmann, L.: "Studien über das Gleichgewicht der lebendigen Kraft zwschen bewegten materiallen Punkten". Wissenschaftliche Abhandlungen, vol. 1, 1868.
- Boltzmann, L.: "Ueber das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen". Wissenschaftliche Abhandlungen, vol. 1, 1871.
- Botzmann, L.: "Weitere Studien über das wärmegleichgewicht unter Gasmolekulen". Wissenschaftliche Abhandlungen, vol. 1, 1872.
- Boltzmann, L.: "Ueber das Wärmegleichgewicht von Gasen, auf welche äussere Kräfte wirken". Wissenschftliche Abhandlungen, vol. 2, 1875.
- Boltzmann, L.: "Ueber die Eigenschaften monozyklischer und damit verwandter Systeme". Wissenschaftliche Abhandlungen, vol. 3, 1884.
- Boltzmann, L.: "Ueber die mechanischen Anbalogien des zweiten Hauptsatzes der Thermodynamik". Wiisenschafliche Abhandlungen, vol. 3, 1887.
- Borodin, A.N.; Salminen, P.: "*Handbook of Brownian motion. Facts and formulas*". Birkhäuser, 1996.
- Born, M.: "Zur Quantenmechanik der Stossvorgänge". Zeitschrift für Physik, vol. 37, 1926.
- Born, M.: "Quantenmechanik der Stossvorgänge". Zeitschrift für Physik, vol. 28, 1926.
- Borel, E.: "*Leçons sur la théorie des fonctions*". Gauthier-Villars, 1898.
- Borel, E.: "Remarques sur certaines questions de probabilité". Oeuvres de Emile Borel, vol. 2, 1905.
- Borel, E.: "Les probabilités dénombrables et leur applications arithmétiques". Rendiconti Circolo Matematico di Palermo, vol. 27, 1909.
- Box, G.E.P.; Cox, D.R.: "An analysis of transformations". JRSS, Serie B, vol. 26, 1964.
- Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.: "*Time series analysis: Forecasting ans control*". Prentice Hall, 1994.
- Box, G.E.P.; Pierce, D.A.: "Distribution of residual autocorrelations in autorregresive moving average time series model. JASA, vol. 65, 1970.
- Box, G.E.P.; Tiao, G.C.: "A change in level of a non-stationary time series". Biometrika, vol. 52, 1965.
- Box, G.E.P.; Tiao, G.C.: "Intervention analysis with applications to economic and environmental problems". JASA, vol. 70, 1975.
- Box, G.E.P.; Tiao, G.C.: "Comparison of forecast and actuality". Applied Statistics, vol. 25, 1976.
- Box, G.E.P.; Tiao, G.C.: "A canonical analysis of multiple time series". Biometrika, vol. 64, 1977.
- Boyle, P.P.; Ananthanarayanan, A.L.: "The impact of variance estimation in option valuation models". Journal of Financial Economics, vol. 5, 1977.
- Brace, A.; Gątarek, D.; Musiela, M.: "The market model of interest rate dynamics". Mathematical Finance, vol. 4, 1997.
- Brailsford, T.J.; Faff, R.W.: "An evalution of volatility forecasting techniques". Journal of Banking and Finance, vol. 20, 1996.
- Brandorff-Nielsen, O.: "*Information and exponential families in statistical theory*". J.Wiley & Sons, 1978.
- Breiman, L.: "*Probability*", Addison-Wesley, 1968.
- Brenner, M.; Subrahmanyan, M.G.: "A simple formula to compute the implied standard deviation". Financial Analysts Journal, vol. 44, 1988.
- Breusch, T.S.: "Testing for autocorrelation in dynamic linear models". Australian Economic Papers, vol. 17, 1978.
- Breusch, T.S.; Pagan, A.R.: "A simple test for heteroscedasticity and random coefficient variation". Econometrica, vol. 47, 1979.
- Brigo, D.; Mercurio, F.: "*Interest rate models: Theory and practice*". Springer, 2001.

- Brillinger, D.R.; Tiao, G.C. (eds.): "***Directions in time series***", MIT Press, 1980.
- Brock, W.A.; Hseih, D.A.; LeBaron, B.: "*A test for non-linear dynamics*". MIT Press, 1990.
- Brock, W.A.; Lakonishock, J.; Le Baron, B.: "Simple technical trading rules and the stochastic properties of stock returns". Journal of Finance, vol. 47, 1992.
- Brockwell, P.J.; Davis, R.A.:"*Time series: Theory and methods*". Springer-Verlag, 1987.
- Brockwell, P.J.; Davis, R.A.:"*Introduction to time series analysis and forecasting*". Springer-Verlag, 1996.
- Brown, B.M.: "Martingale central limit theorems". Annals of Mthematical Statistics, vol. 42, 1971.
- Brown, B.M.: "Cramér-von Mises distributions and permutation tests". Biometrika, vol. 69, 1982.
- Brown, B.M.; Eagleson, G.K.: "Martingale convergence to infinitely divisible laws with finite variances". Transactions of American Mathematical Society, vol. 162, 1971.
- Brown, D.; Jennings, R.: "On technical analysis". Review Financial Studies, vol. 2, 1989.
- Brown, R.L.; Durbin, J.; Evans, M.: "Techniques for testing the constancy of regression relationships over time". JRSS, Serie B, vol. 35, 1975.
- Bru, B.: "Un hiver en campagne". Comptes Rendus de l'Academie des Sciences, Serie I, vol. 331, 2000.
- Bru, B.; Yor, M.: "Comments on the life and mathematical legacy of Wolfgang Doeblin". Financial Stochastics, vol. 6, 2002.
- Brusch, S.G.: "*The kind of motion we call heat*", North-Holland, 1976.
- Buff, R.: "*Uncertain volatility models*". Springer, 2001.
- Bunge, M.: "*Filosofia de la fisica*". Ariel, 1978.
- Bury, K.V.: "*Statistical models in applied sciences*". Wiley, 1975.
- Butler, J.S.; Schachter, B.: "Unbiased estimation of the Black-Scholes formula". Journal of Financial Economics, vol. 15, 1986.
- Campbell, J.Y.; Lo, A.W.; MacKinlay, A.C.: "*The econometrics of financial markets*". Princeton University Press, 1997.
- Campbell, J.Y.; Hentschel, L.: "No news good news: An symmetric model of changing volatility". Journal of Financial Economics, vol. 31, 1992.
- Campbell, J.Y.; Mankiw, N.G.: "Permanent and transitory components in macroeconomic fluctuations". American Economic Review, vol. 77, 1987.
- Campbell, J.Y.; Mankiw, N.G.: "Are output fluctuations transitory?". Quarterly Journal of Economics, vol. 102, 1987.
- Campbell, J.Y.; Perron , P.: "Pitfalls an opportunities: What macroeconomists should know about unit roots". Technical Working Paper 100, NBER, 1991.
- Casella, G.; Berger, R.L.: "*Statistical inference*". Brooks-Cole, 1990.
- Casti, J.L.: "*Searching for certainty*". Morrow, 1991.
- Casti, J.L.; Karlqvist, A.: "*Randomness, prediction and explanation in science*". CRC Press, 1991.
- Caudill, S.B.: "The necessity of mining data". Atlantic Economic Review, vol. 26, 1988.
- Chaitin, G.: "*Algorithmic information theory*". Cambridge University Press, 1987.
- Chaitin, G.: "Randomness in arithmetic". Scientific American, vol. 259, 1988.
- Chaitin, G.: "*Informatiobn, randomness and incompleteness*". World Scientific, 1990,
- Chambers, J.M.; Cleveland, W.S.; Kleiner, B.; Tukey, P.A.: "*Graphical methods for data analysis*". Wadsworth-Duxbury, 1983.
- Chan, K.; Karoly, A.; Longstaff, F.; Sanders, A.: "An empirical comparison of alternative models of the short-term interest rate". Journal of Finance, vol. 47, 1992.
- Chan, N.H.; Wei, C.Z.: "Limiting distributions of least squares estimates of unstable autoregressive processes". Annals of Statistics, vol. 16, 1988.
- Chapman, S.: "On the Brownian displacements and thermal difussion of grains suspended in a non-uniform fluid". Proceedings of the Royal Society, Serie A, vol. 119, 1928.
- Charemza, W.W.; Deadman, D.F.: "*New directions in econometric practice*". Elgar, 1997.
- Charlier, C.V.L.; Wicksell, S.D.: "On the dissection of frequency functions". Arkiv für Matematik och Fysik, vol. 18, 1924.
- Chatfield, C.: "***The analysis of time series: Theory and practice***". Chapman & Hall, 1975.
- Chen, L.: "*Stochastic mean and stochastic volatility - A three factor model of the term structure of interest rates and application to the pricing of interest rate derivatives*". Blackwell, 1996.
- Chen, R.; Scott, L.: "Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the term structure". Review of Financial Studies, vol. 5, 1992.
- Chen, R.; Scott, L.: "Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rate".Journal of Fixed Income, vol. 3, 1993.
- Chen, R.; Scott, L.: "Interest rate options in multifactor Cox-Ingersoll-Ross models of the term structure". Journal of Derivatives, vol. 3, 1995.

- Cherian, K.C.: "A bivariate corelated gamma-type distribution function". Journal of the Indian Mathematical Society, vol. 5, 1941.
- Chesney, M.; Elliot, R.; Madan, D.; Yang, H.: "Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying". Mathematics of Finance, vol. 3, 1993.
- Chesney, M.; Scott, L.O.: "Pricing European currency options: A comparison of the modified Black- Scholes model and a random variance model". Journal financial Quantitative Analysis, vol. 24, 1989.
- Cheung, Y.W.: "Tests for fractional integration: A Monte Carlo investigation". Journal of Time Series Analysis, vol. 14, 1993.
- Cheung, Y.W.; Diebold, F.X.: "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean". Journal of Econometrics, vol. 62, 1994.
- Chevychev, P.L.: "Sur deux théoremes relatifs aux probabilités". 1887.
- Chiang, Ch.L.: "A birth-illness-death processes". Annals of Mathematical Statistics, vol. 35, 1964.
- Chiang, Ch.L.: "On the expectation of the reciprocal of a random variable". The American Statistician, vol. 20, 1966.
- Choi, I.: "Durbin-Hausmann tests for a unit root". Oxford Bulletin of Economics and Statistics, vol. 54, 1992.
- Choi, I.; Chung, B.S.: "Sampling frequency and the power of tests for a unit root: A simulation study". Economic Letters, vol. 49, 1995.
- Choi, S.; Hall, W.J.; Schick, A.: "Asymptotically uniformly most powerful tests in parametric and semiparametric models". Annals of Statistics, vol. 24, 1996.
- Chow, G.C.: "Tests of equality between sets of coefficients in two linear regressions". Econometrica, vol. 52, 1960.
- Chow, Y.S.; Teicher, H.: "*Probability theory: Independence, interchangeability, martingales*". Springer-Verlag, 1978.
- Christ, C.F.: "*Econometric models and methods*". J. Wiley & Sons, 1966.
- Christie, A.A.: "The stochastic behavior of common stock variances: Value, leverage and interest rate effects". Journal of Financial Economics, vol. 10, 1982.
- Chu, J,T,; Hotelling, H.: "The moments of the sample median". Annals Mathematical Statistics, vol. 26, 1955.
- Chung, C.F.: "A note on calculating the autocovariances of fractionally integrated ARMA nodels". Economic Letters, vol. 45, 1994.
- Chung, C.F.: "Estimating a generalized long memory process". Journal of Econometrics, vol. 73, 1996.
- Chung, K.L.: "*Markov chains with stationary transition probabilities*". Springer-Verlag, 1960.
- Chung, K.L.; Williams, R.J.: "*Introduction to stochastic integration*".Birkhäuser, 1990.
- Ciesielski, Z.: "*Lectures on Brownian motion. Heat conduction and potential theory*". Aarhus, 1965.
- Claeskens, G.; Hort, N.I.: "Goodness of fit via parametric likelihood ratios". JRSS, vol. 48, 1986.
- Clark, P.K.: "A subordinated stochastic process model with finite variances for speculative prices". Econometrica, vol. 41, 1973.
- Cleveland, W.S.: "The inverse autocorrelations of a time series and their applications". Technometrics, vol. 14, 1972
- Cline, D.B.H.: "Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data". Statistics Department, Colorado State University, 1983.
- Cline, D.B.H.; Brockwell, P.J.: "Linear prediction of ARMA processes with infinite variance". Stochastic Processes and their Applications, vol. 19, 1985.
- Cochrane, D.; Orcutt, G.H. (1949): "Application of least squares regressions to relationships containing autocorrelated error terms". JASA, vol. 44, 1949.
- Cochrane, J.H.: "How big is the random-walk in GNP". Journal of Political Economy, vol. 96, 1988.
- Cochrane, J.H.: "A critique of the application of unit root tests". Journal of Economic Dynamics and Control, vol. 15, 1991.
- Cogburn, R.: "Asymptotic properties of stationary sequences". University of California, Publications in Statistics, vol. 3, 1960.
- Cohen, A.C.: "Estimation in mixtures of two normal distributions". Technometrics, vol. 9, 1967.
- Coles, S.: "*An introduction to statistical modeling of extreme values*". Springer-Verlag, 2001.
- Conley, T.G.; Hansen, L.P.; Luttmer, E.G.J.; Scheikman, J.A.: "Short term interest rates as subordinated diffusions". Review of Financial Studies, vol. 10, 1997.
- Cont, R.: "Empirical properties of asset returns: Stylized facts and statisticas Issues". Quantitative Finance, vol. 1, 2001.
- Cooley, T.F.; LeRoy, S.F.: "A theoretical macroeconometrics". Journal of Monetary Economics, vol. 16, 1985.
- Cootner, P.H. (ed.): "*The random character of stock market prices*". MIT Press, 1964.
- Corrado, C.J.; Miller, T.W.: "A note on a simple, accurate formula to compute implied standard deviations". Journal Banking and Finance, vol. 20, 1996.
- Cowles, A.: "Can stock market forecasters forecast?". Econometrica, vol. 1, 1933.
- Cox, D.R.; Hinkley, D.V.: "*Theoretical statistics*". Chapman & Hall, 1974.
- Cox, J.C.: "Notes on options pricing I: Constant elasticity of variance diffusions". Working Paper, Stanford University,

- Cox, J.C.; Ingersoll, J.E.; Ross, S.: "A theory of the terrm structure of interest rates". Econometrica, vol. 53, 1985.
- Cox, J.C.; Ross, S.A.; Rubinstein, M.: "Option pricing: A simplified approach". Journal of Financial Economics, vol. 7, 1979.
- Cox, J.C.; Rbinstein, M.: "*Option markets*". Prentice-Hall, 1985.
- Cramér, H.: "*Random variables and probability distributions*". Cambridge University Press, 1937.
- Cramér, H.: "On some classes of non-stationary stochastic processes". Proc. 4th. Berkeley Symposium, vol. 2, 1961
- Cragg, J.G.: "Estimation and testing in time-series regression models with heteroscedasticity disturbances". Journal of Econometrics, vol. 20, 1982.
- Cragg, J.G.: "More efficient estimation in the presence of heteroscedasticity of unknown form". Econometrica, vol. 51, 1983.
- Cramér, H.: "*Mathematical methods of statistics*". Princeton University Press, 1945.
- Cramér, H.: "On the structure of purely non-deterministic stochastic processes". Ark. Mathematics, vol. 19, 1961.
- Csörgö, M.: "On the strong law of large numbers and the central limit theorem for martingales". Transactions of the American Mathemartical Society, vol. 131, 1968.
- Cumby, R.; Figlewski, S.; Hasbrouk, J.: "Forecasting volatility and correlations with EGARCH models". Journal of Derivatives, vol. 1, 1993.
- Cuthbertson, K.: "*Quantitative financial economics: Stocks, bonds and foreign exchange*". Wiley, 1996.
- Dahlhaus, R.: "Efficient parameter estimation for self-similar processes". The Annals of Statistics, vol. 17, 1989.
- Dambis, K.E..: "On the descomposition of continuous submartingales". Theory of Probability and Applications, vol. 10, 1965.
- Dana, R.A.; Jeanblanc, M.: "*Financial markets in continuous time*". Springer, 2003.
- Danielsson, J.; Vries, C.G.: "Tail index and quantile estimation with very high frequency data". Journal of Empirical Finance, vol. 4, 1997.
- Darnell, A.C.; Evans, J.L.: "*The limits of econometrics*". Elgar, 1990.
- David, F.; Barton, D.: "*Combinatorial chance*". Hafner, 1962.
- Davidson, J.: "*Stochastic limit theory*". Oxford University Press, 1994.
- Davies, N.; Triggs, C.M.; Newbold, P.: "Significance levels of the Box-Pierce portmanteau statistic in finite samples". Biometrika, vol. 64, 1977.
- Davis, M.: "*Computability and insolvability*". Mc.Graw-Hill, 1958.
- Davis, M.; Weyuker, E.: "*Computability and languages*". Academic Press, 1983.
- Davis, R.A.; Resnick, S.I.: "Limite theory for the sample covariance and correlation functions of moving averages". Annals of Statistics, vol. 14, 1986.
- Davis, R.C.: "On the theory of prediction of non-stationary processes". Journal of Applied Physics, vol. 23, 1952.
- Davydov, Y.A.: "The invariance principle for stationary processes". Theory of Probability and applications, vol. 15, 1970.
- Dawson, J.: "Kurt Gödel in sharper focus". Mathematical intelligencer, vol. 6, 1984.
- Day, D.H.; Groves, T. (eds.): "*Adaptive economic models*". Academic Press, 1975.
- de Finetti, B.: "Sulle funzioni ad incremento aleatorio". Rendiconti della Reale Academia Nazinale Lincei, vol. 10, 1929.
- de Finetti, B.: "Sulla possibilitá di valori eccezionali per una legge di incrementi aleatori". Rendiconti della Reale Academia Nazionale Lincei, vol. 10, 1929.
- de Finetti, B.: "Integrazione delle finzioni a incremento aledatorio". Rendiconti della Reale Academia Nazionale Lincei, vol. 10, 1929.
- de Finetti, B.: "Sui passaggi al limite nel calcolo delle probabilitá". Rendiconti del Reale Istituto Lombardo di Scienze e Lettere, vol. 63, 1930.
- de Haan, L.; Resnick, S.I.: "A simple asynptotic estimate for the index of a stable distribution". JRSS, Serie B, vol. 42, 1980.
- de Haan, L.; Resnick, S.I.; Rootzén, H.; de Vries, C.G.: "Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes". Stochastic Processes and their Applications, vol. 32, 1989.
- de Helguero, F.: "Sui massima delle curve dimorfiche". Biometrika, vol. 3, 1904.
- de Helguero, F.: "Per la risoluzione delle curve dimorfiche". Bometrika, vol. 4, 1905.
- de Jong, D.N.; Nankervis, J.C.; Savin, N.E.; Whiteman, C.H.: "Integration versus trend stationarity in macroeconomic time series". Working Paper n° 89-99, University of Iowa, 1989.
- Dekker, A.L.M.; De Haan, L.: "On the estimation of extreme value index and large quantile estimation". Annals of Statistics, vol. 17, 1989.
- Delbaen, F.; Schachermayer, W.: "*The mathematics of arbitrage*". Springer, 2006.
- de Lima,P.J.F.: "Nonlinearities and nonstationarities in stock returns". Journal of Business and Economic Statistics, vol. 15, 1997.

- De Moivre, A.: "*Approximatio ad summam terminorum binomii (a+b)" in seriem expansi*", Londres, 1733.
- Denton, F.T.: "Data mining as an industry". Review of Economics and Statistics, vol. 67, 1985.
- Devaney, R.L.: "An introduction to chaotic dynamical systems". Addison-Wesley, 1989.
- Dezhbaksh, H. (1990): "The inappropiate use of serial correlation tests in dynamic linear models". The Review of Economics and Statistics, LXXII, 1990.
- Dhrymes, P:J: "On the game of maximizing R^2·. Australian Economic Papers, vol. 9, 1970.
- Dhrymes, P:J: "***Distributed lags: Problems of estimation and formulation***". Holden-Day, 1971.
- Dhrymes, P.J.: "***Econometrics: Statistical foundations and applications***". Springer, 1974.
- Dhrymes, P.J.: "*Time series, unit roots and cointegration*". Academic Press, 1998.
- Dickey, D.A.: "Estimation and hypothesis testing in nonstationary rime series". Ph.D. Dissertation, Iowa State University, 1976.
- Dickey, D.A.; Bell, W.R.; Miller, R.B.: "Unit roots in time series models: Tests and implications". American Statistician, vol. 15, 1986.
- Dickey, D.A.; Fuller, W.A.: "Distribution of the estimators for autorregressive time series with a unit root". JASA, vol.74, 1979.
- Dickey, D.A.; Fuller, W.A.: "Likelihood ratio statistics for autorregressive time series with a unit root". Econometrica, vol. 49, 1981.
- Dickey, D.A.; Pantula, S.G.: "Determining the order of differencing in autorregressive processes". Journal of Business and Economic Statistics, vol. 15, 1987.
- Diebold, F.X.; Mariano, F.S.: "Comparing predictive accuracy". Journal of Business and Economic Statistics, vol. 13, 1995.
- Diebold, F.X.; Rudebusch, G.D.: "On the power of Dickey-Fuller test against fractional alternatives". Economic Letters, vol. 35, 1991.
- Dimson, E.; Marsh, P.: "Volatility forecasting without data snooping". Journal of Banking and Finance, vol. 14, 1990.
- Doan, Th.: "*RATS user's manual*". Estima, 1992.
- Doeblin, W.: "Sur l'equation de Kolmogoroff". Comptes Rendus de l'Academie des Sciences, Serie I, vol. 331, 2000.
- Dolado, J.J.; Jenkinson, T.; Sosvilla-Rivero, S.: "Cointegration and unit roots". Journal Economic Surveys, vol. 4, 1990.
- Dolby, J.I.: "A quick method for choosing a transformation". Technometrics, vol. 5, 1963.
- Domowitz, J.: "The linear model with sochastic regressors and heteroscedastic dependent errors". Center for Mathematical Studies in Economics and Management Science, Discussion Paper n° 543, Northwestern University, Illinois, 1982.
- Donsker, M.: "An invariance principle for certain probability limit theorems". Memoirs of the American Mathematical Society, n° 6, 1951.
- Doob, J.L.: "Note on probability". Annals of Mathematics, vol. 37, 1936.
- Doob, J.L.: "Stochastic processes depending on a continuous parameter". Transactions of the American Mathematical Society, vol. 42, 1937.
- Doob, J.L.: "Stochastic processes with an integer-valued parameter". Transactions of the American Mathematical Society, vol. 44, 1938.
- Doob, J.L.:"*Stochastic processes*". Wiley, 1953.
- Donsker, M.: "An invariance principle for certain probability limit theorems". Memoirs of the American Mathematical Society, vol. 6, 1951.
- Donsker, M.: "Justification and extension of Doob's neuristic approach to the Kolmogorov-Smirnov theorems". Annals of Mathematuical Statistics, vol. 23, 1952.
- Doornik, J.A.; Hansen, H.: "An omnibus test for univariate and multivariate Normality". Working Paper, Nuffield College, Oxford, 1994.
- Dothan, M.U.: "*Prices in financial markets*". Oxford University Press, 1990.
- Draper, N.R.; Hunter, W.G.: "Transformations: Some examples revisited". Technometrics, vol. 11, 1969.
- Drogin, R.: "An invariance principle for matingales". Annals of Mathematical Statistics, vol. 43, 1972.
- Drost, F.C.; Nijman, T.E.: "Temporal aggregation of GARCH processes". Econometrica, vol. 61, 1993.
- Dubey, S.D.: "Asymptotic properties of several estimators of Weibull parameters". Technometrics, vol. 7, 1965.
- Dubey, S.D.: "Hyper-efficient estimator of the location parameter of the Weibull laws". Naval Research Logistics Quarterly, vol. 13, 1966.
- Dubey, S.D.: "Asymptotic efficiencies of the moment estimators for the parameters of the Weibull laws". Naval Research Logistics Quarterly, vol. 13, 1966.
- Dubey, S.D.: "Normal and Weibull distributions". Naval Research Logistics Quarterly, vol. 14, 1967.
- Dubins, L.E.: "On a theorem of Skorohod". Annals of Mathematical Statistics, vol. 39, 1968.
- Dubins, L.E.; Schwartz, G.: "On continuous martingales". Proc. National Academy of Sciences, USA, vol. 53, 1965.

- Duffie, D.: "*Dynamic asset pricing theory*". Princeton University Press, 1992.
- Duffie, D.; Kan, R.: "A yield-factor model of interest rates". Mathematical Finance, vol. 6, 1994.
- Duffie, D.; Protter, P.: "From discrete-to continuous-time finance: Weak convergence of the financial gain process". Mathematical Finance, vol. 2, 1992.
- Duffie, D.; Singleton, K.: "Simulated moments estimation of Markov models of asset prices". Econometrica, vol. 61, 1993.
- Dunis, C.: "*Forecasting financial markets*". Wiley, 1996.
- Dunis, C.; Zhou, B. (eds.): "*Non-linear modelling of high frequency financial time series*". Wiley, 1998.
- Durbin, J.: "The fitting of time series models". Review International Statistical Institute, vol. 28, 1960.
- Durbin, J.: "Testing for serial correlation in regression analysis based on the periodogram of least squares residuals". Biometrika, vol. 56, 1969.
- Durbin, J.: "Testing for serial correlation in least-squares regresssion when some of the regressors are lagged dependent variables". Econometrica, vol. 38, 1970.
- Durbin, J.: "An alternative to the bounds tests for testing for serial correlation in least-squares regression". Econometrica, vol. 38, 1970.
- Durbin, J.: "Distribution theory for tests based on the sample distribution function". Regional Conference Applied Mathematics, SIAM, 1973.
- Durbin, J.; Watson, G.S.: "Testing for serial correlation in least-squares regression I". Biometrika, vol. 37, 1950.
- Durbin, J.; Watson, G.S.: "Testing for serial correlation in least-squares regression II". Biometrika, vol. 38, 1951.
- Durrett, R.; Resnick, S.I.: "Functional limit theorems for dependent variables". Annals of Probability, vol. 6, 1978.
- Dvoretzky, A.: "Central limit theorems for dependent random variables and some applications". Annals of Mathematical Statistics, vol. 40, 1969.
- Dvoretzky, A.: The central limit problem for dependent random variables". Actges du Congres International des Mathématiciens, Niza, 1971.
- Dvoretzky, A.: "Asymptotic normality for sums of dependent random variables". Proc. 6th. Berkeley Symposium on Mathematics and Statistics, vol. 2, 1972.
- Earnshaw, R.A. (ed.): "*Fundamental algorithms in computer graphics*". Springer-Verlag, 1985.
- Eberlein, E.; Keller, U.: "Hyperbolic distributions in finance". Bernoulli Society, vol. 1, 1995.
- Edgeworth, F.Y.: "Observations and statistics: an essay on the theory of error of observation and the first principles of statistics". Transactions of the Cambridge Philosophical Society, vol. 14, 1885.
- Edwards, R.: "*Technical analysis of stock trends*". J. Magee, 1966.
- Eggenberger, F.: "Die Wahrscheinlichkeitsanstecknung". Mitteilungen der Vereiningung Schweizerischer Versicherungs Mathematiker, 1924.
- Eggenberger, F.; Pólya, G.: "Über die Statistik verketteter Vorgänge". Zeitschrift für Angewandte Mathematik und Mechanik, vol. 1, 1923.
- Eichhorn, W.: "*Functional equations in economics*". Addison-Wesley, 1978.
- Einstein, A.: "Eine Theorie der Grundlagen der Thermodynamik". Annalen der Physik, vol. 11, 1903.
- Einstein, A.: "Ueber die von der molekularkinetischen Theorie de Wärme goforderte Bewegung von in Flüssigkeiten suspendierten Teilchen". Annalen der Physik, vol. 17, 1905.
- Einstein, A.: "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristichen Gesichtpunkt". Annalen der Physik, vo. 17, 1905.
- Einstein, A.: "Zur Thoerie der Brownschen Bewegung". Annalen der Physik, vol. 19, 1906.
- Einstein, A.: "Über die Gültigkeitsgrenze des satzes von thermodynamischen Gleichgewitch und über die Möglichkeit einer neuen Bestimmung der Elementarquanta". Annalen der Physik, vol. 22, 1907.
- Einstein, A.: "Zur Quantentheorie der Strahlung". Mitteilungen der physikalischen Gesellschaft Zürich, vol. 16, 1916.
- Einstein, A.: "Zur Quantentheorie der Strahlung". Physikalischen Zeitschrift, vol. 18, 1917.
- Eisenberger, I.: "Genesis of bimodal distributions". Technometrics, vol. 6, 1964.
- Elderton, W.P.; Johnson, N.L.: "*Systems of frequency curves*". Cambridge University Press, 1969.
- Elerian, O.; Chib, S.; Shephard, N.: "Likelihood inference for discretely observed non-linear diffusions". Econometrica, vol. 69, 2001.
- Elliot, G.; Rothenberg, T.J.; Stock, J.H.: "Efficient tests for an autoregressive unit root". Econometrica, vol. 64, 1996.
- Elsevier, F. (ed.): "*Nonparametric statistics and related topics*". Wiley, 1992.
- Elston, R.E.: "On additivity in the analysis of variance". Biometrics, vol. 17, 1961.
- Embrecht, P.; Klüppelberg, G.; Mikosch, T.: "*Modelling extremal events for insurance and finance*". Springer-Verlag, 1997.
- Engle, R.F.: "A general approach to Lagrange multiplier model diagnostics". Journal of Econometrics. vol. 20, 1982.
- Engle, R.F.; Granger, C.W.J.: "Cointegration and error correction: Representation, estimation and testing". Econometrica, vol. 55, 1987.

- Engle, R.F.; Granger, C.W.J. (eds.): "*Long-run economic relationships*". Oxford University Press, 1991.
- Engle, R.F.; Hendry, D.F.; Richard, J.F.: "Exogeneity, causality and structural invariance in econometric modelling". CORE Discussion Paper 80-38, 1980.
- Engle, R.F.; Hendry, D.F.; Richard, J.F.: "Exogeneity". Econometrica, vol. 51, 1983.
- Engle, R.F.; McFadden, D.L. (eds.): "*Handbook of econometrics*". North-Holland, 1994.
- Engle, R.F.; Russell, J.R.: "Forecasting the frequency of changes in the quoted foreign exchange prices with the autoregressive conditional duration model". Journal of Empirical Finance, vol. 4, 1997.
- Engle, R.F.; Russell, J.R.: "Autoregressive conditional duration: A new model for irregularly spaced transaction data". Econometrica, vol. 66, 1998.
- Engle, R.F.; Sam Yoo, B.: "Forecasting and testing in co-integrated systems". Journal of Econometrics, vol. 35, 1987.
- Epstein, R.J.: "*A history of econometrics*". North-Holland, 1987.
- Eraker, B.: "MCMC analysis of diffusion models with application to finance". Journal of Business and Economic Statistics, vol. 19, 2001.
- Erdös, P.; Kac, M.: "On certain limit theorems in the theory of probability". Bulletin of the American Mathematical Society, vol. 52, 1946.
- Erdös, P.; Kac, M.: "On the number of positive sums of independent random variables". Bulletin of the American Mathematical Society, vol. 53, 1947.
- Fama, E.F.: "The behaviour of stock market prices". Journal of Business, vol. 38, 1965.
- Fama, E.F.: "Efficient capital markets: A review of theory and empirical work". Journal of Finance, vol. 25, 1970.
- Fama, E.F.: "Efficient capital markets: II". Journal of Finance, vol. 46, 1991.
- Fama, E.F.; Blume, M.: "Filter rules and stock market trading profits". Journal of Business, vol. 39, 1966.
- Fama, E.F.; Roll, R.: "Parameter estimates for symmetric stable distributions". JASA, vol. 66, 1971.
- Fang, K.T.; Kotz, S.; Ng, K.W.: "*Symmetric multivariate and related distributions*". Chapman-Hall, 1990.
- Farebrother, R.W.: "The Durbin-Watson test for serial correlation when there is no intercept in the regression". Econometrica, vol. 48, 1980.
- Faust, J.: "When are variance ratio tests for serial dependence optimal?". Econometrica, vol. 60, 1992.
- Feferman, S. (ed.): "*Kurt Gödel: Collected works*". Oxford University Press, 1986.
- Feigl, H.; Brodbeck, M. (eds.): "*Readings in the philosophy of science*". Appleton-Century-Crofts, 1953.
- Feller, W.: "Über den Zentralen Grenzwarsatz der Wahrscheinlichkeitsrechnung, I". Mathematische Zeitschrift, vol. 40, 1935.
- Feller, W.: "Zur Theorie der stochastischen Prozesse". Mathematische Annalen, vol. 113, 1936.
- Feller, W.: "Über den Zentralen Grenzwartsatz der Wahrscheinlichkeitsrechnung, II". Mathematische Zeitscrhift, vol. 42, 1937.
- Feller, W.: "Die Grundlagen der Volterraschen Theorie des Kampfes ums dasein in Wahrscheinlichkeits theoretiischer Bedhandlung". Acta Biotheoretica, vol. 5, 1939.
- Feller, W.: "On the integrodifferential equations of purely discontinuous Markoff processes". Transactions of American Mathematical Society, vol. 48, 1940.
- Feller, W.: "On the Kolmogorov-Smirnov limit theorems for empirical distributions". Annals of Mathematical Statistics, vol. 19, 1948.
- Feller, W.: "Diffusion processes in genetics". Proc. 2nd. Berkeley Symposium on Mathematicas Statistics and Probability, 1951.
- Feller, W.: "***An introduction to probability theory and its applications: Vol. II***". Wiley, 1966.
- Fielitz, B.: "Further results on asymmetric stable distributions of stock prices changes". Journal of Financial and Quantitative Analysis, vol. 11, 1976.
- Fielitz, B.; Rozell, J.: "Stable distributions and mixtures of distributions hypotheses for common stock returns". JASA, vol. 78, 1983.
- Figlewski, S.: "Forecasting volatility". Fianancial Markets, Institutions and Instruments, vol. 6, 1997.
- Fisher, R.A.: "On the mathematical foundations of theoretical statistics". Philosophical Transactions Royal Society, vol.222, 1922.
- Fisher, R.A.: "*Statistical methods for research workers*". Griffin, 1925.
- Fisher, R.A.: "Applications of 'Student' distribution". Metron, vol. 5, 1925.
- Fisher, R.A.: "Expansion of 'Student' integral in powers of n^{-1}". Metron, vol. 5, 1925.
- Fisher, R.A.: "Theory of statistical estimation". Proceedings Cambridge Philosophical Sciety, vol. 22, 1925.
- Fisher, R.A.: "Moments and producto moments of sampling distributions". Proceedings London Mathematical Society, vol. 30, 1928.
- Fisher, R.A.: "The general sampling distribution of the multiple correlation coefficient". Proceedings Royal Society, vol. 121, 1928.
- Fisher, R.A.: "The distribution of gene ratio for rare mutations". Proc. Royal Society of Edinburgh, vol. 50, 1930.

- Fisher, R.A.; Corbett, A.S.; Williams, C.B.: "The relation between the number of species and the number of individuals in a random sample of an animal population". Journal of Animal Ecology, vol. 23, 1943.
- Fisher, R.A.; Mather, K.: "The inheritance of style length 'Lythrum Salicano'". Annals of Eugenesic, vol. 12, 1943.
- Fisher, R.A.; Tipett, L.H.C.: "Limiting forms of the frequency distributions of the largest or smallest member of a sample". Proceedings Cambridge Philosophical Society, vo. 24, 1928.
- Fishman, M.B.; Barr, D.S.; Loick, W.J.: "Using neural nets in market analysis". Technical Analysis of Stocks and Commodities, vol. 9, 1991.
- Fisk, P.R.: "The graduation of income distributions". Econometrica, vol. 29, 1961.
- Fix, E.; Neyman, J.: "A simple stochastic model of recovery, relapse, death and loss of patients". Human Biology, vol. 23, 1951.
- Florens, J.P; Mouchart, M.: "Initial and sequential reduction of Bayesian experiments". CORE Discussion Paper 8015. Université Catholique de Louvain, Louvain-la-Neuve, 1980.
- Fox, R.; Taqqu, M.S.: "Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series". Annals of Statisitics, vol. 14, 1986.
- Fréchet, M.: "Sur la loi de probabilité de l'ecart maximum". Annales de la Societé Polonnaise de Mathématique, vol. 6, 1927.
- French, K.R.; Roll, R.: "Stock return variances: The arrival of information and the reaction of traders". Journal of Financial Economics, vol. 17, 1986.
- Frennberg, P.; Hansson, B.: "An evaluation of alternative models for predicting stock volatility: Evidence from a small stock market". Journal of International Financial Markets, Institutions and Money, vol. 5, 1996.
- Friedman, J.H.; Rafsky, L.C.: "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests". Annals of Statistics, vol. 7, 1979.
- Friedman, J.H.; Rafsky, L.C.: "Graphics for the multivariate two-sample problem". JASA, vol. 76, 1981.
- Frisch, R.: "Autonomy of economic relations". Cambridge Conference of Econometric Society, 1938.
- Froelich, B.R.: "Some estimators for a random coefficient regression model". JASA, vol. 68, 1973.
- Fuller, W.A.: "*Introduction to statistical time series*". Wiley, 1976.
- Furry, W.H.: "On fluctuation phenomena in the passage of high energy electrons through lead". The Physical Review, vol. 52, 1937.
- Galai, D.: "Tests of market efficiencyof the Chicago Board Options Exchange". Journal of Business, vol. 50, 1972.
- Galambos, J.: "*Advanced probability theory*". Dekker, 1995.
- Galbraith, J.W.; Hendry, D.F.: "*Co-integration, error correction and the econometric analysis of non-stationary data*". Oxford University Press, 1993.
- Gallant, A.R.; Long, J.R.: "Estimating stochastic diffsion equations efficiently by minimum chi-squared". Biometrika, vol. 84, 1997.
- Gallant, A.R.; Tauchen, G.: "The relative efficiency of mthod of moments estimators". Working Paper, Economics Department, University of North Carolina, 1997.
- Gallant, A.R.; White, H.: "*A unified theory of estimation and inference for nonlinear dynamics*". Blackwell, 1988.
- Galton, F.: "Typical laws of heredity". Proceedings of the Royal Institution of Great Britain, vol. 8, 1877.
- Galton, F.: "Regression towards mediocrity in hereditary stature". Journal of the Antropological Institute, vol. 14, 1885.
- Galton, F.: "Family likeness in stature". Proceedings of the Royal Society, vol. 40, 1886.
- Galton, F.: "Co-relations and their measurement, chiefly from anthropometric data". Proceedings of the Royal Society, vol. 35, 1888.
- Gänsler, P.; Haüsler, E.: "Remarks on the functional central limit theorem for martingales". Z. Wahrsc. Verw. Gebiete, vol. 50, 1980.
- Gänsler, P.; Strobel, J.; Stute, W.: "On central limit theorems for martingale triangular arrays". Acta Mathematica Academy of Sciences of Hungary, vol. 31, 1978.
- Gardner, M.: "*The ambidestroux universe*". Scribner & Sons, 1979.
- Garwood, F.: "Fiducial limits for the Poisson distribution". Biometrika, vol. 28, 1936.
- Gastwirth, J.L.; Selwyn, M.R.: "The robustness properties of two tests for serial correlation". JASA, vol. 75, 1980.
- Geweke, J.; Porter-Hudak, S.: "The estimation and application of long memory time series models". Journal of Time Series Analysis, vol. 4, 1983.
- Ghose, D.; Kroner, K.F.: "Temporal aggregation of high frequency financial data". First International Conference in High-Frequency Data in Finance. Olsen & Asoc., 1995.
- Ghysels, E.; Gourieroux, C.; Jasiak, J.: "Trading patterns, time deformation and stochastic volatility in foreign exchange markets". Working Paper, Université de Montréal, 1995.
- Ghysels, E.; Jasiak, J.: "Stochastic volatility and time deformation: An application to trading volume and everage effects". Working Paper, Université de Montréal, 1995.

- Gibbs, J.W.: "*Elementary principles in statistical mechanics*". Dover, 1902.
- Gigerenzer, G.; Swijtink, Z.; Porter, T.; Daston, L.; Beatty, J.; Krüger, L.: "*The empire of chance*". Cambridge University Press, 1989.
- Gikhman, I.J.; Sokorokhod, A.V.: "*The theory of stochastic processes*". Springer, 1975.
- Gilbert, C.L.; "Professor Hendry's econometric methodology". Oxford Bulletin of Economics and Statistics, vol. 48, 1986.
- Giot, P.: "Time transformations, intraday data and volatility models". Journal of Computational Finance, vol. 4, 2001.
- Girasinov, I.V.: "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures". Theory of Probability and Applications, vol. 5, 1960.
- Giriatis, L.; Surgalis, D.: "A central limit theorem for quadratic forms in strongly sependent linear variables and its application to asymptotical Normally of Whittle's estimate". Probability Theory and Related Fields, vol. 86, 1990.
- Glejser, H.: "A new test for heteroscedasticity". JASA, vol. 64, 1967.
- Gnedenko, B.V.: "Sur la distribution limite du terme maximum d'une série aléatoire". Annals of Mathematics, vol. 44, 1943.
- Godfrey, L.G.: "Testing for multiplicative heteroskedasticity". Journal of Econometrics, vol. 8, 1978.
- Godfrey, L.G.: "Testing against general autoregressive and moving average error models when the regressors include dependent variables". Econometrica, vol. 46, 1978.
- Godfrey, L.G.: "*A diagnostic check of the variance model in regression equations with heteroskedastic disturbances*". University of York, 1979.
- Goldberger, A.S.: "*Econometric Theory*". Wiley, 1964.
- Goldberger, A.S.: "*A course in econometrics*". Harvard University Press, 1991.
- Goldfeld, S.M.; Quandt, R.E.: "*Nonlinear methods in econometrics*". North-Holland, 1972.
- Goldie, C.M.; Smith, R.L.: "Slow variation wuth remainder: Theory and applications". Quarterly Journal of Mathematics, vol. 38, 1987.
- Gompertz, B.: "On the nature of the function expressive of the law of human mortality". Philosophical Transactions of the Royal Society, vol. 115, 1825.
- Good, I.J.: "A causal calculus". British Journal for the Philosophy of Science, vol. 11, 1961.
- Goodhart, C.A.E.; Figliouli, L.: "Every minute counts in financial markets". Journal of International Money and Finance, vol. 10, 1991.
- Gouriéroux, C.: "*ARCH models and financial applications*". Springer-Verlag.
- Gouriéroux, C.; Jasiak, J.: "*Financial econometrics*". Princeton University Press, 2001.
- Gouriéroux, C.; Monfort, A.: "*Statistique et modéles économitriques*". Economica, 1989.
- Gouriéroux, C.; Monfort, A.: "*Time series and dynamic models*". Cambridge University Press, 1990.
- Govindarajulu, Z.: "Best linear estimates under symmetric censoring of the parameters of a double exponential population". JASA, vol. 61, 1966.
- Gradszteyn, I.S.; Ryzhnik, M.: "*Tables and integrals, series and products*". Academic Press, 1980.
- Granger, C.W.J.: "Economic processes involving feedback". Information and Control, vol. 5, 1963.
- Granger, C.W.J.: "Long memory relationships and the aggregation of dynamic models". Journal of Econometrics, vol. 14, 1980.
- Granger, C.W.J.: "*Modelling economic series: Readings in econometric methodology*". Oxford University Press, 1990.
- Granger, C.W.J.; Andersen, A.P.: "*An introduction to bilinear time series models*". Vandenhoeck & Ruprecht, 1978.
- Granger, C.W.J.; Joyeux, R.: "An introduction to long memory time series models and fractional differencing". Journal of Time Series Analysis, vol. 1, 1980.
- Granger, C.W.J.; Hughes, A.D.: "A new look at some old data: The beveridge wheat price series". JRSS, Serie A, vol. 134, 1971.
- Granger, C.W.J.; Morgenstern, O.: "*Predictability of stock returns prices*". Heath-Lexington, 1970.
- Granger, C.W.J.; Newbold, P.: "Spurius regressions in econom trics". Journal of Econometrics, vol. 2, 1974.
- Granger, C.W.J.; Newbold, P.: "Forecasting transformed series". JRSS, Serie B, vol. 38, 1976.
- Granger, C.W.J.; Newbold, P.: "Forecasting economic time series". Academic Press, 1977.
- Granger, C.W.J.; Ramanathan, R.: "Improved methods of combining forecasts". Journal of Forecasting, vol. 3, 1984.
- Granger, C.W.J.; Teräsvirta, T.: "*Modelling nonlinear economic relationships*". Oxford University Press, 1993.
- Gray, H.L.; Zhang, N.F.; Woodward, W.A.: "On generalized fractional processes". Journal of Time Series Analysis, vol. 10, 1989.
- Greenberg, E.; Webster, C.E.: "*Advanced econometrics: A bridge to the literature*", Wiley, 1983.
- Greenwood, J.A.; Olkin, I.; Savage, I.R.: "Index to Annals of Mathematical Statistics-vol.1-31, 1930-1960". 1962.
- Greenwood, M.; Yule, G.U.: "The statistics of anti-thypoid and anti-cholua innoculations and the interpretation of such statistics in general". Proceedings of the Royal Society of Medicine, vol. 8, 1915.

such statistics in general". Proceedings of the Royal Society of Medicine, vol. 8, 1915.
- Greenwood, M.; Yule, G.U.: "An inquiry into the nature of frequency distribuion representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidentes". JRSS, vol. 83, 1920.
- Grenander, U.: "*Probability and statistics: The Harald Cramér volume*". Almqvist & Wicksell, 1959.
- Grenander, U.; Szego, G.: "*Toeplitz forms and their applications*". University of California Press, 1958.
- Griliches, Z.; Intriligator, M.D. (eds.): "*Handbook of econometrics*". North-Holland, 1984.
- Grimett, G.R.; Stirzaker, D.R.: "*Probability and random processes*". Clarendon Press, 1994.
- Groenendijk, P.A.; Lucas, A.; de Vries, C.G.: "A note on the relationship between GARCH and symetric stable processes". Journal Empirical Finance, vol. 2, 1995.
- Grünbichler, A.; Longsstaff, F.A.: "Valuing futures and options on volatility". Journal of Banking and Finance, vol. 20, 1996.
- Gumbel, E.J.: "Ranges and midranges". Annals of Matehematical Statistics, vol. 15, 1944.
- Gumbel, E.J.: "*Statistics of extremes*". Columbia University Press, 1958.
- Gumbel, E.J.: "Bivariate exponencial distributions". JASA, vol. 55, 1960.
- Gumbel, E.J.; Keeney, R.D.: "The extremal quotient". Annals of Matehematical Statistics, vol. 21, 1950.
- Haavelmo, T.: "The probability approach in econometrics". Econometrica, Suplemento vol. 12, 1944.
- Hagerman, R.: "More evidence on the distribution of security returns". Jounral of Finance, vol. 33, 1978.
- Hagstroem, K.G.: " Remarks on Pareto distributions". Skandinavisk Aktuarietidskrift, 1960.
- Hahn, G.J.; Shapiro, S.S.: "*Statistical models in engineering*". Wiley, 1967.
- Haldane, J.B.S.: "The approximate normalization of a class of frequency distributions". JASA, vol. 57, 1937.
- Hall, A.: "Testing for a unit root in the presence of moving average errors". Biometrika, vol. 76, 1989.
- Hall, P.: "On some simple estimates of an exponent of regular variation". JRSS, Serie B, vol. 44, 1982.
- Hall, P.: "Tailor-made tests of goodness of fit". JRSS, Serie B, vol. 47, 1985.
- Hall, P.; Heyde, C.C.: "*Martingale limit theory and its application*". Academis Press, 1980.
- Hall, P.; Tajvidi, N.: "Permutation tests for equality of distributions in high-dimensional settings". Biometrika, vol. 89, 2002.
- Hall, P.; Welsh, A.H.: "Adaptive estimates of parameters of regular variation". Annals of Statistics, vol. 13, 1985.
- Hamersley, J.M.; Handscomb, D.C.: "*Monte Carlo methods*". Methuen, 1964.
- Hamilton, J.D.: "*Time series analysis*". Princeton University Press, 1994.
- Hannan, E.J.: "*Multiple time series*". Wiley, 1970.
- Hannan, E.J.: "The estimation of the order of an ARMA process". The Annals of Statistics, vol. 8, 1980.
- Hannan, E.J.; Krishnaiah, P.R.; Rao, M.N. (eds.): "*Handbook of statistics*". North-Holland, 1985.
- Hannan, E.J.; Quinn, B.G.: "The determination of the order of an autoregression". JRSS, Serie B, vol. 41, 1979.
- Hannan, E.J.; Rissanen, J.: "Recursive estimation of mixed atoregressive-moving average order". Biometrika, vol. 69, 1982.
- Hansen, B.E.: "Testing for parameterinstability in linear models". Journal of Policy Modeling, vol. 14, 1992.
- Harding, G.: "The use of probability paper for the graphical analysis of polynomial frequency distributions". Journal of Marine Biology, vol. 28, 1949.
- Hardy, G.H.; Littlewood, J.E.; Polya, G.: "*Inequalities*". Cambridge University Press, 1952.
- Hareer, H.L.: "A bibliography on extreme-value theory". International Statistical Review, vol. 46. 1978.
- Harper, C.P.: "Testing for the existence of a lagged relationship within Almon's method". Review of Economic and Statistics, vol. 50, 1977.
- Harrison, M.J.; Pliska, S.R.: "Martingales and stochastic integrals in the theory of continuous trading". Stochastic Processes Applied, vol. 11, 1981.
- Hart, B.I.: "Significance levels for the ratio of the mean-square successive difference to the varfiance". Annals Mathematical Statistics, vol. 13, 1942.
- Harter, H.L.; Moore, A.H.: "Conditional maximum-likelihood estimators, from singly censored samples, of the scale parameters of type II extreme value distributions". Technometrics, vol. 10, 1968.
- Hartman, P.; Winter, A.: "On the law of iterated logarithm". American Journal of Mathematics, vol. 63, 1941.
- Harvey, A.C.: "Estimation of parameters in heteroscedastic regression model". European Meeting of the Econometric Society, Grenoble, 1974.
- Harvey, A.C.: "*Forecasting, structural time series models and the Kalman filter*". Cambridge University Press, 1989.
- Harvey, A.C.: "*The econometric analysis of time series*". Allan-Hempstead, 1990.
- Harvey, C.R.: "Predictable risk and returns in emerging markets". Review Financial Studies, vol. 8, 1995.
- Hassler, U.; Wolters, J.: "On the power of unit roots tests against fractionally integrated alternatives". Economic Letters, vol. 45, 1994.
- Hastings, N.A.J.; Peacock, J.B.: "*Statistical distributions: A handbook for students and practitioners*". Wiley, 1975.

- Haubrich, J.E.; Lo, A.W.: "The sources and nature of long-term memory in the business cycle". NEBR Working Paper, n° 2951.
- Hausman, J.: "Specification tests in econometrics". Econometrica, vol. 46, 1978.
- Heckerman, D.; Madani, A. (eds.): "Causality an Bayesian belief networks". Proc. 9th. Conference on Uncertainty in Artificial Intelligence, 1993.
- Heisenberg, W.: "Ueber Quantunmechanische Umdeutung kinematischer und mechanischer Beziehungen". Zeitschrift für Physik, vol. 33, 1925.
- Helson, J.; Sarason, Y.: "Past and future". Mathematica Scandinavia, vol. 21, 1967.
- Hendry, D.F.: "On asymptotic theory and finite sample experiments". Economica, vol. 160, 1973.
- Hendry, D.F.; Harrison, R.W.: "Monte Carlo methodology and the finite sample behaviour of ordinary and two-stage least squares". Journal of Econometrics, vol.2, 1974.
- Hendry, D.F.: "*Dynamics econometrics*". Oxford University Press, 1995.
- Hendry, D.F.; Morgan, M.S.: "*The foundations of econometric analysis*". Cambridge University Press, 1996.
- Hendry, D.F.; Wallis, K.: "*Econometrics and quantitative economics*". Blackwell, 1984.
- Hey, J.D.; Winch, O. (Eds.): "*A century of economics: Years of the Royal Economic Society and the Economic Journal*". Blackwell, 1990.
- Henze, N.: "A multivariate two sample test based on the number of nearest neighbor type coincidences". Annals of Statistics, vol. 16, 1988.
- Hida, T.: "*Brownian motion*". Springer, 1980.
- Hiemstra, C.; Jones, J.D.: "Another look at long memory in common stock returns". Journal of Empirical Finance, vol. 4, 1997.
- Hildebrand, W. (ed.): "*Advances in econometrics*". Cambridge University Press, 1982.
- Hildreth, C.; Houck, J.P.: "Some estimators for a linear model with random coefficients". JASA, vol. 65, 1968.
- Hill, B.M.: "A simple general approach to inference about the tail of a distribution". Annals of Statistics, vol. 3, 1975.
- Hill, W.J.; "Statistical techniques for model building". Ph. D. Thesys, University of Wisconsin, 1966.
- Hillmer, S.C.; Tiao, G.C.: "Likelihood function of stationary multiple autoregressive moving average models". JASA, vol. 74, 1979.
- Hipel, K.W.; McLeod, A.I.; Lennox, W.C.: "Advances in Box-Jenkins modelling: 1. Mode construction". Water Resources Research, vol. 13, 1977.
- Ho, T.; Lee, S.: "Term structure movements and pricing interest rate contingent claims". Journal of Finance, vol. 41, 1986.
- Hoaglin, D.C.; Mosteller, F.; Tukey, J.W. (eds.): "*Exploring data tables, trends and shapes*". Wiley, 1985.
- Hoeffding, W.: "A class of statistics with asymptotically normal distribution". Annals of Mathematical Statistics, vol. 19, 1948.
- Hoffmann-Jorgensen, J.: "*Probability with a view toward statistics*". Chapman & Hall, 1994.
- Hogg, R.V. (ed.): "*Studies in statistics*". Mathematical Association of America, 1979.
- Holland, D.W.; Rosenbaum, P.R.: "Conditional association and unidimensionality in monotone latent variable models". Annal of Statistics, vol. 14, 1986.
- Hood, W.C.; Koopmans, T.C. (eds.): "*Studies in econometric method*". Cowles Commission Monograph, n°14, Wiley, 1953.
- Hopf, E.: "On causality, startistics and probability". Journal of Mathematics and Physics, vol. 13, 1934.
- Hosking, J.R.M.: "The multivariate portmanteau statistic". JASA, vol. 75, 1980.
- Hosking, J.R.M.: "Fractional differencing". Biometrika, vol. 68, 1981.
- Hosking, J.R.M.: "Lagrange-multiplier tests of multivariate time series models". JRSS, Serie, B, vol. 43, 1981.
- Hostinsky, B.: "Sur la théorie générale de phénomènes de la diffusion". Comptes Rendus du 1er. Congrés des Mathématiciens des Pays Slaves, Varsovia, 1929.
- Hostinsky, B.: "Ein allgemeiner Satz über die Brown'sche Bewegung". Physikalische Zeitswchrift, vol. 30, 1929.
- Hostinsky, B.: "Diskussion über wahrscheinlichkeit". Erkenntnis, vol. 1, 1930.
- Hoyle, M.H.: "Transformations: An introduction and a bibliography". International Statistical Review, vol. 41, 1973.
- Hsu, D.R.; Miller, R.; Wichern, D.: "On the stable Paretian behavior of stock market prices". JASA, vol. 69, 1974.
- Huang, C.F.; Litzenberger, R.: "*Foundations for financial economics*". North-Holland, 1988.
- Hull, J.: "*Options, futures and other derivative securities*". Prentice-Hall, 2002.
- Hull, J.; White, A.: "Taking account of the kurtosis in market variables when calculating VaR". Journal of Derivatives, vol. 5, 1998.
- Hurst, H.E.: "Long-term storage capacity of reservoirs". Transactions American Society Civil Engineers, vol. 116, 1951.
- Hurst, H.E.; Black, R.P.; Sinaika, Y.M.: "*Long-term storage: An experimental study*". Constable, 1965.
- Hurvich, C.M.; Tsai, C.L.: "Regresssion and time series model selection in small samples". Biometrika, vol. 76, 1989.

- Ibragimov, I.A.: "A central limit theorem for a class of dependent random variables". Theory of Probability and Applications", vol. 8, 1963.
- Ikeda, N.; Watanabe, S.: "*Stochastic differential equations and diffusion processes*". North-Holland, 1981.
- Imhof, J.P.: "Computing the distribution of quadratic forms in Normal variables". Biometrika, vol. 48, 1961.
- Ingster, Y.I.: "Asymptotically minimax hypothesis testing for nonparmetric alternatives". Mathematical Methods of Statistics, vol. 2, 1993.
- Isserlis, L.: "On certain probable error and correlation coefficients of multiple frequency distributions with skew regression". Biometrika, vol. 11, 1916.
- Itô, K.: "Stochastic integral". Proc Imperial Academy, Tokyo, vol. 20, 1944.
- Itô, K.: "On stochastic differential equations". Memoirs American Mathematical Society, vol. 4, 1951.
- Itô, K.: "On a formula concerning stochastic differentials". Nagoya Mathematical Journal, vol. 3, 1951.
- Janssen, A.: "Global power functions of goodness-of-fit tests". Annals of Statistics, vol. 28, 2000.
- Jarque, C.M.; Bera, A.K.: "Efficient tests for Normality, homoskedasticity and serial dependence of regression residuals". Economic Letters, vol. 6, 1980.
- Jarque, C.M.; Bera, A.K.: "A test of normality of observations and regression residuals". International Statistical Review, vol. 55, 1987.
- Jenkins, G.M.: "Tests of hypotheses in the linear autorregresive model, I". Biometrika, vol. 41, 1954.
- Jenkins, G.M.: "Tests of hypotheses in the linear autorregresive model, II". Biometrika, vol. 43, 1956.
- Jenkins, G.M.; Watts, D.G.: "*Spectral analysis and its applications*". Holden-Day, 1969.
- Jensen, M.; Ruback, R.: "The market for corporate control: The scientific evidence". Journal Financial Economics, vol. 11, 1983.
- Jerison, D.; Singer, I.; Strook, D.: "*The 1994 Wiener Symposium Proceedings*". American Mathematical Society, 1996.
- Johnson, N.L.: "Systems of frequency curves generated by methods of translation". Biometrika, vol. 36, 1949.
- Johnson, N.L.; Kotz, S.: "*Continuous univariate distributions*". Wiley, 1970.
- Johnson, R.A.; Wichern, D.W.: "*Applied multivariate statistical analysis*". Prentice Hall, 1998.
- Johnston, J.: "*Econometric models*". McGraw-Hill, 1984.
- Judge, G.G.; Griffiths, W.E.; Carter-Hill, R.; Lütkepohl, H.; Lee, T.: "*The theory and practice of econometrics*". Wiley, 1985.
- Kahn, J.A.; Ogaki, M.: "A consistent test for the null of stationarity against the alternative of a unit root". Economic Letters, vol. 39, 1992.
- Kakwani, N.C.: "The unbiased of Zellner's seemingly unrelated regression equations estimators". JASA, vol. 62, 1967.
- Kallenberg, O.: "*Foundations of modern probability*". Springer, 1997.
- Karatzas, I.; Shreve, S.: "*Brownian motion and stochastic calculus*". Springer, 1988.
- Karatzas, I.; Shreve, S.: "*Methods of mathematical finance*". Springer, 1998.
- Karlin, S.; McGregor, J.: "The differential equations of birth-and-death processes and the Stieltjes' moment problem". Transactions American Mathematical Society, vol. 85, 1957.
- Karlin, S.; McGregor, J.: "The classification of birth and death processes". Transactions American Mathematical Society, vol. 86, 1957.
- Karlin, S,; Taylor, H.M.: "*A first course in stochastic processes*". Academic Press, 1975.
- Karlin, S,; Taylor, H.M.: "*A second course in stochastic processes*". Academic Press, 1981.
- Karst, O.J.; Polowy, H.: "Sampling properties of the median of a Laplace distribution". JRSS, Serie A, vol. 74, 1963.
- Katz, J.O.: "Developing neural network forecasters for trading". Technical Analysis of Stocks and Commodities, vol. 10, 1992.
- Kean, J.: "Using neural nets for intermarket analysis". Technical Analysis of Stocks and Commodities, vol. 10, 1992.
- Keiding, N.: "Estimation in the birth process". Biometrika, vol. 61, 1974.
- Keiding, N.: "Maximum likelihood estimation in the birth and death processes". Annals of Statistics, vol. 3, 1975.
- Kendall, D.G.: "On the generalized `birth-and-death` processes": Annals of Mathematical Statistics, vol. 19, 1948.
- Kendall, D.G.: "Stochastic processes and popultaion growth". JRSS, Serie B, vol. 11, 1949.
- Kendall, M.G.: "*Time series*". Griffin, 1973.
- Kendall, M.G.; Stuart, S.: "*The advanced theory of statistics*". Griffin, 1976.
- Kennedy, D.: "The distribution of the maximum Brownian excursion". Journal of Applied Probability, vol. 13, 1976.
- Kennedy, W.J. (Jr.); Gentle, J.E.: "*Statistical computing*". Dekker, 1980.
- Kessler, M.: "Estimation and ergodic diffusion from discrete observations". Scandinavian Journal of Statistics, vol. 24, 1997.
- Keynes, J.M.: "The principal averages and the laws of error which lead to them". JRSS, Serie A, vol. 74, 1911.
- Khinchin, A.Y.: "Über einen Satz der Wahrscheinlichkeitsrechnung". Mathamtische Sbornik, vol. 2, 1924.
- Khinchin, A.Y.: "Zu Birkoffs lösung des Ergodenproblem". Mathematische Annalen, vol. 107, 1932.
- Khinchin, A.Y.: "Korrelationstheorie der stationaren zufalligen Prozesse". Mathematische Annalen, vol. 109, 1934.

- Khinchin, A.Y.: "*Mathematical foundations of information theory*". Dover, 1957.
- Kibble, W.F.: "A two-variate gamma type distribution". Sankhỹa, vol. 5, 1941.
- Kim, K.; Schmidt, P.: "Some evidence on the accuracy of Phillips-Perron tests using alternative estimates of nuisance parameters". Economic Letters, vol. 34, 1990.
- Kimball, B.F.: "The bias in certain estimates of the parameters of the extreme value distribution". Annals of Mathematical Statistics, vol. 27, 1956.
- King, M.L.: "Fourth-order autocorrelation; further significance points for the Wallis test". Journal of Econometrics, vol. 8, 1978.
- King, M.L.: "A note on Szroeter's bound test". Oxford Bulletin of Economics and Statistics, vol. 43, 1981.
- King, M.L.: "The alternative Durbin-Watson test: An assesment of Durbin and Watson's choice of test statistic". Journal of Econometrics, vol. 17, 1981.
- King, M.L.: "The Durbin-Watson bounds test and regressions without an intercept". Australian Economic Papers, vol. 20, 1981.
- King, M.L.: "The Durbin-Watson test for serial correlation: Bounds for regressions with trnd and/or seasonal dummy variables". Econometrica, vol. 49, 1981.
- King, M.L.: "A bound test for heteroscedasticity". Working Paper n°5/82, Department of Economics and Operational Research, Monash University, 1982.
- King, M.L.: "A locally optimal bounds test for autorregressive disturbances". Working Paper n°21/82, Department of Economics and Operational Research, Monash University, 1982.
- King, M.L.; Hillier, G.: "Locally best invariant tests of the error covariance matrix of the linear regression model". JRSS, Serie B, vol. 47, 1985.
- Kitagawa, G.: "On a search procedure for the optimal AR-MA order". Annals of the Institute of Statistical Mathematics, vol. 29, 1977.
- Kiviet, J.F.: "On the rigour of some misspecification tests for modelling dynamic relationships". Review of Economic Studies, vol. 53, 1986.
- Kleijnen, J.P.C.: "*Statistical techniques in simulation*". Dekker, 1974.
- Kloeden, P.E.; Platen, E.: "*Numerical solution of stochastic differential equations*". Springer, 1992.
- Kloeden, P.E.; Platen, E.; Schurz, H.: "*Numerical solution of SDE through computer experiments*". Springer, 1994,
- Kmenta, J.: "*Elements of econometrics*". MacMillan, 1986.
- Koedijk, K.G.; Schafgans, M.M.A.; de Vries, C.G.: "The tail index of exchange rate returns". Journal of International Economics, vol. 29, 1990.
- Koenker, R.: "A note on Studentizing a test for heteroscedasticity". Journal of Econometrics, vol. 17, 1981.
- Koerts, J.: "New estimates of disurbances in regression analysis". JASA, vol. 66, 1971.
- Koerts, J.; Abrahamse, A.P.J.: "*On the theory and application of the general linear model*". Rotterdam University Press, 1969.
- Kolata, G.: "Does Gödel's theorem matter to mathematics?". Science, vol. 218, 1982.
- Kolmogorov, A.N.: "Über die Summen durch den Zufall bestimmter unabhangiger Grossen". Mathematische Annalen, vol. 99, 1928
- Kolmogorov, A.N.: "Sur une formule limité de M.A.Khintchine". Comptes Rendus Adademie des Sciences, Paris, vol. 186, 1928.
- Kolmogorov, A.N.: "Über des Gesetz des iterierten logarithmus". Mathematische Annalen, vol. 101, 1929.
- Kolmogorov, A.N.: "Bemerkungen zu meiner Arbeit `Ueber die summen zufälliger Grössen`". Mathematische Annalen, vol. 102, 1929.
- Kolmogorov, A.N.: "Sur la loi forte des grands nombres". Comptes Rendus Academie des Sciences, Paris, vol.191, 1930.
- Kolmogorov, A.N.: "Über die analytischen Methoden der Wahrscheinlichkeitsrechnung". Matematische Annalen, vol. 104, 1931.
- Kolmogorov, A.N.: "Sulla determinazione empirica di una legge di distribuzione". Giornale Istituto Italiano degli Attuari, vol. 4, 1933.
- Kolmogorov, A.N.: "Grundbegriffe der Wahrscheinlichkeitsrechnung". Springer, 1933.
- Kolmogorov, A.N.: "Zur Theorie der Markoffschen Ketten". Mathematische Annalen, vol. 112, 1935.
- Kolmogorov, A.N.: "Anfangsgründe der Theorie der markoffschen Ketten mit Unedlich Vielen möglichen Zustäden". Mathematische Sbornik, vol. 1, 1936.
- Kolmogorov, A.N.: "Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum". Doklady, vol. 26, 1940.
- Kolmogorov, A.N.: "Interpolation and extrapolation von Stationären Zufälligen Folgen". Bulletin Academy of Sciences (NAUK), USRR, Serie Matemática, vol. 5, 1941.
- Kolmogorov, A.N.: "*Foundations of the theory of probability*". Chelsea, 1956.

- Kolmogorov, A.N.: "The approaches to the quantitative definition of information". Problems in Information transmission, vol. 1, 1965.
- Kon, S.: "Models of stock returns: A comparison". Journal of Finance, vol. 39, 1984.
- Kondepudi, D.; Progogine, I.: "*Modern thermodynamics, from heat engines to dissipative structures*". Wiley, 1998.
- Koopman, B.O.: "Uniqueness conditions for a class of stochastic differential equations". Bulletin American Mathematical Society, vol. 63, 1957.
- Koopman, S.J.; Harvey, A.C.; Doornik, J.A.; Shephard, N.: "*Stamp %.=: Structural time series analyser, modeller and predictor*". Chapman-Hall, 1995.
- Koopmans, T.C. (ed.): "*Statistical inference in dynamic economic models*". Cowles Commission, Wiley, 1950.
- Kotz, S.; Johnson, N.L. (eds.): "*Encyclopedia of statistical sciences*". Wiley, 1983.
- Kou, S.: "A jump diffusion model for option pricing". Management Science, vol. 48, 2002.
- Koutrouvelis, I.A.: "Regression-type estimation of the parameters of stable laws". JASA, vol. 75, 1980.
- Krishnaiah, P.R. (ed.): "*Multivariate analysis*". North-Holland, 1977.
- Krüger, L.; Gigerenzer, G.; Morgan, M.: "*The probabilistic revolution. Ideas in the sciences*". MIT Press.
- Kuan, C.M.; Liu, T.: "Forecasting exchange rates using feedforward and recurrent neural networks". Journal of Applied Econometrics, vol. 10, 1995.
- Kuiper, N.H.: "Test concerning random points on a circle". Koninklijke Nederlandse Akademie van Wetenschappen, Serie A, vol. 63, 1960.
- Kullback, S.: "*Information theory and statistics*". Wiley, 1959.
- Kullback, S.; Leibler, R.A.: "On information and sufficiency". Annals of Mathematics Statistics, vol. 22, 1951.
- Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y.: "Testing the null hypothesis of stationary against the alternative of a unit root". Journal of Econometrics, vol. 54, 1992.
- Lamberton, D.; Lapeyre, B.: "*Introduction to stochastic calculus applied to finance*". Chapman & Hall, 1996.
- Landro, A.H.: "*Acerca de la probabilidad*". Ediciones Cooperativas, 2002.
- Landro, A.H.; González, M.L.: "*El problema de la predicción en el mercado financiero*". Impresos Centro, 1997.
- Lange, O.: "*Studies in mathematical economics an econometrics. In memory of Henry Schultz*". University of Chicago Press, 1942.
- Langevin, P.: "Sur la théorie du mouvement Brownien". Comptes Rendus Academie Sciences, vol. 146, 1908.
- Laplace, P.S.: "Mémoire sur la probabilité des causes par les évènemens". Memoires de Mathématique et de Physique, vol. 6, 1774.
- Laplace, P.S.: "Memoire sur les approximations des formules qui sont fonctions de trés grands nombres et sur leur application aux probabilités". Memoires de l'Academie des Sciences, Parin, 1810.
- Leadbetter, M.R.; Lindgreen, G.; Rootzén, H.: "*Extremes and related properties of random sequences and processes*". Springer-Verlag, 1983.
- Leamer, E.E.: "Let's take the con out of econometrics". American Economic Review, vol. 23, 1983.
- Le Baron, B.: "Technical trading rule profitability and foreign exchange intervention". Working Paper, 5505, NBER, Cambridge, 1996.
- Lebesgue, H.: "Leçons sur l'intégration et la recherche des fonctions primitives". Gauthier-Villars, 1904.
- LeCam, L.; Neyman, J. (eds.): "*Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*". University of California Press (9167).
- Leclerc, B.: "*Distributions statistiques et lois de probabilité*". Mouton, Gauthier-Villars, 1972.
- Lee, D.; Schmidt, P.: "On the power of the KPSS test of stationarity against fractionally integrated alternatives". Journal of Econometrics, vol. 73, 1996.
- Lee, H.S.; Amsler, C.: "Consistency of the KPSS unit root test against fractionally integrated alternative". Economic Letters, vol. 55, 1997.
- Lee, I.C.; Judge, G.G.; Zellner, A.: "*Estimating the parameters of the Markov probability model from aggreegate time series data*". North-Holland, 1970.
- Lehmann, E.L.: "*Testing statistical hypotheses*". Wiley, 1959.
- Lévy, P.: "*Calcul des probabilités*". Gauthier-Villars, 1925.
- Lévy, P.: "La thèoréme fondamental de la théorie des erreurs". Annales Institut H. Poincaré, vol. 1, 1930.
- Lévy, P.: "Sur les séries dont les termes sont des variables éventuelles indépendantes". Studia Mathematica, vol. 3, 1931.
- Lévy, P.: "Sur les integrales dont les éléments sont des variables aleatoires". Annali Scuola Normale Superiore, Pisa, Serie 23, 1934.
- Lévy, P.: "L'addition des variables aléatoires enchainées et la loi de Gauss". Bulletin Societé Mathématique, vol. 62, 1934.
- Lévy, P.: "Proprietés asymptotiques des sommes de variables aléatoires enchainées", Comptes Rendue Academie des Sciences, vol. 199, 1935.

- Lévy, P.: "Proprietés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées". Journal Mathématiques Pures et Appliqués, 1935.
- Lévy, P.: "*Théorie de l'addition des variables aléatoires*". Gauthier-Villars, 1937.
- Lindeberg, J.W.: "Eine neue Herleitung des Exponential gesetz in der Wahescheinlichkeitsrechnung". Mathematische Zeitschrift, vol, 15, 1901.
- Lindley, D.V.: "*Introduction to probability and statistics from a Bayesian viewpoint*". Cambridge U. Press, 1969.
- Ljung, G.M.; Box, G.E.P.: "On a measure of lackoff fit in time series model". Biometrika, vol. 66, 1978.
- Lo, A.W.: "Long term memory in stock market prices". Econometrica, vol. 59, 1991.
- Lo, A.W. (ed.): "*Market efficiency: Stock market behaviour in theory and practice*". Edward Elgar Publishing, 1996.
- Lo, A.W.; McKinlay, A.C.: "Stock market prices do not follows random walks: evidence from a simple especification test". Review Financial Studies, vol. 1, 1988.
- Lo, A.W.; McKinlay, A.C.: "The size and power of the variance ratio test in finite samples: A Monte Carlo investigation". Journal of Econometrics, vol. 40, 1989.
- Loève, M.: "*Probability theory*". Springer, 1977.
- Longin, F.M.: "The asymptotic distribution of extreme stock market returns". Journal of Business, vol. 69, 1996.
- Loretan, M.; Phillips, P.C.B.: "Testing covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets". Journal of Empirical Finance, vol. 1, 1994.
- Lovell, M.C.: "Data mining". The Review of Economics and Statistics, vol. 65, 1983.
- Loynes, R.M.: "The central limit theorem for backwards martingales". Z. Wahrsch. Verw. Gebiete, vol. 13, 1969.
- Loynes, R.M.: "An invariance principle for reversed martingales". Proc. American Mathematical Society, vol. 25, 1970.
- Lukacs, E.: "*Stochastic convergence*". Academic Press, 1975.
- Lukkonen, R.: "Testing for a moving average unit root in autoregressive integrated moving average models". JASA, vol. 88, 1993.
- Lundberg, F.: "*On random processes and their applications to sickness and accident statistics*". Almqvist & Wicksell, 1940.
- Lütkepohl, H.: "*Introduction to multiple time series analysis*". Springer-Verlag, 1991.
- Lyapunov, A.M.: "Sur une proposition de la théorie des probabilités". Izv. Akad. Nauk., Serie 5, vol. 13, 1900.
- Lyapunov, A.M.: "Nouvelle forme du théoreme sur la limite de probabilité". Notes Academy Sciences Phys-Mathematics, Serie 8, vol. 12, 1901.
- Lyapunov, A.M.: "Sur un théoreme du calcul del probabilités". Comptes Rendus Academie des Sciences, vol. 132, 1901.
- Lyapunov, A.M.: "Une proposition générale du calcul des probabilités". Comptes Rendus Academie des Sciences, vol. 132, 1901.
- Maddala, G.S.: "Model selection". Journal of Econometrics, vol. 16, 1981.
- Maddala, G.S.; Rao, C.R. (eds.): "*Handbook of statistics - Vol. 14: Statistical methods in finance*". North-Holland, 1996.
- MacKinnon, J.G.: "Numerical distribution functions for unit roots and cointegration tests". Journal of Applied Econometrics, vol. 11, 1996.
- MacNeill, I.: "Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression al unknown times". Annals of Statistics, vol. 6, 1978.
- Makeham, W.M.: "On the law of mortality and the construction of annuity tables". Journal Institute of Actuaries, vol. 8, 1860.
- Makridakis, S.: "Accuracy measures: Theoretical and practical concerns". International Journal of Forecasting, vol. 9, 1993.
- Makridakis, S.; Wheelwright, S.C. (eds.): "*Forecasting studies in management sciences*". North-Holland, 1979.
- Malik, H.J.: "Estimation of the parameters of the Pareto distributions". Metrika, vol. 10, 1960.
- Malliavin, P.: "*Stochastic analysis*". Springer-Verlag, 1997.
- Malliavin, P.; Thalmaier, A.: "*Stochastic calculus of variations in mathematical finance*". Springer-Verlag, 2006.
- Manaster, S.; Koehler, G.: "The calculation of implied variances from the Black-Scholes model: A note". Journal of Finance, vol. 37, 1982.
- Mandelbrot, B.B.: "The Pareto-Lévy law and the distribution of income". International Economic Review, vol. 1, 1960.
- Mandelbrot, B.B.: "New methods in statistical economics". Journal of Political Economy, vol. 71, 1963.
- Mandelbrot, B.B.: "The variation of certain speculative prices". Journal of Business, vol. 36, 1963.
- Mandelbrot, B.B.: "The variation of some other speculative prices". Journal of Business, vol. 40, 1967.
- Mandelbrot, B.B.: "Statistical metodology for non periodic cycles: From the covariance to R/S analysis". Annals Economic and Social Measurement, vol. 1, 1972.

- Mandelbrot, B.B.: "A fast fractional Gaussian noise generator". Water Resources Research, vol. 7, 1975.
- Mandelbrot, B.B.; Taqqu, M.S.: "Robust R/S analysis of long run serial correlation". Bulletin of the International Statistical Institute, vol. 48, 1979.
- Mandelbrot, B.B.: "*The fractal geometry of nature*". Freeman, 1983.
- Mandelbrot, B.B.: "Self-affine fractal and fractal dimension". Physical Scr., vol. 32, 1985.
- Mandelbrot, B.B.: "*Fractals and the rebirth of iteration theory*". Richter & Peitgen, 1986.
- Mandelbrot, B.B.; Taylor, H.: "On the distribution of stock price differences". Operations Research, vol. 15, 1967.
- Mandelbrot, B.B.; Van Ness, J.W.: "Fractional Brownian motions, fractional Brownian noises and applications". SIAM, vol. 10, 1968.
- Mandelbrot, B.B.; Wallis, J.: "Noah, Joseph and operational hydrology". Water Resources Research, vol. 4, 1968.
- Mandelbrot, B.B.; Wallis, J.: "Computer experiments with fractional Gaussian noises". Water Resources Research, vol. 5, 1969.
- Mann, H.B.; Wald, A.: "On the statistical treatment of linear stochastic difference equations". Econometrics, vol. 11, 1943.
- Mann, N.R.: "Point and interval estimation procedures for two parameter Weibull and extreme-value distributions". Technometrics, vol. 10, 1968.
- Mardia, K.V.: "A non-parametric test for the bivariate two-sample location problem". JRSS, Serie B, vol. 29, 1967.
- Mardia, K.V.: "Multivariate Pareto distributions". Annals of Mathematical Statistics, vol. 33, 1970.
- Markov, A.A.: "*Markov's selected works*". Academy of Sciences, Moscú, 1951.
- Marsh, T.A.; Rosenfeld, E.R.: "Non-trading, market making, and estimates of stock price volatility". Journal of Financial Economics, vol. 15, 1986.
- Martin, J.J.: " *Bayesian decision problems and Markov chains*". Wiley, 1967.
- Martin, R.D.; Samorov, A.; Vandaele, W.: "Robust methods for ARIMA models". Proc. Conference in Applied Time Series Analysis of Economic Data, Arlington, 1981.
- McAleer, M.; Veall, M.R.: "Data mining and the con in econometrics. The US demand for money revisited". Mathematics and Computers in simulation, vol. 39, 1995.
- McClave, J.: "Subset autoregression". Technometrics, vol. 17, 1975.
- McClave, J.T.: "Estimating the order of moving average models: The max χ^2 method". Communications in Statistics, vol. A7, 1978.
- McElroy, F.W.: "A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased". JASA, vol. 62, 1967.
- McGabe, B.; Tremayne, A.: "*Elements of modern asymptotic theory with statistical applications*". Manchester University Press, 1993.
- McKay, A.T.: "Sampling from batches". JRSS, Serie B, vol. 1, 1934.
- McKendrick, A.G.: "Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of progressive nature". Proceedings of London Mathematical Society, Serie II, vol. 13, 1914.
- McLeish, D.L.: "A maximal inequality and dependent strong laws". Annals of Probability, vol. 3, 1975.
- McLeod, A.I.; Hipel, K.W.: "Preservation of the rescaled adjusted range, 1: A reassessment of the Hurst phenomenon". Water Resources Research, vol. 14, 1978.
- McQueen, J.B.: "A linear extension of the martingale convergence theorem". Annnals of Probability, vol. 1, 1973.
- Melino, A.; Turnbull, S.M.: "Pricing foreign currency options with stochastic volatility". Journal of Econometrics, vol. 45, 1990.
- Menon, M.V.: "Estimation of the shape and scale parameters of the Weibull distribution". Technometrics, vol. 5, 1963.
- Merrington, M.: "Numerical approximation to percentage points of the χ^2 distribution". Biometrika, vol. 32, 1941.
- Merton, R.C.: "An analytic derivation of the efficient portfolio frontier". Journal of Financial and Quantitative Analysis, vol. 7, 1972.
- Merton, R.C.: "An intertemporal capital asset pricing model". Econometrica, vol. 41, 1973.
- Merton, R.C.: "Rational theory of option pricing". Bell Journal of Economics and Management Sciences, vol. 4, 1973.
- Merton, R.C.: "Option pricing when underlying stock returns are discontinuous". Journal of Financial Economics, vol. 3, 1976.
- Merton, R.C.: "*Continuous-time finance*". Blackwell, 1990.
- Metropolis, N.; Ulam, S.: "The Monte Carlo method". JASA, vol. 44, 1949.
- Mikhail, W.M.: "A comparative Monte Carlo study of the properties of econometric estimators". JASA, vol. 70, 1975.
- Mikosch, T.: "*Elementary stochastic calculus*". World Scientific, 2006.
- Miller, D.W.: "Degree of normality of maximumlikelihood estimates of the shape parameters of the Weibull failure distributions". Technometrics, vol. 5, 1966.
- Miller, I.; Freund, J.E.: "*Probability and statistics for engineers*". Englewood-Cliffs, 1965.
- Mills, T.C.: "*Time series techniques for economists*". Cambridge University Press, 1990.

- Mills, T.C.: "Modelling skewness and kurtosis in the London Stock Exchange FTSE index return distributions". The Statistician, vol. 44, 1995.
- Mintz, S.L.; Perlmutter, A. (eds.): "*Information processing in biological systems*". Proc. of the 20th. Annual Orbis Scientiae, 1985.
- Mittnik, S.; Rachev, S.T.: "Modeling asset returns with alternative stable distributions". Econometric Reviews, vol. 12, 1993.
- Mizon, G.E.; Hendry, D.F.: "An empirical application and Monte Carlo analysis of tests of dynamic specification". Review of Economis Studies, vol. 47, 1980.
- Molenaar, W.: "Survey of separation methods for two normal distributions". Statistica Neerlandica, vol. 19, 1965.
- Monroe, I.: "On embedding right continuous martingales in Brownian motion". Annals of Mathematical Statistics, vol. 43, 1972.
- Monroe, I.: "Processes that can be embedded in Brownian motion". Annals of Probability, vol. 6, 1978.
- Mood, A.: "the distribution theory of runs". Annals of Mathematical Statistics, vol. 11, 1940.
- Moran, P.A.P.: "Estimation methods for evolutive processes". JRSS, Serie B, vol. 13, 1951.
- Moran, P.A.P.: "The estimation of the parameters of a birth and death process", JRSS, Serie B, vol. 15, 1953.
- Moran, P.A.P.: "*An introduction to probability theory*". Oxford University Press, 1968.
- Moore, P.G.; Tukey, J.W.: "Answer to query 112". Biometrics, vol. 10, 1954.
- Morgan, M.S.: "*The history of econometric ideas*". Cambridge University Press, 1990.
- Morgan, W.A.: "A test for the significance of the difference between the two variances in a sample from a Normal bivariate population". Biometrika, vol. 31, 1939-40.
- Muirhead, R.J.: "*Aspects of multivariate statistical theory*". Wiley, 1982.
- Muniruzzaman, A.N.M.: "On measures of location and dispersion an tests if hypotheses on a Pareto population". Bulletin of Calcutta Statistical Association, vol. 7, 1957.
- Murphy, J.: "*Technical analysis of the future markets*". New York Institute of Finance, 1986.
- Muth, J.F.: "Optimal properties of exponentially weighted forecasts". JASA, vol. 55, 1960.
- Nabeya, S.; Tanaka, K.: "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative". Annals of Statisitcs, vol. 16, 1988.
- Nagel, E. (ed.): "*Logic, metodology and the philosophy of science*". Stanford University Press, 1962.
- Nankervis, J.C.; Savin, N.E.; Whiteman, C.H.: "The power problems of unit root tests in time series with autoregressive errors". Journal of Econometrics, vol. 53, 1992.
- Neave, H.R.: "Comparison of lag windows generators". JASA, vol. 67, 1972.
- Neftci, S.: "Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: A study of 'technical analysis'". Journal of Business, vol. 64, 1991.
- Nelson, C.R.: "The interpretation of R^2 in autoregressive-moving average time series models". American Statistician, vol. 30, 1976.
- Nelson, C.R.; Plosser, C.I.: "Trends and random walks in macroeconomic time series: Some evidence and implications". Journal Monetary Economics, vol.10, 1982.
- Nelson, H.L.; Granger, G.W.J.: "Experience with using the Box-Cox transformation when forecasting economic time series". Journal of Econometrics, vol. 10, 1979.
- Nelson, P.I.: " A class of orthogonal series related to martingales". Annals of Mathematical Statistics, vol. 41, 1970.
- Nerlove, M.: "*Analysis of ecnomic time series*". Academic Press, 1979.
- Nerlove, M.; Wallis, K.F.: "Use of the Durbin-Watson statistic in inappropiate situations". Econometrica, vol. 34, 1966.
- Neudecker, H.: "Abrahamse and Koerts new estimator of disurbances in regression analysis". Journal of Econometrics, vol. 5, 1977.
- Neuhaus, G.; Zhu, X.: "Permutation tests for multivariate location problems". Jounal of Multivariate Analysis, vol. 69, 1999.
- Neveu, J.: "*Mathematical foundations of the calculus of probability*". Holden-Day, 1965.
- Neveu, J.: "*Martingales à temps discret*". Masson, 1972.
- Newbold, P.: "The exact likelihood function for a mixed autoregressive-moving average process". Biometrika, vol. 61, 1974.
- Newbold, P.; Agiakloglou, C.: "Bias in the sample autocorrelations of fractional white noise". Biometrika, vol. 80, 1993.
- Newey, W.K.; West, K.D.: "A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix". Econometrica, vol. 55, 1987.
- Newman, P.; Milgate, M.; Eatwell, J. (eds.): "*New Palgrave dictionary of money and finance*". MacMillan, 1992.
- Neyman, J.: "On a new class of contagious distribution, applicable in entomology and bacteriology". Annals of Mathematical Statistics, vol. 10, 1939.

- Ng, S.; Perron, P.: "Unit roots in ARIMA models with data dependent methods for the selection of the truncation lag". JASA, vol. 72, 1995.
- Nguyen, T.T.; Dinh, K.T.: "Characterizations of Normal distributions and EDF goodness-of-fit tests". Metrika, vol. 58, 2003.
- Nisio, M.: "On polynomial approximation for strictly stationary processes". Journal of Mathematical Society of Japan, vol. 12, 1960.
- Nisio, M.: "Remarks on the canonical representation of strictly stationary time series". American Statistician, vol. 37, 1961.
- Nualart, D.: "*The Malliavin calculus and related topics*". Springer-Verlag, 1995.
- Officer, R.: "The distribution of stock returns". JASA, vol. 67, 1972.
- Øksendahl, B.: "*Stochastic differential equations: An introduction with applications*". Springer, 1985.
- Oldfield, G.S.; Rogalski, R.J.; Jarrow, R.A.: "An autorregresive jump process for common stock returns". Journal of Financial Economics, vol. 5, 1977.
- Olkin, I. (ed.): "*Probability and statistics: Essays in honor of Harold Hotelling*". Stanford University Press, 1960.
- Orcutt, G.H.; Cochrane, D.: "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis". JASA, vol. 44, 1949.
- Ord, J.K.: "*Families of frequency distributions*". Griffin, 1972.
- Ouspensky, P.D.: "*Tertium organum. El tercer canon del pensamiento*"- Kier, 1982.
- Ozaki, T.: "On the order determination of ARIMA models". Applied Statistics, vol. 26, 1977.
- Pagan A.; Nicholls, D.: "Estimating prediction erros and their standard deviations using constructed variables". Journal of Econometrics, vol. 24, 1984.
- Pantula, S.G.; Hall, A.: "Testing for a unit roots in autoregressive moving average models: An instrumental variable approach". Journal of Econometrics, vol. 48, 1991.
- Pareto, V.: "*Cours d'economie politique*". Rouge et Cie., 1897.
- Park, J.Y.; Choi, B.: "A new approach to testing for a unit root". Working Paper nº 88-23, Cornell University, 1988.
- Park, J.Y.; Phillips, P.C.B.: "Statistical inference in regression with integrated processes: Part I". Econometric Theory, vol. 4, 1988.
- Parkinson, M.: "The extreme value method for estimating the variance of the rate of return". Journal of Business, vol. 53, 1980.
- Parzen E.: "*Stochastic processes*". Holden-Day, 1962.
- Parzen , E.: "*Time series analysis papers*". Holden-Day, 1967.
- Parzen, E.: "An approach to time series analysis". IEEE Transactions on Automatic Control, AC-19, 1974.
- Parzen, E.: "Some recent advances in time series modeling". IEEE Transactions on Automatic Control, AC-19, 1974.
- Parzen, E.: "ARARMA models for time series analysis and forecasting". Journal of Forecasting, vol.1, 1982.
- Patel, J.K.; Kapadia, C.H.; Owen, D.B.: "*Handbook of statisticas distributions*". Dekker, 1976.
- Patil, V.T.: "The consistency and adequacy of the Poisson-Markoff model for density fluctuations". Biometrika, vol. 44, 1957.
- Patnaik, P.B.: "The non-central χ^2 and F distributions and their applications". Biometrika, vol. 36, 1949.
- Pau, L.: "Technical analysis of portfolio trading by syntactic pattern recognition". Journal of Economic Dynamics and Control, vol. 15, 1991.
- Pearl, R.: "*Medical biometry and statistics*". Sanders, 1940.
- Pearson, E.S.: "Comparison of tests for randomness of points on a line". Biometrika, vol. 50, 1963.
- Pearson, E.S.; Hartley, H.O.: "*Biometrika-Tables for statisticians*". Cambridge University Press, 1972.
- Pearson, E.S.; Stephens, M.A.: "The goodness of fit tests based on W_N^2 y U_N^2". Biometrika, vol. 49, 1962.
- Pearson, K.: "Contributions to the mathematical theory of evolution, I: On the dissection of asymetrical frequency curves". Philosophical Transactions of the Royal Society, Serie A, vol. 185, 1894.
- Pearson, K.: "Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material". Philosophical Transactions of the Royal Society, Serie A, vol. 186, 1895.
- Pearson, K.: "Contributions to the mathematical theory of evolution, III: Regression, heredity and panmixia". Philosophical Transactions of the Royal Societey, Serie A, vol. 187, 1896.
- Pearson, K.: "On the dissection og asymetrical frequency curves". Philosophical Transactions of the Royal Society, Serie A, vol. 185, 1894.
- Pearson, K.: "On a criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen in random sample".Philosophical Magazine, vol. 50, 1900.
- Pearson, K.: "Contributions to the mathematical theory of evolution, VII: On the correlation of characters not quantitatively measurable". Philosophical Transactions of the Royal Society, Serie A, vol. 195, 1901.
- Pearson, K.: "On the systematic fitting of curves to observations an measurement". Biometrika, vol.1, 1902.

- Pearson, K.: "Contributions to the mathematical theory of evolution, VIII: On homotyposis in homologous but differentiated organs". Proceedings of the Royal Society, vol. 71, 1903.
- Pearson, K.: "Contributions to the mathematical theory of evolution, XIII: On contingency and its relation to association and normal correlation". Drapers' Company Reasearch Memoirs, Biometric Sereies, I, 1904.
- Pearson, K.: "Contributions to the mathematical theory of evolution, XIV: On the general theory of skew correlation and non-linear regression". Drapers' Company Research Memoirs, Biometric Series, II, 1905.
- Pearson, K.: "On a new method of determining correlation when one variable is given by alternative and the other by multiple categories". Biometrika, vol. 7, 1910.
- Pearson, K.: "Notes on the history of correlation". Biometrika, vol. 13, 1920.
- Pearson, K.: "Notes on skew frequency surfaces". Biometrika, vol. 15, 1923.
- Pearson, K.: "Notes on non-skew frequency surfaces". Biometrika, vol. 15, 1923.
- Pearson, K.: "On a certain double hypergeometric series and its representation by continuous frequency surfaces". Biometrika, vol. 16, 1924.
- Pearson, K.: "The fifteen constant bivariate frequency surface". Biometrika, vol. 17, 1925.
- Perks, W.F.: "On some experiments in the graduation of mortality statistics". Journal Institute of Actuaries, vol. 58, 1932.
- Perron, P.: "Trends and random walks in macroeconomic time series: Further evidence from a new approach". Journal of Economic Dynamics and Control, vol. 12, 1988.
- Perron, P.: "Test consistency with varying sampling frequency". Econometric Theory, vol. 7, 1991.
- Perron, P.; Ng, S.: "Useful modifications to som unit root tests with dependent errors and their local asymptotic properties". Review of Economic Studies, vol. 63, 1996.
- Petrov, B.N.; Csáki, F. (eds.): "Proc. 2nd. International Symposium on Information Theory". Akadémiai Kiadó, Budapest, 1973.
- Phillips, G.D.A.; Harvey, A.C.: "A simple test for serial correlation in regression analysis". JASA, vol. 69, 1974.
- Phillips, P.C.B.: "Understanding spurius regressions in econometrics". Journal of Econometrics, vol.33, 1986.
- Phillips, P.C.B.: "Time series regression with a unit root". Econometrica, vol. 55, 1987.
- Phillips, P.C.B.; Durlauf, S.N.: "Multiple time series regression with integrated processes". Review Economic Studies, vol. 53, 1986.
- Phillips, P.C.B.; McFarland, J.W.; McMahon, P.C.: "Robust tests of forward exchange market efficiency with empirical evidence from the 1920's". Journal Applied Econometrics, vol. 11, 1996.
- Phillips, P.C.B.; Ouliaris, S.: "Testing for cointegration using principal components methods". Journal of Economic Dynamics and Control, vol. 12, 1988.
- Phillips, P.C.B.; Perron , P.: "Testing for a unit root in time series regression". Biometrika, vol. 75, 1988.
- Pickands, J.: "Statistical inference using extreme order statistics". Annals of Statistics, vol. 3, 1975.
- Pierce, D.; Haugh, L.: "Causality in temporal systems". Journal of Econometrics, vol. 5, 1977.
- Pierse, R.G.: Snell, A.G.: "Temporal aggregation and the power of tests for a unit root". Journal of Econometrics, vol. 65, 1995.
- Pigou, A.C.: "*The economic of Welfare*". MacMillan, 1932.
- Pindyck, R.S.; Rubinfeld, D.L.: "*Econometric models and economic forecasts*". McGraw-Hill, 1991.
- Plackett, R.L.: "Linear estimation from censored data". Annals of Mathematical Statistics, vol. 29, 1958.
- Plosser, C.I.; Schwert, G.W.: "Money, income and suspots: Measuring economic relationships and the effects of differencing". Journal of Monetary Economics, vol.4, 1978.
- Poincaré, H.: "*Calcul des probabilités*". Gauthier-Villars, 1912.
- Poisson, S.D.: "*Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédes des règles générales du calcul des probabilités*". Bachelier, Paris, 1837.
- Pólya, G.: "Sur quelque points de la théorie des probabilités". Annals Institute H. Poincaré, vol. 1, 1931.
- Pólyam G.; Szego, G.: "*Aufgaben und Lehrsätze aus der Analysis*". Springer-Verlag, 1964.
-Poole, W.: "Speculative prices as a random walk: An analysis of ten time series of flexible exchange rates". Southern Economic Journal, vol. 33, 1967.
- Poskitt, D.S.; Tremayne, A.R.: "The selection and use of linear and bilinear time series models". International Journal of Forecasting, vol. 2, 1986.
- Poskitt, D.S.; Tremayne, A.R.: "Determining a portfolio of linear time series models". Biometrika, vol. 74, 1987.
- Poterba, J.M.; Summers, L.H.: "The persistence of volatility and stock market fluctuations". American Economic Review, vol. 76, 1986.
- Pötscher, B.M.: "Order estimation of ARMA systems by the Lagrange multiplier test". Institut für Ökonometrie und Operations Research, Technische Universität Wien, Informe n°5, 1982.
- Praetz, P.D.: "The distribution of share prices changes". Journal of Business, vol. 45, 1972.

- Praetz, P.D.: "The detection of omitted variables by Durbin-Watson statistics in multiple regression models". Australian Journal of Statistics, vol. 21, 1979.
- Prakasa Rao, B.L.S.: "Statistical inference for stochastic processes". Technical Report CRM-465, Centre de Recherches Mathèmatiques, Université de Montreal, 1974.
- Prakasa Rao, B.L:S.: "On some problems in inference for Markov processes". Proceedings of General Mathematical Seminar, University of Patras, 1978.
- Prasad, A.: "Bi-modla distributions derived from the normal distribution". Sankhyā, vol. 14, 1955.
- Press, S.J.: "A compound events model for security prices". Journal of Business, vol. 40, 1967.
- Pretorius, S.J.: "Skew bivariate frequency surfaces, examined in the light of numerical illustrations". Biometrika, vol. 22, 1930.
- Priestley, M.B.: "***Spectral analysis and time series***". Academic Press, 1981.
- Prigent, J.L.: "*Weak convergence of financial markets*". Springer, 2003.
- Prigogine I.: "*Nonequilibrium statistical mechanics*". Wiley, 1962.
- Progogine, I.: "*From being to becoming: Time and complexity in the physical science*". Freeman, 1980.
- Prigogine, I.: "*Les lois des chaos*". Flamarion, 1994.
- Prigogine, I.: "*L'homme devant l'incertain*". Odile Jacob, 2001.
- Prigogine, I.: "*Is future given?*". World Scientific, 2003.
- Prigogine, I.; Nicolis, G.: "*Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*". Wiley, 1977.
- Prigogine, I.; Nicolis, G.: "*Exploring complexity*". Freeman, 1989.
- Prigogine, I.; Stengers, I.: "*La nouvelle alliance. Ler mètamorphoses de la science*". Gallimard, 1979.
- Prigogine, I.; Stengers, I.: "*Entre le temps et l'eternité*". Librairie Fayard, 1988.
- Prigogine, I.; Stengers, I.: "*The end of certainty. Time's flow and the laws of nature*". Free Press, 1997.
- Prokhorov, Y.V.: "Convergence of random processes and limit theorems in probability theory". Theory of Probability and Applications, vol. 1, 1956.
- Protter, P.: "*Stochastic integration and differential equations*". Springer, 1990.
- Puri, P.S.: "Inteconnected birth and death processes". Journal of Applied Probability, vol. 5, 1968.
- Pynn, R.; Skjeltrop, A. (eds.): "*Scaling phenomena in disordered systems*". Plenum Press, 1985.
- Quandt, R.E.: "Old and new methods of estimation on the Pareto distribution". Metrika, vol. 16, 1966.
- Quenouille, M.H.: "A large-sample test for the goodness of fit of autorregresive schemes". JRSS, vol. 110, 1947.
- Quenouille, M.H.: "Approximate tests of correlation in time series". JRSS, Serie B, vol. 11, 1949.
- Quenouille, M.H.: "*Introductory statistics*". Butterworth, 1950.
- Quenouille, M.H.: "*Fundamentals of statistical reasoning*". Griffin's Statistical Monografs and Courses, vol. 3 1965.
- Quetelet, L.A.J.: "*Letters on the theory of probability*". Layton, 1849.
- Raiffa, H.; Schlaifer, R.: "*Applied statistical decision theory*". Harvard University Press, 1961.
- Rainville, E.D.: "***Special functions***". MacMillan, 1960.
- Ramsey, J.B.: "Tests for specification errors in classical linear least square regression analysis". JRSS, Serie B, vol. 31, 1969.
- Ramsey, J.B.; Gilbert, R.: "A Monte Carlo study of some small sample properties of tests for specification error". JASA, vol. 67, 1972.
- Ramsey, J.B.; Schmidt, P.: "Some further results on the use of OLS y BLUS residuals in specification error tests". JASA, vol. 71, 1976.
- Randles, R.; Wolfe, D.: "*Introduction to the theory of nonparametric statistics*". Wiley, 1979.
- Rao, C.R.: "Estimation of heteroscedastic variances in linear models". JASA, vol. 67, 1970.
- Rao, C.R.: "*Linear statistical inference and its applications*". Wiley, 1973.
- Rappoport, P.; Reichlin, I,: "Segmented trends and non-stationary time series". Economic Journal, vol. 99, 1989.
- Refenes, A.; Abu-Mustafa, Y.; Moody, J.; Weigenda, A. (eds.): "*Neural networks in financial engineering*". Proceedings 3rd. Inrternational Conference on Neural Networks in the Capital Markets, World Scientific, 1996.
- Reichenbach, H.: "Stetige Wahrscheinlichkeitsrechnung". Zeitschrift für Physik, vol. 53, 1929.
- Reichenbach, H.: "*The direction of time*". University of California Press, 1956.
- Reinsel, G.C.: "*Elements of multivariate time series analysis*". Springer-Verlag, 1993.
- Renshaw, E.: "Interconnected population processes". Journal of Applied Probability, vol. 10, 1973.
- Renton, G.A. (ed.): "*Modelling the economy*". Heinemann, 1975.
- Rényi, A.: "On the distribution function L(z)". Selected Translations in Mathematical Statistics and Probability, vol. 4, American Mathematical Society, 1963.
- Resnick, S.I.: "*Extreme values, regular variation and point processes*". Springer-Verlag, 1987.
- Resnick, S.I.: "*Adventures in stochastic processes*". Birkhäuser, 1992.

- Reuter, G.E.H.; Lederman, W.: "On the differential equations for the transition probabilities of Markov processes with enumerably many states". Proc. Cambridge Philosophical Society, vol. 49, 1953.
- Revuz, D.; Yor, M.: "*Continuous martingales and Brownian motion*". Springer, 1991.
- Rhodes, G.F.; Fomby, T.B. (eds.): "*Advances in econometrics*". JAI Press, 1990.
- Richard, J.F.: "Models with several regimes and changes in exogeneity". Review of Economic Studies, vol. 47, 1980.
- Richardson, M.: "Temporary components of stock prices: A skeptic's view". Journal of Business and Economic Statistics, vol. 11, 1993.
- Richardson, M.; Smith, T.: "A test for multivariate Normality in stock returns". Journal of Business, vol. 66, 1993.
- Richardson, M.; Smith, T.: "A unified approach to testing for serial correlation in stock returns". Journal of Business, vol. 67, 1994.
- Richardson, M.; Stock, J.: "Drawing inferences from statistics based on multi-year asset returns". Journal of Financial Economics, vol. 25, 1989.
- Rissanen, J.: "Modeling by shortest data description". Automatica, vol. 14, 1978.
- Roberts, H.: "Statistical versus clinical prediction of the stock market". Center for Research in Security Prices, University of Chicago, 1967.
- Robinson, P.M.: "Testing for strong serial correlation and dynamic conditional heterokedasticity in multiple regression". Journal of Econometrics, vol. 47, 1991.
- Robinson, P.M.: "Log-periodogram regression of time series with long-run dependence". Annals of Statistics, vol. 23, 1995.
- Rogers, L.C.G.; Satchell, S.E.: "Estimating variance from high, low and closing prices". Annals of Applied Probability, vol. 1, 1991.
- Romano, J.; Thombs, L.: "Inference for autocorrelations under weak assumptions". JASA, vol. 91, 1996.
- Rootzén, H.: "On the functional central limit theorem for martingales, I". Zeitschrift für Wahrscheinleichkeits Theorie und Verwandte Gebiete, vol. 38, 1977.
- Rootzén, H.: "On the functional central limit theorem for martingales, II". Zeitschrift für Wahrscheinleichkeits Theorie und Verwandte Gebiete, vol. 41, 1980.
- Rosén, B.: "On the central limit theorem for sums of dependent random variables". Zeitschrift für Wahrscheinleichkeits Theorie und Verwandte Gebiete, vol. 7, 1967.
- Rosén, B.: "On asymptotic normality of sums of dependent random variables". Zeitschrift für Wahrscheinleichkeits Theorie und Verwandte Gebiete, vol. 7, 1967.
- Rosén, B.: "On the central limit theorem fo a class of smapling procedures". Zeitschrift für Wahrscheinleichkeits Theorie und Verwandte Gebiete, vol. 7, 1967.
- Rosen, R.: "Church's thesis and its relation to the concept of reliability in biology and physics". Bulletin of Mathematical Biophysics, vol. 24, 1962.
- Rosen, R.: "*Anticipatory systems*". Pergamon Press, 1985.
- Rosen, R.: "Complexity and information". Journal of Computation and Applied Mathematics, vol. 22, 1988.
- Rosenbaum, P.; Rubin, D.: "The central role of propensity score in observational studies for causal effects". Biometrika, vol. 70, 1983.
- Rosenblatt, M.: "A central limit theorem and strong mixing condition". Proceedings of National Academy of Sciences, vol. 42, 1956.
- Rosenblatt, M.: "*Markov processes. Structure and asymptotic behavior*". Springer-Verlag, 1971.
- Rosenthal, J.S.: "*A first look at rogorous probability theory*". World Scientific, 2006.
- Rozanov, Y.A.: "*Stationary random processes*". Holden-Day, 1967.
- Ruelle, D.; Takens, F.: "On the nature of turbulence". Communications Mathematical Physiscs, vol. 20, 1971.
- Said, S.E.; Dickey, D.A.: "Testing for unit-roots in autorregressive-moving average models with unknown order". Biometrika, vol. 71, 1984.
- Said, S.E.; Dickey, D.A.: "Hypothesis testing in ARIMA (p,1,q) models". JASA, vol. 80, 1985.
- Sakai, H.; Tokumaru, H.: "Autocorrelations of certain chaos". IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 1, 1980.
- Salkever, D.: "The use of dummy variables to compute predictions, prediction errors, and confident intervals". Journal of Econometrics, vol. 4, 1976.
- Samorodnitsky, G.; Taqqu, M.S.: "***Stable non-Gaussian random processes***". Chapman & Hall, 1994.
- Samuelson, P.A.: "Proof that properly anticipated prices fluctuate randomly". Industrial Management Review, vol. 6, 1965.
- Samuelson, P.A.: "Rational theory of warrant pricing". Industrial Management Review, vol. 6, 1965.
- Samuelson, P.A.: "Efficient portfolio selection for Pareto-Lévy investments". Journal of Financial and Quantitative Analysis, vol. 2, 1967.

- Samuelson, P.A.: "Proof that properly discounted present values of assets vibrate randomly". Bell Journal, vol. 4, 1973.
- Samuelson, P.A.: "Limited liability, short selling, bounded utility and infinity stable distributions". Journal of Financial and Quantitative Analysis, vol. 12, 1976.
- Sargan, J.D.; Bhargava, A.: "Testing residuals from least squares regression for being generated by de Gaussian random walk". Econometrica, vol. 51, 1983.
- Sarhan, A.E.: "Estimation of the mean and standard deviation by order statistics-Part I". Annals of Mathematical Statistics, vol. 25, 1954.
- Savin, N.E.; White, K.J.: "The Durbin-Watson test for serial correlation with extreme sample size or many regressors". Econometrica, vol. 45, 1977.
- Scala, C.: "La funzione di potenza di un test di omogeneità fra indici paretiani". Quaderni dell'Istituto di Statistica, Università di Roma, vol. 2, 1967.
- Scala, C.: "La distribuzione delle stime campionarie di massima verosimiglianza dei parametri di universi paretiani nel product testing". Quaderni dell'Istituto di Statistica, Università di Roma, vol. 3, 1967.
- Schmidt, P.: "*Econometrics*". Dekker, 1976.
- Schmidt, P.; Phillips, P.C.B.: "LM test for a unit root in the presence of deterministic trends". Oxford Bulletin of Economics and Statistics, vol. 54, 1992.
- Schneider, I.: "Rudolph Clausius' erster Beitrag zur Einführung Wahrscheinlichkeitstheoretischer Methoden in der Physisk nach 1856". Archive for History of Exact Sciences, vol. 4, 1974/75.
- Schwarz, G.: "Estimating the dimension of a model". Annals of Statistics, vol. 6, 1978.
- Schwert, G.W.: "Effects of model specification on tests for unit roots in macroeconomic data". Journal of Monetary Economics, vol.20, 1987.
- Scott, L.O.: "Option pricing when the variance changes randomly: Theory, estimation and an application". Journal of Financial Quantitative Analysis, vol. 22, 1987.
- Scott, L.O.: "Random-variance option pricing: Empirical tests of the model and delta-sigma hedging". Advances in Futures and Options Research, vol. 5, 1991.
- Scott, M.F.G.: "*Relative share prices and yields*". Oxford Economic Papers, Oct-1962.
- Seier, E.: "Comparison of tests for univariate Normality". International Statistics, vol. 1, 2002.
- Serfling, R. : "*Approximation theorems of mathematical statistics*". Wiley, 1980.
- Shaman, P.: "An approximate inverse for the covariance matrix of moving average and autorregressive processes". Annals of Statistics, vol. 3, 1975.
- Shaman, P.: "Approximations for stationary covariances matrices and their inverses with applications to ARMA models". Annals of Statistics, vol. 4, 1976.
- Shanker, S. (ed.): "*Gödel's theorem in focus*". Croom & Helm, 1988.
- Shapiro, S.S.; Gross, A.J.: "*Statistical modeling techniques*". Dekker, 1981.
- Shapiro, S.S.; Wilk, M.B.: "An analysis of variance test for normality". Biometrika, vol. 52, 1965.
- Shibata, R.: "Selection of the order on an autoregressive model by Akaikes's information criterion". Biometrika, vol. 63, 1976.
- Shibata, R.: "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process". The Annals of Statistics, vol. 8, 1980.
- Shih, Y.L.: "Neural nets in technical analysis". Technical Analysis of Stocks and Commodities, vol. 9, 1991.
- Shiller, R.J.: "Testing the random walk hypothesis: Power versus frequency of observation". Economic Letters, vol. 18, 1985.
- Shoham, Y.: "*Reasoning about change: time and causation from the standpoint of artificial intelligent*". MIT Press, 1988.
- Shreve, S.E.: "*Stochastic calculus for finance*". Springer, 2004.
- Siddiqui, M.: "The asymptotic distribution of the range and other functions of partial sums of stationary processes". Water Resources Research, vol. 12, 1976.
- Simkowitz, M.; Beedles, W.: "Diversification in a three moment world". Journal of Financial and Quantitative Analysis, vol. 13, 1978.
- Simkowitz, M.; Beedles, W.: "Asymmetric stable distribuited security returns". JASA, vol. 75, 1980.
- Simon, H.: "On the definition of the causal relation". Journal of Philosophy, vol. 49, 1952.
- Sims, C.A.: "Money, income and causality". American Economic Review, vol. 62, 1972.
- Sims, C.A.: "A note on exact tests for serial correlation". JASA, vol. 70, 1975.
- Sims, C.A.: "*New methods in business cycle research: Proceedings from a conference*". Federal Reserve Bank of Minneapolis, 1977.
- Sims, C.A.: "Macroeconomis and reality". Econometrica, vol.48, 1980.

- Singleton, J.C.; Wingender, J.: "Skewness persistence in common stock returns". Journal of Financial and Quantitative Analysis, vol. 21, 1986.
- Skinner, D.J.: "Options markets and stock return volatility". Journal of Financial Economics, vol. 23, 1989.
- Slutzsky, E.: "Über Stochastiche Asymptotem und Grenzwerte". Metron, vol. 5, 1925.

- Slutzsky, E.: "Sur un théoreme limite relatif aux séries des quantités éventuelles". Comptes Rendus Academie des Sciences, vol. 185, 1927.
- Slutzsky, E.: "Sur l'extension de la theorie de periodogrammes aux suites des quantités dependentes". Comptes Rendus Academie des Sciences, vol. 189, 1929.
- Slutzsky, E.: "Alcuni applicazioni di coefficienti di Fourier al analizi di sequenze eventuali coherenti stazionarii". Giornale Istituto Italiano degli Attuari, vol. 5, 1934.
- Slutzsky, E.: "The summation of random causes as the source of cyclic processes". Econometrica, vol. 5, 1937.
- Smirnov, N.: "Über die Verteiling des allgmeinen Gliedes in der Variationsreine". Metron, vol. 12, 1935.
- Smirnov, N.: "On the estimation of the discrepancy between empirical curves of distribution for two independent samples". Boletín Matemático, Universidad de Moscú, vol. 2, 1939.
- Smirnov, N.: "On the Cramér-von Mises criterion". Uspekhi Matematischeskikh Nauk (N.S.), vol. 4, 1949 (original en ruso).
- Smith, V.K.: "The estimated power of several tests for autocorrelation with non-first-order alternatives". JASA, vol. 71, 1976.
- Snedecor, G.W.; Cochran, W.G.: "*Statistical methods*". Iowa State University Press, 1980.
- Sobol', I.M.: "*The Monte Carlo method*". University of Chicago Press, 1974.
- Sowell, F.B.; "*Fractionally integrated vector time series*". Duke University, 1986.
- Sowell, F.B.: "The fractional unit root distribution". Econometrica, vol. 58, 1990.
- Sowell, F.B.: "Maximum likelihood estimation of stationary univariate fractionelly integrated time series models". Journal of Econometrics, vol. 53, 1992.
- Spanos, A.: "*Statistical foundations of econometric modelling*". Cambridge University Press, 1986.
- Spanos, A.: "On the theory testing in econometrics: modelling with nonexperimental data". Journal of Econometrics, vol. 67, 1995.
- Spanos, A.: "On Normality and the linear regression model". Economic Reviews, vol. 14, 1995.
- Spanos, A.: "*Probability theory and statistical inference*". Cambridge University Press, 1999.
- Spearman, C.: "The proof and measurement of association between two things". American Journal of Psychology, vol. 15, 19094.
- Srinivasan, S.K.; Mehata, K.M.: "*Stochastic processes*". McGraw-Hill, 1988.
- Stephens, M.A.: "The distribution of the goodness-of-fit statistic $U^2_{(n)}$, I". Biometrika, vol. 50, 1963.
- Stephens, M.A.: "Random walk on a circle". Biometrika, vol. 50, 1963.
- Stephens, M.A.: "The distribution of the goodness-of-fit statistic $U^2_{(n)}$, II". Biometrika, vol. 51, 1964.
- Steindl, J.: "*Random processes and the growth of firms*". Hafner, 1965.
- Stigler, S.M.: "Mathematical statistics in the early states". American Statistician, vol. 6, 1978.
- Stigler, S.M.: "Stigler's law of eponymy". Transactions of the New Yok Academy of Sciences, Serie II, vol. 39, 1980.
- Stigler, S.M.: "*Statistics on the table*". Harvard University Press, 1999.
- Stigler, S.M.: "*American contributions to mathematical statistics in the nineteenth century*". ARNO, 1980.
- Stock, J.H.: "Asymptotic properties of least squares estimators of cointegrating vectors". Econometrica, vol.55, 1987.
- Stock, J.H.; Watson, M.W.: "Testing for common trends". JASA, vol. 83, 1988.
- Stone, M.: "Comments on model selection criteria of Akaike and Schwarz". JRSS, Serie B, vol. 41, 1979.
- Stout, W.E.: "The Hartman-Wintner law of the iterated logarithm for martingales". Annals of Mathematical Statistics, vol. 41, 1970.
- Stout, W.E.: "*Almost sure convergence*". Academic Press, 1974.
- Strang, G.: "*Linear algebra and its applications*". Harcourt-Brace.Jovanovich, 1980.
- Strasse, V.: "Almost sure behaviour of sums of independent random variables and martingales". Proc. 5th. Berkeley Symposium on Mthematics, Statistics and Probability, vol. 2, 1967.
- Stratonovich, R.L.: "A new representation for stochastic integrals and equations". SIAM Journal of Control, vol. 4, 1966.
- Student: "On the probable error of the mean". Biometrika, vol. 6, 1908.
- Subba-Rao, T.: "The fitting of non-stationary time series with time-dependent parameters". JRSS, Serie B, vol. 32, 1970.
- Suppes, P.: "*A probabilistic theory of causality*". North-Holland, 1970.
- Suppes, P.; Zanotti, M.: "When are probabilistic explanations possible?". Syntheses, vol. 48, 1981.

- Swales, G.S.; Yoon, Y.: "Applying artificial neural networks to investment analysis". Financial Analysts Journal, vol. 48, 1992.
- Szroeter, J.: "A class of parametric tests for heteroscedasticity in linear econometric models". Econometrica, vol. 46, 1978.
- Taillie, C.; Patil, G.P.; Baldessari, B.A. (eds.): "*Statistical distributions in scientific work*". Reidel-Dodrecht, 1981.
- Talacko, J.: "Perks' distributions and their role in the theory of Wiener's stochastic variables". Trabajos de Estadística, vol. 7, 1956.
- Tanaka, K.: "Testing for a moving average unit root". Econometric Theory, vol. 6, 1990.
- Taqqu, M.S.: "Weak convergence to fractional Brownian motion and to the Rosenblatt process". Zeitschrift für Wahrscheinleichkeits theorie und Verwandte Gebiete, vol. 31, 1975.
- Taqqu, M.S.: "Law of iterated logarithms for sums of non-linear functions of Gaussian variables the exhibit a long range dependence". Zeitschrift für Wahrscheinleichkeits theorie und Verwandte Gebiete, vol. 40, 1977.
- Tauchen, G.E.; Pitts, M.: "The price variability-volume relationship on speculative markets". Econometrica, vol. 51, 1983.
- Taylor, B.J.R.: "The analysis of polymodla frequency distributions", Journal of Animal Ecology, vol. 34, 1965.
- Taylor, M.; Allen, H.: "The use of technical analysis in the foreign exchange market". Journal of International Money and Finance, vol. 11, 1992
- Taylor, S.J.: "*Modelling financial time series*". Wiley, 1986.
- Taylor, S.J.; Xu, X.: "The incremental volatility information in one million foreign exchange quotation", Journal of Empirical Finance, vol. 4, 1997.
- Teichroew, D.: "The mixture of normal distributions with different variances". Annals of Mathematical Statistics, vol. 28, 1957.
- Teräsvirta, T.: "Power properties of linearity tests for time series". University of California, San Diego, Department of Economics, Discussion Paper, n° 90, 1990.
- Theil, H.: "*Principles of econometrics*". Wiley, 1971.
- Thode, H.C.: "*Testing for Normality*". Dekker, 2002.
- Thursby, J.G.: "Alternative specification error tests: A comparative study". JASA, vol. 74, 1979.
- Thursby, J.G.: "Misspecification, heteroscedasticity, ande the Chow and Goldfeld-Quandt tests". Review of Economics and Statistics, vol. 64, 1982.
- Thursby, J.G.; Schmidt, P.: "Some properties of tests for specification error in linear regression model". JASA, vol. 72, 1977.
- Tiago de Oliveira, J.: "*Statistics for Gumbel and Fréchet distributions*". Pergamon Press, 1972.
- Tiao, G.C.; Box, G.E.P.: "Modeling multiple time series with applications". JASA, vol. 76, 1981.
- Tiao, G.C.; Tsay, R.S.: "Model especification in multivariate time series".JRSS, Serie B, vol. 51, 1989.
- Tiao, G.C.; Tsay, R.S.; Wang, T.: "Usefulness for linear transformations in multivariate time series analysis". Empirical Economics, vol. 18, 1993.
- Tiku, M.L.: "Estimating the parameters of normal and logistic distributions for censored samples". Australian Journal of Statistics, vol. 10, 1968.
- Tinbergen, J.: "Econometric business cycle research". Review of Economic Studies, vol. 7, 1940.
- Tinbergen, J.: "*Econometrics*". Allen & Unwin, 1951.
- Tintner, G.: "*The variate difference method*". Principia-Press, 1940.
- Tintner, G.: "*Econometrics*". Wiley, 1952.
- Tipett, L.H.C.: "Statistical methods in textil research. Part 2: Uses of the binomial and a Poisson distributions". Shirley Institute Memories, vol. 13, 1934.
-Toe, D.T.: "*Statistical analysis of Markovian and related time series*". La Trobe University, 1976.
- Tong, H.: "*Non-linear time series analysis: A dynamical systems approach*". Oxford University Press, 1990.
- Tong, H. (ed.): "*Chaos and forecasting*". World Scientific, 1995.
- Toranzos, F.I.: "*Teoría estadística y aplicaciones*". Kapelusz, 1962.
- Treynor, J.; Ferguson, R.: "In defense of technical analysis". Journal of Finance, vol. 40, 1985.
- Tsay, R.S.: "Time series model specification in the presence of outliers". JASA, vol. 81, 1986.
- Tsay, R.S.: "Outliers level shifts and variance changes in time series". Journal of Forecasting, vol. 7, 1988.
- Tsay, R.S.: "Two canonical forms for vector ARMA processes". Statistica Sinica, vol. 1, 1991.
- Tsay, R.S.; Tiao, G.C.: "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models". JASA, vol. 79, 1984.
- Tucker, A.: "A reexamination of finite -and infinite- variance distributions as models of daily stock returns". Journal of Business and Economic Statistics, vol. 10, 1992.
- Tukey, J.W.: "On the comparative anatomy of transformations". Annals of Mathematics and Statistics, vol. 28, 1957.
- Tukey, J.W.: "The true purpose of transformations". Biometrics, vol. 24, 1958.

- Tukey, J.W.: "***Exploratory data analysis***". Addison-Wesley, 1977.
- Uhlenbeck, G.E.; Ornstein, L.S.: "On the theory of the Brownian motion". Physical Review, vol. 36, 1930.
- Urbisaia, H.L.; Brufman, J.Z.: "*Análisis de series de tiempo. Univariadas y multivariadas*". Ediciones Cooperativas, 2000.
- van Putten, B.: "On the construction of multivariate permitation tests in the two sample case". Statistica Neerlandika, vol. 41, 1987.
- van Uven, M.J.: "Extensions of Pearson's probability distributions to two variables, I". Ned. Akad. Wetensch. Prog., vol. 50, 1947.
- van Uven, M.J.: "Extensions of Pearson's probability distributions to two variables, II". Ned. Akad. Wetensch., vol. 50, 1947.
- van Uven, M.J.: "Extensions of Pearson's probability distributions to two variables, III". Ned. Akad. Wetensch., vol. 51, 1948.
- van Uven, M.J.: "Extensions of Pearson's probability distributions to two variables, IV". Ned. Akad. Wtensch., vol. 51, 1948.
- Vasicek, O.: "An equilibrium characterization of the term structure". Journal of Financial Economics, vol. 5, 1977.
- Velasco, C.: "Non-Gaussian log-periodogram regression". Econometric Theory, vol. 36, 1998.
- Velasco, C.: "Non-stationary log-periodogram regression". Econometric Theory, vol. 36, 1998.
- Velasco, C.: "Gaussian semiparametric estimation of nonstationary time series". Journal of Time Series Analysis, vol. 1, 1999.
- Verhulst, P.F.: "Recherches mathématiques sur la loi d'accroissement de la population". Académie de Bruxelles, vol. 18, 1845.
- Vinod, H.D.: "Generalization of the Durbin-Watson statistic for higher order autorregressive processes". Communications in Statistics, vol. 2, 1973.
- Vogt, H.: "Zur Parameter-und Prozentpunktschätzung von Lebensdaverteilungen bei kleinen Stichprobenumfang" Metrika, vol. 14, 1968.
- Volterra, V.: "Diffusion processes in genetics". Proceedings 2nd. Berkeley Symposium on Mathematical Statistics and Probability, 1851.
- von Mises, F.: "Ausschaltung der Ergondenhypothese in der physikalischen Statistik". Physikalische Zeitschrift, vol. 21, 1920.
- von Mises, L.: "Ueber die Wahrscheinlichkeit seltener Ereignisse". Zitschrift für Angewandte Mathematik und Mechanik, vol. 1, 1921.
- von Mises, L.: "Ueber die gegenwärtige Krise der Mechanik". Zeitschrift für Angewandte Mathematik und Mechanik, vol. 1, 1921.
- von Mises, L.: "*Wahrscheinlichkeit, Statistik und Wahrheit*". Speringer, 1928.
- von Mises, L.: "Ueber kausale und statistiche Gesetzmässigkeit in der Physik". Die Naturwissenschaften, vol. 18, 1930.
- von Mises, L.: "Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theorischen Physik". Deuticke, 1931.
- von Mises, L.: "Ueber Zahlenfolgen die ein Kollektiv-ähnliches Verhalten zeigen". Mathematische Annalen, vol. 108, 1933.
- von Mises, L.: "Théorie des probabilités, fondements et applications". Istitut Henri Poincaré, vol. 3, 1934.
- von Neumann, J.: "Distribution of the ratio of the mean-square successive difference to the variance". Annals Mathematical Statistics, vol. 12, 1941.
- von Smoluchowsky, M.: "Zur kinetische theorie der Brownschen Molekularbewegung und der Suspensionen". Annalen der Physik, vol. 21, 1906.
- von Smoluchowski, M.: "Ueber Brownsche Molekularbewegung unter Einwirkung äusserer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffussionsgleichung". Annalen der Physik, vol. 48, 1915.
- von Smoluchowski, M.: "Drei Vorträge über diffusio Brownsche Molekularbewegung und Koagulation von Kolloidteilchen". Physikalische Zeitschrift, vol. 17, 1916.
- von Smoluchowski, M.: "Ueber den Begriff des Zufalls und den Ursprung der wharscheinlichkeitsgesetze in der Physik". Die Naturwissenschaften, vol. 6, 1918.
- Wallace, T.D.; Ashar, V.G.: "Sequential methods in model construction". Review of Economics and Statistics, vol. 54, 1972.
- Wallis, K.F.: "Testing the fourth-order autocorrelation in quarterly regression equations". Econometrica, vol. 40, 1972.
- Wallis, WE.; Roberts, H.: "*Statistics: A new approach*". Free Press, 1956.
- Wang, H.: "*Reflections on Kurt Gödel*". MIT Press, 1987.
- Weaver, W.: "*Lady luck*". Dover, 1982.

- Weber, J.E.; Monarchi, D.E.: "Performance of the Durbin-Watson test and WLS estimation when the disturbance term includes serial dependence in addition to first-order autocorrelation". JASA, vol. 77, 1982.
- Weibull, W.: "A statistical theory of the strength of material". Ing. Vetenskaps Akad. Handl., vol. 151, 1939.
- Weibull, W.: "Moment about smallest sample value". Technical Report AFML-TR-67-105, Air Force Material Laboratory, 1967.
- Weichselberger, K.: "Über ein graphisches Verfahren zur Trennung von Mischverteilungen und zur Identifikation kapierter Normalverteilungen bei grossem Stichprobenumfang". Metrika, vol. 4, 1961.
- Weida, F.M.: "On certain distribution functions when the law of the universe is Poisson's first law of error". Annals of Mathematical Statistics, vol. 6, 1935.
- Wellner, J.A.: "Permutation tests for directional data". Annals of Statistics, vol. 7, 1979.
- Wessels, J.: "Multimodality in a family of probabilitye densities with application to a linear mixture of two normal densities". Statistica Neerlandica, vol. 18, 1964.
- West, K.D.; Cho, D.: "The predictive ability of several models of exchange rate volatility". Journal of Econometrics, vol. 69, 1995.
- White, H.: "A heterokedasticity-consistent covariance matrix estimator and a direct test for heterokedasticity". Econometrica, vol. 48, 1980.
- White, H.: "Instrumental variables regression with independent observations". Econometrica, vol. 50, 1982.
- White, H.: "*Asymptotic theory for econometricians*". Academic Press, 1984.
- White, H.: "Economic predictions using neural networks: The case IBM daily stock returns". Proceedings of IEEE International Conference on Neural Networks, 1988.
- White, H.; Domowitz, I.: "Nonlinear regression with dependent observations". Econometrica, vol. 52, 1984.
- Whittle, P.: "Gaussian estimation in stationary time series". Bulletin of the International Statistician Institute, vol. 39, 1962.
- Whittle, P.: "*Prediction and regulation*". English Universities Press, 1963.
- Wiener, N.: "Differential-space". Journal of Mathematics and Physics, vol. 2, 1923.
- Wiener, N.; "Un probléme de probabilités denombrables". Bulletin Societé Mathématique de France, vol. 52, 1924.
- Wiener, N.: "*The extrapolation, interpolation and smoothing of stationary time series with engineering applications*". Wiley, 1949.
- Wiggins, J.B.: "Option values under stochastic volatility: Theory an empirical estimates". Journal Financial Economics, vol. 19, 1987.
- Wilk, M.B.; Gnanadesikan, R.: "Probability plotting methods for the analysis data". Biometrika, vol. 55, 1968.
- Wilks, S.S.: "*Mathematical statistics*". Wiley, 1962.
- Willinger, W.; Taqqu, M.: "Towards a convergence theory for continuous stochastic securities market models". Mathematical Finance, vol.1, 1991.
- Willmott, P.: "*Derivatives*". Wiley, 1998.
- Williams, D.: "*Diffusion Markov processes and martingales*". Springer-Verlag, 1979.
- Wilson, E.B.; Hilferty, M.M.: "The distribution of chi-square". Proc. of the National Academy of Sciences, vol. 17, 1931.
- Wilson, G.: "Factorization of the covariance generating function of a pure moving average process". SIAM, vol. 6, 1969.
- Wold, H.: "*A study in the analysis of stationary time series*". Almqvist & Wicksell, 1938.
- Wold, H.: "Statistical estimation of economic relationships". Econometrica, vol. 17, 1949.
- Wold, H.: "*Demand analysis: A study in econometrics*". Wiley, 1953.
- Wold, H.: "Causality and econometrics". Econometrica, vol. 22, 1954.
- Wold, H.: "*A study in the analysis of stationary time series*". Almqvist-Wiksell, 1954.
- Wold, H.: "Causal inference from observational data: A review of ends and means". JRSS, Serie A, vol. 119, 1956.
- Wong, F.S.: "Fuzzy neural systems for stock selection". Fianancial Analysts Journal, vol. 48, 1992.
- Woodbury, M.A.: "On a probability distribution". Annals of Mathematical Statistics, vol. 20, 1949.
- Working, H.: "The investigation of economic expectation". American Economic Review. vol. 39, 1949.
- Wu, P.: "Testing fractionally integrated time series". Working Paper, Victoria University, Wellington, 1992.
- Yaglom, A.M.: "*An introduction to the theory of stationary random functions*". Prentice-Hall, 1962.
- Yajima, Y.: "On estimation of long-memory time series models". Australian Journal of Statistics, vol. 27, 1985.
- Yor, M.: "Présentation du pli cacheté". Comptes Rendus del'Academie des Sciences, Serie I, vol. 331, 2000.
- Yor, M.: "*Exponential functionals of Brownian motion and related processes*". Springer, 2001.
- Young, W.E.: "Random walk of stock prices: A test of the variance-time function". Econometrica, vol. 39, 1971.
- Yule, G.U.: "On the correlation of total pauperism with proportion of out-relief". Economic Journal, vol. 5, 1895-96.
- Yule, G.U.: "Notes on the history of pauperism in England and Wales from 1850, treated by the method of frequency curves, with an introduction on the method". JRSS, vol. 59, 1896.

- Yule, G.U.: "On the theory of correlation". JRSS, vol. 60, 1897.
- Yule, G.U.: "*An introduction to the theory of statistics*". Griffin, 1911.
- Yule, G.U.: "A mathematical theory of evolution based on the conclusion of Dr. Willis F.R.S.". Philosophical Transactions of the Royal Society, Serie B, vol. 213, 1924.
- Yule, G.U.: "Why do we sometimes get nonsence-correlation between time series? A study in sampling and the nature of time-series". JRSS, vol. 89, 1926.
- Yule, G.U.: "On a method of investigating periodicities in disturbed series with special reference to Wolfer's suspot numbers". Philosophical Transactions of the Royal Society, vol. 226, 1927.
- Yule, G.U.; Kendall, M.G.: "*An introduction to the theory of statistics*". Arnold, 1950.
- Zarembka, P. (ed.): "*Frontiers in econometrics*". Academic Press, 1974.
- Zellner, A. (ed.): "***Seasonal analysis of economic time series***". Bureau of Census, 1978.
- Zhou, B.: "High frequency data and volatility in foreign exchange markets". Journal of Business and Economic Statistics, vol.14, 1996.
- Zivot, E.; Andrews, D.W.K.: "Further evidence on the great crash, the oil price shock, and the unit root hypothesis". Cowles Foundation Discussion Paper, nº 944, 1990.
- Zyskind, G.: "On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models". Annals of Mathematical Statistics, vol. 38, 1967.

Indice temático

- Abdrabo, N.A., 350
- Adecuacidad de un modelo, 258-276
- Afinidad
    transformación de, 429
- AIC, 162-168, 185, 188, 250
- AIC-corregido, 168-170, 185, 188, 250
- AIC-insesgado, 170-172, 176, 188, 250
- Akaike, H., 157, 162, 170, 172, 173, 180, 250
- Algoritmo
    de Durbin-Levinson, 151, 519
    de Sowell, 524
- Allen D.M., 158
- Amplitud, 204, 219
- Amsler, C., 536
- Antle, C.E., 476
- Análisis
    de intervención, 346
    técnico, 388
- Andersen, A., 143
- Anderson, T.W., 188, 269, 271
- Andrews, D.W.K., 346, 418, 532
- Anscombe, F.J., 291, 300, 301
- Aproximación
    de Bhargava, A., 332
    de Kolmogorov, 281
    de Phillips-Ouliaris, 332
- Arley, N., 88
- Aroian, L.A., 475
- Atractor, 4
    caótico, 4
    extraño, 4
    quasiperiódico, 4
    simple, 4
- Autoafinidad, 429
    estadística, 429
    transformación de, 429
- Autosimilitud
    cualitativa, 429
    estadística, 429
    estricta, 429
- Avram, F., 548
- Axiomas
    sobre el tiempo, 6

- Bachelier, L., 8, 10, 422, 424, 440, 443, 460
    modelo de, 8
    representación de, 443
- Badrinath, S.G., 463
- Bai, I., 265
- Baillie, R.I., 131
- Bain, I.J., 476
- Balke, N.S., 346

- Barnard, G.A., 317
- Barrera
    absorbente, 51, 52, 53, 505
    parcialmente reflectora, 51, 53
    reflectora, 51, 53
- Bartlett, M.S., 295
- Beal, G., 301
- Bera, A.K., 264
- Beran, J., 517
- Bergström, B.M., 415
- Bernoulli, J., 27
- Bernstein, S., 28, 417
- Bhansali, R.J., 175
- BIC, 180
- Billingsley, P., 415, 421
- Birkhoff-Khinchin, teorema de, 26
- Black, F., 411
- Blattberg, R.C., 496
- Boltzmann, L., 5, 10, 27
- Borel, E., 10, 30
- Borel-Cantelli
    teorema de, 26
- Born, M., 8
- Box, G.E.P., 137, 293, 302
- Brockwell, P.J., 259, 517
- Brown, R., 422
- Brush, S., 422

- Cadena
    con mecanismo de transición estacionario, 39
    de Markov, 14, 37-53, 389, 439
    ergódica, 49
    estacionaria, 48
    estrictamente estacionaria, 40
    homogénea, 39, 40
    homogénea finita, 42
    no-periódica, 49
    no-reducible, 40, 49
    recurrente positiva, 49
    recurrente, 49
    reducible, 47
    regular, 49
    reversible, 48
- Cálculo de Stratonovic, 438
- Cambio de medida, 427
- Campbell, J.Y., 377, 457, 520
- Cantidad aleatoria, 10
- Caos, 449
- Cardano, G., 8
- Cartera de modelos, 251
- Cauchy-Schwartz, desigualdad de, 34
- Chaitin, G., 4
- Charlier, C.V.L., 465
- Chapman, S., 14
- Chatterjee, S., 463
- Chebychev, P.L., 27, 32
- Cheung, Y.W., 527

- Chu, J.T., 470
- Chung, K.L., 436, 522
- Church, A., 157
- Ciesielski, Z., 423
- Clark, P.K., 493, 494, 495
- Clase
    cerrada, 44
    comunicante, 44
- Clausius, R., 5
- Clausura, 44
- Cleveland, W.S., 255
- Cochran, J.H., 347, 348, 530
- Cociente
    de información, 458
    de verosimilitudes, 458
    señal-ruido, 159, 165, 166, 169, 170, 171, 174-175, 176, 179, 181-182, 186
- Coeficiente
    de determinación, 257, 307-308
    de Hurst, 518, 529
    de información de Kullback-Leibler, 159
    δ, 408
    delta, 461
    de rangos de Spearman, 408
    de riesgo deocurrencia de un evento, 66
    gamma, 461
    kappa, 461
    λ, 291, 293, 295-298
    R/S, 528
    vega, 461
- Cogburn, R., 29
- Cohen, A.C., 465
- Cohorte, 104
- Comportamiento
    espectral típico de una variable económica, 519
    fat.tailed, 462, 491, 493, 494
- Condición
    de autorregresividad de primer orden, 54
    de Brockwell-Davies-Beran, 517
    de compatibilidad, 7
    de convergencia de las autocovarianzas de McLeish, 34
    de determinación, 194, 225
    de Doeblin, 55
    de Hausman, 401
    de heterocedasticidad de White, 407
    de homogeneidad absoluta, 17
    de invariancia con respecto al tiempo, 25
    de Kolmogorov-Chapman, 14,38, 39, 55, 67
    de Lee-Schmidt, 514
    de la martingala, 58, 66
    de la mixingala, 418
    de Lindeberg, 414
    de linealidad, 54, 189-191
    de Lipschitz, 438
    de liquidez, 440
    de Markov, 54, 424, 448
    de la martingala, 58

- Condición
de no-recurrencia, 46
de recurrencia, 46
de simetría, 7
de Sowell, 514
de Taqqu-Lo, 515
- Condiciones
de consistencia de Kolmogorov, 7, 8
de estacionariedad del AR(p), 144-145, 206-208
de estacionariedad del ARMA(p,q), 238, 239
de estacionariedad del MA(q), 227
de estacionariedad en proceso periódicos, 353-355
de Gnedenko, 471
de invertibilidad del AR(p), 145
de invertibilidad del ARMA(p,q), 240
de invertibilidad del MA(q), 227
de invertibilidad en procesos periódicos, 353-355
de Lee-Schmidt, 514, 534
de recurrencia (en procesos de nacimiento-fallecimiento), 74
de Phillips, 131
mixing, 29, 31, 32, 408, 419, 518
mixing fuertes, 29
- Conjunto fractal, 449
- Constante
de Euler, 164, 472, 490
de Pareto, 480
- Cont, A., 439
- Convergencia
casi con certeza, 25, 31
completa, 26
débil, 26, 315
de Lee-Schmidt, 131
en-distribución, 26, 315
en-probabilidad, 25, 31
en valor medio cuadrático, 37
estocástica, 26
- Correlograma, 136
- Cowles, A., 389
- Cox, D.R., 293, 422, 452
- Cox, J.C., 506
- Cp, 177-180
- Cramér, H.,11, 303, 350
- Criterios
consistentes, 157-159
de Kolmogorov, 413
de predicción de dispersión mínima, 488
eficientes157-159
- Cuadrados mínimos ordinarios, 35
- Curva smile, 462

- Darling, D.A., 269, 271
- Darwin, Ch., 5
- Davies, R.A., 517
- Davis, R.A., 259, 486
- Davis, R.C., 350
- Davisson, I.D., 172
- Delta de Kronecker, 252

- De Moivre, A., 8, 32
- de Finetti, B., 10
- Definición
    de invertibilidad de Granger-Andersen, 143
    de McLeod-Hipel, 141
- de Forest, E.L., 317
- De Helguero, F., 465
- De Jong, D.N., 334
- Dependencia
    de martingala, 34
    Markoviana, 422
- Desarrollo de Volterra, 140
- Descomposición de la varianza, 206
- Desigualdad
    de Cauchy-Schwartz, 34, 126
    de Markov, 414
    de Jensen, 403
    de Schwartz-Hölder, 252
- Devaney, R.I., 449
- De Vries, C.G., 491
- Diagrama
    P-P, 260
    P-P estandarizado, 261
    Q-Q, 260, 262, 263, 484
- Dickey, D.A., 317, 319, 320, 329, 333
- Diferencia periódica, 354
- Dimensión fractal, 449
- Dinámica
    individual, 5
    estadística, 5
- Discrepancia de Kullback-Leibler, 159
- Distribución
    α-estable simétrica, 481
    Beta, 260, 262, 467, 496, 519
    binomial negativa, 98
    contaminada, 493
    de Bernoulli, 482
    de Cauchy, 262, 476, 482, 486, 489, 514
    de contagio de Neyman, 102
    de contagio verdadero, 100
    de Dirichlet, 61
    de dos colas, 469
    de Frechét, 476-480
    de Gumbel, 471, 473, 476, 478
    de Laplace, 467, 469
    del proceso, 8
    de Neyman-tipo A, 104
    de Neyman- tipo B, 103, 104
    de Neyman-tipo, C, 103, 104
    de Pareto, 262, 467, 470
    de Pareto de primera clase, 481
    de Pareto de segunda clase, 481
    de Pareto-Lévy, 480-484
    de Poisson, 100
    de probabilidades de equilibrio (en procesos de nacimiento-fallecimiento), 75
    de valores extremos, 471
    de variación lenta en el infinito, 483
    de variación regular en el infinito, 482, 483

- Distribución
de Weibull, 262, 473-476
de Weibull-Gnedenko, 473
doble exponencial, 469
exponencial, , 59, 105, 262
exponencial bilateral, 469
Gamma, 60, 66, 100, 467
hiperbólica, 466
latente, 463, 465
log-Beta, 163
logística, 467-469
log_Normal, 441, 443, 463, 495
logística, 262
log Weibull, 472
Max-estable, 481
Normal compuesta, 463
Normal inversa, 466
Normal mezclada, 463, 464
Normal mezclada contaminada, 464
poli-Gamma, 472
t, 466-467, 496
Theta de Jacobi, 85
uniforme, 261
- Doeblin. W., 432, 435
condición de, 55
- Dominio
de las fases, 3
de las variables, 3, 9, 244
de los estados, 3
del tiempo, 3, 244
- Donsker, M., 421
- Doob, J.L., 11, 409, 417, 432
matriz Q de, 72, 75, 85, 88, 92, 97, 104
- Doornik, J.A., 265, 509
- Downham, D.Y., 175
- Drogin, R., 421
- Dubey, S.D., 476
- Duración de la vida, 105
- Durbin, J., 276
- Dutka, J., 423
- Dvoretzky, A., 415

- Ecuación
backward, 56, 67, 92
característica del AR(p), 144, 521
característica del MA(2), 234
del AR(2), 198
de Cauchy, 191
de Chapman, 10
de difusión de Fokker-Planck, 10, 431
de Kolomogorov-Chhapman, 54, 70
de Langevin, 450
de Puri, 111
de Stratonovich, 438
diferencial estocástica de Itô, 436, 437
diferencial estocástica de Vasicek, 452
estocástica lineal general, 437

- Ecuación
  forward, 56, 67
  integral, 432
  integral estocástica, 432, 447
  lineal con perturbaciones aditivas, 437
  lineal homogénea con perturbaciones multiplicativas, 437
  reflexiva, 446, 447
- Edad del proceso al momento de la extinción, 96
- Efecto José, 519
- Eficiencia, 449
  débil, 448
  del mercado, 443, 448
  fuerte, 448
  semi-fuerte, 448
- Eggenberger, F., 99
- Elliot, G., 332
- Elston, R.E., 291
- Einstein, A., 5, 8, 10, 422
- Engle, R.F., 292, 457
- Entropía,159
- Erdös, P., 421
- Ergal, 27
- Ergode, 27
- Ergodisch, 27
- Error
  de la predicción, 279, 285, 352, 488
  medio absoluto, 458
  medio cuadrático, 35, 458
  medio cuadrático estandarizado, 458
- Esqueleto, 143
- Esquema
  de Bernoulli, 99
  de contagio débil, 100
  de contagio fuerte, 100
- Estabilidad asintótica, 4
- Estcionalidad, 353
  determinística, 356
- Estacionariedad
  concepto de, 17
  estricta, 18
- Estadístico
  de Box-Pierce, 137
  de Cowles-Jones, 390-391
  de Dickey-Fuller, 319, 320
  de Jarque-Bera, 264, 265
  de Kuiper, 267
  de Ljung-Box, 137, 259
  de McLeod-Li, 259
  de Wald, 265
  Q, 398
  R/S, 530
  τ, 316
  z de Phillips, 331
- Estado
  absorbente, 44, 75-87
  accesible, 43
  comunicado, 43
  comunicado en ambos sentidos, 43

- Estado
de enfermedad, 115
de fallecimiento, 115
doblemente comunicado, 43, 46
efímero, 43
esencial, 43
infinitely often, 50
no-esencial, 46
no-recurrente, 46, 47, 50, 51
periódico, 44
persistente, 45
recurrente, 45, 46, 50
recurrente nulo, 45, 47
recurrente positivo, 45
- Estimador
autorregresivo de la función de densidad espectral, 132
de Geweke-Porte Hudak, 527
de Hill, 477
de la función de autocorrelaciones, 134-138
de la función de autocovarianzas, 128-131
de la función de densidad espectral, 132-133
de la volatilidad, 457-459
de los coeficientes del AR(p), 146-149
de los coeficientes del AR(1), 192-193
de los coeficientes del ARFIMA, 523-528
de los coeficientes del ARMA(p,q), 245-249
de los coeficientes del MA(q), 228-232
de los coeficientes del MA(1), 233
de los coeficientes del MA(2), 234
de los coeficientes en procesos periódicos, 368-372
de los ódenes p y q, 249-251
de máxima verosimilitud (del AR(p)), 148
de máxima verosimilitud (cadenas de Markov finitas), 42, 43
de máxima verosimilitud de ρ (en procesos de Poisson homogéos), 61, 62, 63,64
de máxima verosimilitud (en procesos de nacimiento-fallecimiento lineal), 87
de máxima verosimilitud (en procesos puros de nacimiento lineal), 88-91
de máxima verosimilitud aproximado de Whittle, 524, 525
de Newey-West, 534
de Pickands, 480
de Yule-Walker, 147, 208, 210, 211
GPH, 527
máximo-verosímil, 192-193, 210-211
mínimo-cuadrático, 147-148, 209-210, 211
preliminar, 228
- Estructura
causal, 3, 4
de la heterocedasticidad, 292
temporal, 143, 145, 198, 203, 344
temporal de la tasa de interés, 451, 454
- Extensión de Cramér del teorema de Wold, 303
- EWMA, 456, 457

- Factor de amortiguación, 203
- Fama, E.F., 448
- Familia de transformaciones de Johnson, 264, 294
- Fase, 205, 216, 220
- Feller-Lundberg, teorema de, 85

- Feller, W., 33, 38, 73, 93, 268, 316, 397
- Fenómeno
    de Hurst, 517
    de multiplicación de la aleatoriedad, 449
    dinámico, 3
    no-integrables, 6
- Fidis, 7
- Filtración, 12
    aumentada, 12
    natural, 8, 12, 126
- Filtro lineal, 140
- Fisher, R.A., 295, 466, 472
- Fix, E., 115
- Flujo
    de información, 12
    de información natural, 8, 12
- Fomby, T.B., 346
- Fórmula
    de Black-Scholes-Merton, 411, 426, 446, 459, 461
    de Cameron-Martin, 428
    de De Moivre-Stirling, 515, 516
    de Itô-Doeblin, 435, 452, 498, 505
    de Smirnov, 267
    recursiva de Gradshtein-Ryzhik, I.M., 547
- Fourier, J-J., 5
    transformación de, 18
- FPE, 158, 172-175, 188
- FPE-insesgado, 175-177, 188, 250
- Frechét, M., 472
- Frecuencia, 204, 216, 219
    G, 546
- Freund, J.E., 476
- Fuller, W.A., 317, 319, 320
- Función
    aleatoria, 7
    Beta, 262
    complementaria, 309
    de autocorrelaciones, 133-138, 145-146, 192, 228, 232, 234, 251, 286, 355-366, 366-368, 517, 518
    de autocorrelaciones extendida, 242
    de autocorrelaciones inversas, 252-258
    de autocorrelaciones parciales, 149-155, 207, 228, 233, 234, 519
    de autocovarianzas, 126-127, 128-131, 191-192, 227-228, 232, 251, 518, 521-522, 547-548, 549
    de autocovarianzas inversas, 252-258
    de Cauchy, 262
    de densidad espectral, 131-133, 194, 233, 305, 520, 522-523, 526-527
    de difusión, 431
    de discrepancia, 158
    de distancia, 157, 158
    de distribución espectral, 131
    de Euler, 164
    definida no-negativa, 131
    de información, 158
    de orden asintótico igual, 55
    de orden asintótico menor, 55
    de Pareto, 262
    de penalización, 157
    de predicción, 279, 351
    de probabilidades de transición, 8

- Función
de respuesta a un impulso, 140
de transferencia, 138
de transferencia autorregresiva, 178
de Weibull, 262
de Whittle, 524
exponencial, 262
g, 463
Gamma incompleta, 502
generatriz de autocorrelaciones inversas, 253
generatriz de autocovarianzas inversas, 252
generatriz de probabilidades (en procesos de enfermedad-fallecimiento), 120
generatriz de probabilidades (en procesos de Greenwood-Yule-Egggenberger-Pólya), 98
generatriz de probabilidades (en procesos de nacimiento fallecimiento con intensidades dependientes del tiempo y del tamaño de la población), 93
generatriz de probabilidades (en procesos de nacimiento-fallecimiento lineal), 85
generatriz de probabilidades (en procesos de nacimiento-fallecimiento lineal bidimensional), 113, 114
generatriz de probabilidades (en procesos de nacimiento-inmigración-emigración-fallecimiento lineal unidimensional), 107
generatriz de autocorrelaciones, 254
generatriz de autocovarianzas, 127, 254
generatriz de probabilidades (en procesos de nacimiento-migración-fallecimiento bidimensional), 110
generatriz de probabilidades (en procesos puros de fallecimiento lineal), 92
generatriz de probabilidades (en procesos puros de nacimiento lineal), 88
generatriz de probabilidades (en procesos de nacimiento fallecimiento con intensidades dependientes del tiempo y lineales con respecto al tamaño de la población), 97
h, 463
intensidad de fallecimiento, 72
intensidad de nacimiento, 72
intensidad de ocurrencia, 54, 69
intensidad de transición en los estados de enfermedad, 116
uniforme, 261
varianza-tiempo, 404
z de Riemann, 468
- Furry, W.H., 88

- Galambos, J., 481
- Galton, F., 260
- Gänsler, P., 415, 421
- Garwood, F., 295
- Geweke, J., 526, 527
- Gibbs, J.W., 5, 38
- Girasinov, I.V., 428
- Gnedenko, B.V., 471, 472, 473
- Gödel, G., 4
- Gompertz, B., 105
- Gonedes, W.J., 496
- Gosset, W.S., 466
- Gouriéroux. C., 259
- Grado de distorsión, 218
- Gradshtein, I.S., 547
- Granger, C.W.J., 143, 190, 440, 514, 519
- Gray, H.I., 545
- Green, J.R., 274
- Greenwood, M., 100
- Grünbichler, A., 507
- Gumbel, E.J., 468, 476

- Haas-Lorentz, G.L., 422
- Haldane, J.B.S., 158, 295
- Hall, A., 332
- Hall, P., 491
- Hannan, E.J., 183, 250, 255
- Hansen, H., 265
- Harding, H., 465
- Hartman, P., 415
- Harvey, A.C., 509
- Harvey, C.R., 266
- Häussler, E., 421
- Hegazy, Y.A.S., 274
- Hegel, G.F.W., 5
- Heisenberg, G.W., 8
- Hentschel, I., 457
- Heráclito, 5
- Heterocedasticidad
  - estructura de la, 292
- Hilferty, M.M., 295
- Hill, B.M., 479, 480
- Hillier, G.; 533
- Hipel, K.W., 141, 255
- Hipótesis
  - clásicas sobre el mercado de capitales, 439
  - de martingala, 398
  - RW, 440
  - varianza-tiempo, 404
- Hoeffding, W., 408
- Hopf, E., 10
- Hosking, J.R.M., 514, 545
- Hostinsky, B., 10
- Hotelling, H., 470
- HQ, 183-185, 188, 250
- HQ-corregido, 185-187, 188, 250
- Hull, J., 465
- Hurst, H.E., 513, 517, 528
- Hurvich, C.M., 168

- Ibragimov, I.A., 415
- Inclusión de Skorohod, 427
- Indice
  - de determinismo lineal, 255
  - de femineidad, 112
  - de la cola, 480
  - esperado de femineidad, 115
- Ingersoll, J.E., 452
- Integral
  - complementaria, 309
  - de Itô, 433, 434, 450
  - de Stratonovich, 438
  - estocástica de Itô, 315, 413
  - estocástica de Stratonovich, 438
- Intensidad de reacción, 456
- Intercambiabilidad, 13, 17
- Interpretación
  - constructivista, 8
  - distribucionista, 8

- Invariancia con respecto al tiempo, condición de, 25
- Itô, K., 432, 436

- Jacobi, distribución Theta de, 85
- Jarque, C.M., 264
- Jarrow, R.A., 502
- Jenkins, G.M., 302
- Jensen, M., 403
- Johnson, N.L., 264
- Jones, H., 389
- Joyeux, R., 514

- Kac, M., 421
- Keenan, D.M., 190
    test de linealidad de, 190
- Keeney, R.D., 468
- Kendall, D.G., 94, 106
- Kendall, M.G., 440
- Khinchin, A.N., 10, 28, 30, 38, 143
- Kim, K., 331
- King, M.L., 533
- Koedijk, K.J., 491
- Kolmogorov, A.N., 7, 10, 30, 143, 266, 267, 281, 413, 422, 443
    condiciones de consistencia de, 7, 8
    criterio de, 413
    sistema de ecuaciones estocásticas diferenciales-en diferencias de, 55, 69, 73, 74, 85, 92, 94, 97, 113
    teorema de existencia de, 8
- Kolmogorov-Chapman
    condición de, 14, 38, 39, 55
    ecuación de, 54
- Kolmogorov-Daniels
    teorema de, 8
- Koopman, S.J., 509
- Kou, S.G., 500
- Kuiper, N.H., 267
- Kurtosis, 263, 265
- Kwaiatkowski, D., 533

- Langevin, P., 451
- Laplace, P.S., 32, 469
- Laurent, matriz de, 133
- Lebesgue, H., 10
- Lee, D., 131
- Lee, H.S., 536
- Lee, P., 514
- Lema
    de Itô-Doeblin, 435, 443, 446, 450
    de la transformación de la probabilidad integral, 261
- Lennox, W.C., 255
- Lévy, P.P., 11, 422, 466, 481, 482
- Ley
    débil de los grandes números, 27, 32
    del logaritmo iterado, 30, 32
    del movimiento de Newton, 451

- Ley
de los grandes números, 27
de mortalidad de Makeham, 105
de T a la ½, 528
fuerte de los grandes números, 27, 30
- Límite
de Cesaro, 28
- Lin, F., 190
- Lindeberg, J.W., 414
- L-mixingala, 418
- Lo. A.W., 349, 377, 398, 400, 514, 515, 532
- Longsstaff, F.A., 507
- Loretan, M., 492, 493
- Loynes, R.M., 415
- Lindeberg, J.W., 33
- Ljung, G.M., 137
- Lundberg, O., 10, 83
- Lyapunov, A.M., 32

- Makeham, W.M., 105
ley de mortalidad de, 105
- Malkiel, B., 448
- Mallows, C.L., 157, 177
- Mandelbrot, B.B., 513, 518, 480, 519
- Mankiw, N.G., 520
- Mann, H.B., 148
- Markov, A.A., 38
cadena de, 14, 37
proceso de, 9, 37
- Martingala, 9, 16, 34, 409-417
backwards, 412
condición de, 58
continua en el dominio del tiempo, 417
de cuadrado integrable, 410
de orden s, 410
dependencia de, 34
de segundo orden, 410
inversa, 412
- Matriz
de autocovarianzas, 211, 212
de autocovarianzas de Toeplitz, 36
de Laurent, 133
de probabilidades de transición (cadenas de Markov finitas), 42
de probabilidades de transición (cadenas de Markov reducibles), 47
de probabilidades de transición en dos pasos, 39
de probabilidades de transición en n pasos, 39
de probabilideades de transición infinitesimal (en procesos de nacimiento-fallecimiento), 72
estocástica, 38
Q de Doob, 72, 75, 85, 88, 92, 97, 104
- McClave, J.T., 255
- McFarland, J.W., 491, 492
- McKendrick, A.G., 83
- McKinlay, A.C., 349, 377, 398, 400
- McKinnon, J.G., 329
- McLeish, D.L.
condición de convergencia de las autocovarianzas de, 34

- McLeod, A.I., 141, 255
- McMahon, P.C., 491, 492
- McNeill, I., 535
- Medidas
    cociente de información, 458
    cuadrado de coeficiente de contingencia, 408
    de dependencia temporal, 31
    de diferencia de información, 408
    de distancia, 159-161
    de distancia dirigida de Kullback-Leibler, 157, 159, 163, 187
    de distancia L_2, 160, 187
    de distancia de Lévy, 315
    de eficiencia esperada, 160, 161
    de eficiencia observada, 160, 161
    de martingala equivalentes, 428
    de persistencia, 140, 347-349
    de persistencia no-paramétrica de Cochran, 347
    de probabilidad equivalentes, 428
    de resistencia al fallecimiento, 105
    Δ de Hoeffding, 408
- Memoria del proceso, 12
- Menon, M.V., 476
- Merrington, M., 295
- Merton, R.C., 411, 422, 497, 503
- Método
    de adición de variable, 332
    de la variable instrumental de Hall, 332
    Delta, 391
    de Monte Carlo, 317, 327, 331, 439
    de Quenouille, 188-189
- Metropolis, N., 317
- Mezcla de distribuciones Normales, 463
- Miller, D.W., 475
- Miller, I., 476
- Mixingala, 34, 417-420
    condición de, 34
- Modelo
    ADC, 493
    autorregresivo de duración condicionada, 493
    con umbral, 489
    de Bachelier, 8, 459, 460
    de Black-Scholes-Merton, 460
    de Clark, 493-496
    de Cox-Ingersoll-Ross, 452
    de Fisher-Corbett-Williams, 106
    de Kou, 499-501
    de Merton, 503-505
    de Oldfield-Rogalski-Jarrow, 502-503
    de Praetz-Blattberg-Gonedes, 496-497
    de Press, 501-502
    determinístico, 147
    de redes neuronales, 190, 407
    de Vasicek, 451, 453
    estático, 194
    estocástico, 147, 225
    GARCH
- Molenaar, W., 465
- Monode, 27

- Mood, A., 392
- Moore, D.S., 272
- Moore, P.J., 291
- Morán, P.A.P., 91
- Morgan, W.A., 406
- Morgenstern, O., 440
- Movimiento
    Browniano, 10, 33, 53, 54, 314, 315
    hiperbólico de Lévy, 466

- Nabeya, S., 534
- Nankervis, J.C., 334
- Newey, W.K., 534
- Neyman, J., 102, 115
    distribución de contagio de, 102
    distribución tipo B de, 103
    distribución tipo C de, 103, 104
- Ng, S., 265
- No-estacionariedad
    en el valor medio, 301-312
    en la varianza, 291-301
    homogénea, 302

- Oldfield, G.S., 502
- Operador
    autorregresivo, 143, 144
    backward, 138, 226, 239
    densidad, 5
    de promedios móviles, 138, 227
    diferencial en valor medio cuadrático, 11
    forward, 141, 253
- Orden de autorregresividad, 155-189
- Ornstein, L.S., 451
- Ouliaris, S., 332
- Ouspensky, P.D., 6, 125

- Palm, C., 95
- Pantula, S., 333
- Paradigma
    de la reversibilidad, 3
- Pareto, V., 480
- Parménides, 5
- Parzen, E., 350, 517
- Pearson, K., 263, 265, 271, 423, 465
- Percentil, 260
- Período, 204, 216, 220
    básico, 353
- Perron, 322, 330, 331, 344, 345
- Persistencia, 456
- Phillips, P.C.B., 131, 322, 327, 330, 331, 332, 514, 491, 492, 493, 533
- Pickands, J., 479, 480
- Pierce, D.A., 137
- Poincaré, H., 443
- Poisson, S.D., 27

- Plakett, R.I., 468
- Polinomios de Gegenbauer, 546
- Pólya, F., 99, 100
- Poole, W., 404
- Porter-Hudak, S., 526, 527
- Poskitt, D.S., 251
- Potencia del contagio, 99
- Praetz, P., 496
- Predicción en procesos no-estacionarios, 349-353
- Predictor, 279, 280, 488
- PRESS, 158
- Press, S.J., 501
- Piestley, M.B., 350
- Primera ley de Laplace, 469
- Principio
    de entropía cruzada mínima de Kullback, 163
    de invariancia, 131, 421, 421, 514
    de juego equitativo, 409
    directo de optimización de la entropía, 159
    minx-ent, 163
- Probabilidad
    de absorción (en procesos de nacimiento-fallecimiento), 79
    de arribo al cabo de n transiciones, 44, 446
    de arribo alguna vez, 45
    de equilibrio, 48
    de equilibrio estacionaria, 49
    de extinción (en procesos de nacimiento-fallecimiento), 77, 78, 79, 80, 81
    de Greenwood-Yule-Eggenberger-Pólya, 97, 102
    de Itô, 436
    de los tiempos de espera (en procesos de Poisson homogéneos), 58, 59, 60, 61
    de los tiempos de espera (en proceso de Poisson generalizados), 68, 69
    del tiempo a transcurrir en un estado al cabo de n transiciones, 49
    del tiempo de ocupación de un estado al cabo de n transiciones, 49
    del tiempo promedio, 27
    de ocupación de un estado un tiempo infinito, 50
    de recurrencia, 45
    de recurrencia al cabo de n transiciones, 45
    de recurrencia por segunda vez, 47
    de recurrencia por segunda vez a un estado recurrente al cabo de n transiciones, 48
    de retorno alguna vez, 45
    de retorno por primera vez al cabo de n transiciones, 45
    de sobreestimación (AIC), 166-168
    de sobreestimación (AICc), 169-170
    de sobreestimacion (AICu), 171-172
    de sobreestimación (Cp), 179-180
    de sobreestimación (FPE), 175
    de sobreestimación (FPEu), 177
    de sobreestimación (HQ), 184-185
    de sobreestimación (HQc), 186-187
    de sobreestimación (SIC), 182-183
    de subestimación, 187-188
    de transición a los estados de fallecimiento, 116
    de transición en los estados de enfermedad, 116
    de transición (en RW), 52
    de transición (en RW simétrico), 53
    de transición (en proceso de Poisson homogéneo), 54
    de transición en un paso, 37
    de transición en dos pasos, 38

- Probabilidad
de transición infinitesimal (en procesos de Poisson homogénos), 55
de transición simple, 12
de transición simple (proceso de Markov), 37, 38
tiempo esperado de extinción (en procesos de nacimiento-fallecimiento), 82
inicial, 40
- Problema ergódico, 27
- Proceso
de convergencia cuadrática, 231245, 246
linealmente convergente, 230, 245
linealmente deteminístico, 16
- Proceso estocástico
adaptado, 12
anticipatorio, 157
antipersistente, 519, 529
AR(1), 191-197
AR(2), 197-226
ARARMA, 517
ARCH, 457
ARFIMA, 305, 334, 457, 513, 520-528, 490, 548
ARMA(1,1), 251-252
ARMA de Gegenbauer, 545
asintóticamente estrictamente autorregresivo de orden 0, 13
asintóticamente incorrelacionado, 15
autorregresivo, 140
autorregresivo de orden p, 144-226
autorregresivo de promedios móviles de orden (p,q), 238-251
autosimilares en el dominio del tiempo, 428
CADLAG, 53
CAGLAD, 53
caótico determinístico, 59
causal, 140
completamente homogéneo, 14
componente, 11
compuesto, 244
con dominancia femenina, 114
con salto, 53, 54
continuo en el dominio de las variables, 9
continuo en el dominio del tiempo, 10
continuo en valor medio cuadrático, 11
débilmente autorregresivo de orden 1, 14
débilmente autorregresivo de orden m, 12
débilmente estacionario de orden s, 18
de colisión molecular, 85
de crecimiento divergente, 83
de difusión, 436, 504
de difusión bidimensional, 505
de difusión con elasticidad de la varianza constante, 506
de difusión con salto, 497-499
de enfermedad-fallecimiento, 115-121
de Fix-Neyman, 115
de heterogeneidad separable de segundo orden, 376
de incrementos estacionarios, 17
de incrementos incorrelacionados, 16, 378
de incrementos independientes, 15, 378
de inmigración-fallecimiento, 111
de Itô, 431-439

- Proceso
de la familia de Koopman-Darmois, 180
de la ruina de los jugadores, 51
de las innovaciones, 35
de las perturbaciones, 35
de Markov, 9, 37-53, 377
de Markov débil, 14, 191-197
de Markov estricto, 14
de memoria corta, 131, 514, 517, 527
de memoria extendida, 519
de memoria intermedia, 519
de memoria larga, 131, 141, 513, 517, 519
de nacimiento-fallecimiento, 71-105
de nacimiento-fallecimiento con funciones intensidad dependientes del tamaño de la población, 71-82
de nacimiento-fallecimiento con funciones intensidad dependientes del tiempo y del tamaño de la población, 94-96
de nacimiento-fallecimiento lineal, 85-88, 93, 504
de nacimiento-fallecimiento lineal bidimensional, 112-115
de nacimiento-fallecimiento con estados absorbentes, 75-87
de nacimiento-fallecimiento no-recurrente de tipo-1, 74
de nacimiento-fallecimiento no-recurrente de tipo-2, 74
de nacimiento-fallecimiento recurrente-nulo, 74
de nacimiento-fallecimiento recurrente-positivo, 74
de nacimiento-inmigración-emigración-fallecimiento lineal unidimensional, 105-108
de nacimiento-migración-fallecimiento, 105-112
de nacimiento-migración-fallecimiento bidimensional, 108-112
de parámetro discreto, 11
de Poisson, 53-71, 98, 497
de Poisson compensado, 58
de Poisson compuesto, 64-66
de Poisson compuesto compensado, 66
de Poisson con función intensidad de ocurrencia constante, 53-64
de Poisson generalizado, 67-71
de Poisson homogéneo, 53-64
de Poisson no-homogéneo, 67-71
de Poisson ponderado, 66, 67, 98
de promdios móviles de orden q, 226-238
deshonesto, 54
de Uhlembeck-Ornstein, 449-451. 452, 506
de variables simples, 9
de Yule, 14, 197-226
de Yule-Furry, 88-91
diferenciable en valor medio cuadrático, 11
diferencia de una martingala, 410
diferencia de una mixingala, 419
directriz, 447
discreto en el dominio de las variables, 9
discreto en el dominio del tiempo, 10
EGARCH, 454
ergódico, 11, 26, 141
ergódico de primer orden, 27, 28
ergódico de segundo orden, 27, 30
ergódico en el valor esperado, 27
ergódico en las autocovarianzas, 27, 30
estacionario, 10, 11
estacionario de orden s en sentido amplio, 18
estacionario en las covarianzas, 18, 19, 141
estacionario en sentido estricto, 16

- Proceso
estrictamente autorregresivo de orden 0, 13
estrictamente autorregresivo de orden 1, 14
estrictamente autorregresivo de orden 2, 14
estrictamente autorregresivo de orden m, 12
estrictamente estacionario, 18, 19
explosivo, 83, 302
GARCH, 407, 454
GARCH-exponencial 454
GARMA, 545
H-autosimilar, 428
homogéneo en el tiempo, 16
honesto, 54
incorrelacionado de orden q, 15
independiente del futuro, 140
intercambiable, 13
invariante, 16
invertible, 143
invertible en el sentido de Granger-Andersen, 143
k-dimensional, 10, 20
linealmente dependiente, 16
MA(1), 232-233
MA(2), 233-238
marginal, 10
movimientgo Browniano, 131, 314, 315, 404, 420-431, 514, 447, 548
movimiento Browniano aritmético, 377, 422
con tendencia lineal, 423, 443
bidimensional, 447
movimiento Browniano de Wiener, 422
movimiento Browninao estandarizado, 424, 425, 426, 536
movimiento Browniano exponencial con tendencia, 426
movimiento Browniano fraccionario, 549
movimiento Browniano geométrico, 377, 426, 437, 443, 497
movimiento Browniano geométrico estandarizado, 426
movimiento Browniano integrado, 425
movimiento Browniano restringido estandarizado, 426
no-autorregresivo, 37
no-ergódico, 11
no-estacionario, 11
no-lineal, 142
periódico, 353-372
persistente, 529
puro de fallecimiento con función intensidad dependiente del tiempo y lineal con respecto al tamaño de la población, 104-105
puro de fallecimiento lineal, 91-93
puro de Gegenbauer, 546
puro de nacimiento con función intensidad dependiente del tamaño de la población, 83-93
puro de nacimiento con función intensidad dependiente del tiempo y lineal con respecto al tamaño de la población, 97-99
puro de nacimiento lineal, 88-91
puro de salto, 54
random flight, 423
random walk, 51, 375-385
random walk con ruido, 382
random walk con tendencia, 381
random walk con tendencia general y comportamiento irregular, 384
random walk con tendencia lineal local, 384
random walk con tendencia y con ruido, 383

- Proceso
    random walk débil, 378
    random walk de Bernoulli, 376
    random walk de segundo orden, 376
    random walk estricto, 377
    random walk Normal, 376, 377
    random walk puro, 376
    random walk reescalado, 376
    random walk simétrico, 376
    random walk simple, 376
    random walk tipo 1, 376, 420
    random walk tipo 2, 377
    random walk tipo 3, 351, 378, 411
    RCLL, 53
    reversible en el dominio del tiempo, 18, 142
    sen sible a las condiciones iniciales, 517
    simple, 434
    sobre-diferenciado, 518
    subordinado, 494
    uniformemente integrable, 418
    unidimensional, 10
    vector de orden k, 10
- Procesos
    conjuntamente estacionarios de segundo orden, 20
    conjuntamente estacionarios en sentido estricto, 20
    independientes, 20, 58
    individualmente estacionarios de segundo orden, 20
    individualmente estacionarios en sentido estricto, 20
- Prokhorov, Y.V., 421
- Promedio
    de fase, 27
    móvil exponencialmente ponderado, 455
- Propiedad
    de ausencia de memoria, 57, 59
    de invariancia con respecto a la adición de variables aleatorias, 485
    de isometría, 434
    de la martingala, 409
    de linealidad, 434, 442
    de retorno al valor medio, 140, 452
    de riesgo limitado, 377
    de sensibilidad a las condiciones iniciales, 449
    de sesgo limitado, 442
    de similitud, 318
    de transitividad topológica, 449
    solidaria, 46, 47
- Protter, P., 436
- Puente Browniano, 426
    de segundo nivel, 535
    estacionario, 535
- Punto
    de soporte, 463
    de truncamiento, 133
- Puri, P.S., 111
    ecuaciones de, 111

- Quantil, 260, 262

- Quenouille, M.H., 152, 157, 188, 225
    método de, 188
- Quinn, B.G., 183, 250

- Racha, 392
- Raíces unirtarias múltiples, 332-344
- Raíz
    negativa dominante, 198
    unitaria, 145
- Ramsey, J.B., 190
    test de, 190
- Random walk, 9, 16, 33, 51
    débil, 16
    homocedástico, 16
    no-recurrente, 52
    recurrente, 52
    recurrente-nulo, 52
    simétrico, 53
    tipo-3, 16
    unidimensional, 51
- Rango
    estandarizado, 528
    reescalado, 528
    reescalado modificado, 530
    rango reescalado modificado de Haubrich-Lo, 535
    sobre el desvío estándar, 528
- Rappoport, P., 346
- RASET, 190
- Regla
    de encadenamiento de la diferenciación, 434
    de $\tau^{1/2}$, 404-405
- Reichenbach, H., 10
- Reichlin, I., 346
- Reinsel, G.C., 330
- Relación
    ln(varianza)-ln(valor medio)
    rango-mediana, 295
    rango-valor medio, 295
    tiempo-error, 404
    varianza-valor medio, 295
- Renshaw, E., 108
- Rentabilidad continua, 441
- Representación
    AR, 142, 143, 311-312
    ARCH(p), 292
    ARFIMA, 389
    ARV, 507
    autorregresiva, 143
    autorregresiva de la varianza, 507
    autorregresiva integrada de promedios móviles, 302
    bilateral, 141
    cuántica, 3
    de Bachelier, 443
    de Grünbicher-Longstaff, 507
    de Lévy, 423
    de Lévy-Ciesielki, 423
    de Paley-Wiener, 423

- Representación
  en promedios móviles, 138-142, 144
  espectral, 131
  GARCH, 459, 465, 478
  lineal bilateral, 142
  MA, 138-142, 308-310
  MA del proceso inverso, 253
  mecánica, 3
  mediante el espectro evolutivo, 350
  multiplicativa, 354, 355
  no-lineal doble logarítmica, 495
  no-multiplicativa, 366
  polinomial bilateral, 142
  truncada, 240-241, 309
- RESET, 190
- Resnick, S.I., 486
- Richardson, M., 266, 398
- Riedwyl, H., 269
- Roberts, H., 448
- Robinson, P.M., 131, 527, 533
- Rogalski, R.J., 502
- Romano, J., 398
- Rootzén, H., 421
- Rosén, P., 415, 421
- Rosen, R., 157
- Rosemblatt, M., 29
- Ross, S., 422, 452
- Rothenberg, T.J., 332
- Royston, P., 273
- Ruback, R., 403
- Rubinstein, M., 422
- Ruido blanco
  coloreado, 431
  condicionado, 15
  con heterocedasticidad condicionada, 15
  débil, 15
  del criterio de selección, 158, 165
  estricto, 14
  Normal, 14, 449
- Ryzhik, I.M., 547

- Sadi Carnot, N.L., 5
- Said, S., 329
- Samorodnitsky, G., 481
- Samuelson, P., 412, 422, 459
- Savin, N.E., 334
- Schafgans, N.M.A., 491
- Schmidt, P., 131, 327, 331, 514, 533
- Scholes, M., 411
- Schwarz, G., 180, 250
- Schwert, G.W., 331
- Segunda ley de Laplace, 469
- Selección del orden de autorregresividad, 155-189
- Señal del criterio de selección, 158, 164, 168-169, 171, 174, 175, 178-179, 181, 183-184, 185
- Serie
  cronológica, 11, 125
  cronológica lineal, 142

- Shapiro, S.S., 273
- Shepard, N., 509
- Shibata, R., 158
- Shin, Y., 533
- SIC, 180-183, 188, 250
- Sistema
    de ecuaciones en-diferencias, 145, 205
    de ecuaciones estocásticas diferenciales-en diferencias de Kolmogorov, 55, 69, 73, 74, 85, 92, 94, 97, 113
    de funciones de distribución, 7
    de probabilidades de teransición infinitesimal, 55, 94
    de Yule-Walker, 146, 150
    dinámico, 4, 10
    dinámico caótico, 449
    integrable en el sentido de Poincaré, 4
    log-Normal, 264
    S_β de Pearson, 264
    S_U de Pearson, 264
- Situación neutralmente estable, 145
- Slutsky, E., 143
    definición en el sentido de, 8
- Smirnov, N., 266, 267
- Smith, T., 266, 398
- Sobre-diferenciación, 304
- Sowell, F., 514, 533
- Spearman, C., 408
- Sprenkle C.M., 404, 406
- Srinivasan, R., 268
- Stephens, M.A., 268, 270
- Stock, J.H., 332
- Stout, W.F., 415
- Stratonovich, R.L., 438
- Strobel, J., 415
- Stute, W., 415
- Stylized facts, 439
- Subergode, 27
- Submartingala, 415, 416
    continua en el dominio del tiempo, 417
- Sucesión
    absolutamente sumable, 140
    de cuadrado sumable, 140
    de procesos independientes, 20
    de procesos marginales, 25
- Sugiura, N., 168
- Suma de Riemann-Stieltjes, 315, 433
- Superficie de volatilidades, 454
- Supermartingala, 415
- Supuesto de Phillips, 514

- Talacko, J., 472
- Tanaka, K., 534
- Taylor, B.J.R., 465
- Taqqu, M., 515, 481, 548
- Tasa
    de rentabilidad, 441
    de rentabilidad compuesta, 441
    de rentabilidad simple, 441
    instantánea de ocurrencia, 54

- Tasa
instantánea de reproducción, 86
- Tendencia
determinística, 292, 303, 327, 383, 384
estocástica, 292, 301, 327, 375, 382, 383, 384
segmentada, 346
- Teorema
central del límite, 27, 32, 127
central del límite de Avram-Taqqu, 548
central del límite funcional, 33, 315, 421
de Birkhoff-Khinchin, 26
de Borel-Cantelli, 26
de convergencia de la martingala, 413, 417
de Davis-Resnick, 486
de descomposición de Wold, 34, 303
de Donsker, 314
de existencia de Kolmogorov, 8
de Feller-Lundberg, 84
de Girasinov, 428, 445
de Herglotz, 131
de Kolmogorov-Daniels, 8
de Mann-Wald, 148
de Nabeya-Tanaka, 534
de Reuter-Ledermann, 73
de Slutsky, 167
de Wold, 34-37, 143
- Teoremas
asintóticos de Gnedenko, 472
de Strook-Taylor, 460
de Wiener-Khinchin, 131
ergódicos, 27, 127
- Teoría
de la predicción de Kolmogorov-Wiener, 276-287, 350
de la predicción en procesos no-estacionarios, 349-353
del caos, 449
ergódica, 10
- Teräsvirta, T.C., 190
- Tesis de Turing-Church, 157
- Test
de Anderson-Darling, 271-272
de Cowles-Jones, 389-392
de Cranér-von Mises, 269-270, 271
de diagnóstico, 137
de Dickey,-Fuller, 317-332, 344
de Dickey-Fuller aumentado, 318, 332, 533
de Dickey-Pantula, 333
de Doornik-Hansen, 265
de Durbin-Hausmann, 327
de Hegazy-Green, 274-275
de Jarque-Bera, 264-265
de King-Hillier, 533
de Kolmogorov-Smirnov, 266, 269, 272
de Kwaiatkowski-Phillips-Schmidt-Shin, 533-545
de linealidad de Keenan, 190
de Ljung-Box, 259
de Lo.McKinlay, 400-402
de Monte Carlo, 317
de Mood, 392-393

- Test
de Moore, 272-273
de no-estacionariedad, 313
de Phillips-Perron, 322, 344
de raíces unitarias, 312-332
de Ramsey, 190
de Rappoport-Reichlin, 346
de Robinson, 533
de Royston, 273
de Said-Dickey, 329
de Schmidt- Phillips, 327
de Shapiro-Wilk, 273
de Sherman, 270
de Smirnov, 269
de Stephens, 270
de Wilk-Francia, 275-276
de Wu, 533
de Yap-Reinsel, 330
de Young, 405-407
DF-GLS, 332
KPSS, 533, 534, 535, 536
Q, 394-398
τ de correlación de Kendall, 388
χ^2, 271-272
- Thombs, L., 398
- Tiago de Oliveira, J., 477, 478
- Tiku, M.I., 468
- Tiempo
a transcurrir en un estado al cabo de n transiciones, 49
a transcurrir hasta la extinción, 96
a transcurrir hasta la primera ocurrencia de un evento, 58
a transcurrir hasta la absorción (en procesos de nacimiento-fallecimiento), 80, 81
de ocupación de un estado al cabo de n transiciones, 49
de recurrencia por segunda vez, 48
de recurrencia por k-ésima vez, 48
esperado de recurrencia, 45, 49
promedio,probabilidad de, 27
- Tintner, G., 304
- Tipett, L.H.C., 472
- Toeplitz,
matriz de autocovarianzas de, 36
- Transformación
angular, 300-301
arco seno, 300
de autoafinidad, 429
de Box-Cox, 293, 443
de Fourier, 18
de Granger-Hughes, 300
de Koopman-Harvey-Doornik-Shepard, 509
de potencia, 291-300
de similitud, 429
desvío equivalente, 301
de Tukey, 293-294
función de distribución, 301
función de distribución inversa, 301
hiperbólica, 301
logarítmica, 294
logística, 294

- Transformación
    logit, 301
    log-log, 301
    monótona, 301
    probit, 301
    score exponencial, 301
    score Normal, 301
- Trayectoria, 3, 125
    reflejada, 446
- Tremayne, A.R., 251
- Tsai, C.L., 168
- Tsay, R.S., 190
- Tukey, J.W., 291, 293, 294
- Turing, A., 157

- Uhlenbeck, G.E., 451
- Ulam, S., 317

- Valor medio
    estimación del 127-128

- Van Ness, J.W., 513
- Variable
    de estado, 3
    dummy de nivel, 345
    fat-tailed, 532
    input determinística, 345
- Varianza
    del error de la predicción, 279, 280, 281, 284
    descomposición de la, 146
    inversa, 255
    no condicionada, 454
- Variate difference method, 304, 405
- Velazco, C., 528
- Ventana
    de Bartlett, 531
    de Parzen, 133
- Verosimilitud de la predicción, 458
- Vibraciones Normales, 497
- Vida media, 283
- Ville, J., 409
- Vogt, H., 476
- Volatilidad
    condicionada, 454
    histórica, 454-459
    implícita, 459-462
    promedio, 462
- Volterra, V., 85
    desarrollo de, 140
- von Mises, R., 38
- von Smoluchowski, M., 10, 38, 422, 423

- Wald, A., 148
- Wallis, K.F., 294, 519
- Weibull, W., 476

- Weichselberger, K., 465
- Welsh, A.H., 491
- West, K.D., 534
- White, A., 465
- White, H., 407, 410
- Whiteman, C.H., 334
- Whittle, P., 350, 524, 527
- Wicksell, S.D., 465
- Wiener, N., 143, 422
- Wiggins, J.B., 506
- Wilk, M.B., 273
- Williams, R.J., 436
- Wilson, E.B., 295
- Winter, A., 415
- WN
    - condicionado, 15
    - con heterocedasticidad condiconada, 15
    - con orden de diferenciación fraccionario, 429, 514-520
    - estricto, 14
    - fraccionalmente integrados, 514
    - Normal, 14
- Wold, H., 34, 143
    - teorema de descomposición de, 34-37, 143
- Woodbury, M.A., 100
- Woodward, W.A., 545
- Working, H., 404
- Wu, P., 533

- Yap, S.F., 330
- Young, W.E., 405
- Yule, G.U., 83, 88, 100, 143, 145

- Zhang, N.F., 545
- Zivot, E., 346