Tests de Dominancia Estocástica en base a Estadísticos de Kolmogorov-Smirnov Multivariados, con Aplicaciones al Estudio de la Desigualdad Económica Multidimensional

Perez, Luciano

2016

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS ECONOMICAS
DOCTORADO

Tesis

Tests de Dominancia Estocástica en base a Estadísticos de Kolmogorov-Smirnov Multivariados, con Aplicaciones al Estudio de la Desigualdad Económica Multidimensional

Alumno: Luciano Perez
Directora de Tesis: Juana Z. Brufman
Miembros del Tribunal de Tesis: Luis A. Beccaria, Alberto Landro y María Emma Santos
Fecha de defensa de la Tesis: 3 de Agosto de 2016

Índice general

Prefacio

El presente trabajo de Tesis Doctoral se realizó en el Instituto de Investigaciones en Administración, Contabilidad y Matemática (*IIADCOM*) de la Facultad de Ciéncias Económicas de la Universidad de Buenos Aires, bajo la dirección de la Dra. Juana Z. Brufman.

El mismo pudo ser llevado a término principalmente gracias al financiamiento del CONICET, en la forma de una Beca de Posgrado Tipo I (2007-2010) y una Beca de Posgrado Tipo II (2010-2012), también dirigidas por la Dra. Brufman.

Algunos agradecimientos son de rigor: a Hernán Benítez, por innumerables aportes bibliográficos, a Sergio Gevatschnaider, por interesantes discusiones sobre los diversos alcances y aplicaciones del presente trabajo, a Germán Chiappe por su rápida respuesta a consultas específicas.

Sobre el contenido del trabajo, sin entrar en detalles técnicos, quiero dejar clara una filosofía subyacente. La premisa ha sido en todo momento diseñar herramientas que sean de utilidad práctica. Detrás de los detalles estadísticos y el rigor matemático, siempre se tuvo en cuenta al *usuario* final de estas páginas, léase el econometrista empírico y el economista aplicado, sobre todo aquellos dedicados al diseño y evaluación cuantitativa de la política económica.

En este sentido agradezco nuevamente a la Dra. Brufman que siempre hizo hincapié en que esa es la tarea final del econometrista teórico, brindar un herramental con aplicaciones concretas, y no simplemente demostrar teoremas. Espero haber podido en estas páginas convertirme en fiel expositor de dicha guía.

Capítulo 1

Introducción, Planteo y Objetivos

1.1. Desigualdad y Bienestar Económico

1.1.1. Distribución del Ingreso y Bienestar

Desde los inicios de la Ciencia Social la desigualdad entre los individuos de una determinada sociedad ha sido un tema de importancia fundamental. La principal característica del enfoque científico de la desigualdad radica en que, a diferencia del enfoque filosófico que se basa en la noción de *justicia* como un fin en si mismo, la Ciencia Social ha encarado el estudio de la desigualdad (en alguna de sus acepciones) como causa potencial de ciertos fenómenos sociales.

J-J. Rousseau [181], en su *Discurso sobre el Origen de la Desigualdad*, plantea que la desigualdad generada a partir de la propiedad es la causa principal del *conflicto social*, hipótesis ésta que ha sido reevaluada bajo diversas metodologías y con distintas intepretaciones a lo largo de los años (cfr. Perez [166]).

La teoría marxista del conflicto social (*la lucha de clases*), por ejemplo, parte de una idea básica que es la desigualdad en la propiedad de los medios de producción. Desde al menos el siglo XIX entonces, el foco de estudio en la desigualdad se ha puesto en las variables o dimensiones *económicas*, es decir en condiciones materiales tales como los recursos (productivos y/o de consumo) al alcance de los individuos.

Por su parte, la literatura económica sobre Argentina hasta los años ochenta (e.g. Llach [134]) hacía hincapié en el conflicto en su forma de puja distributiva entre diversos sectores económicos. Por un lado entre trabajadores y empresarios (distribución funcional) y por otro entre sectores vinculados de forma diferencial con el mercado internacional (exportadores de productos agroindustriales vs. empresarios del mercado local). Tristemente, este fenómeno y los análisis al respecto han vuelto a surgir en los últimos años, siendo una característica típica de economías de alta inflación.

Frente a este enfoque de carácter *positivo* (es decir sobre las *consecuencias*) de la desigualdad económica, la economía moderna del bienestar (Boadway [29]) recupera en cierta forma el carárcter *normativo* involucrado en los enfoques filosóficos.

Esto se deriva del hecho de que la economía del bienestar, trabajando con funciones de bienestar social agregado (funcionales de Bergson-Samuelson), incorpora implícitamente alguna noción de justicia social determinada, como señala Sen [199]. Según dicho autor, la principal diferencia entre los conceptos absolutos de bienestar o justicia social radica en sus distintas *bases de información*. La base de información de una teoría de la justicia social consiste en el conjunto de variables y condiciones que entran en los juicios de valor.

Lo más usual en la práctica, sobre todo en estudios cuantitativos, es un supuesto *utilitarista* en el cual los funcionales son integrales (o sea *sumas*) de funciones de utilidad crecientes en el ingreso, como señalamos en la Sec.1.2.1 (Deaton [46], Lambert [126]).

La idea fundamental entonces, en este marco, es evaluar cómo impacta la desigualdad en la asignación de recursos sobre el *bienestar* agregado de la sociedad. Para llevar esto a la práctica, sin embargo, es necesario especificar de manera concreta de qué forma se está evaluando cuantitativamente la idea de desigualdad. El trabajo de Atkinson [13] desde la década de 1970 ha sido fundamental en este sentido, relacionando la desigualdad en la distribución del ingreso con el bienestar agregado para clases bastante amplias de funciones de utilidad individuales, en base a ciertas medidas de desigualdad.

En el enfoque de Atkinson una medida de desigualdad (p.ej. la dominancia Lorenz, o el coeficiente de Gini) es útil en la medida en que permita hacer juicios relativos de bienestar. El resultado principal de este enfoque es el siguiente [13][126]:

Teorema de Atkinson. *Supongamos dada una función de bienestar cualquiera que agrega utilidades individuales que son crecientes con el nivel de ingreso. Entonces si la distribución A domina a la distribución B en el sentido de Lorenz, el bienestar derivado de la distribución A es mayor que el derivado de la distribución del ingreso B.*

El resultado vale siempre que se mantega igual el ingreso medio (o sea mayor en la economía menos desigual). El Teorema de Atkinson data de 1970 y dio origen a buena parte de la teoría moderna de la distribución del ingreso en el marco de la economía del bienestar.[1]

Por supuesto, las consecuencias en términos positivos de la desigualdad no son ajenas al estudio del economista actual. La idea de la desigualdad como raíz del conflicto social es parte de la razón por la cual tiene interés su estudio, aun cuando

[1] Amartya Sen, al comentar las *Radcliffe Lectures* en la Univesidad de Warwick en 1972, señala en primer lugar la importancia del trabajo de Arrow y agrega: *"From a different direction, A.B. Atkinson's works on inequality measurement much influenced the 1972 Radcliffe Lectures..."* (Sen [199], p.xi).

esté *mediado* por el recurso de las nociones de bienestar social. Como señala el propio Sen [199], la naturaleza misma del problema se relaciona de manera singular con la emergencia de conflicto social:

> *"The relation between inequality and rebellion is indeed a close one, and it runs both ways. That a perceived sense of inequity is a common ingredient of rebellion in societies is clear enough..."* (Sen [199], p.1)

Puesta a prueba rigurosamente, sin embargo, la hipótesis de la desigualdad como origen del conflicto social no supera el contraste empírico, al menos no en el enfoque usual de la desigualdad en la distribución del ingreso. El trabajo de P. Justino [112] contrasta empíricamente diversas hipótesis en torno a la relación entre desigualdad y conflicto social, utilizando datos de India (1973-2000), no encontrando una correlación concluyente entre dichos fenómenos.

Frente a ello, algunos autores plantean la inadecuación de las *medidas* de desigualdad en la distribución del ingreso para predecir el conflicto social. Los trabajos de Esteban y Ray [68] y de Duclós, Esteban y Ray [57] desarrollan el concepto de *polarización del ingreso*, en un marco que definen como de *identificación-alienación*.

Según ellos, una sociedad en la cual existen dos picos (*identificación*) de la distribución del ingreso, uno bajo y otro alto (*alienación*) presentará mayor conflictividad que otra con mayor variabilidad de ingresos. Pero las medidas de desigualdad usuales, al menos hasta el punto que satisfagan el axioma de transferencias de Dalton-Pigou, no siempre identificarán el problema (cfr. Perez [166]).

Otros autores (Midlarsky [151]) han puesto énfasis en la *desigualdad extrema* como origen del conflicto social.

El mencionado trabajo de Justino [112], señala la particular estructura de castas de la India, junto con factores religiosos, como posible fuente de cohesión social. **Esto implica que para entender la relación entre la desigualdad y el conflicto social, no alcanza con quedarnos en la mera distribución del ingreso, sino que debemos incorporar otras variables.**

Si bien nuestro presente trabajo de Tesis no apunta a la relación entre desigualdad económica y conflicto social, creemos que el desarrollo de un herramental cuantitativo que contemple diversas dimensiones será de utilidad para el contraste de hipótesis multidimensionales en dicho ámbito.

Por otro lado, al expandirse el alcance de la teoría económica, el impacto de la desigualdad ha sido evaluado en diversas ramas, particularmente en la teoría del crecimiento económico. En su trabajo de 1991 Alessina y Rodrik [5] presentan un modelo de crecimiento endógeno con conflicto distributivo y llegan a la conclusión de que una mayor desigualdad, en una sociedad democrática (donde se cumple el *Teorema del Votante Mediano*), llevará a una mayor presión impositiva y a una menor tasa de crecimiento.

P. Aghion [3] estudia la dinámica de la desigualdad en los salarios en el marco de la teoría schumpeteriana del crecimiento, destacando la desigualdad entre grupos educacionales y su relación con los ingresos sectoriales. La desigualdad *entre* grupos educacionales, según dicho autor, explica gran parte del crecimiento de la desigualdad en USA y Gran Bretaña[2] en las dos últimas décadas del siglo XX. El efecto del componente de variabilidad *intra* grupo es de carácter transitorio sobre la desigualdad en los ingresos.

Destaca entonces, el papel de la desigualdad sobre el crecimiento, pero a la vez, tanto en el enfoque *endógeno* como *schumpeteriano*, no se puede dejar de lado la relación básica entre ingresos y educación, sea como determinantes uno de otro, o como determinantes del crecimiento. **Frente a ello, juzgamos de gran utilidad introducir en la presente Tesis herramientas econométricas para el estudio conjunto de ambas dimensiones de la desigualdad económica.**

Por último, señalemos que el estudio de la desigualdad ha cobrado relevancia también en el análisis de la economía mundial. El trabajo de 2001 de Cornia y Court [41] da cuenta del aumento que ha tenido la desigualdad en la distribución del ingreso en tres niveles:

- *Global*, es decir entre todos los individuos del mundo.
- *Internacional*, es decir entre distintos países del mundo.
- *Nacional*, es decir entre los individuos de un mismo país.

Por supuesto los niveles nacional y global están estrechamente ligados, en el sentido de que el aumento de la desigualdad global se explica en gran parte por la creciente desigualdad dentro de los países.

Es de esperar que los distintos niveles de desigualdad observados, y sobre todo su dinámica a través del tiempo (en base a lo mencionado más arriba sobre la teoría del crecimiento) tengan que ver con la correlación entre las distintas dimensiones y no sólo con el nivel de ingreso. En este sentido, los tests introducidos en nuestro trabajo podrán encontrar utilidad en la comparación internacional.

1.1.2. Dimensiones de la Desigualdad Económica

Si bien en la Sec.1.1.1 hemos tratado con mayor énfasis el análisis de la distribución del ingreso en la teoría económica, lo cierto es que la ciencia social ha considerado la desigualdad como un fenómeno multidimensional.

El concepto fundamental subyacente es el de *estratificación* de los individuos de una sociedad en base a uno o más atributos objetivos. En el estudio de la distribución del ingreso tal como lo vinimos tratando, por ejemplo, se ordena a los

[2]Relativo al efecto de las diferencias educacionales en la distribución del ingreso, el trabajo de Gasparini [78] que citamos más abajo (Sec.1.6.5) concluye algo similar para el caso argentino.

individuos según el valor monetario de su ingreso neto en determinado período de tiempo.

Sin embargo, éste no tiene por qué ser el único criterio de estratificación. Max Weber [229] introdujo el concepto de *estamento* en el cual juegan un rol fundamental el *modo de vida*, distintas formas de educación formal y el *prestigio* hereditario o profesional.

El concepto de estamento complementa al de *clase social* y permite explicar situaciones de conflicto en estructuras sociales no necesariamente correspondientes a lo que llamaríamos capitalismo moderno, sino también en sociedades tradicionales (sociedad de castas en India o Japón) o hierocráticas (en diversas sociedades islámicas).

Más contemporáneamente, el trabajo de McAll [148] analiza el papel de los factores *étnicos* en la desigualdad social, integrándolos con el concepto marxista de clase social. Como señalamos más arriba (Sec.1.1.1), todos estos factores pueden ser relevantes a la hora de comprender fenómenos sociales, en particular las situaciones de conflicto.

La sociología y la economía aplicadas han reducido el universo de dimensiones posibles mediante la idea de *status socioeconomico* (SES), una jerarquía entre los individuos de una sociedad que resume las diferentes fuentes de estratificación. Los sub-conceptos que indican el estatus socioeconomico de un individuo, y que se utilizan en la práctica en las investigaciones empíricas para determinarlo, se encuadran en tres dimensiones (Marks et al. [144]):

- Características ocupacionales
- Nivel de educación formal
- Ingreso y riqueza

Dichas tres dimensiones son las que usualmente se emplean para llevar a cabo la estratificación (ordenamiento) de los individuos y a partir de ella estudiar, por ejemplo, la desigualdad presente en la sociedad.

Por razones técnicas (que señalamos en las Sec.1.3.1, 1.3.2 y 1.5.1) el estudio empírico suele reducirse a un enfoque unidimensional. La economía del bienestar, como hemos señalado en la Sec.1.1.1, ha encarado con prioridad el estudio de la distribución del ingreso. La sociología (Marks [144]) ha optado en general por resumir las tres dimensiones en una variable numérica a través de alguna normalización.

En el presente trabajo de Tesis se plantea como objetivo general[3] abordar la problemática del estudio *empírico-cuantitativo* de la desigualdad económica en su carácter intrínsecamente multidimensional, sin la necesidad de reducir

[3]Como detallamos en la Sec.1.6.1.

los vectores de atributos individuales a una única variable numérica como el ingreso o el SES.

La relevancia de esta ampliación del estudio de la desigualdad radica, a nuestro juicio, en que de esta forma se incorporan dimensiones sin duda relevantes en cuanto al bienestar social, y a la vez se evitan las necesarias arbitrariedades involucradas en la reducción de las distintas dimensiones a un simple índice numérico.

El método que hemos elegido para este desarrollo (descrito en las Sec.1.2.2, 1.3.2 y 1.6.1) es consistente con el punto de vista de la economía del bienestar, puesto que la idea básica que relaciona desigualdad y bienestar en el Teorema de Atkinson se mantendrá en el caso multidimensional para la caracterización cuantitativa de la desigualdad escogida (Atkinson y Bourguignon [14]).

Por último, debemos señalar que, una vez definidas las dimensiones de estratificación, existen al menos dos enfoques posibles sobre la desigualdad: la desigualdad *vertical* y la *horizontal*. El primero se enfoca en el nivel de las variables seleccionadas como relevantes y ordena a los individuos sólo en base a ella. La segunda, por su parte, estudia la desigualdad entre grupos sociales que se forman de acuerdo a características comunes no directamente relacionadas con la variable en estudio (e.g. raza, género, religión, etc.).

En el presente trabajo de Tesis planteamos concretamente la utilización de las herramientas econométricas desarrolladas para el estudio de la desigualdad en el sentido vertical (cfr. Sec.1.6.5). La razón principal es que se trata del enfoque más abordable a través de los datos relevados en encuestas de hogares. Por otro lado, si bien existen casos como el de las comunidades indígenas en el norte del país donde puede plantearse la relevancia de la desigualdad horizontal, la realidad de los aglomerados urbanos que se toman en las muestras es mucho más homogénea.

1.1.3. Capital Humano

Hemos señalado en la Sec.1.1.2 que el estudio la desigualdad económica abarca típicamente un abanico de tres dimensiones de estratificación. Las mismas, sin embargo, claramente **no son independientes**. La capacidad de un individuo para generar ingresos, por ejemplo, depende naturalmente a corto plazo de su situación laboral.

Por otro lado, si bien la dimensión dada por el ingreso permite darnos una idea del bienestar de los individuos en un instante determinado, las diferencias en sus niveles de educación formal y experiencia laboral determinarán situaciones distintas a mediano y largo plazo. Por tanto, en un enfoque de *ingreso permanente* o de ciclo vital, es necesario tener en cuenta las diferencias entre los individuos que se derivan de las demás dimensiones además del ingreso corriente.

Esta idea dinámica del efecto de las características individuales sobre el bienestar del agente económico a lo largo de su ciclo vital es incorporada en la noción de *capital humano* introducida desde la década de 1970 por los trabajos de Mincer

[152] y Becker [18] entre otros.

La idea básica de Mincer es la de una *función de ingresos potenciales* (*earning power*) según la cual la capacidad de un individuo para percibir ingresos a lo largo de su vida se deriva de los conocimentos adquiridos en su educación formal y experiencia. En base a ello, por ejemplo, un individuo elegirá racionalmente pasar cierto lapso de su vida con menores ingresos corrientes (cuando trabaja menos horas para dedicar tiempo a su educación) a cambio de mayores ingresos futuros derivados de su mayor productividad (*retornos a la educación*). En este sentido, el agente económico está *invirtiendo en Capital Humano*, para poder percibir los frutos de dicho capital a lo largo de su ciclo vital.

Las variables incluidas en las funciones de ingreso mincerianas han pasado de las básicas (educación formal, experiencia laboral) de Mincer en 1974, hasta un amplio abanico que incluye diversas variables de control como género, salud, etc (Polachek [170]). Debe destacarse, sin embargo, que aun manteniendo las variables básicas, el capital humano no sólo impacta de manera implícita en el bienestar a través de su efecto sobre el ingreso, sino de manera directa. Por ejemplo, un nivel mayor de capital humano en el sentido de años de educación formal está en general positivamente correlacionado con mejores niveles de salud (Wößmann [238]) y en consecuencia con el nivel de utilidad que el individuo puede obtener de su ingreso.

Por estas razones, en la presente Tesis, hemos optado por llevar a la práctica nuestro objetivo general (desarrollo de tests multivariados de desigualdad) en la forma de tests *bivariados* en base a las dimensiones de *ingreso* y *capital humano*.

La idea básica tras este enfoque es que, si bien una determinada sociedad puede tener una estructura de ingresos corrientes más igualitaria que otra, esta situación puede no ser sostenible a largo plazo si la distribución de las capacidades de los individuos para generar esos ingresos (capital humano) están distribuidas de manera desigual.

El Cap.2, Sec.2.1.2 desarrollamos más detalladamente el concepto de capital humano y su interpretación en la economía del bienestar. En el Cap.6, Sec.6.2 damos detalles sobre la forma operacional que toman dichas variables en nuestro trabajo empírico, así como las fuentes de datos utilizadas.

1.1.4. Economía Positiva, Normativa y Política Económica

Como señalamos en la Sec.1.1.1, la ciencia social en general, y la economía en particular, diferencia dos aspectos básicos de la desigualdad: el *normativo* y el *positivo*; ambos aspectos, sin embargo, se encuentran relacionados.

Por un lado, el *origen* de la desigualdad es un aspecto en principio positivo (histórico) pero sobre el cual se realizan asimismo juicios de valor normativos.

> *"Telle fut, ou dut être, l'origine de la société et des lois, qui donnèrent de nouvelles entraves au faible et de nouvelles forces au riche, detrisirent sans retour la liberté naturelle, fixèrent pour jamais la loi de la propriété et de l'inégalité, d'une adroite usurpation firent un droit irrévocable, et pour le profit de quelques ambitieux assujettirent désormais tout le genre humain au travail, à la servitude et à la misère"*. J.J.Rousseau [181].

En el enfoque marxista, el proceso de *acumulación originaria* da origen a la estructura de propiedad de los medios de producción necesaria para el modo de producción capitalista, el cual la reproduce y amplía (Marx [145]).

El trabajo pionero de Sir Francis Galton (Slottje [211]) se centró en las diferencias de *productividad* de los trabajadores, y en consecuencia esperaba encontrar una distribución normal en los ingresos, mientras Pigou señalaba una asimetría derivada de la riqueza heredada. En la década del 30 los estudios de Gibrat y Chapernowne, introducen el *origen estocástico* de la desigualdad y mediante modelos markovianos llegan a la distribución de Pareto, aun usada en el modelado de las curvas de Lorenz (Nygård y Sandström[158]).

Midlarsky [151] deriva *geométricamente* la distribución de Pareto, en base a un modelo de *apropiación colonial* de los recursos económicos.

Dentro de nuestro marco conceptual, el de la moderna economía del bienestar, si bien el *origen* de la desigualdad no es un objeto de estudio en sí mismo, las *consecuencias* en términos de pérdida de bienestar social son el centro de atención.

Pero si estas consecuencias son juzgadas desde un marco normativo, es porque en última instancia se plantea la necesidad de mejorar la situación, es decir, la necesidad de una *política económica*.

A su vez, esta política económica tiene que tomar en consideración las respuestas estratégicas de los agentes económicos para poder cumplir con su objetivo, es decir, debe enfrentar un problema de *implementación*. La teoría del *diseño de mecanismos*, la *ingenieria de juegos* y las *finanzas públicas* (entre otras ramas) tratan los problemas derivados de la implementación de las soluciones que se consideran deseables desde lo normativo, en distintos arreglos institucionales (Bowles [31]).

Debe recalcarse además, como señala Sen [199], que subyacente a los enfoques tanto normativos y filosóficos, como positivos e históricos de la desigualdad, existe un ineludible componente *intuitivo* en el concepto. Y gran parte de la tarea de los mencionados enfoques es justamente operacionalizar este componente intuitivo, algo no tan sencillo como podría suponerse.[4]

De manera general, el concepto *intuitivo* de igualdad en la distribución tiene principalmente que ver con la *simetría* y con la *envidia*. Estos conceptos se formalizan en la economía matemática y el concepto de una asignación *justa* es el de una

[4]En el *Fedro* de Platón [168], Sócrates nos dice: *"Pero que se nos hable de lo* justo *y de lo* injusto *y estas palabras despiertan ideas diferentes y nos ponemos en el momento en desacuerdo con los demás y con nostros mismos"*.

asignación *libre de envidia* (*envy free*, cfr. Thomson y Varian [219]), o sea aquella en la que ningún agente económico prefiere el resultado de otro.

Por caso, si el problema de la *distribución* lo ilustramos como el reparto de una torta entre dos personas, el concepto intuitivo de *igualdad* o más bien de *justicia* nos dirá que asignemos la mitad a cada uno.[5] Sin embargo, si contamos además con la información de que uno de ellos compró los ingredientes y la cocinó sin ayuda del otro, probablemente la asignación *igualitaria* no nos parezca necesariamente la más *justa*.

Por otra parte, darle al primer individuo la totalidad de la torta *porque le pertenece legítimamente* puede no ser lo óptimo desde el punto de vista *colectivo*: probablemente al comer un poco más de la mitad de la torta esté saciado y ya no quiera o disfrute el resto, que se desperdicia.

Los economistas diremos que esta última asignación no es *Pareto eficiente*, en el sentido de que el primer individuo podría haber quedado igual de satisfecho dando cierta parte al segundo, que estaría mucho mejor que con las manos vacías. En este sentido, probablemente tampoco llamaríamos *justo* al reparto que implica un desperdicio.

Lo que esta metáfora intenta mostrar, es que los juicios sobre igualdad o justicia, en el enfoque de la *distribución* de los recursos económicos, son complejos. Por un lado, la distribución no puede dejar de lado a la *producción* en su consideración de justicia.

Por otro, resulta innegable que los juicios de justicia o igualdad, deben ser *de carácter colectivo* y tener como guía no simplemente una noción *filosófica de justicia* sino una idea de *bien común*. En este sentido, dentro de la ciencia económica, el estudio de la desigualdad se enmarca en la *teoría de la elección social* (Sen [196]).

Ahora bien, aún en el caso en que estuvieramos de acuerdo en una definición de la igualdad y sólo nos preocupara la distribución, queda aun por resolver el *problema de implementación* que señalamos más arriba. Volvamos al ejemplo de la torta con dos comensales.

Si le damos el cuchillo a uno de ellos, y la posibilidad de decidir por sí solo como repartir el pastel, lo más probable es que se quede, si no con todo, al menos con la mayor parte del mismo. No por un egoísmo patológico, sino por su mera racionalidad como consumidor, ya que su utilidad crece al consumir mayor cantitdad del bien.

Pero este reparto, intuitiva o técnicamente, sabemos que no sería justo. Un *mecanismo* clásico que lidia con esta situación es el de *"I cut, you choose"*, en el cual le damos al primer individuo la potestad de elegir cómo partir en dos la torta (i.e. le damos el cuchillo) y al segundo la potestad de elegir cuál de los dos pedazos

[5] Aunque parezca elemental, el problema de la división justa de una torta, introducido por el matemático Hugo Steinhaus en 1948 [214], es todavía una rama activa tanto de la ciencia económica (Chen [38]) como de las matemáticas (Hill [101]).

quiere. Este mecanismo *resuelve* el problema, porque el primer individuo se verá obligado a cortar la torta en partes iguales si no quiere correr el riesgo de quedarse, casi con seguridad, con la parte más pequeña.

Pero la *realidad* puede tener otras variantes que no controlamos tan fácilmente. Supongamos que el corte de la torta no es igualitario. Una vez que el primer individuo acaba de cortar la torta, el segundo elegirá el que más le conviene. Pero, a la hora de decidir cuál de las dos porciones es la *mejor para él*, no dejará de notar que el primer individuo todavía tiene un cuchillo muy grande y afilado en la mano, por lo cual probablemente no querrá dejarlo demasiado descontento con el resultado.

Este ejemplo puede parecer una simple nota de color, pero ilustra otra inherente complejidad de los problemas de distribución de los recursos económicos: una vez que aceptamos la necesidad de *juicios colectivos de bienestar*, no podemos dejar de considerar los mecanismos de *acción pública*, es decir la intervención de un agente económico con potestad de imponer una asignación a los demás.

Los manuales clásicos de Finanzas Públicas, de hecho, señalan al problema de la distribución del ingreso o de la igualdad en general, como una de los fundamentos *racionales* para la existencia del Estado (Stiglitz [215]).

Pero si la distribución de los *recursos económicos* se considera inaceptable en base a algún concepto de igualdad o justicia, también debería utilizarse el mismo juicio sobre la distribución del *poder político*. Darle al Estado la potestad de imponer una solución pública en la distribución (o sea, darle el cuchillo en nuestro ejemplo) puede de todas maneras no tener como resultado un reparto justo.[6]

En la presente Tesis, no tratamos de manera extensiva el *diseño* de la política económica orientada a la reducción de la desigualdad, sino que apuntamos al desarrollo de medidas cuantitativas para su *evaluación*. Es decir, no ahondamos técnicamente sobre las medidas necesarias de política para aumentar el bienestar social pero podremos decir si una determinada medida tiene el efecto deseado o no.

Por supuesto, no consideramos menos relevante el diseño de las políticas, pero varias razones nos llevan a centrarnos en un objetivo más limitado. Por un lado, dichos problemas darían tema para una o más tesis doctorales por sí mismos, abarcando áreas desde la economía de la educación, la economía laboral, las finanzas públicas, etc, hasta la macroeconomía del crecimiento y el desarrollo.

Pues si bien la teoría económica de los impuestos/transferencias y su efectividad como ecualizadores de *ingresos* es conocida (Lambert [126], Salanie [187]), no ocurre lo propio con las demás dimensiones de la desigualdad. Por caso, no existe una analogía directa de la *transferencia de ingresos* para algo como la educación o el capital humano. Por ello, la política en un enfoque multidimensional de la de-

[6]En un extremo, un marxista nos dirá que el Estado es un instrumento de explotación del proletariado por parte de la clase capitalista. En el otro, un liberal libertario declarará cualquier avance del Estado más allá de un mínimo indispensable, como una violación de sus derechos individuales. En el medio, cada uno de nosotros estará más o menos satisfecho con la provisión de bienes públicos, y probablemente *disatisfecho* a la hora de pagar impuestos.

sigualdad es mucho más compleja y debe ser encarada desde varias ramas de la economía, recalcando el papel fundamental de las interacciones entre las distintas dimensiones[7].

Por otro lado, creemos que la existencia de herramientas cuantitativas de evaluación de la política es de utilidad fundamental aún en el proceso de diseño de la misma, porque al combinarse con los métodos de *simulación* permite una optimización de las políticas en base al resultado esperable de las mismas.

[7]En el Cap.8 damos algunos detalles más.

1.2. Economías de Distribución, Desigualdad y Bienestar como Preordenes

1.2.1. Rankings, Utilitarismo e Igualitarismo

La principal finalidad de la economía del bienestar es proveer una forma razonable de dirimir qué situaciones económicas son socialmente preferidas a otras. Es decir, apunta a conformar una *ordenación* o *ranking* de las asignaciones según las *preferencias sociales*.

Hasta qué punto es posible hacer esto con el mayor grado de razonabilidad (e.g. *transitividad* y *completitud*) y generalidad (e.g. *dominio universal*) de una manera aceptable (e.g. *no dictatorial*) ha sido tema de debate científico desde el trabajo seminal de Arrow [10] hasta el día de hoy (Maskin y Sen [147]).

Los *órdenes de bienestar social* (o *SWO* por sus siglas en inglés) parten de una serie de juicios previos de valor que permiten hacer comparaciones significativas. Naturalmente estos *supuestos* varían según lo que cada investigador considere relevante, y tendrán mayor o menor aceptabilidad entre los demás (Boadway y Bruce [29]).

Un supuesto *débil* como el de Pareto: *"Si todos los individuos prefieren el estado A al estado B, la sociedad también lo hará"*, es en general mucho más comunmente aceptado que uno *fuerte* como el de Rawls: *"Una reasignación de recursos es justa si y sólo si opera a favor del individuo en peor condición"*.

Por otra parte, como dijimos más arriba, Sen [198] muestra como la principal diferencia entre los conceptos absolutos de bienestar o justicia social radica en sus distintas *bases de información*.

Nuestro enfoque es el de la economía moderna del bienestar, que resulta esencialmente *utilitarista* o *bienestarista* (*welfarist*) pues incorpora como base de información las utilidades individuales de los agentes económicos (cfr. Johansonn [110] Ch.3).

Si bien existen diversas razones para considerar otros elementos a la hora de juzgar el bienestar social (Sen [199]), la profesión se ha centrado en aquellas. Esto es particularmente visible en los estudios cuantitativos, y queda plasmado en las principales obras de referencia, como el notable libro de Deaton [46] sobre el estudio microeconométrico de las políticas de desarrollo.

Nótese también que este enfoque adopta entonces el *individualismo metodológico*, en el sentido de que las preferencias individuales no son ignoradas al definir las de la sociedad.

Para fijar ideas vamos a definir algunos conceptos abstractos, pero de utilidad práctica, siguiendo a Perez [166].

Definición 1.2.1.1 (Economías de Distribución)**.** *Llamamos* economía de distribución *a un par* $E = (S, \mathfrak{a})$ *donde* $S = (\Omega, \Sigma, \mu)$ *es un espacio de medida y* $\mathfrak{a} : \Omega \to D$ *es una función medible.* □

Aquí D es algún espacio de Banach que usualmente llamamos *espacio de bienes*, porque incluye al dominio de las utilidades individuales, y lo vamos a considerar fijo para comparar economías. Al conjunto Ω se lo denominará típicamente el conjunto de *individuos* u *hogares* de la economía E. La función $\mathfrak{a}$ será una *asignación* y se considera medible respecto al álgebra de Borel en D.

Nosotros trabajaremos siempre el caso en que $D = \mathbb{R}^n$, pero en principio podría ser cualquiera, incluso de dimensión infinita como en Khan y Yannelis [113]. La medida μ la tomaremos tal que $\mu(\Omega) = 1$ (y en general la llamaremos P) y entonces la asignación $\mathfrak{a}$ es un vector aleatorio de n dimensiones.

Consideremos la clase $\mathscr{E} = \{E$: *E es una economía de distribución*$\}$. Un *orden de bienestar social* (SWO) será un preorden sobre $\mathscr{E}$, es decir una relación binaria reflexiva y transitiva.

Existen infinitas maneras de definir un SWO; la que adoptamos aquí, siguiendo la práctica usual, es la de los *funcionales de Bergson-Samuelson*[8] que agregan utilidades individuales en una *función* de bienestar. Para ello consideramos dada una familia de funciones de utilidad $\mathscr{U}$ con las propiedades usuales (Mascollel et al. [146]) y una aplicación $u : \Omega \to \mathscr{U}$ que asigna a cada individuo una función de utilidad determinada $u(\omega) = u_\omega : D \to \mathbb{R}$, que pueden estar definidas no necesariamente sobre todo D sino sólo sobre la imagen de $\mathfrak{a}$ si es necesario.

Ahora para cada $\omega \in \Omega$ tenemos un número real $u_\omega(\mathfrak{a}(\omega))$ y así queda definida una función $f_{u,E} \in \mathbb{R}^\Omega$ vía u. El principio de *utilitarismo* o *bienestarismo* (Boadway y Bruce [29]) nos dice la función de bienestar social W_{BS} sólo debe depender de la asignación $\mathfrak{a}$ vía las utilidades que los individuos ω obtienen de dicha asignación.

Para ello entonces es necesario contar con una función $W : \mathbb{R}^\Omega \to \mathbb{R}$ y diremos que el *bienestar social* de la economía de distribución $E = (S, \mathfrak{a})$, bajo el supuesto de las utilidades u es:

$$W_{BS}(E) = W(f_{u,E}) \tag{1.1}$$

Si el planteo parece demasiado abstracto, pensemos en una economía finita, digamos con n agentes, cuyas funciones de utilidad son $u_1, ..., u_n$. La economía asignará a cada uno un vector $\vec{x}_i$ de bienes de cual obtienen una utilidad $u_i(\vec{x}_i)$ cada uno. El planteo que acabamos de hacer es que debe existir una función $W : \mathbb{R}^n \to \mathbb{R}$ tal que el bienestar social viene dado por $W(u_1(\vec{x}_1), ..., u_n(\vec{x}_n))$.

Ahora bien, desde el punto de vista microeconómico, las funciones de utilidad no tienen interpretación cardinal ni tiene sentido comparar la utilidad de agentes distintos. En la práctica, sin embargo, es usual dar por supuesto que aunque no tenga una interpretación cardinal absoluta, las utilidades de los distintos agentes pueden si no compararse, al menos agregarse *simétricamente*, lo cual implica que los agentes son anónimos para el bienestar social.

[8] Curiosamente, en español, las iniciales BS de Bergson y Samuelson, se pueden asociar también a *bienestar social*.

En el planteo más simple, el *benthamita* (Johansonn [110]), por ejemplo, podríamos simplemente sumar las utilidades para obtener el bienestar social. Esto nos lleva a funcionales del tipo de Atkinson [13], en los cuales suponemos una única función de utilidad común a todos los individuos, y que son los que suelen tenerse en cuenta al analizar la política económica (Deaton [46]).

Supongamos dada la economía de distribución $E = ((\Omega, \Sigma, P), \mathfrak{a})$ y abusando de la notación digamos que u es la función de utilidad común para todo $\omega \in \Omega$. Entonces:

$$W(E) = \int_{\Omega} u(\mathfrak{a}(\omega))dP \tag{1.2}$$

Para darle una forma más conocida notemos en principio que, como dijimos más arriba, $\mathfrak{a}$ es un vector aleatorio en el caso típico, llamémoslo $\vec{y}$, digamos con función de distribución $F(y)$. En ese sentido podemos reinterpretar 1.2 como:

$$W(E) = \int_{\Omega} u(\mathfrak{a}(\omega))dP = \int_{\Omega} u(y)dF(y) = E(u(\vec{y})) \tag{1.3}$$

donde por una desafortunada notación la última E no es la economía sino la *esperanza matemática* de la variable aleatoria $u(\vec{y})$. Ahora, el resultado es que **una economía tendrá mayor bienestar que otra si la *utilidad esperada* es mayor**. Lo bueno de esta formulación es que con ella, tenemos por añadidura un link directo al concepto de *dominancia estocástica* de la Sec.1.2.2.

Dado que en la mayor parte del resto de la Tesis trabajaremos con este concepto y no con la idea general de bienestar social, puede parecer redundante la introducción que hemos hecho al concepto. Sin embargo, la idea era mostrar como el mismo surge naturalmente al definir e interpretar con cuidado los elementos básicos de la teoría económica del bienestar.[9]

Lo fundamental es ver que en este planteo *à la Atkinson* el concepto de dominancia surgirá como fruto del análisis del bienestar social, y no como un concepto asociado simplemente a loterías y funciones de utilidad de von Neumann-Morgenstern en la teoría del riesgo.

Retomando, una vez que tenemos una función de bienestar social resulta inmediata la construcción del SWO: diremos que $E_1 \preceq E_2$ sii $W(E_2) \leq W(E_1)$. Aquí hemos puesto *antes* en el ranking al de mayor bienestar, o sea, el *primero* será el que mejor esté.

Ahora bien, es evidente que si tenemos cualquier función $T : \mathscr{E} \to \mathbb{R}$ (o sobre cualquier *cpo*) podemos definir un orden en $\mathscr{E}$, inducido de la misma forma que el orden dado por las funciones W_{BS}. Pero es claro que no nos interesa un orden cualquiera dado de esta forma, sino uno que tenga propiedades particulares.

[9] Por otra parte, el planteo usual en las referencias (e.g. Boadway y Bruce [29] o Johansonn [110]) es demasiado simplificado como para una aplicación directa, por eso se juzgó de utilidad hacerlo de esta forma.

El orden dado por 1.3, por ejemplo es claramente *paretiano*, pues si $u(\vec{y}) \geq \tilde{u}(\vec{y})$ para cada $\vec{y}$, la esperanza será obviamente mayor.

Es también *simétrico* en el sentido de que si π es una biyección de Ω en sí mismo (una *permutación*) entonces el bienestar derivado de la economía $(S, \mathfrak{a})$ y el de la economía $(S, \mathfrak{b})$ donde $\mathfrak{b}(\omega) = \mathfrak{a}(\pi(\omega))$ son los mismos, porque la distribución del vector $\vec{y}$ es la misma.

Una familia de funciones con propiedades particulares deseables es también la de los llamados *índices de desigualdad*:[10] $\mathscr{I} = \{I : \mathscr{E} \to \mathbb{R}\}$.

Típicamente se hace el supuesto de que la desigualdad depende de la economía E sólo a través del resultado de su asignación de recursos,[11] o sea del conjunto $D_E = \mathfrak{a}(\Omega)$.

O sea que el dominio de I (abusando de la notación, esta I es el índice y la función a la vez) será en principio el conjunto $\mathscr{D} = \bigcup_E D_E$ donde la unión es sobre todas las economías de distribución E.

Estos índices se usan típicamente para estudiar distribuciones *unidimensionales* de atributos individuales (e.g. el ingreso), y en ese caso se tiene que D_E es un elemento $\vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ o bien una sucesión si la economía es infinita. Considerando economías finitas, Chakravarty [37] define entonces el dominio de I como la unión $\bigcup_n \mathbb{R}^n_{\geq 0}$.

Entre las propiedades que se piden usualmente a los índices de desigualdad, destaca el llamado *principio de transferencias* o de Dalton-Pigou.

Definición 1.2.1.1 (Transferencia)**.** *Sean* $E = (S, \mathfrak{a}))$ *y* $\tilde{E} = (S, \mathfrak{b}))$ *economías de distribución (difieren sólo en la asignación). Supongamos que existen* ω_1 *y* ω_2 *en* Ω *tales que:*

- $\mathfrak{a}(\omega) = \mathfrak{b}(\omega) \ \ \forall \omega \in \Omega - \{\omega_1, \omega_2\}$
- $\mathfrak{a}(\omega_1) \neq \mathfrak{b}(\omega_1)$ *y* $\mathfrak{a}(\omega_2) \neq \mathfrak{b}(\omega_2)$
- $\mathfrak{a}(\omega_1) + \mathfrak{a}(\omega_2) = \mathfrak{b}(\omega_1) + \mathfrak{b}(\omega_2)$

entonces diremos que al pasar *de la economía* E *a* $\tilde{E}$ *hubo una* "transferencia simple" *entre los individuos* ω_1 *y* ω_2.

□

Con economías finitas, y en el caso unidimensional, tendremos una asignación $\mathfrak{a}(\Omega) = (y_1, ..., y_n)$ para E y $\mathfrak{b}(\Omega) = (\tilde{y}_1, ..., \tilde{y}_n)$ para $\tilde{E}$, con y_j representando por ejemplo el ingreso del individuo j. En ese caso, sólo cambiarán un par de y_j en una transferencia simple y la última condición implica que lo que alguno de los dos

[10] Seguimos en principio el planteo de Chakravarty [37], aunque hay que tener cuidado con algunas desprolijidades en el trabajo de dicho autor.

[11] En este sentido uno tendería a llamar a este supuesto también *utilitarista*, no porque haga foco en las utilidades, sino porque dicha escuela siempre abogó por la evaluación de los procesos económicos por sus *resultados* antes que por sus objetivos.

gane, el otro lo *perderá*. Se suele decir entonces que el individuo que termina con menor ingreso del que empezó *le transfirió* ingresos al que terminó con más.

Dado un índice de desigualdad I el principio de Dalton-Pigou nos dice entonces que si en E teníamos $y_i < y_j$ y el individuo i le trasfiere ingresos al individuo j al pasar a $\tilde{E}$, entonces $I(\tilde{y}_1,...,\tilde{y}_n) > I(y_1,...,y_n)$. Es decir, la desigualdad aumenta con las *transferencias regresivas*.

El principio de transferencias de Dalton-Pigou también se le puede pedir a las funciones W, pero en este caso debe ser $W(\tilde{E}) < W(E)$ al darse una transferencia regresiva.

A los índices también se les pide *simetría*, en el mismo sentido que le dimos para W, es decir si $\mathfrak{b} = \mathfrak{a} \circ \pi$ para alguna permutación π de Ω, entonces $I(E) = I(\tilde{E})$.

La *invariancia relativa en escala* nos dice que si $\mathfrak{b} = \alpha \cdot \mathfrak{a}$ para algún número real positivo fijo α entonces $I(E) = I(\tilde{E})$. La invariancia *absoulta* en escala (o invariancia *por traslaciones*) se da cuando la desigualdad es constante ante el cambio $\mathfrak{b} = \alpha + \mathfrak{a}$, de nuevo para algún α real fijo.

El *principio poblacional* es para economías finitas: supongamos que E tiene una asignación $\{y_1,...,y_n\}$ y construyamos una economía $\tilde{E}$ agregando un *gemelo* a cada individuo, es decir una economía con $2n$ agentes y con asignación $\{y_1,y_1,y_2,y_2,...,y_n,y_n\}$. Decimos que $\tilde{E}$ es una *2-réplica* de E.

Si en lugar de 1 añadimos más gemelos de forma que la nueva economía tiene y_j repetido k veces para cada i, decimos que obtuvimos una k- réplica de E. El principio poblacional nos dice que si $\tilde{E}$ es una k-réplica de E, entonces $I(E) = I(\tilde{E})$. Se aplica a las funciones de bienestar W de la misma forma.

Tal como lo enunciamos, el principio poblacional tiene sentido también para economías con infinito numerable agentes, pero en este caso la economía y su k réplica tienen el mismo cardinal, con lo cual la interpretación se puede poner algo abstracta. Para un continuo de agentes habría que formularlo de otra manera.

Un índice que cumple Dalton-Pigou, simetría, normalización y principio poblacional se dice *regular*. En general, los principios que le pedimos a los I los llamamos *axiomas* y decimos que la clase $\mathscr{I}$ de todos los I que cumplen determinada lista de principios está *axiomáticamente determinada* (Chakravarty [37]).

Otras propiedades que se le piden a los I tiene que ver con la *descomponibilidad*. Si $\Omega = \dot{\cup}_{i=1}^{N}\Omega_j$ (el punto indica una unión disjunta) podemos definir economías E_j donde las asignaciones se restringen a los correspondientes a Ω_j. El índice I se dice *descomponible* si $I(E) = B + \sum_{j=}^{N} \mu(\Omega_j)I(E_j)$. El sumando B depende de la partición y $\mu(\Omega_j) \geq 0$ fijos. Esta descomposición replica la idea de la desigualdad *entre* grupos (asociada a B) y la desigualdad *intra* grupos (asociada a los términos de la sumatoria) de ANOVA.

Si la asignación de la economía $\mathfrak{a}$ se escribe como la suma de otras dos funciones $\mathfrak{a}(\omega) = \mathfrak{b}(\omega) + \mathfrak{c}(\omega)$, decimos que $\mathfrak{a}$ se *descompone por fuentes*.

La clase de *entropías generalizadas* de Theil se define como:

$$I_\alpha(\vec{y}) = \frac{1}{\alpha(1-\alpha)}\frac{1}{n}\sum_{i=1}^{n}\left(1-\left(\frac{y_i}{\mu}\right)^\alpha\right) \quad \alpha \in (0,1) \tag{1.4}$$

$$I_1(\vec{y}) = T(\vec{y}) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu}\log\frac{y_i}{\mu} \tag{1.5}$$

$$I_0(\vec{y}) = D(\vec{y}) = \frac{1}{n}\sum_{i=1}^{n}\log\frac{\mu}{y_i} \tag{1.6}$$

Shorrocks demostró que un índice de desigualdad I que sea Lorenz consistente. normalizado, continuo y aditivamente descomponible tiene que ser un múltiplo de positivo de un miembro de la familia de entropías generalizadas (Sen [199]).

Como dijimos más arriba, dada una función I tendremos definido un SWO inducido por ella, por ejemplo $E_1 \preceq E_2$ sii $I(E_1) \leq I(E_2)$, donde el orden de precedencia es entonces para aquella economía con *menor desigualdad*.

Vemos entonces que tanto las funciones de bienestar social como los índices de desigualdad permiten definir un orden parcial sobre el conjunto $\mathscr{E}$ de economías de distribución. Una distinción que debe hacerse entre estas dos clases de órdenes es que usualmente los índices no miden la desigualdad en un sentido *absoluto*, sino por comparación con una economía *igualitaria* de referencia E_0, típicamente una con un ingreso constante (*total igualdad*).

Si a esa economía se le asigna un valor $I(E_0) = 0$ se suelen definir las funciones de forma que a la *total desigualdad* (i.e. una con el ingreso concentrado en un solo agente) tenga un valor 1. Esto se suele llamar *principio de normalización* de los índices de desigualdad.

Es posible, de todas formas, pasar de un índice de desigualdad a una función de bienestar social y viceversa. Por caso, dado un índice I que podemos podemos asignar a una economía E cuyo ingreso medio sea μ el valor $W(E) = \mu(1-I)$ cuya interpretación intuitiva es que al *tamaño de la torta* μ l corregimos hacia abajo por efecto de la desigualdad.

En la otra dirección, la familia de *medidas de Atkinson* se puede pensar que vienen de la forma $I(E) = 1 - \frac{W(E)}{W(E_0)}$. Para que esta familia tenga algunas de las propiedades deseables que señalamos recién, es necesario hacer algunos supuestos sobre la función W, que vamos a pensar como evaluada en $\vec{y} \in \mathbb{R}^n$. Atkinson [13] probó que para el caso de funciones de bienestar homotéticas[12] alcanza con las medidas:

[12] Los economistas (los matemáticos no) llaman *homotética* a una función de la forma $f(x) = g(T(x))$ donde T es homogenea de grado 1 ($T(\lambda x) = \lambda T(x)$) y g es una función real creciente (Simon & Blume [209]).

$$I_{\varepsilon}^{At}(\vec{y}) = \begin{cases} 1 - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu_y}\right)^{\varepsilon}\right)^{1/\varepsilon} & \varepsilon \in (0,1] \\ 1 - \prod_{i=1}^{n}\left(\frac{y_i}{\mu_y}\right)^{1/n} & \varepsilon = 0 \end{cases} \tag{1.7}$$

Si fijamos $\vec{y}$ entonces $I_{\varepsilon}^{At}(\vec{y})$ es decreciente con ε. Una transferencia de un individuo i a uno más rico j va a hacer aumentar I_{ε}^{At} en mayor proporción cuánto más rico sea j. A medida que ε decrece, $I_{\varepsilon}^{At}(\vec{y})$ se vuelve más sensible a las transferencias de ingresos en el rango de ingresos bajos que en el de ingresos medios y altos.

Por otra parte, $\text{lím}_{\varepsilon\to 0} I_{\varepsilon}^{At}(\vec{y}) = 1 - \frac{\text{mín}\, y_i}{\mu_y}$, es decir, tiende al índice de desigualdad relativo *maximin* (Chakravarty [37]).

Señalemos por último que buena parte de los axiomas que se piden a los índices se deben a que se plantean como funciones de $(y_1,...,y_n)$. Si la *base de información* fueran directamente las funciones de distribución F, entonces principios como el de simetría, traslaciones y poblacional son de cumplimiento inmediato, mientras que la invariancia a escala muchas veces se cumple en el sentido de dar los mismos ránkings.

La filosofía detrás de la presente Tesis es que la base de información al hacer un juicio de bienestar no deben ser los *valores* obtenidos sino las *distribuciones*.

1.2.2. Dominancia Estocástica y Lorenziana

Noción de Dominancia Estocástica y su SWO

El concepto de *dominancia estocástica* fue introducido en el análisis de riesgo a fines de la década de 1960 por Hadar y Russell [95]. Posteriormente Atkinson [14] incorporó dicha noción en su análisis pionero sobre la relación entre desigualdad y bienestar económico[13].

Para introducir la idea de dominancia estocástica consideremos una variable aleatoria X con función de distribución acumulativa F_X, cuyo soporte supondremos compacto y conexo en $\mathbb{R}$, p.ej. el intervalo $[0,1]$. Definimos la familia de operadores integrales:

$$\mathcal{T}_1(z,F_X) = F_X(z) \qquad \mathcal{T}_{j+1}(z,F_X) = \int_0^z \mathcal{T}_j(t,F_X)dt$$

En particular tenemos $\mathcal{T}_2(z,F_X) = \int_0^z F_X(t)dt$ y $\mathcal{T}_3(z,F_X) = \int_0^z \int_0^t F_X(s)dsdt$.

Definición. *Dadas dos variables aleatorias X e Y con funciones de distribución F_X y F_Y de soporte $[0,1]$ diremos que X domina estocásticamente a orden j si $\forall z \in [0,1]$ se cumple $\mathcal{T}_j(z,F_X) \leq \mathcal{T}_j(z,F_Y)$. Notamos esta situación con $X\ SD_j\ Y$.*

[13] Desarrollo histórico y estado del arte en Sec.1.5.1

Es trivial comprobar que $X\ SD_j\ Y$ implica $X\ SD_{j+1}\ Y$ pero el recíproco no vale en general. De hecho, la idea básica en la utilización práctica de este concepto para rankear distribuciones empíricas es que si uno no puede determinar la dominancia entre X e Y a orden j de manera concluyente, se evalúe SD_{j+1}.

A decir verdad, estas son condiciones equivalentes a la noción intuitiva de dominancia estocástica que nos dice que la F domina a G si la utilidad esperada para la primer distribución es mayor o igual que la esperada para la segunda, para una dada clase de funciones. La dominancia a primer orden tal como la acabamos de definir, resulta la dominancia sobre la clase de funciones $u(y)$ con $u'(y) > 0$, la de segundo orden para la clase que además cumple $u''(y) < 0$ y así sucesivamente cada orden se asocia con una condición adicional sobre la derivada $n-$ésima.

En la práctica sin embargo, no podríamos chequear algo sobre una clase infinita por simple inspección, así que se encuentran las condiciones equivalentes mencionadas en la definición, que solo usan la información de las cdf.

Resulta entonces que si F domina estocásticamente a G para algún grado entonces $E_F(u) \geq E_G(u)$ para *todas* las u de una cierta clase dada. Escrito de otra forma:

$$\int_0^1 u(t)dF(t) \geq \int_0^1 u(t)dG(t)$$

Pero esto nos remite a la Ec.1.3, donde definimos los funcionales de bienestar social aditivos. Si llamamos W_u a la función de bienestar social asociada con la función de utilidad individual $u(y)$, tenemos que $W_u(F) \geq W_u(G)$, sobre una clase completa de funciones $u(y)$, por ejemplo las derivables crecientes para la dominancia de primer orden.

Si bien ya habíamos mostrado en la Sec.1.2.1 que existen formas más o menos canónicas de pasar de las funciones de bienestar social a las mediadas de desigualdad y viceversa, **la dominancia estocástica implica *unanimidad* porque el bienestar social es mayor en *toda la clase de funcionales de bienestar social de cierto tipo.***

La dominancia estocástica induce entonces de manera natural un preorden entre las economías de distribución, pues es trivialmente reflexivo y transitivo. No es un *orden* en el sentido de que FSD_jG y GSD_jF al mismo tiempo no implican $F = G$ sino a lo sumo la equivalencia sobre una clase de funciones de utilidad.

Este preorden tiene la buena cualidad de que si FSD_jG sobre una clase de funciones, entonces todos los funcionales de bienestar social asociados con alguna función de dicha clase rankearan de la misma forma a las distribuciones F y G.

Otro aspecto que debe notarse del preorden de dominancia estocástica es que no es *lineal* o *completo*, en el sentido de que dadas F y G no siempre serán *comparables*, es decir, no siempre se dará que FSD_jG o GSD_jF. Esto en principio parece una molestia, pero en realidad nos permite diferenciar los casos en que las distribuciones son *genuinamente equivalentes* de aquellos en que no se puede rankear las distribuciones por un entrecruzamiento compensado de los funcionales $\mathscr{T}_j(z, F_X)$. Eso no se podría hacer con medidas resumen como los índices de desigualdad.

Nótese que, como dijimos más arriba, es usual que si no se tiene resultado concluyente de la dominancia a orden j se analice a orden $j+1$. La razón por la cual es de esperar *mayor decisividad* de los órdenes superiores de dominancia, es que al pasar de uno al otra, vamos *achicando* o *anidando* la clase de funciones sobre las que se domina: para SD_1 las crecientes, para SD_2 las crecientes **y** cóncavas, etc.

De manera que la mayor decisividad de los órdenes superiores de dominancia se logra a expensas de una menor *unanimidad* del resultado.

Esto se puede relacionar con el hecho de que los *índices de desigualdad axiomáticamente derivados* (Chakravarty [37]) son más aceptables cuando tenemos pocos axiomas generales y van siendo más *personales* a medida que uno le va exigiendo más cosas.

Una propiedad que carateriza a la SD_2 respecto a la SD_1 es que resulta *más igualitaria*. Esto quiere decir que si la distribución F está concentrada en único valor μ, entonces para cualquier otra distribución G con esperanza μ se tiene que FSD_jG. Esto es fácil de demostrar usando integración por partes.

Esta propiedad de *igualitarismo* es en principio preferible, pero puede ser discutible en ciertos contextos, como veremos en el Cap.7, Sec.7.1.

Para sacarnos la dependencia de las comparaciones con la esperanza μ lo usual es trabajar siempre con la distribución *normalizada*, es decir, dividir todos los ingresos por la media (Sen [199], Zheng et al. [242]). Esto es equivalente a trabajar con las curvas de Lorenz generalizadas que mencionamos más abajo en la presente Sección.

Orden de Lorenz

Si bien la curva de Lorenz fue introducida a principios del siglo XX, la forma más práctica de definirla es la que introdujo Gastwirth [79] en 1971. Dada una distribución univariada con cdf F, y esperanza finita μ_F la *curva de Lorenz* de F es la función $L_F : [0,1] \to [0,1]$ dada por:

$$L_F(z) = \frac{1}{\mu_F} \int_0^z F^{-1}(t)dt \tag{1.8}$$

donde $F^{-1}(t) = \text{ínf}\{x : F(x \geq t)\}$ sabemos que está bien definida para cualquier función de distribución (Yohai [241]).

Decimos que la distribución F *domina Lorenz* a otra distribución G de la misma media si para cada z se tiene la desigualdad $L_F(z) \geq L_G(z)$. Para independizarnos de la media dividimos cada elemento de cada distribución por la correspondiente, o lo que es lo mismo usamos la *curva de Lorenz generalizada* dada por $GL_F(z) = \mu_F L_F(z)$. Cuando digamos que una distribución domina Lorenz a otra nos referiremos siempre a la comparación normalizada, al igual que en el caso de la dominancia estocástica.

Un resultado que se conoce como Teorema de Shorrocks [204] (aunque los especialistas lo adjudican a Kolm unos años antes, cfr. Lambert [126]) es el que da la equivalencia entre la dominancia de Lorenz (que notamos $F\ LD\ G$) y la dominancia estocástica a orden 2. Esto implica en particular, que la dominancia de Lorenz también es *igualitaria*, en el sentido de que la distribuciones concentradas en un valor dominaran Lorenz a cualquier otra, o lo que es lo mismo tendrán su curva de Lorenz estará siempre más cerca de la diagonal.

Este resultado es en realidad un caso particular de otro más general de Muliere y Scarsini [154] para lo que llaman *dominancia estocástica inversa de orden n*. Dada una variable aleatoria X con distribución asociada F_X definamos los operadores:

$$\mathscr{T}_1^{-1}(z,F_X) = F_X^{-1}(z) \qquad \mathscr{T}_{j+1}^{-1}(z,F_X) = \int_0^z \mathscr{T}_j^{-1}(t,F_X)dt$$

diremos que *F domina inversa estocásticamente a G a orden j* si $\mathscr{T}_j^{-1}(z,F) \geq \mathscr{T}_j^{-1}(z,G)$ y lo notamos $FSD_j^{-1}G$. El interesante resultado de Muliere y Scarsini [154] es que $FSD_jG \Leftrightarrow FSD_j^{-1}G$. El Teorema de Shorrocks se sigue de inmediato para $j = 2$.

Como mencionamos en varias oportunidades, los índices de desigualdad (o sea funciones que asignan un número real a cada distribución) determinan un preorden sobre las economías de distribución. Los índices regulares (p.16) I tales que $F\ LD\ G$ implica $I(F) \leq I(F)$ se denominan *Lorenz consistentes* y al conjunto de todas las medidas Lorenz compatibles se lo denomina $\mathscr{L}$.

Este conjunto de índices resulta además igual al conjunto de índices regulares que cumple con el axioma de Dalton-Pigou (Shorrocks [205]). En este sentido debe destacarse que si $I(F) \leq I(G)\ \forall I \in \mathscr{L}$ entonces $F\ LD\ G$, o sea la clase *completa* o unánime de índices es determinante del orden de Lorenz.

Pero del hecho de que sepamos que para *algún* o algunos índices de la clase $I(F) \leq I(G)$ no podemos inferir que $F\ LD\ G$. En la práctica, por supuesto no puede chequearse un número infinito de índices, así que se busca un número pequeño que caracterice con buena probabilidad el orden de Lorenz, aunque estos no resultan ser los más habituales en la investigación empírica (Shorrocks [205]).

Uno de los problemas del paso al caso multidimensional es que no existe una buena clase de índices aproximantes de los órdenes usuales, como el de dominancia estocástica multidimensional.

El más famoso de los índices Lorenz compatibles es el coeficiente de Gini:

$$I_G(F) = 1 - 2\int_0^1 L(z)dz \tag{1.9}$$

Este índice es consistente con la dominancia Lorenz, es decir la dominancia estocástica de orden 2, y también lo es con la de orden 3 (Muliere y Scarisini

[154]). Esto último demuestra que el Gini por sí solo no puede ser determinante del orden de Lorenz, basta tomar un par de distribuciones comparables a orden 3 que no lo sean a orden 2.

Señalemos por último que para poder comparar en el orden de Lorenz dos distribuciones, es necesario que las curvas de Lorenz no se crucen. En la práctica empírica, parece haber una diferencia de opiniones sobre si este fenómeno es común o no. Lambert [126] señala cierta literatura que afirma que es un fenómeno usual, mientras que Sen [199] cita otra con la conclusión contraria.

En teoría, no hay nada que garantice la decisividad del orden de Lorenz sobre un par dado de distribuciones. En la práctica, sin embargo, nuestra experiencia (Perez [166]) es que al comparar distribuciones de ingreso de un mismo país, tanto el orden SD_1 como el de Lorenz resultan bastante completos. Al comparar entre distintos países u otras estructuras, puede ser que no siga valiendo esta observación.

Principios de Transferencias

La dominancia de Lorenz cumple con el llamado *principio de transferencias* o de Dalton-Pigou (cfr. Sec.1.2.1), en el sentido de que si al pasar de F a G sólo hubo una transferencia de ingresos de un individuo más rico a uno más pobre entonces $G\ LD\ F$. En consecuencia tanto la SD_2 como los índices de la clase $\mathscr{L}$ cumplen el principio.

Sin embargo, en el orden de Lorenz no impactan necesariamente más las transferencias en el sector de menores ingresos que entre los más altos. Esto sería deseable, porque se entiende que el efecto de la desigualdad entre los ricos no debe ser igual de importante que entre ingresos medios o bajos.

Si al pasar de F a G hubo transferencias progresivas y regresivas en lugares distintos, el ranking de F y G por Lorenz será inconclusivo.

El principio de las transferencias disminuidas (*diminishing transfers principle*) fue introducido por Kolm en los 70 y dice básicamente que las transferencias en los rango más bajos de la distribución deben tener mayor impacto en el bienestar.

Introduciendo técnicamente la noción de *sensibilidad a las transferencias*, Shorrocks y Foster [206] prueban que la dominancia estocástica a tercer orden permite rankear sin ambigüedad distribuciones que el orden de Lorenz no podía.

Una forma práctica de darle mayor peso a los rangos bajos de ingresos es literalmente añadir un factor de peso en la construcción de las curvas:

$$\tilde{L}(z) = \frac{1}{\mu}\int_0^z w(z)F^{-1}(z)dz \tag{1.10}$$

y trabajar a partir de ellas. Esto tiene relación con la noción de funciones duales de bienestar social o funcionales de Yaari, que son un enfoque alternativo al bienestarismo (Zoli [244]).

Interpretación Económica de la Dominancia Estocástica

El resultado de Hadar y Russell [95] y que retoma Atkinson [13] lo podemos resumir en los siguientes dos lemas (Deaton [46], Perez [166]).

Lema 1. *Supongamos que X SD_1 Y y supongamos que la función de bienestar social viene dada por $W(H) = \int U(z)dH(z)$ donde H es la función de distribución asociada a la distribución del ingreso y U es cualquier función creciente de z. Entonces el bienestar en la situación X no es menor que en la Y, es decir $W(F_X) \geq W(F_Y)$.*

Lema 2. *Supongamos que X SD_2 Y y supongamos que la función de bienestar social viene dada por $W(H) = \int U(z)dH(z)$ donde H es la función de distribución asociada a la distribución del ingreso y U es cualquier función creciente y cóncava de z. Entonces el bienestar en la situación X no es menor que en la Y, es decir $W(F_X) \geq W(F_Y)$.*

Estos dos resultados son una expresión más concreta del Teorema de Atkinson que mencionabamos en la Sec.1.1.1.

El trabajo de Hadar y Russell [96] generaliza la noción de dominancia estocástica a varias variables y el trabajo de Atkinson y Bourguignon [14] lo aplica a la economía del bienestar. Al pasar a varias variables, las condiciones de monotonía y concavidad de las funciones del integrando deben complementarse con condiciones sobre las derivadas cruzadas.

El resultado final es análogo al caso univariado: las distribuciones que presentan mayor desigualdad (en el sentido de ser dominadas estocásticamente) serán las de menor bienestar agregado.

En el Cap.3 presentamos demostraciones sencillas de estos resultados y las generalizaremos al caso discontinuo. Con ellos tendremos condiciones suficientes para la dominancia multivariada y su interpretación en términos de la economía del bienestar.

Por último mencionemos que la noción de dominancia estocástica es relevante en el estudio de la pobreza además de la desigualdad. Lambert [126] muestra como para poder lograr rankings de bienestar basados en el *head count* de pobreza o en la *brecha* de pobreza que no se inviertan por un simple salto en la línea de pobreza, es necesario pedir la condición de dominancia estocástica a orden 2 de las distribuciones.

1.3. El Enfoque Estadístico de La Desigualdad Económica

1.3.1. Necesidad del Enfoque, Limitaciones

Necesidad del Enfoque Estadístico

La mayoría de los índices de desigualdad y curvas de concentración desarrollados en la primera mitad del siglo XX (e.g. coeficiente de Gini, entropía de Theil, etc) se presentaron como estadísticas descriptivas para resumir en información numérica ciertas propiedades de las economías en estudio, y algunos trabajos actuales aun los utilizan de esta forma.

Sin embargo, a partir de la relación que señalamos en la Sec.1.1.1 entre la desigualdad económica y el bienestar agregado, se plantea la necesidad de poner base firme a las inferencias estadísticas sobre las medidas de desigualdad.

Si uno trabajara en *teoría*, las distribuciones de ingreso u otros atributos se conocerían con certeza y las medidas de desigualdad se pueden computar con arbitraria precisión, incluso de manera analítica.

Pero siendo en la práctica las *encuestas de hogares* la fuente básica de datos para la economía cuantitativa del bienestar (Deaton [46]), se plantea la necesidad ineludible de tomar en cuenta los efectos de los errores muestrales a la hora de obtener estimadores para las curvas o índices de concentración estudiados. Sólo así se obtendrán inferencias estadísticamente válidas a partir de la comparación de los datos provistos para distintas economías o períodos de tiempo.

La confiabilidad de las conclusiones en las comparaciones de bienestar agregado, y en especial las que se basan en el estudio de la desigualdad distributiva, es de vital importancia pues en ellas se basarán gran parte de las decisiones de política económica, principalmente las relacionadas al diseño del sistema impositivo y de transferencias de ingresos (redistribución del ingreso).

El enfoque estadístico de la desigualdad económica se ha centrado entonces en la obtención de las distribuciones muestrales y asintónticas de los índices y curvas de concentración involucrados en el análisis de la desigualdad. El trabajo de Brufman & Perez [33] pasa revista a los desarrollos en este sentido que han tenido lugar en las últimas décadas, en particular en el enfoque no paramétrico.

Limitaciones del Enfoque: el Caso Multivariado

La dificultad más inmediata que enfrenta cualquier enfoque multivariado de la desigualdad (no sólo el estadístico) radica en la inexistencia de una versión multivariada obvia de algunos inputs básicos del estudio unidimensional. Como detallamos en la Sec. 3.1, la curva de Lorenz, y los cuantiles en general, no tienen una versión multivariada unívoca o siquiera generalmente aceptada por los investigadores.

En cuanto a las herramientas estadísticas que puedan generalizarse (e.g. medidas de entropía, funciones de distribución) la dificultad aparece en los aspectos técnicos y computacionales. Por un lado, es un problema teóricamente más dificultoso derivar las matrices de covarianzas para estadísticos multivariados.

Por otro lado, desde el punto de vista puramente práctico, la complejidad implicada en los esquemas de muestreo y en el cómputo de los valores de los estadísticos de prueba para una muestra dada, crecen exponencialmente con la dimensión (*curse of dimensionality*).

En este sentido hemos elegido para nuestro trabajo de Tesis la construcción de test bidimensionales, tanto por el hecho de que las variables ingreso y capital humano resumen los principales aspectos de la desigualdad económica a nivel dinámico y de bienestar, como por la necesidad de restringir la dimensionalidad por los aspectos computacionales.[14]

1.3.2. Dominancia Estocástica en el Enfoque Estadístico

Como vimos en la Sec.1.2.2 en el caso de la dominancia estocástica tenemos condiciones necesarias y suficientes que se expresan como operadores integrales aplicados sobre las distribuciones.

Esto implicará que al pasar a la etapa de testeo estadístico de las mismas, tanto la hipótesis nula como la alternativa se expresan en términos de ciertos funcionales regulares de las distribuciones subyacentes, y sus correlatos muestrales serán entonces incorporados en la construcción de estadísticos de prueba apropiados.

En este caso, como señalan Hájek et al. [97] cuando la hipótesis nula corresponde a la hipótesis de invariancia, se pueden encontrar estadísticos de prueba con distribuciones exactas y totalmente no paramétricas[15]. Este es el caso con los estadísticos de Kolmogorov-Smirnov, que para el caso univariado continuo resultan *pivotales* (cfr. Sec.1.5.3).

El trabajo de McFadden [149] muestra cómo los estadísticos de Kolmogorov-Smirnov surgen de manera natural al tratar el problema de testear estadísticamente la dominancia estocástica. Dada la hipótesis $H_0 : F_X(z) \geq F_Y(z)$ $\forall z \in [0,1]$ contra $H_1 : F_X(z) < F_Y(z)$ para algún $z \in [0,1]$, supongamos dos muestras $(x_1,...,x_n)$ y $(y_1,...,y_n)$.

Vamos a considerar el problema general, donde F_X y F_Y no están en familias paramétricas. Para llevar a cabo el test definimos la función de distribución empírica $F_X^n(x) = \frac{1}{n}\sum_{i=1}^n 1_{(x_i \leq x)}$ o sea la fracción de observaciones que son $\leq x$. Análogamente se define la $F_Y^n(x)$.

A partir de H_0 un estadístico de prueba natural resulta el análogo empírico: $D_n^* = \text{máx}_{w\in[0,1]} D_n(w)$ donde $D_n(w) = \sqrt{n}(F_Y^n(w) - F_X^n(w))$. Como muestra McFadden [149], este estadístico es justo del tipo de Kolmogorov-Smirnov y tiene una

[14] Sobre todo al implementar rutinas de re-muestreo como el bootstrap y el jackknife.

[15] *Distribution free* en el sentido de que la distribución del edp no depende de la distribución común subyacente.

distribución muestral exacta bajo la hipótesis nula y una distribución asintótica que son independientes de la distribución común subyacente. Ambas distribuciones son conocidas y se encuentran tabuladas por lo que su utilización es inmediata.

Para el caso de dominancia a segundo orden, la distribución subyacente no se encuentra tan inmediatamente porque hay un operador integral involucrado. A partir del estadístico de prueba provisto por McFadden: $S_n^* = \max_{w \in [0,1]} S_n(w)$ donde $S_n(w) = \sqrt{n} \int_0^w (F_Y^n(z) - F_X^n(z))dz$ Barret y Donald [15] desarrollan un método de bootstrap para obtener valores críticos dependientes de los datos para el mismo.

Un objetivo particularmente importante del presente trabajo de Tesis es entonces la generalización del método de Barret y Donald [15] para el caso multivariado. Si bien dicho trabajo presenta el método para testear dominancia de orden arbitrario, **en este trabajo de Tesis trataremos los casos $j = 1, 2$ debido a las complejidades computacionales que conllevan las integrales numéricas sucesivas en varias dimensiones.**

Como señalamos a lo largo de la presente introducción, nuestro objetivo implica el desarrollo de tests de dominania estocástica en dos dimensiones, y del tipo de Kolmogorov-Smirnov.

En este punto se presenta entonces la problemática técnica desde el punto de vista estadístico de nuestro trabajo: el test de Kolmogorov-Smirnov, que en el caso univariado se presenta totalmente no paramétrico (*distribution free*), pierde su condición pivotal al pasar a cualquier dimensión mayor. Simpson [210] muestra esta dependencia de la distribución subyacente con un ejemplo elemental.

Planteada esta problemática, en el Cap.5 detallamos como la introducción de ciertos elementos de la teoría de probabilidades y la estadística matemática, la llamada *teoría de procesos empíricos*, nos permite de alguna forma sortear los obstáculos y obtener distribuciones asintóticas para nuestros edp multivariados.

1.4. El Análisis Microeconométrico de las Encuestas de Hogares

1.4.1. Encuestas y Microdatos

El estudio de la desigualdad económica tal como lo hemos descrito en las Sec.1.1 y 1.3 se lleva a cabo en principio para datos de los agentes económicos como *individuos*. En la práctica, sin embargo, los datos tienen más sentido cuando la unidad de relevamiento es el *hogar*.

Como señala Amartya Sen:

> *"Las rentas que gana uno o más miembros de una familia son compartidas por todos, tanto por los que no perciben ningún ingreso como por los que perciben alguno. Por lo tanto, la familia es la unidad básica para examinar las rentas desde el punto de vista de su uso."*
> (Sen [196], Cap.3)

Para volver a la distribución *individual* la práctica profesional usual es la aplicación de un *factor adulto equivalente* (Duclos, Esteban y Ray [57], Gasparini [78]) sobre todo en el análisis de la pobreza, o bien la *regla de la raíz cuadrada* más común en el análisis de la desigualdad (OECD [162]; en la Sec.6.2 damos detalles).

La fuente de datos natural para el estudio de la desigualdad económica es entonces la *encuesta de hogares*. Este tipo de encuestas son relevadas con una estructura y periodicidad similar en todo el mundo (UN [225], Deaton [46]) lo cual permite una comparación entre distintas poblaciones o a lo largo del tiempo.

Por otra parte, dichas encuestas recogen información relevante para el diseño y evaluación de las políticas públicas, no sólo en cuanto a ingresos y gastos, sino al acceso a bienes básicos como educación y salud, características ocupacionales, demográficas, etc. Nótese que este tipo de datos también se releva en los *censos poblacionales* pero con una periodicidad mucho menor debido a su costo, uno cada 10 años frente a al menos un par de ondas al año para las encuestas de hogares.

En base a ello entonces, como señalamos en la Sec.1.3.1, debe tomarse siempre en cuenta que trabajamos con datos *muestrales* y no *poblacionales* de manera que las inferencias derivadas de los resultados deben tomar debida cuenta de la variabilidad muestral. De esta manera, los juicios de bienestar en base a encuestas de hogares deben entonces apoyarse de manera general en lo que llamamos *enfoque estadístico* de la desigualdad.

Una observación que debe hacerse en cuanto a dichos juicios de bienestar, es que se verán afectados por la *calidad de los datos*. Como señalan Cowell y Victoria-Feser [42], los juicios sobre bienestar a partir de la dominancia estocástica, basados en la interpretación de Atkinson, deben ser tomados con cautela al tratar con datos muestrales. La existencia de datos contaminados (e.g rangos faltantes de datos, etc.) pueden afectar la robustez de la medida de dominancia estocástica en relación al bienestar.

En la medida en que la contaminación de datos provenga de la *no respuesta*, se han diseñado soluciones parciales, que van desde el diseño de encuestas que reduzcan la falta de respuesta hasta el ajuste de los datos para tomar en cuenta el problema (cfr. Bethlehem, Cobben y Schouten [24]). Cuando la contaminación de datos proviene de otras fuentes, el problema puede ser mucho más difícil de solucionar.[16]

Otra fuente de problemas, particularmente cuando se estudia el ingreso o el consumo, es el efecto de los períodos inflacionarios. Gasparini [78] cita el efecto para la EPH del INDEC, y Paxson [165] para el estudio de los ahorros en Tailandia. En nuestro caso (Sec.6.2), en el período 1996-2006 el principal efecto puede haberse dado en el año 2002, pero más allá del ajuste por IPC, no realizamos correcciones de los datos.

En base a los microdatos de las encuestas de hogares, el investigador podrá adoptar un amplio abanico de técnicas disponibles de acuerdo a los objetivos que se plantee. Dichos métodos aputan por un lado al *diagnóstico* de las situaciones sociales en base a la teoría adoptada y por otro lado al modelado de comportamientos y respuestas a cambios en las variables, base para el diseño y análisis de la política económica.

Describimos brevemente algunos de estos métodos en las siguientes dos subsecciones.

1.4.2. Métodos Paramétricos

Métodos Generales

Los métodos de regresión clásicos, como el de cuadrados mínimos, presentan algunos inconvenientes a la hora de analizar microdatos. El problema básico de la no identificación hace que no se puedan estimar de manera directa las ecuaciones estructurales y la correlación entre el término de error y las variables explicativas implica la falta de consistencia del estimador OLS. La subidentificación surge con frecuencia al trabajar con microdatos, ya sea que su fuente es la *simultaneidad* (e.g. al estimar curvas de demanda) o la *retroalimentación* (e.g. al analizar ingresos y gastos, Deaton [46], Ch.2).

La respuesta usual a este problema (Dufour [64]) es la de encontrar *variables instrumentales* (IV) que estén correlacionadas con las variables explicativas pero incorrelacionadas con el término de error, de manera que la estimación de los parámetros estructurales resulte consistente.

Desde el punto de vista econométrico debe tenerse en cuenta, sin embargo, que la dispersión de los estimadores IV es mayor que la de los estimadores OLS, de

[16] Esta es la razón principal por la cual en la Sec.6.2 no utilizamos datos para la EPH de INDEC después del año 2006.

manera que lo que se gana en consistencia puede perderse en precisión al trabajar con datos muestrales. Más aun, la teoría detrás de los estimadores IV es asintótica, así que puede tener problemas de precisión en muestras finitas.

Al trabajar con gran cantidad de datos individuales, es común que los mismos se agrupen naturalmente en *sectores* en el sentido de que las relaciones entre variables tomen diferentes formas para distintos subconjuntos de la población. Por ejemplo, es probable que la elasticidad ingreso de la demanda (a partir de curvas de Engel) sea distinta para los hogares urbanos que para los rurales. En estos casos, el método de *regresiones con pesos* [46] permite evitar la inconsistencia del estimador OLS.

Los *métodos de datos de panel* (Arellano [8]) aprovechan las estructuras de datos con observaciones repetidas para un conjunto de individuos (pueden ser personas, sectores, países, etc) de manera que cada uno puede funcionar en cierta forma como su propio individuo de control (en el sentido experimental). Un ejemplo de la utilidad de estos métodos la de incorporar en los modelos variables que no se miden directamente. Por caso, para estudiar el problema de *retornos a la educación* (la relación entre el ingreso individual y los años de escuela) uno puede plantear un *efecto fijo* (para cada individuo) de la *inteligencia o habilidad natural*. Los modelos de efectos fijos estimados en diferencias eliminan la necesidad de medir en la práctica dicha variable, por ejemplo mediante el IQ.

En el análisis de datos de encuestas de hogares, sin embargo, los métodos de panel deben tomarse con cuidado por su particular sensibilidad al *ruido*, es decir a los errores de medición, que introducen un sesgo a la estimación (Deaton [46]).

Métodos Paramétricos en la Distribución del Ingreso

Como señala Slottje [211], desde la década de 1930 la llamada ecuela *estocástica* de la desigualdad planteaba modelos que, como el de cadenas de Markov de Chapernowne, llevaban a una forma determinada de la función de distribución y/o la curva de Lorenz, por ejemplo la de Pareto.

De manera general, como señalamos en la Sec.1.2.2, la curva de Lorenz se relaciona con la función de distribución vía:

$$L(z) = \frac{1}{\mu} \int_0^z F^{-1}(t)dt \tag{1.11}$$

con lo cual si tenemos una forma funcional determinada para una de ellas (F o L) tendremos una para la otra.

El enfoque *paramétrico indirecto* de la curva de Lorenz consiste entonces en encontrar la forma funcional de F que satisfaga algún criterio conveniente, típicamente por ajuste de los datos dentro de una familia paramétrica (Ryu y Slottje [185]). Una vez estimada la cdf se obtiene inmediatamente la curva de Lorenz, según fórmulas fácilmente deducibles, siendo algunos ejemplos conocidos (Barry Arnold [9]):

Exponencial: $F(x) = 1 - e^{-\lambda x} \Rightarrow L(z) = z + (1-z)ln(1-z)$

Pareto: $F(x) = 1 - \left(\frac{x}{ax}\right)^{-\alpha} \Rightarrow L(z) = 1 - (1-z)^{\frac{\alpha-1}{\alpha}}$

Lognormal: $F(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-1/2\sigma^2(\log x - \mu)^2} \Rightarrow L(z) = \frac{\log z - \mu}{\sigma} - \sigma$

Este método indirecto requiere entonces encontrar aquella F que mejor satisfaga los criterios de evaluación elegidos para los datos dados (e.g. desviación media cuadrática, variación total, etc, Devroye & Lugosi [51]), lo cual es un problema en sí mismo. Sen [199] señala que a su juicio la forma dada por la *lognormal* se ajusta bien a ciertos países, aunque en el rango de ingresos altos es mejor el ajuste de una paretiana.

Por otra parte, el enfoque paramétrico *directo* consiste en proponer una forma funcional para la curva de Lorenz y ajustar sus parámetros, sin pasar por la cdf. Nótese que no cualquier función $L(z)$ tiene sentido como curva de Lorenz, pues algunas propiedades obvias que debe cumplir son:

- $L(0) = 1 - L(1) = 0$
- $L(z) \leq z$

Kakwani y Podder [114], proponen la forma simple $L(z) = ze^{-h(1-z)}$. Basmann et al. [17] proponen una familia de curvas con la forma funcional:

$$L(z) = z^{az+b} e^{-g(1-z)^2 - h(1-z) + \mu} \tag{1.12}$$

o más prolijamente en logaritmos:

$$\log L(z) = (az+b)\log z - g(1-z)^2 - h(1-z) + \mu \tag{1.13}$$

ecuación esta última que proceden a estimar con OLS. Una observación que debe hacerse es que estos ajustes proceden primero a *agrupar* los datos en deciles o percentiles, de manera que la interpretación que le dimos en términos de bienestar a la curva de Lorenz (que se basa en los ingresos individuales) se pierde.

Es cierto que la definición de $L(z)$ con una formulación funcional simple puede tener algunas ventajas teóricas, por ejemplo, poner de manifiesto ciertas relaciones entre datos en un modelo dado de distribución del ingreso que de otra forma no serían visibles.

Lamentablemente, sin embargo, la realidad no se lleva bien con los ajustes paramétricos de las curvas de Lorenz.

Ya en 1972 Gastwirth [80] señalaba para el modelo paretiano:

> *"It appears that the Pareto fit was not bad as late as 1955 but is no longer appropiate."*

Dicho autor procede entonces a dar un procedimiento no paramétrico para la estimación no ya de la curva de Lorenz completa sino para cotas superiores e inferiores del coeficiente de Gini.

Schader y Schmid [194] en 1994 pusieron a prueba una serie formas funcionales para $L(z)$, entre ellas las mencionadas de Kakwani y Podder [114] y las de Basmann et al. [17], encontrando en **todos** los casos que su performance era bastante mala, produciendo estimaciones siempre sobre las cotas del método no paramétrico de Gastwirth.

Debe notarse además, que las formulaciones paramétricas desde Kakwani y Podder [114] hasta Ryu y Slottje [185], imponen condiciones de suavidad a la $L(z)$ (derivadas de suponer una densidad para la distribución) que son un poco exigentes ($L(z) \in \mathscr{C}^2$) y que pueden ser la raíz de su mala performance empírica.

De manera general, entonces, los enfoques paramétricos de la curva de Lorenz no son muy confiables.

1.4.3. Métodos no Paramétricos

Las Cotas de Gastwirth para el Coeficiente de Gini

El trabajo de Gastwirth [80] parte de una definición práctica de la curva de Lorenz en base a:

- Límites de Clase: $0 = a_0 < a_1 < \cdots < a_k = \infty$ (la última clase se considera abierta).
- Frecuencias de Clase: $n_1, n_2, ..., n_k$ con $n = \sum_{i=1}^{k} n_i$ el número total de individuos.
- Medias de Clase: μ_i para $i = 1, ..., k$ con $\mu = \sum_{i=1}^{k} \frac{n_i}{n} \mu_i$ la media de la distribución completa.

Llamemos (x_i, y_i) al par que tiene la porción acumulada de la población en la primera coordenada y la proporción acumulada de ingreso en la segunda, es decir:

$$x_0 = y_0 = 0$$

$$x_k = y_k = 1$$

$$x_i = \sum_{j=1}^{i} \frac{n_j}{n} \qquad i = 1, ..., k-1$$

$$y_i = \sum_{j=1}^{i} \frac{n_j \mu_j}{n\mu} \qquad i = 1, ..., k-1$$

Las cotas de Gastwirth para el coeficiente de Gini vienen dadas por:

$$G_L = 1 - \sum_{i=1}^{k} (x_i - x_{i-1})(y_i + y_{i-1}) \qquad G_U = G_L + \Delta \tag{1.14}$$

donde Δ, conocido como *efecto de agrupamiento*, viene dado por:

$$\Delta = \frac{1}{\mu} \left(\sum_{i=1}^{k-1} \left(\frac{n_i}{n}\right)^2 \frac{(\mu_i - a_{i-1})(a_i - \mu_i)}{a_i - a_{i-1}} + \left(\frac{n_k}{n}\right)^2 (\mu_k - a_{k-1}) \right)$$

Se ve con facilidad que G_L es dos veces el área entre la curva de Lorenz basada en los datos agrupados y la diagonal. Gastwirth mostró que el verdadero valor del coeficiente de Gini de la distribución subyacente debe caer entre G_L y G_U para cualquier distribución.

Debe notarse que éste es un método que trabaja con *datos agrupados*, de manera que los juicios de bienestar que se basan en la agregación de preferencias *individuales* no aplican de manera directa. Lo mismo corre para los métodos paramétricos de estimación que citamos recién.

En la presente Tesis, por el contrario, queremos desarrollar herramientas que comparen la totalidad de las distribuciones, brindando entonces mayor consistencia a los estimadores y una interpretación directa en términos de bienestar social. En este sentido puede decirse que buscamos tener las propiedades del test de Barret y Donald [15], llevadas al caso multidimensional.

De manera general, los trabajos sobre la distribución muestral asintótica de estimadores de índices de desigualdad como el de Gini y la entropía de Theil fueron desarrollados en los trabajos de Nygård y Sandström en la década de 1980 (cfr. [158, 159, 160]). La teoría subyacente es el resultado clásico de Hoeffding sobre estadísticos U para funciones simétricas (Dudley [63]).

El paper de Brufman y Perez [33] resume estos y otros resultados sobre el enfoque no paramétrico para obtener distribuciones asintóticas de los estimadores muestrales de los índices de desigualdad clásicos.

Método *KDS*

Un enfoque no paramétrico clásico, pero que permite conservar supuestos de suavidad sobre $L(z)$ puede consistir en realizar una *estimación de densidad por el método de kernel* o KDS por sus siglas en inglés.[17]

Dada una muestra de ingresos $X_1, \ldots, X_n$ una primera idea para aproximar la función de densidad $f(x)$ puede ser a través de la medida empírica (o histograma) $\mu_n(A) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i\}}$. Sin embargo, no se puede aproximar f en L^1 por μ_n ya que la distancia en variación total entre cualquier densidad y una medida atómica

[17] Sobre KDS y Wavelets seguimos lo expuesto en Perez [166], donde hay más detalles.

es siempre 1 (Devroye y Lugosi [51] Ch.9). Lo que necesitamos es una función densidad f_n para la medida empírica y el método de kernel nos provee la misma.

La idea detrás del método de kernel es usar las aproximaciones de la identidad en L^p que son básicamente familias de funciones (*kernels* o núcleos) que convergen en cierto sentido a la unidad de convolución (o sea a la δ de Dirac). Dada una función $K(x)$ definida en $\mathbb{R}^d$ y un número ε (*ancho de banda*) llamemos:

$$K_\varepsilon(x) = \varepsilon^{-d} K\left(\frac{x}{\varepsilon}\right)$$

El siguiente teorema, demostrado en Wheeden y Zygmund [234] Ch.9, nos presenta el aspecto fundamental de las aproximaciones de la identidad, y su consiguiente utilidad en el método de kernel.

Teorema (Aproximación de la Identidad)**.** *Sea $K(x) \in L^1(\mathbb{R}^d)$ tal que $\int_{\mathbb{R}^d K(x)} dx =$ 1 (o sea K es un* kernel*), y llamemos f_ε a la convolución $f * K$. Si $f \in L^p(\mathbb{R}^d)$ con $1 \leq p < \infty$ entonces $\| f_\varepsilon - f \|_p \xrightarrow[\varepsilon \to 0]{} 0$.*

$\square$.

Ahora a partir de dicha propiedad fundamental de los kernels $\{K_\varepsilon\}$, podemos definir una estimación de f dada la muestra $X_1, ..., X_n$ convolucionándolos con la medida empírica (*estimador de Akaike-Rosenblatt-Parzen*):

$$f_{n,\varepsilon}(x) = \mu_n * K_\varepsilon = \int \mu_n(t) K_\varepsilon(x-t) dt =$$
$$= \frac{1}{n} \sum_{i=1}^{n} K_\varepsilon(x - x_i) = \frac{1}{n\varepsilon^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{\varepsilon}\right) \quad (1.15)$$

Y ahora simplemente por el teorema y las propiedades de los kernels se obtiene la consistencia en L^1 del estimador de la densidad (Devroye [51], Ch.9):

Teorema (Consistencia de la KDS)**.** *Sea K un kernel fijo y se $\varepsilon = \varepsilon(n) > 0$. Si $\varepsilon \to 0$ y $n\varepsilon^d \to \infty$ con $n \to \infty$ entonces $E(\int |f_{n,\varepsilon} - f)| \to 0$*

$\square$.

Si bien el resultado de consistencia es válido de manera general (o sea cualquier densidad y cualquier kernel base) en la práctica la estimación de la densidad requiere de una elección inteligente de K y del ε que se considera aceptable. Un criterio como la minimización del error cuadrático medio es en principio razonable, pero su aplicación práctica requeriría el conocimiento de la densidad que se quiere estimar.

Pese a ello, bajo algunos supuestos se puede calcular un ancho de banda óptimo en ese sentido, por ejemplo usando un kernel gaussiano y suponiendo que f es asimismo gaussiana con varianza σ el ancho de banda óptimo resulta $1{,}06\sigma N^{-1/5}$ (Deaton [46] Ch. 3). Una solución práctica suele ser también la comparación de varias estimaciones con distinto ancho y en todo caso la selección de diversos anchos de banda para el rango de la densidad en base a métodos iterativos.

Ahora, una vez que se tiene la estimación de la densidad basta observar que la proporción de individuos con un ingreso menor que x es $F_1(x) = \frac{1}{\mu}\int_0^x f(t)dt$ y con ello (mediante la relación entre F y F_1) tenemos una estimación no paramétrica de la curva de Lorenz.

Wavelets

Los *wavelets* son familias numerables $\{\phi_k, \psi_{jk}\}$ de funciones ortonormales en $L^2(\mathbb{R})$ con ciertas propiedades que permiten expresar cualquier función $g \in L^2(\mathbb{R})$ como:

$$g(x) = \sum_{k=-\infty}^{\infty} \alpha_k \phi_k(x) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \beta_k \psi_{jk}(x) \tag{1.16}$$

Cuando lo usamos para estimar densidades f, notamos que $\int f\phi_k = E(\phi_k)$ y por lo tanto un estimador razonables es el de *plug in* $\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} \phi_k(X_i)$, y lo análogo para los $\hat{\beta}_{jk}$.

El método de wavelets con *thresholding* (Perez [167]) es una suavización del estimador de wavelets simple mediante la cancelación de los términos cuyos coeficientes estimados son pequeños. Este método incorpora entonces información adicional que permite mejorar la estimación desde el punto de vista práctico.

Esto es así porque la parte lineal del estimador describe las oscilaciones de baja frecuencia (o promedio) mientras que la parte no lineal (debida al thresholding) le permite adaptarse a las fluctuaciones locales, como discontinuidades y oscilaciones de alta frecuencia. Como resultado de estas propiedades, la estimación de densidades por wavelets con thresholding resulta de mejor performance que otros métodos cuando existen discontinuidades u oscilaciones locales.

En su análisis del capital humano Perez [166] construye la variable H a partir de un polinomio cuadrático en la experiencia y de una función no lineal en los aõs de educación. Si bien la parte polinómica en la experiencia (vía el proxy de la edad) le da cierta suavidad a la distribución de H , la contribución de los años de educación le aporta una marcada presencia de picos o modas locales.

Esto se debe en principio a la forma en que traduce el nivel educativo en años de educación mediante una conversión discreta y podría en principio considerarse simplemente una suavización de la distribución en base a un método de kernel con mayor ancho de banda o algo similar.

Sin embargo, en el estudio del capital humano existen razones para preferir quedarse con la versión más discontinua de la distribución. Esto se debe a que dicho trabajo no incluye de manera explícita la condición laboral en la construcción de H, y en este sentido debe tomarse en cuenta el *efecto de señal* que tiene el diploma de cada nivel educativo.

A modo de ejemplo, la condición *educación secundaria completa* suele ser un filtro del mercado laboral, y asimismo un *título universitario* suele deparar, aun para el mismo cargo, una remuneración distinta (e.g. en el sector público).

Por ello puede considerarse apropiado que estos efectos de señal se mantengan, y no imponer un oversmoothing de las densidades de capital humano, optando entonces por la metodología de wavelets cuyo mejor desempeño es de esperar en este caso por las razones previamente expuestas.

Bootstrap Empírico

A partir de una *muestra* de tamaño n a partir de una distribución F, dada por $(x_1, x_2, ..., x_n)$, entonces podemos definir la *distribución empírica* $\hat{F}$ como aquella que asigna a cada x_i una probabilidad de $1/n$, es decir aquella que define la probabilidad $\hat{P} = (1/n)\sum_i^n \delta_{x_i}$.

Un *parámetro* es una función de la *distribución*, $\theta = t(F)$, mientras que un *estadístico* es una función de la *muestra*, o lo que es lo mismo, de la distribución empírica $\hat{\theta} = s(F)$. Los parámetros son en general números (e.g. $E(X)$) mientras que los estadísticos son *variables aleatorias (e.g. $\bar{x}$)*.

Definición. *Dado un parámetro $\theta = t(F)$, el* estimador de Plug-In *del mismo está dado por $\hat{\theta} = t(\hat{F})$.*

Es decir, que el estimador de plug-in de un parámetro es el estadístico que se obtiene reemplazando a la distribución F con la empírica $\hat{F}$ como argumento en la misma función que lo define.

Por ejemplo, si $\theta = E(X)$, entonces será $\hat{\theta} = E_{\hat{F}}(x) = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$, la media muestral.

El estimador de plug-in es en general el más razonable, y es de hecho el mejor si no tenemos otra información sobre F que la obtenida de la muestra. En este, caso, el estimador $\hat{\theta} = t(\hat{F})$ no puede ser mejorado, al menos en el límite asintótico $(n \to \infty)$.

El *bootstrap* nos permite estudiar el sesgo y el error estándar de los estimadores de plug-in, aun para los casos complicados de $\theta = t(F)$. Veamos cómo funciona. Una *muestra de bootstrap* es $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ se obtiene muestreando aleatoriamente con reposición n veces, a partir de la muestra original $(x_1, ..., x_n)$.

Dada una muestra de tamaño n, generamos un número relativamente grande (típicamente 100) de muestras de bootstrap $x^{*1}, x^{*2}, ..., x^{*B}$, cada una también de tamaño n. Para cada muestra de bootstrap tenemos la correspondiente *réplica* del estimador, $t(\hat{F}_{x^{*b}})$, el valor del estadístico evaluado en x^{*b}. Por ejemplo, si $t(F)$ es la esperanza o la mediana de la distribución, entonces $t(\hat{F}_{x^*})$ es la correspondiente media o mediana de la muestra de bootstrap.

La *estimación de bootstrap del error estándar* es la desviación estándar de las réplicas de bootstrap.

$$\hat{se}_{boot} = \left(\sum_{b=1}^{B} \frac{(t(\hat{F}_{x^*}) - t(\cdot))^2}{B-1} \right)^{1/2}$$

donde $t(\cdot) = \sum_{b=1}^{B} t(\hat{F}_{x^*})/B$.

El bootstrap empírico (o no paramétrico) trabaja entonces de una manera curiosa: para obtener la estimación de un paraémetro *poblacional*, toma la *muestra* como si fuera la población total y procede a tomar muestras independientes a partir de la misma, todas del mismo tamaño.

Nótese que al muestrear **con** reposición, los n datos originales se convierten en *infinitos*, pues cada uno puede aparecer, con la misma probabilidad en cada extracción. Concretamente, el vector aleatorio $\vec{X} = (x_1, ..., x_n)$ lo transformamos en la familia infinita de vectores aleatorios $\{Y_\sigma\}$ con $Y_\sigma^i = X_{\sigma(i)}$ para alguna función aleatoria de $\{1, 2...n\}$ en sí mismo.

El hecho de que este procedimiento tenga buenas propiedades de convergencia en la mayoría de los casos parece en principio sorprendente. De alguna forma nos dice que la *muestra* tiene mucha más información de la que el mero principio de plug-in parecería indicar.

Esta apariencia *contra-intuitiva* de la técnica de bootstrap generó cierto escepticismo en los inicios de su aplicación (a fines de los 70) y dio lugar al propio nombre con el que se la bautizó[18]. El momento histórico de su introducción en la forma moderna de Efron (la idea general ya existía hacía tiempo) fue clave para su popularidad subsiguiente, pues coincidió con el comienzo de la revolución de la computadoras personales, haciendo que la técnica estuviera disponible para investigadores de cualquier índole (teóricos, aplicados, técnicos, etc.) sin necesidad de equipamiento especial.

Los fundamentos matemáticos precisos del bootstrap empírico se enmarcan dentro de la teoría de *procesos empíricos* (van der Vaart y Wellner [228]) y de *teoremas uniformes del límite central* (Dudley [63]). En el Cap.5 damos los detalles de la teoría que será el centro de nuestros propios desarrollos, para las demás aplicaciones del bootstrap véase por ejemplo Efron y Tibshirani [66] o Hall [98].

[18]"Bootstraps" son las pequeñas orejas con que se jalan las botas militares para calzarse. En una de las hiperbólicas historias del Barón de Münchhausen, el personaje, al estar hundiéndose en un lago, idea la siguiente salida a su entuerto: tomarse de las orejas de las botas y jalarse a sí mismo hacia arriba hasta salir a flote. Naturalmente, dicha operación está, en el mundo físico real, destinada al fracaso (por mera conservación del impulso) pero en el cuento, el protagonista consigue salir ileso.

1.5. Evolución Histórica y Estado del Arte

1.5.1. Desigualdad Económica Multidimensional

Como señalamos en la Sec.1.1.1, la ciencia social ha contemplado desde sus inicios (Rousseau [181]) a la desigualdad como un fenómeno complejo y donde diversas dimensiones deben ser contempladas, tanto por consideraciones de justicia distributiva como por la interacción entre ellas, que impone su carácter dinámico a la distribución de recursos.

En la práctica, como vimos, dicho carácter multidimensional se ha plasmado en técnicas de estratificación que ordenan a los miembros de una sociedad a partir de variables cuantitativas (ordinales y cardinales) que representan cada dimensión.

En este marco, sin embargo, la teoría económica ha centrado su análisis en la estratificación por niveles de ingreso. Diversas razones han marcado dicha tendencia; por un lado el ingreso monetario juega un rol único como unidad de cuenta para dar una idea de los conjuntos de bienes que un agente económico tiene a su alcance (poder adquisitivo). Dado que en microeconomía el input básico de las funciones de utilidad individuales son los conjuntos de bienes, es de esperar que se piense en el ingreso como la variable *natural* a considerar y se aproveche su carácter de numerario para facilitar las comparaciones interpersonales aún con utilidades ordinales.

La economía moderna del bienestar (Boadway & Bruce [29]), en consecuencia, al tratar el problema del bienestar social como un mecanismo de *agregación* de utilidades individuales, ha considerado a la *distribución del ingreso* como el problema fundamental al estudiar la desigualdad económica (Lambert [126]).

Adicionalmente, la adopción de un enfoque multidimensional de la desigualdad, al menos en el enfoque estadístico, da surgimiento a los problemas técnicos que señalamos en las Sec.1.3.1, 1.3.2 y 1.5.1.

El trabajo de Kolm [121] de 1977 introduce una generalización del criterio de Dalton-Pigou al caso multivariado, que se conoce como *mayorización uniforme* y el concepto de *mayorización direccional*, según el cual se multiplican las distribuciones multidimensionales por vectores de precios y se estudian las distribuciones unidimensionales resultantes (List [133]).

El aporte de Atkinson y Bourguignon [14] de 1982 incorpora la generalización del concepto de dominancia estocástica de Hadar y Russell [96] al caso multidimensional. A partir de ello, expanden la idea de Kolm para dar lugar a la interacción de dimensiones, dada por la correlación entre las distribuciones de los diferentes atributos.

El artículo de Maasoumi [140] de 1986 generaliza la clase de entropías de Theil y otras medidas populares para el caso N dimensional.

En su paper de 1995, Tsui [223] generaliza el enfoque axiomático de los índices de desigualdad, según el cual se encuentran índices que cumplen con ciertos axiomas básicos para su aceptabilidad como criterio de desigualdad (simetría, transferencias, etc, cfr. Perez [166] Cap. 6).

El trabajo de List [133] de 1999, propone un método general por el cual diseñar índices que respetan las ideas básicas de Kolm-Atkinson-Bourguignon-Sen-Tsui.

La mayoría de los aportes al campo de la desigualdad económica multidimensional han estado enfocados en el área axiomático-teórica y en la práctica se han basado en la presentación de las medidas propuestas como estadísticas descriptivas antes que en el cómputo de las distribuciones muestrales y asintóticas.

El trabajo de 2007 de Nilsson [157] aplica varios de los enfoques multidimensionales más conocidos al estudio de la economía de Zambia. Para el índice multidimensional de Maasoumi presenta varianzas muestrales en base a remuestreo, pero no para el Gini. No encara, asimismo, la comparación punto a punto de distribuciones.

Dado este panorama, hemos juzgado como un aporte original y relevante al presentar este trabajo de Tesis, el desarrollo de tests que permitan: i) comparar distribuciones multidimensionales en el enfoque *estadístico* (distribuciones muestrales, asintóticas); ii) realizar comparaciones de las distribuciones completas y no en base a medidas de concentración o índices.

1.5.2. Tests de Dominancia Estocástica

El concepto de dominancia estocástica fue introducido a fines de la década de 1960 por Hadar y Russell [95], Hanoch y Levy [99], y Rothschild y Stiglitz [179] en el marco de la teoría del análisis de riesgo. Básicamente, una distribución domina estocásticamente a otra si conlleva una mayor utilidad esperada para todas las funciones de utilidad de una clase determinada, en el enfoque de las funciones de utilidad de von Neumann y Morgenstern [199] [128].

El trabajo de Hadar y Russell [95], presenta la dominancia estocástica y además la caracteriza para el caso univariado, y *a primer y segundo orden*, mediante un teorema que compara ciertos operadores integrales de las funciones de distribución involucradas al comparar dos variables aleatorias. Esto es importante, pues da un criterio práctico, mediante el cual la comparación se reduce a evaluar los operadores con el sólo argumento de la función de disribución, en lugar de las infintas posibles funciones de utilidad involucradas en la clase dada.

El *orden* (1° o 2°) se refiere a la restricción de la comparación, dentro del universo de las funciones de utilidad, a aquellas que son crecientes (1° orden) y a aquellas que además presentan aversión al riesgo, i.e. concavidad (2° orden). Esta reducción de la clase de funciones de utilidad involucradas no es demasiado restrictiva porque involucra la mayoría de las que se consideran relevantes en la teoría del riesgo o la teoría del consumidor. Hablamos de órdenes de grado sucesivo porque existe una jerarquía trivial, dada por el hecho de que la dominancia de menor orden implica la de órdenes sucesivos (simplemente porque las clases de funciones están anidadas).

A principios de la década de 1970 el concepto de dominancia estocástica fue introducido en el estudio de la desigualdad en la distribución del ingreso y sus efectos sobre el bienestar económico. La teoría se relacionó rápidamente con las medidas de desigualdad basadas en la determinación axiomática de índices de desigualdad (Gini, Theil, Atkinson, var-log) y se probaron resultados de equivalencia entre la dominancia estocástica y la dominancia de Lorenz. A segundo orden la equivalencia fue probada en 1984 por Shorrocks [204] y de manera más general, Muliere y Scarisini [154] probaron la equivalencia con la noción de *dominancia estocástica inversa* a cualquier orden.

Las investigaciones más importantes en esta etapa estuvieron guiadas por el trabajo de Atkinson ([13] de 1970), *On the measurement of inequality*, pionero en el estudio cuantitativo moderno de la desigualdad económica en el marco de la economía del bienestar, seguido por trabajos de Rothschild y Stiglitz [180] y A. Sen [199], entre otros.

El aporte de McFadden ([149] de 1989) al estudio estadístico de la dominancia estocástica estableció una de las líneas básicas en que se ha desarrollado la literatura, a saber, el uso de los estadísticos del tipo Kolmogorov-Smirnov. Por su parte, la línea de Anderson ([7] de 1996) utiliza el herramental de los tests χ^2 de bondad de ajuste.

Siguiendo el enfoque de McFadden, el trabajo de Barret y Donald ([15] de 2003, [16] de 2004) desarrolla tests consistentes de dominancia estocástica univariada a orden arbitrario, basándose en la teoría de procesos empíricos.

El enfoque *multidimensional* de la dominancia estocástica (o sea la comparación de vectores aleatorios) fue introducido en líneas análogas al caso univariado por Hadar y Russell ([96] de 1974), probando caracterizaciones de forma similar, sólo que con condiciones adicionales sobre el hessiano de las funciones de utilidad. Levy y Paroush [129] y Russell y Seo [184] encontraron las primeras caracterizaciones de las clases de dominancia multivariadas.

Atkinson y Bourguginon [14] utilizan la generalización de la dominancia estocástica al caso multivariado encontrando condiciones de dominancia suficientes sobre ciertas clases de funciones (cfr. Cap.3) y aplicándolas a la economía del bienestar para extender los resultados de Atkinson [13] al caso multidimensional.

Scarsini [193] presenta condiciones suficientes para la dominancia estocástica multivariada, pero solo en el caso en que las dimensiones son *independientes*.

Epstein y Tanny [67] introducen la noción clave de *transferencias incrementadoras de la correlación* y su efecto sobre la dominancia bivariada. Este concepto permite extender los principios de Dalton-Pigou y Kolm para las transferencias al caso multidimensional.

Tchen [218] encuentra desigualdades que caracterizan las dominancias para distribuciones pero para el caso en que las distribuciones comparadas tienen marginales idénticas.

O'Brien y Scarisini [161] caracterizan las dominancias multivariadas de orden n en base a desigualdades en los momentos de las distribuciones, generalizando los resultados que Fishburn [72] obtuviera para el caso unidimensional.

El estudio estadístico del problema, es decir el diseño de estadísticos de prueba y el problema de hallar su distribución asintótica, sin embargo, no ha sido demasiado desarrollado. La literatura relacionada con la desigualdad económica multidimensional, como veremos en el Capítulo 2, ha estado principalmente ligada a la generalización de los índices de concentración clásicos antes que en la comparación directa de las funciones de distribución.

En consecuencia los tests de dominancia estocástica multivariados son muy pocos en la literatura.

El trabajo de Crawford ([43] de 2005), *A nonparametric Test of Stochastic Dominance in Multivariate Distributions*, es una generalización del enfoque de Anderson [7]. En el presente trabajo, por las razones que expondremos en el Capítulo 5, seguiremos la línea de Barret y Donald, encarando el estudio de la dominancia estocástica con estadísticos de Kolmogorov-Smirnov, generalizando los resultados al caso multivariado.

1.5.3. Kolmogorov-Smirnov

El resultado clásico de convergencia de la función de distribución empírica a la cdf real F es el conocido teorema de *Glivenko Cantelli*, una ley uniforme de los grandes números que data de 1933:

$$\sup_x |(F_n - F)(x)| \xrightarrow[n\to\infty]{} 0$$

Al mismo tiempo, Kolmogorov estudió la convergencia con $n \to \infty$ del estadístico $\alpha_n = \sqrt{n}(F_n - F)$ y más tarde Smirnov estudió las distribuciones límite de las variables aleatorias involucradas en la aplicación del criterio de Kolmogorov. Estos trabajos dieron origen al conocido test de Kolmogorov-Smirnov.

Puede decirse que los *teoremas funcionales del límite central*, y más específicamente los *teoremas del límite central uniformes* (TLCU), nacieron ligados fuertemente a dicho test. Los trabajos de Doob ([56] de 1949), *Heuristic Approach to the Kolmogorov-Smirnov Theorems*, y Donsker ([55] de 1952), *Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Theorems*, que dan origen a los TLCU, surgen como una generalización de los resultados mencionados. De hecho, como señala Billingsley [27], el Teorema de Donsker (también conocido como *principio de invariancia* o más correctamente *principio de invariancia débil*) representa para el resultado de Glivenko-Cantelli lo que el Teorema del Límite Central a la ley de los grandes números.

A partir de la década de 1970, los TLCU han sido objeto de continua investigación. El punto de partida de la nueva generación de investigadores fue el trabajo ya clásico de Billingsley ([27] de 1968) *Convergence of Probability Measures*. El mismo resumía gran parte de los resultados conocidos hasta el momento en cuanto a los TLCU, incluyendo los desarrollos relativamente recientes de Donsker y Skorohod.

Los TLCU pasaron a ser parte fundamental de una rama de la teoría de probabilidades y procesos estocásticos, la teoría de procesos empíricos. Los principales investigadores entre las décadas de 1970 y 1990 fueron R. M. Dudley (del MIT), E. Giné y J. Zinn (Texas A&M University). Sus trabajos generalizaron los teoremas para clases más generales de funciones, siendo las principales las llamadas *clases de Donsker*, las cuales lograron caracterizar (al menos en la suficiencia) en base a condiciones sobre la combinatoria de Vapnik-Černovenkis de las clases de funciones involucradas, y ciertas condiciones de medibilidad (*admisibilidad Suslin*).

Tal vez los resultados más conocidos de Giné y Zinn sean los *teoremas bootstrap* [62][81].

Entre los principales trabajos de dichos autores encontramos: Dudley [58], *Central limit thorems for empirical measures*, Dudley [60], *Universal Donsker classes and metric entropy*, Dudley [61], *Fréchet differentiability, p-variation and uniform donsker classes*, Zinn [243], *A note on the central limit theorem in Banach spaces*, Giné y Zinn [82], *Some limit theorems for empirical processes*, Giné y Zinn [83], *Bootstrapping general empirical measures*, Giné y Zinn [84], *Gaussian characterization of uniform Donsker clases of functions*.

Por su parte, el Trabajo de D. Pollard a principios de los 80 permitió caracterizar clases de Donsker en base a condiciones de entropía, lo cual ha resultado en una fructifera herramienta en las demostraciones de teoremas límite ([171], [62], [228]). El trabajo de Sheeny y Wellner ([202] de 1992), presentó condiciones suficientes para los TLCU y la equivalencia de éstos con las formulaciones*funcionales* de los teoremas límite.

En épocas más recientes el trabajo de Talagrand [217] relacionó las clases de Donsker con la geometría aleatoria y el de Rudelson y Vershynin [183] con la dimensión combinatoria. La línea más clásica de Dudley es continuada en los trabajos de J. Wellner y A. van der Waart, cuyo libro ([228], 1° ed. [1996], 2° ed. [2000]) *Weak convergence and empirical processes*, es una referencia obligada para el estudio de los procesos empíricos, reuniendo gran parte de los resultados mencionados a la manera en que Billingsley lo hiciera en 1968.

Los tests de Kolmogorov-Smirnov han seguido a su vez una evolución significativa. El problema fundamental resultó ser el de pasar de una a varias variables, porque el estadístico de Kolmogorov pierde en este caso su condición pivotal[19]. Como muestra un ejemplo simple de Simpson ([210] de 1951), la distribución del estadístico de prueba (edp) Kolmogorov-Smirnov para el caso univariado continuo es la misma para cualquier función de distribución acumulativa (cdf) subyacen-

[19]Es decir su independencia distribucional de la cdf subyacente.

te F. Sin embargo, al pasar al caso bivariado, dicho autor muestra que ese no es necesariamente el caso.

Para el caso discontinuo, el trabajo de Schmid [195] encuentra distribuciones límite para los estadísticos KS que dependen del valor de F en los puntos de discontinuidad.

El paper de Bickel [25] *"A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case"* encuentra distribuciones límite para el caso multivariado del estadístico KS en el caso continuo, mediante las técnicas clásicas de convergencia débil de Billingsley [27]. Algo importante de este paper es que la estructura general de la demostración del resultado es bastante apropiada para ver las posibles generalizaciones al caso de procesos intercambiables.

Con su trabajo de 1971, *"Convergence Criteria for Multiparameter Stochastic Processes and Some Applications* [26], Bickel y Wichura dieron las herramientas básicas para varios resultados de convergencia en el caso multivariado, incluyendo el de Sen [200] sobre procesos empiricos.

Pranab Sen [201] encuentra un *principio de invariancia* para estadísticos KS multivariados y lo usa para para encontrar ciertas expresiones asintóticas para su distribución de probabilidad.

El problema no paramétrico *de las dos muestras* (i.e. decidir si dos muestras vienen de variables aleatorias con la misma cdf) ha sido encarado desde diversos ángulos a fin de generalizar los resultados al caso multivariado. Friedman y Rafsky en su trabajo ([74] de 1979) *Multivariate generalization of the Wald-Wolfowitz and Smirnov two sample tests*, derivan una generalización con el método del *árbol de expansión minimal* (minimal spanning tree).

Un trabajo más reciente (y más cercano a nuestra línea de investigación) es el de Cabaña y Cabaña ([36] de 1997) que diseñan un test de bondad de ajuste para el caso multivariado a través de la introducción de los *procesos empíricos transformados*.

Como veremos en la Sección 5.2.1, la introducción de los teoremas de Donsker es el paso fundamental para aplicar la teoría de procesos empíricos al problema de las dos muestras. Si bien la pivotalidad del estadístico de Kolmogorov-Smirnov para comparar dos distribuciones (P,Q) de probabilidad se pierde al pasar de la clase de semirrectas en $\mathbb{R}$ a clases más generales, la condición de *clase de Donsker* para la clase de funciones de base hace posible determinar cotas para las distribuciones en base a los teoremas bootstrap.

Por otra parte, para clases de Donsker $\mathscr{F}$, el test para $P = Q$ resulta asintóticamente consistente contra toda alternativa (P,Q) con $\|P - Q\|_{\mathscr{F}} > 0$.

1.6. Aportes Originales del Presente Trabajo

1.6.1. Objetivo General

El presente trabajo tiene como objetivo realizar un aporte relevante a la economía del bienestar cuantitativa, desarrollando una herramienta de estudio empírico de la desigualdad económica multidimensional que permita comparar distribuciones punto a punto, sin necesidad de recurrir a medidas resumen.

En este sentido se diferencia de otros enfoques multivariados que desde el trabajo pionero de Kolm [121] y el desarrollo posterior de Atkinson y Bourguignon [13] han abordado el estudio de la desigualdad multidimensional con el objetivo de obtener una generalización de los índices típicamente utilizados en el análisis de la distribución del ingreso, tales como el coeficiente de Gini, la entropía de Theil, la familia de medidas de Atkinson, etc[20].

Dicho enfoque se basa en la idea de obtener *caracterizaciones axiomáticas* que permitan determinar de manera única índices con ciertas propiedades que se consideran deseables desde el punto de vista de la economía del bienestar, como por ejemplo la simetría y el principio de transferencias de Dalton-Pigou (Tsui [223]).

Desde el punto de vista estadístico, la obtención de las distribuciones muestrales de dichos índices [21] se puede encarar, por ejemplo, a partir de la teoría de estadísticos U y V tal como en el caso univariado (cfr. Brufman & Perez [33]), aunque la dificultad del caso multidimensional será mayor tanto en lo computacional, como por efecto de la dependencia entre las dimensiones.

El enfoque de nuestro trabajo también se diferencia de otros enfoques bien difundidos tales como el *Índice de Desarrollo Humano (HDI)* del PNUD (Ray [177]) o el enfoque de las *capacidades* de Sen (Sen [199]), que buscan combinar diversas variables del bienestar económico asignando *pesos* a las distintas dimensiones para combinarlas en una variable numérica y elaboran luego juicios de bienestar sobre medidas de desigualdad unidimensional similares a las usuales en distribución del ingreso. Lo mismo las ya mencionadas (Sec.1.1.2) medidas de *status socioeconomico* (SES).

Asimismo, enfoques como el de Maasoumi y Heshmati [141] plantean un *control* de los resultados de dominancia en base a otros atributos de los hogares e individuos, como educación, género, etc. Sin embargo, dicho enfoque se realiza por comparación de celdas agrupadas por dichas variables (una idea del tipo tratamiento diferencial o ANOVA) y no mediante el estudio directo de la distribución multivariada.

Nuestro enfoque evita la necesidad de recurrir a índices o medidas resumen, mediante la introducción de los estadsticos de Kolmogorov-Smirnov multivariados para el contraste empírico de la dominancia estocástica multidimensional. A su vez,

[20]Desarrollamos más sobre este enfoque en el Cap.2.

[21]Necesidad ineludible si se trabaja con Encuestas de Hogares.

el hecho de utilizar la dominancia estocástica como criterio de comparación de las distribuciones nos permite, aprovechando los resultados clásicos de Hadar y Russell (cfr. McFadden [149]) y Atkinson [13]) establecer una conexión directa con la teoría estándar del bienestar económico (funcionales de Bergson-Samuelson) mediante los métodos microeconómicos usuales (cfr. Boadway & Bruce [29], Lambert [126]).

De esta manera, la conexión entre los juicios de desigualdad y bienestar derivados de la dominancia estocástica nos permite mantener las conclusiones esenciales del mencionado enfoque de Atkinson y Bourguignon [14], pero a la vez ampliar las posibilidades del herramental estadístico.

El aporte de la presente Tesis en este sentido se resume, pues, en el desarrollo de un herramental que permite:

- Generalizar el resultado de Barret y Donald [15], pasando del caso unidimensional al multidimensional.

- Presentar un enfoque alternativo al de Crawford [43] para testear estadísticamente la dominancia estocástica multimensional.

Por otro lado, considerando el aporte dentro del *enfoque estadístico* de la desigualdad económica (Giorgi [85]) debe tenerse en cuenta que la comparación multivariada de distribuciones presenta problemas teóricos específicos que nuestro enfoque permite superar.

Por ejemplo, el método de la *regresión por cuantiles* (Koenker [120]) está ampliamente difundido para el estudio de distribuciones univariadas. Al pasar al caso multivariado, sin embargo, no existe forma satisfactoria de generalizar el enfoque. De hecho, no hay una definición universalmente aceptada de algo tan básico como la mediana multivariada, y las diversas propuestas (*mediana de Oja, regresión direccional por cuantiles*, etc.) no han ganado una aceptación amplia[22].

La utilización de los estadísticos de Kolmogorov-Smirnov multivariados brinda entonces la posibilidad de una comparación puntual de las distribuciones sin la necesidad de una *ordenación arbitraria* que sería necesaria si quisieramos trabajar con versiones multivariadas de la curva de Lorenz.

Por otra parte el enfoque del estadístico de Kolmogorov-Smirnov, combinando con la técnica de bootstrap y la noción de proceso empírico de permutación, nos brinda una herramienta que es de naturaleza totalmente no paramétrica.

Esto presenta una ventaja por cuanto, aun en el caso univariado, los enfoques paramétricos suelen presentar resultados menos satisfactorios. El trabajo de Schader & Schmid [194], por ejemplo, estudia los resultados de varios ajustes paramétricos de la curva de Lorenz y los compara con el método no paramétrico de

[22] *"The search for a satisfactory notion of multivariate quantiles has become something of a quest for the statistical holy grail in recent years. Despite generating an extensive literature, it is fair to say that not general agreement has emerged"*. Koenker [120] p. 272

Gastwirth para cotas de variabilidad del coeficinete de Gini, encontrando que en general las curvas parmétricas son poco confiables (cfr. Sec.1.4.2).

Sumado este hecho a la problemática intrínseca de definir curvas de Lorenz multivariadas, vemos la ventaja del enfoque adoptado. Esta posibilidad de un enfoque no paramétrico viene dada por las propiedades específicas que presenta el test de Kolmogorov-Smirnov. Para el caso univariado, es conocida la condición de *pivotalidad* bajo H_0 del estadístico de prueba KS clásico para comparar distribuciones no atómicas de variables aleatorias reales sobre la clase de funciones indicadoras de intervalos en $\mathbb{R}$. Esto significa que la distribución asintótica del edp no depende de las medidas subyacentes y se puede computar la distribución del mismo a partir del teorema de convergencia débil del proceso empírico al puente browniano (Billingsley [27] p. 141).

Al pasar al caso multivariado y discontinuo, dicha propiedad de pivotalidad se pierde (cfr. Simpson [210]). Sin embargo, la introducción del herramental de la teoría de procesos empíricos nos permite mostrar que la sucesión de estadísticos KS para testear $H_0 : P = Q$ dada por el bootstrap de dos muestras (o bien por el proceso empírico de permutación), es asintóticamente convergente contra toda alternativa (P,Q) con $||P-Q||_{\mathscr{F}} > 0$ para la cual $\mathscr{F}$ es $P-Donsker$ y $Q-Donsker$.

En base a ello, el test de dominancia estocástica multivariada que introducimos en este trabajo resulta totalmente no paramétrico y con propiedades estadísticas que nos permiten en la práctica encontrar valores p o valores críticos dependientes de los datos, tales que el test basado en ellos resulta consistente contra las alternativas, recobrando en cierto sentido un carácter pivotal.

Como detallamos en el Cap.2, el estudio de la desigualda multidimensional tiene dificultades intrínsecas por sí mismo. Si a ello le sumamos las que se generan por el hecho de que la *estadística* multidimensional tiene los propios, nos encontramos con una situación en la cual el *enfoque estadístico* de la desigualdad multidimensional está muy poco desarrollado.

Los Capítulos 3 a 7 de la presente Tesis buscan dar respuesta a esta particular combinación del problema económico y econométrico en la desigualdad multivariada.

1.6.2. Dominancia Multivariada

Lo que hace relevante en la práctica la noción de dominancia estocástica es que existen resultados que permiten reducirla a la comparación de ciertos operadores integrales aplicados sobre las funciones de distribución.

En las Sec.3.2 presentamos una generalización de lo que denominamos Teorema de Hadar y Russell (que engloba resultados de Hadar y Russell [96], Levy y Paroush [129] y Russell y Seo [184]), pasando del caso en que las distribuciones son absolutamente continuas al caso general de cualquier distribución con soporte compacto.

En la Sec.3.3 hacemos lo propio con el resultado de Atkinson y Bourguignon [14].

Para plantear los tests de dominancia requerimos además una generalización de los operadores introducidos por Davidson y Duclós [45] y Barret y Donald [15] desde el caso univariado al multivariado. Ello lo hicimos en la Sec.5.3.7, donde además introdujimos una sucesión de operadores $\tilde{T}_k$ para facilitar la demostración de los resultados estadísticos.

1.6.3. Tests Multivariados

Probablemente las contribuciones más importantes del presente trabajo estén en la Sec.5.3. A partir de los resultados del Cap.3, en base a las condiciones suficientes de dominancia estocástica bivariada para los casos super y submodular, construimos en la Sec.5.3.1 estadísticos de prueba de dos muestras al estilo de McFadden.

Enunciamos y demostramos luego el Teo.5.3.4 que muestra la convergencia de los mismos en distribución, y un método para hallar valores críticos del test.

En la Sec.5.3.5 hacemos lo propio para la dominancia de segundo orden y en la Sec.5.3.7 mostramos como se demuestra en general para cualquier orden de dominancia.

Debe notarse que los resultados aquí referidos permiten la generalización no solo al caso multidimensional, sino al caso discontinuo. En este sentido se combinan con los resultados del Cap.3 y permiten generalizar completamente el trabajo de Barret y Donald [15].

1.6.4. Variables Dependientes

En la Sec.5.4 extendemos primeramente (Teo.5.4.1.1) el resultado de los tests de dominancia al caso de *procesos intercambiables* (*exchangeable processes*).

Pasando luego a la consideración de porcesos dependientes estacionarios más generales, nos centramos en los llamados *procesos de mezcla* y **presentamos un resultado heurístico (Heurística 5.4.2.1) que permite construir tests para comparar series de tiempo vectoriales en base al método de *bootstrap de dos muestras en bloque* (*2SMBB*).**

1.6.5. Aplicación: Estudio de la Desigualdad Económica en Argentina

Como señalamos en la Sec.1.1.1, la literatura sobre desigualdad económica en Argentina se centró principalmente en la distribución *funcional* (salario-ingresos de capital) y el concepto de *puja distributiva* asociado al mismo (Llach [134], Diamand [52]).

Durante la década de 1990, las reformas estructurales en Argentina y América Latina dieron paso a un menor interés en dichos conceptos y la literatura se centró en la distribución *personal* del ingreso, es decir la distribución del ingreso tal como la entendemos en economía del bienestar[23].

El trabajo de Bocco et al. [30] de 1997 analiza la distribución del ingreso en Argentina, en particular el impacto que la política impositiva tiene sobre la misma. Si bien se trata de un análisis cuantitativo, su enfoque es más bien de tipo descriptivo, no haciendo hincapié en la variabilidad muestral al presentar sus resultados.

En el mismo sentido se orienta el trabajo de 1999 de Brufman y Urbisaia, que analiza el impacto de la reforma del sistema de jubilaciones y pensiones sobre la distribución del ingreso.

El interesante artículo de Altimir y Beccaria [6] del año 2000, estudia los cambios en la distribución del ingreso en las décadas de 1980-90, utilizando medidas clásicas como el coeficiente de Gini o la entropía de Theil. Pese a trabajar con datos de EPH, se omite el tratamiento de las varianzas muestrales, de manera que el enfoque no es *estadístico* en el sentido que le dimos a este término en la Sec. 1.2.

Perez [166] en su trabajo de 2005, estudia la desigualdad en el enfoque estadístico, utilizando dominancia estocástica univariada en base a un test de Barret y Donald [15]. Este es el enfoque que intentamos generalizar en la presente Tesis.

Si bien la mayor parte de la literatura se enfoca en la distribución del ingreso, el análisis se vuelve necesariamente multidimensional al considerar los *factores* que determinan el cambio en la desigualdad.

El trabajo **clásico** de 2001 de Gasparini, Marchionni y Sosa Escudero [78] utiliza la metodología de las descomposiciones microeconométricas para identificar las fuentes del cambio en la desigualdad. Sus resultados implican que el cambio en la desigualdad durante la década de 1990 se deriva en primer lugar del incremento en los retornos a la educación formal, luego del retorno a factores inobservables y en tercer lugar a la reducción de horas trabajadas por los individuos menos educados[24].

Nuestro trabajo de Tesis entonces, con un enfoque multidimensional directo, puede considerarse complementario a dicho tipo de análisis. La relación

[23] Por supuesto existen relaciones entre la distribución funcional y la personal. Al respecto remitimos al trabajo de Camilo Dagum [44]. En ese artículo, el autor introduce una función generadora del ingreso, a partir de la cual el ingreso se explica como una función no lineal del capital humano y la riqueza.

[24] En su análisis de la distribución del ingreso dichos autores computan los índices clásicos (Gini, Theil, etc) en base a EPH sin considerar la variabilidad muestral. Optan por hacerlo así en base a un trabajo previo de Gasparini y Sosa Escudero que juzga importante dicho efecto, pero considera que los resultados del estudio de la desigualdad son robustos ante esas variaciones. Tal vez una consideración similar subyace al mencionado trabajo de Altimir y Beccaria [6]

que en la descomposición microeconométrica se da *factorialmente*, en nuestro análisis estará presente en la *correlación* de las dimensiones que surgirá al construir las distribuciones conjuntas de vectores de atributos individuales.

Otros trabajos que consideran diversos factores de desigualdad económica, como el de Salvia y Vera [188] de 2011, lo hacen también en base a la descomponibilidad de índices como el de Gini en diversos factores, y trabajan también en un enfoque no estadístico.

En este marco, se considera entonces que la utilización de un enfoque estadístico, que compare las distribuciones multivariadas (ingreso y capital humano) punto a punto en base al concepto de dominancia estocástica, es un aporte original y relevante de la presente Tesis a la literatura sobre desigualdad económica en Argentina.

1.6.6. Axiomática y la Macrojusticia de Kolm

A lo largo de la investigación que culmina con la presente Tesis, se encontraron además ciertas observaciones de carácter general, no directamente relacionadas con el desarrollo de tests de dominancia estocástica, aunque sí con ciertas propiedades conceptuales detrás de ellos.

En la Sec.7.1 hacemos algunas observaciones sobre la necesidad de interpretar los juicios sobre bienestar social en un marco más amplio que el de la mera desigualdad, haciendo referencia a la existencia de un *contrato social* subyacente.

En la Sec.7.2 argumentamos que los tests desarrollados en la presente Tesis pueden resultar de utilidad en el estudio estadístico de las políticas derivadas del concepto de *macrojusticia* de Kolm.

Capítulo 2

Desigualdad Económica Multidimensional

2.1. Dimensiones de la Desigualdad Económica

2.1.1. Estratificación en Ciencia Social

Como señalamos en la Sec.1.1.2, la ciencia social usualmente *estratifica* a los individuos de una sociedad en base a tres tipos de atributos individuales: ingreso, educación y condición de empleo. El resultado de dicha estratificación es una estructura que típicamente un sociólogo denominará de *clases* o *estamentos*, mientras el economista puede interesarse sobre *pobres*, *ricos* o *desigualdad*.

Debe tenerse cuidado, por supuesto, en no confundir los niveles epistemológicamente distintos que pueden tener los mismos vocablos. Lo que el científico denomina una *clase*, por ejemplo, será un concepto de trabajo para la ordenación de los datos que releva. Pero para el individuo en sí, puede resultar totalmente ajeno su pertenencia a *esa clase* y considerarse *miembro* de otra. Claramente lo que uno y otro consideran determinate a la hora de definir una clase puede diferir ampliamente, no teniendo por qué corresponder la propuesta del sociólogo con la visión del individuo estudiado.

Por otra parte, en el enfoque marxista (McAll [148]), la *clase* es un concepto que no sirve sólo para partir el conjunto de individuos de una sociedad, sino que en cierto sentido debe considerarse como un ente en sí mismo, con una dinámica de interacción con otras entidades del mismo estilo y afectando a sus propios *miembros*. El problema de la identificación del individuo con su clase en este caso se conoce como el de la *conciencia de clase*, y un analista marxiano no dudará de su importancia *histórica*, a diferencia de lo que recién veíamos como una mera cuestión de diferencia en perspectivas.

Talcott Parsons [164], por su parte, considera que cierto tipo de estratificación de la sociedad en base a la desigualdad ha venido a reemplazar a la *aristocracia* en la sociedad moderna.

Nosotros adoptamos el enfoque corriente, según el cual la estratificación (y los *estratos* derivados de ella) es un concepto de trabajo del científico y no adquiere entidad *per se*. De esta forma, la valoración de dicho proceso de estratificación estará dada por la utilidad que la misma nos provea a la hora de estudiar un determinado proceso social.

Los sociólogos, por ejemplo, encuentran que los conceptos tradicionales de clase social (*big classes*) son de poca ayuda para estudiar fenómenos en los cuales las características ocupacionales están altamente institucionalizadas[1]. Proceden entonces a dividir aquellas en *micro-clases* de ocupaciones detalladas, que permiten capturar gran parte de la variabilidad en oportunidades, actitudes y comportamientos (Weeden & Grusky [230]).

Los modelos sociológicos pueden también volverse mucho más sofisticados, encontrándose por ejemplo en el enfoque de Haller que la estratificación no necesariamente es de caracter discreto, sino que el status varía en un continuo a través de los cuales el individuo se mueve en una carrera a lo largo del tiempo. El trabajo de Woelfel y Murero [237] presenta una versión de este enfoque que se conoce como *sistema Galileo* y lo aplican encontrando diversas *leyes de movimiento* para las mencionadas trayectorias.

Desde nuestro enfoque, basado en la teoría moderna de la economía del bienestar, nuestra valoración del proceso de estratificación se debe a la posibilidad de realizar juicios de valor sobre distintas situaciones sociales. Este es el enfoque de Atkinson, según el cual una sociedad tendrá menor bienestar cuanto más desigual sea.

Como dijimos en la Introducción, el enfoque principal ha estado siempre en la distribución del *ingreso*, por diversas razones. Por empezar, el ingreso es una *buena* variable en el enfoque *bienestarista*, porque a falta de una posibilidad de medir las *utilidades* de los individuos el ingreso nos da una idea del conjunto de consumo a su disposición. Es de esperar que individuos con un conjunto de consumo más amplio tengan la posibilidad de acceder a un nivel de utilidad mayor.[2]

Esta es la idea detrás de funcionales de Bergson-Samuelson del tipo $\int u(y)dF$, en la cual unificamos las funciones de utilidad en alguna u representativa y nos fijamos únicamente en el efecto de la distribución de y, dada por F.

Por otra parte, el estudio *unidimensonal* de la desigualdad se ve facilitado por la posibilidad de definir ciertas herramientas básicas, como la curva de Lorenz, ausentes en el caso multidimensional. Por ello las técnicas de estudio de la desigualdad en el ingreso se adoptaron también para otras variables permitiendo extender la

[1]La clase como *ente* también ha ido perdiendo importancia: *"The thesis I should like to state as underlying the present analysis is that class in this sense represents a transitional phase in the development of the stratification systems which have become prominent in modern societies since the industrial revolution."* Parsons, T [164].

[2]Aunque como señalamos más abajo, Sen [199] aporta ciertas ideas en contrario, en el enfoque de las *capacidades*.

idea práctica de desigualdad económica a otras dimensiones, aunque sea *una por una* o con un promedio ponderado como el SES.

En principio, sin embargo, aun cuando supongamos un única u puede tener sentido preguntarnos cuál es el efecto de la distribución de los *bienes* que son *argumentos* de la misma. Si el espacio de bienes es n dimensional, nos estaríamos preguntando que pasa con funcionales de la forma $\int u(x_1,..,x_n)dF$, donde F es ahora la función de distribución n-variada.

Aquí es donde entra el problema clásico de la *maldición de la dimensionalidad* en una de sus versiones. Como veremos en el Cap.3, a falta del conocimiento de la u, y sin querer imponer una forma funcional arbitraria, se pueden hacer juicios de bienestar para clases bastante amplias de funciones al comparar dos distribuciones F y G. Pero estos juicios se basan en desigualdades de operadores integrales actuando sobre las distribuciones, y como es sabido (Novak & Hinrichs [106]) al computar integrales la necesidad de puntos y operaciones crece exponencialmente con la dimensión.

Por ello es que debemos *reducir* las dimensiones de estratificación a un número relativamente razonable para que sea posible la implementación práctica.

Pero el abanico de dimensiones es muy amplio y, fuera del *ingreso* por las razones antedichas, no es obvio cuáles serán las fundamentales. La respuesta concreta, por supuesto, depende del caso: como señalamos, factores etno-religiosos pueden ser importantes a la hora de explicar el conflicto social (Justino [112]) o la persistencia de ciertas segregaciones (McAll [148]).

Para darnos una idea de cuáles pueden ser relevantes en nuestro caso, podemos aprovechar los estudios desde otras aproximaciones al problema de la desigualdad. Los enfoques *no bienestaristas*, si bien a veces no tienen una interpretación económica directa, al dejar de lado los funcionales de Bergson Samuleson, no caen en los problemas técnicos de la dimensionalidad y pueden trabajar con muchas variables a la vez.

El enfoque de las *capacidades* de Sen [197] cambia la base de información del bienestarismo, pasando del espacio de bienes (o su resumen en el ingreso) a un espacio de *funcionamientos* (*functionings*), que engloba las variadas cosas que una persona disfruta hacer o ser. Estos funcionamientos son en principio muy variados: desde el consumo de bienes básicos y salud, hasta factores tan abstractos como el respeto por uno mismo.

Las *capacidades* de un individuo están dadas por el conjunto de funcionamientos a su alcance, y estos pueden variar según las condiciones socioeconómicas del individuo (e.g. el ingreso) y según características personales (e.g. un inválido no tiene a su alcance ciertas actividades).

La *elección* de los funcionamientos entonces, plantea el primer reto del investigador, pues se basa en sus propias preferencias:

"There is no escape from the problem of evaluation in selecting a

> *class of functionings and in the corresponding description of capabilities"* (Sen [197] Ch.3).

Por otra parte, aún cuando estemos de acuerdo en cuáles son los funcionamientos o capacidades relevantes, en la práctica, al operacionalizar el enfoque de las capacidades de Sen, terminamos con un *índice* basado en pesos relativos al estilo del *SES* para ciertas variables o indicadores que el investigador encuentra relevantes. Es decir, no se trabaja sobre las distribuciones de los vectores de funcionamientos sino que se los reduce a uno o más índices de manera más o menos razonable, pero en última instancia arbitraria.

El primer ejemplo práctico de este enfoque, el *Human Development Index* o *HDI* fue introducido en 1990 por el UNDP y reduce el estudio a un índice numérico (Fukuda-Parr [76]), sobre el cual se puede estudiar la desigualdad unidimensional.

Otro ejemplo es el trabajo de Grasso [89], que es un poco más sofisticado y en su versión operativa del enfoque de Sen, el CFM (*Conversion Factors Model*) reduce el estudio a tres sub-niveles: *salud*, *educación* e *interacciones sociales*. Pero el estudio no hace hincapié en las distribuciones conjuntas de las dimensiones, sino más bien en tres índices unidimensionales separados.

Los estudios sobre *pobreza multidimensional* también suelen trabajar con una amplia gama de variables o indicadores. El trabajo de Asselin [12] analiza los sistemas de medición de la pobreza *CBMS* (*community based monitoring system*) para Filipinas y Burkina Faso, que trabajan con 37 y 40 variables respectivamente. Estas variables se engloban a su vez en ciertas categorías: *ingreso, educación, salud, nutrición, acceso al agua, empleo, vivienda, acceso a mercados, participación social*.

Cada categoría tiene por supuesto una importancia diferente, reflejada el número de variables que engloba. En ambos casos, alrededor de la mitad de los 40 items se pueden agrupar en las categorías de ingreso, salud y educación.

Vemos entonces que pese a la amplitud inicial de las variables abarcadas, la mayoría de los enfoques multidimensionales están dominados por el efecto de dimensiones que podemos resumir en ingreso y lo que hemos llamado *capital humano*.

Atkinson y Bourguignon [14], en un trabajo que resulta la base fundamental de parte de nuestro desarrollo (Cap.3), plantean un enfoque desde la economía del bienestar, y a la hora de ejemplificarlo toman en consideración las variables ingreso y *expectativa de vida*.

Nosotros trabajaremos, también en un enfoque bivariado de las funciones de bienestar social, y siguiendo la práctica usual de la disciplina, con una construcción de la variable capital humano en base a la educación y experiencia laboral. Esta variable tiene importancia por sí misma en el estudio de la estructura social de un país y permite en principio juicios de bienestar univariados por comparación (Kensuke Miyazawa [153]).

Por otro lado, como señalamos en la Sec.1.1.3, la variable capital humano *interacciona* de diversas formas con la variable ingreso, siendo la más obvia el concepto minceriano de *retornos a la educación*. Justamente esta *correlación* entre las variables es la base de la interpretación económica de la dominancia estocástica multivariada, tal como desarrollaremos en el Cap.4.

A su vez, debe señalarse que el estudio de la estratificación social por sí misma no es totalmente satisfactoria si no se puede dar una idea de la capacidad de cambio de la misma, algo que los sociólogos llaman *movilidad social* (Goldthorpe [86]) y en este sentido resulta evidentemente clave el papel de la educación formal.

Por estas razones, creemos que la combinación de las dimensiones ingreso y capital humano da una buena perspectiva dinámica de la desigualdad económica, al tiempo que el enfoque de Atkinson-Bourguignon nos permite juicios normativos basados en la economía del bienestar.

2.1.2. Capital Humano: Ingreso y Educación

La mayoría de los autores (Slottje [211], Becker [18], Neal y Rosen [155]) coincide en que la teoría del capital humano, como explicación de la variabilidad de los salarios entre los individuos, tiene su origen en la obra de Adam Smith:

> *"A man educated at the expence of much labour and time to any of those employments which require extraordinary dexterity and skill, may be compared to one of those expensive machines. The work which he learns to perform, it must be expected, over and above the usual wages of common labour, will replace to him the whole expence of his education, with at least the ordinary profits of an equally valuable capital."* (A. Smith [213] p.118)

La idea básica, entonces, es que un mayor nivel de educación nos hace más productivos o, lo que es lo mismo, nos brinda un mayor salario en el mercado. Se plantea entonces un problema de equilibrio: por una lado el tiempo pasado en la escuela nos brindará un mayor salario futuro, pero por otro, tiene un costo directo, al pagar el costo de producción de la educación, y otro indirecto, debido a que el tiempo que pasemos educándonos es tiempo que no estaremos en el mercado laboral ganando un salario.

A fines de la década de 1950 Mincer retoma el problema de las diferencias de salarios como producto de los años de educación (Polachek [170]). En su versión más elemental, podemos encontrar la *función de ganancias de Mincer* (*Mincer earnings function*) mediante la condición de equilibrio entre un empleo que paga un salario w_0 y requiere n_0 años de educación y otro que paga el salario w_1 pero requiere $n_0 + s$. En los años que van de n_0 a $n_0 + s$, el trabajador que se está educando no obtiene ingresos.

Dada una vida activa esperada por el trabajador de n períodos, si la tasa de descuento de ingresos futuros es ρ, para que el asalariado sea indiferente entre los dos empleos debemos tener:

$$\int_{n_0}^{n} w_0 e^{-rt} dt = \int_{n_0+s}^{n} w_1 e^{-rt} dt$$

o lo que es lo mismo:

$$\frac{w_1}{w_0} = \frac{e^{-rn} - e^{-rn_0}}{e^{-rn} - e^{-r(n_0+s)}}$$

y ahora con $n \to \infty$:

$$\frac{w_1}{w_0} = e^{rs} \Rightarrow \log w_1 = \log w_0 + rs \tag{2.1}$$

Por supuesto, esta ecuación presupone cierta homogeneidad en otras variables como habilidad personal, preferencias laborales, etc, pero en su forma elemental es una buena idea de cómo la dispersión en los salarios puede venir dada por la dispersión en la educación adquirida. Más aun, en toda su simplicidad, alcanza para ver que los años de educación s añaden *potencial de ingreso* (*earning power*) al trabajador.

Aunque está claro que este resultado del modelo simple con horizonte infinito no es demasiado ajustable al mundo real, unas pocas modificaciones permiten predecir patrones bastante realistas.

El modelo de Ben-Porath[3] trabaja con la idea de *ciclo vital*, y parte del efecto clave que un horizonte *finito* tiene sobre los comportamientos, modelando la decisión de invertir en capital humano como un problema de maximización del valor presente de los ingresos futuros.

Básicamente, una vida finita implica que las ganancias derivadas de la inversión declinan a medida que el trabajador envejece. A su vez, esto implica entonces una declinación progresiva de las inversiones en capital humano, lo que significa que el stock de capital humano y el potencial de ingreso van creciendo a una tasa positiva pero cada vez menor.

Entre las consecuencias *realistas* del modelo de Ben-Porath está que explica por qué las personas adquieren su educación en la juventud, la mayor movilidad geográfica de los jóvenes, y por qué los ingresos crecen rápidamente al comienzo de la vida activa pero van moderando su variación al envejecer el trabajador.

Llamemos $h(t)$ al stock de capital humano de un individuo a tiempo t. Si suponemos una renta del capital humano (salario) w, entonces $wh(t)$ es el *ingreso potencial*. El ingreso real es el potencial menos los costos $c(t)$ que en principio incluyen los gastos en bienes y servicios para obtener el capital humano más los

[3] Seguimos a Polachek [170], aunque el lector queda advertido de los múltiples errores de tipeo e inconsistencias notacionales si va a consultar dicha referencia, sin duda producto de un mal equipo editorial.

ingresos perdidos por no trabajar. Simplificando, sólo tenemos en cuenta estos últimos y si llamamos $s(t)$ al tiempo gastado en invertir en capital humano en el período t tendremos un ingreso neto:

$$y(t) = wh(t) - s(t)wh(t) = (1 - s(t))wh(t)$$

El objetivo del trabajador es entonces, como dijimos al principio de esta sección, maximizar el valor presente de los ingresos a lo largo del ciclo de vida:

$$\max_{s(t)} \int_0^T e^{-rt} y(t) dt \tag{2.2}$$

A su vez la dinámica del capital humano está sujeta a la ecuación de estado $\dot{h} = q(t) - \sigma h(t)$, donde σ es la tasa de depreciación del capital humano y $q(t)$ es el capital humano creado en el período t.

El capital humano creado a tiempo t depende naturalmente del tiempo empleado en ello $s(t)$ y del stock acumulado hasta entonces $h(t)$, de manera que $q(t) = f(s(t), h(t))$ para alguna función $f(x, y)$ que suponemos suave, siendo el ejemplo típico la de Cobb-Douglas $f(x, y) = b(xy)^{\beta}$ para algunas constantes b, β.

El hamiltoniano descontado resulta entonces:

$$\begin{aligned}\mathscr{H}(h(t), s(t)) &= e^{-rt} y(t) - \lambda(t)(f(s(t), h(t)) - \sigma h(t)) = \\ &= e^{-rt}(1 - s(t))wh(t) - \lambda(t)(f(s(t), h(t)) - \sigma h(t))\end{aligned} \tag{2.3}$$

de aquí que la condición de primer orden dada por el principio del mínimo ($\mathscr{H}_{s(t)} = 0$) implica que $\lambda(t) f_{s(t)} = wh(t)e^{-rt}$, es decir que en la trayectoria de inversión óptima los individuos igualan el costo marginal descontado de invertir su tiempo ($wh(t)e^{-rt}$), con las ganancias marginales ($\lambda(t) f_{s(t)}$).

La dinámica exacta depende entonces de la forma de $f(s(t), h(t))$ y bajo ciertas hipótesis se deriva algebráicamente una dinámica completa[4]. Lo que se sigue de este análisis es una dinámica con varias fases de inversión en capital humano.

La fase inicial resulta de una solución de *esquina* al problema de optimización es una etapa de especialización pura de la inversión. Esta es la etapa de *escuela*, en la cual el individuo dedica todo su tiempo al aprendizaje y no genera ingresos, aunque obviamente sus *ingresos* potenciales vana aumentado con su acumulación de h. En esta etapa el stock de capital humano crece y a una tasa también creciente.

En la siguiente etapa, la *inversión post-escuela* o *entrenamiento en el trabajo* el individuo trabaja e invierte. Esta fase se caracteriza por el hecho de que el stock

[4]Sobre este tema, está bueno el artículo de Weiss [231] y aun mejor las clases de Acemoglu y Autor [1] (Ch1. Sec.8).

de h crece, pero a una tasa cada vez menor, dado que el tiempo dedicado por el trabajador a la inversión ed capital humano es cada vez menor[5].

Debe tenerse en cuenta que este modelo parte de una serie de supuestos que no siempre son demasiado realistas (y cuya modificación ha dado lugar a gran parte de la literaura sobre capital humano en los últimos años):

- El modelo asume neutralidad al riesgo de los trabajadores. Dado que se trata de una *decisión de inversión* esto es poco probable y debe más bien adecuarse a la teoría de inversión bajo incertidumbre (Dixit & Pindyck [54]).
- Los individuos saben con certeza cuántos años deben trabajar a lo largo de su vida.
- Se asume que los individuos trabajan durante la extensión completa de su vida activa. Los modelos más elaborados contemplan la ausencia temporal del mercado de trabajo, sobre todo porque resulta empíricamente más relevante para las mujeres.
- Los individuos tienen todos las mismas *habilidades*, las mismas preferencias entre trabajo y descanso, y entre consumo actual y futuro.
- Los trabajadores sólo difieren en la *cantidad* de h, pero no en el *tipo* de capital de que se trata. Desarrollos modernos que incorporan este efecto (Willis [236]) brindan una mejor comprensión, pues corrigen el desbalance en los estudios de capital humano que centran su atención casi siempre en la *oferta* del mismo y no en su *demanda*.
- El w está fijo exógenamente para toda la trayectoria de inversión y no se estudia cómo se determina.

Mincer [152] deriva una *función de ingresos* (*earning functions*) a partir de la trayectoria de inversión en capital humano del individuo. Empecemos trabajando en tiempo discreto y llamemos E_t al ingreso potencial, es decir, sin tomar en cuenta el costo de invertir en capital humano. A tiempo 0 suponemos un E_0 derivado de la habilidad innata del individuo; a $t = 1$ tenemos entonces $E_1 = E_0 + rk_0$ donde k_0 es la inversión del período en capital humano, y en general $E_{t+1} = E_t + rk_t$.

Empíricamente es difícil medir (o mejor dicho estimar) k_t directamente así que definimos el factor $s_t = k_t/E_t$, la proporción de ingreso potencial destinada a la obtención de capital humano y que *grosso modo* podemos asimilar con el tiempo pasado en le escuela en el período t. Dado que $k_t = s_t E_t$ tenemos de inmediato la relación $E_{t+1} = E_t(1 + rs_t)$ y por inducción es fácil probar que:

[5] A medida que pasa el tiempo tiene menos sentido seguir invirtiendo en capital humano porque queda menor tiempo para disfrutar de los beneficios en términos de mayor salario. Por supuesto, esto no tiene en cuenta beneficios indirectos del capital humano sobre el bienestar, individual y social.

$$E_t = E_0 \prod_{i=0}^{t-1} (1 + rs_i) \tag{2.4}$$

Tomando logaritmos se tiene entonces:

$$\log E_t = \log E_0 + \sum_{i=0}^{t-1} \log(1 + rs_i) \tag{2.5}$$

Aproximando $\log(1 + rs_i) \approx rs_i$, la última ecuación se convierte en:

$$\log E_t = \log E_0 + \sum_{i=0}^{t-1} rs_i \tag{2.6}$$

Ahora la clave está entonces en el patrón o *trayectoria* de s_t dado por el *ciclo vital* del que hablamos más arriba. Durante los años de escuela, $s_t = 1$, porque esencialmente sólo ocupa su tiempo en educarse. A medida que pasa el tiempo, s_t va cayendo hasta anularse en el momento del retiro. Por otro lado, si bien vinimos considerando una única tasa de retorno del capital humano, es natural esperar que los *años de escuela* y los de *entrenamiento en el trabajo* tengan distinto valor de mercado, digamos r_s y r_p.

Tenemos entonces, si el número de años de escuela es S:

$$\log E_t = \log E_0 + r_s S + r_p \sum_{i=S}^{t-1} s_i \tag{2.7}$$

o en tiempo continuo para la inversión post escuela:

$$\log E_t = \log E_0 + r_s S + r_p \int_S^t s(u) du \tag{2.8}$$

Esta es *casi* la ecuación elemental con la que comenzamos, pero ahora tenemos un término nuevo, derivado de la estructura de inversión en capital humano a lo largo del ciclo de vida, en particular después de la escuela (i.e. para $t > S$). La teoría (e.g. Ben Porath) nos dice que $k(t)$ va a ir cayendo con t pero no nos dice *cómo*, y aquí debemos proponer algún modelo y ver cuál elegir en términos de utilidad empírica y simpleza. Mincer trabajó con varios, pero el más difundido es el de decaimiento lineal:

$$s(t) = s_0 - \frac{s_0}{T}(t - S)$$

donde T es el horizonte temporal, i.e. el periódo máximo donde la inversión en h es positiva. Esto nos lleva a un ingreso potencial:

$$\log E_t = \log E_0 + r_s S + r_p \left(s_0(t - S) - \frac{s_0(t - S)^2}{2T} \right) \tag{2.9}$$

y dado que el ingreso efectivo es $y_t = E_t(1 - s_t)$ tenemos:

$$
\begin{aligned}
\log y_t = \log E_0 + r_s S + r_p \left(s_0(t-S) - \frac{s_0(t-S)^2}{2T} \right) + \log(1-s_t) \approx \\
\approx \log E_0 + r_s S + r_p \left(s_0(t-S) - \frac{s_0(t-S)^2}{2T} \right) + s_t + s_t^2 = \\
= \log E_0 + r_s S + r_p \left(s_0(t-S) - \frac{s_0(t-S)^2}{2T} \right) + s_0 - \frac{s_0(t-S)}{T} + \left(s_0 - \frac{s_0(t-S)}{T} \right)^2 = \\
= \log E_0 - \left(s_0 + \frac{1}{2}s_0^2 \right) + r_s S + \left(r_p s_0 + \frac{s_0}{T} + \frac{s_0^2}{T} \right)(t-S) - \left(\frac{r_p s_0}{2T} + \frac{s_0^2}{2T^2} \right)(t-S)^2
\end{aligned}
\tag{2.10}
$$

ecuación esta última que tiene la forma clásica de la *función de Mincer*: es lineal en S (los años de escuela) y es cuadrática en $(t-S)$ (la experiencia laboral), o sea es de la forma:

$$
\log y_t = \beta_0 + \beta_1 S + \beta_2(t-S) + \beta_3(t-S)^2 \tag{2.11}
$$

Esta ecuación nos remite a un problema clásico de la microeconometría, conocido como el de *retornos a la educación*, que consiste básicamente en estimar el valor del coeficiente β_1. Este problema presenta algunas peculiaridades que lo hacen particularmente interesante en econometría (Griliches [92]).

Supongamos que intentamos estimar los coeficientes en base a *datos de corte transversal*, i.e. datos de ingresos, educación, edad, etc., para un conjunto de individuos a un tiempo dado.

Para empezar, si miramos la ecuación 2.10, vemos que existe un efecto del término $\log E_0$ que depende de la *habilidad innata* del trabajador. Esta variable es en principio muy difícil de definir, no digamos ya de *medir*. Si el econometrista opta por simplemente *ignorar* dicho factor de *heterogeneidad inobservable*, cae sin remedio en el conocido *sesgo por variable omitida*, que dará una estimación del coeficiente asociado a S mayor que la realidad.[6]

Una opción es usar alguna variable que tenga una interpretación plausible como habilidad, como puede ser el *IQ*, pero esto en general no funciona muy bien por los propios problemas de medición de dichas variables y la poca relación que parecen tener con la habilidad para generar ingresos (Griliches [92]).

Por otro lado, el efecto de la heterogeneidad inobservable nos dice además que la verdadera estimación de estos retornos a la educación debería hacerse no con datos de corte transversal, sino con series temporales para cada individuo, pues queremos conocer el impacto de los años de educación sin que se mezclen las

[6]Si existen además *errores de medición* en la variable S, puede que ni siquiera tengamos certeza sobre el signo del sesgo (Griliches [91, 92]).

diferencias de *habilidad*, etc. Esto se conoce como *datos de cohorte* y son pocas las afortunadas veces que podemos contar con bases de este tipo[7].

En base a ello, si contaramos con dichos datos, parece ser que la técnica que mencionamos en la Sec.1.4.2, los métodos de *datos de panel* pueden darnos una estimación de coeficientes fijos sobre las *diferencias* de la ecuación 2.11. Lamentablemente, en este caso no podemos aprovecharnos de la estructura del panel, porque en la práctica, los años de educación presentan muy poca variabilidad en la serie temporal, y no se pueden realizar regresiones confiables sobre el modelo diferenciado (Arellano [8]).

Otras variantes incluyen el uso de *variables instrumentales* (Wooldridge [239]) que permitan tratar el problema de la endogeneidad de la variable S si omitimos la variable E_0, y el particularmente ingenioso uso de los *datos de gemelos* (Ashenfelter y Krueger [11]). Con este último tipo de datos, disponiendo de un panel completo para muchos pares de gemelos, sí se podrá aplicar el modelo en diferencias y obtener una regresión sin sesgo[8].

Otros aspecto interesante de este resultado es que nos da una pista, al mostrar como van variando los ingresos con S y con t, de cómo se acumula el capital humano. En particular, el modelo con decaimiento lineal de la tasa de inversión nos lleva a una función cuadrática en la experiencia laboral y lineal en los años de educación. Sin embargo, como veremos en el Cap.6, Sec.6.2, en la práctica se encuentra que los modelos de acumulación individual de capital humano que mejor funcionan son los *no lineales*.

Debemos señalar, además, que el capital humano como tal tiene efectos positivos que van más allá del retorno individual como mayor salario. En este sentido, la teoría del capital humano no sólo es un enfoque relativamente exitoso para explicar la dispersión de salarios, sino que puede aportar a otros estudios y ramas de la economía.

Por caso la moderna teoría macroeconómica del crecimiento endógeno pone a la acumulación de capital humano como motor del crecimiento (Lucas [138]). Es cierto que la literatura que testea empíricamente esta relación deja bastante lugar a la duda (Lange y Topel [127]), pero de todas formas es un proposición generalmente aceptada por la profesión.

Pero existen además muchas otras *externalidades* derivadas del capital humano. Como señala Michael Grossman, al hablar de los resultados socialmente deseables de la educación más allá del salario (*nonmarket outcomes*):

[7] El ejemplo paradigmático de este tipo de base, al que todos los econometristas estamos agradecidos, es por supuesto el del *Panel Study of Income Dynamics (PSID)*. Otras bases de este tipo se describen en el apéndice de Rubinstein & Weiss [182].

[8] Por supuesto, esto está sujeto a ciertas condiciones, como la del *error clásico* (error independiente de la verdad). Cuando estas condiciones no se cumplen, el método puede dar un resultado sesgado (Flores-Lagunes & Light [73]).

> "*Examples of outcomes considered include general consumption patterns at a moment in time, savings and the rate of growth of consumption over time, own (adult) health and inputs into the production of own health, fertility, and child quality or well-being reflected by their health and cognitive development. They are distinguished from the labor market outcomes of education in terms of higher earnings and wage rates.*" (Grossman [94])

Otros efectos positivos de la educación que han interesado a los economistas tienen que ver con la reducción del crimen y la mayor participación política (Lochner [135]). **En este sentido es de esperar que el efecto social de la externalidad tenga que ver no sólo con el *stock* de capital humano sino también con su mayor o menor desigualdad.**

Por último, señalemos que en cuanto a los efectos del capital humano, en particular de la educación, sobre la desigualdad social, no deben ser ignorados los *efectos intergeneracionales*. Este es otro *efecto externo* de la educación: más allá del efecto positivo sobre el ingreso en la vida del trabajador, tiene efectos positivos sobre los ingresos de sus hijos (Plug [169]).

Elegimos entonces para nuestro trabajo sobre la desigualdad bidimensional la variable *capital humano*, porque presenta doble interés. Por un lado su interrelación con la variable ingreso nos da una visión dinámica más acabada de la desigualdad económica.

Por otro, el hecho de que tenga efectos *sociales* más allá de lo individual, la pone en el centro de atención para poder explicar diversos fenómenos complejos.

Por ejemplo, estamos convencidos de que es de esperar la desigualdad en la distribución de capital humano y su efecto de correlación con el ingreso tenga efectos sobre el crecimiento.

En este sentido pensamos que se debe pasar de trabajos que centran su estudio en el *stock* de capital humano a aquellos que tomen en cuenta la estructura subyacente de su distribución con mayor o menor desigualdad. Lo mismo puede aplicarse a los efectos positivos del capital humano, relacionados con la salud, el crimen, la participación política y la mobilidad social intergeneracional.

2.2. La Desigualdad Multivariada en el Enfoque Axiomático

2.2.1. El *zonoide* de Lorenz

Como señalamos en la Introducción, el paso del caso unidimensional al multidimensional en el estudio de la desigualdad encuentra un problema grave enseguida: una de las principales herramientas utilizadas, la Curva de Lorenz, no está unívocamente definida para distribuciones de vectores de dimensión 2 o más.

Esto es claro si recordamos que la curva de Lorenz se obtiene integrando la *función cuantil* $F^{-1}(z)$ asociada con una distribución F. Pues si bien en una dimensión F^{-1} está siempre bien definida (Yohai [241]) por la monotonía y continuidad a derecha de F, en dos dimensiones esto ya no pasa. Por caso, supongamos dada una cdf $F(s,t)$ con soporte en $[0,1] \times [0,1]$ y preguntemos por $F^{-1}(0,5)$. Es claro que $F(s,1) = F^X(s)$ y $F(1,t) = F^Y(t)$ son cdfs univariadas así que tienen bien definidas las respectivas inversas, y en consecuencia su mediana o cuantil 0.5. Pero entonces vamos a encontrar un punto de la forma $(s,1)$ y otro de la forma $(1,t)$ tales que *hasta ellos* se acumula la mitad de la probabilidad, es decir tenemos al menos dos medianas. Es fácil ver que esto se puede hacer cualquier dirección y que pasa lo mismo para toda dimensión ≥ 2.

Dentro del estudio de la desigualdad, tal vez el trabajo más importante para generalizar la curva de Lorenz al caso multidimensional sea el de Koshevoy y Mosler [123] que construyen lo que llaman *zonoide de Lorenz* y en base a ello una *curva* o superficie de Lorenz como una variedad de d dimensiones en un espacio de $d+1$.

Sea $\mathscr{F}$ el conjunto de cdfs sobre $\mathbb{R}^d_+$ tales que tienen esperanzas finitas positivas, es decir $\mu_j = \int_{\mathbb{R}^d_+} x_j dF > 0$ para $j = 1,...,d$, y dado un punto $\vec{x} = (x_1,...,x_d)$ definamos su normalización como siempre $\psi_F(\vec{x}) = (\tilde{x}_1,...,\tilde{x}_d)$ siendo $\tilde{x}_j = x_j/\mu_j$.

Ahora definamos la función ζ que a cada función medible $h : \mathbb{R}^d_+ \to [0,1]$ le asigna el vector $\zeta(h) \in \mathbb{R}^{d+1}_+$ según:

$$\zeta(h) = \left(\int_{\mathbb{R}^d_+} h(\vec{x}) dF, \int_{\mathbb{R}^d_+} h(\vec{x}) \psi_F(\vec{x}) dF \right)$$

Ahora para una F dada podemos definir el *zonoide* de F y el *zonoide de Lorenz* para F como:

$$Z(F) = \left\{ \vec{y} \in \mathbb{R}^d_+ : \vec{y} = \int_{\mathbb{R}^d_+} h(\vec{x}) \psi_F(\vec{x}) dF, \text{ para alguna } h \text{ medible} \right\} \tag{2.12}$$

$$LZ(F) = \left\{ \vec{z} \in \mathbb{R}^{d+1}_+ : \vec{z} = \zeta(h), \text{ para alguna } h \text{ medible} \right\} \tag{2.13}$$

La construcción de la *superficie de Lorenz* para F se realiza definiendo primero la *función inversa de Lorenz*: $l_F : Z(F) \to \mathbb{R}_+$ como:

$$l_F(\vec{y}) = \text{máx}\{t \in \mathbb{R}_+ : (t, \vec{y}) \in LZ(F)\} \tag{2.14}$$

y ahora la superficie de Lorenz es simplmente el gráfico de la función l_F, que es una variedad d dimensional.

Hemos elegido presentar aquí esta construcción de Koshevoy y Mosler por dos razones. La primera es porque se trata de una construcción interesante, que comparte propiedades fundamentales con la curva de Lorenz unidimensional: por ejemplo, si la distribución es completamente igualitaria (i.e. concentrada en un único punto) entonces la superficie de Lorenz es la diagonal principal del cubo $[0,1]^{d+1}$.

Otra propiedad interesante es que se prueban la equivalencia del *orden lorenziano* dado por $LZ(F) \subset LZ(G)$ con la *mayorización direccional*, que indica que para cada $\vec{p} \in \mathbb{R}^d$ la curva de Lorenz generalizada para $F_{\vec{p}}$ está por arriba de la de $G_{\vec{p}}$ donde estas son las distribuciones de las variables construidas como $\vec{p} \cdot \vec{x}$.

Pero la segunda razón para presentar este enfoque ha sido mostrar lo abtruso que puede ser generalizar un concepto tan elemental como la curva de Lorenz al caso multivariado. Esta seguramente es una de las razones por las cuales esta construcción no ha tenido, aun 20 años después, una adopción más o menos extendida dentro de la literatura especializada.

Más aun, si en la práctica queremos utilizar este concepto vamos a tener que encontrar distribuciones muestrales para los *cuantiles* del zonoide, algo en principio nada obvio y probablemente muy complejo.

Frente a ello, nuestro enfoque basado en la dominancia estocástica n dimensional tiene la ventaja de no requerir más que la cdf multivariada, y en consecuencia resulta muy fácil de trasladar al estudio empírico-estadístico de la desigualdad.

2.2.2. Enfoque Multidimensional Clásico

Llamamos *enfoque clásico* al introducido por Kolm [121], que introduce dos nociones de dominancia, la llamada *mayorización uniforme* y la ya mencionada *mayorización direccional o por precios*.

Este enfoque busca ordena distribuciones de vectores de n dimensiones, y para hacerlo parte de matrices de tamaño $m \times n$, en las cuales cada fila representa el vector (la *canasta de bienes* si se quiere) asignado a un individuo, y hay m individuos. Kolm define la mayorización uniforme de una distribución (o sea una matriz, llamémosla $\mathbf{X}$) sobre otra ($\mathbf{Y}$), cuando $\mathbf{Y} = B \cdot \mathbf{X}$, para alguna matriz bi-estocástica B (o sea que tiene columnas y filas, no negativas, que suman 1).

Para el otro tipo de mayorización, observemos que dado cualquier vector $\vec{p} \in \mathbb{R}^n_+$ (*vector de precios*, que pensamos como columna) obtenemos vectores $\vec{x}_p, \vec{y}_p$ de m componentes mediante el producto $\mathbf{X} \cdot \vec{p}$ y lo análogo para $\vec{y}_p$. Ahora tenemos vectores de m componentes reales, cada una asociada a uno de los individuos, o lo que es lo mismo, distribuciones unidimensionales. Podemos decir entonces que **X** mayoriza a **Y** para el vector de precios $\vec{p}$ si la curva de Lorenz de la distribución $\vec{x}_p$ está por sobre la de $\vec{y}_p$.

La mayorización uniforme de **X** sobre **Y**, implica que para cualquier $\vec{p}$ se dará que **X** mayoriza a **Y** en dirección $\vec{p}$, de aquí el nombre de *uniforme*.

Nótese que la noción con sentido económico es la de mayorización por precio, mientras que la uniforme nos da una condición de *unanimidad*. Pero en la práctica, no es elemental *factorizar* para chequear si una matriz ha sido obtenida de otra mediante el producto por una estocástica. Con lo cual no sabemos a priori si una distribución es mejor que otra para *cualquier* vector de precios o solo para algunos.

Elegir un $\vec{p}$ arbitrariamente es poco satisfactorio, pero suele ser usado en la práctica. Una idea alternativa sería un enfoque tipo *montecarlo* probando la desigualdad para una distribución uniforme al azar de vectores $\vec{p}$, aunque esto no es usual en la literatura.

Pero más aun, como señala Savaglio [190], **el principal defecto de este enfoque es que sus resultados son válidos solo en el caso en que las *relaciones* entre los distintos componentes del bienestar (las dimensiones) son irrelevantes para las comparaciones de desigualdad.**

El trabajo de Rietveld [178] trabaja sobre la descomposición de la desigualdad en factores componentes del ingreso, y muestra como si se define una curva de Lorenz para cada subcomponente del ingreso, entonces la desigualdad en el ingreso total no es mayor que la del componente más desigual. En general, la curva de Lorenz para el ingreso total está por sobre el promedio ponderado de las curvas de los componentes.

Trabajando con medidas o índices de desigualdad homegeneas de grado cero y que se expresan como suma de funciones convexas, Rietveld muestra que se producen *efectos mitigadores* de la desigualdad, en el sentido de que el valor de la medida es menor en el ingreso total que en el promedio ponderado de sus componentes. Esto vale también para algunos índices que no son de ese tipo, por ejemplo el coeficiente de Gini. La conclusión general de su trabajo, es que las interrelaciones entre los componentes del bienestar son relevantes para las comparaciones de desigualdad.

El trabajo de Atkinson y Bourguignon [14] generaliza al caso multidimensional las ideas de Atkinson [13]. El concepto de *dominancia estocástica* que utilizan para comparar bienestar[9] tiene varias ventajas.

Por un lado, a diferencia del enfoque de Kolm, en este caso se destaca el papel de la *interrelación* entre las dimensiones de la desigualdad. Por otro, resulta de interpretación directa en términos de bienestar como un orden *unánime* dentro de una clase de funciones de bienestar social.

[9]Este es el concepto clave que utilizamos en esta Tesis, para un desarrollo completo cfr. nuestra Sec.1.2.2 y nuestros Cap.3 y 4.

El hecho de que en dimensión 1 el orden SD_2 sea equivalente al lorenziano, le da además una aceptabilidad bastante natural entre los especialistas en desigualdad. Además, en términos prácticos, es bastante sencillo de caracterizar en base a operadores integrales sobre las funciones de distribución, aun para el caso n dimensional general.

Nótese que en este enfoque la *interrelación* entre las dimensiones se da no sólo en términos "estadísticos" sino que tiene vital importancia la *complementariedad* de los bienes que entran en la función de utilidad, algo que se caracteriza según condiciones de *modularidad* (cfr. Sec.3.2). Esto permite juicios de bienestar social con más significado y en consecuencia posibilita una mayor aplicabilidad a la evaluación de la política económica en términos de bienestar.

Otro enfoque que podemos llamar clásico de la desigualdad multidimensional es el de la *agregación de dos etapas* de Maasoumi (Maasoumi [140], Weymark [233]). El mismo consiste en elegir primero una *función de utilidad* $u : \mathbb{R}^n \to \mathbb{R}$ que permite reducir el vector de cada individuo (su canasta) a un número real (la utilidad) y luego en una segunda etapa utilizar un índice de desigualdad para estudiar la distribución de utilidades.

Maasoumi propone para la segunda etapa usar algún miembro de la clase de entropías generalizadas, que contiene a la clase de índices de Atkinson (Sec.1.2.1) y sus equivalentes ordinales.

El enfoque de Maasoumi es interesante pero tiene ciertas desventajas. Por empezar, hay que elegir, de manera más o menos arbitraria, una función de utilidad para los individuos. En esto contrasta con la dominancia estocástica, que es de caractér unánime sobre clases amplias de funciones.

Por otro lado, se agregan las distribuciones en un índice numérico, lo que implica pérdida de información. De nuevo, esto contrasta con la dominancia estocástica, que utiliza la totalidad de la distribución para las comparaciones.

Otro problema con este enfoque es que no cumple la *mayorización uniforme* de Kolm. Dardanoni (cfr. Savaglio [190] o Weymark [233]) prueba con un ejemplo que ciertos índices construidos con el método de Maasoumi, caen al transformar una distribución multivariada con una matriz biestocástica.

List [133] presenta generalizaciones multivariadas de los índices de Gini y de Atkinson. Estos índices, muestra el autor, tienen la propiedad de respetar a la vez la condición de mayorización uniforme de Kolm y la condición de Atkinson-Bourguignon de la dependencia en la correlación entre las dimensiones.

La clase de índices resultantes es interesante y tiene valor práctico pero debe resaltarse que son *no determinantes*. Es decir, una menor desigualdad medida por alguno o todos estos índices no implica un mayor bienestar para clases amplias de funcionales de bienestar social, como es el caso de la dominancia estocástica. Vale decir, que la comparación de la desigualdad medida por alguno de los índices de List no implica el mismo resutado para otros índices que satisfagan también los principios de Kolm y de Atkinson-Bourguignon.

2.2.3. Axiomática

Si definimos un índice I de desigualdad sobre distribuciones n dimensionales de atributos individuales como una familia de funciones $I^m : \mathbb{R}^{m\times n} \to [0,1]$ (donde m sería el número de individuos), podemos traducir los axiomas de regularidad de los índices de desigualdad univariados al caso multidimensional. Siguiendo a Lugo [139] enunciamos algunos de ellos.

i. *Continuidad:* I^m es continua para cada m.

ii. *Simetría (o anonimidad):* Si P es una matriz de permutaciones, $I^m(\mathbf{X}) = I^m(P\mathbf{X})$.

iii. *Normalización:* Si todas las columnas de $\mathbf{X}$ son iguales $I^m(\mathbf{X}) = 0$.

iv. *Poblacional:* Si armamos una matriz con k copias de $\mathbf{X}$ entonces $I^m(\mathbf{X}) = I^{km}(\mathbf{X},\mathbf{X},...,\mathbf{X})$.

v. *Invariancia de Escala:* Si D es una matriz escalar positiva, entonces $I^m(\mathbf{X}) = I^m(D\mathbf{X})$.

vi. *Invariancia Traslacional:* si $\tilde{X}_{ij} = X_{ij} + a$ para cada par ij con a fija, entonces $I^m(\mathbf{X}) = I^m(\tilde{\mathbf{X}})$

El trabajo de Diez et al. [53] reemplaza el principio de invariancia de escala por el de la *consistencia en unidades*. En dicho trabajo, argumentan siguiendo a Zheng que lo que debe permanecer invariante tras un cambio de escala (unidades) no es el valor numérico del índice, sino el orden derivado del mismo, y caracterizan la clase de medidas de desigualdad multidimensional que son *consistentes en unidades*.

Para dar un principio equivalente al de Dalton-Pigou definimos primero las matrices $T = \lambda\mathbf{1_n} + (1-\lambda)Q$ donde $\lambda \in (0,1)$, $\mathbf{1}_n$ es la identidad y Q es una matriz de permutación de filas. Decimos que $(\mathbf{X},\mathbf{X}T) \in UPD$ si T es una matriz de ese tipo que no sea una matriz de permutaciones.

vi. *Mayorización Uniforme de Dalton Pigou:* $(\mathbf{X},\mathbf{X}T) \in UPD$ entonces $I^m(\mathbf{X}) > I^m(\mathbf{X}T)$.

vi. *Mayorización Uniforme de Kolm:* Si B es una matriz biestocástica (que no sea una permutación) $I^m(\mathbf{X}) > I^m(\mathbf{X}B)$.

El paper de Tsui [224] utiliza la idea de *transformación de correlación creciente* de Epstein y Tanny, para dar un principio que capte la idea de Atkinson y Bourguginon sobre la importancia de la correlación entre las dimensiones. Dados v y w vectores de d coordenadas, llamemos $v \wedge w = (\text{mín}\{v_1,w_1\},...,\text{mín}\{v_d,w_d\})$ y $v \vee w = (\text{máx}\{v_1,w_1\},...,\text{máx}\{v_d,w_d\})$. Ahora dadas matrices $\mathbf{X}$ e $\mathbf{Y}$ diremos que $\mathbf{Y}$ puede ser obtenida a partir de $\mathbf{X}$ por una transferencia de correlación creciente, si para algún par de filas $i < j$ se tiene $\mathbf{y}_i = \mathbf{x}_i \wedge \mathbf{x}_j$ y $\mathbf{y}_j = \mathbf{x}_i \vee \mathbf{x}_j$ mientras que para las demás $\mathbf{y}_k = \mathbf{x}_k$.

vii. *Mayorización de Correlación Creciente:* Si $\mathbf{Y}$ se obtiene de $\mathbf{X}$ por una transformación de correlación creciente, entonces $I^m(\mathbf{Y}) \geq I^m(\mathbf{X})$.

Debemos hacer notar que este último axioma sólo capta la idea de Atkinson y Bourguginon [14] en el caso *submodular* (cfr. nuestra Sec.3.2) es decir cuando los atributos se consideran *substitutos* entre sí.

Este no tiene por qué ser el caso general, y de hecho nuestra aplicación principal, en el presente trabajo es para atributos que consideramos *complementarios*, a saber, el ingreso y el capital humano. En este sentido, es preferible nuestro enfoque de tratar directamente con la dominancia estocástica y no con índices axiomáticamente derivados.

Como señalamos en la Sec.1.2.1, una forma de cumplir de manera inmediata con los axiomas de simetría, poblacional, y traslaciones es trabajar con la F como base de información. Nótese que si estamos usando dominancia estocástica, entonces los resultados son *consistentes en unidades* pues si $F(t) = P(y \leq t)$ entonces para $\alpha > 0$ $P(\alpha y \leq t) = P(y \leq t/\alpha) = F(t/\alpha)$ y la condición de comparación seguirá siendo la misma.

2.3. La Desigualdad Multivariada en el Enfoque Estadístico

Si el enfoque axiomático o matemático de la desigualdad multidimensional ha tenido relativamente menos importancia que el unidimensional, en el enfoque *estadístico* de la situación es todavía peor. A ello se debe probablemente que los trabajos que necesitan *medir* desigualdad multidimensional en la práctica (e.g. los informes del PNUD) recurran a medidas de agregación unidimensional como el SES o el HDI.

La gran mayoría de los trabajos que estudia empíricamente desigualdad multivariada lo hace dando valores para las distintas medidas de desigualdad elegidas, pero sin tomar en cuenta la variabilidad muestral, algo que como hemos dicho no resulta totalmente satisfactorio.

El trabajo de Barry Arnold [9] presenta versiones paramétricas de curvas de Lorenz a partir generalizar definiciones de la curva de Lorenz univariada que están libres de referencias a los cuantiles. Siguiendo este enfoque, el reciente trabajo de Sarabia y Jordá [189] especifica modelos paramétricos bivariados de la curva de Lorenz con marginales beta, gama, Pareto y lognormal, y estudian la distribución del coeficiente de Gini derivado de ellas.

Por supuesto, como ya hemos señalado, un problema con los enfoques paramétricos es que nada garantiza la *consistencia* de los mismos, en el sentido de que asintóticamente tiendan a la verdadera distribución subyacente.

En nuestro trabajo, hemos optado por utilizar la noción de dominancia estocástica multivariada directamente, sin pasar por generalizaciones de la curva de Lorenz que por un lado resultan abtrusas (como las de Koshevoy y Mosler [123]) y por otro recaen en modelos paramétricos de posible inconsistencia.

El principal problema de nuestro camino es entonces poder saltear la *no pivotalidad* del estadístico de Kolmogorov-Smirnov en el caso multivariado, algo que encaramos en el Cap.5. También aprovechamos el potencial de la teoría utilizada para generalizar los resultados al caso en que las subyacentes no son absolutamente continuas, para lo cual tenemos que generalizar los rsultados de Atkinson y Bourguignon en el Cap.3.

Creemos que el enfoque aquí adoptado presenta una serie de ventajas que describimos a continuación.

i. **Como se basa en la dominancia estocástica, es fácil de interpretar directamente en términos de la economía del bienestar, como en Atkinson y Bourguginon [14].**

ii. **Por la misma razón, es un método que toma plena cuenta de la *interacción* entre las dimensiones de la desigualdad.**

iii. **Se puede aplicar tanto a los casos super como submodulares, es decir al caso de atributos *complementarios* como *sustitutos*.**

iv. **Es fácil de usar en la práctica, porque solo requiere de la función de distribución multivariada $F(x_1,...,x_n)$ y de sus integrales, algo conocido por todos los investigadores, aceptable y elemental de calcular.**

v. **Por usar sólo las F, es casi inmediata su formulación estadística, mediante las distribuciones empíricas. Y una vez que tenemos éstas se puede usar el enfoque de McFadden [149] o el de Anderson [7] para definir tests.**

En este sentido creemos que el presente trabajo aportará herramientas de utilidad a la literatura aplicada y a la evaluación de la política económica relacionada con la desigualdad multidimensional.

Capítulo 3

Dominancia Estocástica Multivariada

3.1. Acoplamientos y Teorema de Strassen

Supongamos que tenemos dos distribuciones de ingreso dadas por las variables aleatorias y_1 e y_2 definidas sobre la misma población $\Omega = \{\omega\}$. Si la segunda distribución implica una mejora respecto a la primera en el sentido *paretiano*, es decir si para cada individuo ω se cumple $y_1(\omega) \leq y_2(\omega)$ (o más precisamente si $P\{y_1 \leq y_2\} = 1$), entonces naturalmente el *bienestar* en la segunda economía de distribución debe ser mayor que en la primera.

Lógicamente, este resultado de *sentido común* está cubierto por la noción de dominancia estocástica. Pues si $\forall \omega : y_1(\omega) \leq y_2(\omega)$ entonces inmediatamente $\forall t$: $P(y_2 \leq t) \leq P(y_1 \leq t)$, es decir (por el Teo.de Hadar y Russell) que y_2 domina estocásticamente a primer orden a y_1.

El resultado recíproco, a saber, que la dominancia estocástica implica una mejor posición individual, es trivialmente falso: pensemos por ejemplo, en simplemente permutar los ingresos de una distribución cualquiera no perfectamente igualitaria (la dominancia puntual cambia, pero la distribución es la misma).

Sin embargo, dadas dos distribuciones de ingreso y_1, y_2 que cumplen, por ejemplo, $y_2\ SD_1\ y_1$ podemos encontrar nuevas variables $\hat{y}_1, \hat{y}_2$, estrechamente relacionadas con las anteriores, y tales que $\hat{y}_1 \leq \hat{y}_2$ puntualmente.

El concepto que debemos introducir para ello es el de *acoplamiento* de una familia de variables aleatorias: dado un proceso estocástico $\{X_i : i \in \mathbb{I}\}$, decimos que la familia $\{\hat{X}_i : i \in \mathbb{I}\}$ es un *acoplamiento* de la primera, si cada $\hat{X}_i$ tiene la misma distribución que el correspondiente X_i. Nótese que cada $\hat{X}_i$ es una copia (en el sentido probabilístico) de cada X_i, pero la familia $\{\hat{X}_i : i \in \mathbb{I}\}$ no necesariamente es una copia de $\{X_i : i \in \mathbb{I}\}$ pues la distribución *conjunta* puede ser distinta.

Dada una función de distribución $F(t)$ cualquiera, podemos encontrar una variable aleatoria X tal que su distribución viene dada por $F(t)$, mediante una técnica típica de la simulación[1]. Definamos para cada $s \in \mathbb{R}$ el conjunto $A_s = \{t \in \mathbb{R} : F(t) \geq s\}$.

A partir de ello, podemos definir una *inversa* $F^{-1} : (0,1) \to \mathbb{R}$ de F aun en el caso en que F no es inyectiva, poniendo $F^{-1}(s) = \text{ínf} A_s$. A partir de ella, tomando una variable aleatoria U uniforme en el $[0,1]$, definimos $\hat{X} = F^{-1}(U)$, variable que tiene la misma distribución de X (dada por F) y usualmente se conoce como *acoplamiento cuantil* de X. Ahora, dadas X_1 y X_2 tales que X_1 SD_1 X_2, entonces si llamamos $\hat{X}_1$ y $\hat{X}_2$ a los respectivos acoplamientos cuantil, tenemos la desigualdad *puntual* $\hat{X}_1 \geq \hat{X}_2$.

Hemos probado entonces que la dominancia estocástica de X_1 sobre X_2 es equivalente a la existencia de un acoplamiento $(\hat{X}_1, \hat{X}_2)$ con dominancia puntual. La importancia de este resultado radica, en primera instancia en la posibilidad de demostrar fácilmente el Teorema de Hadar y Russell para dominancia a primer orden que ya hemos mencionado.

Pues si $\hat{X}_1 \geq \hat{X}_2$ puntualmente, entonces para toda función real creciente y acotada ϕ se tiene puntualmente $\phi(\hat{X}_1) \geq \phi(\hat{X}_2)$, y tomando esperanza resulta $E(\phi(\hat{X}_1)) \geq E(\phi(\hat{X}_2))$. Pero como X_i tiene la misma distribución que $\hat{X}_i$, se sigue que $E(\phi(X_1)) \geq E(\phi(X_2))$, es decir que X_1 domina estocásticamente a X_2.

Por otro lado, si suponemos que $E(\phi(X_1)) \geq E(\phi(X_2))$ para toda función creciente y acotada ϕ, entonces dado $t \in \mathbb{R}$ basta tomar $\phi = 1_{(t,\infty)}$ para obtener $P(X_1 > t) \geq P(X_2 > t)$, y en consecuencia $F_1(t) \leq F_2(t)$, como indica el Teorema de Hadar y Russell.

Por supuesto, no sólo nos interesa esta construcción como una forma de demostrar un teorema que ya conocíamos. Lo que realmente resulta interesante es que el resultado anterior motiva una definición de dominancia estocástica, no ya de variables aleatorias, sino de manera general para distribuciones de probabilidad en espacios parcialmente ordenados, y como caso particular, naturalmente, los vectores aleatorios que son de nuestro interés.

Definición 3.1.0.1 (Dominancia Estocástica)**.** *Sea (E,Ω) un espacio medible, que es al mismo tiempo un conjunto parcialmente ordenado y un espacio polaco (i.e. metrizable y topológicamente completo) y sean P_1 y P_2 dos medidas de probabilidad sobre el mismo. Diremos que P_1* domina estocásticamente *a P_2 (y lo notamos P_1 SD_1 P_2) si para toda función medible $\phi : E \to \mathbb{R}$ que sea creciente y acotada se cumple*

$$\int \phi dP_1 \geq \int \phi dP_2$$

□

Por otra parte, el concepto de *acoplamiento* que mencionamos más arriba también puede generalizarse a medidas de probabilidad sobre espacios arbitrarios: dadas dos medidas de probabilidad P_1 y P_2 sobre un espacio medible (E,Ω), un

[1] Thorisson [220], Yohai [241]

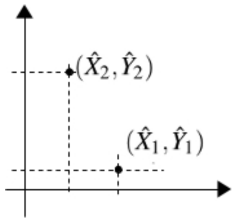

Figura 3.1: Los puntos son incomparables para $\preceq_1$. Por otro lado, nótese que $(\hat{X}_2,\hat{Y}_2) \preceq_2 (\hat{X}_1,\hat{Y}_1)$ pero $\hat{Y}_2 \geq \hat{Y}_1$,.

acoplamiento de las mismas es una medida de probabilidad $\hat{P}$ sobre (E^2,Ω^2) tal que aquellas resultan de las proyecciones canónicas de la última, es decir tal que $P_1 = \hat{P}\pi_1^{-1}$ y $P_2 = \hat{P}\pi_2^{-1}$.

El resultado que generaliza el caso univariado al más general posible se conoce como *Teorema de Strassen* y es uno de los fundamentos básicos de la teoría de acoplamientos[2] (Lindvall [130]).

Teorema 3.1.0.1 (Strassen). *Sea (E,Ω) un espacio medible, que es al mismo tiempo un conjunto parcialmente ordenado (mediante $\preceq$) y un espacio polaco, y sea $M = \{(x_1,x_2) \in E^2, x_1 \preceq x_2\}$. Sean P_1 y P_2 dos medidas de probabilidad sobre (E,Ω), tales que P_1 SD_1 P_2. Entonces existe un acoplamiento $\hat{P}$ de P_1 y P_2 tal que $\hat{P}(M) = 1$.* □

El caso de la dominancia estocástica univariada está claramente implicado por el Teorema 3.1.0.1, pues las variables aleatorias dan medidas de probabilidad sobre $(\mathbb{R}, \mathscr{B})$ y las *proyecciones* del acoplamiento (como medidas de probabilidad) inducen variables aleatorias sobre $\mathbb{R}$, que se ordenan puntualmente. Es decir, que el Teorema de Hadar y Russell para SD_1 es consecuencia de Strassen.

Al pasar al caso multivariado (o a cualquier caso abstracto más general), sin embargo, no es tan directa la deducción de condiciones *suficientes* que resulten en desigualdades fácilmente interpretables estadísticamente. Lo que *sí* es posible encontrar de manera directa son condiciones *necesarias*, aplicando el Teorema de Strassen.

Para ello, consideremos el *orden parcial* dado en $\mathbb{R}^2$ por la relación: $(x_1,y_1) \preceq_1 (x_2,y_2)$ siempre que $x_1 \geq x_2$ y $y_1 \geq y_2$, es decir, cuando el primer par domina al segundo en ambas coordenadas al mismo tiempo. Nótese que este orden es *parcial*, en sentido estricto, porque claramente hay puntos incomparables entre sí (Fig. 3.1), pero ello es lo único que hace falta según el Teo.3.1.0.1.

[2]La demostración del Teorema de Strassen radica en una clásica aplicación del Teorema de Hahn-Banach, con algunas sutilezas técnicas (cfr. Lindvall [130]). Las aplicaciones, por supuesto, van más allá de las que presentamos aquí (cfr. e.g. Dudley [63]), donde tratamos principalmente la dominancia estocástica.

A partir de ello, supongamos que tenemos (X_1,Y_1) (a la que asociamos la medida de probabilidad P_1 en $\mathbb{R}^2$) y (X_2,Y_2) (análogamente P_2) $P_1 SD_1 P_2$. Por Strassen (Teo.3.1.0.1) existe un acoplamiento $\hat{P}$ tal que en $\mathbb{R}^4$ se cumple:

$$P(\{(x,y,z,w) : (z,w) \preceq_1 (x,y)\}) = 1$$

con P_1 y P_2 como distribuciones marginales. Pero esto implica la existencia de un par de vectores aleatorios $(\hat{X}_1,\hat{Y}_1)$ y $(\hat{X}_2,\hat{Y}_2)$, con las respectivas distribuciones de los originales, tales que, en casi todo punto $\omega \in \Omega$, cumplen $(\hat{X}_2,\hat{Y}_2) \preceq_1 (\hat{X}_1,\hat{Y}_1)$, lo que implica a su vez $\{\hat{X}_1 \geq \hat{X}_2, \hat{Y}_1 \geq \hat{Y}_2\}$ en casi todo punto.

De esto último derivamos inmediatamente que, para cada par $(s,t) \in \mathbb{R}^2$ se cumple $\{X_1 \leq s, Y_1 \leq t\} \subseteq \{X_2 \leq s, Y_2 \leq t\}$, de manera que tomando probabilidades tenemos $F_1(s,t) \leq F_2(s,t)$. Esta condición es exactamente la misma que teníamos en el caso *unidimensional*, pero allí era *suficiente*, además de *necesaria*.

Para entender por qué surge esta asimetría en las condiciones de dominancia al pasar al caso multivariado debemos recordar que para demostrar la suficencia utilizamos concretamente el *acoplamiento cuantil* de las distribuciones que queriamos comparar. ¡Pero sabemos que al pasar al caso *multidimensional* no contamos con un concepto análogo al de los cuantiles de una distribución univariada!

En el caso unidimensional, la dominancia estocástica era equivalente al ordenamiento puntual de las funciones de distribución, pero, si recordamos la demostración dada más arriba, hay en ella un paso *difícil*. Éste es justamente la *suficiencia*, es decir, a partir de la desigualdad puntual de las funciones de distribución, demostrar la dominancia estocástica de las variables.

Para probarlo, usamos allí el acoplamiento cuantil, que se basa en la definición unívoca de $F^{-1}(x)$, aun en el caso no continuo o no inyectivo de $F(t)$, utilizando la monotonía. Ahora bien, en dos dimensiones ($2D$, o en ND en general), no se puede definir de manera unívoca $F^{-1}(x)$ para $x \in [0,1]$.

En principio, $F(s,t)$ puede no ser estrictamente creciente en ninguna de las direcciones, pero esto también podía pasar en $1D$ y es posible solucionarlo. El problema intrínseco del caso ND es el siguiente: consideremos $(s,t) = (s_0,t_0)$ tal que su correspondiente imagen es $F(s_0,t_0) \neq 0$; ahora, sabemos que por ser F una función de distribución, debe cumplir $\lim_{s\to-\infty} F(s,t_0) = 0$ y $\lim_{t\to-\infty} F(s_0,t) = 0$. Pero entonces, si F es continua, cualquier $x \in (0, F(s_0,t_0))$ tiene al menos *dos* preimagenes distintas, es decir, el cuantil x no estará unívocamente definido.

Estamos entonces restringidos por un problema que hemos señalado desde los Cap. 1 y 2, a saber, la inexistencia de un instrumento análogo a la curva de Lorenz para el caso multivariado, que nos impide replicar las condiciones suficientes del caso $1D$ para el caso general ND.

En la Sección 3.2 veremos cómo, si consideramos condiciones adicionales sobre las funciones de prueba para la dominancia estocástica (i.e. las funciones de utilidad individual, o en general las ϕ de la Def.3.1) entonces es posible encontrar condiciones *suficientes* para la dominancia estocástica de los vectores, a partir de comparar las funciones de distribución puntualmente (o integradas para dominancias de orden superior), destacándose el efecto contrapuesto de las contribuciones de las distribuciones marginales (las univaridas) y de las distribuciones conjuntas en el interior del dominio.

Por otra parte, podría pensarse que la imposibilidad de definir cuantiles multidimensionales puede derivarse del hecho de que el orden elegido para R^2 es estrictamente *parcial*, dejando puntos incomparables. Para ver que este no es el caso, podemos analizar el siguiente ejemplo.

Si consideramos un caso simple bivariado, digamos $\mathbb{R}^2$ con el orden del diccionario, entonces la condición de dominancia estocástica de un vector (X_1,Y_1) sobre otro vector (X_2,Y_2), implicaría la dominancia puntual en el orden del diccionario de un acoplamiento $(\hat{X}_1,\hat{Y}_1)$ sobre $(\hat{X}_2,\hat{Y}_2)$. Sin embargo, esta desigualdad no implica que para cada (s,t) de $\mathbb{R}^2$ se cumpla $F_1(s,t) \leq F_2(s,t)$ para las respectivas funciones de distribución.

Pues, dado que con probabilidad 1 se tiene $(\hat{X}_2,\hat{Y}_2) \preceq_2 (\hat{X}_1,\hat{Y}_1)$ en el orden del diccionario, las desigualdades $\{\hat{X}_1 \leq s, \hat{Y}_1 \leq t\}$ implican que $\hat{X}_2 \leq s$ casi seguro, pero no necesariamente $\hat{Y}_2 \leq t$ (ver Figura 3.1). De manera que $F_2(s,t) = P(\hat{X}_2 \leq s, \hat{Y}_2 \leq t)$ puede no ser mayor o igual que $F_1(s,t) = P(\hat{X}_1 \leq s, \hat{Y}_1 \leq t)$.

Lo que sí queda claro de ambos ejemplos es que la dominancia estocástica bivariada implica las dominancias unidimensionales, pues como vimos, las distribuciones marginales se ordenan como en el caso univariado. Pero ello las convierte en condiciones *necesarias*, no *suficientes*, pues del mero cumplimiento ellas no se sigue que las distribuciones bivariadas se ordenen de la misma forma.

De una parte es claro que los ordenamientos dados por cada dimensión pueden ser contradictorios, lo cual se relaciona de manera directa con la mencionada inexistencia del cuantil multidimensional.

Pero aun más importante, como hemos señalado (cfr. Cap. 1 y 2) la *correlación* entre las variables podría generar un efecto compensatorio, de manera que, sea porque la *distribución* conjunta compensa las diferencias marginales, o bien porque la *función de utilidad* bivariada modera el efecto de dichas diferencias (via las utilidades marginales diferentes en cada dimensión), el resultado total es un orden de *bienestar* distinto al que presenta el simple estudio de las desigualdades univaridas.

En este sentido, estos efectos son los que justifican de manera más acabada la necesidad de estudiar la desigualdad económica multidimensional en base al análisis de las distribuciones multivariadas, sin reducirla al caso unidimensional mediante algún índice o variable resumen, por más que estos sean de cierta utilidad en algunos procesos de clasificación y estratificación.

3.2. Condiciones de Modularidad y Efectos de Borde

3.2.1. Clases de Modularidad y Teorema de Hadar y Russell

Desde el punto de vista de las aplicaciones prácticas debemos tener en cuenta que, a los efectos de testear dominancia estocástica en términos estadísticos, es de vital importancia contar con condiciones suficientes.

Si tenemos dos vectores aleatorios (X_1,Y_1) y (X_2,Y_2), con funciones de distribución respectivas $F_1(s,t)$ y $F_2(s,t)$ hemos visto en la Sec.3.1 que la condición $F_1(s,t) \leq F_2(s,t)$ puntualmente es *necesaria* pero no *suficiente* para que el primer vector domine estocásticamente al segundo. La condición es necesaria como consecuencia directa del Teorema 3.1.0.1, pero no es suficente, al menos no en la misma forma que lo era para el caso unidimensional por la simple aplicación del acoplamiento cuantil.

Para obtener condiciones suficientes debemos restringir el espacio de funciones de prueba sobre el cual consideramos la dominancia de las distribuciones. Recordemos de la Sec.1.2.2 que las condiciones que imponemos sobre las funciones de prueba (funciones de utilidad reales cuyo argumento es el ingreso) venían dadas en términos de las sucesivas derivadas: $\phi' > 0$ (crecientes), $\phi'' < 0$ (cóncavas)... y siguiendo con la alternancia de signo para los sucesivos órdenes de dominancia.

Estas condiciones, si se quiere, se presentan como suficientes de manera evidente al comparar las integrales $E(\phi(X))$ y $E(\phi(Y))$ mediante la fórmula de integración por partes. Cuando pasamos al caso bivariado, la fórmula de integración por partes nos obliga a tomar en cuenta, ya a primer orden, derivadas de segundo orden para las funciones de prueba $\phi(x,y)$, en particular las derivadas cruzadas. En base a los signos de dichas derivadas, las condiciones suficientes para la dominancia estocástica surgirán del balance de los efectos del interior y del borde en la integral.

La siguiente definición presenta las diferentes clases de funciones para las cuales es posible encontrar condiciones suficientes de dominancia estocástica en base a desigualdades sobre las funciones de distribución bivariadas y/o las distribuciones marginales. Dados dos puntos $x = (x_1,...,x_n), \tilde{x} = (\tilde{x}_1,...,\tilde{x}_n) \in \mathbb{R}^n$, definimos[3] otros dos puntos de $\mathbb{R}^n$, llamados *juntura*: (*join*) $x \vee \tilde{x} = (max\{x_1,\tilde{x}_1\},...,max\{x_n,\tilde{x}_n\})$; y *encuentro* (*meet*): $x \wedge \tilde{x} = (min\{x_1,\tilde{x}_1\},...,min\{x_n,\tilde{x}_n\})$.

Definición 3.2.1.1 (Clases de Modularidad)**.** *Sea $U \subseteq \mathbb{R}^n$ y sea $f : U \to \mathbb{R}$. Decimos que f es supermodular si*

$$f(x) + f(\tilde{x}) \leq f(x \wedge \tilde{x}) + f(x \vee \tilde{x})$$

Decimos que f es submodular *si $-f$ es supermodular.*

Definimos la clase $\mathscr{I}^+$ de funciones $f : U \to \mathbb{R}$ como aquella conformada por las f que son crecientes en cada variable (recordar que $U \subseteq \mathbb{R}^n$) y además

[3] En la definición de las clases de modularidad y la caracterización seguimos a Topkis [222].

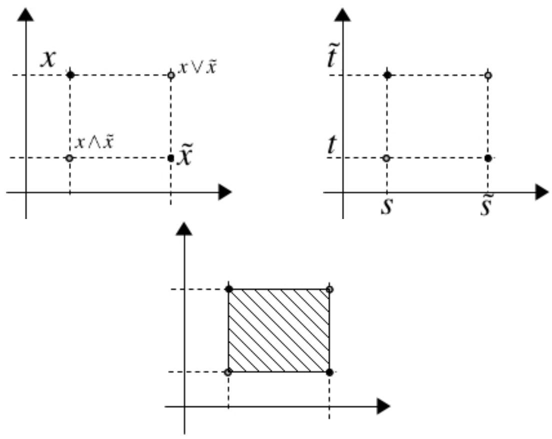

Figura 3.2: Las funciones supermodulares presentan *diferencias crecientes*, tal como las funciones de distribución de vectores aleatorios y las cópulas. Esta propiedad nos permite medir *áreas* con sus diferencias.

son supermodulares. La clase $\mathscr{I}^-$ será la de las funciones, crecientes en cada variable, pero que son submodulares. □

La definición de las clases de modularidad es, a primera vista, poco intuitiva. Para darnos una idea de qué significa la supermodularidad, podemos mirar el caso más simple, en $\mathbb{R}^2$. Notemos que en este caso, si $x \preceq_1 \tilde{x}$ (en la notación de la Sec.3.1), entonces resulta que $x \wedge \tilde{x} = x$, $x \vee \tilde{x} = \tilde{x}$, y la desigualdad es trivialmente satisfecha. Lo mismo en el caso $\tilde{x} \preceq_1 x$. Pero en el caso en que los puntos no son comparables (Fig. 3.2) tenemos una desigualdad no trivial.

Concretamente, la condición de supermodularidad en este caso implica que la función es *creciente en diferencias*: para cada $t \leq \tilde{t}$ (cfr. Fig.3.2), la función *diferencia* $D(s) = f(s,\tilde{t}) - f(s,t)$ es creciente en s. Esto implica que para $s \leq \tilde{s}$ debe ser $D(s) \leq D(\tilde{s})$, es decir que $f(s,\tilde{t}) - f(s,t) \leq f(\tilde{s},\tilde{t}) - f(\tilde{s},t)$, y reordenando los términos: $f(\tilde{s},\tilde{t}) + f(t,s) - f(s,\tilde{t}) - f(\tilde{s},t) \geq 0$.

Podemos recordar que esta propiedad la tienen por ejemplo las funciones de distribución de vectores aleatorios (Yohai [241], Cap.4), lo cual nos permite a partir de ellas medir *áreas*, o más específicamente probabilidades, de subconjuntos de $\mathbb{R}^2$ (Fig.3.2). Las *cópulas* (Nelsen [156]) también cumplen esta propiedad, que se conoce a veces como condición de *2-crecientes*.

Puede probarse (Topkis [222]) que las funciones $f : \mathbb{R}^n \to \mathbb{R}$ son supermodulares si y sólo si tiene diferencias crecientes en cada par de variables (x_i, x_j). Para nosotros, sin embargo, es más relevante la siguiente caracterización, basada en las derivadas de la función, que si bien requiere diferenciabilidad, nos alcanza para estudiar las condiciones suficientes de dominancia estocástica.

Lema 3.2.1.1 (Caracterización de las Clases de Modularidad)**.** *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función diferenciable. Entonces f es supermodular si y sólo si $\frac{\partial f}{\partial x_i}(x)$ es no decreciente en x_j para cada par de índices $i \neq j$. Si f es de clase $\mathscr{C}^2$ entonces f es supermodular si y sólo si $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \geq 0$ para cada par de índices $i \neq j$ y cada $x \in \mathbb{R}^n$, y análogamente, será submodular si $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \leq 0$.*

Demostración. Veamos el caso de $\mathbb{R}^2$ que es sencillo, el caso general se sigue de manera inmediata simplemente utilizando índices. Si f es supermodular, entonces para cada $x_2 > x_1$, $y_2 > y_1$, se cumple que $f(x_2,y_2) + f(x_1,y_1) - f(x_1,y_2) - f(x_2,y_1) \geq 0$. Entonces, reordenando, vemos que $[f(x_2,y_2) - f(x_2,y_1)] - [f(x_1,y_2) - f(x_1,y_1)] \geq 0$ y dividiendo por $(y_2 - y_1)$ y tomando límite con $y_2 \to y_1$ resulta $\frac{\partial f}{\partial y}(x_2,y_1) - \frac{\partial f}{\partial y}(x_1,y_1) \geq 0$, de manera que $\frac{\partial f}{\partial y}(x_2,y_1) \geq \frac{\partial f}{\partial y}(x_1,y_1)$. Análogamente para la derivada en x.

Para la vuelta, basta usar el Teorema del Valor Medio de Lagrange y el hecho de que cada derivada parcial es creciente en la otra variable (cfr. Simchi [208] Ch.2, Topkis [222] Ch.2). Nótese que el caso en que la f es $\mathscr{C}^2$ es mucho más fácil de probar, simplemente con el cociente incremental. $\square$

La condición de supermodularidad (o submodularidad) puede en principio parecer demasiado exigente. Pero debemos tener en cuenta que al estudiar dominancia estocástica las funciones de prueba ϕ son en general *funciones de utilidad*. Entre los tipos básicos, por ejemplo, las funciones $u : \mathbb{R}^n_+ \to \mathbb{R}$ del tipo Cobb-Douglas, $u(x_1,...,x_n) = x_1^{\alpha_1}...x_n^{\alpha_n}$ (donde $\alpha_i \geq 0$) son supermodulares, así como otras clases típicas de la microeconomía (Simchi et al. [208]).

De manera general, puede relacionarse la condición de supermodularidad con la *complementariedad* de los bienes involucrados como input de la función de utilidad (Topkis [222], Ch.2.6). En este sentido, cuando dos bienes (i, j) cumplen para una función de utilidad u, la condición $\frac{\partial^2 u}{\partial x_i \partial x_j} \geq 0$, se dice que los mismos son *complementos estratégicos*. Es decir, los bienes estratégicamente complementarios son aquellos tales que la utilidad marginal de uno de ellos *crece* al aumentar el consumo del otro[4].

En el caso que nos interesa más concretamente, las funciones de utilidad individual $u(y,h)$ tienen como argumentos dos tipos de *bienes*, a saber, el ingreso (y) y el capital humano (h). En este caso, la relación de compementariedad estratégica es una suposición bastante sensata, según hemos discutido en el Cap.2. Por ello, la condición de supermodularidad $\frac{\partial^2 u}{\partial y \partial h} \geq 0$ está debidamente justificada en nuestras aplicaciones.

El siguiente teorema, debido a Hadar y Russell [96], nos permite sin embargo, encontrar condiciones suficientes de dominancia estocástica tanto para el caso

[4]La palabra *complementos*, o bienes *complementarios* a secas suele usarse en microeconomía para aquellos pares tales que la *demanda* de uno de ellos es decreciente en el *precio* del otro (Mas Colell et al. [146]). Los conceptos evidentemente están relacionados, pero al no depender de la decisión *restringida* por el presupuesto, la complementariedad estratégica se juzga más fundamental.

super como submodular. Presentamos el enunciado del mismo para el caso bivariado, siguiendo a Atkinson y Bourguignon [14], con una demostración en la que introducimos algunas modificaciones propias, tales como la aplicación del Teorema de Integración por partes en $\mathbb{R}^n$, que nos permite lograr una prueba más simple y elegante. El resultado se generaliza al caso $N-$dimensional de manera directa (tal como lo presentaron Hadar y Russell originalmente).

Recordemos que para elementos aleatorios Z y W sobre $\mathbb{R}^n$, decimos que Z domina estocásticamente a W a primer orden si $E(\phi(Z)) \geq E(\phi(W))\ \ \forall \phi$ creciente y acotada.

Teorema 3.2.1.1 (Hadar y Russell (1974))**.** *Sean (X_1,Y_1) y (X_2,Y_2) vectores aleatorios con funciones de distribución F_1 y F_2 respectivamente. Supongamos además que las distribuciones son absolutamente continuas respecto a la medida de Lebesgue y que tienen soporte común con clausura compacta $\bar{U} \subseteq \mathbb{R}^2$. Entonces:*

1. *Si $F_1(s,t) \leq F_2(s,t)$ en todo $(s,t) \in U$, entonces se cumple que $E(\phi(X_1,Y_1)) \geq E(\phi(X_1,Y_1))$ para toda $\phi \in \mathscr{I}^-$ de clase $\mathscr{C}^2$ y que es creciente en ambos argumentos.*

2. *Llamemos F_1^X a la distribución marginal de X_1 en el primer vector, F_1^Y a la de Y_1, y de manera análoga F_2^X y F_2^Y para el segundo vector aleatorio. Definamos también $K_1(s,t) = F_1^X(s) + F_1^Y(t) - F_1(s,t)$ y de manera análoga $K_2(s,t)$ en U.*

   Si suponemos que $F_1^X(s) \leq F_2^X(s)$, $F_1^Y(t) \leq F_2^Y(t)$ y $K_1(s,t) \leq K_2(s,t)$ en todo $(s,t) \in U$, entonces se cumple que $E(\phi(X_1,Y_1)) \geq E(\phi(X_1,Y_1))$ para toda $\phi \in \mathscr{I}^+$ de clase $\mathscr{C}^2$ y que es creciente en ambos argumentos.

Demostración. Dado que (X_1,Y_1) tiene distribución absolutamente continua, existe una función de densidad $f_1(s,t)$ tal que en casi todo punto de U se tiene $\frac{\partial^2 F_1}{\partial x \partial y}(s,t) = f_1(s,t)$. Para calcular la esperanza $E(\phi(X_1,Y_1))$ simplemente computamos la integral doble sobre $U \subseteq \mathbb{R}^2$:

$$E(\phi(X_1,Y_1)) = \int_U \phi(s,t) f_1(s,t)dsdt \tag{3.1}$$

Recordemos por otra parte la Fórmula de Integración por Partes en $\mathbb{R}^n$ (Evans [69], App. C) nos dice, para un par de funciones $u, v \in \mathscr{C}^1(\bar{U})$, donde $U \subseteq \mathbb{R}^n$ es un abierto acotado, se cumple:

$$\int_U \frac{\partial u}{\partial x_i} v dx = -\int_U u \frac{\partial v}{\partial x_i} dx + \int_{\partial U} uv\nu^i dS$$

donde ∂U es el borde de U orientado con la normal exterior $\vec{\nu}$, y $\nu^i = \vec{\nu} \cdot \hat{e}_i$ es la componente en la dirección x_i de dicha normal. En nuestro caso, U es un abierto acotado de $\mathbb{R}^2$, de manera que sin pérdida de generalidad podemos suponer $\bar{U} = [0,1] \times [0,1]$, y la normal exterior viene dada justamente en cada dirección por el versor canónico $\hat{e}_i$ (Fig.3.3)

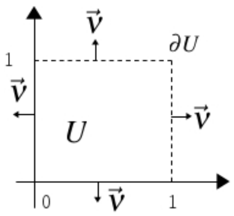

Figura 3.3: En $\mathbb{R}^2$ el borde de U es una curva.

Tenemos entonces:

$$E(\phi(X_1,Y_1)) = \int_U \phi(s,t)\frac{\partial^2 F_1}{\partial x \partial y}(s,t)dsdt =$$
$$= \underbrace{-\int_U \frac{\partial \phi}{\partial x}(s,t)\frac{\partial F_1}{\partial y}(s,t)dsdt}_{I_1} + \underbrace{\int_{\partial U}\phi(s,t)\frac{\partial F_1}{\partial y}(s,t)v^x dS}_{I_2} \quad (3.2)$$

y podemos volver aplicar nuevamente la fórmula de integración por partes para calcular cada uno de esos términos:

$$I_1 = \underbrace{\int_U \frac{\partial^2 \phi}{\partial x \partial y}F_1(s,t)dsdt}_{I_3} - \underbrace{\int_{\partial U}\frac{\partial \phi}{\partial x}F_1(s,t)v^y dS}_{I_4}$$

y teniendo en cuenta la orientación de $\vec{v}$ a lo largo de ∂U, así como también que $F_1(0,t) = F_1(s,0) = 0$:

$$I_2 = \int_{\partial U}\phi\frac{\partial F_1}{\partial y}(s,t)v^x ds =$$
$$= \int_0^1 \phi(1,t)\frac{\partial F_1}{\partial y}(1,t)dt - \int_1^0 \phi(0,t)\frac{\partial F_1}{\partial y}(0,t)dt = \int_0^1 \phi(1,t)\frac{\partial F_1}{\partial y}(1,t)dt =$$
$$= -\int_0^1 \frac{\partial \phi}{\partial y}(1,t)F_1(1,t)dt + \phi(1,t)F_1(1,t)|_0^1 =$$
$$= -\int_0^1 \frac{\partial \phi}{\partial y}(1,t)F_1^y(t)dt + \phi(1,1) \quad (3.3)$$

Por otra parte, haciendo uso de las mismas herramientas y resultados computamos I_4:

$$I_4 = -\int_0^1 \frac{\partial \phi}{\partial x}F_1(s,1)ds = -\int_0^1 \frac{\partial \phi}{\partial x}F_1^X(s)ds$$

de manera que recopilando todos los cálculos resulta:

$$E(\phi(X_1,Y_1)) = \underbrace{\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1(s,t)dsdt}_{A_1} \overbrace{- \int_0^1 \frac{\partial\phi}{\partial y}(1,t)F_1^Y(t)dt}^{B_1} +$$

$$+\phi(1,1) \underbrace{- \int_0^1 \frac{\partial\phi}{\partial x}F_1^X(s)ds}_{C_1} = \phi(1,1) + A_1 + B_1 + C_1 \quad (3.4)$$

De manera totalmente análoga llegamos al resultado:

$$E(\phi(X_2,Y_2)) = \underbrace{\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_2(s,t)dsdt}_{A_2} \overbrace{- \int_0^1 \frac{\partial\phi}{\partial y}(1,t)F_2^Y(t)dt}^{B_2} +$$

$$+\phi(1,1) \underbrace{- \int_0^1 \frac{\partial\phi}{\partial x}F_2^X(s)ds}_{C_2} = \phi(1,1) + A_2 + B_2 + C_2 \quad (3.5)$$

Ahora si $\phi(s,t) \in \mathscr{I}^-$ es de clase $\mathscr{C}^2$, entonces según hemos visto más arriba en el Lema 3.2.1.1, se debe cumplir $\forall (s,t) \in U: \ \frac{\partial^2\phi}{\partial x\partial y}(s,t) \leq 0$.

Supongamos entonces que para casi todo $(s,t) \in U$ se cumple $F_1(s,t) \leq F_2(s,t)$. Esto implica en primer lugar, dado que $\frac{\partial^2\phi}{\partial x\partial y}(s,t) \leq 0$, que la desigualdad se invierte, es decir: $\frac{\partial^2\phi}{\partial x\partial y}(s,t)F_2(s,t) \leq \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1(s,t)$, de manera que $A_2 \leq A_1$.

Por otro lado, tomando límite con t fijo y $s \to 1$, la desigualdad $F_1(s,t) \leq F_2(s,t)$ se convierte en $F_1^Y(t) \leq F_2^Y(t)$, y con s fijo y $t \to 1$ en $F_1^X(s) \leq F_2^X(s)$. Dado que la función $\phi(s,t)$ es creciente en ambos argumentos, entonces $\frac{\partial\phi}{\partial x} \geq 0$ y $\frac{\partial\phi}{\partial y} \geq 0$, de manera que las desigualdades se mantienen al multiplicar, es decir $\frac{\partial\phi}{\partial y}F_1^Y(s) \leq \frac{\partial\phi}{\partial y}F_2^Y(s)$ y $\frac{\partial\phi}{\partial y}F_1^Y(s) \leq \frac{\partial\phi}{\partial y}F_2^Y(s)$, y lo mismo al integrar, pero por último el factor -1 invierte las desigualdades y resulta $B_1 \geq B_2$ y $C_1 \geq C_2$.

A partir de dichas desigualdades, simplemente sumando, se tiene que la condición $F_1(s,t) \leq F_2(s,t)$ es **suficiente** para asegurar, dada cualquier $\phi \in \mathscr{I}^-$ de clase $\mathscr{C}^2$ y que es creciente en ambos argumentos:

$$E(\phi(X_1,Y_1)) \geq E(\phi(X_2,Y_2))$$

es decir, hemos probado la parte (1) del Teorema.

Para probar la parte (2) notemos primero que para cada $t \in [0,1]$ se tiene $\frac{\partial\phi}{\partial y}(1,t) = \int_0^1 \frac{\partial^2\phi}{\partial x\partial y}(s,t)ds + \frac{\partial\phi}{\partial y}(0,t)$, con lo cual podemos reescribir B_1 como:

$$B_1 = -\int_0^1 \frac{\partial\phi}{\partial y}(1,t)F_1(1,t)dt = -\int_0^1 \left(\int_0^1 \frac{\partial^2\phi}{\partial x\partial y}(s,t)ds + \frac{\partial\phi}{\partial y}(0,t)\right)F_1(1,t)dt =$$
$$= -\int_0^1\int_0^1 \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1^Y(t)dsdt - \int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt =$$
$$= -\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1^Y(t)dsdt - \int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt \quad (3.6)$$

De manera similar, se llega a la expresión:

$$C_1 = -\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1^X(s)dtds - \int_0^1 \frac{\partial\phi}{\partial x}(s,o)F_1^X(s)ds$$

y entonces:

$$E(\phi(X_1,Y_1)) = \int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)\left(F_1(s,t) - F_1^X(s) - F_1^Y(t)\right)dsdt -$$
$$-\int_0^1 \frac{\partial\phi}{\partial x}(s,0)F_1^X(s)ds - \int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt =$$
$$= -\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)K_1(s,t)dsdt - \int_0^1 \frac{\partial\phi}{\partial x}(s,0)F_1^X(s)ds - \int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt = \quad (3.7)$$

Por un lado, observemos que, como ϕ es creciente en cada variable, sus derivadas parciales primeras son positivas, de manera que, por la hipótesis $F_1^X(s) \leq F_2^X(s)$, se sigue que $-\int_0^1 \frac{\partial\phi}{\partial x}(s,0)F_1^X(s)ds \geq -\int_0^1 \frac{\partial\phi}{\partial x}(s,0)F_2^X(s)ds$, y lo mismo $-\int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt \geq -\int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_2^Y(t)dt$.

Pero por otro lado, como estamos en el caso supermodular, es $\frac{\partial^2\phi}{\partial x\partial y}(s,t) \geq 0$, de manera que la condición $K_1(s,t) \leq K_2(s,t)$ implica $E(\phi(X_1,Y_1)) \geq E(\phi(X_2,Y_2))$, es decir, hemos probado la parte (2) del Teorema.

□

Al analizar el significado de las condiciones de suficiencia, puede llamarnos la atención la *asimetría* que presentan las condiciones de dominancia estocástica en el caso sub y supermodular, en al menos dos aspectos. Por un lado, si miramos las expresiones involucradas al comparar puntualmente:

- Caso Submodular $\Rightarrow$ comparar $F(s,t)$
- Caso Supermodular $\Rightarrow$ comparar $K(s,t) = F^X(s) + F^Y(t) - F(s,t)$

vemos de manera inmediata que el término dado por la función de distribución conjunta $F(s,t)$ cambia de signo. Esto se debe, matemáticamente, a que en el término integral del interior de U, esto es A_1 en (3.4), el signo de la derivada segunda

$\frac{\partial^2\phi}{\partial x\partial y}(s,t)$ se invierte. Por otro lado, tal vez más importante, en el segundo caso pasan a ser totalmente relevantes los *efectos del borde*, es decir, la contribución de las distribuciones marginales de X e Y.

Esto puede interpretar económicamente en el sentido que venimos señalando desde los Capítulos 1 y 2 y que ahora podemos poner en otros términos: al existir complementariedades estratégicas entre los argumentos de la función de bienestar, el efecto de la desigualdad *univariada* en cada variable por separado (distribución marginal) puede verse compensado por el *efecto conjunto* de ambos.

Es decir, la *interacción* entre las distintas dimensiones, cuando se cumple la complementariedad (supermodularidad), tiene un efecto *contrario* al efecto de los bordes, mientras que en el caso submodular, estos últimos tienen el mismo sentido que la contribución conjunta y por ello no hacía falta tenerlos en cuenta en la comparación.

Es importante interpretar estas condiciones **suficientes** desde el punto de vista de la **economía del bienestar**: lo que estamos viendo, en el caso de las complementariedades estratégicas, es cómo una menor desigualdad en *cada* variable no es suficiente para asegurar un mayor bienestar en una economía, pues *el efecto* ***combinado*** *de ambos argumentos* sobre la utilidad total puede compensar el efecto de aquella desigualdad marginal.

Por ello, a la hora de estudiar la desigualdad económica, debemos considerar las distribuciones *conjuntas* de las variables, y nunca reducir el análisis a la estratificación unidimensional, pues la *interacción* entre las variables tiene un posible impacto mayor sobre el bienestar de una economía que cada variable por separado.

3.2.2. Generalización del Teorema de Hadar y Russell

En lo que sigue, presentaremos un versión más general del Teorema de Hadar y Russell[5], en cuya demostración se hará visible cómo contribuyen los efectos del borde a la integral total, destacándose la diferencia entre las distintas clases de modularidad. Por otra parte, nos permitirá aplicar el teorema cuando las funciones de distribución conjuntas subyacentes no sean absolutamente continuas, de hecho, nisiquiera se requiere la continuidad[6].

Teorema 3.2.2.1 (Generalización de Hadar y Russell)**.** *El resultado del Teorema de Hadar y Russell (Teo. 3.2.1.1) es válido en general para cualquier par de vectores con distribuciones F_1 y F_2 de soporte compacto.*

[5]Tal como lo hemos enunciado, el Teorema de Hadar y Russell 3.2.1.1 incluye resultados de Levy y Paroush [129] y Russell y Seo [184]. En este sentido estamos generalizando dichos trabajos también.

[6]Notemos que, a la hora de *testear estadísticamente* las hipótesis planteadas por las condiciones suficientes de dominancia, sí sería necesario suponer continuidad (para usar Kolmogorov-Smirnov clásico), aunque no necesariamente deberá existir una densidad (Barret y Donald [15], Bickel [25]). La necesidad de continuidad, por su parte, finalmente quedará descartada en nuestro enfoque cuando usemos los resultados de la teoría de procesos empíricos, en particular el bootstrap de dos muestras.

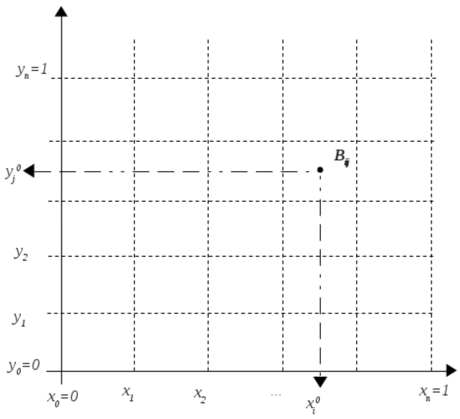

Figura 3.4: Para una suma parcial de Riemann Stieljes, formamos una partición $\{B_{ij} : i, j = 1, \ldots, n\}$ de U, y en cada bloque elemental escogemos un punto (x_i^0, y_j^0).

Demostración. Para empezar, notemos que como las funciones de prueba $\phi(s,t)$ son crecientes y acotadas sobre el soporte de las distribuciones, entonces la esperanza matemática existe, es finita, y viene dada por la integral de Riemman-Stieljes[7]:

$$E(\phi(X_1,Y_1)) = \int_U \phi(s,t)dF_1(s,t) \tag{3.8}$$

Ahora bien, para encontrar un resultado análogo al de Hadar y Russell, uno podría intentar replicar la demostración que hemos dado más arriba, adaptando la fórmula de integración por partes al caso de la integral de Stieljes, pero este procedimiento no es tan directo como para ser fácilmente adaptable a nuestra situación (cfr. Hewitt [105], o Saks [186]).

Sin embargo, es mucho más sencillo e ilustrativo plantear la integración de manera elemental, como el límite de las sumas parciales de Stieljes. Para ello supondremos que el soporte compacto y común de F_1 y F_2 es $U = [0,1] \times [0,1]$ y consideraremos una partición del mismo: $\{0 = x_0 < x_1 < x_2 < \ldots < x_n = 1\} \times \{0 = y_0 < y_1 < \ldots < y_n = 1\}$. Dentro de cada bloque elemental $B_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ elegimos un punto (x_i^0, y_j^0) (cfr. Fig.3.4); como la integral 3.8 existe, podemos tomar siempre el punto medio del rectángulo, sin perder generalidad.

Cada bloque elemental B_{ij} tiene un *cuasi-volumen* $\sigma(B_{ij})$ dado por la variación doble de la función de distribución (Fig. 3.5):

$$\sigma(B_{ij}) = F(x_i,y_j) + F(x_{i-1},y_{j-1}) - F(x_i,y_{j-1}) - F(x_{i-1},y_j)$$

En cada rectángulo B_{ij} de la partición elegimos además un punto (x_i^0, y_j^0) donde

[7] Barry James [109], Cap. 3.

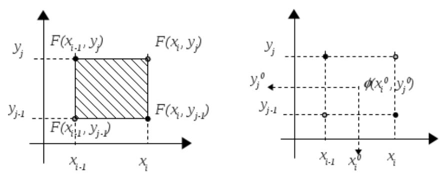

Figura 3.5: El cuasi-volumen de B_{ij} viene dado por la variación doble de F. Para las sumas de Stieljes evaluamos $\phi(x_i^0, y_j^0)$

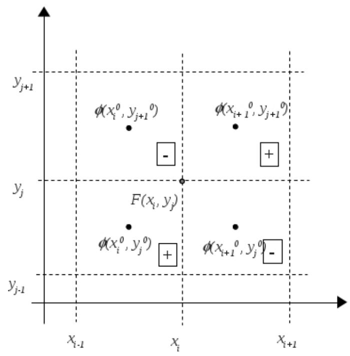

Figura 3.6: Si (x_i, y_j) es un punto interior de la partición, $F(x_i, y_j)$ aparece en 4 términos de la suma, dos veces sumando y dos restando.

evaluamos ϕ (Fig.3.5). Con ello, formamos la suma parcial de Riemann-Stieljes para nuestra partición Π_n y los puntos de evaluación escogidos:

$$S_{\Pi_n}(\phi) = \sum_{i,j=1}^{n} \phi(x_i^0, y_j^0)\sigma(B_{ij}) \tag{3.9}$$

Ahora notemos que, si usamos la definición de $\sigma(B_{ij})$ tenemos:

$$S_{\Pi_n}(\phi) = \sum_{i,j=1}^{n} \phi(x_i^0, y_j^0)(F(x_i, y_j) + F(x_{i-1}, y_{j-1}) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j)) \tag{3.10}$$

Para entender mejor cómo influyen las integrales del borde y del interior de U, observemos que si (x_i, y_j) no es un punto del borde de la partición (o sea $x_i \neq 0; 1$, $y_j \neq 0; 1$), entonces el factor $F(x_i, y_j)$ aparece en 4 términos de la suma, dos veces con signo positvo y dos con signo negativo (Fig.3.6). Esto es así porque al recorrer toda la partición, el punto (x_i, y_j) cumple el rol de los 4 vértices en 4 rectángulos distintos.

Los únicos puntos que aparecen sólo como 2 vertices de rectángulos distintos son los del borde, es decir aquellos de la forma $\{(x_i, 0), (x_i, 1), (0, y_j), (1, y_j)\}$. Ahora bien, como $F(s,t)$ es una función de distribución, tenemos $F(x_i, 0) = F(0, y_j) =$

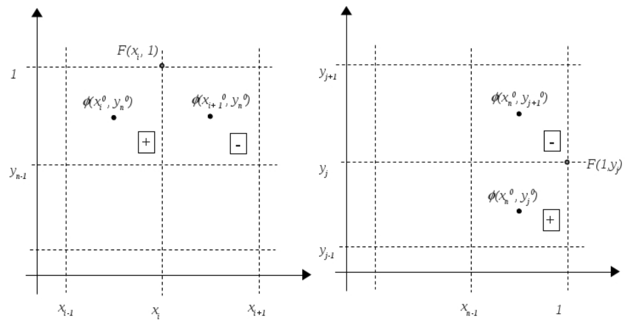

Figura 3.7: Los puntos del borde aportan factores $F(x_i,1)$ y $F(1,y_j)$ que aparecen en 2 términos, una vez con signo positivo y otra negativo.

0. De manera que sólo nos interesa el efecto de los factores $F(x_i,1)$ y $F(1,y_j)$, que aparecen en 2 términos, una vez con signo positivo y otra negativo (Fig.3.7).

Tenemos entonces que para los puntos interiores podemos agrupar los términos que llevan el factor $F(x_i,y_j)$:

$$F(x_i,y_j)*(\phi(x_i^0,y_j^0)+\phi(x_{i-1}^0,y_{j-1}^0)-\phi(x_{i-1}^0,y_j^0)-\phi(x_i^0,y_{j-1}^0)) = \\ = F(x_i,y_j)\Delta_{i,j}\phi \quad (3.11)$$

y para los puntos del borde tenemos:

$$F(x_i,1)*(\phi(x_i^0,y_n^0)-\phi(x_{i+1}^0,y_n^0)) = F(x_i,1)\delta_{i,n}\phi$$
$$F(1,y_j)*(\phi(x_n^0,y_j^0)-\phi(x_n^0,y_{j+1}^0)) = F(1,y_j)\delta_{n,j}\phi$$

Y por último, el factor $F(1,1)=1$ aparece en único término, con lo cual, reemplazando estas expresiones podemos reescribir la fórmula 3.10 como:

$$S_{\Pi_n}(\phi) = \overbrace{\sum_{i,j=1}^{n-1} F(x_i,y_j)\Delta_{i,j}\phi}^{A} + \overbrace{\sum_{i=1}^{n-1} F(x_i,1)\delta_{i,n}\phi}^{B} + \\ + \underbrace{\sum_{j=1}^{n-1} F(1,y_j)\delta_{n,j}\phi}_{C} + \phi(1,1) \quad (3.12)$$

Notemos en principio que, como estamos suponiendo en todos los casos que las funciones $\phi(s,t)$ con crecientes en cada uno de los argumentos, entonces las variaciones que entran en los términos de borde, esto es las $\delta_{n,j}\phi$ y $\delta_{i,n}\phi$ son siempre **negativas**, pues el término más grande está restando[8].

Supongamos ahora que estamos comparando las esperanzas de ϕ para las distribuciones F_1 y F_2. De acuerdo a 3.12 tendremos

$$S^1_{\Pi_n} = A_1 + B_1 + C_1 + \phi(1,1)$$

$$S^2_{\Pi_n} = A_2 + B_2 + C_2 + \phi(1,1)$$

Consideremos que estamos en el caso submodular, es decir $\frac{\partial^2\phi}{\partial x_i \partial x_j}(x) \leq 0$. En este caso, sabemos por lo visto luego de la Def.3.2.1, que $\Delta_{ij}\phi \leq 0$. Supongamos ahora que se cumple la condición $F_1(s,t) \leq F_2(s,t)$, entonces $\Delta_{ij}\phi F_2(x_i,y_j) \leq \Delta_{ij}\phi F_1(x_i,y_j)$, y sumando tenemos $A_2 \leq A_1$. Por otro lado, en este caso tenemos de manera trivial que $F_1(x_i,1) \leq F_2(x_i,1)$ y como $\Delta_{n,j}\phi \leq 0$ se sigue que $\Delta_{n,j}\phi F_2(x_i,1) \leq \Delta_{n,j}\phi F_1(x_i,1)$, y entonces $B_2 \leq B_1$.

De manera totalmente análoga se tiene que $C_2 \leq C_1$ y entonces es claro que $S^2_{\Pi_n} \leq S^1_{\Pi_n}$. Luego, tomando límite con el diámetro de la partición tendiendo a 0, tenemos que:

$$\int_U \phi(s,t)dF_2(s,t) \leq \int_U \phi(s,t)dF_1(s,t) \tag{3.13}$$

es decir $E(\phi(X_2,Y_2)) \leq E(\phi(X_1,Y_1))$.

Para el caso supermodular, nuvamente tendremos que estudiar con un poco más de detalle las contribuciones del borde, es decir, los términos B y C de las sumas de Riemann-Stieljes. Notemos para ello en primer lugar que, dada la notación que usamos para $\Delta_{i,j}$ con $i,j = 1,...,n-1$ por un lado y $\delta_{n,j}$, $\delta_{i,n}$ por otro, tenemos las igualdades:

$$\sum_{i=1}^{n-1} \Delta_{i,j}\phi_{1,j} = \phi_{1,j-1} - \phi_{1,j} + \phi_{n,j} - \phi_{n,j-1} = \phi_{1,j-1} - \phi_{1,j} - \delta_{n,j}\phi$$

$$\Rightarrow \delta_{n,j}\phi = \phi_{1,j-1} - \phi_{1,j} - \sum_{i=1}^{n-1} \Delta_{i,j}\phi_{1,j} = \delta_{1,j}\phi - \sum_{i=1}^{n-1} \Delta_{i,j}\phi_{1,j} \tag{3.14}$$

donde los términos intermedios se fueron cancelando por cómo se solapan los $\Delta_{i,j}\phi$ (Fig. 3.8).

Análogamente, tenemos:

$$\delta_{i,n}\phi = \phi_{i-1,1} - \phi_{i,1} - \sum_{j=1}^{n-1} \Delta_{i,j}\phi_{1,j} = \delta_{i,1}\phi - \sum_{j=1}^{n-1} \Delta_{i,j}\phi_{1,j}$$

[8]Nótese que los δ minúscula, tienen **dos** términos y los Δ tienen **cuatro**, por eso la notación distintiva.

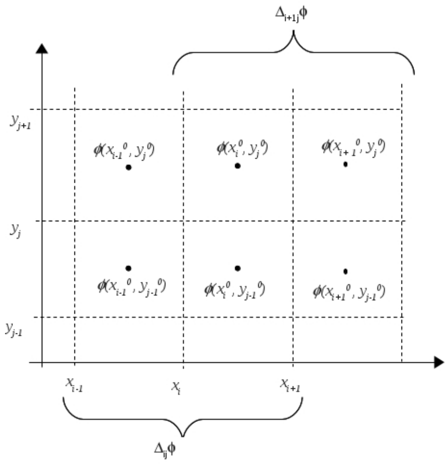

Figura 3.8: Los $\Delta_{i,j}\phi$ se solapan, de manera que la suma resulta *telescópica*.

Ahora reemplazando en B de 3.12 tenemos:

$$B = \sum_{i=1}^{n-1} F(x_i,1)\delta_{i,n}\phi = \sum_{i=1}^{n-1} F(x_i,1)\left(\delta_{i,1}\phi - \sum_{j=1}^{n-1}\Delta_{i,j}\phi_{1,j}\right) =$$

$$= \sum_{i=1}^{n-1} F(x_i,1)\delta_{i,1}\phi - \sum_{i,j=1}^{n-1} F(x_i,1)\Delta_{i,j}\phi \quad (3.15)$$

y de la misma forma para C de 3.12:

$$C = \sum_{j=1}^{n-1} F(1,y_j)\delta_{1,j}\phi - \sum_{i,j=1}^{n-1} F(1,y_j)\Delta_{i,j}\phi$$

Reemplazando todo en 3.12:

$$S_{\Pi_n}(\phi) = \sum_{i,j=1}^{n-1}\left(F(x_i,y_j) - F(x_i,1) - F(1,y_j)\right)\Delta_{i,j}\phi +$$

$$+ \sum_{j=1}^{n-1} F(1,y_j)\delta_{1,j}\phi + \sum_{i=1}^{n-1} F(x_i,1)\delta_{i,1}\phi + \phi(1,1) = -\sum_{i,j=1}^{n-1} K(x_i,y_j)\Delta_{i,j}\phi +$$

$$+ \sum_{j=1}^{n-1} F(1,y_j)\delta_{1,j}\phi + \sum_{i=1}^{n-1} F(x_i,1)\delta_{i,1}\phi + \phi(1,1) \quad (3.16)$$

y ahoro sólo hace falta notar que los $\delta_{1,j}\phi$ y $\delta_{i,1}\phi$ son **negativos** (pues $\phi(s,t)$ es creciente en cada variable) y $\Delta_{i,j}\phi \geq 0$ (pues $\phi(s,t)$ es supermodular). Ahora con ello, y dados los supuestos de que $K_1(s,t) \leq K_2(s,t)$, y $F_1^X(s) = F_1(s,1) \leq F_2(s,1) = F_2^X(s)$, junto con la otra desigualdad marginal, podemos afirmar, para cada n:

$$S^1_{\Pi_n}(\phi) \geq S^2_{\Pi_n}(\phi)$$

y en consecuencia, pasando al límite, tenemos que para cada ϕ supermodular:

$$\int_U \phi(s,t)dF_1(s,t) \geq \int_U \phi(s,t)dF_2(s,t)$$

es decir

$$E(\phi(X_1,Y_1)) \geq E(\phi(X_2,Y_2))$$

que es lo que queríamos demostrar. □

Nótese que con esta demostración se tiene no sólo la validez del teorema para F general (continua o no) sino además, para el caso discreto, una *fórmula*, dada por una buena reinterpretación de las sumas parciales, exácta en el caso finito.

Observemos por último que la clave de toda la complicación está en la *correlación* entre las dimensiones. Si se supusiera independencia, entonces no es difícil probar que las dominancias marginales son suficientes para la bivariada (Scarsini [193])

3.3. Dominancia a Segundo Orden

3.3.1. Clases Superiores de Modularidad

La necesidad de introducir ordenes superiores de dominancia estocástica en el caso multivariado surge del mismo modo que en el caso univariado. Como señalamos en la Sección 1.1, cuando tenemos un concepto operativo de desigualdad económica, podemos utilizarlo para *rankear* distribuciones, es decir, para definir un orden parcial en cierto conjunto de las economías de distribución.

La problemática que enfrenta la dominancia estocástica a este respecto, es la misma en 1 o N dimensiones: las condiciones suficientes dadas en términos de la función de distribución pueden dar resultados ambiguos, pues las desigualdades puntuales para $F(s,t)$ y $K(s,t)$ (cfr. Teo.3.2.1.1) pueden darse en una dirección para ciertos pares (s,t) y en la contraria para otros. En este sentido, la dominancia estocástica a primer orden resulta a veces *inconclusiva* y debe recurrirse a dominancias de orden superior para poder rankear un conjunto dado de distribuciones (cfr. Sec. 1.2.2).

La idea es entonces, al igual que en el caso univariado, *achicar* la clase de funciones de utilidad que entran en la definición de la dominancia, para poder diferenciar más *finamente* las distribuciones que operan sobre ellas. En el caso univariado, agregábamos a las condiciones de monotonía (i.e. conndiciones sobre ϕ') nuevas condiciones de concavidad (i.e. sobre ϕ'').

En el caso multivariado, vimos que ya las condiciones suficientes para dominancia de primer orden, al requerir condiciones de modularidad, imponían el estudio de derivadas de segundo orden (cruzadas). Esto, sumado al hecho de que dicha condición (la submodularidad) se menciona a menudo como *aversión al riesgo* (Scarsini [192]), puede llevarnos a confusión con el caso univariado, donde la dominancia a *segundo* orden tenía que ver con la aversión al riesgo y la derivada segunda.

Pero recordemos que la noción de dominancia a segundo orden tiene que ver con la *concavidad* de las funciones de utilidad. Ahora, si bien la noción de aversión al riesgo es difícil de generalizar de manera obvia al caso $N-dimensional$, la noción de concavidad es simplemente la misma; $\phi : E \to \mathbb{R}$ es cóncava si:

$$\phi(\lambda x+(1-\lambda)y) \geq \lambda\phi(x)+(1-\lambda)\phi(y) \quad \forall\lambda \in [0,1]$$

Con ello, tenemos la siguiente definición que generaliza la noción de dominancia estocástica de segundo orden al caso *ND*.

Definición 3.3.1.1 (Dominancia Estocástica a Segundo Orden)**.** *Sea (E,Ω) un espacio medible, donde E es un espacio vectorial topológico parcialmente ordenado, Ω su álgebra de Borel, y sean P_1 y P_2 dos medidas de probabilidad sobre el mismo. Diremos que P_1* domina estocásticamente a segundo orden *a P_2 (y lo notamos P_1 SD_2 P_2) si para toda función medible $\phi : E \to \mathbb{R}$ que sea creciente, acotada y* *cóncava se cumple:*

$$\int \phi dP_1 \geq \int \phi dP_2$$

□

Un inconveniente básico con la dominancia a segundo orden es que no contamos con algo como el teorema de Strassen de manera directa.[9] Esta situación dificulta la tarea de encontrar condiciones necesarias y suficientes para la dominancia, aun en el caso univariado.

Russell & Seo [184] presentan condiciones necesarias y suficientes para la dominancia a segundo orden, pero el resultado es de poca utilidad práctica. La condición que encuentran, comparando integrales sobre una clase infinita de conjuntos pues no se traduce en desigualdades fáciles de probar en un caso concreto, menos aun de llevar al estudio estadístico.

Como en el caso de la dominancia de primer orden, sin embargo, podemos conseguir condiciones operativas mediante restricciones en los signos de las derivadas de orden superior. Atkinson y Bourguignon [14] restringen las clases de modularidad $\mathscr{I}^-$ y $\mathscr{I}^+$ (cfr. Sec.3.2) obteniendo condiciones suficientes en base a desigualdades de operadores integrales aplicados a las funciones de distribución.

Este resultado, como cabe esperar, es análogo al de la dominancia estocástica de segundo orden en el caso unidimensional (cfr. Sec.1.2.2), pero naturalmente las dimensiones adicionales requieren una nueva serie de condiciones sobre las derivadas cruzadas que surgen como resultado de la aplicación sucesiva de la fórmula multivariada de integración por partes.

Las clases de modularidad de segundo orden introducidas por Atkinson y Bourguignon [14] son las siguientes:

Definición 3.3.1.1 (Clases de Modularidad de Segundo Orden)**.** *Sea $U \subseteq \mathbb{R}^n$ y sea $f : U \to \mathbb{R}$.*

- *Decimos que $f \in \mathscr{I}^{--}$ si $f \in \mathscr{I}^-$ y además se cumplen en U:*

$$\begin{cases} \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y^2} \leq 0 \\ \frac{\partial^3 f}{\partial x^2 \partial y}, \frac{\partial^3 f}{\partial x \partial y^2} \geq 0 \\ \frac{\partial^4 f}{\partial x^2 \partial y^2} \leq 0 \end{cases}$$

- *Decimos que $f \in \mathscr{I}^{++}$ si $f \in \mathscr{I}^+$ y además se cumplen en U:*

$$\begin{cases} \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y^2} \leq 0 \\ \frac{\partial^3 f}{\partial x^2 \partial y}, \frac{\partial^3 f}{\partial x \partial y^2} \leq 0 \\ \frac{\partial^4 f}{\partial x^2 \partial y^2} \geq 0 \end{cases}$$

[9] Esta puede ser una interesante línea de investigación, al menos dentro de la teoría de probabilidades o más específicamente de acoplamientos. Parece ser un problema de resolución posible, pero no dentro del alcance de la presente Tesis.

□

Aunque parecen complicadas, en la demostración del teorema de Atkinson y Bourguignon veremos que estas condiciones surgen de manera evidente para fijar la dirección de las desigualdades cuando uno integra por partes. Por otro lado, en la Sec.4 estudiaremos el significado económico de estas condiciones.

Notemos que en la definición de las clases de modularidad de segundo orden hemos enunciado directamente las condiciones sobre las derivadas superiores de las funciones, antes que sobre la doble diferencia como hacíamos en el caso de orden uno (Def.3.2.1). Se observa inmediatmente, sin embargo, que podríamos haber usado la misma condición, pero aplicada sobre la derivada segunda cruzada de f. Es decir, se podía enunciar pidiendo que $\frac{\partial^2 f}{\partial x \partial y}$ sea (sub)supermodular y las demás desigualdades, que toman en cuenta la monotonía y la concavidad.

Antes de enunciar el teorema, recordemos del Teo.3.2.1.1 que para una función de distribución bivariada $F(x,y)$ teníamos definida una nueva función $K(x_1,x_2) = -[F(x_1,x_2) - F_1(x_1) - F_2(x_2)]$, que era útil en el caso supermodular.

En base a estas dos funciones, definimos los operadores integrales:

$$H(x,y,F) = \int_0^x \int_0^y F(s,t)dtds \tag{3.17}$$

$$L(x,y,F) = \int_0^x \int_0^y K(s,t)dtds \tag{3.18}$$

Si pensamos fija la F, tanto H como L son funciones de (x,y), en este sentido, podemos pensar en operadores que asignan a cada función de distribución F la función de clase $\mathscr{C}^2$ $H(x,y)$ y $L(x,y)$ respectivamente. Ello nos permite mirar las funciones *marginales* derivadas de estos operadores:

$$H^X(x,F) = \int_0^x F^X(s)ds \tag{3.19}$$

$$H^Y(y,F) = \int_0^y F^Y(t)dt \tag{3.20}$$

y análogamente para el operador L.

Teorema 3.3.1.1 (Atkinson y Bourguignon (1982)). *Sean (X_1,Y_1) y (X_2,Y_2) vectores aleatorios con funciones de distribución F_1 y F_2 respectivamente. Supongamos además que las distribuciones son absolutamente continuas respecto a la medida de Lebesgue y que tienen soporte común con clausura compacta $\bar{U} \subseteq \mathbb{R}^2$ (sin perder generalidad, pensamos $U = [0,1] \times [0,1]$).*

Llamemos $H_1(x,y)$ y $H_2(x,y)$ a los respectivos H para cada distribución, y lo mismo $L_1(x,y)$ y $L_2(x,y)$, cada uno con sus correspondientes versiones marginales con la notación recién expuesta.

1. *Supongamos que*

$$\begin{cases} H_1^X(x) \leq H_2^X(x) & \forall x \in [0,1] \\ H_1^Y(y) \leq H_2^Y(y) & \forall y \in [0,1] \end{cases} \tag{3.21}$$

y que $H_1(x,y) \leq H_2(x,y)$ *en* U*. Entonces se cumple que* $E(\phi(X_1,Y_1)) \geq E(\phi(X_1,Y_1))$ *para toda* $\phi \in \mathscr{I}^{--}$ *de clase* $\mathscr{C}^4$ *y que es cóncava y creciente en ambos argumentos.*

2. *Supongamos que*

$$\begin{cases} H_1^X(x) \leq H_2^X(x) & \forall x \in [0,1] \\ H_1^Y(y) \leq H_2^Y(y) & \forall y \in [0,1] \end{cases}$$

y que $L_1(x,y) \leq L_2(x,y)$ *en* U*. Entonces se cumple que* $E(\phi(X_1,Y_1)) \geq E(\phi(X_1,Y_1))$ *para toda* $\phi \in \mathscr{I}^{++}$ *de clase* $\mathscr{C}^4$ *y que es cóncava creciente en ambos argumentos.*

Demostración. La demostración es sencilla porque recae en lo que ya hicimos para el caso de primer orden. Recordemos que teníamos, según la ecuación 3.4

$$E(\phi(X_1,Y_1)) = \underbrace{\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1(s,t)dsdt}_{A_1} \overbrace{-\int_0^1 \frac{\partial\phi}{\partial y}(1,t)F_1^Y(t)dt}^{B_1} +$$

$$+\phi(1,1) \underbrace{-\int_0^1 \frac{\partial\phi}{\partial x}F_1^X(s)ds}_{C_1} = \phi(1,1) + A_1 + B_1 + C_1 \tag{3.22}$$

y análogamente A_2, B_2, C_2. El término $\phi(1,1)$ obviamente no nos interesa porque es el mismo para las dos distribuciones. De los otros, empecemos comparando los B y el caso de los C saldrá de la misma forma.

Integrando por partes tenemos:

$$\begin{aligned} B_1 &= -\int_0^1 \frac{\partial\phi}{\partial y}(1,t)F_1^Y(t)dt = \\ &= -\left[\frac{\partial\phi}{\partial y}(1,1)\int_0^1 F_1^Y(t)dt - \int_0^1 \frac{\partial^2\phi}{\partial y^2}(1,t)\left(\int_0^t F_1^Y(u)du\right)dt\right] = \\ &= \frac{\partial\phi}{\partial y}(1,1)\int_0^1 F_1^Y(t)dt + \int_0^1 \frac{\partial^2\phi}{\partial y^2}(1,t)H_1^Y(t)dt \end{aligned} \tag{3.23}$$

Ahora bien, podemos suponer sin perder generalidad que las ϕ tienen soporte compacto en U (y usar un argumento de densidad), así que el primero término se anula (por la derivada de ϕ en el borde). Y en cuanto al segundo término, como sabemos que $H_1^Y(t) \leq H_2^Y(t)$ para todo t, y además que $\frac{\partial^2\phi}{\partial y^2} \leq 0$, entonces se da vuelta la desigualdad y tenemos $B_1 \geq B_2$. La cuenta para los C es obviamente igual.

Estas desigualdades nos sirven para ambos casos pero para los A tenemos que separar entre ambos. Para el primero, supongamos $\phi \in \mathscr{I}^{--}$ y llamemos $\tilde{\phi}(x,y) = \frac{\partial^2\phi}{\partial x\partial y}(x,y)$. Notemos además que $F(x,y) = \frac{\partial^2 H}{\partial x\partial y}(x,y)$.

Tenemos entonces:

$$\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)F_1(s,t)dsdt = \int_U \tilde{\phi}(s,t)\frac{\partial^2 H}{\partial x\partial y}(s,t)dsdt$$

Pero esta integral es exáctamente de la misma forma que la 3.1 , sólo que ahora tenemos $\tilde{\phi}$ y H en lugar de ϕ y F, por lo cual es evidente que:

$$\int_U \tilde{\phi}(s,t)\frac{\partial^2 H}{\partial x\partial y}(s,t)dsdt = \int_U \frac{\partial^2\tilde{\phi}}{\partial x\partial y}(s,t)H_1(s,t)dsdt - \int_0^1 \frac{\partial\tilde{\phi}}{\partial y}(1,t)H_1^Y(t)dt + \\ + \tilde{\phi}(1,1) - \int_0^1 \frac{\partial\tilde{\phi}}{\partial x}H_1^X(s)ds \quad (3.24)$$

Ahora observemos que en esta expresión entran las derivadas terceras de la ϕ pues, por ejemplo:

$$\frac{\partial\tilde{\phi}}{\partial x} = \frac{\partial^3\phi}{\partial x^2\partial y}$$

y éstas son positivas en el caso $\phi \in \mathscr{I}^{--}$, así que la demostración de este caso sigue de forma totalmente análoga al teorema de Hadar y Russell en el caso $\phi \in \mathscr{I}^{-}$.

Para el caso $\phi \in \mathscr{I}^{++}$, retomemos en la ecuación 3.7:

$$E(\phi(X_1,Y_1)) = -\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)K_1(s,t)dsdt - \\ - \int_0^1 \frac{\partial\phi}{\partial x}(s,0)F_1^X(s)ds - \int_0^1 \frac{\partial\phi}{\partial y}(0,t)F_1^Y(t)dt = \quad (3.25)$$

Por un lado podemos dar por demostrada la desigualdad para los dos últimos términos, pues las condiciones sobre las marginales de $H(x,y)$ son las mismas en este caso que en el anterior. Por otro, el primer término se transforma, usando integración por partes, de manera equivalente a lo que hicimos recién:

$$-\int_U \frac{\partial^2\phi}{\partial x\partial y}(s,t)K_1(s,t)dsdt = -\int_U \tilde{\phi}(s,t)\frac{\partial^2 L}{\partial x\partial y}(s,t)dsdt =$$
$$= -\int_U \frac{\partial^2\tilde{\phi}}{\partial x\partial y}(s,t)L_1(s,t)dsdt + \int_U \frac{\partial\tilde{\phi}}{\partial y}(1,t)L_1^Y(t)dt + \int_U \frac{\partial\tilde{\phi}}{\partial x}(s,1)L_1^X(s)ds \tag{3.26}$$

Por hipótesis es $L_1(x,y) \leq L_2(x,y)$, lo que implica las marginales $L_1^X(x) \leq L_2^Y(x)$ y la otra. Tenemos también, al estar en el caso $\phi \in \mathscr{I}^{++}$, que $\frac{\partial\tilde{\phi}}{\partial x} = \frac{\partial^3\phi}{\partial x^2\partial y} \leq 0$, así que las desigualdades se dan vuelta al multiplicar y quedan igual al integrar. Eso da cuenta de los dos últimos sumandos.

Por otra parte, en el caso $\phi \in \mathscr{I}^{++}$ tenemos que $\frac{\partial^2\tilde{\phi}}{\partial x\partial y} = \frac{\partial^4\phi}{\partial x^2\partial y^2} \geq 0$ y el signo menos de adelante nos da vuelta las desigualdades. Con todo ello, tenemos:

$$E(\phi(X_1,Y_1)) \geq E(\phi(X_2,Y_2))$$

como queriamos demostrar. □

Es fácil ver que si impusieramos otros signos sobre las derivadas terceras y cuarta, llegaríamos a condiciones suficientes de dominancia similares, mediante el mismo procedimiento. Nos quedamos con las clases $\mathscr{I}^{++}$ y $\mathscr{I}^{--}$ por ser las que se consideran relevantes en la literatura (Atkinson & Bourguignon [14]), en base a su interpretación económica. La implementación estadística que presentamos en la Sec.5.3 puede aplicarse a cualquiera de estos casos, con modificaciones elementales.

3.3.2. Generalización del Teorema de Atkinson y Bourguignon

Tal como en el caso de la dominancia a primer orden, salta a la vista que tanto en la definición de dominancia estocástica a orden dos (Def.3.3.1) como en las condiciones suficientes dadas por el Teo.3.3.1.1, entran desigualdades que se pueden plantear para cualquier distribución F, no solamente para aquellas que son absolutamente continuas. La absoluta continuidad (i.e. la existencia de densidad) entra en la *demostración* del teorema, pero esto esencialmente porque la forma más fácil de probar el resultado es mediante la fórmula de partes.

Al no entrar de manera evidente las propiedades de absoluta continuidad en el planteo, cabe esperar que, al igual que en el caso de primer orden, podamos generalizar el resultado a distribuciones arbitrarias. Esto es lo que hacemos en el siguiente teorema, cuya demostración sigue un curso similar al que tomamos en el caso absolutamente continuo: tomamos las igualdades que teníamos en la dominancia de primer orden, y las reescribimos para poder volver a empezar pero ahora pensando en una $\tilde{\phi}$ y una H (o una K) en lugar de una ϕ y una F.

Teorema 3.3.2.1 (Generalización de Atkinson y Bourguignon)**.** *El resultado del Teorema de Atkinson y Bourguignon (Teo. 3.3.1.1) es válido en general para cualquier par de vectores con distribuciones F_1 y F_2 de soporte compacto.*

Demostración. No vamos a dar una demostración pormenorizada como lo hicimos en el caso de primer orden. La misma resulta innecesaria si mostramos cómo, mediante una reescritura de las sumas parciales, estamos en un caso análogo al anterior.

Pensemos en el caso $\phi \in I^{--}$ y retomemos en la suma parcial de Riemann Stieljes 3.12:

$$S_{\Pi_n}(\phi) = \overbrace{\sum_{i,j=1}^{n-1} F(x_i, y_j)\Delta_{i,j}\phi}^{A} + \overbrace{\sum_{i=1}^{n-1} F(x_i, 1)\delta_{i,n}\phi}^{B} + \\ + \underbrace{\sum_{j=1}^{n-1} F(1, y_j)\delta_{n,j}\phi}_{C} + \phi(1,1) \quad (3.27)$$

Empecemos por el término A y recordemos que:

$$\Delta_{i,j}\phi = \phi(x_i^0, y_j^0) + \phi(x_{i-1}^0, y_{j-1}^0) - \phi(x_{i-1}^0, y_j^0) - \phi(x_i^0, y_{j-1}^0)$$

Haciendo uso repetido del teorema del valor medio de Lagrange, podemos reescribir este factor como:

$$\Delta_{i,j}\phi = \frac{\partial^2 \phi}{\partial x \partial y}(\tilde{x}_i^0, \tilde{y}_j^0)\Delta x \Delta y + \frac{\partial^2 \phi}{\partial x^2}(\hat{x}_i^0, \hat{y}_j^0) O(\Delta x)^2$$

para ciertos puntos $(\tilde{x}_i^0, \tilde{y}_j^0)$ y $(\hat{x}_i^0, \hat{y}_j^0)$ en el interior del rectángulo $[x_{i-1}^0, x_i^0] \times [y_{j-1}^0, y_j^0]$ y donde llamamos $\Delta x = x_i^0 - x_{i-1}^0$ y $\Delta y = y_i^0 - y_{i-1}^0$. Introduciendo esto en el factor A tenemos:

$$\sum_{i,j=1}^{n-1} F(x_i, y_j)\Delta_{i,j}\phi = \sum_{i,j=1}^{n-1} F(x_i, y_j)\left(\frac{\partial^2 \phi}{\partial x \partial y}(\tilde{x}_i^0, \tilde{y}_j^0)\Delta x \Delta y + \frac{\partial^2 \phi}{\partial x^2}(\hat{x}_i^0, \hat{y}_j^0) O(\Delta x)^2\right) = \\ = \sum_{i,j=1}^{n-1} F(x_i, y_j)\frac{\partial^2 \phi}{\partial x \partial y}(\tilde{x}_i^0, \tilde{y}_j^0)\Delta x \Delta y + \sum_{i,j=1}^{n-1} F(x_i, y_j)\frac{\partial^2 \phi}{\partial x^2}(\hat{x}_i^0, \hat{y}_j^0) O(\Delta x)^2 \quad (3.28)$$

Ahora bien, el factor cuadrático de la última sumatoria hace que esta se anule cuando el diámetro de la partición tienda a 0, así que éste término no nos preocupa[10].

En cuanto a la primera sumatoria, observemos que:

$$F(x_i, y_j)\Delta x \Delta y \approx \Delta_{i,j} H \quad (3.29)$$

[10]Esto es elemental, pero para convencernos, podemos pensar que la partición es regular, así que los Δx son todos iguales, pudiendo sacar uno como factor común. Como el otro factor es acotado (para y_j fijo tiende a una integral parcial, por Fubini) el producto tiende a cero.

donde la igualdad es exacta si en lugar de H tomamos las sumas parciales de la integral doble que la define. De esta forma, tenemos que en el límite, las sumas de Riemann Stieljes convergen a lo mismo que las sumas:

$$\sum_{i,j=1}^{n-1} \tilde{\phi}(x_i, y_j)\Delta_{i,j}H \tag{3.30}$$

¡Pero esto es haber vuelto a empezar con $\tilde{\phi}$ y H en lugar de ϕ y F!. Dados los supuestos sobre las desigualdades de H_1 (correspondiente a F_1) y H_2 (correspondiente a F_2) así como de sus marginales, se prueba de manera totalmente análoga al caso de primer orden que $A_1 \geq A_2$. Si se hace con cuidado la reordenación de los términos, no hay problema en rehacer la misma demostración para el caso $\phi \in I^{++}$.

Los términos B y C no ofrecen dificultad alguna, como en el caso absolutamente continuo. Pues basta observar que:

$$F(x_i, 1)\delta_{i,n}\phi = F(x_i, 1)\frac{\partial \phi}{\partial x}(x_i^0, 1)\Delta x \approx \frac{\partial \phi}{\partial x}(x_i^0, 1)\Delta_i H^X$$

y dado que sabemos $\frac{\partial \phi}{\partial x} \geq 0$, la desigualdad marginal para H nos da el resultado buscado, es decir $B_1 \geq B_2$. Para C simplemente se usa el mismo argumento. □

De la misma forma que obtuvimos condiciones suficientes para la dominancia a segundo orden, podemos obtenerlas para orden n-ésimo arbitrario. La dificultad, sin embargo, crece continuamente pues debemos utilizar sucesivamente la fórmula de integración por partes.

Esto implica por un lado que para definir las clases de funciones de utilidad sobre las que vale la dominancia debemos imponer condiciones en los signos de las derivadas hasta orden $2n$.

Por otro, las condiciones suficientes serán dadas como desigualdades de operadores que involucran integrales sucesivamente más complicadas, al anidarse por efecto de la integración por partes. Pensemos por ejemplo en que la dominancia de orden 3 involucra a $\int_0^x \int_0^y H(s,t)dsdt$ pero ya teníamos $H(s,t) = \int_0^s \int_0^t F(u,v)dudv$.

Por otra parte, como mencionamos al definir las nociones de dominancia, las mismas tienen sentido en un espacio de bienes de cualquier dimensión. Los enunciados de los teoremas, las definiciones involucradas y las condiciones impuestas pueden generalizarse al caso n-dimensional[11] de manera bastante evidente, teniendo en cuenta la validez de la fórmula de integración por partes en $\mathbb{R}^n$. Los tests, sin embargo, sufrirán de un aumento exponencial de la necesidad de datos, por usar entre otras cosas integraciones numéricas.

Es conocido el efecto correspondiente en dificultad para la integración numérica: la *maldición de la dimensionalidad* una vez más, que hace crecer el número de puntos de evaluación (y la complejidad del algoritmo, i.e. el número de operaciones necesaries) exponencialmente con la dimensión (Hinrichs & Novak [106]).

[11] El caso de dimensión infinita puede resultar más complicado.

En consecuencia, y dada la mayor relevancia en términos de la economía del bienestar, nos conformamos con estudiar sólo los primeros dos órdenes de dominancia y el caso bivariado.

Capítulo 4

Interpretación Económica de la Dominancia Estocástica Multivariada

4.1. Bienestar Social y Complementariedad Estratégica

La interpretación económica básica de la dominancia estocástica multivariada debe ser la misma que teníamos en una dimensión, a saber, una idea de comparación de bienestar *à la Atkinson*. Es decir, considerando los funcionales de Bergson-Samuelson uniformes:

$$W(F) = \int_U \phi(x_1, x_2, ..., x_n) dF \tag{4.1}$$

donde la ϕ es una función de utilidad sobre $\mathbb{R}^n$ y F es una función de distribución N-dimensional. En este sentido, decir que F domina estocásticamente a G es decir que brinda un mayor bienestar social para la clase de funciones de utilidad que corresponda.

En el caso de la dominancia a primer orden bivariada, vimos (Teo.3.2.1.1 y su generalización) que el ordenamiento puntual de las funciones de distribución implicaba un mayor bienestar para la clase de funciones de utilidad creciente en cada variable y submodular. El ordenamiento puntual de la función $K(x,y)$ implicaba un mayor bienestar para la clase supermodular. Esta última se asocia en sentido económico con la noción de *complementariedad estratégica* de los bienes, mientras la primera con el concepto de bienes *sustitutos*.

En el caso que tratamos en la presente Tesis, los *bienes* (o sea los argumentos de la función de utilidad ϕ) que entran en la definición son el ingreso corriente y el capital humano. Si pensamos en la derivada cruzada $\frac{\partial^2 \phi}{\partial x \partial y}$ (donde x es el ingreso corriente e y el capital humano) como el **efecto de un incremento del bien** y **sobre la utilidad marginal del bien** x entonces es de esperar que el signo de dicha derivada sea positivo en nuestro caso.

De manera general, por ejemplo, un mayor nivel de capital humano se asocia a un mejor nivel de salud de los miembros del hogar (Groot y van den Brink [93]). Esto es válido para cualquier medida razonable del capital humano, incluso si sólo tenemos en cuenta la educación como variable (Savvides & Stengos [191], Ch.9). Naturalmente, un mejor nivel de salud implica una mejor *capacidad* (si se quiere incluso en el sentido de Sen [199]) para obtener utilidad de un ingreso dado, es decir que tiene un impacto positivo sobre la utilidad marginal del ingreso.

Por otro lado, un mejor nivel de capital humano permite un refinamiento de los gustos y en consecuencia mayores *aspiraciones* de un individuo. Esto implica entonces una posible menor *satisfacción* de un nivel de ingreso dado fijo, pero en consecuencia un mayor efecto marginal positivo sobre la utilidad de un aumento del mismo (Clark [40]).

Puesto de otra forma, se espera razonablemente que el ingreso y el capital humano sean *complementos estratégicos*. Por esta razón, cuando trabajemos con datos empíricos, nuestro foco estará en testear las condiciones de dominancia sobre las clases I^+ e I^{++}.

Nótese sin embargo que esta complementariedad no la estamos poniendo a prueba, la estamos tomando de la literatura. Si la misma no fuera aceptable tras alguna investigación empírica específica, nuestros resultados econométricos no sufren ningún revés, simplemente usamos el test para el caso submodular a primer y segundo orden.

4.2. Ambigüedad de Rankings y Principios de Transferencias

Dado un vector aleatorio (X,Y) sobre el compacto $U=[0,1]\times[0,1]$ con función de distribución $F(x,y)$, decimos que $(\tilde{X},\tilde{Y})$ se obtiene de aquella mediante una *transformación de correlación decreciente* si la distribución $\tilde{F}(x,y)$ de este último es la misma que $F(x,y)$ en todo U, excepto en un subrectángulo $[x_0,y_0]\times[x_0+h,y_0+k]$, donde $\tilde{F}(x,y)\leq F(x,y)$.

Una forma de obtener una transformación de este tipo es trabajando con las frecuencias, al menos en el caso discreto, que es el más relevante computacionalmente. Si F tiene asociada la frecuencia f_{ij} y G la g_{ij}, entonces G viene de una *transformación elemental de correlación decreciente* si para cierto $i_1<i_2$, $j_1<j_2$:

$$g_{ij}-f_{ij}=\begin{cases}-\varepsilon & \text{si } (i.j)=(i_1,j_1) \;\; o \;\; (i.j)=(i_2,j_2)\\ \varepsilon & \text{si } (i.j)=(i_1,j_2) \;\; o \;\; (i.j)=(i_2,j_1)\\ 0 & \text{en otro caso}\end{cases} \tag{4.2}$$

Epstein y Tanny [67] demuestran que si una distribución se obtiene de otra a partir de una sucesión de estas transformaciones en el mismo sentido, entonces se pueden rankear sin ambigüedad mediante el concepto de dominancia estocástica de primer orden.

También prueban que basta que haya habido un par de estas transformaciones de dirección opuesta en dos puntos distintos para que no se puede determinar la dominancia sin ambigüedad.

Esto traza una analogía de la dominancia estocástica multivariada con el *principio de transferencias* o de Dalton-Pigou que vimos para la dominancia univariada a primer orden en la Sec.1.2.2.

Por otro lado, resulta también en este caso que para poder rankear de manera definitiva un par de distribuciones, puede ser necesario tener en cuenta un peso específico diferente de las transferencias, según sean a niveles más altos o bajos de las variables.

Atkinson y Bourguignon [14] prueban que la dominancia estocástica bivariada a segundo orden permite ordenar distribuciones sin ambigüedad al asignar un mayor peso a las transferencias en los valores más bajos las variables. Este resultado es equivalente al de las *transferencias disminuidas* de Kolm, que vimos para la dominancia univariada en la Sec.1.2.2.

Estos resultados dan cuenta de las buenas propiedades del concepto de dominancia estocástica desde el punto de vista *axiomático*, aun cuando pasamos al caso multivariado.

4.3. La *Interacción* de Dimensiones y su Significado Económico

Hemos señalado ya en la Sec.1.5.1 y el Cap.2 que el paso de la desigualdad unidimensional a la multidimensional trae aparejado ciertas dificultades. El núcleo de esas dificultades está bien resumido por Savaglio [190]:

> *"...it is difficult to extend the ranking principles and measures from the univariate to the multivariate case. The main reason for this difficulty regards the interaction between income and non-income attributes."*(Savaglio [190])

En el caso bivariado de dominancia estocástica, es claro que el efecto de la *interacción* que menciona Savaglio entra en juego por dos motivos:

- La función de utilidad $\phi(x,y)$: el *signo* del efecto de la interacción viene dado por el de la derivada segunda $\frac{\partial^2\phi}{\partial x\partial y}$. Es el efecto de la complementariedad estratégica.
- La distribución bivariada $F(x,y)$: la correlación estadística entre las variables nos da la *fuerza* del efecto.

es claro que esto se generaliza a n dimensiones, evaluando las derivadas segundas cruzadas y la correlación en base a la función $F(x_1,...,x_n)$.

Recordemos además el hecho demostrado en el Cap.3 de que la dominancia en las marginales es *necesaria* (las desigualdades marginales están implicadas por las de $F(x,y)$ y $K(x,y)$) pero no *suficiente* para la dominancia bivariada. Esto tiene una interpretación desde la economía del bienestar y es que la interacción es que puede llevar a efectos del tipo *mitigación* o *agravamiento* mencionado por Rietveld [178].

Una economía puede tener mejor distribución del ingreso que otra, e incluso mejor distribución de capital humano, pero la *distribución conjunta* (es decir el producto de su correlación) puede ser tal que el bienestar no sea mayor en la primera. Esto nos dice desde un punto de vista técnico que las dominancias marginales no alcanzan para dar una ordenación conclusiva de las distribuciones.

Y desde el punto de vista económico, tomando en cuenta que el ingreso y el capital humano son complementos estratégicos, se podría decir que para una mejora del bienestar social no sólo importa que sea mejor la distribución del ingreso y de la educación, sino que las correlación entre dichas variables debe ser mayor.[1]

[1] Recordemos del Cap.3 (y de Epstein y Tanny [67]) que si para pasar de F a G hubo una *transformación de correlación creciente* entonces $F \leq G$ puntualmente. Esto implica que la primera domina en el caso submodular, pero la segunda en el supermodular, pues como la transformación deja las marginales fijas, resulta $K_G \leq K_F$.

Ahora bien, económicamente, la mayor o menor correlación entre las variables ingreso y capital humano va a estar determinada por la estructura del mercado laboral. Consideremos por ejemplo una economía en la cual la creación de empleo se de principalmente en trabajos poco calificados, como ser la construción y el empleo público en diversos niveles.[2]

En este caso, el retorno a los años invertidos en educación superior cae en relación con los de los niveles medios e inferiores. De manera que con iguales marginales (o incluso una mejora en la desigualdad en el ingreso) se puede tener un bienestar social menor.

Otro problema derivado de ello es por supuesto la *sustentabilidad* de un modelo basado en la creación de empleo público de baja productividad. Si pensamos en base a los modelos endógenos que el crecimiento de la economía depende a largo plazo de la acumulación de capital humano (por la capacidad de generar y *utilizar* tecnología) entonces dicho esquema de creación de empleo no es sostenible, porque se retroalimenta con un incentivo al trabajador para acumular menos capital humano.

Esto necesariamente generará estancamiento y en la medida que no se revierta puede dar lugar a situaciones de crisis (colapso, ruptura de contratos, etc.[3]) que generan saltos en la desigualdad capaces de compensar lo que se hubiera ganado en las marginales.

Estas observaciones son generales, y se aplican a la dominancia estocástica de cualquier orden. Para el caso de la dominancia a segundo orden, Atkinson y Bourguignon introducen una función $R(x,y)$ que permite ver claramente el efecto de la correlación sobre la dominancia, al menos en el caso absolutamente continuo.

Definamos la *media incompleta* y la *covarianza incompleta*:

$$\bar{x}(x,y) = \frac{1}{F(x,y)} \int_0^x \int_0^y sf(s,t)dtds \tag{4.3}$$

$$\overline{cov}(x,y) = \frac{1}{F(x,y)} \int_0^x \int_0^y (st - \bar{x}(x,y)\bar{y}(x,y))f(s,t)dtds \tag{4.4}$$

y a partir de ello la función de correlación parcial:

$$R(x,y) = \overline{cov}(x,y) + (x - \bar{x}(x,y))(y - \bar{y}(x,y)) \tag{4.5}$$

Mediante esta función, Atkinson y Bourguignon [14] reescriben la función $L(x,y)$ de la Sec.3.3 de la forma:

[2]Algunos analistas opinan que éste es el caso de la Argentina desde 2009, con el Estado como único creador de empleo desde 2011/12. No hay fuentes de estadísticas realmente confiables, referimos a notas periodísticas como la de Ismael Bermúdez [21] y la de López Gottig [136] que trabajan sobre números oficiales del Ministerio de Trabajo e INDEC.

[3]Por supuesto, esto depende de otros factores macroeconómicos, pero en la mayoría de los casos esta situación llevará aparejados probablemente un déficit fiscal y/o una presión impositiva creciente para financiar el empleo improductivo.

$$L(x,y) = xy\left(F^X(x)\left(1-\frac{\bar{x}(x,b)}{x}\right)+F^Y(x)\left(1-\frac{\bar{y}(a,y)}{y}\right)\right)-F(x,y)R(x,y) \tag{4.6}$$

y evaluando en (a,b) resulta:

$$L(a,b) = (ab-\mu_x\mu_y)-cov \tag{4.7}$$

En esta fórmula es entonces posible ver cómo una mayor correlación implica un menor valor de $L(x,y)$ y en consecuencia la dominancia de segundo orden. En el caso de iguales medias, la de menor covarianza no puede dominar. Pero en el caso en que las medias de una distribución sean mayores, la condición sobre las covarianzas será necesaria pero no suficiente.

Por último, señalemos que hemos usado la idea básica de que la *interdependencia* de las dimensiones determinada por la condición de supermodularidad de las funciones base de la dominancia. El trabajo de Meyer y Strulovici [150] presenta diversas nociones de interdependencia y demuestra que en el caso *bidimensional* todas son equivalentes, pero con más dimensiones pueden dar origen a distintos preordenes.

Capítulo 5

Tests de Dominancia Estocástica en Base a Estadísticos de Kolmogorov-Smirnov

5.1. Dominancia Estocástica via Kolmogorov-Smirnov

5.1.1. Teorema de Hadar y Russell (1969)

Desde que introdujimos la noción de dominancia estocástica en la introducción (Sec.1.2.2) venimos trabajando con las condiciones dadas por las desigualdades del tipo:

$$\begin{cases} F(z) \leq G(z) \quad z \in [0,1] \\ \int_0^z F(t)dt \leq \int_0^z G(t)dt \end{cases}$$

En el Cap.3 vimos que estas condiciones son necesarias y suficientes para la dominancia a primer y segundo orden en el caso univariado. Una de las implicaciones del caso $SD1$ es trivialmente demostrada integrando por partes. Para la otra, es necesario algún resultado más general, como el Teo. de Strassen, y entonces sale fácil (lo vimos en Sec.3.1).

Para la condición de SD_2, podemos demostrar la suficiencia fácilmente integrando por partes. Para ver que la condición es necesaria construyamos la siguiente función:

$$u(y) = y(1 - H(y)) + \int_0^y t dH$$

donde H es una función de distribución de probablilidad concentrada en cierto $x \in [0,1]$ fijo. Es fácil ver que la función $u(y)$ es creciente y cóncava, de manera que si FSD_2G debe ser:

$$\int_0^1 udF \geq \int_0^1 udG$$

Pero $\int_0^1 u dF = x - \int_0^x F(t)dt$ así que la condición es necesaria.

Este resultado fue probado por Hadar y Russell [95] en 1969 en su trabajo seminal sobre la dominancia estocástica. De hecho su demostración de la suficiencia es válida para funciones crecientes y cóncavas en general, sin usar integración por partes.

Es el resultado más importante porque convierte la idea intuitiva de una mayor utilidad de von Neumann-Morgenstern para una clase de funciones (unanimidad sobre una clase infinita) en una comparación puntual de funciones que puede hacerse de forma inmediata. Además las condiciones son prácticas porque al usar sólo la función de distribución, se pueden testear estadísticamente a partir de muestras, lo cual nos lleva al Test de McFadden.

5.1.2. Tests de McFadden (1989) y Barret & Donald (2003)

Como vimos en la Introducción (Sec.1.3.2) McFadden busca testear la hipótesis $H_0 : F_X(z) \geq F_Y(z)\ \forall z \in [0,1]$ contra $H_1 : F_X(z) < F_Y(z)$ para algún $z \in [0,1]$, a partir de dos muestras $(x_1, ..., x_n)$ y $(y_1, ..., y_n)$.

Nótese que la condición $F_X(z) \geq F_Y(z)\ \forall z \in [0,1]$ es equivalente a pedir que $\sup_{z\in[0,1]}(F_Y - F_X)(z) \leq 0$, por lo cual resulta natural definir el análogo empírico o de plug-in $D_n(z) = \sqrt{n}(F_Y^n(z) - F_X^n(z))$ y definir el estadístico de prueba:

$$D_n^* = \max_{w\in[0,1]} D_n(w)$$

Lo que resta para tener un test de dominancia univariada a primer orden es hallar la distribución de D_n^*, que McFadden [149] demuestra ser del tipo de Smirnov.

Para la dominancia a segundo orden, surge análogamente el estadístico:

$$S_n^* = \max_{w\in[0,1]} S_n(w)$$

donde $S_n(w) = \sqrt{n}\int_0^w (F_Y^n(z) - F_X^n(z))dz$. McFadden no caracterizó completamente el test a segundo orden.

Barret y Donald [15] generalizan los tests de McFadden de varias formas. En primer lugar, proponen una serie operadores integrales para la condición de dominancia a orden j:

$$\begin{cases} S_1(z,f) = f(z) \\ S_{j+1}(z,f) = \int_0^z S_j(t,f)dt \end{cases}$$

La dominancia a orden j de una distribución F sobre otra distribución G si $S_j(z,F) \leq S_j(z,G)$. Ahora dada la distribución empírica $\hat{F}_N = \frac{1}{N}\sum_{i=1}^N 1_{\{x_i \leq z\}}$ obtenemos, como Davidson y Duclos [45], Barret y Donald [15], los estimadores muestrales para los operadores:

$$S_1(z,\hat{F}_N) = \hat{F}_N(z)$$

$$S_2(z,\hat{F}_N) = \int_0^z \hat{F}_N(t)dt = \int_0^z \frac{1}{N}\sum_{i=1}^N 1_{\{x_i \le t\}}dt = \frac{1}{N}\sum_{i=1}^N \int_0^z 1_{\{x_i\le t\}}dt =$$
$$= \frac{1}{N}\sum_{x_i\le z}(z-x_i) = \frac{1}{N}\sum_{i=1}^N (z-x_i)1_{\{x_i\le z\}}(z) \quad (5.1)$$

$$S_3(z,\hat{F}_N) = \frac{1}{N}\sum_{i=1}^N \int_0^z 1_{\{x_i\le t\}}(t-x_i)dt = \frac{1}{N}\sum_{i=1}^N 1_{\{x_i\le z\}}(z)\frac{1}{2}(z-x_i)^2$$

$$\Rightarrow S_j(z,\hat{F}_N) = \frac{1}{N}\sum_{i=1}^N 1_{\{x_i\le z\}}(z)\frac{(z-x_i)^{j-1}}{j-1!} \quad (5.2)$$

el estadístico de prueba a orden j se construye ahora como en el problema general de dos muestras de Kolmogorov permitiendo, a diferencia de lo que hacía McFadden [149] que el tamaño de las muestras de X e Y sea diferente. Definimos primero:

$$D^j_{n,m}(z) = \sqrt{\frac{nm}{n+m}}(S_j(F_X,z) - S_j(F_Y,z)) \quad (5.3)$$

y proponen como antes $D^j_* = \sup_{z\in[0,1]} D^j_{n,m}(z)$.

La técnica usada por Barret y Donald [15] consiste en encontrar *valores-p dependientes de los datos*. Para ello proceden a remuestrear con reposición de $\{x_1,\ldots,x_m\}$ obteniendo muestras de boostrap $\{x^*_1,\ldots,x^*_m\}$ y computan:

$$D^j_{F^X} = \sqrt{m}\sup_{z\in[0,1]}\left(S_j(z,F^{X*}_m) - S_j(z,F^X_m)\right) \quad (5.4)$$

$$D^j_{F^X_m,F^Y_n} = \sqrt{\frac{nm}{n+m}}\sup_{z\in[0,1]}\left(S_j(z,F^X_m) - S_j(z,F^Y_n)\right) \quad (5.5)$$

donde F^*_m es la cdf de la muestra bootstrapeada. Repiten un número B de veces y cuentan la proporción de veces que ocurre $D^j_{F^X} > D^j_{F^X_m,F^Y_n}$, obteniendo de esta forma el *p-value*.

Otra forma de implementar este contraste en lugar de computar los valores p es computar los deciles de la distribución de bootstrap, al estilo de Efron [66].

5.1.3. Desarrollos Recientes

Algunos desarrollos recientes tratan aspectos que no incluimos en la presente Tesis. Un aspecto clave de nuestras comparaciones es que suponemos que las distribuciones involucradas tienen soporte compacto. Esto es razonable para las variables usuales en el estudio de la desigualdad, pues el ingreso, el capital humano y demás variables relevantes están claramente dentro de límites finitos.

Pero dado que los tests de dominancia tienen en otras aplicaciones además del estudio de la desigualdad, y en algunas de ellas la compacidad de los soportes no es un supuesto inmediatamente aceptable, cabe preguntarse qué ocurre si se lo deja de lado.

El trabajo de Horvath, Kokoszka y Zitikis [107] muestra que si no se supone soporte compacto, entonces el estadístico de tipo KS construido por ejemplo para el test de Barret y Donald [15], en el caso de dominancia de orden 3 o superior, resulta no acotado.

La solución que proponen en dicho paper es entonces la utilización de un estadístico de Kolmogorov-Smirnov *con pesos*. El mismo, para $j \geq 3$ y para cada $q : \mathbb{R} \to \mathbb{R}$ se define como:

$$T^j_{n,m}(q) = \sup_{z \in \mathbb{R}} q(z) D^j_{n,m}(z)$$

Los autores demuestran que si el *peso* $q(z)$ cumple que:

$$\sup_{z \in \mathbb{R}} q(z)(1+x)^{j-2} < \infty$$

entonces $T^j_{n,m}$ está bien definido. Por otra parte, siendo $D(z) = \sqrt{\frac{nm}{n+m}}(S_j(F,z) - S_j(G,z))$ prueban que el siguiente estadístico:

$$\sup_{z \in \mathbb{R}} q(z)(D^j_{n,m}(z) - D^j(z))$$

converge en distribución al proceso gaussiano $\Gamma_j(z)$ dado por:

$$\Gamma_j(z) = \int_{-\infty}^{z} (z-y)^{-2} \left(\sqrt{\eta} B_1(F(y)) + \sqrt{1-\eta} B_2(G(y)) \right) dy$$

donde $B_{1,2}$ son puentes brownianos unidimensionales. Una vez caracterizado el proceso límite, encuentran valores críticos mediante un método de bootstrap, de la manera de Barret y Donald.

5.2. Tests de Kolmogorov-Smirnov Multivariados

5.2.1. Clases de Donsker y Teoremas del Límite Central Uniforme

Generalidades y Notaciones

Dado un espacio métrico M llamemos $\mathscr{C}_b(M)$ al conjunto de funciones $f : M \to \mathbb{R}$ continuas y acotadas, y $\mathscr{B}_M$ a su σ-álgebra de Borel, es decir la generada por los abiertos de M.

Una *medida de Borel* sobre M es una medida μ sobre $\mathscr{B}_M$ que es finita sobre compactos; si además $\mu(M) < \infty$ la medida se dice *finita* y si $\mu(M) = 1$ es una medida de Borel de probabilidad. Si para cada $\varepsilon > 0$ existe un compacto K tal que $\mu(K) \geq 1 - \varepsilon$, la medida se dirá *ajustada*.

Dado un espacio de medida (Ω, Σ),una función $X : \Omega \to M$ se dice *ajustada* si la medida que induce sobre $\mathscr{B}_M$ ($P_X = P \circ X^{-1}$)) es ajustada. A la medida inducida por X la llamamos su *ley*. Si existe un conjunto medible separable con medida 1, X o P_X se dicen *separables*.

Dado un espacio de probabilidad (Ω, Σ, P) y una función $f : \Omega \to \bar{\mathbb{R}}$ medible, su *esperanza* viene dada por:

$$Ef = Pf = \int_\Omega f dP \tag{5.6}$$

donde introdujimos la notación Pf para dicha operación. Si $T : \Omega \to \bar{\mathbb{R}}$ es arbitraria (i.e. no necesariamente medible) entonces no siempre se puede calcular su esperanza, pero sí la *esperanza exterior*:

$$E^*T = \text{ínf}\left\{Ef : f \geq T, f : \Omega \to \bar{\mathbb{R}} \textit{ medible y } Ef \textit{ existe}\right\} \tag{5.7}$$

donde la existencia de Ef se entiende en el sentido usual de que no sean infinitas al mismo tiempo su parte positiva y negativa (Wheeden y Zygmund [234]).

Análogamente, sabemos que si $A \subseteq \Omega$ arbitrario, su probabilidad puede no estar definida, pero la *probabilidad exterior* se define como:

$$P^*(B) = \text{ínf}\{P(A) : B \subseteq A, A \in \Sigma\} \tag{5.8}$$

Un conjunto dirigido es un par $(\mathscr{A}, \leq)$ tal que $\leq$ es un preorden sobre $\mathscr{A}$ y para cada par de elementos a, b existe c con $a \leq c$ y $b \leq c$. Una *red* sobre alguna clase es una aplicación con dominio en un conjunto dirigido e imagen en la clase.

Definición 5.2.1.1 (Convergencia Débil)**.** *Sea* $(\Omega_\alpha, \Sigma_\alpha, P_\alpha)$ *una red de espacios de probabilidad, y sean* $X_\alpha : \Omega_\alpha \to M$ *funciones arbitrarias. Decimos que la red* X_α converge débilmente a una medida de Borel μ *(y lo notamos con* $X_\alpha \rightsquigarrow \mu$*) si:*

$$E^* f(X_\alpha) \to \int f d\mu \quad \forall f \in \mathscr{C}_b(M)$$

□

Análogamente, una red $\{\mu_\alpha\}$ de medidas borelianas sobre M se dice que converge débilmete a μ (notado $\mu_\alpha \rightsquigarrow \mu$) si $\int f d\mu_\alpha \to \int f d\mu$ para las funciones de $\mathscr{C}_b(M)$.

Decimos que la red X_α *converge en probabilidad exterior* a cierto $c \in M$ si $P^*(d(X_\alpha, c) < \varepsilon) \to 0$ para cada $\varepsilon > 0$. Decimos que X_α converge en probabilidad exterior a X si $P^*(d(X_\alpha, X) > \varepsilon) \to 0$.

Un *proceso estocástico* es una familia indexada de elementos aleatorios $\{X_t\}_{t\in T}$ definidas sobre el mismo espacio de probabilidad (Ω, Σ, P). Si estamos indexando con los reales positivos podemos definir una *filtración* como una colección de σ-álgebras $\{\Sigma_t\}_{t\geq 0}$ tal que $\Sigma_t \subseteq \Sigma_u \subseteq \Sigma$ si $0 \leq t \leq u$.

Un proceso estocástico $\{X_t\}_{t\geq 0}$ se dice *adaptado* a $\{\Sigma_t\}_{t\geq 0}$ si X_t es Σ_t-medible.

Nótese que si en la definición de convergencia débil $\{X_\alpha\}$ está adaptado, i.e. X_α es Σ_α medible, entonces la convergencia débil de X_α y la de P_{X_α} es equivalente. Como éste no es el caso general, introdujimos las probabilidades y esperanza exterior.

Dado un conjunto arbitrario S definimos $l^\infty(S)$ como el espacio de todas las funciones $f : S \to \mathbb{R}$ tal que:

$$\| f \|_S = \sup_{s\in S} |f(s)| < \infty$$

Proceso Empírico

Para representar un proceso de muestreo estadístico, queremos contar con variables $X_1, X_2, \ldots$ iid con cierta ley P común a todas. El *modelo estándar* para hacerlo es tomar el producto numerable Ω^∞ de copias de (Ω, Σ, P) y proponer X_i como sus coordenadas (Dudley [62]). Si no se aclara lo contrario, supondremos que (Ω, Σ) es un espacio métrico separable con su álgebra de Borel.

Las *medidas empíricas* vienen dadas por:

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

donde δ_x es la medida de Dirac, es decir una medida sobre los conjuntos en el espacio de llegada de las X_i dada por $\delta_x(A) = 1$ si $x \in A$ y 0 en otro caso.

Nótese que tal como las definimos, las medidas empíricas son *medidas aleatorias*. Para que esto sea claro, supongamos que las X_n toman valores en $\mathbb{R}^d$; entonces dado B un boreliano de $\mathbb{R}^d$ (notado a veces como $B \in \mathscr{B}_d$) $P_n(B)$ depende de ω. Es decir, para cada $\omega \in \Omega$ (i.e. para cada *resultado del experimento aleatorio*) tendremos la medida:

$$P_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}$$

y ésta sí se la podemos aplicar a B, obteniendo la *proporción de observaciones que cayeron en B*. El *proceso empírico* viene definido como $G_n = \sqrt{n}(P_n - P)$.

Dada una clase de funciones $\mathscr{F}$ el proceso empírico lo podemos pensar como una función aleatoria de $\mathscr{F}$ operando según:

$$G_n f = \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - Pf) \tag{5.9}$$

La *ley de los grandes números* y el *teorema del límite central* implican para el caso en que sean variables reales:

$$P_n f \xrightarrow{as} Pf \tag{5.10}$$

$$G_n f \rightsquigarrow N(0, P(f - Pf)^2) \tag{5.11}$$

siempre que exista Pf y sea $Pf^2 < \infty$ respectivamente.

Dada una medida de probabilidad Q y una clase de funciones medibles $\mathscr{F}$ notamos $\| Q \|_{\mathscr{F}} = \sup\{|Qf| : f \in \mathscr{F}\}$. Con esta notación, la *ley uniforme de los grandes números* sobre la clase $\mathscr{F}$ se enuncia:

$$\| P_n - P \|_{\mathscr{F}} \to 0 \tag{5.12}$$

donde la convergencia es en probabilidad experior o casi segura. El caso clásico sobre $\mathbb{R}$ es el Teorema de Glivenko-Cantelli y en consecuencia decimos en general que una clase de funciones $\mathscr{F}$ es *P-Glivenko-Cantelli* (o simplemente Glivenko-Cantelli si la P está clara) si se cumple 5.12.

Si suponemos que $\sup_{f \in \mathscr{F}} |f(x) - Pf| < \infty$ para todo x, entonces el proceso empírico resulta una aplicación acotada sobre $\mathscr{F}$, es decir podmos pensar:

$$G_n : \Omega \to l^\infty(\mathscr{F})$$

es decir un proceso estocástico, y en consecuencia tiene sentido preguntarse bajo qué condiciones:

$$G_n = \sqrt{n}(P_N - P) \rightsquigarrow G \quad \textit{en} \;\; l^\infty(\mathscr{F}) \tag{5.13}$$

donde el límite G es un elemento borel medible ajustado en $l^\infty(\mathscr{F})$. Una clase $\mathscr{F}$ para la que se cumple esto s dice una *clase P-Donsker* o simplemente Donsker si no hay confusión sobre el P. El nombre viene del hecho de que si estamos en el caso en que $\mathscr{F}$ son las indicadoras de semirrectas en $\mathbb{R}$, entonces el resultado es válido y es el conocido Teorema de Donsker (Billingsley [27]). Una clase $\mathscr{F}$ que es de Donsker para cualquier P se dice una *clase universal de Donsker*.

El proceso límite G se conoce como *P-puente Browniano* y se suele notar G_P para acentuar la dependencia en P. Su naturaleza se puede entender mirando las distribuciones marginales, siendo que por el Teorema del Límite Central en $\mathbb{R}^k$:

$$(G_n f_1, ..., G_n f_k) \rightsquigarrow N_k(0, \Lambda) \tag{5.14}$$

donde $\Lambda \in \mathbb{R}^{k\times k}$ con $\Lambda_{ij} = P(f_i - Pf_i)(f_j - Pf_j)$.

Como la convergencia en $l^\infty(\mathscr{F})$ implica la convergencia marginal, el proceso límite $\{Gf : f \in \mathscr{F}\}$ debe ser un proceso gaussiano de media nula y función de covarianzas:

$$EGf_1Gf_2 = P(f_1 - Pf_1)(f_2 - Pf_2) = Pf_1f_2 - Pf_1Pf_2 \tag{5.15}$$

Dado que en general supondremos que el espacio muestral es un espacio métrico separable, entonces la imagen del puente browniano, supuesta la hipótesis del continuo, será separable (Dudley [63]).

Entropía Métrica y Bracketing

Las clases de funcioness $\mathscr{F}$ que nos van a interesar son aquellas que vienen dadas como subconjuntos de algún espacio normado, usualmente espacios $L^p(Q)$ para alguna medida de probabilidad Q. Tenemos entonces la estructura $(\mathscr{F}, \|\|)$ y con ella de alguna forma podemos *medir* qué tan grande es nuestra clase de funciones.

En el extremo de que tenga una sola función, sabemos que resultan Donsker, porque es el teorema del límite central clásico. Lo sabemos también para las indicadoras de semirrectas en $\mathbb{R}$, pero para clases demasiado *grandes* puede resultar falsa la afirmación 5.13 (Dudley [62]).

El *número de cubrimiento* $N(\varepsilon, \mathscr{F}, \|\|)$ es el número mínimo de bolas $\{g : \| g - f \| < \varepsilon\}$ necesarias para cubrir $\mathscr{F}$. La *entropía* de $\mathscr{F}$ es el número $\log(N(\varepsilon, \mathscr{F}, \|\|))$.

El *bracket* $[l, u]$ entre dos funciones l y u es el conjunto de funciones f con $l \leq f \leq u$. Un ε-bracket es un bracket con $\| u - l \| < \varepsilon$. El *número de bracketing* de $\mathscr{F}$ notado $N_{[]}(\varepsilon, \mathscr{F}, \|\|)$ es el número mínimo de ε-brackets necesario para cubrir $\mathscr{F}$. La *entropía con bracketing* es $\log\left(N_{[]}(\varepsilon, \mathscr{F}, \|\|)\right)$.

En muchos casos, por ejemplo para el caso $L^p(Q)$, la desigualdad $|f| \leq |g|$ implica $\| f \| \leq \| g \|$ y esto implica a su vez $N(\varepsilon, \mathscr{F}, \|\|) < N_{[]}(\varepsilon, \mathscr{F}, \|\|)$. En general no hay desigualdades para el otro lado, excepto para el caso de la norma uniforme en que son iguales.

Cuando estamos en el caso $\mathscr{F} \subseteq L^p(Q)$ se suele notar $N(\varepsilon, \mathscr{F}, L^p(Q))$ y el respectivo con bracketing. Una *función envolvente* para $\mathscr{F}$ es cualquier $F(x)$ tal que $|f(x)| < F(x)$ para cada x y cada f. La función envolvente minimal es la que a cada x asigna $\sup_f |f(x)|$, y se asume (excepto que se diga lo contrario) que es finita.

La existencia de una envolvente finita nos permite definir el *número uniforme de entropía* relativo a $L^p(Q)$ como:

$$\sup_Q \log N(\varepsilon, \mathscr{F}, L^p(Q)) \tag{5.16}$$

donde el supremo se toma sobre las medidas de probabilidad Q que cumplen $0 < QF^r < \infty$.

Teoremas de Glivenko-Cantelli, Teoremas de Donsker

En el libro de van der Vaart y Wellner [228] se demuestran los siguientes teoremas que caracterizan clases de Glivenko-Cantelli:

Teorema de Glivenko-Cantelli (1): si $\mathscr{F}$ es tal que $N_{[]}(\varepsilon, \mathscr{F}, L^p(Q)) < \infty$ entonces $\mathscr{F}$ es una clase de P-Glivenko-Cantelli.

Teorema de Glivenko-Cantelli (2): si $\mathscr{F}$ es $P-$medible con envolvente F tal que $P^*F < \infty$ y además $\sup_Q \log N(\varepsilon \parallel F \parallel_{Q,1}, \mathscr{F}, L^p(Q)) < \infty$ para todo $\varepsilon > 0$.

Teorema de Donsker (1): si $\mathscr{F}$ es un clase de funciones medibles tales que:

$$J_{[]}(\infty, \mathscr{F}, L^p(Q)) = \int_0^\infty \sqrt{N_{[]}(\varepsilon \parallel F \parallel_{Q,1}, \mathscr{F}, L^p(Q))} d\varepsilon < \infty \tag{5.17}$$

entonces $\mathscr{F}$ es P-Donsker.

Teorema de Donsker (2): si $\mathscr{F}$ es tal que

$$\int_0^\infty \sup_Q \sqrt{N(\varepsilon \parallel F \parallel_{Q,1}, \mathscr{F}, L^p(Q))} d\varepsilon < \infty \tag{5.18}$$

entonces $\mathscr{F}$ es P-Donsker para cada P con $P^*F^2 < \infty$ para alguna envolvente F (bajo ciertas hipótesis de medibilidad, cfr. van der Vaart y Wellner [228]).

Combinatoria de Vapnik-Červonenkins

La clave entonces para tener clases que podamos afirmar que cumplen las versiones uniformes de las leyes asintóticas es poder acotar los números de entropía de $\mathscr{F}$. Dado que vamos a trabajar muchas veces con funciones que son indicadoras de conjuntos, una idea es usar desigualdades combinatorias para conjuntos.

Sea A un conjunto y sea $\mathcal{C} \subseteq 2^A$ o sea una colección de subconjuntos. Dado un conjunto finito $\{x_1, ..., x_n\} \subseteq A$ y un cierto $B \subseteq \{x_1, ..., x_n\}$ decimos que la clase $\mathcal{C}$ *levanta* a B si existe $C \in \mathcal{C}$ tal que $B = \{x_1, ..., x_n\} \cap C$. Decimos que la clase $\mathcal{C}$ *pulveriza* a $\{x_1, ..., x_n\}$ si levanta cada uno de sus 2^n subconjuntos.

El *índice VC* de la clase $\mathcal{C}$, notado $V(\mathcal{C})$, es el mínimo n tal que ningún conjunto de cardinal n es pulverizado por $\mathcal{C}$. Una colección $\mathcal{C}$ de conjuntos se llama una *clase VC* si $V(\mathcal{C}) < \infty$.

Llamemos $\Delta_n(\mathcal{C}, x_1, ..., x_n)$ al número de subconjuntos de $\{x_1, ..., x_n\}$ levantados por $\mathcal{C}$. El *lema de Sauer* afirma que si $\mathcal{C}$ es una clase VC entonces:

$$\max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, ..., x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j} \leq \left(\frac{ne}{V(\mathcal{C})-1} \right)^{V(\mathcal{C})-1} \tag{5.19}$$

Ahora podemos pensar en las clases $\mathcal{C}$ como representando una clase de funciones $\mathscr{F} = \{1_C : C \in \mathcal{C}\}$ y usar las cotas combinatorias para acotar los números de cubrimiento de la misma, en el caso de $L^1(Q)$ para medidas empíricas. Utilizando el Lema de Sauer, van der Vaart y Wellner [228] prueban que existe una constante universal K tal que para cualquier clase VC $\mathcal{C}$:

$$N(\varepsilon, \mathcal{C}, L^p(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\varepsilon} \right)^{p(V(\mathcal{C})-1)} \tag{5.20}$$

para cualquier medida de probabilidad Q, $p \geq 1$ y $\varepsilon \in (0,1)$. Esto significa que las clases VC son *polinomiales* en el sentido de que sus números de cubrimiento están acotados por un polinomio en $1/\varepsilon$.

La acotación muestra que las clases VC satisfacen las condiciones suficientes para los teoremas de Glivenko-Cantelli y Donsker enunciados más arriba.

5.2.2. El Problema de Dos Muestras

Consideremos dos sucesiones de elementos aleatorios $\{X_i\}_{i\in\mathbb{N}}$, $\{Y_j\}_{j\in\mathbb{N}}$, que toman valores en el mismo espacio medible (E, Σ) (e.g. en $\mathbb{R}^N$ con los borelianos). Los subconjuntos ordenados del tipo $(X_1, X_2, ..., X_m)$ o $(Y_1, Y_2, ..., Y_n)$ son las *muestras*, cada una dando lugar a las siguientes medidas sobre Σ:

$$P_m = \frac{1}{m} \sum_{i=1}^{m} \delta_{X_i} \qquad Q_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{Y_j} \tag{5.21}$$

El *problema de dos muestras* en estadística consiste en determinar si las medidas P_m y Q_n son *medidas empíricas* de la misma distribución subyacente desconocida P sobre (E, Σ).

Al ser desconocida la distribución P, no podemos comparar cada una de las distribuciones *muestrales* con ella[1] mediante $\sqrt{m}(P_m - P)$ y $\sqrt{n}(Q_n - P)$ sino que debemos comparar las distribuciones muestrales entre sí.

Esto nos lleva al planteo del siguiente estadístico:

$$\nu_{m,n} = \left(\frac{mn}{m+n} \right)^{1/2} (P_m - Q_n) \tag{5.22}$$

A su vez, la *medida empírica agrupada* H_N se define como la medida empírica obtenida sampleando (con o sin reemplazo según el caso) de los *datos agrupados*:

[1] En ese caso hablamos del *problema de una muestra*, Siegel [207].

$$(Z_{N1}, Z_{N2}, ..., Z_{NN}) = (X_1, ..., X_m, Y_1, ..., Y_n)$$

es decir, si llamamos $\alpha_N = m/n$:

$$H_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{Z_{Ni}} = \alpha_N P_m + (1-\alpha_N)Q_n \tag{5.23}$$

La *hipótesis básica* de independencia (que dejaremos de lado en la Sec.5.4) es que sobre el espacio producto numerable:

$$(\mathscr{E}, \Omega) = \prod_{i=1}^{\infty}(E_i, \Sigma_i) \times \prod_{j=1}^{\infty}(F_j, \Lambda_j) \text{ con } E_i = F_j = E \;\; \Sigma_i = \Lambda_j = \Sigma \;\; \forall (i,j) \tag{5.24}$$

tenemos definida una medida de probabilidad con dos medidas Q y P que son sus factores, es decir, la medida $\mathscr{P} \times \mathscr{Q}$ sobre $(\mathscr{E}, \Omega)$ es tal que cada una es la medida producto de copias identicas (E, Σ, P) y (E, Σ, Q) respectivamente.

Sea ahora Θ una clase de medidas de probabilidad sobre (E, Σ); la (Θ)*hipótesis nula* es que $P = Q \in \Theta$. Bajo estas condiciones, Dudley [62] Ch.11, demuestra el siguiente resultado:

Teorema 5.2.2.1 (Problema de Dos Muestras)**.** *Supongamos que $\mathscr{F}$ es una clase de funciones Θ-universal de Donsker sobre (E, Σ). Entonces para cada $P \in \Theta$, bajo la* (Θ)hipótesis nula *se tiene $\nu_{m,n} \rightsquigarrow G_P$ con $m, n \to \infty$* □

Notemos que al ser $P_m - H_N = (1-\alpha_N)(P_m - Q_n)$ el test de Kolmogorv-Smirnov basado en $\nu_{n,m}$ es equivalente a un test basado en la discrepancia entre P_m y la medida empírica agrupada.

El Teo.5.2.2.1 generaliza el caso usual para procesos empíricos univariados, que se obtiene como caso particular cuando $\mathscr{F}$ es la clase de indicadoras de semirectas $1_{(-\infty,x]}$ y los elementos aleatorios son variables aleatorias reales. El resultado clásico de Kolmogorov y Smirnov (cfr. Billingsley [27], Ch.2) nos dice que cuando las distribuciones son continuas el estadístico es *pivotal* (su distribición límite es independiente de quién sea P) y de hecho su distribución es conocida, fue tabulada por Smirnov [212] en 1939 y su derivación puede verse en los conocidos libros de Feller [70], Billingsley [27] o de Hájek et al. [97].

De manera general, el estadístico es pivotal (*distribution free*) si las medidas P y Q no tienen átomos (van der Vaart & Wellner [228]). Sin embargo, para espacios muestrales más generales y otras clases de funciones, la distribución asintótica bajo la hipótesis nula dependerá de la medida común subyacente $P = Q$ (desconocida).

Ello implica que necesitamos alguna forma alternativa de encontrar valores críticos para este test. Existen dos formas básicas de hacerlo, que son el *proceso empírico de permutación* y el *bootstrap de dos muestras*. En las Sec.5.3 y 5.4 utilizaremos éstas para encontrar *valores críticos dependientes de los datos* de manera general para Kolmogorov-Smirnov y en base a ello para los tests de dominancia estocástica que allí diseñamos.

5.2.3. Bootstrap Empírico

Es útil contar con métricas que caractericen la convergencia débil en ciertas situaciones; una de ellas es la *métrica β de Lipschitz*.

Si (M,d) es un espacio métrico, y $f : M \to \mathbb{R}$ definimos su *norma de Lipschitz* como:

$$\| f \|_L = \sup \left\{ \frac{|f(x) - f(y)|}{d(x,y)} : x \neq y \right\}$$

y la *norma acotada de Lipschitz* como $\| f \|_{BL} = \| f \|_L + \| f \|_\infty$. Una función tal que $\| f \|_L < \infty$ se dice función de Lipschitz y una tal que $\| f \|_{BL} < \infty$ se dice función de Lipschitz acotada.

Sean $(\Omega_0, \Sigma_0, Q_0)$ y $(\Omega_1, \Sigma_1, Q_1)$ espacios de probabilidad y sea (M,d) un espacio métrico. Sean $f_0 : \Omega_0 \to M$ y $f_1 : \Omega_1 \to M$ donde f_0 es medible y tiene imagen separable. Sea $P = Q_0 \circ f_0^{-1}$ y llamemos:

$$\beta(f_1, f_2) = \sup_{\|\psi\|_{BL} \leq 1} \{E^* \psi(f_1) - E\psi(f_0)\} \tag{5.25}$$

donde el supremo lo tomamos sobre las funciones $\psi : M \to \mathbb{R}$ que tienen norma de Lipschitz acotada menor que 1. Dudley [62] prueba que dada una sucesión $(\Omega_k, \Sigma_k, Q_k)$ de espacios de probabilidad y (M,d) métrico, funciones $f_k : \Omega_k \to M$, donde f_0 tiene rango separable y es medible, entonces son equivalentes:

$$f_k \rightsquigarrow f_0 \tag{5.26}$$

y

$$\beta(f_k, f_0) \xrightarrow[k \to \infty]{} 0 \tag{5.27}$$

Antes de pasar al bootstrap definamos la seminorma $\rho_P(f) = (P(f - Pf)^2)$ y llamemos *pre-gaussiana* a una clase de funciones $\mathscr{F}$ tal que:

- $\mathscr{F}$ es totalmente acotada para ρ_P
- Existe una versión del puente browniano G con trayectorias muestrales $f \to Gf$ uniformemente ρ_P-continuas.

La idea intuitiva del Bootstrap, desde Efron [65] es usar las *muestras* como si fueran *la población*. La justificación matemática rigurosa de esta técnica está en los Teoremas Bootstrap de Giné-Zinn [82, 83, 84].

Dada una muestra con su medida empírica P_n, iteramos ese procedimiento y obtenemos una nueva muestra P_n^B muestreando n veces con reposición. Los teoremas bootstrap nos dicen que bajo ciertas condiciones el *proceso empírico de bootstrap* $\sqrt{n}(P_n^B - P_n)$ se comporta asintóticamente como el proceso empírico $\sqrt{n}(P_n - P)$, y ambos como el puente browniano G_P.

Para formalizarlo, sea (Ω, Σ, P) un espacio de probabilidad y $\mathscr{F}$ una clase de funciones sobre Ω, reales y medibles. $X_1, X_2, \ldots$ vienen dadas como en el *modelo estándar*, es decir como las coordenadas del producto numerable de infinitas copias de (Ω, Σ, P). Sean a su vez $X_{n1}^B, \ldots, X_{nn}^B$ independientes con distribución P_n. La *medida empírica de bootstrap* y el *proceso empírico de bootstrap* son:

$$P_n^B = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_{nj}^B} \tag{5.28}$$

y

$$G_n^B = \sqrt{n}(P_n^B - P_n) \tag{5.29}$$

Notemos por $\beta_{\mathscr{F}}$ a la métrica de Lipschitz 5.25 cuando d es la metrica dada por $\|\|_{\mathscr{F}}$.

Decimos que para P y $\mathscr{F}$ vale el *teorema del límite central de bootstrap* en probabilidad si $\mathscr{F}$ es pre-gaussiana para P y $\beta_{\mathscr{F}}(G_n^B, G_P) \xrightarrow[n\to\infty]{} 0$ condicional a $X_1, \ldots, X_n$.

El resultado de Gine y Zinn [83] da condiciones suficientes para el teorema de bootstrap, que son simplemente las de Donsker. La demostración se puede encontrar en Dudley [62] Ch.11.

Teorema 5.2.3.1 (Giné-Zinn)**.** *Sea (Ω, Σ, P) cualquier espacio de probabilidad. Entonces si $\mathscr{F}$ es una clase $P-$Donsker se cumple el teorema boostrap del límite central para P y $\mathscr{F}$.* □

5.3. Tests Multivariados de Dominancia Estocástica

5.3.1. Dominancia Estocástica a Primer Orden

Estadísticos de Prueba

En base a los resultados de la Sección 3.2, recordando que $K(x,y) = -(F(x,y) - F^X(x) - F^Y(y))$, tenemos condiciones necesarias y suficientes para la dominancia estocástica sobre clases de modularidad:

$$SD_1 : \begin{cases} \Delta F \leq 0 \Rightarrow F_1 SD_1 F_2 & \text{sobre la clase } \mathscr{I}^- \\ \Delta K \leq 0 \Rightarrow F_1 SD_1 F_2 & \text{sobre la clase } \mathscr{I}^+ \end{cases} \tag{5.30}$$

donde $\Delta F = F_1 - F_2$ y la expresión $\Delta F \leq 0$ significa $\Delta F(x,y) \leq 0 \ \forall (x,y)$ y análogamente para K.

Para la implementación estadística necesitamos construir versiones empíricas de estas funciones, es decir los *estadísticos* asociados a una muestra de tamaño n. De la misma forma que en el caso unidimensional, tenemos la medida empírica dada por $P_n(B) = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i,y_i)}(B)$ para cada boreliano B de $\mathbb{R}^2$.

Con ella obtenemos a su vez la función la distribución empírica:

$$\hat{F}_n(s,t) = P_n((-\infty,s] \times (-\infty,t]) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \leq s\}}(s) 1_{\{y_i \leq t\}}(t) \tag{5.31}$$

y las marginales:

$$\hat{F}_n^X(s) = P_n((-\infty,s] \times \mathbb{R}) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \leq s\}}(s)$$

$$\hat{F}_n^Y(s) = P_n(\mathbb{R} \times (-\infty,t]) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{y_i \leq t\}}(t)$$

La versión empírica para $K(x,y)$ resulta de esta manera:

$$\hat{K}(s,t) = -\left(\hat{F}(s,t) - \hat{F}^X(s) - \hat{F}^Y(t)\right) \tag{5.32}$$

A partir de 5.30 seguimos la lógica de McFadden [149] y proponemos los estadísticos de Kolmogorov-Smirnov para testear dichas condiciones:

$$KS : \begin{cases} \lambda_{m,n} & \text{para la clase } \mathscr{I}^- \\ \kappa_{m,n} & \text{para la clase } \mathscr{I}^+ \end{cases} \tag{5.33}$$

Para el primer caso, hemos retomado el estadístico del *problema de dos muestras* (Sec.5.2.2.1) $\nu_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2}(P_m - Q_n)$, lo aplicamos sobre los conjuntos de la forma $[0,s] \times [0,t]$, y tomamos el supremo para obtener la condición uniforme del tipo Kolmogorov-Smirnov, quedando entonces definido:

$$\lambda_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{(s,t)\in[0,1]\times[0,1]} (\hat{F}_m - \hat{G}_n)(s,t) \tag{5.34}$$

Para la clase supermodular hemos introducido un estadístico análogo, que toma en cuenta los efectos de borde:

$$\kappa_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{[0,1]\times[0,1]} (\hat{K}_m^F - \hat{K}_n^G)(s,t) \tag{5.35}$$

donde $K^F(x,y) = -(F(x,y) - F^X(x) - F^Y(y))$ y lo propio para G.

El Bootstrap de Dos Muestras

Nuestro enfoque de los tests de dominancia estocástica sigue la línea de Barret y Donald [15], en el sentido de que vamos a trabajar con comparaciones punto a punto de las distribuciones (vía Kolmogorov-Smirnov) y utilizaremos técnicas de remuestreo.

Para testear estadísticamente la dominancia estocástica a primer orden, necesitamos obtener la distribución asintótica de nuestros *edp* $\lambda_{m,n}$ y $\kappa_{m,n}$ bajo la hipótesis nula. Dada la no pivotalidad del estadístico de Kolmogorov-Smirnov para el caso general, con distribuciones multivariadas F y G arbitrarias, dicha distribución asintótica dependerá en general de la subyacente, común bajo $\mathscr{H}_0$.

La *consistencia* del *bootstrap* (tal como dada, por ejemplo, en probabilidad por el teorema de Giné y Zinn 5.2.3) nos permitirá, sin embargo, encontrar valores críticos para los estadísticos de prueba mediante una técnica de *remuestreo*, de manera que el test resulte consistente contra cualquier hipótesis alternativa ($\mathscr{H}_1$: $P \neq Q$).

Aplicada al problema de dos muestras, la técnica requiere un remuestreo con reposición de muestras de tamaño $N = n + m$ a partir de la *muestra agrupada*, tal como definimos en la Sec.5.2.2:

$$(Z_{N1}, Z_{N2}, \ldots, Z_{NN}) = (X_1, \ldots, X_m, Y_1, \ldots, Y_n) \tag{5.36}$$

siendo a su vez la *muestra de bootstrap* $(\hat{Z}_{N1}, \hat{Z}_{N2}, \ldots, \hat{Z}_{NN})$ obtenida muestreando N veces con reposición de 5.36. Cada muestra de bootstrap permite entonces definir sus propias medidas empíricas, las *medidas empíricas del bootstrap de dos muestras*:

$$\hat{P}_{m,N} = \frac{1}{m}\sum_{i=1}^{m} \delta_{\hat{Z}_{Ni}} \qquad \hat{Q}_{n,N} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{Z}_{N,m+i}} \tag{5.37}$$

y la ya mencionada *distribución empírica agrupada*:

$$H_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{Z_{Ni}} = \alpha_N P_m + (1-\alpha_N)Q_n \tag{5.38}$$

Al muestrear con reposición, las muestras de bootstrap resultan independientes.

Los siguientes resultados son demostrados por van der Vaart y Wellner [228] y resultan básicos para la demostración de nuestro propio teorema sobre la distribución de los edps (Teo.5.3.4).

Teorema 5.3.2 (Bootstrap de Dos Muestras (convergencia en probabilidad))**.** *Sea $\mathscr{F}$ una clase de funciones medibles que es al mismo tiempo P-Donsker y Q-Donsker y que cumple $||P||_{\mathscr{F}} < \infty$, $||Q||_{\mathscr{F}} < \infty$. Si $m,n \to \infty$ de manera que $m/N \to \alpha \in (0,1)$, entonces $\sqrt{m}(\hat{P}_{m,N} - H_N) \rightsquigarrow G_H$ dadas $X_1, X_2, ...Y_1, Y_2, ...$ en probabilidad.*

Aquí G_H es un proceso de puente Browniano ajustado, correspondiente a la medida $H = \alpha P + (1-\alpha)Q$.

Demostración. Es el Teo.3.7.6 de van der Vaart y Wellner [228]. □

Teorema 5.3.3 (Bootstrap de Dos Muestras (convergencia casi segura))**.** *Si además $\mathscr{F}$ tiene una función envolvente F con $P^*F^2 < \infty$ y $Q^*F^2 < \infty$, entonces $\sqrt{m}(\hat{P}_{m,N} - H_N) \rightsquigarrow G_H$ dada casi cualquier sucesión $X_1, X_2, ...Y_1, Y_2, ...$*

Demostración. Es el Teo.3.7.7 de van der Vaart y Wellner [228]. □

Nótese que si se cumple que $P = Q$, entonces obviamente $H = P = Q$ y el puente browniano es G_P.

Por un simple argumento de continuidad, estos teoremas implican que la sucesión:

$$\begin{aligned}\tilde{D}_{mn} = \sqrt{\frac{nm}{n+m}} \parallel \hat{P}_{m,N} - \hat{Q}_{n,N} \parallel_{\mathscr{F}} = \\ = \parallel \sqrt{m(1-\alpha_N)}(\hat{P}_{m,N} - H_N) - \sqrt{n\alpha_N}(\hat{Q}_{n,N} - H_N) \parallel_{\mathscr{F}}\end{aligned} \tag{5.39}$$

converge en distribución[2] a la variable:

$$\parallel \sqrt{1-\alpha} G_H - \sqrt{\alpha}\tilde{G}_H \parallel_{\mathscr{F}} \tag{5.40}$$

para puentes brownianos independientes $\tilde{G}_H$ y $\tilde{G}_H$, y el proceso $\sqrt{1-\alpha}G_H - \sqrt{\alpha}\tilde{G}_H$ es una versión del H-puente browniano también.

De esta manera, los β cuantiles de la distribución condicional de $\tilde{D}_{mn}$:

$$\hat{c}_{mn} = \text{ínf}\left\{t : P_{\hat{Z}}(\tilde{D}_{mn} > t) \leq \beta\right\} \tag{5.41}$$

[2]Tómese en cuenta que el valor de una expresión como $\parallel P_N \parallel_{\mathscr{F}}$ (o su análogo de bootstrap) depende de la *muestra* , es decir de $\omega \in \Omega$. En este sentido, es una variable aleatoria, y hablamos de la convergencia en distribución.

se usan como *valores críticos dependientes de los datos* para el test de Kolmogorov-Smirnov.

La *consistencia* del test viene dada por el hecho de que los anteriores teoremas implican que se tiene la convergencia (en probabilidad y casi segura respectivamente):

$$\hat{c}_{mn} \to c_H = \text{ínf}\{t : P(\| G_H \|_{\mathscr{F}} > t) \leq \beta\} \tag{5.42}$$

En esta versión el test resulta a *dos colas*, pues se estudia en valor absoluto la diferencia. En nuestro estudio de la dominancia estocástica, sin embargo, usaremos naturalmente la versión del test a *una cola* (Hájek et al. [97]) pues nos interesan, por ejemplo, hipótesis del tipo $F(s,t) - G(s,t) \leq 0$, $\forall s,t$.

Testeando Dominancia a Primer Orden

En lo que sigue, vamos a considerar la clase de funciones indicadoras[3]:

$$\mathscr{F} = \left\{ 1_{[0,s]\times[0,t]} : s,t \in [0,1] \right\}$$

y el espacio $l^\infty(\mathscr{F})$ de todas las funciones reales acotadas sobre $\mathscr{F}$. Un elemento típico de $l^\infty(\mathscr{F})$ es por ejemplo una medida signada finita, actuando sobre los elementos de $\mathscr{F}$ por integración.

Será práctico asignar a cada función $f \in \mathscr{F}$ un par de funciones adicionales también de $\mathscr{F}$, dadas por:

$$f = 1_{[0,s]\times[0,t]} \longrightarrow \begin{cases} f^x = 1_{[0,s]\times[0,1]} \\ f^y = 1_{[0,1]\times[0,t]} \end{cases}$$

Consideremos ahora los siguientes dos operadores, $T_1, T_2 : l^\infty(\mathscr{F}) \to \mathbb{R}$, dados por:

$$T_1(P) = \sup_{f\in\mathscr{F}} Pf \tag{5.43}$$

$$T_2(P) = \sup_{f\in\mathscr{F}} (Pf^x + Pf^y - Pf) \tag{5.44}$$

Observemos que si P y Q son medidas de probabilidad sobre $[0,1] \times [0,1]$, entonces tienen asociadas funciones de distribución $F(s,t)$ y $G(s,t)$ (y funciones $K^F(s,t), K^G(s,t)$ dadas como en la Sec.3.2) y resulta:

$$T_1(P - Q) = \sup_{(s,t)\in[0,1]\times[0,1]} (F - G)(s,t) \tag{5.45}$$

$$T_2(P - Q) = \sup_{(s,t)\in[0,1]\times[0,1]} (K^F - K^G)(s,t) \tag{5.46}$$

[3]Trabajamos en $\mathbb{R}^2$ porque en la práctica nos enfocamos en el caso bivariado. Todo lo hecho en la presente Sección se generaliza sin dificultad a $\mathbb{R}^d$.

Teorema 5.3.4 (Test de Dominancia Estocástica Multivariada a Primer Orden). *Consideremos los estadísticos $\lambda^B_{m,n}$ y $\kappa^B_{m,n}$ dados usando las distribuciones empíricas del bootstrap de dos muestras (dadas por 5.37) en las fórmulas 5.34 y 5.35. Entonces si $m,n \to \infty$ de tal forma que $m/N \to \alpha \in (0,1)$:*

(i) $\lambda^B_{m,n} \rightsquigarrow T_1(G_H)$

(ii) $\kappa^B_{m,n} \rightsquigarrow T_2(G_H)$

dada casi cualquier sucesión $X_1, X_2, ...Y_1, Y_2, ...$, donde $H = \alpha P + (1-\alpha)Q$ y G_H es una versión del puente browniano ajustado correspondiente a la medida H.

Demostración. Nuestra clase de funciones $\mathscr{F}$ es una clase de Donsker porque es una subclase de los rectángulos de $\mathbb{R}^2$ y ésta lo es (Dudley [63], van der Vaart & Wellner [228]). Por otro lado, la función constante 1 es una envolvente que cumple lo pedido en el Teo.5.3.3. De esta manera tenemos que:

$$\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N}) \rightsquigarrow G_H$$

Ahora bien, por un lado, tenemos que $\lambda^B_{m,n} = T_1(\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N}))$ y $\kappa^B_{m,n} = T_2(\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N}))$. Recordemos a su vez que las medidas empíricas (y sus versiones bootstrapeadas) son elementos aleatorios que toman valores en el espacio métrico $(l^\infty(\mathscr{F}), \|\|_{\mathscr{F}})$, pues para cada $\omega \in \Omega$ dan una medida $P_N(\omega)$ sobre los borelianos de $\mathbb{R}$.

Luego, por la conservación de la convergencia débil por transformaciones continuas (van de Geer [226], Teo.6.4, van der Vaar & Wellner Teo.1.11.1, Pollard [172] Sec.IV.2) se seguirá lo que queremos si los operadores T_1 y T_2 son continuos.

Ahora bien, si P y Q son elementos de $l^\infty(\mathscr{F})$:

$$|T_1(P) - T_1(Q)| = |\sup_{f\in\mathscr{F}} Pf - \sup_{f\in\mathscr{F}} Qf| \leq \sup_{f\in\mathscr{F}} |(P-Q)f| = \| P - Q \|_{\mathscr{F}}$$

lo que implica entonces que T_1 es continua y esto prueba (i).

Ahora sean P y Q elementos de $l^\infty(\mathscr{F})$, entonces:

$$\begin{aligned}|T_2(P) - T_2(Q)| &= |\sup_{f\in\mathscr{F}} (Pf^x + Pf^y - Pf) - \sup_{f\in\mathscr{F}} (Qf^x + Qf^y - Qf)| \\ &\leq \sup_{f\in\mathscr{F}} \{|(Pf^x + Pf^y - Pf) - (Qf^x + Qf^y - Qf)|\} \leq \\ &\leq \sup_{f\in\mathscr{F}} \{|(P-Q)f^x| + (P-Q)f^y| + |(Q-P)f)|\}\end{aligned}$$

Pero es claro que $\sup_{f\in\mathscr{F}} \{|(P-Q)f^x|\} \leq \| P-Q \|_{\mathscr{F}}$, porque estamos barriendo una clase más chica de funciones al tomar las diferencias. Lo mismo se puede decir del término con $\{|(P-Q)f^y|\}$ y entonces $|T_2(P)-T_2(Q)| \leq 3 \| P-Q \|_{\mathscr{F}}$, o sea que T_2 es continuo, y esto termina la demostración de (ii).

□

Como consecuencia del teorema, tenemos una sucesión de valores críticos dada por los cuantiles de la distribución de bootstrap, es decir:

$$\hat{c}^-_{nm} = \text{ínf}\left\{t : P_{\hat{Z}}(\lambda^B_{mn} > t) \leq \beta\right\} \tag{5.47}$$

para el caso submodular, y:

$$\hat{c}^+_{nm} = \text{ínf}\left\{t : P_{\hat{Z}}(\kappa^B_{mn} > t) \leq \beta\right\} \tag{5.48}$$

para el supermodular.

Resulta entonces la consistencia del test, en el sentido de que:

$$\hat{c}^+_{nm} \longrightarrow c_H = \text{ínf}\{t : P(T_1(G_H) > t) \leq \beta\} \tag{5.49}$$

Nuestro test será de una cola, rechazando la hipótesis PSD_1Q (que se deriva de $K^F \leq K^G$ uniformemente) si $l_{m,n} > \hat{c}^+_{nm}$.

Dada la muestra agrupada $X_1,...,X_m,Y_1,...,Y_n$, existen N^N posibles (re)muestras de bootstrap, todas equiprobables y en consecuencia N^N posibles valores de κ^B_{mn} (contando con multiplicidad los empates).

Nótese que el rechazo de PSD_1Q **no** implíca que se acepte QSD_1P, por la simple razón de que $sup_{s,t}(F(s,t)-G(s,t)) \nleq 0$ no implica $sup_{s,t}(G(s,t)-F(s,t)) \leq 0$ (al cambiar el signo habría que mirar el ínfimo, no el supremo). Por ello, no tiene sentido usar un test a dos colas, sino que hay que testear por separado las hipótesis PSD_1Q y QSD_1P. Esto es similar a lo que ocurre con el test de Barret y Donald [15].

Nuestro test rechaza la hipótesis si el valor de κ_{mn} está en la fracción β superior de los valores de κ^B_{mn}, resultando entonces en una probabilidad de error de tipo I menor que β, y la sucesión de tests resulta asintóticamente de nivel β[4] .

5.3.5. Dominancia a Segundo Orden

En la Sección 3.3 encontranos condiciones necesarias y suficientes para la dominancia estocástica en clases superiores de modularidad. Para ello definimos los operadores $\mathscr{H},\mathscr{L} : \mathscr{D} \to \mathscr{C}^2([0,1]\times[0,1])$, donde $\mathscr{D}$ es la clase de todas las funciones de distribución sobre $[0,1]\times[0,1]$ y:

[4] Bajo la hipótesis $P = Q$ todos los valores del edp en el remuestreo son igualmente probables. Bajo la hipótesis PSD_1Q naturalmente se puede correr la distribución a la izquierda, no a la derecha.

$$\begin{cases} \mathscr{H}(F)(x,y) = H(x,y,F) = \int_0^x \int_0^y F(s,t)dtds \\ \mathscr{L}(F)(x,y) = L(x,y,F) = \int_0^x \int_0^y K(s,t)dtds \end{cases} \tag{5.50}$$

y para las condiciones marginales:

$$\begin{cases} H^X(x,F) = \int_0^x F^X(s)ds \\ H^Y(y,F) = \int_0^y F^Y(t)dt \end{cases} \tag{5.51}$$

donde, como siempre $F^X(s) = \text{lím}_{t\to 1} F(s,t)$, y análogamente $F^Y(t)$. Las condiciones de dominancia son, entonces:

$$SD_2 : \begin{cases} \Delta H \leq 0 \ \ \Delta H^{X,Y} \leq 0 \Rightarrow F_1 SD_1 F_2 \ \text{ sobre la clase } \mathscr{I}^{--} \\ \Delta L \leq 0 \ \ \Delta H^{X,Y} \leq 0 \Rightarrow F_1 SD_1 F_2 \ \text{ sobre la clase } \mathscr{I}^{++} \end{cases} \tag{5.52}$$

y aquí nuevamente las desigualdades se interpretan puntualmente. En el enfoque de McFadden [149] vamos a estudiar la condición uniforme de Kolmogorov-Smirnov, y para ello introduciremos:

$$KS : \begin{cases} \mu_{m,n} & \text{para la clase } \mathscr{I}^{--} \\ \gamma_{m,n} & \text{para la clase } \mathscr{I}^{++} \end{cases} \tag{5.53}$$

Para definir estos estadísticos comenzamos con las versiones empíricas $\hat{H}, \hat{L}$ de los operadores, evaluando en la función de distribución empírica:

$$\hat{H}(x,y) = H(x,y,\hat{F}_N) = \int_0^x \int_0^y \hat{F}_N(s,t)dsdt \tag{5.54}$$

Obtenemos una fórmula más operativa del mismo si partimos de la definición de la distribución empírica $\hat{F}_N(s,t) = \frac{1}{N}\sum_{i=1}^N 1_{\{x_i \leq s\}}(s) 1_{\{y_i \leq t\}}(t)$ y la integramos obteniendo:

$$\begin{aligned} \hat{H}(x,y) = \int_0^x \int_0^y \hat{F}_N(s,t)dsdt = \int_0^x \int_0^y \frac{1}{N}\sum_{i=1}^N 1_{\{x_i \leq s\}}(s) 1_{\{y_i \leq t\}}(t)dsdt = \\ = \frac{1}{N}\sum_{i=1}^N \left(\int_0^x 1_{\{x_i \leq s\}}(s)ds\right)\left(\int_0^y 1_{\{y_i \leq t\}}(t)dt\right) = \\ = \frac{1}{N}\sum_{i=1}^N (x-x_i)(y-y_i) 1_{\{x_i \leq x\}}(x) 1_{\{y_i \leq y\}}(y) \end{aligned} \tag{5.55}$$

y para $\hat{L}(x,y)$ podemos hacer lo propio, observando que ya hemos realizado la mayor parte del trabajo puesto que:

$$L(x,y) = \int_0^x \int_0^y K(s,t)dt = \int_0^x \int_0^y (F^X(s) + F^Y(t) - F(s,t))dsdt =$$
$$= -H(x,y) + yH^X(x,F) + xH^Y(y,F) \quad (5.56)$$

y entonces usar las fórmulas empíricas de $H(x,y)$ y de las marginales.

Nótese que cuando usamos en la práctica las funciones de distribución empíricas, usamos la fórmula 5.31 sobre una malla de valores en $[0,1] \times [0,1]$. Para usar la versión empírica del H dada por 5.55 (y la análoga para L), podemos integrar la función sobre esa malla de valores. Esto puede ahorrar algo de tiempo de computación y funcionará muy bien si las muestras son grandes.

Nuevamente, emplearemos estadísticos típicos del *problema de dos muestras* e impondremos la condición uniforme del tipo Kolmogorov-Smirnov. Partiendo de muestras $X_1, X_2, ..., X_m$ y $Y_1, Y_2, ..., Y_m$ definimos entonces para el caso submodular:

$$\mu_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{[0,1]\times[0,1]} (\hat{H}_m^F - \hat{H}_n^G)(s,t) \quad (5.57)$$

donde el supraíndice se refiere a las distribuciones de las muestras correspondientes. El caso supermodular es completamente análogo, resultando:

$$\gamma_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{[0,1]\times[0,1]} (\hat{L}_m^F - \hat{L}_n^G)(s,t) \quad (5.58)$$

Nótese que estos estadísticos de prueba generalizan al caso multidimensional los de Barret y Donald [15, 16].

Trabajamos con las misma definiciones de la Sec.5.3.1 y definimos dos operadores, $T_3, T_4 : l^\infty(\mathscr{F}) \to \mathbb{R}$.

Para definir T_3, empecemos por por asignar a cada elemento $P \in l^\infty(\mathscr{F})$ una función $F_P : [0,1] \times [0,1] \to \mathbb{R}$ dada por:

$$F_P(s,t) = P1_{[0,s]\times[0,t]} \quad (5.59)$$

Esta función es acotada, pues $P \in l^\infty(\mathscr{F})$, así que resulta integrable[5] Lebesgue sobre el compacto $[0,1] \times [0,1]$.

En consecuencia podemos definir:

$$T_3(P) = \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y F_P(s,t)dsdt \quad (5.60)$$

Notemos que si P es una medida de probabilidad, F_P resulta su función de distribución y tenemos:

[5]Si no llegara a ser medible (porque el P sea un elemento raro de $l^\infty(\mathscr{F})$), usamos la integral exterior. De todas formas en el caso de las medidas finitas signadas se ve que es medible, porque es como una función de distribución.

$$T_3(P-Q) = \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y (F_P - F_Q)(s,t)dsdt \tag{5.61}$$

donde adentro de la integral podríamos haber puesto $F-G$ si esos son los nombres que dimos a las respectivas funciones de distribución respectivas de P y Q.

Para el caso supermodular empecemos definiendo para cada $P \in l^\infty(\mathscr{F})$ la función $K_P : [0,1]\times[0,1] \to \mathbb{R}$ según:

$$K_P(s,t) = P1_{[0,s]\times[0,1]} + P1_{[0,1]\times[0,t]} - P1_{[0,s]\times[0,t]} \tag{5.62}$$

y como recién, podemos definir:

$$T_4(P) = \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y K_P(s,t)dsdt \tag{5.63}$$

de manera que para el caso de medidas de probabilidad:

$$T_4(P-Q) = \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y (K_P - K_Q)(s,t)dsdt \tag{5.64}$$

y estos $K(s,t)$ son los mismos de las secciones anteriores, pudiendo de nuevo llamarlos K^F y K^G.

Vamos a usar el bootstrap de dos muestras para encontrar los valores críticos de los estadísticos de prueba. El siguiente teorema es análogo al que probamos para la dominancia a primer orden.

Teorema 5.3.6 (Test de Dominancia Estocástica Multivariada a Segundo Orden)**.** *Consideremos los estadísticos $\mu^B_{m,n}$ y $\gamma^B_{m,n}$ dados usando las distribuciones empíricas del bootstrap de dos muestras (dadas por 5.37) en las fórmulas 5.57 y 5.58. Entonces si $m,n \to \infty$ de tal forma que $m/N \to \alpha \in (0,1)$:*

(i) $\mu^B_{m,n} \rightsquigarrow T_3(G_H)$

(ii) $\gamma^B_{m,n} \rightsquigarrow T_4(G_H)$

dada casi cualquier sucesión $X_1, X_2, \ldots Y_1, Y_2, \ldots$, donde $H = \alpha P + (1-\alpha)Q$ y G_H es una versión del puente browniano ajustado correspondiente a la medida H.

Demostración. Estamos en las mismas condiciones del Teo.5.3.4, así que alcanza con ver que T_3 y T_4 son continuos, puesto que $\mu^B_{m,n} = T_3(\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N}))$ y $\gamma^B_{m,n} = T_4(\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N}))$.

Ahora bien, si P y Q son elementos de $l^\infty(\mathscr{F})$:

$$
\begin{aligned}
&|T_3(P) - T_3(Q)| = \\
&= \left| \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y F_P(s,t)dsdt \;\; - \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y F_Q(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \left| \int_0^x \int_0^y F_P(s,t)dsdt \;\; - \int_0^x \int_0^y F_Q(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \left| \int_0^x \int_0^y (F_P - F_Q)(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y |(F_P - F_Q)(s,t)|dsdt \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y \parallel P - Q \parallel_{\mathscr{F}} dsdt \leq \parallel P - Q \parallel_{\mathscr{F}} \quad (5.65)
\end{aligned}
$$

lo cual implica que T_3 es continua y prueba (i).

La misma técnica permite demostrar (ii) usando la desigualdad triangular como en (ii) de Teo.5.3.6. □

Una vez establecido el Teorema, seguimos la misma lógica de la Sec.5.3.1 y obtenemos valores críticos para la sucesión de tests, que resulta consistente en el sentido antedicho.

5.3.7. Dominancias Superiores

Como señalamos al final de la Sec.3.3.2, mediante un uso repetido de la fórmula de integración por partes se pueden obtener condiciones suficientes para la dominancia estocástica de orden $n-$ésimo sobre clases de funciones anidadas.

Estas clases de funciones quedan determinadas por los signos impuestos a las derivadas de orden $2n$ y las condiciones suficientes estarán dadas por desigualdades de operadores integrales aplicados a cada distribución.

En todos los casos, si bien para cada combinación de signos de las derivadas cruzadas superiores los operadores específicos tienen pequeñas diferencias que reflejan los *efectos del borde* (como el $K(x,y)$ y el $L(x,y)$ en nuestro desarrollo), la clave del asunto es la generalización de la familia de operadores del tipo de Davidson-Duclos dada en 5.2.

En realidad, ya vimos como generalizar los operadores $S_1(z,F)$ y $S_2(z,F)$ pues sus generalizaciones no son más que la función de distribución bivariada $F(s,t)$ y la función $H(x,y)$ que introdujimos en la Sec.3.3 y venimos usando repetidamente desde entonces.

Para el caso bivariado (el N-dimensional es totalmente análogo) podemos proceder a partir de la definición de $H(x,y,F)$ por recurrencia y obtendremos una familia de operadores:

$$\begin{cases} H_1(x,y,F) = F(x,y) \\ H_{j+1}(x,y,F) = \int_0^x \int_0^y H_j(s,t,F)dsdt \quad j \geq 1 \end{cases}$$

El estadístico muestral se obtiene, como siempre, evaluando en la distribución empírica y resulta:

$$\hat{H}_j(x,y) = H_j(x,y,\hat{F}_N) = \\ = \frac{1}{N}\sum_{i=1}^{N}(x-x_i)^{j-1}(y-y_i)^{j-1}\frac{1}{(j-1!)^2}1_{\{x_i \leq x\}}(x)1_{\{y_i \leq y\}}(y) \quad (5.66)$$

Debe notarse que la generalización a N dimensiones es análoga, aunque naturalmente la profusión de subíndices hace más complicadas las fórmulas.

En nuestro enfoque para testear dominancia estocástica a primer y segundo orden recurrimos a condiciones uniformes, y en base a ello fue que definimos los operadores $T_k : l^\infty(\mathscr{F}) \to \mathbb{R}$ con $k = 1,2,3,4$.

Ahora, para seguir con esta lógica imitaremos lo hecho con la dominancia a segundo orden aprovechado que ya hemos definido para cada elemento $P \in l^\infty(\mathscr{F})$ una función $F_P : [0,1] \times [0,1] \to \mathbb{R}$ y entonces podemos definir una sucesión de funciones asociadas a P que llamaremos $S_j(P) : [0,1] \times [0,1] \to \mathbb{R}$ dada por:

$$\begin{cases} S_1(P)(x,y) = F_P(x,y) \\ S_{j+1}(P)(x,y) = \int\int_{[0,x]\times[0,y]} S_j(P)(s,t)dsdt \quad j \geq 1 \end{cases} \quad (5.67)$$

y con ellas podemos definir la familia de operadores $\tilde{T}_k : l^\infty(\mathscr{F}) \to \mathbb{R}$ para cualquier $k \in \mathbb{N}$ como:

$$\tilde{T}_k(P) = \sup_{(x,y)\in[0,1]\times[0,1]} S_j(P)(x,y) \quad (5.68)$$

Le pusimos una tilde al operador $\tilde{T}_k$ porque los T_k como los habíamos definido incluían a los utilizados para las clases supermodulares también (usando $K(x,y)$), así que T_2 y T_4 no están exáctamente en esta nueva familia.

A partir de allí podemos evaluarlos en los elementos de $l^\infty(\mathscr{F})$ que vienen dados por $\sqrt{\frac{mn}{m+n}}(\hat{P}_{m,N} - \hat{Q}_{n,N})$ y podremos usar el mismo razonamiento de los Teoremas 5.3.4 y 5.3.6 si los mismos resultan continuos.

Lema 5.3.8 (Continuidad de la Familia $\left\{\tilde{T}_k\right\}_{k\in\mathbb{N}}$). *: El operador $\tilde{T}_k : l^\infty(\mathscr{F}) \to \mathbb{R}$ es continuo.*

Demostración. Vimos en la demostración del Teo.5.3.4 que para cualquier par P,Q de elementos en $l^\infty(\mathscr{F})$ se tenía $|T_1(P) - T_2(Q)| \leq \parallel P - Q \parallel_{\mathscr{F}}$, y ya sabemos que $T_1 = \tilde{T}_1$.

Ahora trabajamos por inducción y suponemos que $|\tilde{T}_j(P) - \tilde{T}_j(Q)| \leq \| P - Q \|_{\mathscr{F}}$. Bastará ver que esto implica lo propio para $\tilde{T}_{j+1}$ y lo tendremos por inducción para todo $j \in \mathbb{N}$. Tenemos:

$$
\begin{aligned}
&|\tilde{T}_{j+1}(P) - \tilde{T}_{j+1}(Q)| = \\
&= \left| \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y \tilde{T}_j(P)(s,t)dsdt \; - \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y \tilde{T}_j(Q)(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \left| \int_0^x \int_0^y \tilde{T}_j(P)(s,t)dsdt \; - \int_0^x \int_0^y \tilde{T}_j(Q)(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \left| \int_0^x \int_0^y (\tilde{T}_j(P) - \tilde{T}_j(Q))(s,t)dsdt \right| \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y |(\tilde{T}_j(P) - \tilde{T}_j(Q))(s,t)|dsdt \leq \\
&\leq \sup_{(x,y)\in[0,1]\times[0,1]} \int_0^x \int_0^y \| P - Q \|_{\mathscr{F}} \, dsdt \leq \| P - Q \|_{\mathscr{F}}
\end{aligned}
\tag{5.69}
$$

Lo cual prueba la continuidad del operador $\tilde{T}_{j+1}$ y por inducción la de la familia completa. □

Este lema permite entonces generalizar los métodos de obtención de valores críticos dependientes de los datos a cualquier orden de dominancia, siendo las adaptaciones para las diferentes condiciones de modularidad superior bastante evidentes.

De esta manera, la familia de estadísticos de prueba de Barret y Donald [15] queda generalizada al caso multivariado y la validez del test extendida al caso no necesariamente continuo para cualquier orden de dominancia.

5.4. Procesos Dependientes

Tanto la teoría de procesos empíricos, como en particular el estudio del problema de dos muestras y los estadísticos de Kolmogorov-Smirnov, se tratan típicamente en el caso de muestras iid. Existen sin embargo casos en los cuales es deseable no hacer dicho supuesto, y permitir una dependencia entre las observaciones, ya sea porque se ha recurrido a técnicas de muestreo que así lo imponen o porque la dependencia es un aspecto fundamental que no quiere perderse al procesar los datos (e.g. si son series temporales).

El trabajo de Linton et al. [131] presenta tests de dominancia univariados bajo diversos esquemas de muestreo, lo cual implica distintos tipos de dependencia. También permite la dependencia entre *prospectos*, es decir, entre las distribuciones a comparar, lo cual lo hace interesante por ejemplo al estudiar distribuciones antes y después de impuestos. La principal diferencia con nuestro enfoque es que los autores suponen condiciones muy exigentes de suavidad sobre la F, por ejemplo diferenciabilidad. Por otra parte, no trabajan con bootstrap sino con otros métodos de remuestreo.

Klecan, McFadden y McFadden [118] desarrollaron tests a primer y segundo orden con dependencia, entre prospectos y entre las muestras, obteniendo valores críticos mediante una simulación de MonteCarlo siempre en el caso univariado.

Nosotros vamos a presentar dos tipos de dependencia de las variables que conforman la muestra. La primera es el caso más favorable de dependencia, principalmente para el caso de *tests de permutaciones* y se conoce como *intercambiabilidad*. Los muestreos aleatorios simples *sin* reposición de poblaciones finitas dan lugar a este tipo de dependencia (Kochar & Korwar [119]).

El otro tipo de dependencia que vamos a tratar, es el caso de los *procesos de mezcla* que son sucesiones en las cuales la dependencia se concentra en un unos pocos lags de la variable X_i. Es decir, procesos en los cuales la correlación de las variables va decayendo a medida que pasa el *tiempo* o lo que es lo mismo, a medida que nos alejamos de un subíndice fijo.

Esta propiedad es típica de las series temporales, y en este sentido es útil extender los tests a dicho caso. La noción de dominancia estocástica como mayor o menor *desigualdad económica* puede no tener sentido en este caso de manera directa, pero los tests que permiten diferenciar una sucesión de otra serán herramientas útiles para señalar, por ejemplo, cambios de la estructura subyacente.

Por otra parte, en un enfoque de *ingreso permanente* o de *LCH* puede pensarse que un patrón de ingresos que domina estocásticamente a otro proporciona un mayor bienestar por ser menores las disparidades de ingresos a lo largo del ciclo vital.

5.4.1. Simetrías Probabilísticas y Generalización del Test de Bickel

Los resultados clásicos de convergencia del proceso empírico son válidos bajo la condición de *independencia* de la sucesión de elementos aleatorios $X_1, X_2, \ldots$. En situaciones más generales, la *invariancia* implícita en el teorema de Donsker no se mantiene, y la convergencia del proceso empírico a un puente browniano ajustado no está garantizada.

Como consecuencia, resultados que dependen fuertemente de la convergencia del proceso empírico, como puede ser el teorema bootstrap de Giné-Zinn, no valen de manera universal para procesos $\{X_i\}$ con dependencia genérica.

La clave para una generalización del resultado está en la forma en que la dependencia incide sobre las condiciones que hacen de una cierta clase de funciones $\mathscr{F}$ una clase de Donsker. Esto puede sonar tautológico, pero a lo que nos referimos es al hecho de que las técnicas usuales de procesos empíricos funcionan bien cuando las sucesión de funciones indicadoras $\{1_{(s,t]}(X_i)\}_{i\in\mathbb{N}}$ hereda buenas propiedades del proceso original $\{X_i\}_{i\in\mathbb{N}}$.

Esta situación se da, por ejemplo, en los casos de procesos de *mezcla* (Sec.5.4.2) porque entonces la sucesión $\{1_{(s,t]}(X_i)\}_{i\in\mathbb{N}}$ tiene la misma propiedad. En casos más generales de procesos de Markov y sistemas dinámicos es necesario ser más exigente con la clase de funciones (cfr. Dehling et al. [48, 49]).

El caso más favorable de dependencia, la condición de *intercambiabilidad* (*exchangeability*), ofrece la posibilidad de generalizar los resultados de convergencia del proceso empírico de manera bastante directa.

El trabajo de Aldous [4] trata con los *procesos de urna*[6] y los trabajos de Berti et al. [22, 23] generalizan dicho enfoque para procesos idénticamente distribuidos en general, obteniendo para el caso concreto de la intercambiabilidad un resultado básico al estilo de Donsker.

Definición 5.4.1.1 (Procesos Intercambiables). *Un proceso estocástico $\{X_i\}_{i\in\mathbb{N}}$ es* intercambiable *si $\{X_i\}_{i\in\mathbb{N}} \sim \{X_{k_i}\}_{i\in\mathbb{N}}$ para toda permutación finita $k_1, k_2, \ldots$ de los naturales.* □

Dicho de otra forma, un proceso intercambiable es aquel en el cual la distribución conjunta de un subconjunto de $(X_{k_1}, \ldots, X_{k_m})$ de esas variables depende sólo de m y no de cuales sean los k_i.

La condición de intercambiabilidad es una de las *simetrías probabilísticas* básicas y el resultado más conocido respecto a ella es el famoso Teorema de de Finetti: *una sucesión es intercambiable sii es una mezcla de variables iid*. Puesto de otra forma, una sucesión es intercambiable si es iid condicionalmente respecto a alguna distribución subyacente (Kallenberg [115, 116], Chow [39]).

[6] El ejemplo más elemental de sucesión intercambiable no iid es la conocida *urna de Polya*: una urna con bolas blancas y rojas, retirando sin reposición, $X_n = 1$ si la n-ésima es blanca, 0 si roja. Claramente las X_n no son independientes, pero cualquier subconjunto de tamaño k tiene la misma distribución.

Este tipo de procesos está asociado en estadística con el *muestreo aleatorio simple* y se entiende típicamente en el sentido de que las muestras futuras se comportan como las muestras presentes [39].

Para entender como se aplica la idea de intercambiabilidad en nuestro trabajo, debemos comenzar observando que el proceso de bootstrap de dos muestras está íntimamente relacionado con el *proceso empírico de permutación*.

De hecho, si miramos los edps que presentamos en la Sec.5.3, vemos que son típicas expresiones de los tests de permutaciones del estilo de Fisher. No podemos dejar de notar además que, en el bootstrap de dos muestras, remuestreamos también de la *muestra agrupada* como en dichos tests.

Más concretamente (van der Vaart y Wellner [228], 3.7) los tests derivados del proceso empírico de permutación y del bootstrap de dos muestras tienen el mismo tipo de comportamiento asintótico, y en este sentido es de esperar que compartan propiedades estadísticas equivalentes.

Pero la teoría estadística señala que los tests de permutaciones son exactos para observaciones intercambiables [87][7]. Esto se debe esencialmente a que la distribución empírica es *suficente* (en el sentido estadístico) para cualquier conjunto de medidas intercambiables (Dudley [62], Ch.5).

El trabajo de Bickel [25] introduce el test de Kolmogorov-Smirnov en el caso multivariado para distribuciones subyacentes continuas, y su estructura es la de un test de permutaciones. El proceso empírico de permutaciones (van der Vaart y Wellner [228], 3.6) es básicamente una forma de extender dicho resultado al caso más general.

Estudiando la demostración del Test de Bickel [25] entonces, podemos ver que el planteo general no requiere necesariamente más que la intercambiabilidad, siendo ineludible la independencia sólo en la Sec.3.7 (el Teo.3.2) de dicho paper[8], donde se hace uso del Teorema de Glivenko-Cantelli.

Ahora bien, nosotros sabemos que la condición de Glivenko-Cantelli es corolario del Teorema de Donsker, y por lo tanto, si podemos afirmar la convergencia del proceso empírico para el caso de la dependencia intercambiable, tendremos que el resultado de Bickel es válido para dichos procesos[9].

El trabajo de Berti et al. [22] brinda condiciones para la convergencia del proceso empírico en el caso dependiente pero con distribución idéntica. Para un proceso estocástico $\{X_i\}_{i\in\mathbb{N}}$, adaptado a un filtrado $\mathscr{G} = \{\mathscr{G}_i\}_{i\in\mathbb{N}}$, dicho trabajo demuestra primero un *ley de los grandes números*, que garantiza para toda función f tal que $E(|f(X_1)|) < \infty$ la existencia de una variable aleatoria V_f tal que

[7] *"Permutation tests are exact only if the data points that are rearranged are exchangeable under the null hypothesis, that is, if the joint distribution of the observations remains unchanged under rearrangements of the data labels when the null hypothesis is true."* (P. Good [87]).

[8] En el Apéndice A damos algunos detalles más sobre la construcción básica del test de Bickel.

[9] De manera más restrictiva, el teorema de Glivenko-Cantelli se puede probar para procesos estacionarios ergódicos [48].

$E(f(X_{n+1})|\mathscr{G}_n) \to V_f$ (Berti et al. [22], Lemma 2.1).

A partir de allí, definen el operador W, equivalente al proceso empírico para el caso dependiente idéntico[10]:

$$W_n f = \frac{1}{\sqrt{n}}(\sum_{i=1}^{n} f(X_i) - nV_f) \tag{5.70}$$

y prueban el siguiente teorema, que esencialmente es un teorema de Donsker para este caso:

Teorema 5.4.1.1 (Convergencia del Proceso Empírico Intercambiable)**.** *Sea $\{X_i\}_{i\in\mathbb{N}}$ un proceso estocástico intercambiable, con medida P y sea G_P una versión del puente browniano ajustado para el espacio con medida de probabilidad P. Entonces:*

$$W_n \rightsquigarrow G_P \tag{5.71}$$

Demostración. Es el Teo.4.5 de Berti et al. [22]. □

El Teo.5.4.1.1 es la clave para generalizar nuestros tests al caso intercambiable. La clave está en que el hecho de que la estructura general del test de permutaciones usa la condición de intercambiabilidad antes que la independencia de las variables.

Corolario 5.4.1.1 (Tests de Dominancia, Caso Intercambiable)**.** *Los Teo.5.3.4, 5.3.6 y 5.3.8 son válidos para procesos generadores de datos intercambiables.*

Demostración. La demostración del proceso empírico de permutaciones de van der Vaart y Wellner [228] y la del bootstrap de dos muestras (Sec.3.6 y 3.7 respectivamente) se pueden adaptar usando el resultado de Berti et al. (Teo.5.4.1.1).

Concretamente, la desigualdad de Hoeffding sigue pudiendo aplicarse (van de Geer [227]) para encontrar las marginales y la desigualdad de poissonización y el teorema del límite central multiplicador permiten probar la convergencia a un proceso gaussiano ajustado. Respecto a este último teorema (Teo.2.9.2 de [228]), nótese que el resultado recién enunciado (Teo.5.4.1.1) de Berti et al. se aplica para generalizarlo al caso intercambiable.

□

5.4.2. Procesos α, β-mixing

La condición de *intercambiabilidad* que usamos en la Sec.5.4.1 implica que las muestras futuras se comportarán como las muestras presentes. Como resultado de ello, dichos procesos resultan estacionarios y el muestreo a partir de ellos da

[10] Claramente para el caso independiente 5.70 es el proceso empírico. Para el caso dependiente, sin embargo, la medida empírica va a incorporar los efectos de la dependencia y no va operar directamente como el W_n (póngase el caso extremo en que sea la misma variable repetida, y la medida se concentra en la diagonal).

buenas propiedades a las medidas empíricas derivadas, especialmente en el caso de tests de permutaciones.

El siguiente caso más favorable para el estudio de las medidas empíricas es aquel en que la dependencia se va debilitando a medida que pasa el tiempo. Algo así como:

$$E|E(X_{k+n}|X_1, X_2, ..., X_k)| \xrightarrow[n\to\infty]{} 0$$

Esta condición no caracteriza todos los procesos débilmente dependientes, pero da una idea del comportamiento general. Dentro de estos, la clase de los *procesos de mezcla* (*mixing processes* [27]) se diferencian de acuerdo a la forma específica en que dicha dependencia se va debilitando, y se los etiqueta como α, β, ϕ(etc)-mixing.

Esta propiedad (debilitamiento progresivo de la dependencia) es típica, por ejemplo, de las series temporales y ha dado lugar al desarrollo del método de *bootstrap en bloque* (Efron & Tibshirani [66], Bühlmann [35]).

El trabajo de Pranab Sen [200] generaliza el de Billingsley [27] y presenta un resultado de convergencia del proceso empírico para el caso multivariado, de vectores ϕ*-mixing*, permitiéndole computar cotas para las colas de la distribución de estadísticos del tipo de Kolmogorov-Smirnov.

Si bien el resultado de convergencia es similar al del proceso empírico independiente en su estructura básica, el efecto de la dependencia se presenta en las *covarianzas* del proceso ajustado que es el límite. El paper de Tae Yoon Kim [117] relaja las condiciones de Sen [200] sobre la sucesión de constantes de mixing y obtiene también colas de la distribución del estadístico KS. En ambos casos se trata de resultados válidos con supuestos de continuidad de las distribuciones subyacentes.

Un enfoque más relacionado con el nuestro es el que tienen los trabajos de Dehling et al. [48, 49], Radulović [175] y Kosorok [124], que estudian la teoría de procesos empíricos para el caso dependiente.

La idea básica detrás de este enfoque es replicar los resultados obtenidos en el caso independiente o intercambiable, es decir, comenzar con resultados de convergencia del proceso empírico a un proceso gaussiano ajustado (*à la Donsker*) y luego probar un teorema de consistencia del bootstrap (*à la Gine-Zinn*).

Por supuesto el *proceso límite* en el caso dependiente tendrá una estructura de covarianzas más complicada que el clásico, en algunos casos intratable desde el punto de vista teórico. Pero si el bootstrap es consistente tendremos una herramienta para obtener distribuciones dependientes de los datos que repliquen el comportamiento de la distribución subyacente con un grado arbitrario de precisión.

Comencemos definiendo los distintos tipos de procesos de mezcla según la literatura (Dehling et al. [48]):

Definición 5.4.2.1 (Procesos de Mezcla)**.** *Sea* $\{X_k\}_{k\in\mathbb{N}}$ *un proceso estocástico y llamemos* $\mathscr{M}_a^b$ *a la sigma álgebra generada por los elementos aleatorios* $X_a, ..., X_b$.

i. El proceso se llama α-mixing si:

$$\alpha(n) = \sup_{k\geq 1}\left\{|P(A\cap B) - P(A)P(B)| : A \in \mathscr{M}_1^k, B \in \mathscr{M}_{k+n}^{\infty}\right\} \xrightarrow[n\to\infty]{} 0 \quad (5.72)$$

ii. El proceso se llama β-mixing si:

$$\beta(n) = E\left(\sup_{k\geq 1}\left\{|P(B|\mathscr{M}_1^k) - P(B)| : B \in \mathscr{M}_{k+n}^{\infty}\right\}\right) \xrightarrow[n\to\infty]{} 0 \quad (5.73)$$

iii. El proceso se llama ρ-mixing si:

$$\rho(n) = \sup_{k\geq 1}\left\{\frac{|EXY|}{\parallel X \parallel_2 \parallel X \parallel_2} : X \in L^2(\mathscr{M}_1^k), Y \in L^2(\mathscr{M}_{k+n}^{\infty}), EX = EY = 0\right\} \xrightarrow[n\to\infty]{} 0 \quad (5.74)$$

iv. El proceso se llama ϕ-mixing si:

$$\phi(n) = \sup_{k\geq 1}\left\{\frac{|P(A\cap B) - P(A)P(B)|}{P(A)} : A \in \mathscr{M}_1^k, P(A) > 0, B \in \mathscr{M}_{k+n}^{\infty}\right\} \xrightarrow[n\to\infty]{} 0 \quad (5.75)$$

v. El proceso se llama ψ-mixing si:

$$\psi(n) = \sup_{k\geq 1}\left\{\frac{|P(A\cap B) - P(A)P(B)|}{P(A)P(B)} : A \in \mathscr{M}_1^k, P(A) > 0, B \in \mathscr{M}_{k+n}^{\infty}, P(B) > 0\right\} \xrightarrow[n\to\infty]{} 0 \quad (5.76)$$

□

Nótese que el coeficiente ρ es el *coeficiente de correlación máxima*. Por otra parte, no es difícil probar (Dehling [48]) que $\alpha(n) \leq \beta(n) \leq \phi(n) \leq \psi(n)$ lo cual implica que un proceso ψ-mixing es ϕ-mixing, y éste a su vez β-mixing, etc.

Recordemos que llamamos *estacionario* al proceso $\{X_k\}_{k\in\mathbb{N}}$ si dados $m_1 \leq m_2 \leq ... \leq m_k \in \mathbb{N}$, la distribución conjunta de $(X_{m_1+j}, ..., X_{m_k+j})$ es la misma para cualquier entero $j \geq -m_1 + 1$. Supongamos que el proceso $\{X_k\}_{k\in\mathbb{N}}$ toma valores en un espacio polaco con distribución P y llamemos $\mathbb{G}_n f = \sqrt{n}\sum_{i=1}^{n}(f(X_i) - Pf)$.

El siguiente teorema (que tomamos de Kosorok [124], Ch.11) nos brinda un resultado de convergencia del proceso empírico para un tipo de procesos β-mixing. Resultados similares se obtienen para las demás familias de procesos (Dehling et al. [48], Ch.1).

Teorema 5.4.2.1 (Convergencia del Proceso Empírico, Caso β-Mixing)**.** *Sea $\{X_k\}_{k\in\mathbb{N}}$ un proceso estacionario en un espacio polaco con distribución marginal P y sea $\mathscr{F} \subseteq L^2(P)$. Supongamos que $\exists p \in (2,\infty)$ tal que:*

$$\sum_{k=1}^{\infty} k^{2/(p-2)}\beta(k) < \infty \quad (5.77)$$

$$J_{[]}(\infty, \mathscr{F}, L^p(P)) < \infty \tag{5.78}$$

Entonces: $\mathbb{G}_n \rightsquigarrow \mathbb{H}$ *en* $l^\infty(\mathscr{F})$*, donde* $\mathbb{H}$ *es un proceso gaussiano ajustado con media cero y covarianzas:*

$$\Gamma(f,g) = \lim_{k\to\infty} \sum_{i=1}^{\infty} cov(f(X_k), g(X_i)) \ \ \forall f,g \in \mathscr{F} \tag{5.79}$$

Demostración. Es el Teo.11.22 en Kosorok [124]. □

Una vez que tenemos un resultado de convergencia del proceso empírico, la consistencia del boostrap es el próximo paso. Ahora bien, cuando existe dependencia entre las observaciones, la técnica estándar del boostrap de Efron (remuestrear aleatoriamente de la mustra original) no tiene sentido, puesto que en el remuestreo se perdería la estructura de correlaciones de las variables.

La técnica de *bootstrap en bloques* (*moving blocks bootstrap* o *MBB*) permite mantener la estructura de dependencia para los casos de procesos de mezcla. La misma (Bühlmann [35]) comienza con una muestra $X_1, X_2, ..., X_n$ a la que partimos en bloques de tamaño b:

$$B_1 = \{X_1, X_2, ..., X_b\} \ \ B_2 = \{X_1, X_2, ..., X_{b+1}\} \ ... \ B_{n-b+1} = \{X_{n-b+1}, X_2, ..., X_n\} \tag{5.80}$$

Ahora repetimos el proceso de remuestreo aleatorio iid del bootstrap de Efron, pero no sobre la muestra original sino sobre el conjunto de bloques $\{B_1, ..., B_{n-b+1}\}$. Muestreamos $m = n/b$ bloques y armamos una muestra de bootstrap concatenando los bloques obtenidos

$$\{X_i\}_{i=1}^n = B_1^*, B_2^*, ..., B_m^* \tag{5.81}$$

Los siguientes resultados de Radulović [175] nos brindan lo pedido para los casos α y β-mixing. Recordemos la notación $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ y para la muestra de bootstrap $P_n^* = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^*}$, en este caso obtenida por el método *MBB*.

Teorema 5.4.2.1 (Consistencia del Bootstrap, Caso α-Mixing)**.** *Sea* $\{X_i\}_{i\in\mathbb{N}}$ *un proceso* α*-mixing tal que* $\alpha(n) = O(n^{-\eta})$ *con* $\eta > 1$*, y elijamos para cada n el tamaño de bloque b de forma que* $b = O(n^{\varepsilon})$ *con* $0 < \varepsilon < \text{mín}\left(\frac{1}{3}, \frac{\eta}{4} - \frac{1}{6}\right)$*. Entonces:*

$$\sqrt{n}\left(P_n^* 1_{\{x\le t\}} - P_n 1_{\{x\le t\}}\right)_{t\in\mathbb{R}^n} \rightsquigarrow G_P 1_{\{x\le t\}} \tag{5.82}$$

en probabilidad.

Demostración. Es el Teo.2.4 en Radulović [175]. □

Teorema 5.4.2.1 (Consistencia del Bootstrap, Caso β-Mixing)**.** *Sea* $\mathscr{F}$ *una clase de funciones VC con envolvente F tal que* $EF^p < \infty$ *para algún* $p > 2$*. Sean* X_i^* *las muestras de bootstrap en bloque con tamaño de bloque* $b = O(n^\rho)$ *tal que*

$0 < \rho < \frac{p-2}{2(p-1)}$. *Si el proceso original* X_i *es* β*-mixing con* $\beta(i) = O(i^{-q})$ *para un* $q > \frac{p}{p-2}$. *Entonces:*

$$\sqrt{n}(P_n^* f - P_n f)_{f\in\mathscr{F}} \rightsquigarrow G_P(f)_{f\in\mathscr{F}} \tag{5.83}$$

en probabilidad.

Demostración. Es el Teo.2.5 en Radulović [175]. □

Ahora en base a estos resultados, podemos intuir cierto formato para un test de dominancia estocástica basada en estadísticos de Kolmogorov-Smirnov. La idea es rehacer lo que hicimos para el bootstrap de dos muestras independientes, pero utilizando la técnica *MBB*.

Debe notarse, sin embargo, que el método tal y como lo planteamos no tiene una consistencia demostrada por la sola aplicación de los Teo.5.4.2.1,5.4.2.1. Hacen falta posteriores desarrollos de la teoría matemática de los procesos empíricos para una justificación rigurosa del mismo.

En este sentido, es que llamamos *heurística* a la construcción, confiando en la posible extensión de los resultados de Radulović al problema de dos muestras, aunque en este caso no contamos con una modificación directa de las demostraciones, porque al no ser el caso *intercambiable*, hay muchos elementos que no permanecen invariantes al introducir la dependencia.

Heurística 5.4.2.1 (Test de Dominancia, Caso α,β-mixing)**.** *Supongamos dadas muestras* $(X_1,...,X_m)$ *y* $(Y_1,...,Y_n)$. *Elegimos dos tamaños de bloque* b_1, b_2 *y formamos* $m - b_1 + 1$ *bloques* B_i^X *y* $n - b_2 + 1$ *bloques* B_j^Y. *Conformamos ahora una* muestra agrupada de bloques *(*pooled block sample*) con todos ellos y tomamos elementos de ella al azar.*

Con los primeros m/b_1 *bloques sampleados de la muestra agrupada conformamos una muestra de bootstrap* $\{X_i^*\}_{i=1}^m$ *y con los segundos* n/b_2 *conformamos la muestra* $\{Y_j^*\}_{j=1}^n$.

Ahora construimos con estas muestras de bootstrap los estadísticos $\lambda_{m,n}^B$, $\kappa_{m,n}^B$, *etc, como en la Sec.5.3. Repetimos la técnica de remuestreo y utilizamos los cuantiles de la distribución de bootstrap obtenida como valores críticos dependientes de los datos.*

□

Debemos observar que el planteo es heurístico, pero así fue como nació la propia técnica del bootstrap de Efron y diversas técnicas de remuestreo y estimación estadística. Del planteo de Efron [65] hasta la rigurosa justificación de Giné & Zinn [83] pasó más de una década, y para las técnicas particulares derivadas del bootstrap (el bootstrap intercambiable, con pesos, etc) aún más (Giné [81]).

En este sentido, es necesario tener en cuenta que la teoría de procesos empíricos dependientes es una rama muy reciente de la ciencia probabilística. El primer tomo que reune diversos resultados (ni siquiera un libro de texto) es el de Dehling et al. [48] y desde allí se han ido sumando otros aportes (e.g. Dehling et al. [49], Marinucci [143]) pero de ninguna manera se tiene el amplio arsenal de resultados que conocemos para el caso independiente, resumidos por ejemplo en van der Vaart y Wellner [228], van de Geer [226] o Wellner [232].

Por ello es lógico esperar avances en la materia, sobre todo teniendo en cuenta que para nuestro caso contamos con factores bastante favorables. Por ejemplo, la clase de funciones que utilizamos para el bootstrap de dos muestras, las indicadoras de rectángulos en $\mathbb{R}^n$ es una de las que presenta mejores propiedades (es obviamente VC, es Donsker en el caso independiente, etc.).

Entre dichas propiedades no podemos dejar de destacar que se cumplen teoremas de convergencia del proceso empírico para distintos procesos de mezcla (Teo.5.4.2.1) y para el bootstrap (Teo.5.4.2.1, Teo.5.4.2.1), por lo cual es de esperar un buen comportamiento del método bajo las condiciones de mixing (y entropía, por esto es buena nuestra $\mathscr{F}$) de dichos resultados.

Por último, como señalaremos en la Sec.5.5.1, el propio método de bootstrap nos permite muchas veces hacer una valoración (*assessment*) de la potencia de un test por simulación.

5.5. Propiedades y Ventajas del Herramental Desarrollado

5.5.1. Propiedades Estadísticas

Consistencia: los valores críticos dependientes de los datos obtenidos para nuestros tests por bootstrap convergen a los cuantiles de la distribución subyacente.

Potencia: En análisis de potencia del test no es tan sencillo, porque tanto la hipótesis nula ($H_0 : PSD_jQ$) como la alternativa (H_1 :*P no domina a Q*) son compuestas. No podemos entonces definir de manera inmediata una probabilidad de aceptar H_1 porque no sabemos la ley subyacente.

Una técnica usual en el estudio de la potencia de un test es la de las *hipótesis contiguas*. Estas son hipótesis alternativas Q_n que asintóticamente son absolutamente continuas (en el sentido de Radon-Nykodim) respecto a la ley subyacente bajo H_0, algo que vale por ejemplo si en la distancia de Heilinger $d_h(P,Q_n) \to 0$ (van der Vaart y Wellner [228]).

En la práctica sin embargo, los resultados asintóticos obtenidos de este forma[11] son muy difíciles de evaluar de manera explícita.

Pero como señalan Efron y Tibshirani [66]:

> *"Bootstrap tests are useful in situations where the alternative hypothesis is not well-specified"* (Efron y Tibshirani [66]).

La idea es entonces, siguiendo a Philip Good [88] hacer un *assessment local* de potencia de la siguiente forma: partiendo de una distribución base P, la modificamos de forma tal sepamos que la nueva ley Q es tal que no se cumple $H_0 : PSD_jQ$. Si la modificación no es demasiado grande, estaremos con una ley lo suficientemente cercana, pero a la vez distinta, como para poder evaluar la potencia del test.

La mecánica es: (i) tomamos P; (ii) modificamos y obtenemos Q tal que no vale H_0; (iii) muestreamos de cada una y tenemos P_n y Q_n; (iv) implementamos el test de dominancia al orden que queramos; (v) repetimos un cierto número de veces el procedimiento, la potencia estimada será la proporción de veces que rechazamos H_0.

Dada una P, en principio no parece trivial decidir cómo modificarla para obtener una Q tal que podamos asegurar de manera efectiva que no está dominada por P. Sin embargo, podemos usar el concepto de *transformación*

[11] En van der Vaart y Wellner [228] p.408 se obtienen la potencia del test KS de *una muestra*. Estos no nos sirven, porque nosotros no tenemos una hipótesis con una ley fija, pero además como señalan los autores, el resultado no se puede evaluar explícitamente ni siquiera en este caso sencillo.

de correlación creciente de Epstein y Tanny [67] para modificar a P y obtener la Q deseada.

Nótese que la P dada y la Q así obtenida son las *distribuciones exáctas subyacentes*. Ahora lo que hacemos es muestrear a partir de ellas y obtener P_n y Q_n y luego aplicar el test de dominancia a estas muestras.

Trabajamos con una distribución normal bivariada simulada en Matlab con 40000 observaciones, y aplicamos una serie de transformaciones de correlación creciente que serían como las de 4.2 pero con $-\varepsilon$. En la práctica en realidad lo que hacemos es mover puntos de forma que la frecuencia bivariada cambie de la forma requerida, y luego calculamos el ε, con el programa *corrinc2.m*, Ap.C. Típicamente cambiamos el entre el 1 % y 2 % de los puntos, para ser estrictos en la localidad de la evaluación de potencia.

Muestreamos 20000 de cada una para obtener P_n y Q_n, testeamos con un nivel $\alpha = 0{,}05$. Repitiendo este procedimiento unas 50 veces, encontramos que para el test a primer orden se rechaza alrededor de un 70 % de las veces, mientras que el test de segundo orden es un poco más potente. Estos están muy cerca del 80 % de potencia que se considera ideal en diseño de experimentos (Suresh y Chandrashekara [216]).

Nótese que tanto Bickel [25] como Barret y Donald [15] prueban para sus tests que $P\{\textit{rechazar } H_0 \textit{ si es falsa}\} \to 1$ asintóticamente. Pero en ambos casos la demostración depende esencialmente de la continuidad de las distribuciones subyacentes.

Caso Dependiente: Los test de dominancia valen para el caso dependiente intercambiable. Nótese que Linton et al. [131] prueban resultados para dependencias más generales, pero sólo en el caso univariado y con hipótesis muy restrictivas sobre la suavidad de las F.

Para el caso de procesos de mezcla creemos que es razonable esperar que la heurística funcione bien, dada la convergencia del proceso empírico y la consistencia del bootstrap. En la práctica, se puede usar y confiar en un assessment de potencia por bootstrap similar al que describimos en el item anterior, adaptándolo al método 2*SMBB*.

5.5.2. Ventajas Sobre el Test de Crawford (2005)

Comparación Punto a Punto: El test de Crawford [43] se basa en una generalización bivariada del test de Anderson [7]. El mismo se basa en un test de bondad de ajuste de Pearson, pero trabaja sobre datos *agrupados* en percentiles. Es decir, se comparan las funciones de distribución y sus integrales en un número finito de puntos fijos. Esto reduce la complejidad del problema pero como señalan Davidson y Duclos [45]y Barret y Donald [15, 16] los tests basados en datos agrupados implican una probabilidad de inconsistencia.

Por el contrario, nuestros tests se basan en la comparación punto a punto de las funciones de distribución y resultan entonces más consistentes. Por lo mismo, resultan interpretables de manera directa en términos de la economía del bienestar, mientras que los tests basados en datos agrupados no lo son [199].

Hipótesis de Suavidad: el test de Crawford sólo es válido para funciones de distribución suaves, al menos con derivadas de segundo orden continuas. Nuestros tests no están limitados de esa forma.

5.5.3. Ventajas de Aplicabilidad

Simpleza: Tanto los edps como el método de bootstrap para obtener valores críticos son fáciles de implementar mediante rutinas en cualquier paquete estadístico o matemático como Matlab o Stata.

Sumado a la directa interpretación del resultado en términos económicos creemos que resultan herramientas prácticas y útiles para el econometrista empírico y el economista aplicado.

Capítulo 6

Aplicación: Argentina 1996-2006

6.1. Función Minceriana No Lineal

La Ec.2.11 implica que los retornos a la educación son los mismos para todos los años pasados en la escuela. En la práctica sin embargo, es usual que los mismos varíen, típicamente de manera que el retorno marginal es decreciente.

La forma general de las funciones mincerianas no lineales fue introducida por Bils y Klenow [28] en su estudio de la relación entre educación y crecimiento. Para un trabajador de edad a con s años de escuela se especifica:

$$h_i(a) = h_i(a+n)^{\phi} e^{f(s)+g(a-s)} \qquad \forall a > s \tag{6.1}$$

o tomando logaritmos:

$$\log h_i(a) = \phi \log h_i(a+n) + f(S) + g(a-S) \tag{6.2}$$

Nótese que el capital humano de un trabajador de edad a a depende del $h_t(a+n)$; esto incorpora el efecto del capital humano de los *maestros* en la función.

Para dar una especificación concreta tenemos que fijar un valor para el parámetro ϕ y dar una forma concreta para f y para g. Nótese que el caso con $\phi = 0$, $f(S) = r_s S$ y g un polinomio de grado 2 es justo la Ec.2.11.

Las formas más usualmente adoptadas por los investigadores empíricos [153] para f y g son:

$$g(a-s) = \gamma_1(a-s-6) + \gamma_2(a-s-6)^2 \tag{6.3}$$

$$f(s) = \frac{\alpha}{1-\psi} s^{1-\psi} \tag{6.4}$$

con $\psi > 0$ para dar retornos decrecientes. También es usual que se tome $\phi = 0$.

6.2. Fuentes de Datos y Variables Utilizadas

Nuestro objetivo en este Capítulo es aplicar los de tests *bivariados* de dominancia estocástica desarrollados en el Cap.5 para encontrar rankings de bienestar. La serie de muestras que compararemos será la de las ondas de la EPH para los años 1996 a 2002 en el conjunto de los aglomerados urbanos.

De acuerdo a lo que hemos desarrollado a lo largo de la Tesis, el par de variables que hemos elegido para implemetar el test ha sido el del *ingreso individual* y_i y el capital humano h_i.

La fuente de datos para el estudio de la desigualdad económica en Argentina ha sido la Base Usuaria de la EPH del INDEC, en el período de 1996 a 2006. Las razones de tal elección de período son diversas. Por un lado, para dicho rango se tiene la disponibilidad en internet de las bases usuarias brindadas por el INDEC para el total de aglomerados urbanos del país.

Debemos justificar por qué elegimos cortar la serie con el fin de 2006. Entre los años 2007 y 2010 la EPH no fue publicada por el INDEC, debido a que su directora se econtraba bajo *licencia psiquiatrica*. Si bien en 2010 se publicaron resultados retroactivos de la EPH para los años anteriores (para los cuales estimamos que las bases usuarias deben estar disponibles a pedido o a medida que se suban a la web del instituto) hemos de todas formas optado, por razones metodológicas, por no considerar la serie desde 2007 en adelante. Ello se debe a que la veracidad y confiabilidad de los datos de la misma se ha visto seriamente comprometida, no sólo por la situación irregular de la EPH, sino por la de todo el INDEC[1].

En dichas circunstancias, lamentablemente, elegimos preservar la representatividad de los datos y restringir el rango temporal de la investigación. Eso resulta muy lamentable porque justamente desde 2007 se han venido presentando fenómenos (alta inflación, restricciones a la actividad comercial, planes asistenciales, etc.) cuyo impacto en la desigualdad sería muy interesante observar.

Para construir la variable ingreso trabajamos como Perez [166]. La variable que hemos utilizado como ingreso es $y_i = \log Y_i$, donde $Y_i = \frac{ITF}{\sqrt{\sigma}}$. Aquí ITF es el Ingreso Total Familiar de la EPH, σ es el número de personas en el hogar ($R01$ en la EPH) y $\sqrt{\sigma}$ es el factor de normalización.

Respecto al factor de normalización, si bien en el análisis empírico de la pobreza se utiliza la suma del *factor adulto equivalente* en lugar de σ, en la práctica del análisis de la desigualdad está muy difundida la normalización dada por la *regla de la raíz cuadrada* que adoptamos aquí (Duclos, Esteban y Ray [57], OECD [162]). Ajustamos los ingresos por la serie del IPC de INDEC.

Debe tomarse en cuenta que el ingreso declarado en la EPH no tiene en cuenta muchos impuestos ni provisión de bienes públicos. En este sentido no es exáctamente el ingreso neto disponible o sea del cual se obtiene la utilidad (Gasparini

[1]Atestiguado, por ejemplo, en por el informe del consejo de universidades para el entonces Ministro de Economía Amado Boudou en 2010.

[78]). No es común tener en cuenta estos efectos en la práctica.

Con respecto a la variable *capital humano*, adoptamos el enfoque de Miyazawa Kensuke [153] en su medición del capital humano de Japón en 2011. De acuerdo con ello, transformamos los *años de educación* s_i del individuo i en capital humano h_i a partir de una función minceriana de salario no lineal como la que presentamos en la Sec.6.1:

$$\log h_i = \gamma_1(a_i - s_i - 6) + \gamma_2(a_i - s_i - 6)^2 + \frac{\alpha}{1-\psi} s_i^{1-\psi}$$

Los parámetros α y ψ los tomamos de la estimación de Bils y Klenow [28] que trabajan con una meta-regresión de los retornos a la educación para 56 países estimados por Psacharopoulos [173]. Los mismos fueron $\gamma_1 = 0{,}0512$, $\gamma_2 = -0{,}00071$, $\alpha = 0{,}32$ y $\psi = 0{,}28$.

No realizamos una estimación propia de los mismos para el caso argentino porque para ello se necesitan microdatos de cohortes, algo que la EPH no presenta. Margot [142] realiza una estimación de rendimientos a la educación aprovechando que la EPH repite cierto porcentaje de las unidades al pasar de una onda a la siguiente, pero lo hace con una especificación lineal. Nosotros tomamos los parámetros de Bils y Klenow para el caso no lineal, algo usual en la literatura (e.g. Miyazawa Kensuke [153]).

Los años de educación los definimos a partir de la variable categórica de la EPH $NIVEL-ED$ según la equivalencia que tomamos de Margot [142]:

Tabla 6.1: Conversión de *NIVEL-ED* en

Categoría	**Nivel Educativo Alcanzado**	**Años** (s_i)
1	Primaria Incompleta	4
2	Primaria Completa	7
3	Secundaria Incompleta	10
4	Secundaria Completa	12
5	Superior Incompleta	15
6	Superior Completa	17

Observemos que, en los casos en que se trabaja con microdatos de educación suele darse la peculiaridad de que una de las variables asociadas al individuo, el número de años de educación s_i tiene poca variación de una medición a la siguiente. Esto trae problemas con algúnas técnicas de estimación, como los modelos de panel con coeficientes fijos (Arellano [8]).

En nuestro caso esto no es relevante porque estamos trabajando con datos de muestras sucesivas para distintas unidades de observación, y por otro lado si bien la distribución del h puede no cambiar demasiado de una onda a la siguiente, no hay nada que diga que las *correlaciones* entre dicha variable y otra con variación rápida como el ingreso se mantengan constantes.

De la Base de datos de Individuos de la EPH seleccionamos aquellos de entre 18 y 65 años, con ingresos positivos.

Observemos que el relevamiento *puntual* de las encuestas se dio hasta la Onda Mayo 2003, y apartir de allí la EPH se relevón en *forma continua*. Más allá de la conveniencia o no de tal método de relevamiento con inflación medianamente alta, consideramos los datos comparables y a partir de 2004 utilizamos el relevamiento del primer trimestre y el del tercer trimestre en reemplazo de las ondas. Para el 3er trimestre de 2003 la base usuaria no contó con un relevamiento de todas las variables por lo cual no pudo ser utilizada.

6.3. Implementación del Test

La implementación de los tests de dominancia es bastante sencilla. Trabajamos con los edp del Cap.5 para el caso supermodular, por considerar que ingreso y capital humano son complementos estratégicos:

$$\kappa_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{[0,1]\times[0,1]} (\hat{K}_m^F - \hat{K}_n^G)(s,t) \tag{6.5}$$

$$\gamma_{m,n} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{[0,1]\times[0,1]} (\hat{L}_m^F - \hat{L}_n^G)(s,t) \tag{6.6}$$

donde 6.5 es para la dominancia a primer orden y 6.6 para segundo orden. La mecánica del test es la siguiente, que implementamos en la práctica con rutinas diseñadas en Matlab (cfr. Ap.C):

- Dadas dos muestras, es decir, los datos completos de la EPH para ingreso y capital humano tal como los definimos en la Sec.6.2, calculamos el valor del edp muestral correspondiente.
- Agrupamos las dos muestras y remuestreando aleatoriamente con reposición vamos generando pares de muestras del mismo tamaño. Calculando el edp para cada una de los pares vamos construyendo la distribución de bootstrap del mismo.
- Fijado un nivel α de significatividad, comparamos el valor crítico obtenido de la distribución de bootstrap con el calculado para el edp, rechazando la hipótesis FSD_jG si este último resulta mayor.

Trabajamos con 100 iteraciones de bootstrap y $\alpha = 0{,}05$. Dado el tamaño de las muestras (usualmente mas de 25.000 datos individuales de cada variable) las 100 iteraciones se consideran un número aceptable para una buena consistencia (Efron y Tibshirani [66]). Por otro lado, dado que el tamaño de la muestra impacta de manera no lineal sobre el número de operaciones necesarias para calcular el valor crítico, al trabajar con muestras tan grandes no es practicable usar muchas repeticiones más.

En nuestro caso, con las rutinas implementadas mediante programación en paralelo con procesadores de 2 o 4 núcleos y de 2 a 8GB de RAM, cada implementación realizando el doble contraste de FSD_jG y GSD_jF llevó alrededor de 1h. Para darse una idea de porqué esto es relevante, nótese que comparamos 21 distribuciones entre sí, dando un número total de 210 contrastes.

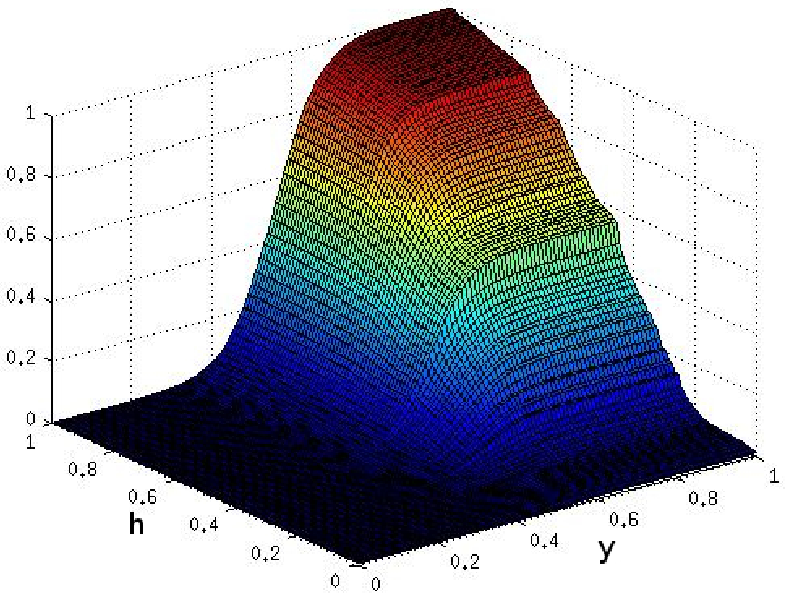

Figura 6.1: Distribución de Conjunta de Ingreso y Capital Humano. Argentina-Onda Mayo 1996.

6.4. Resultados

Las Distribuciones y sus Marginales

Antes de presentar los resultados de los tests de dominancia, podemos mirar brevemente las distribuciones obtenidas como paso intermedio. En la Fig.6.1 presentamos la cdf conjunta normalizada de ingreso y capital humano a partir de los datos de la Onda Mayo de 1996. En el mismo podemos observar como las cdf marginales son distintas, presentando la cdf del capital humano ciertos *saltos* frente a la mayor suavidad de la del ingreso.

Esto es algo esperable, por la discretización implícita en la definicion del h que hemos construido a partir de las categorías de la tabla 6.1 y se nota bien si miramos las cdfs marginales (Fig.6.2, p.147). Si se trabajara con la necesidad de distribuciones continuas, una posibilidad es la estimación no paramétrica de las densidades, que presentamos para este ejemplo en la Fig.6.3 de p.147.

Se ve claramente en la Fig.6.3 (p.147) que las densidades son muy distintas. El aspecto *ondulante* de la densidad de h se debe a cierto *oversmoothing* debido al método kds. El trabajo de Perez [167] utiliza un método de *wavelets con thresholding* para evitar esa suavización escesiva.

De cualquier forma nuestros tests no requieren como input distribuciones con-

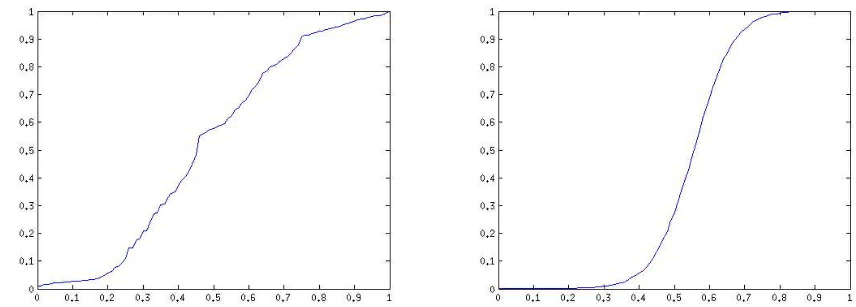

Figura 6.2: Cdfs - Capital Humano (izq.) e Ingreso (der.). Argentina-Onda Mayo 1996.

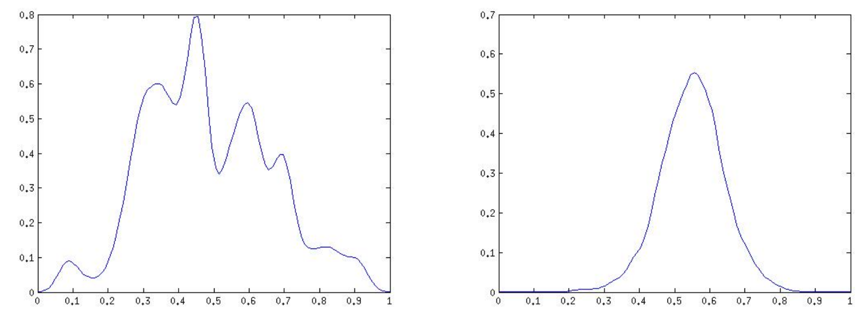

Figura 6.3: Densidad - Capital Humano (izq.) e Ingreso (der.). Argentina-Onda Mayo 1996.

tinuas así que podemos trabajar con las cdf empíricas sin inconvenientes.

Más allá del aspecto escalonado de la densidad del h, otro hecho característico que se puede ver en las densidades es que la del ingreso tiene una forma acampanada asimétrica (la cola izquierda es más ancha que la derecha) y con cierta curtosis positiva. Esto es bastante típico de las distribuciones de ingreso individual (Perez [166]).

Resultados de los Tests

Aplicamos los tests para dominancia estocástica bivariada a primer y segundo orden al conjunto de las ondas de la EPH que van de Mayo de 1996 al tercer trimestre de 2006. Cada aplicación consiste en tomar un par de distribuciones, realizar el testeo y decidir si podemos afirmar con significatividad estadística la dominancia de una sobre la otra, es decir, se testea tanto $H_0 : FSD_jG$ como $\tilde{H}_0 : GSD_jF$.

Los tablas completas de resultados con los valores de los edp y los valores críticos en cada comparación (para ambos sentidos y para las marginales) se dan en el Apéndice B.

Para presentar los resultados construimos las matrices presentadas en las tablas 6.4 de la p.152 y 6.5 de la p. 153.

En las mismas debe interpretarse un valor 1 como el hecho de que la onda representada por la fila domina a aquella representada por la columna. Un valor -1 significa que la columna domina sobre la fila. Un valor 0 resulta si no hay dominancia en uno u otro sentido o si la hay en ambos[2].

Con dichos resultados podemos construir entonces ránkings de dominancia estocástica que deben interpretarse como ránkings de bienestar social. Presentamos estos ordenamientos en las tablas 6.2 de la p.149 y 6.3 de la p.150, en las cuales trazamos líneas horizontales para agrupar las distribuciones entre las que no se estableció una dominancia definitiva a dicho orden.

Más allá de la interpretación obvia del ránking, es decir, de qué distribuciones presentan mayor bienestar social que otras, podemos realizar una serie de observaciones a partir de los resultados.

- Si bien las matriz de SD1 (Tabla 6.4, p.152) presenta cierta cantidad de 0s (descontando obviamente los de la diagonal), el hecho es que no son tantos sobre el total. Esto refleja el hecho de que la dominancia estocástica suele ser más decisiva al comparar distribuciones de un mismo país a lo largo del tiempo.

[2]Esto último en cierta forma *pierde información*, pero lo que queremos destacar es que no se puede afirmar con significatividad estadística que el bienestar de una distribución sea superior al de la otra, que es lo requerido para hacer el ranking. Añadir una notación cuaternaria sólo complicaría la comprensión sin agregar a la interpretación económica.

Tabla 6.2: Ránking de Bienestar. Dominancia Estocástica a Primer Orden.

Onda	**Ránking**
Octubre 1999	1
Octubre 1996	2
3er Trim. 2006	3
Mayo 1996	4
Mayo 1998	4
Mayo 1999	4
Mayo 2000	4
Mayo 2001	4
Octubre 2001	4
Octubre 1998	10
Mayo 1997	11
Octubre 1997	12
Octubre 2000	12
1er Trim. 2005	14
1er Trim. 2004	15
3er Trim. 2004	15
3er Trim. 2005	15
1er Trim. 2006	15
Mayo 2003	19
Mayo 2002	20
Octubre 2002	20

- La matriz de SD2 (Tabla 6.5, p.153) presenta destacados en color las comparaciones entre distribuciones en las que no fue decisiva la SD1. Se observa en este caso que el nuevo orden aumenta la decisividad, pero no totalmente, pasando a estar bien definida la dominancia en un 60 % de los casos antes indefinidos. Esto también es típico de la SD unidimensional (Perez [166]).

- **Un resultado característico muy destacable es la existencia de una estructura de *bloques* en la matriz de dominancia**. En efecto, se ve que la totalidad de las distribuciones que van desde Mayo de 2002 al primer trimestre de 2006 se encuentran dominadas por *cualquier* distribución previa a esa fecha.

  Esto debe interpretarse como un reflejo de un fenómeno que ha sido observado en el análisis de la desigualdad unidimensional del ingreso (Perez [166]) que es el de los *saltos del umbral de desigualdad* en las crisis severas como la orginada al salir de la convertibilidad.

- Un aspecto que puede ser interesante es el hecho de que a partir de un *pico de*

Tabla 6.3: Ránking de Bienestar. Dominancia Estocástica a Segundo Orden.

Onda	**Ránking**
Octubre 1999	1
Octubre 1996	2
3er Trim. 2006	3
Mayo 1999	4
Mayo 1998	5
Mayo 1996	6
Mayo 2000	6
Mayo 2001	6
Octubre 2001	6
Octubre 1998	10
Mayo 1997	11
Octubre 1997	12
Octubre 2000	12
1er Trim. 2005	14
3er Trim. 2004	15
3er Trim. 2005	15
1er Trim. 2006	17
1er Trim. 2004	18
Mayo 2003	19
Mayo 2002	20
Octubre 2002	20

desigualdad en Mayo de 2002, la desigualdad unidimensional en el ingreso va cayendo, pero las dominancias bivariadas no son igualmente monótonas. Esto impicaría un fenómeno por el cual la desigualdad en el ingreso cae pero no se incrementa su correlación con el capital humano, como puede ser un crecimiento del empleo en sectores de menor productividad como la construcción o el sector público.

- El hecho de que la distribución del 3er trimestre de 2006 esté tan arriba en el ránking puede estar reflejando un cambio del patrón o bien puede ser un outlier por razones estadísticas o de otro tipo.[3]

- Las distribuciones de Mayo y Octubre de 2002, el año de la crisis son dominadas por todas las demás. En el estudio de la desigualdad en el ingreso, Perez [166] encontraba con el test de Barret y Donald que la Onda Octubre 2002 era dominante.

[3]Si bien la intervención *de facto* del INDEC se dio en enero de 2007, las presiones en grado ascendente comenzaron desde un año antes (Francisco Jueguen [111], en La Nación).

Si bien la marginal del ingreso sigue dando una dominancia (Apéndice B), la introducción de la variable capital humano y la correlación hacen que no se pueda establecer un ránking decisivo entre ellas.

- Además de este último hay varios resultados permiten ver la diferencia entre los ránkings de dominancia unidimensional y la bidimensional, y el efecto de considerar los dos órdenes de dominancia. En el ránking SD1, la Onda Mayo de 1999 domina a la de Mayo 1996 en las márginales, pero no en la bivariada, por efecto de la correlación. Al pasar al segundo orden, la dominancia de la primera sí resulta decisiva.

- Lo mismo ocurre con la comparación Mayo 1996 vs Mayo 1998 y algunos otros casos (Apéndice B).

- En otros casos, pese a la dominancia marginal, no se da la bivariada a ninguno de los dos órdenes (e.g. Mayo 96 vs Mayo 2000).

- La comparación de Mayo 96 contra mayo de 2001 es interesante: la primera domina en el ingreso pero la segunda en el capital humano. Más allá de que la dominancia no resulta decisiva a ningún orden, uno puede pensar que una distribución menos igualitaria en capital humano conlleva una *desigualdad potencial* mayor a largo plazo.

- Son muy pocas las comparaciones en las que ninguno domina en ningún aspecto (e.g. Octubre 2000 vs Octubre 1997).

Pese los casos en que la correlación vuelve no decisiva la dominación marginal, creemos de manera general que el ordenamiento de las distribuciones bivariadas brinda una mejor información en términos de bienestar social que el de la sola distribución del ingreso.

Por supuesto, una interpretación completa de los resultados requiere un trabajo adicional de historia económica, revisando para los años cubiertos tanto los cambios de política económica como los cambios demográficos, de estructura del mercado laboral, etc. Ello permitirá evaluar los posibles mecanismos de cambio en la estructura de correlación entre y y h, así como la evaluación cuantitativa del impacto de las políticas y las tendencias demográficas, los eventos de crisis externa, etc.

Las hipótesis que pudieran plantearse en esos sentidos, se pueden comparar con los resultados de bienestar social a partir de la interpretación de los rankings obtenidos. En la presente Tesis no profundizamos sobre todos estos interesantes aspectos, sino que apuntamos más bien a mostrar como se puede utilizar el herramental econométrico desarrollado para construir un ránking de bienestar social entre las distintas distribuciones.

Onda	M96	O96	M97	O97	M98	O98	M99	O99	M00	O00	M01	O01	M02	O02	M03	P04	T04	P05	T05	P06	T06
M96	0	-1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O96	1	0	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1
M97	-1	-1	0	1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	1	1	1	1
O97	-1	-1	-1	0	-1	-1	-1	-1	-1	0	-1	-1	1	1	1	1	1	1	1	1	-1
M98	0	-1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O98	-1	-1	1	1	-1	0	-1	-1	-1	1	-1	-1	1	1	1	1	!	1	1	1	-1
M99	0	-1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O99	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M00	0	1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O00	-1	-1	-1	0	-1	-1	-1	-1	-1	0	-1	-1	1	1	1	1	1	1	1	1	-1
M01	0	1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O01	0	1	1	1	0	1	0	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
M02	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
O02	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
M03	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	-1	-1	-1	-1	-1	-1
P04	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	0	0	-1	0	0	-1
T04	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	0	0	-1	0	0	-1
P05	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	0	1	1	-1
T05	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	0	0	-1	0	0	-1
P06	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	0	0	-1	0	0	-1
T06	1	-1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	0

Tabla 6.4: **Dominancia Estocástica Bivariada a Primer Orden. Argentina 1996-2006**

Onda	M96	O96	M97	O97	M98	O98	M99	O99	M00	O00	M01	O01	M02	O02	P03	P04	T04	P05	T05	P06	T06
M96	0	-1	1	1	-1	1	-1	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O96	1	0	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1
M97	-1	-1	0	1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	1	1	1	1
O97	-1	-1	-1	0	-1	-1	-1	-1	-1	0	-1	-1	1	1	1	1	1	1	1	1	-1
M98	1	-1	1	1	0	1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	-1
O98	-1	-1	1	1	-1	0	-1	-1	-1	1	-1	-1	1	1	1	1	!	1	1	1	-1
M99	1	-1	1	1	1	1	0	-1	1	1	1	1	1	1	1	1	1	1	1	1	-1
O99	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
M00	0	1	1	1	-1	1	-1	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O00	-1	-1	-1	0	-1	-1	-1	-1	-1	0	-1	-1	1	1	1	1	1	1	1	1	-1
M01	0	1	1	1	-1	1	-1	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
O01	0	1	1	1	-1	1	-1	-1	0	1	0	0	1	1	1	1	1	1	1	1	-1
M02	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
O02	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
M03	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	0	-1	-1	-1	-1	-1	-1
P04	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	0	-1	-1	-1	-1	-1
T04	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	0	-1	0	0	-1
P05	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	0	1	1	-1
T05	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	0	-1	0	1	-1
P06	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	0	-1	-1	0	-1
T06	1	-1	1	1	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	0

Tabla 6.5: **Dominancia Estocástica Bivariada a Segundo Orden. Argentina 1996-2006**

Capítulo 7

Algunos Aspectos Adicionales en el Estudio de la Desigualdad

7.1. Desigualdad, Riesgo y Contrato Social

Hemos trabajado a lo largo de esta Tesis con la idea básica de que la desigualdad y el bienestar social generan *rankings* de sociedades (economías de distribución). Estos rankings reflejan de manera general las bases de información y los principios de *justicia* que suponemos implícitos en las *preferencias sociales*.

A la hora de *evaluar* o *comparar* las economías mediante estos enfoques, estamos comparando sus **resultados**, que son a su vez producto de la *interacción* o *juego* entre los miembros de la sociedad. Se plantea entonces la pregunta de cuán justas son las reglas de juego o mejor dicho cuál es una forma justa de *asignar* las mismas al juego.

Teorías como la de Rawls [176] y Buchanan y Tullock [34] plantean que las *reglas justas* del juego serán aquellas que surjan de un acuerdo unánime en una situación primitiva completamente igualitaria, en el sentido de una total *incertidumbre* sobre el lugar que cada individuo en la sociedad una vez adoptadas las reglas. Dichos autores presentan entonces algunas reglas que consideran aptas para representar dicho concepto de justicia.

En nuestro estudio, vamos a interpretar el *bienestar social* como un orden de preferencias sobre las distintas *sociedades posibles* en este estado de absoluta incertidumbre o aleatoriedad. Vale decir, dadas economías de distribución E_1 y E_2 podemos juzgar que $W(E_1) > W(E_2)$ si un individuo, puesto a elegir entre un *boleto de loteria* que represente la posibilidad de *ser* cualquiera de los individuos de la primera economía u otro boleto correspondiente a la segunda, elige el primero. Nótese que estamos hablando de *un* individuo, que podemos interpretar representativo como al trabajar con bienestar social usabamos una única u en la agregación, y también porque es una situación *a priori* en la que le puede tocar cualquier lugar en la sociedad con la misma probablilidad.

Vamos a hacer claro esto con un ejemplo. Supongamos que la economía E_1 tiene 10 individuos y la siguiente distribución del ingreso: $x_k = 3$ para $k = 1,...,10$, mientras que E_2 también con 10 individuos tiene la distribución $y_1 = 0$; $y_2 = y_3 = 1$, $y_k = 4$ para $k = 4,...,10$.

Se le ofrece ahora al individuo un boleto de lotería T_1 tendrá el ingreso x_k con probabilidad 1/10 y otro T_2 donde tendrá y_k con probabilidad 1/10. Se le da a elegir entre ellos, es decir, se le deja elegir en qué economía prefiere ser un *individuo al azar*. La idea es que si uno prefiere ser un individuo al azar en una economía antes que en otra, el bienestar social se considera mayor.

¿Cómo elige entre los boletos? Si piensa según las ideas de la introducción entonces puede construir un ranking de bienestar a partir de una medida de bienestar social o de un índice de desigualdad. Si elige el orden de Lorenz, como las economías tienen el mismo ingreso medio y la primera es de *total igualdad* entonces será la elegida. Lo mismo es válido en general para cualquier índice que cumpla Dalton-Pigou porque es claro que para pasar de E_1 a E_2 sólo hacen falta transferencias regresivas.

Pero, ¿podemos afirmar que cualquier individuo elegirá E_1? Una idea alternativa para comparar las economías es un criterio que podemos llamar *dominancia de Mann-Whitney*[1] y que sería el siguiente: al comparar dos variables aleatorias X e Y diremos que $X\ MWD\ Y$ si $P(X > Y) > 1/2$.

Es decir, al elegir entre cada billete de loteria (variable aleatoria) tiene sentido que elijamos una en la cual la probabilidad de tener un ingreso mayor es mayor que 1/2. De lo contrario, simplemente elegiríamos entre ellas tirando una moneda.

En nuestro ejemplo, es clara la relevancia de este enfoque: si llamamos X a la variable aleatoria que toma cada valor x_k con $p = 1/10$ y llamamos Y a la que toma los valores y_k con $p = 1/10$, entonces $P(Y > X) = P(Y = 4) = 0{,}7$. Esto significa que eligiendo la segunda economía, uno tiene una probabilidad del 70 por ciento de tener un mayor ingreso que en la primera, y se trata de una suba no despreciable de $33,3\,\%$. ¿Rechazaríamos todos este segundo billete en favor del primero?

La clave de la resupuesta es obvia: **depende de nuestra aversión al riesgo**. Cuando empezamos a hablar de dominancia estocástica (Sec.1.2.2) dijimos que lo que iba cambiando de un orden de dominancia al siguiente era la clase de funciones de utilidad sobre la que basamos las comparaciones. Esto se entiende claramente si pensamos los funcionales del tipo 1.3 en el sentido de von Neumann-Morgenstern, es decir, que el bienestar social de la distribución viene dado por la *utilidad esperada* de la misma.

Resulta entonces claro lo que dijimos, pues la dominancia estocástica a segundo orden la elegimos precisamente para las funciones con derivada segunda negati-

[1] Algunos textos (Hájek et al. [97]) llaman a este concepto "dominancia estocástica", pero no es una nomenclatura usual y sólo nos traería confusión ya que usamos ese nombre a lo largo de toda la Tesis para otro concepto, no totalmente ajeno, pero distinto. Le damos el nombre del test estadístico usual para contrastar la condición.

va, es decir las cóncavas, basados en la idea de que la utilidad marginal del ingreso es decreciente, concretando entonces un orden de las distribuciones compatibles con un agente *averso al riesgo*.

Pero a partir del ejemplo creemos que no resulta obvio que esta noción sea la única razonable. Recalquemos que estas comparaciones no se basan en lo que el individuo *efectivamente ganó*, sino en el ingreso *potencial*, es decir en lo que *podría* ganar de ser un miembro de cada sociedad.

Una forma de interpretar esto sería que la primer economía tiene mayor igualdad, pero la segunda ofrece *mayores oportunidades*. En la primera uno tiene *asegurado* un ingreso mínimo más alto, pero a la vez está *condenado* a un ingreso máximo más bajo.

Hemos elegido un ejemplo lo suficientemente simple como para que sea visible la contradicción entre los conceptos, por supuesto se podrían elaborar otros donde las distribuciones hicieran más dudosa la comparación, incluso aquellos en los que el orden de Lorenz no sea decisivo pero que sí tengan una dominancia Mann-Whitney.

Ahora bien, yendo un paso más allá, esta idea de que la aversión al riesgo es determinante al discutir un ranking de bienestar social podría explicar la persistencia de desigualdad en sociedades avanzadas. Sin entrar en detalles estadísticos, podemos pensar que una economía como la de USA es del segundo tipo, mientras una como la de Noruega o Suecia es del primero[2]. En el primer país, tenemos una mayor desigualdad, y una chance más alta de caer en la pobreza que en el segundo; pero también una chance más alta de volvernos millonarios. ¿Resulta acaso evidente pensar cuál de los efectos es determinante?

En el enfoque *bienestarista* o en cualquier *igualitarismo* la respuesta es que sí, pero esto parece suponer arbitrariamente la aversión al riesgo o alguna función de bienestar social cóncava. En términos técnicos, no está claro por qué descartar de plano las funciones de utilidad de tipo Friedman-Savage [75], sobre todo al comparar economías con altos niveles de ingresos medios.

No resulta absurdo suponer que hubiera preferencias sociales que hicieran elegir al segundo tipo de economía.[3] En base a ello, una conclusión que podemos esbozar, es que **al comparar dos distribuciones mediante una función de bienestar social o un índice de desigualdad, estamos suponiendo una aversión al riesgo similar como base del contrato social de cada economía, y esto puede no reflejar las preferencias sociales adecuadamente.**

Por supuesto, hasta qué punto puede llevarse a la práctica una *prueba* para contrastar las preferencias sociales y la aversión al riesgo no es algo trivial. La *persis-*

[2]Obvio que *no* lo son en sentido estricto, pues nisiquiera tienen el mismo ingreso esperado. Se entiende que nos referimos a características generales de las distribuciones.

[3]El efecto de las diferentes aversiones al riesgo sobre la elección de carrera del trabajador y el mercado laboral en su conjunto ha sido estudiado por Bergstrom [20]. Es un trabajo interesante de releer y reinterpretar con las ideas que acabamos de presentar en esta Sección.

tencia que señalamos recién es un dato que bien podría reflejar esto, o simplemente reflejar que la segunda economía tiene un menor bienestar.

Algo destacable sobre las nociones de bienestar social que hemos trabajado a lo largo de esta Tesis, es que la noción de dominancia estocástica a primer orden (al menos en el caso 1D) no depende de la aversión al riesgo, sino que es unánime sobre todas las funciones crecientes en el ingreso. Esto implica que si $E_1 SD_1 E_2$ entonces estaría mejor sin importar la aversión al riesgo subyacente.

Esto es una ventaja de la SD_1 por sobre las demás nociones de dominancia. El problema, claro está, es que al ser demasiado *unánime*, resulta menos veces *decisiva*. En el ejemplo dado, sin ir más lejos, no hay dominancia para ningún lado.

Sí tenemos en este caso dominancia en el sentido de Mann-Whitney, y una línea interesante de trabajo a futuro es estudiar este concepto en términos de bienestar social y los órdenes derivados, principalmente para extenderlos al caso multidimensional.

7.2. El Enfoque Estadístico de la Desigualdad en la *Macrojusticia* de Kolm

Sin duda uno de los principales aportes a la economía normativa de los últimos años ha sido el famoso libro de Serge-Christophe Kolm [122] *"Macrojustice: The political economy of fairness"*.

El alcance y profundidad del enfoque de la obra es notable, cubriendo desde los aspectos filosóficos más fundamentales de la noción de justicia hasta las reglas de implementación necesarias para una política fiscal *justa*. Su *modesto* objetivo es claramente expuesto por el autor:

> *"...it is to investigate in depth the solution to the core problem of distributive justice, "global or overall distributive" justice in "macro-justice" in a society."* (Kolm [122] p.11)

El adjetivo *modesto* que antepusimos al objetivo, es por supuesto en principio una ironía, por la importancia fundamental que presenta el mismo. Pero además, pese a lo que tradicionalmente se piensa, Kolm apunta que la justicia social entendida en términos de la distribución de bienes, es un *third best* en términos filosófico-históricos:

> *"Hence, justice is certainly not "the first virtue of society," as Aristotle and Rawls put it. Resorting to justice only is a third best, making up for the lack of sufficient personal awareness and control, and of other-looking and integrating social sentiments. It even is a more distant value if one considers the place of culture and the respect of cultural and natural heritage."* (Kolm [122] ibid.)

Un punto de partida básico de obra es una crítica al *bienestarismo* (welfarism) del que adolece la economía normativa moderna. En este sentido, Kolm distingue tres niveles básicos de juicios de equidad en la distribución de recursos: 1) *microjusticia*; 2) *mesojusticia*; 3) *macrojusticia*. Estos niveles se diferencian esencialmente en base a las necesidades de *información* que requiere el juicio.

El nivel de microjusticia trabaja con la información de las utilidades de los agentes individuales. Este nivel, señala Kolm, sólo puede aplicarse a casos muy concretos y acotados (pocos agentes, pocos bienes), donde es posible evaluar las necesidades y utilidades individuales involucradas. Por ejemplo, la adjudicación de un órgano vital entre pacientes esperando un transplante. La mesojusticia también trabaja con ciertos bienes específicos, pero que son utilizados por todos los agentes económicos (salud, educación, etc.).

La noción de macrojustica, por su parte, considera que las funciones de utilidad individuales como fuente de información están fuera del alcance del hacedor de política. Este es el caso, por ejemplo, de la distribución del ingreso, donde los perfiles de utilidad de cada individuo serán distintos, y donde además se trata típicamente con poblaciones muy numerosas. De esta forma, Kolm altera el enfoque básico de la economía del bienestar.

Combinado con esta crítica, la teoría de la macrojusticia de Kolm plantea la necesidad de un enfoque de la tributación que se base en el ingreso potencial que se podría obtener en base al uso completo de las capacidades de trabajo en el mercado. En este sentido el autor desarrolla su propuesta de *ecualización de ingresos por igual trabajo* (*Equal-Labor Income Equalization, ELIE*). Para entender cómo esta idea puede pasar sencillamente a la práctica de política tributaria podemos citar a Kolm:

> *"Actual policies show the material possibilities (for instance the exemption of overtime labour earnings from the income tax that amounts to basing transfers on capacities). The result is a simple and richly meaningful distributive structure that means, jointly, equal real liberty; adding an egalitarian and a classical liberal parts of income; reciprocity by providing each other with the product of the same labour; and an equal basic income financed by an equal partial labour of each."* ([77], p.35)

En base a las ideas de Kolm, diversos autores han planteado los necesarios pasos de evaluación econométrica de modelos que permitan simular los efectos de las política propuestas. El trabajo de Lubrano [137] estudia los efectos redistributivos de la ELIE en base a ejercicios de simulación.

El modelo básico de ELIE que plantea Lubrano parte de la idea de Kolm de tomar como base imponible el *ingreso potencial*, notado para el individuo i por ω_i. Notemos que en el enfoque del capital humano que dimos en la Sec.2.1.2 este ingreso potencial sería igual a wh_t^i, donde hemos añadido el supraíndice i para identificar al agente.

Ahora según el enfoque de Kolm, una cantidad trabajo k se toma de cada individuo y es medida en términos de su productividad de manera que el impuesto será $k\omega_i$. Al mismo tiempo, un monto fijo $k\tilde{\omega}$ a determinar se transfiere a cada individuo, de manera que la transferencia neta para el individuo i será $\tau_i = k(\tilde{\omega} - \omega_i)$.

El valor de $\tilde{\omega}$ se determina en base al autofinanciamiento del sistema, es decir se pide $\sum_i \tau_i = 0$, de donde es fácil deducir que $\tilde{\omega} = (1/n)\sum_i \omega_i$. Nótese que $\tilde{\omega}$ juega el papel de *pivot* porque para $\omega_i > \tilde{\omega}$ la transferencia neta será negativa (se pagará impuesto) y para el caso contrario positiva.

El ingreso neto disponible entonces, para un trabajador de productividad ω_i que trabaja l_i horas será:

$$y_i = \omega_i l_i + \tau_i = \omega_i l_i + k(\tilde{\omega} - \omega_i) = k\tilde{\omega} + \omega_i(l_i - k) \tag{7.1}$$

así que en cierta forma el $k\tilde{\omega}$ es una suerte de *ingreso mínimo universal*, al menos para todos los que trabajen más de k horas.

A partir de eso, Lubrano [137] supone una distribución sobre las productividades ω de tipo $\Gamma_{\nu,s}$. Para calibrarla, ajusta sus parámetros de forma que el coeficiente de Gini de la distribución de salarios corresponda a los datos reales, en su caso Francia en 1998.

Se procede entonces a considerar la aplicación de ELIE con tres valores de k : 0.2, 0.3 y 0.4, siendo un valor $k = 0{,}2$ el equivalente a 1 día de trabajo en una semana de 5 días, etc. Se computa luego, a partir de la Ec.7.1, la distribución de los ingresos netos disponibles y se estudia el efecto que ha producido la introducción del impuesto, usando el coeficiente de Gini.

Esto merece un par de observaciones de nuestra parte. En primer lugar, como señalamos en reiteradas ocasiones, el autor usa un coeficiente de Gini *exacto* y no un enfoque estadístico. Esto se debe a que en el modelo simulado se conoce con exactitud la distribución subyacente, pero no es algo que pueda considerarse satisfactorio si se aplicara en términos empíricos, y de hecho la comparación con el coeficiente de Gini francés (empírico) que hace el autor no parace tener demasiado sentido, porque no da un intervalo de incerteza.

Pero en segundo lugar, mucho más importante, el efecto que se estudia es completamente unidimensional (sobre el ingreso), cuando el punto de impácto del impuesto está directamente relacionado con el capital humano. Es entonces inevitable la necesidad de considerar el efecto de ELIE sobre esta variable para poder hacer una evaluación de la posible mejora de bienestar derivada de su aplicación.

Para entender el efecto de este impuesto, debemos recordar de la Sec.2.1.2 que la inversión en capital humano se modela bastante bien como un proceso de decisión optimizadora del ingreso a lo largo del ciclo vital. Es clave así estudiar el efecto de ELIE sobre la decisión del individuo al elegeir la trayectoria óptima h_t.

Delacroix y Lubrano [50] presentan un modelo de generaciones solapadas con agentes heterogéneos, con el cual expanden el análisis de Kolm para estudiar el impacto de ELIE en la distribución de capital humano. Una primera conclusión interesante de dicho trabajo es que la aplicación elemental de ELIE genera una baja en el nivel agregado de educación y un menor crecimiento.

En este sentido, el enfoque de la macrojusticia parece enfrentar un problema que se asigna comunmente a los *igualitarismos* que es el efecto negativo sobre los incentivos a la productividad (Putterman et al. [174]).

Pero hay dos cosas que deben señalarse como posibles mejoras al trabajo de Delacroix y Lubrano [50]. En primer lugar, el modelo de generaciones solapadas es muy útil para estudiar el crecimiento pero tal como está plantado pierde la posibilidad de modelar el ciclo de Ben Porath (cfr. Sec.2.1.2) en plenitud.

Huggett et al. [108] presentan un modelo dinámico de la distribución de capital humano y los ingresos que capta mejor la estructura de ciclo vital del proceso de inversión. El problema decisorio se plantea como un programa dinámico y no como un hamiltoniano, pero la idea es totalmente análoga a la que presentamos en la Sec.2.1.2.

Volvemos la notación de la Sec.2.1.2 y entonces el ingreso potencial de un individuo en el período t será $w_t h_t$ (dando lugar a que w cambie con el tiempo). Suponemos que en el período t el trabajador invierte una fracción l_t de su ingreso potencial en la producción de capital humano.

El agente económico resuelve entonces:

$$\text{máx} \sum_{t=1}^{J} \frac{w_t h_t (1 - l_t)}{(1+r)^{t-1}} \tag{7.2}$$

$$s.a. \quad \begin{cases} l_t \in [0,1] \\ h_{t+1} = h_t(1-\sigma) + a(h_t l_t)^{\alpha} \end{cases}$$

donde hemos incorporado la *depreciación* σ y usamos una función de producción Cobb-Douglas para el capital humano. El factor a es un factor que capta la *habilidad* del trabajador, y es heterogéneo entre los individuos.

Huggett et al. encuentran la solución del problema de programación dinámica por inducción hacia atrás sobre la ecuación de Euler y simulan trayectorias para distintas a a partir de una distribución inicial del capital humano.

Lo que queremos destacar respecto a este modelo es que permite la introducción de ELIE de manera natural en el marco de la dinámica de decisión intertemporal del trabajador. Pues si el impuesto es $kh_t w_t$ entonces el ingreso del período se transforma en[4]:

$$h_t w_t (1 - l_t) - k h_t w_t = h_t w_t (1 - l_t - k) \tag{7.3}$$

y en consecuencia se ve claramente la modificación introducida en el programa de optimización 7.2. Es de esperar que este modelo capte mejor el efecto de ELIE sobre el capital humano a lo largo del ciclo vital.

Una vez que se tiene el modelo se puede proceder en primer lugar a contrastar su capacidad de replicar los datos empíricos. Si el resultado es aceptable, se puede utilizar el modelo para la simulación del efecto de las políticas económicas, como el ELIE simple o su combinación con otra serie de medidas para lograr efectos deseados, por ejemplo, una mayor igualdad con menor efecto negativo sobre el crecimiento (Delacroix y Lubrano [50]).

Pero por lado, no podemos dejar de insistir en que tanto Lubrano [137] como Delacroix y Lubrano [50] a la hora de evaluar los resultados de la política, trabajan con análisis unidimensionales de los resultados.

Aun cuando el último de los trabajos mencionados estudia el efecto sobre el capital humano, lo hace en sentido *marginal*, **nunca estudia la distribución con-**

[4]Nótese que por simplicidad no estamos usando el subíndice i para indicar que se trata del individuo i-ésimo, pero está implícito

junta del ingreso y el capital humano, perdiéndo entonces la posibilidad de captar el efecto de interacción entre ambas dimensiones.

Una forma de captar estos efectos, como sabemos, es utilizar tests bivariados de dominancia estocástica. Por otro lado, al trabajar con datos empíricos, se deberá recurrir necesariamente a datos muestrales. Es aquí donde entra la necesidad de un enfoque estadístico de la desigualdad, y hemos visto que las herramientas desarrolladas en el Cap.5 son sencillas y con buenas propiedades estadísticas.

Creemos entonces que los tests bivariados de dominancia estocástica desarrollados en la presente Tesis pueden ser una herramienta de gran importancia para el estudio econométrico de la Macrojusticia de Kolm.

Capítulo 8

Conclusiones

8.1. Conclusiones Técnicas

Generalización de Teoremas de Caracterización de la Dominancia Estocástica: En el Cap.3 mostramos cómo los resultados de Hadar y Russell (Teo.3.2.1.1) y de Atkinson y Bourguginon (Teo.3.3.1.1), que caracterizan la dominancia estocástica en base a desigualdades de operadores aplicados sobre las funciones de distribución, pueden ser generalizados al caso en que la distribución subyacente es arbitraria, es decir, no necesariamente absolutamente continua o siquiera continua.

Esto es relevante por cuanto al trabajar con desigualdad muchas veces los argumentos parten de distribuciones discretas. En la *axiomática* de los índices (Sec.1.2.1 y 2.2.3) esto es muy claro: el *principio poblacional*, por ejemplo, carece de sentido en el caso continuo.

Si bien nuestro planteo de la introducción (Sec.1.2.1) permite darle sentido a muchos de las propiedades y de los axiomas con conjuntos de agentes arbitrarios, en algunos casos concretos como el mencionado, la economía debe considerarse finita.

La dominancia estocástica, al tener como base de información la función de distribución F cumple trivialmente el principio poblacional, pues la misma no cambia al pasar a una k-réplica. Pero si la caracterización sólo fuera válida para distribuciones absolutamente continuas, esta propiedad se pierde de vista.

Generalización de los tests al caso discontinuo: La generalización anterior se vería opacada si al pasar al enfoque estadístico tuvieramos que considerar sólo el caso continuo, por ejemplo si quisieramos usar la distribución tabulada del estadístico KS unidimensional clásico.

Pero como demostramos en la Sec.5.3, nuestros resultados son válidos de manera general, incluyendo medidas subyacentes P, Q con átomos.

Así podemos combinar este resultado con el anterior y tener una herramienta válida incluso en el caso de distribuciones discretas.

Generalización Multidimensional de los Operadores: En la Sec.1.3.2 generalizamos los operadores H_j de Davidson y Duclós [45] al caso multidimensional. Presentamos entonces una familia de funciones S_j y una de operadores $\tilde{T}_k$ que permiten una rápida generalización de los resultados estadâísticos al caso N dimensional, generalizando entonces el trabajao de Barret y Donald [15].

Tests Multivariados de Dominancia Estocástica: En el Cap.5 construimos los estadísticos de prueba del tipo KS para la dominancia estocástica a orden 1 y 2, mostrando como se generalizan para un orden arbitrario. Mediante el uso de la teoría de procesos empíricos pudimos enunciar y demostrar teoremas que caracterizan el comportamieto asintótico de los mismos. Los resultados son válidos para el caso multidimensional y discontinuo.

En conjunto, dichos resultados nos permiten afirmar que hemos logrado la generalización al caso multidimensional de los tests de Barret y Donald [15] propuesta como objetivo general del presente trabajo.

Consistencia: Los tests presenados son consistentes en el sentido estadístico por la convergencia de los valores críticos a los correspondientes al puente browniano de medida H (Sec.5.3). Esta es una ventaja de basarnos en el bootstrap empírico por contraposición a simulaciones de Monte Carlo o bootrstrap paramétrico.

Por otra parte, como señalan Barret y Donald [15, 16], los tests de dominancia estocástica en base a KS, al comparar distribuciones punto a punto son en general más consistentes que los tests basados en comparación de percentiles de la curva de Lorenz o medidas resumen.

Además, el hecho de no reducir la comparación a percentiles permite una interpretación más directa en términos de bienestar (Sen [199]). Esto es una ventaja de la dominancia estocástica por sobre los índices o medidas, que en general son *compatibles* pero no *determinantes* del orden de bienestar social.

Ampliación al Caso Dependiente: Las técnicas de procesos empíricos dependientes nos permitieron ampliar la validez de los resultados al caso de procesos dependientes intercambiables, de utilidad cuando se muestrea sin reposición de una población finita.

Por otra parte la *heurística* construida en la Sec.5.4.2 permite extender los tests al caso dependiente de los *procesos de mezcla*. Mediante el diseño del *bootstrap en bloque de dos muestras* (*two sample MBB* o 2*SMBB*) construimos una herramienta de valor cuando los datos provienen de series temporales.

Aplicabilidad Práctica del Herramental Desarrollado: El valor de una herramienta econométrica radica principalmente en su aplicabilidad práctica. En este sentido, como señalamos en el Cap.2, creemos que el enfoque y los resultados de la presente Tesis presentan una serie de ventajas.

A saber:

i. **Como se basa en la dominancia estocástica, es fácil de interpretar directamente en términos de la economía del bienestar, como en Atkinson y Bourguginon [14].**

ii. **Por la misma razón, es un método que toma plena cuenta de la *interacción* entre las dimensiones de la desigualdad.**

iii. **Se puede aplicar tanto a los casos super como submodulares, es decir al caso de atributos *complementarios* como *sustitutos*.**

iv. **Es fácil de usar en la práctica, porque solo requiere de la función de distribución multivariada $F(x_1,...,x_n)$ y de sus integrales, algo conocido por todos los investigadores, aceptable y elemental de calcular.**

v. **Por usar sólo las F, es casi inmediata su formulación estadística, mediante las distribuciones empíricas. Y una vez que tenemos éstas se puede usar el enfoque de McFadden [149] o el de Anderson [7] para definir tests.**

vi. **La implementación mediante estadísticos KS y bootstrap empírico le da consistencia y a la vez elimina la arbitrariedad de la elección de una distribución en los métodos de Monte Carlo**.

vii. **El método para la obtención de valores críticos dependientes de los datos es fácil de programar en cualquier lenguaje o paquete matemático o estadístico, como ser Matlab o Stata.**

vi. **La validez de los tests para el caso de procesos intercambiables y el diseño del $2SMBB$ amplía la aplicabilidad de los resultados a las series temporales.**

8.2. Conclusiones Económicas

Evaluación de la Política Económica: Como señalamos en la Sec.2.2 y sobre todo en la Sec.4.3, la *interacción* entre las dimensiones de la desigualdad es clave para entender el significado de la misma en términos económicos. Por ello, a la hora de evaluar los resultados de la política económica, creemos fundamental encararlo desde un enfoque multivariado.

Para el caso cuando el ingreso y el capital humano como dimensiones, esta interacción cobra particular importancia. Por ello en el análisis de políticas como la ELIE en la macrojusticia de Kolm (Sec.7.2) es indispensable contemplar **a la vez** ambas dimensiones.

En este sentido la dominancia estocástica bivariada y el herramental estadístico aquí desarrollado pueden resultar de gran utilidad práctica.

Diseño de la Política Económica: Dijimos ya en la Introducción que el presente trabajo no tiene como objetivo el diseño, propuesta o sugerencia de determinadas políticas económicas, sino más bien el de brindar herramientas para su evaluación.

Sin embargo, a lo largo del desarrollo del presente trabajo, algunas cuestiones inherentes al diseño de la política no pueden escapar nuestra atención.

Como señalamos en la Sec.4.3, la correlación entre las variables ingreso y capital humano estará determinada fundamentalmente por la estructura del mercado laboral. Dado que según hemos visto, la complementariedad estratégica de las dimensiones hace que una mayor correlación implique un mayor bienestar. Pero, ¿cómo se logra una correlación mayor?

En principio, una idea para subir la correlación es buscar aumentar los retornos a la educación, por ejemplo, fomentando el empleo con mayor calificación. Pero esto no es tan simple, porque para asegurar que el bienestar mejore, deben manternese las marginales. Si los retornos a la educación crecen, pero el capital humano está distribuido desigualmente, entonces lo más probable es que aumente la desigualdad en la distribución del ingreso, y el bienestar en consecuencia no habrá aumentado.

¿Es éste un trade-off inevitable o existen formas de lograr mayor correlación con mejor distribución del ingreso? Encontrar una respuesta a este interrogante es por supuesto muy complejo y daría material, probablemente, para una tesis doctoral adicional.

Como respuesta tentativa, sin embargo, afirmaremos que sí, es posible mejorar la correlación mateniendo o aun mejorando las distribuciones marginales, y señalaremos brevemente cierta literatura especializada que fundamenta dicha afirmación.

El trabajo de Topel [221] presenta un modelo que incorpora las observaciones de Kuznets sobre la relación entre crecimiento y desigualdad[1] y explica la mecánica de la siguiente manera:

> *"In the model, export-driven demand for industrial output raises the demand for skilled labor. In turn, investment in human capital responds to differences in wages between skilled and unskilled labor. Wage inequality spurs investment in human capital and more rapid economic growth, but increased relative abundance of skills serves to reduce inequality"*

Si bien la *curva de Kuznets* no se considera demasiado apoyada por los estudios empíricos más recientes (Fields [71]), el mecanismo descrito por Topel nos da como clave entonces la necesidad de un proceso de ecualización en la distribución de capital humano al tiempo que crece el retorno del mismo. El papel de una *orientación exportadora* del modelo económico también es interesante como guía de política.

El desarrollo tecnológico es la base de la desigualdad generada por las diferencias de capital humano: como señalan Acemoglu y Autor [2], la habilidad del trabajador y la tecnología, que fueron sustitutos hasta el siglo XIX (el obrero industrial de baja calificación reemplazando al artesano) se vuelven complementos desde el siglo XX y la complementariedad aumenta en la época actual. De allí se deriva una mayor demanda de trabajo calificado y un crecimiento relativo de los salarios de ese segmento.

Por ello, una política económica que fomente la inversión en actividades tecnologicas probablemente tenga un impacto positivo sobre la correlación. Sin embargo, para mantener la igualdad sería necesario una ecualización de la distribución de capital humano. ¿Cómo se logra tal resultado? En el caso clásico de la distribución del ingreso, es conocido el papel igualador de los impuestos progresivos (Lambert [126]), combinados con transfrenecias.

Pero no existe de manera obvia una *transferencia* de capital humano, aun cuando se lo pudiera considerar base imponible de algún gravamen como la ELIE. Las políticas de provisión pública de educación buscan de alguna forma dicho efecto, y como señala Neal [155], tienden a disminuir la dispersión entre las *curvas de oferta* de capital humano.

Debe señalarse, sin embargo, que no es directa la asociación entre la provisión pública de un bien y su mayor *consumo*, pues ello depende del mapa de indiferencia del consumidor (la combinación del efecto precio e ingreso).

[1]El trabajo clásico de Kuznets [125] de 1955 observa que a medida que crece el producto de un país, la desigualdad primero crece y luego decrece. Algunos han interpretado equivocadamente esta afirmación como un suerte de *justificación* o explicación de la mayor desigualdad existente en países en vías de desarrollo relativa a los desarrollados. Esto es equivocado porque los datos de Kuznets eran de serie temporal, no de corte transversal.

Por otra parte, en el segmento de educación superior es probable que un sistema educativo con mayor *diversidad* de oferta sea más flexible para adaptarse a las demandas de un mercado laboral orientado a la tecnología. Si esta diversidad viene dada por una estructura con mayor número de instituciones privadas, una opción sería adoptar algún esquema de subsidios (*grants* o *vouchers*) para garantizar una mayor igualdad, tanto en los segmentos universitarios como técnicos.

De manera general, políticas de corrección a los *mercados financieros incompletos* son recomendables en este sentido (Behrman [19]).

Señalemos que esta diversidad en la demanda de capital humano, es decir la diferenciación de *tipos* del mismo, brinda la posibilidad de un crecimiento en el stock social sin que los retornos caigan.

Por último señalemos que la literatura especializada (e.g. Gregory y Borland [90]) señala que el sector público tiende a pagar salarios por encima de su productividad. Esto implica que una economía en la que la creación de empleo se centra en el sector público tenderá a tener una menor correlación de ingresos y capital humano. Por ello debe considerarse preferente el fomento a la creación de empleo en el sector privado.

Irreversibilidad de las Pérdidas: Con frecuencia el análisis del bienestar social se torna estático, aun en la comparación intertemporal. Se focaliza en la distribución de factores en un tiempo y lugar dado y se la compara con la de otros tiempo y lugar dados.

Pero el análisis del bienestar no puede dejar de ser dinámico si se está pensando en la evaluación y diseño de la política económica. **Esto se debe a que las pérdidas de bienestar social son irreversibles.** Una economía puede empeorar al pasar de la situación a a la situación b en términos de bienestar agregado, y luego mejorar para llegar a una situación c con el mismo bienestar. Pero ¿qué pasó en el medio? ¿ es neutral el resultado, es decir, estamos igual en a que en c?

Naturalmente, la respuesta depende de la noción operativa de bienestar social que adoptemos, pero en términos generales la respuesta es *no*. Esto tiene manifestaciones muy concretas: una economía en la cual el nivel de pobreza salta abruptamente (e.g. Argentina 2002) verá un aumento de la desnutrición infantil que implicará una generación en la cual una porción importante de la población crecerá con menor desarrollo cerebral.

En el caso que nos ocupa, es claro que el efecto de la *no inversión en capital humano* es bastante irreversible. Alguien que a los 15 o 20 años deja de educarse para encontrar un empleo en sectores de baja productividad, casi con seguridad no volverá a la escuela si años más tarde se encuentra desempleado.

La dinámica del mercado laboral puede ser tal que un sector importante de la población activa se encuentre con una acumulación insuficiente de capital humano, y éste es un problema que el diseño de la política económica no puede dejar de lado.

8.3. Posibles Extensiones del Estudio

Teorema de Strassen para SD_2***:*** Como señalamos en el Cap.3, el Teorema de Strassen sobre acoplamientos y dominancia estocástica facilita grandemente la demostración de ciertos resultados. Pero en el caso de la dominancia de segundo orden no hay un resultado similar. Una interesante línea de trabajo futuro es entonces la de buscar un resultado de este tipo, utilizando la teoría de acoplamientos.

Evaluación de Modelos Dinámicos: Los tests de dominancia estocástica básicamente comparan distribuciones. Si dichas distribuciones difieren, entonces los tests de dominancia mostrarán la diferencia. Esto permite comparar los resultados de las predicciones de los modelos dinámicos con la realidad, y en base a ello evaluar su desempeño.

Consecuentemente, una interesante dirección de estudio futuro es comprobar la utilidad de los tests para el *assessment* de modelos de equilibrio, sobre todo en cuanto se usen datos de serie temporal.

Estructura del Mercado Laboral y Política: Todo lo que señalamos en la Sec.8.2 sobre el diseño de la política económica requiere un estudio mucho más profundo. En el mismo es probable que entren técnicas como las señaladas en la Sec.7.2, con modelos de equilibrio general que capturen la esencia del proceso decisorio de la inversión en capital humano.

En base a ellos, se podrá simular los efectos de las políticas sobre la distribución de y y de h y con los tests de dominancia bivariados se podrá evaluar los resultados.

Desigualdad Económica: Las observaciones de la Sec.7.1 pueden desarrollarse mucho más, en las líneas allí mencionadas. Por ejemplo, intentar captar la *forma* de las preferencias sociales relativas al riesgo (tal vez funciones de Friedman-Savage). Por otra parte, la noción de dominancia de Mann-Whitney puede ser una interesante opción dentro de la economía del bienestar, siendo necesaria su generalización al caso multivariado.

Apéndice A

Apéndice: Test de Bickel

El trabajo de Bickel [25] fue el primero en desarrollar un test del tipo Kolmogorov-Smirnov para el caso p-variado, con distribuciones continuas, que demuestra ser consistente contra todas las alternativas y pivotal (*distribution free*). La estructura del mismo se basa en el clásico principio de permutaciones de Fisher, y por lo tanto se presenta apropiado para generalizar al caso intercambiable.

Se plantea como siempre el modelo muestral estándar tomando el espacio producto $\Omega = (\mathbb{R}^p)^\infty \times (\mathbb{R}^p)^\infty$ con el álgebra producto de Borel $\mathscr{B}$ y X_i, Y_i las proyecciones. Suponemos que sobre $(\Omega, \mathscr{B})$ rige una medida de probabilidad P_θ dentro de una familia de distribuciones con θ variando sobre

$$\Theta = \{(F,G) : F, G \textit{ medidas de probabilidad sobre } \mathbb{R}^p\}$$

y si $\theta = (F,G)$, P_θ es la medida producto sobre $(\Omega, \mathscr{B})$ que hace a las X_i iid F y las Y_i iid G. Sea $\Theta_0 = \{\theta \in \Theta : \theta = (F,F) \textit{ para alguna } F\}$. El problema que nos interesa testear es la hipótesis $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \notin \Theta_0$, consistentemente a nivel α.

Para computar las probabilidades condicionales, la σ-álgebra de condicionamiento que toma Bickel es la de los tests de permutaciones:

$$\begin{aligned} \mathscr{A} = \{B \in \mathscr{B} : (x_1,\ldots,x_m,x_{m+1},\ldots;y_1,\ldots,y_n,y_{n+1},\ldots) \in B \Leftrightarrow \\ (T(x_1),\ldots,T(x_m),x_{m+1},\ldots;P(y_1),\ldots,P(y_n),y_{n+1},\ldots) \in B\} \end{aligned} \tag{A.1}$$

para cualquier permutación T sobre $\{x_1,\ldots,x_m,y_1,\ldots,y_n\}$.

Define entonces el test mediante una sucesión de funciones $\phi_{m,n}$ por:

$$\begin{cases} \phi_{m,n}(X_1(\omega),\ldots,X_m(\omega),Y_1(\omega),\ldots,Y_n(\omega)) = 1 & d(F_m(\cdot,\omega),G_m(\cdot,\omega)) > c_N \\ \qquad\qquad = \gamma_N & d(F_m(\cdot,\omega),G_m(\cdot,\omega)) = c_N \\ \qquad\qquad = 0 & \textit{en otro caso} \end{cases}$$

donde la distancia es la uniforme y c y γ definidos con la condición de ser los números más pequeños que cumplen:

$$E\{\phi(X_1,...,X_m,Y_1,...,Y_n)|\mathscr{A}\} = \alpha$$

Bickel prueba que la sucesión cumple:

$$E_\theta(\phi(X_1,...,X_m,Y_1,...,Y_n)) \leq \alpha \ \ \forall\theta \in \Theta_0$$

donde la esperanza se toma bajo el supuesto de que rige θ y además:

$$\lim_{n,m\to\infty} E_\theta(\phi(X_1,...,X_m,Y_1,...,Y_n)) = 1 \ \ \forall\theta \notin \Theta_0$$

es decir, un test consistente y de nivel α.

Ahora bien, analizando el planteo vemos que la σ-álgebra es la misma si hubieramos planteado el caso intercambiable, y que es suficiente en el sentido estadístico (Dudley [63]). Por otro lado, la construcción con las proyecciones es canónica, pero podríamos haber construido un espacio de forma tal que sólo valga la condición de distribuciones idénticas, no necesariamente independientes, sino intercambiables.

Mirando los pasos de la demostración de Bickel, vemos que los resultados sobre muestreo sin reposición son válidos en cualquier caso, no hay incidencia de la condición de independencia. **El único momento en que no se puede dejar de lado la condición de independencia es cuando usa el teorema de Glivenko-Cantelli.**

Pero como vimos en la Sec.5.4, se puede generalizar al caso intercambiable el teorema de Donsker, y la condición $P-$Donsker es más fuerte que la de Glivenko-Cantelli.

Apéndice B

Tablas Completas de Resultados

A continuación pueden consultarse las tablas completas de resultados para los tests SD1+ y SD2++ para el caso de Argentino 1996-2006 (cfr. Sec.6.2).

El sufijo + o ++ se debe a que los tests fueron hechos para el caso supermodular, por razones explicitadas más arriba (Cap.4).

Los números se redondean a 2 decimales para facilitar la presentación, y la interpretación de las columnas es como en los programas de Matlab (Apéndice C).

Concretamente, las primeras dos columnas indícas quienes son P Q para el test *directo*, o sea con hipótesis PSD_jQ. La columna $SD1+$ da el resultado del test de dominancia a primer orden, interpretado como en la Sec.6.2. Análogamente la columna La columna $SD2+$. El símbolo + viene dado para recordar que se trata del caso supermodular.

Las columnas $l1$ y $l2$ dan los resultados de la evaluación de los edps de primer y segundo orden respectivamente, y las columnas $cc+1$ y $cc+2$ dan los *valores críticos dependientes de los datos obtenidos* mediante el método de bootstrap.

Las columnas *bdx*, *lx*, *crx* dan los resultados del test de dominancia unidimensional sobre la distribución marginal del capital humano, y las de sufijo *y* las del ingreso. El resultado de dominancia lo llamamos *bdx* porque el test en las marginales resulta equivalente al de Barret y Donald [15].

Las columnas con un prefijo *i* se refieren al test en el sentido *inverso*, es decir con la hipótesis QSD_jP, esto nos permite recortar la tabla a la mitad de filas.

Tabla B.1: Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
M96	O96	-1	19,48	12,42	1,2	4,58	-1	2,13	1,05	0,05	0,38	-1	2,31	1,39	-1	2,33	1,44
M96	M97	1	3,26	17,85	28,87	27,54	1	0,45	1,71	2,3	2,9	1	0,07	1,74	1	0	1,78
M96	O97	1	3,15	23,57	42,55	15,01	1	0,28	2,3	4,24	1,44	1	-0,01	2,64	1	0	2,75
M96	M98	0	4,02	19,47	20,32	33,27	-1	2,15	1,96	0,78	1,89	-1	1,47	1,27	-1	1,23	1,18
M96	O98	1	3,86	10,76	17,83	5,71	1	0,54	1,08	1,07	0,67	1	0	1,26	1	0	1,26
M96	M99	0	5,31	7,48	3,58	6,37	-1	1,18	0,71	0,03	0,55	-1	0,79	0,64	-1	0,78	0,69
M96	O99	-1	29,94	15,58	1,13	1,42	-1	4,64	1,79	0,03	0,31	-1	4,37	1,93	-1	4,37	1,93
M96	M00	0	6,81	6,65	9,89	6,77	0	1,19	0,64	0,32	0,71	1	0,71	0,79	1	0,34	0,79
M96	O00	1	6,22	9,96	59,48	50,25	1	0,79	1,31	5,63	5,74	1	-0,01	1,44	1	0	1,44
M96	M01	0	7,86	3,65	11,52	5,15	0	1,28	0,23	0,5	0,3	-1	0,7	0,2	1	0,1	0,23
M96	O01	0	9,45	1,66	5,69	1,36	0	2,51	0,34	0,2	0,16	-1	1,5	0,27	0	1,5	0,27
M96	M02	1	9,57	43,61	43,28	38,9	1	1,3	5,06	3,24	4,77	1	0,25	4,81	1	0	4,88
M96	O02	1	7,86	10,33	12,7	1,22	1	1,25	1,37	0,62	0,26	1	0,62	1,3	1	0,1	1,3
M96	M03	1	8,96	10,21	19,46	9,62	1	1,47	1,97	0,92	1,01	1	0,68	1,13	1	0	1,19
M96	P04	1	8,72	52,32	19,21	52,49	1	1,73	5,82	0,99	5,71	1	0,75	5,61	1	0,07	5,64
M96	T04	1	10,61	10,99	20,84	8,38	1	1,43	1,41	1,3	0,85	1	0,48	1,28	1	0	1,28
M96	P05	1	11,15	18,72	10,36	6,43	1	2,15	2,22	0,71	0,63	1	1,33	1,89	1	1,13	1,94
M96	T05	1	11,77	22,77	27,8	9,32	1	2,04	2,7	1,72	1,16	1	0,72	2,59	1	0	2,59
M96	P06	1	9,9	10,64	11,85	2,8	1	2,51	2,63	0,49	0,21	1	1,68	2,05	1	1,4	1,55
M96	T06	-1	32,75	11,91	0,29	11,79	-1	7,59	1,48	0	1,61	-1	6,47	1,39	-1	5,48	1,39
O96	M97	1	3,36	42,1	52,05	46,19	1	0,54	4,95	5,09	5,28	1	0,1	4,8	1	0	4,8

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
O96	O97	1	5,26	13,5	16,48	8,45	1	1,1	4,01	0,97	0,45	1	1,33	2,49	1	0,77	0,29
O96	M98	1	3,18	11,06	19,39	10,16	1	0,76	4,18	1,46	0,58	1	2,08	4,01	1	0,03	0,84
O96	O98	1	4,74	7,58	11,95	5,19	1	0,55	4,34	2,6	1,89	1	0,92	1,61	1	0,96	0,23
O96	M99	1	3,96	8,9	37,92	12,96	1	1,24	3,73	0,97	0,37	1	1,36	5,67	1	0,91	0,07
O96	O99	-1	13,92	12,64	0,22	5,41	1	2,91	1,57	0	0,47	-1	2,55	1	-1	2,88	1
O96	M00	1	3,44	16,32	11,41	3,57	1	0,87	1,5	2,01	1,48	1	3,37	7,51	1	0,38	1,47
O96	O00	1	10,07	14,91	21,87	6,92	1	0,52	2,18	1,86	1,54	1	2,12	5,64	1	3,55	4,12
O96	M01	1	6,39	18,49	9,02	7,39	1	0,84	1,44	3,91	1,79	1	2,24	4,13	1	4,46	5,62
O96	O01	1	3,7	5,7	29,63	20,67	1	1,49	4,56	0,95	0,52	1	2,25	3,52	1	1,06	3,79
O96	M02	1	7,8	15,81	0,85	0,87	1	3,17	3,89	1,13	0,28	1	3	4,53	1	4,11	8,12
O96	O02	1	8,66	17,16	25,96	5,17	1	1,03	4,1	3,34	2,03	1	0,31	0,36	1	1,55	2,64
O96	M03	1	5,93	11,3	17,82	4,56	1	3,95	5,19	4,47	2,17	1	2,98	4,6	1	3,77	4,18
O96	P04	1	13,5	32,01	15,45	7,73	1	2,96	4,49	3,75	3,3	1	3,05	5,1	1	2,95	4,85
O96	T04	1	19,95	22,99	24,04	13,72	1	4,28	7,84	1,3	0,78	1	3,11	5,03	1	3,84	4,38
O96	P05	1	5,48	20,09	23,31	19,3	1	0,46	2,54	6,85	2,64	1	0,19	2,9	1	1,81	4,16
O96	T05	1	19,59	35,17	29,26	10,67	1	1,86	4,22	4,54	1,26	1	3,96	4,13	1	3,79	4,45
O96	P06	1	5,51	20,11	31,35	15,66	1	3,53	5,23	0,85	0,33	1	3,32	3,9	1	0,46	1,14
O96	T06	1	19,61	24	10,41	6,29	-1	6,33	2,33	0	2,84	-1	4,89	2,4	-1	4,57	2,4
M97	O97	1	0,8	1,28	22,19	17,66	1	0	0,14	2,27	1,68	1	-0,01	0,16	1	0	0,16
M97	M98	-1	26,31	16,06	18,89	24,7	-1	3,76	1,69	2,28	4,01	-1	6,22	3,51	-1	4,38	3,96
M97	O98	-1	15,07	14,03	1,89	12,31	-1	1,66	1,62	0,14	1,24	-1	1,76	1,65	-1	1,82	1,65

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
M97	M99	-1	22,96	12,31	9,3	24,03	-1	4,56	1,65	0,4	0,92	-1	4,26	2,44	-1	3,25	1,64
M97	O99	-1	30,58	19,28	11	31,05	-1	5,53	3,47	2,85	3,79	-1	2,84	0,19	-1	5,67	3,54
M97	M00	-1	22,37	21,77	2,09	17,9	-1	3,7	2,33	0,04	1,75	-1	2,9	2,56	-1	2,9	2,56
M97	O00	1	4,84	20,21	44,3	38,5	1	0,47	2,07	3,94	3,16	1	-0,01	2,06	1	0	2,12
M97	M01	-1	24,21	13,68	19,95	26,7	-1	2,62	0,59	3,17	3,84	-1	3,32	1,08	-1	2,35	2,16
M97	O01	-1	14,77	7,39	8,52	15,24	-1	4,22	3,24	2,12	6,65	-1	4,23	3,19	-1	2,12	0,35
M97	M02	1	19,5	30,43	22,58	3,02	1	3,56	7,6	2,29	0,82	1	2,42	5,1	1	2,9	4,05
M97	O02	1	11,68	27,94	22,43	7,13	1	0,57	5,08	5,27	3	1	1,37	4,31	1	3,14	3,28
M97	M03	1	11,31	24,56	8,96	7,76	1	2,62	2,49	2,14	1,58	1	1,89	2,61	1	1,19	2,75
M97	P04	1	15,65	32,71	12,02	5,48	1	1,75	5,77	3,69	2,24	1	2,74	3,89	1	3,49	6,42
M97	T04	1	10,42	14,34	32,31	19,36	1	3,65	8,36	4,81	2,55	1	3,63	6,14	1	1,4	3,27
M97	P05	1	7,45	11,09	35,81	15,34	1	3,8	5,56	6	2,28	1	3,18	5,84	1	2,57	3,28
M97	T05	1	3,98	9,63	20,78	5,57	1	3,09	6,09	4,61	3,76	1	0,12	0,28	1	1,19	2,97
M97	P06	1	5,25	22,97	14,51	10,24	1	3,99	4,93	4,56	2,74	1	2,57	4,29	1	0,7	1,78
M97	T06	-1	33,7	18,08	3,79	11,86	-1	4,45	2,94	1,22	5,92	-1	4,61	2,22	-1	5,33	2,66
O97	M98	-1	25,2	15,34	20,33	40,98	-1	1,26	0,67	3,55	3,96	-1	4,4	3,4	-1	2,87	0,65
O97	O98	-1	29,99	17,11	0,2	11,74	-1	3,84	1,92	0	1,22	-1	3,71	1,8	-1	3,71	1,8
O97	M99	-1	14,56	8,55	19,4	36,99	-1	5,79	3,65	2,42	5,52	-1	4,53	3,98	-1	6,66	3,75
O97	O99	-1	23,46	5,39	12,15	27,21	-1	2,74	1,4	0,08	0,9	-1	1,17	0,78	-1	1,98	0,95
O97	M00	-1	9,29	5,57	11,72	24,08	-1	3,75	2,41	3,75	5,57	-1	3,2	2,83	-1	6,33	3,78
O97	O00	0	13,12	24,9	16,73	23,38	0	1,7	3,58	1,2	4,07	0	2,34	0,63	0	1,17	0,33

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	**F2**	**SD1+**	**l1**	**cc+1**	**il1**	**icc+1**	**SD2++**	**l2**	**ccr+2**	**il2**	**icc+2**	**bdx**	**lx**	**crx**	**bdy**	**ly**	**cry**
O97	M01	-1	24,39	17,33	11,23	20,5	-1	3,4	2,25	1,7	5,05	-1	4,18	2,11	-1	3,68	3,47
O97	O01	-1	24,75	18,43	19,53	37,08	-1	2,47	1,06	3,58	6,38	-1	1,54	0,43	-1	6,5	3,99
O97	M02	1	15,19	35,15	18,63	4,16	1	3,57	8,04	3,91	0,37	1	3,28	3,64	1	3,63	6,4
O97	O02	1	10,86	30,3	19,89	16,71	1	1,94	4,8	5,05	2,15	1	0,17	1,83	1	1,65	4,13
O97	M03	1	14,2	31,58	28,55	13,99	1	1,12	2,57	1,27	0,06	1	3,21	4,56	1	0,97	2,78
O97	P04	1	12,61	32,25	35,77	20,3	1	0,12	3,76	3,57	1,61	1	3,72	5,27	1	0,05	0,94
O97	T04	1	7,59	25,98	30,17	12,88	1	1,17	5,32	7,56	3,11	1	1,12	2,89	1	2,83	4,64
O97	P05	1	12,26	29,21	12,75	5,85	1	3,04	6,26	6,9	2,68	1	3,4	5,06	1	2,97	3,93
O97	T05	1	14,78	27,68	19,68	12,59	1	2,7	7,25	4,06	3,23	1	2,31	4,42	1	1,16	1,63
O97	P06	1	5,51	18,59	16,51	9,67	1	2,01	2,61	1,82	1,58	1	2,82	4,16	1	0,53	1,12
O97	T06	-1	18,13	11,7	18,45	33,33	-1	1,09	0,89	3,64	7,49	-1	1,26	1,11	-1	2,09	0,91
M98	O98	1	5,26	16,63	6,13	5,89	1	0,1	1,78	0,45	0,44	1	-0,01	1,68	1	0	1,68
M98	M99	0	9,68	22,2	9,59	13,94	-1	3,12	0,53	0,19	3,01	-1	0,52	0,1	-1	5,33	3,78
M98	O99	-1	26,49	18,8	4,27	18,29	-1	2,6	0,7	0,25	5,12	-1	2,92	0,75	-1	3,59	1,72
M98	M00	0	23,64	12,87	40,22	20,69	1	2,84	7,1	3,45	3,3	1	2,22	3,72	1	2,18	4,25
M98	O00	1	15,47	33,3	21,31	10,35	1	3,51	7,48	6,41	2,71	1	2,32	4,61	1	2,6	3,78
M98	M01	0	12,7	23,88	20,28	26,77	1	3,53	8,21	2,99	1,78	1	1,67	4,11	1	2,99	4,01
M98	O01	0	14,58	30,21	3,36	8,96	1	3,16	5,1	2,21	1,94	1	0,89	3,84	1	0,98	1,29
M98	M02	1	11,04	31,26	20,86	7,48	1	1,49	1,74	1,15	0,69	1	0,12	1,18	1	3,69	4,94
M98	O02	1	10,67	20,25	33,53	15,23	1	3,65	4,59	8,57	3,92	1	1,19	2,84	1	0,37	3,35
M98	M03	1	9,13	14,76	20,04	4,27	1	1,69	2,28	5,1	3,28	1	0,24	1,08	1	3,15	4,69

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
M98	P04	1	10,19	26,93	11,54	5,81	1	0,06	3,94	4,37	0,7	1	1,22	4,16	1	2,7	4,63
M98	T04	1	20,18	38,59	17,26	11,94	1	0,37	0,67	7,37	3,55	1	3,89	6,55	1	0,05	2,95
M98	P05	1	6,8	17,67	25,19	9,12	1	0,2	4,39	4,56	0,34	1	0,32	2,25	1	1,59	1,96
M98	T05	1	16,22	28,92	29,26	13,58	1	3,26	6,99	4,09	2,29	1	3,87	5,38	1	2,22	3,17
M98	P06	1	9,06	22,85	15,37	4,7	1	2,41	5,3	1,94	0,89	1	1,38	3,53	1	2,46	3,38
M98	T06	-1	25,79	5,82	6,61	22,82	-1	6,33	1,38	3,45	6,07	-1	4,87	3,28	-1	3,54	1,12
O98	M99	-1	25,42	8,49	12,57	24,56	-1	4,12	0,92	0,87	1,92	-1	5,05	2,58	-1	5,4	3,74
O98	O99	-1	12,91	4,59	12,67	18,92	-1	1,62	1,29	0,7	2,12	-1	1,6	0,56	-1	3,69	0,79
O98	M00	-1	21,16	8,17	5,33	10,49	-1	3,67	2,58	3,59	6,36	-1	5,93	3,94	-1	2,95	0,9
O98	O00	1	12,8	28,22	30,22	12,26	1	2,02	2,32	4,21	1,24	1	3,28	3,5	1	3,4	3,59
O98	M01	-1	25,8	5,04	18,88	27,65	-1	3,74	0,79	3,47	4,43	-1	1,53	0,51	-1	4,35	3,28
O98	O01	-1	15,44	4,81	8,65	14,66	-1	4,09	1,64	3,06	4,83	-1	2,79	0,39	-1	2,24	0,71
O98	M02	1	7	14,56	23,02	12,94	1	0,62	1,95	6,25	3,35	1	0,55	0,85	1	2,91	3,62
O98	O02	1	10,89	24,9	21,63	15,33	1	1,57	3,18	3,05	1,99	1	2,5	3,86	1	3,7	4,94
O98	M03	1	19,06	36,26	18,07	5,83	1	2,99	3,07	3,46	0,96	1	0,21	3	1	0,07	2,86
O98	P04	1	5,84	14,13	24,78	13,73	1	3,7	6,38	4,12	3,07	1	2,19	4,5	1	0,83	2,47
O98	T04	1	4,34	23,51	32,62	14,92	1	1,65	1,92	3,69	2	1	2,84	4,41	1	2,91	3,51
O98	P05	1	20,82	36,75	17,57	7,85	1	1,66	4,53	2,3	1,98	1	3,31	4,86	1	2,47	4,23
O98	T05	1	11,87	26,44	11,38	4,97	1	0,36	2,85	3,39	1,83	1	1,68	3,49	1	0,3	1,75
O98	P06	1	9,61	15,12	18,25	11,66	1	1,09	3,32	2,24	0,74	1	3,25	3,89	1	0,39	3,32
O98	T06	-1	16,95	5,9	20,22	27,7	-1	3,1	0,42	1,31	3,82	-1	4,84	3,91	-1	3,01	1,22

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
M99	O99	-1	32,82	12,15	10,63	20,93	-1	5,4	3,4	1,71	4,49	-1	4,31	3,44	-1	3,13	2,63
M99	M00	0	31,22	12,55	22,45	15,66	1	3,5	5,84	8,85	3,91	1	0,64	0,94	1	3,88	4,75
M99	O00	1	14,54	27,1	38,94	19,16	1	3,08	4,12	2,81	2,07	1	3,02	4,22	1	3,59	3,76
M99	M01	0	9,05	28,45	6,49	17,74	1	0,28	0,62	1,5	0,95	1	0,13	0,58	1	2,71	4,99
M99	O01	0	16,64	26,31	7,57	13,1	1	0,74	0,82	0,52	0,28	1	3,89	4,77	1	1,15	1,34
M99	M02	1	16,88	28,17	24,33	7,01	1	0,13	3,11	5,47	1,72	1	1,14	3,29	1	3,59	4,3
M99	O02	1	10,66	22,84	21,92	9,07	1	2,48	2,62	4,85	2,51	1	2,92	4,64	1	2,77	3,5
M99	M03	1	11,14	18,47	28,21	11,33	1	3,61	6,46	5,43	0,77	1	1,22	2,32	1	0,22	2,37
M99	P04	1	18,16	36,71	24,82	11,2	1	0,01	0,91	5,38	1,36	1	2,07	2,84	1	1,97	4,95
M99	T04	1	17,85	38,12	13,37	7,58	1	2,16	2,82	5,47	1,09	1	1,46	4,19	1	0,44	0,51
M99	P05	1	12,86	18,73	29,23	10,75	1	1,17	3,55	3,72	3,49	1	0,29	3,11	1	0,52	1,85
M99	T05	1	17,72	26,01	24,87	16,83	1	2,19	3,98	5,38	2,65	1	0,32	0,44	1	0,9	3,01
M99	P06	1	3,92	11,88	33,73	13,99	-1	4,96	2,6	1,71	2,68	-1	1,67	1,24	-1	5,7	3,42
M99	T06	-1	24,85	18,76	16,16	21,54	1	1,54	4,47	7,56	3,82	1	2,02	4,73	1	2,48	4,73
O99	M00	1	19,61	36,35	18,44	12,43	1	3,72	4,85	1,61	0,55	1	0,78	3,56	1	1,08	3,39
O99	O00	1	8,72	26,19	24,31	13,02	1	1,28	2,44	5,03	1,16	1	2,56	2,95	1	3,8	4,2
O99	M01	1	17,54	20,99	25,3	10,12	1	0,16	0,42	4,77	2,89	1	0,78	2,08	1	2,7	5,57
O99	O01	1	15,39	34,32	19,75	9,76	1	1,8	4,43	7	2,97	1	3,83	4,31	1	1,07	3,38
O99	M02	1	9,8	19,61	25,39	8,37	1	0,3	2,75	5,97	3,69	1	1,68	2,44	1	0,51	2,41
O99	O02	1	7,21	15,91	31,51	15,95	1	3,53	4,81	5,56	0,73	1	2,7	3,29	1	1,04	1,83
O99	M03	1	10,69	14,99	22,07	17,42	1	0,99	3,35	3,74	2,41	1	3	4,05	1	3,86	4,77

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
O99	P04	1	8,19	19,26	31,5	18,29	1	3,21	4,18	6,72	2,16	1	2,02	3,98	1	1,5	2,14
O99	T04	1	20,18	24,8	13,26	7,53	1	3,44	5,24	5,64	3,49	1	2,2	2,56	1	1,6	1,88
O99	P05	1	15,88	25,52	34,19	19,77	1	3,97	6,91	4,61	1,51	1	2,94	5,62	1	1,78	3,57
O99	T05	1	12,62	30,67	12,39	3,07	1	1,12	3,66	5,28	0,48	1	2,4	3,2	1	1,62	3,17
O99	P06	1	8,39	18,22	18,45	10,67	1	2,58	5,56	5,62	3,58	1	1,34	4,11	1	2,77	2,96
O99	T06	1	11,83	21,9	12,15	12,02	1	3,52	3,74	2	1,76	1	2,65	2,73	1	2,65	2,83
M00	O00	1	20,87	28,87	39,17	20,18	1	0	3	1,99	1,51	1	3,37	4,59	1	0,77	2,75
M00	M01	0	11,43	15,84	7,49	22,66	0	0,03	1,8	2,37	3,49	0	3,4	4,28	1	1,68	1,77
M00	O01	0	32,01	15,33	25,93	16,32	0	3,5	2,08	1,15	0,96	0	2,84	0,59	0	0,98	1,94
M00	M02	1	17,17	37,08	21,18	15,81	1	3,02	5,63	2,85	0,39	1	2,05	4,57	1	3,7	5,77
M00	O02	1	20,73	40,42	21,46	12,73	1	3,17	3,34	2,53	0,6	1	1,02	2,7	1	1,66	2,89
M00	M03	1	17,25	29,41	23,18	6,63	1	1,55	2,43	7,46	2,87	1	0,42	2,73	1	2,26	3,37
M00	P04	1	3,82	21,96	27,12	14,1	1	2,23	3,23	7,58	2,69	1	2,45	4,55	1	0,61	1,04
M00	T04	1	5,43	11,93	36,68	20,72	1	3,45	5,24	8,13	3,88	1	1,86	3,54	1	1,86	2,45
M00	P05	1	16,23	35,91	32,74	17,83	1	0,45	2,61	7,87	2,9	1	3,97	6,95	1	2,61	3,65
M00	T05	1	8,92	21,01	20,66	13,67	1	2,12	3,73	1,92	0,73	1	3,36	5,16	1	0,69	0,79
M00	P06	1	9,96	27,92	27,82	14,45	1	1,06	2,4	6,07	1,33	1	3,71	5,34	1	0,59	2,55
M00	T06	-1	15,38	9,72	6,28	18,03	-1	3,79	3,73	0,98	1,36	-1	3,68	3,06	-1	3,36	0,65
O00	M01	-1	23,09	5,17	8,4	20,56	-1	5,89	1,02	1,03	3,54	-1	3,28	2,72	-1	1,37	0,44
O00	O01	-1	22,18	12,29	11,16	29,17	-1	8,23	3,96	1,52	1,76	-1	3,16	2,2	-1	2,55	1,2
O00	M02	1	10,91	28,01	24,67	17,89	1	3,54	4,12	1,61	0,71	1	1,47	4,16	1	0,49	2,05

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
O00	O02	1	6,18	23,23	11,87	3,04	1	3,15	4,81	5,32	1,76	1	0,03	1,62	1	3,83	4,79
O00	M03	1	16,34	23,93	24,63	7,76	1	2,36	4,99	4,91	2,03	1	0,54	1,55	1	2,38	4,91
O00	P04	1	6,97	12,81	22,37	6,77	1	1,63	4,1	2,39	1,41	1	1,05	2,59	1	1,38	2,3
O00	T04	1	14,02	28,63	19,18	6,68	1	3,46	4,84	2,23	0,58	1	1,91	3,76	1	0,97	1,49
O00	P05	1	18,68	29,08	23,43	12,03	1	0,44	3,04	6,7	3,32	1	1,19	2,69	1	1,51	2,94
O00	T05	1	6,83	10,01	37,11	18,42	1	1,93	3,2	3,73	2,67	1	0,07	1,93	1	0,07	2,85
O00	P06	1	14,21	30,2	20,39	5,58	1	2,32	3,44	4,68	2,04	1	0,16	0,16	1	0,56	1,22
O00	T06	-1	29,18	10,66	7,63	14,32	-1	3,74	1,01	3,85	6,63	-1	4,95	2,28	-1	1,99	0,93
M01	O01	0	6,88	27,12	13,37	33,69	0	3,62	6,09	2,98	2,99	0	0,97	2,85	-1	2,89	0,54
M01	M02	1	19,53	22,74	34,17	16,83	1	3,27	5,85	3,55	1,84	1	1,47	3,36	1	1,21	3,42
M01	O02	1	6,74	11,66	29,57	11,52	1	0,38	1,68	2,33	0,46	1	2,99	3,26	1	3,12	4,29
M01	M03	1	16,02	23,3	22,71	16,12	1	1,98	3,42	4,51	1,73	1	1,18	1,52	1	2,43	5,27
M01	P04	1	16,72	23,71	32,25	14,06	1	2,16	4,16	2,72	2,17	1	2,29	2,98	1	3,45	6,06
M01	T04	1	6,01	24,3	25,53	12,17	1	2,79	3,61	2,79	0,18	1	3,05	3,12	1	1,97	4,06
M01	P05	1	9,54	17,23	27,17	12,55	1	0,76	2,97	7,13	2,36	1	3,2	3,53	1	0,11	0,22
M01	T05	1	10,73	18,66	36,84	20	1	1,32	2,75	3,43	0,16	1	0,78	1,44	1	2,05	3,06
M01	P06	1	4,17	24,29	31,71	17,14	1	1,85	3,77	4,97	0,33	1	2,97	3,2	1	2,11	3,64
M01	T06	-1	32,65	13,99	18,06	30,17	-1	1,99	0,6	3,07	3,96	-1	5,35	3,64	-1	3,42	0,45
O01	M02	1	15,18	23,21	31,72	11,9	1	2,73	2,78	5,83	1,73	1	3,72	3,75	1	2,97	5,4
O01	O02	1	4,05	19,86	21,99	10,12	1	2,69	3,59	4,16	1,31	1	1,08	1,47	1	3,83	4,77
O01	M03	1	7,97	26,08	23,04	10,74	1	0,98	3,05	2,79	1,43	1	3,82	4,44	1	1,52	2,08

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
O01	P04	1	5,61	13,43	37,12	16,44	1	0,14	1,66	4,06	3,7	1	1,38	4,23	1	0,27	2,72
O01	T04	1	13,55	24,44	17,9	7,43	1	3,34	3,63	5,74	2,81	1	0,4	1,12	1	2,59	4,13
O01	P05	1	14,35	25,44	11,88	8,34	1	3,07	4,25	1,07	0,22	1	1,05	2,67	1	0,74	3,58
O01	T05	1	17,5	26,23	13,33	5,82	1	2,13	2,53	2,6	1,09	1	1,28	1,89	1	2,28	5,13
O01	P06	1	19,92	24,87	36,37	16,12	1	3,47	5,83	3,3	0,35	1	1,23	1,66	1	1,17	3,81
O01	T06	-1	17,08	5,18	20,26	31,42	-1	5,41	1,71	0,6	2,23	-1	2,61	0,85	-1	4,22	1,99
M02	O02	0	15,55	21,57	7,65	25,29	0	2,75	0,87	2,47	1,99	0	2,03	2,77	1	0,59	1,71
M02	M03	-1	23,1	18,37	6,28	31,97	-1	3,24	2,23	0	4,06	-1	2,92	1,88	-1	2,17	1,62
M02	P04	-1	34,92	15,13	5,17	15,73	-1	2,03	0,74	2,89	5,3	-1	2,99	0,47	-1	5,65	3,62
M02	T04	-1	27,77	10,51	20,69	23,84	-1	2,59	1,79	2,81	3,07	-1	0,92	0,86	-1	2,13	2,13
M02	P05	-1	24,78	14,92	5,71	21,36	-1	7,96	3,91	1,36	3,82	-1	2,52	1,06	-1	5,07	3,13
M02	T05	-1	16,13	6,35	16,29	36,25	-1	4,23	3,84	0,05	2,4	-1	2,06	0,45	-1	3,73	2
M02	P06	-1	20,79	12,65	13,05	24,51	-1	3,14	1,7	1,82	3,17	-1	6,26	3,67	-1	3,46	1,32
M02	T06	-1	22,85	14,73	10,43	27,06	-1	5,14	1,81	0,28	1,44	-1	1,84	1,72	-1	3,53	3,16
O02	M03	-1	14,95	10,46	20,52	39,13	-1	4,04	1,27	1,67	4,34	-1	3,34	2,81	-1	0,23	0,01
O02	P04	-1	19,3	9,3	5,87	11,98	-1	3,5	0,92	2,64	4,43	-1	4,18	3,25	-1	2,82	2,24
O02	T04	-1	21,48	3,22	3,1	18,47	-1	4,85	1,32	0,87	2,98	-1	5,87	3,26	-1	1,69	1,2
O02	P05	-1	21,94	5,15	14,61	18,31	-1	6,73	2,24	0,99	1,76	-1	5,23	3,74	-1	2,25	1,81
O02	T05	-1	23,82	6,9	14,74	23,27	-1	2,58	2,4	2,74	4,01	-1	0,98	0,57	-1	4,69	2,62
O02	P06	-1	18,39	13,1	19,97	37,3	-1	6,76	3,87	1,78	4,68	-1	4,69	3,42	-1	1,44	0,64
O02	T06	-1	19,94	3,03	9,03	13,29	-1	6,38	2,08	3,09	5,15	-1	1,64	0,58	-1	4,23	2,28

Resultados Completos de los Tests de Dominancia . Argentina 1996-2006

F1	F2	SD1+	l1	cc+1	il1	icc+1	SD2++	l2	ccr+2	il2	icc+2	bdx	lx	crx	bdy	ly	cry
M03	P04	-1	20,92	3,13	20,81	30,56	-1	7,2	3,35	3,51	4,39	-1	2,54	1,55	-1	2,56	0,08
M03	T04	-1	34,03	16,09	15,04	34,78	-1	6,71	3,73	3,45	6,3	-1	5,45	3,33	-1	4,43	2,32
M03	P05	-1	19,66	8,21	15,1	33,41	-1	7,8	3,48	3,09	4	-1	1,78	0,41	-1	4,52	1,57
M03	T05	-1	27,49	18,67	18,27	23,21	-1	2,5	0,42	2,89	3,37	-1	4,93	2	-1	1,21	0,94
M03	P06	-1	8,32	3,71	15,66	30,76	-1	3,12	2,62	2,06	4,36	-1	5,84	3,76	-1	2,95	0,54
M03	T06	-1	26	5,95	16,51	21,96	-1	3,24	2,2	3,99	6,64	-1	3,62	3,17	-1	5,23	3,23
P04	T04	0	3,14	18,75	10,7	21,65	-1	6,38	3,69	3,88	6,25	-1	3,88	3,09	-1	5,86	2,96
P04	P05	-1	20,33	9,7	5,49	17,65	-1	4,18	1,76	3,22	4,34	-1	3,26	1,44	-1	3,33	0,77
P04	T05	0	14,9	32,79	14,56	19,27	-1	1,62	1,15	1,56	2,07	-1	2,61	1,13	-1	1,29	0,83
P04	P06	0	5,67	20,94	11,66	21,75	0	3,27	4,87	3,58	6,48	0	0,34	0,68	-1	1,44	3,64
P04	T06	-1	35,85	19,44	19,67	24,69	-1	6,53	3,23	2,3	4,89	-1	3,09	1,16	-1	3,99	3,53
T04	P05	-1	24,98	9,07	8,33	12,94	-1	5,67	1,28	1,55	4,28	-1	1,98	0,21	-1	5,43	3,83
T04	T05	0	11,9	31,14	14,31	25,4	0	0,13	0,32	1,51	4,01	1	3,11	5,46	1	1,65	3,66
T04	P06	0	16,52	35,75	18,01	28,5	1	1,21	2,45	5,17	3,48	1	1,08	1,1	1	1,25	3,7
T04	T06	-1	25,87	15,05	14,23	20,38	-1	5,88	2,45	2,04	4,92	-1	1,27	0,18	-1	4,95	3,91
P05	T05	1	10,67	30,08	16,16	9,03	1	0,33	0,5	5	2,98	1	2,37	2,45	1	2,53	3
P05	P06	1	16,16	32,9	15,07	4,46	1	2,69	5,29	6,11	2,12	1	3,36	3,8	1	1,8	4,16
P05	T06	-1	27,51	20,93	18,87	36,96	-1	3,95	0,73	2,1	3,66	-1	1,77	1,38	-1	4,53	2,53
T05	P06	0	14,02	20,65	3,29	11,26	1	0,61	1,85	3,46	1,76	1	3,15	3,74	1	0,96	1,12
T05	T06	-1	24,88	14,61	12,76	32,64	-1	6,61	3,6	0,84	3,23	-1	1,68	1,12	-1	2,96	0,03
P06	T06	-1	26,69	11,57	13,39	34,05	-1	2,74	2,29	3,45	4,64	-1	4,9	3,37	-1	3,76	0,92

Apéndice C

Códigos de Aplicaciones en MatLab

En cada programa, los inputs son variables estructuradas con 2 subvariables que se referencian de la forma $F1.h$ y $F1.y$, correspondiendo a las series completas de capital humano e ingreso para la onda especificada $(F1, F2)$. Además se tiene como input un nivel de confianza α y un número de iteraciones de bootstrap B.

La rutina *kappamn.m* es la que obtiene la distribución de bootstrap para los edps. Esta rutina trabaja con una estructura de computación en paralelo, a partir del comando *parfor* de Matlab. Se trabajó en algunos casos con procesadores de 2 núcleos (Intel®Core™ 2 Duo) y de 4 núcleos (Intel®Core™ 2 Quad). La utilización de multiples *labs* de Matlab reduce notablemente el tiempo insumido por las réplicas de bootstrap, y se recomienda ésta u otra técnica de computación en paralelo siempre que se deban implementar subrutinas de simulación y/o remuestreo.

La rutina *SD.m* toma los inputs antedichos, computa los edps para la muestra original y llama a la subrutina *kappamn.m* para obtener los valores críticos dependientes de los datos. Luego compara y da los resultados de dominancia.

También se diseñaron e implementaron las rutinas *bivacdf.m* y *intd.pdf*. La primera obtiene la cdf bivariada a partir de la muestra, la segunda integra una función bivariada en el recinto $[0,1] \times [0,1]$.

Nótese que la comprobación de las condiciones marginales en la dominancia estocástica a orden 2 requiere testear dos condiciones sobre distribuciones univariadas. En este sentido, la puesta en práctica del test a segundo orden involucra como subrutina una implementación del test de Barret y Donald [15, 16], aunque vía valores críticos.

Rutina *kappamn.m*

```
function kappamn= kappamn(F1,F2,B)

N1=length(F1.h);
N2=length(F2.y);
N=N1+N2;
h1=F1.h;
y1=F1.y;
h2=F2.h;
y2=F2.y;

H=[h1 y1; h2 y2];%pooled sample
kappamn=[];
parfor k=1:B
S=[];
%inicializa variables del loop hb1=[];
hb2=[];
yb1=[];
yb2=[];
d1=[];
d2=[];
dx=[];
dy=[];
KSmn=[];
lmn=[];
KFB=[];
LFB=[];
KGB=[];
LGB=[];
HFB=[];
HGB=[];

for j=1:N
s=floor(rand(1)*N+1);
S=[S s];
end


for l=1:N1
hb1(l)=H(S(l),1);
yb1(l)=H(S(l),2);
end
```

```
    for l=1:N2
hb2(l)=H(S(N1+l),1);
yb2(l)=H(S(N1+l),2);
end


    FB=bivacdf(hb1,yb1);
GB=bivacdf(hb2,yb2);
FBX=FB(:,101);
FBY=FB(101,:);
GBX=GB(:,101);
GBY=GB(101,:);

for i=1:101
for j=1:101
KFB(i,j)=-[FB(i,j)-FBX(i)-FBY(j)];
KGB(i,j)=-[GB(i,j)-GBX(i)-GBY(j)];
end
end

d1=KFB-KGB;
KSmn1=max(max(d1));
iKSmn1=-min(min(d1));
lmn1=sqrt((N1*N2)/N)*KSmn1;
ilmn1=sqrt((N1*N2)/N)*iKSmn1;


    %SD2

LFB=intd(KFB);
LGB=intd(KGB);
d2=LFB-LGB;
KSmn2=max(max(d2));
iKSmn2=-min(min(d2));
lmn2=sqrt((N1*N2)/N)*KSmn2;
ilmn2=sqrt((N1*N2)/N)*iKSmn2;

%BDmargin

HFB=intd(FB);
HGB=intd(GB);
HFBX=HFB(:,101);
HGBX=HGB(:,101);
```

```
HFBY=HFB(101,:);
HGBY=HGB(101,:);
dxB=HFBX-HGBX;
dyB=HFBY-HGBY;
lxB=sqrt((N1*N2)/N)*max(dxB);
lyB=sqrt((N1*N2)/N)*max(dyB);
ilxB=-sqrt((N1*N2)/N)*(-min(dxB));
ilyB=-sqrt((N1*N2)/N)*(-min(dyB));

%distrib. simulada de los edps

lmn=[lmn1; ilmn1; lmn2; ilmn2; lxB; ilxB; lyB; ilyB];
kappamn=[kappamn lmn];
end
```

Rutina *SD.m*

La rutina toma $F1$ y $F2$ e implementa los tests para ambos órdenes de dominancia y para ambos sentidos.

```
function result=SD(F1,F2,alpha,B)

%Bivariadas

F=bivacdf(F1.h,F1.y);
G=bivacdf(F2.h,F2.y);

M=length(F1.h);
N=length(F2.h);


    %marginales

FX=F(:,101);
FY=F(101,:);
GX=G(:,101);
GY=G(101,:);

%edpSD1

for i=1:101
for j=1:101
KF(i,j)=-[F(i,j)-FX(i)-FY(j)];
```

```
KG(i,j)=-[G(i,j)-GX(i)-GY(j)];
end
end

d1=KF-KG;

kol1=max(max(d1));
ikol1=-min(min(d1));

l1=sqrt((M*N)/(M+N))*kol1;
il1=sqrt((M*N)/(M+N))*ikol1;

%los H
HF=intd(F);
HG=intd(G);

HFX=HF(:,101);
HFY=HF(101,:);
HGX=HG(:,101);
HGY=HG(101,:);

%edpSD2

LF=intd(KF);
LG=intd(KG);

d2=LF-LG;

kol2=max(max(d2));
ikol2=-min(min(d2));

l2=sqrt((M*N)/(M+N))*kol2;
il2=sqrt((M*N)/(M+N))*ikol2;

%BDmargin

dx=HFX-HGX;
dy=HFY-HGY;
lx=sqrt((M*N)/(M+N))*max(dx);
ilx=-sqrt((M*N)/(M+N))*min(dx);
ly=sqrt((M*N)/(M+N))*max(dy);
ily=-sqrt((M*N)/(M+N))*min(dy);

%criticosSD1
```

```
kappa=kappamn(F1,F2,B);

kappa1=kappa(1,:);
ikappa1=kappa(2,:);

kappa1=sort(kappa1);
ikappa1=sort(ikappa1);

Nl=length(kappa1);
ncrit1=floor((1-alpha)*Nl);

ccritmas1=kappa1(ncrit1);
iccritmas1=ikappa1(ncrit1);

%criticosSD2
kappa2=kappa(3,:);
ikappa2=kappa(4,:);

kappa2=sort(kappa2);
ikappa2=sort(ikappa2);

N2=length(kappa2);
ncrit2=floor((1-alpha)*N2);

ccritmas2=kappa2(ncrit2);
iccritmas2=ikappa2(ncrit2);

%criticosBDmargin
kappa3=kappa(5,:);
ikappa3=kappa(6,:);

kappa3=sort(kappa3);
ikappa3=sort(ikappa3);

N3=length(kappa3);
ncrit3=floor((1-alpha)*N3);

crbdx=kappa3(ncrit3);
icrbdx=ikappa3(ncrit3);

kappa4=kappa(7,:);
ikappa4=kappa(8,:);
```

```
kappa4=sort(kappa4);
ikappa4=sort(ikappa4);

N4=length(kappa4);
ncrit4=floor((1-alpha)*N4);

crbdy=kappa4(ncrit4);
icrbdy=ikappa4(ncrit4);

%resultado de los tests

%SD1

if l1<=ccritmas1
SD1mas=1;
elseif il1<=iccritmas1
SD1mas=-1;
else
SD1mas=0;
end

%SD2

%BDmargin

if lx<=crbdx
bdx=1;
elseif ilx<=icrbdx
bdx=-1;
else
bdx=0; end

if ly<=crbdy
bdy=1;
elseif ily<=icrbdy
bdy=-1;
else
bdy=0; end

%joint distrib.
if l2<=ccritmas2 && bdx>=0 && bdy>=0
SD2mas=1;
elseif il2>=iccritmas2 && bdx<=0 && bdy<=0
SD2mas=.11;
```

```
else
SD2mas=0;
end

result=[SD1mas l1 ccritmas1 il1 iccritmas1 SD2mas l2 ccritmas2 il2 iccritmas2
bdx lx crbdx bdy ly crbdy]

end
```

Rutina *bivacdf.m*

```
function funcdist=bivacdf(x,y)

mx=mean(x);
my=mean(y);
x=x/mx;        % normaliza las distribuciones
y=y/my;
x=x-min(x);        % las soporta en [0,1]
x=x/max(x);
y=y-min(y);
y=y/max(y);

s=0:0.01:1;
t=0:0.01:1;
n=length(s);

for i=1:n
for j=1:n
F=0;
for k=1:length(x)
if x(k)<=s(i) && y(k)<=t(j)
F=F+1;
end
cum(i,j)=F/length(x);
end
end
end

figure
surf(s,t,cum)

funcdist=cum;
```

Rutina *intd.m*

```
function I=intd(f)

for i=1:101
for j=1:101
M=[];
for k=1:i
for l=1:j
M(k,l)=f(k,l);
end
end
I(i,j)=(0,01^2)*sum(sum(M)); %el paso es 0.01 en cada variable
end
end
```

Rutina *corrinc.m*

```
function incor= corrinc2(fx,fy,i1,j1,i2,j2)

fx=fx-min(fx);
fx=fx/max(fx);
fy=fy-min(fy);
fy=fy/max(fy);

s=0:0.01:1;
t=0:0.01:1;

n=length(fx);

q11=0;
q12=0;
q21=0;
q22=0;
k11=[];
k12=[];
k21=[];
k22=[];

for m=1:n
```

```
if fx(m)>=s(i1) && fx(m)<=s(i1+10) && fy(m)>=t(j1) && fy(m)<=t(j1+10)
q11=q11+1;
k11=[k11 m];
end
if fx(m)>=s(i1) && fx(m)<=s(i1+10) && fy(m)>=t(j2) && fy(m)<=t(j2+10)
q12=q12+1;
k12=[k12 m];
end
if fx(m)>=s(i2) && fx(m)<=s(i2+10) && fy(m)>=t(j1) && fy(m)<=t(j1+10)
q21=q21+1;
k21=[k21 m];
end
if fx(m)>=s(i2) && fx(m)<=s(i2+10) && fy(m)>=t(j2) && fy(m)<=t(j2+10)
q22=q22+1;
k22=[k22 m];
end
end

q=min(q12,q21);

epsi=q/n;

if q==0
display('no habia nada, ponga otros ij')
else

for l=1:q
fx(k12(l))=s(i2);
end

for l=1:q
fx(k21(l))=s(i1);
end
end

incor=[fx fy];
epsi

end
```

Bibliografía

[1] Acemoglu, D. & Autor, D., *Lectures in Labor Economics*, mimeo, disponible en http://economics.mit.edu/files/4689.

[2] Acemoglu, D. & Autor, D., *Skills, Tasks and Technologies: Implications for Employment and Earnings*, en David Card and Orley Ashenfelter (Eds.), *Handbook of Labor Economics 4, Part B*, North Holland, 2011.

[3] Aghion, P., *Schumpeterian Growth Theory and the Dynamics of Income Inequality*, Econometrica Vol.70, n3, pp. 855-882, 2002.

[4] Aldous, D. J., *Exchangeability and related topics. Ecole d'été de probabilités de Saint-Flour XIII*, Lecture Notes in Math. 1117.

[5] Alessina, A. & Rodrik, D., *Distributive Politics and Economic Growth*, NBER WP. 3668.

[6] Altimir, O. & Beccaria, L., *Distribución del Ingreso en Argentina*, en Heymann & Kosacoff (eds), *La Argentina de los Noventa. T1.*, EUDEBA, 2000. Springer, Berlin, 1985.

[7] Anderson, G., *Nonparametric Tests of Stochastic Dominance in Income Distributions*, Econometrica, vol 64, N°5,1183-1193, 1996.

[8] Arellano, M., *Panel Data Econometrics*, Advanced Texts in Econometrics, Oxford University Press, 2003.

[9] Arnold, B., *Pareto Distributions*, Second Edition, Monographs on Statistics and Applied Probability 140, CRC Press, 2015.

[10] Arrow, K.J., *Social Choice and Individual Values, Second edition*, Cowles Foundation Monographs Series, 1970.

[11] Ashenfelter, O. y Krueger, A., *Estimates of the Economic Return to Schooling from a New Sample of Twins*, The American Economic Review, Vol. 84: 1157-1173, 1994.

[12] Asselin, L-M., *Analysis of Multidimensional Poverty. Theory and Case Studies*, Springer, 2005.

[13] Atkinson, A., *On the measurement of inequality*, Journal of Economic Theory, v. 2, 244-263, 1970.

[14] Atkinson, A. B. & Bourguignon, F., *The Comparison of Multi-Dimensioned Distributions of Economic Status*, The Review of Economic Studies, Vol. 49 No. 2 (Apr. 1982), pp. 183-201.

[15] Barret, G. & Donald, S., *Consistent Tests for Stochastic Dominance*, Econometrica, Vol.73, n1, pp. 71-104, 2003.

[16] Barret, G. & Donald, S., *Consistent Nonparametric Tests for Lorenz Dominance*, Econometric Society Australasian Meeting 2004, 321.

[17] Basmann R.L., Hayes K.H., Slottje D.J. & Johnson J.D., *A general functional form for approximating the Lorenz curve*, Journal of Econometrics 43: 77-90, 1990.

[18] Becker, G., *Human Capital. A Theoretical and Empirical Analysis, with Special Reference to Education, 3rd Ed.*, University of Chicago Press, 1993.

[19] Behrman, J.R., *Labor Markets in Developing Countries*, en *Handbook of Labor Economics 3, Part B*, Orley C. Ashenfelter and David Card (Eds.), North Holland, 1999.

[20] Bergstrom, R., *Soldiers of Fortune?*, en Walter P. Heller, Ross M. Starr, David A. Starrett *Essays in Honor of Kenneth J. Arrow: Volume 2, Equilibrium Analysis*, Cambridge University Press 1986

[21] Bermúdez, I., *El Estado se convirtió en el único creador de empleo*, en iEco, 31/03/2013.

[22] Berti, P., Pratelli, L. & Rigo, P., *Limit Theorems for a Class of Identically Distributed Random Variables*, The Annals of Probability, Vol. 32, No. 3A, pp. 2029-2052, 2004.

[23] Berti, P., Pratelli, L. & Rigo, P., *Limit Theorems for Predictive Sequences of Random Variables*, Quaderni di Dipartimento, #23 (12-02) Dipartimento di economia politica e metodi quantitativi, Università degli studi di Pavia, Dicembre 2002.

[24] Bethlehem, J.,Cobben, F. & Schouten, B., *Handbook of Nonresponse in Household Surveys*, John Wiley and Sons, 2011.

[25] Bickel, P., *A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case*, Ann. Math. Statist., Volume 40, No. 1 (1969), 1-23.

[26] Bickel, P. & Wichura, M., *Convergence Criteria for Multiparameter Stochastic Processes and Some Applications*, Ann. Math. Stat. Vol.42. No.5, pp.1656.1670, 1971.

[27] Billingsley, P., *Convergence of Probability Measures*, Wiley NY, 1968.

[28] Bils, M. & Klenow, P., *Does Schooling Cause Growth*, The American Economic Review, vol. 90, pp. 1160-1183, 2000.

[29] Boadway, R. & Bruce, N., *Welfare Economics*, Basil Blackwell, Oxford, 1984.

[30] Bocco, A. et al., *Regresividad Tributaria y Distribución del Ingreso*, UNICEF-Losada, 1997.

[31] Bowles, S. & Sung-Ha Hwang, *Social Preferences and Public Economics: Mechanism design when social preferences depend on incentives*, University of Massachusetts - Amherst, Economics Department Working Paper Series, Working Paper 2008-06.

[32] Brufman, J. & Urbisaia, H., *Implicancias Macroeconómicas del Sistema Integrado de Jubilaciones y Pensiones*. Ed. Macchi, Bs. As, 1999.

[33] Brufman, J. & Perez, L., *Distribución Muestral de Índices de Desigualdad Económica: Enfoque no Paramétrico*, en XXVI Jornadas Nacionales de Docentes de Matemática de Facultades de Ciencias Económicas y Afines, Facultad de Ciencias Económicas Universidad Nacional de Misiones, Octubre 2011.

[34] James Buchanan & Gordon Tullock, *The Calculus of Consent*, The University of Michigan Press, Ann Arbor, 1965.

[35] Bühlmann, P. (1993), *The Blockwise Bootstrap in Time Series and Empirical Processes*, PhD. Diss. SWISS FEDERAL INSTITUTE OF TECHNOLOGY, ZÜRICH.

[36] Cabaña, A. y Cabaña, E., *Transformed Empirical Processes and Modified Kolmogorov-Smirnov Statistics for Multivariate Distributions*, Annals of Stat. v.25, pp. 2388-2409, 1997.

[37] Chakravarty, S., *Measuring Inequality: the Axiomatic Approach*, en Silber, J. (ed), *Handbook on Income Inequality Measurement*, Kluwer 1999.

[38] Yiling Chen , Laia J., Parkes D., Procaccia, A. *Truth, justice, and cake cutting*, Games and Economic Behavior 77 (2013) 284-297.

[39] Chow, Y. & Teicher, H., *Probability Theory: Independence, Interchangeability, Martingales, 3rd ed.*, Springer Texts in Statistics, Springer 2003.

[40] Clark, A. & Oswald, A., *Satisfaction and Comparison Income*, mimeo, 1995.

[41] Cornia, G. A. & Court, J., *Inequality, Growth and Poverty in the Era of Liberalization and Globalization*, World Institute for Development Economics Research, UNU, Policy Brief N4, 2001.

[42] Cowell, F. & Victoria-Feser, M.P. ,*Welfare Rankings in the Presence of contamined Data*, Econometrica, Vol.70, No.3, 1221-1233, 2002.

[43] Crawford, I., *A nonparametric Test of Stochastic Dominance in Multivariate Distributions*, Discussion Papers in Economics, DP 12/05, University of Surrey, 2005.

[44] Dagum, C., *Linking the Functional and Personal Distributions of Income*, en Silber, J. (ed), *Handbook on Income Inequality Measurement*, Kluwer 1999.

[45] Davidson, R. & Duclos, J-Y., *Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality*, Econometrica, Vol.68, No. 6, pp. 1435-1464, 2000.

[46] Deaton, A., *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*, Johns Hopkins University Press (published for the World Bank), 1997.

[47] Decancq, K., *Multidimensional Inequality and Dependence between its Dimensions*, (Extended Abstract) March 2007.

[48] Dehling, H., Mikosch, T., and Sørensen, M., editors, *Empirical Process Techniques for Dependent Data*, Birkhäuser, Boston, 2002.

[49] Dehling, H., Durieu, O. & Volny, D., *New Techniques for Empirical Processes of Dependent Data*, arXiv:0806.2941v2 [math.PR] 1 Oct 2008.

[50] de la Croix, D. & Lubrano, M., *The Trade-off Between Growth and Redistribution: ELIE in an Overlapping Generations Model*, en Gamel, C. & Lubrano, M. (Eds.), *On Kolm's Theory of Macrojustice. A Pluridisciplinary Forum of Exchange*, Springer, 2011.

[51] Devroye, L. & Lugosi, G., *Combinatorial Methods in Density Estimation*, Springer Series in Statistics, 2001.

[52] Diamand, M., *Hacia la Superación de las Restricciones al Crecimiento Económico Argentino*, Cuadernos del centro de Estudios de Pensamiento Económico y Social, 1988.

[53] Diez, H, Lasso de la Vega, M.C. & Urrutia, A.M., *Unit-Consistent Agregative Multidimensional Inequality Measures: A Characterization*, ECINEQ 2007-66, April 2007.

[54] A. Dixit & R. Pindyck, *Investment Under Uncertainty*, Princeton University Press 1993.

[55] Donsker, M., *Justification and Extension of Doob's Heuristic Approach to the Kolmogorov-Smirnov Theorems*, Annals of Math. Stat. v.23, pp. 277-281, 1952.

[56] Doob, J., *Heuristic Approach to the Kolmogorov-Smirnov Theorems*, Annals of Math. Stat. v.20, pp. 393-403, 1949.

[57] Duclos, J-Y., Esteban, J., Ray, D., *Polarization: Concepts, Measurement, Estimation*, Econometrica, Vol.72, No. 6, 1737-1772, 2004.

[58] Dudley, R., *Central Limit Theorems for Empirical Measures*, Annals of Prob. V.6, pp.899-929, 1978.

[59] Dudley, R., *A Course in Empirical Processes*, Ecole d'été de probabilités de Saint-Flour (1982), *Lecture Notes in Math* 1097, Springer, 1984.

[60] Dudley, R., *Universal Donsker Classes and Metric Entropy*, Annals of Prob. V.15, pp.1360-1326, 1987.

[61] Dudley, R., *Fréchet Differentiability, p-Variation and Uniform Donsker Classes*, Annals of Prob. V.20, pp.1968-1982, 1992.

[62] Dudley, R., *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics 63, Cambridge University Press, 1999.

[63] Dudley, R., *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics 74, Cambridge University Press, 2003.

[64] Dufour, J-M & Hsia, C., *Identification*, en *Microeconometrics*, S. Durlauf & L. Blume (eds.), Palgrave Mac Millan, 2010.

[65] Efron, B., *Bootstrap Methods: Another Look at the Jackkniffe*, Ann. Statist. **7**, 1-26 (1979).

[66] Efron, B. & Tibshirani, R., *An Introduction to the Bootstrap*, CRC Press, 1994.

[67] Epstein, L. & Tanny, S., *Increasing Generalized Correlation: A Definition and Some Economic Consequences*, The Canadian Journal of Economics / Revue Canadienne d'Economique, Vol. 13, No. 1, (Feb., 1980), pp. 16-34.

[68] Esteban, J., Ray, D., *On the Measurement of Polarization*, Econometrica, Vol.62, No4, 819-851, 1994.

[69] Evans, L.C., *Partial Differential Equations*, AMS, Graduate Studies in Mathematics, Vol.19, 1998.

[70] Feller, W., *An Introduction to Probability Theory and its Applications, Vol.II, 3rd. Ed*, John Wiley and Sons, 1971.

[71] Fields, G., *Distribution and Development, A New Look at the Developing World*, MIT Press, 2001.

[72] Fishburn, P., *Stochastic Dominance and Moments of Distributions*, Mathematics of Operations Research, Vol. 5 No. 1, pp. 94-100, Feb. 1980.

[73] Flores-Lagunes, A. & Light, A., *Measurement Error in Schooling: Evidence from Samples of Siblings and Identical Twins*, Department of Economics, University of Arizona, Working Paper 03-05, 2003.

[74] Friedman, J. y Rafsky, L., *Multivariate Generalization of the Wald-Wolfowitz and Smirnov Two Sample Tests*,

[75] Friedman, M. & Savage, L. J. *Utility Analysis of Choices Involving Risk*, Journal of Political Economy 56 (4): 279â304, 1948. doi:10.1086/256692. Annals of Stat. v.7, pp. 697-717, 1979.

[76] Fukuda-Parr, A., *The Human Development Paradigm: Operationalizing Sen's Ideas on Capabilities*, Feminist Economics Vol.9(2-3), pp.301-317, 2003.

[77] Gamel, C. & Lubrano, M. (Eds.), *On Kolm's Theory of Macrojustice. A Pluridisciplinary Forum of Exchange*, Springer, 2011.

[78] Gasparini, L., Marchionni, M. & Sosa Escudero, W., *Distribución del Ingreso en Argentina: Perspectivas y Efectos sobre el Bienestar*, Fundación Arcor, 2001.

[79] Gastwirth, J., *A General Definition of the Lorenz Curve*, Econometrica, Vol. 39, No. 6, pp. 1037-1039, Nov., 1971.

[80] Gastwirth, J., *The Estimation of the Lorenz Curve and Gini Index*, The Review of Economics and Statistics, Vol. 54, No. 3, pp. 306-316, Aug. 1972.

[81] Giné, E., *Lectures on some aspects of the bootstrap*, en *Lectures on Probability Theory and Statistics*, Ecole d'été de probabilités de Saint-Flour (1996), ed. P. Bernard, *Lecture Notes in Math* 1665, Springer, 1997.

[82] Giné, E. & Zinn, J., *Some limit theorems for empirical processes*, Annals of Prob. V.10, pp. 929-989, 1984.

[83] Giné, E. & Zinn, J., *Bootstrapping General Empirical Measures*, Annals of Prob. V.18, pp. 851-869, 1990.

[84] Giné, E. & Zinn, J., *Gaussian Characterization of Uniform Donsker Clases of Functions*, Annals of Prob. V.19, pp. 758-782, 1991.

[85] Giorgi, G., *Income Inequality Measurement: the Statistical Approach*, en Silber, J. (ed), *Handbook on Income Inequality Measurement*, Kluwer 1999.

[86] Goldthorpe, J. H., *Progress in Sociology: The Case of Social Mobility Research*, en Stefan Svallfors, *Analyzing Inequality. Life Chances and Social Mobility in Comparative Perspective*, Studies in Social Inequality, Stanford University Press 2005.

[87] Good, P.I., *Note on the Use and Misuse of Permutation Methods*, http://statisticsonline.info.

[88] Good, P.I., *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Springer series in statistics, 2005.

[89] Grasso, M., *A Dynamic Operationalization of Sen's Capability Approach*, Quaderno CRASL 3/2006.

[90] Gregory, R.G. & Borland, J., *Recent Development on Public Sector Labor Markets*, en *Handbook of Labor Economics, Volume 3C*, Edited by O. Ashenfelter and D. Card, North Holland, 1999.

[91] Griliches, Z., *Errors in Variables and Other Unobservables*, Econometrica, vol. 42, No6, pp. 971-998, 1974.

[92] Griliches, Z., *Estimating the Returns to Schooling: Some Econometric Problems, Econometrica*, vol. 45, No.1, pp. 1-22, 1977.

[93] Groot, W. & Maassen van den Brink, H., *The effects of Education on Health*, en *Human Capital Advances in Theory and Evidence*, edited by Joop Hartog and Henrietë Maassen van den Brink, Cambridge University Press, 2007.

[94] Grossman, M., *Education and Nonmarket Outcomes*, en

[95] Hadar, J. & Russell, W. R., *Rules for Ordering Uncertain Prospects*, Amer. Econ. Rev. 59, pp. 25-34, 1969.

[96] Hadar, J. & Russell, W. R., *Stochastic Dominance in Choice under Uncertainty*, en *Essays on Economic Behavior under Uncertainty*, ed. M. S. Balch, D. L. McFadden & Y. Wu, North Holland, Amsterdam, 1974.

[97] Hájek, J., Šidák, Z. & Sen, P., *Theory of Rank Tests*, Academic Press, 2004.

[98] Hall, P., *Methodology and Theory for the Bootstrap*, en Handbook of Econometrics, v. 4 ed. R. F. Engle & D. L. McFadden ch. 39, 1994.

[99] Hanoch, G. & Levy, H., *The Efficciency Analysis of Choices Involving Risk*, Rev. Econ. Studies 36, pp. 335-346, 1969.

[100] Härdle, W., *Applied Nonparametric Regression*, Econometric Society Monographs 19, Cambridge Univesity Press, 2002.

[101] Hill, T. & Morrison, K., *Cutting Cakes Carefully*, The College Mathematics Journal, Vol. 41, No. 4, pp. 281-288, September 2010.

[102] Hall, P. & Horowitz, J., *Bootstrap Critical Values for Tests Based on the Generalized-Method-of-Moments Estimators*, Econometrica, Vol.64, n°4, 891-916, 1996.

[103] Heer, B. & Maussner, A., *Dynamic General Equlibrium Modelling. Computational Methods and Applications*, Springer 2005.

[104] Herbst, E. & Schorfheide, F., *Evaluating DSGE Model Forecasts of Comovements*, mimeo.

[105] Hewitt, E., *Integration by Parts for Stieljes Integrals*, Am. Math. Monthly, Vol.67, No.5 (May, 1960), pp. 419-423.

[106] Hinrichs, A. & Novak, E., *The Curse of Dimensionality for Numerical Integration of Smooth Functions*, arxiv:1211.0871v2 [math.NA] 16 Apr 2013.

[107] Horváth, L., Kokoszka, P. & Zitikis, R., *Testing for Stochastic Dominance Using the Weighted McFadden Type Statistic*, Journal of Econometrics, 133 (1), pp. 191-205, 2006.

[108] Huggett, M., Ventura, G. & Yaron A., *Human Capital and Earning Distribution Dynamics*, Journal of Monetary Economics 53, pp.265-290, 2006.

[109] James, B., *Probabilidade: um Curso em Nível Intermediário*, IMPA 1981.

[110] Johansonn, P-O., *An Introduction to Modern Welfare Economics*, Cambridge University Press, 2003.

[111] Jueguen, F., *Sin la intervención en el Indec, la inflación oficial serÃa hoy casi cuatro veces mayor*, en La Nación, 1 de Febrero 2014.

[112] Justino, P., *Redistribution, Inequality and Political Conflict*, PRUS WP. No. 18, 2004.

[113] Khan, M. & Yannelis, N., *Equilibria in Markets with a Continuum of Agents and Commodities*, en M. Ali Khan & Nicholas C. Yannelis (Eds.), *Equilibrium Theory in Infinite Dimensional Spaces*, Springer, 1991.

[114] Kakwani, N. & Podder, N., *Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations*, Econometrica, Vol. 44, No. 1, pp. 137-148, Jan. 1976.

[115] Kallenberg, O., *Probabilistic Symmetries and Invariance Principles*, Springer, 2005.

[116] Kallenberg, O., *Foundations of Modern Probabilities 2nd Ed.*, Springer, 2002.

[117] Tae Yoon Kim, *On Tail Probabilities of Kolmogorov-Smirnov Statistics Based on Uniform Mixing Processes*, Statistics & Probability Letters 43 (1999) 217-223.

[118] Klecan, L., McFadden, R. & McFadden, D., *A Robust Test for Stochastic Dominance*, Working Paper, Dept. Economics, MIT, 1991.

[119] Kochar, S. & Korwar, R., *On Random Sampling Without Replacement From a Finite Population*, Ann. Inst. Statist. Math., Vol.53, No.3, pp. 631-646, (2001).

[120] Koenker, R., *Quantile Regression*, Econometric Society Monographs (ESM 38), Cambridge University Press 2005.

[121] Kolm, S.-Ch, *Multidimensional Egalitarianisms*, Quarterly Journal of Economics, 91, No.1, pp. 1-13, 1977.

[122] Kolm, S.-Ch, *Macrojustice: The political economy of fairness*, Cambridge University Press.

[123] G. Koshevoy & K. Mosler, *The Lorenz Zonoid of a Multivariate Distribution*, Journal of American Statistical Association, Vol. 91 No. 434, pp. 873-882, Jun. 1996.

[124] Kosorok, M., *Introduction to Empirical Processes and Semiparametric Inference*, Springer, 2006.

[125] Kuznets, S., *Economic growth and income inequality*, American Economic Review 45:1 28, 1955.

[126] Lambert, P.J., *The Distribution and Redistribution of Income, 3rd Ed.*, Manchester University Press, 2001.

[127] Lange, F. & Topel, R., *The Social Value of Education and Human Capital*, en en Handbook of the Economics of Education Vol.1, E. Hanushek and F. Welch (Eds.), North Holland, 2006.

[128] Levy, H, *Stochastic Dominance. Investment Decision Making under Uncertainty, Second Edition*, Springer, 2006.

[129] Levy, H. & Paroush, J., *Toward Multivariate Efficiency Criteria*, Journal of Economic Theory, 7, pp.129-142, 1974.

[130] Lindvall, T., *Lectures on The coupling Method*, Dover 1992.

[131] Linton, O., Maasoumi, E. & Wang Y-J, *Testing for Stochastic Dominance under General Sampling Schemes*, LSE STICERD Research Paper No. EM/2003/466.

[132] Linton, O., Song, K. & Wang Y-J, *An Improved Bootstrap Test of Stochastic Dominance*, mimeo.

[133] List, C. *Multidimensional Inequality Measurement: A Proposal*, Nuffield College Working Paper in Economics, 19 November 1999.

[134] Llach, J. J., *Reconstrucción o Estancamiento*, Ed. Tesis, Bs. As. 1987.

[135] Lochner, L., *Nonproduction Benefits of Education: Crime, Health, and Good Citizenship*, en Handbook of Economics of Education Vol. 4, Hanushek, E., Machin, S. & Woessmann, L., Elsevier, 2011.

[136] López Gottig, R., *El crecimiento del empleo público en la Argentina kirchnerista*, Infobae, 13/06/2014.

[137] Lubrano, M., *The Redistributive Aspects of ELIE: A Simulation Approach*, en Gamel, C. & Lubrano, M. (Eds.), *On Kolm's Theory of Macrojustice. A Pluridisciplinary Forum of Exchange*, Springer, 2011.

[138] Lucas, R.E., *On the Mechanics of Economic Development*, Journal of Monetary Economics, 22(1), pp.3-42, July 1988.

[139] Lugo, M. A., *Comparing Multidomensional Indices of Inequality: methods and application*, July 2005.

[140] Maasoumi, E., *The Measurement and Decomposition of Multi-Dimensional Inequality*, Econometrica, Vol. 54 No. 4, pp. 991-997, Jul. 1986.

[141] Maasoumi, E. & Heshmati, A., *Evaluating Dominance Ranking PSID Incomes by Various Household Attributes*, mimeo, 2003.

[142] Margot, D., *Rendimientos a la educación en Argentina: Un análisis de cohortes*, Fac. Cs. Ecs. UNLP, Documento de Trabajo n33, Julio 2001.

[143] Marinucci, D., *The Empirical Process for Bivariate Sequences With Long Memory*, mimeo 2002.

[144] Marks G. et al., *The Measurement of Socioeconomic Background for the Reporting of Nationally Comparable Outcomes of Schooling*, Draft Report, National Education Performance Monitoring Taskforce, Australian Council for Educational Research & Sociology Program Research School of Social Sciences Australian National University, March 2000.

[145] Marx, K., *El Capital. Vol.1*, FCE, México, 1992.

[146] Mas-Colell, A., Whinston, M. & Green, J., *Microeconomic Theory*, Oxford University Press, 1995.

[147] Maskin, E. & Sen, A., *The Arrow Impossibility Theorem*, Columbia University Press, 2014.

[148] McAll, C., *Class, Ethnicity and Social Class*, McGill University Press, 1992.

[149] McFadden, D., *Testing for Stochastic Dominance*, en Fomby, Th. B & Tae Kun Seo, *Studies in the Economics of Uncertainty. In Honor of Josef Hadar*, Springer, 1989.

[150] Meyer ,M. & Strulovici, B., *Increasing interdependence of multivariate distributions*, Journal of Economic Theory 147, 1460-1489, 2012.

[151] Midlarsky, M., *A Distribution of Extreme Inequality With Applications to Conflict Behavior: A Geometric Derivation of the Pareto Distribution*, Mathl Comput. Modelling, Vol. 12, No. 4/5, pp. 577-587, 1989.

[152] Mincer, J., *Schooling, Experience and Earnings*, NBER, Columbia University Press, 1974.

[153] Miyazawa Kensuke, *Measuring Human Capital in Japan*, RIETI Discussion Paper Series 11-E-037, March 2011.

[154] Muliere, P. & Scarsini, M., *A Note on Stochastic Dominance and Inequality Measures*, Technical Report N.232, Department of Statistics, Stanford University.

[155] Neal, D. & Rosen, S., *Theories on the Distribution of Earnings*, en Handbook of Income Distribution Vol.1, A.B. Atkinson & F. Bourguignon (eds), North Holland, 2007.

[156] Nelsen, R., *An Introduction to Copulas*, Springer Series in Statistics, 2006.s

[157] Nilsson, T., *Measuring Changes in Multidimensional Inequality - An Empirical Application*, Mimeo, 18 Aug. 2007.

[158] Nygård, F. & Sandström, A., *Measuring Income Inequality*, Almqvist & Wiksell International, Stockholm-Swede, 1981.

[159] Nygård, F, Sanström, A., *The Estimation of the Gini and Entropy Inequality Parameters in Finite Populations*, Journal of Official Statistics V1, pp. 399-412, 1985.

[160] Nygård, F, Sanström, A., *Income Inequality Measures Based on Sample Surveys*, Journal of Econometrics, V42, pp. 81-95, 1989.

[161] O'Brien, G. & Scarsini, M., *Multivariate Stochastic Dominance and Moments*, Mathematics of Operations Research, Vol.16, No.2, may 1991.

[162] OECD, *Growing Unequal? Income Distribution and Poverty in OECD Countries*, Paris, 2008.

[163] Pagan, A. & Ullah, A., *Nonparametric Econometrics*, Themes in Modern Econometrics, Cambridge Univesity Press, 1999.

[164] Parsons, T., *Equality and Inequality in Modern Society, or Social Stratification Revisited*, Sociological Inquiry Volume 40 issue 2 1970.

[165] Paxson, Ch., *Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand*, The American Economic Review, Vol. 82, No. 1, pp. 15-33, Mar. 1992.

[166] Perez, L., *Polarización del Ingreso en la Argentina. 1999-2004*, Seminario de Integración y Aplicación, Facultad de Ciencias Económicas, UBA. Mimeo, disponible en http://finanzas.x10host.com/Perez2005.pdf.

[167] Perez, L., *Estimación de las Distribuciones Conjuntas de Ingreso y Capital Humano en Argentina (1996-2006)*, en Documentos de Trabajo, Año 9 No.1, Publicación de la Sección de Investigaciones en Métodos Cuantitativos para la Gestión Prof. Dr. Fausto I. Toranzos, FCE, UBA, Junio de 2013.

[168] Platón, *Diálogos*, Ed. Porrúa, México, 1991.

[169] Plug, E., *Are successful parents the secret to success?*, en *Human Capital Advances in Theory and Evidence*, edited by Joop Hartog and Henrietë Maassen van den Brink, Cambridge University Press, 2007.

[170] Polachek, S., *Earnings Over the Life Cycle: The Mincer Earnings Function and Its Applications*, Foundations and Trends in Microeconomics, now Publishers Inc., 2008.

[171] Pollard, D., *A Central Limit Theorem for Empirical Processes*, J. Austral. Math. Soc. Ser. A **33**, 235-248.

[172] Pollard, D., *Convergence of Stochastic Processes*, Springer 1984.

[173] Psacharopoulos, G., *Returns to Investment in Education: A Global Update World Development*, September 1994, 22(9), pp. 1325â43.

[174] Putterman, L., Roemer, J. & Silvestre, J, *Does Egalitarianism Have a Future?*, Journal of Economic Literature, Vol. XXXVI, pp.861-902, June 1998.

[175] Radulović, D., *On the Bootstrap and Empirical Processes for Dependent Sequences*, en Dehling, H., Mikosch, T., and Sørensen, M., editors, *Empirical Process Techniques for Dependent Data*, Birkhäuser, Boston, 2002.

[176] Rawls, J., *The Theory of Justice*, Harvard University Press, Cambridge, 1971.

[177] Ray, D., *Economía del Desarrollo*, Antoni Bosch Editor, Barcelona 1998.

[178] Rietveld, P.,*Multidimensional Inequality Comparisons. On Aggravation and Mitigation of Inequalities*, Economics Letters 32, pp.187-192, 1990.

[179] Rothschild, M. & Stiglitz, J., *Increasing Risk I: A Definition*, J. Econ. Theory 2, pp. 225-243, 1970.

[180] Rothschild, M. & Stiglitz, J., *Some Further Results on the Measurement of Inequality*, J. Econ. Theory 6, pp. 188-204, 1973.

[181] Rousseau, J-J., *Dicours sur l'origine el les fondements de l'inegalité parmi les hommes*, Librairie Générale Française, 1996.

[182] Rubinstein, Y. & Weiss, Y, *Post Schooling Wage Growth: Investment, Search and Learning*, en Handbook of the Economics of Education Vol.1, E. Hanushek and F. Welch (Eds.), North Holland, 2006.

[183] Rudelson, M. & Vershynin, R., *Combinatorics of Random Processes and Sections of Convex Bodies*, arXiv:math.FA/0404192 v1, Apr. 2004.

[184] Russell, W. R. & Seo, T.K., *Ordering Uncertain Prospects: The Multivariate Utility Functions Case*, The Review of Economic Studies, Vol.45, No.3 (Oct., 1978), pp. 605-610.

[185] Ryu, H.K. & Slottje, D. J., *Parametric Approximations of the Lorenz Curve*, en Silber, J. (ed), *Handbook on Income Inequality Measurement*, Kluwer 1999.

[186] Saks, S., *Theorie de l'Integrale*, Monografie Matemattczne, Warszawa-Lwów, 1937.

[187] Salanie, B., *The Economics of Taxation*, MIT Press, 2000.

[188] Salvia, A. & Vera, J., *Segmentación Laboral y Desigualdad Económica en la Argentina (1992-2010)*, presentado en la Fac. Cs. Económicas, UBA., Nov. 2011.

[189] Sarabia, J.M. & Jordá, V., *Modeling Bivariate Lorenz Curves with Applications to Multidimensional Inequality in Well-Being*, presentado en Fifth meeting of the Society for the Study of Economic Inequality ECINEQ, Bari, Italy, July 22-24, 2013.

[190] Savaglio, E., *Multidimensional Inequality: A Survey*, Feb 2004, Mimeo.

[191] Savvides, A. & Stengos, T., *Human Capital and Economic Growth*, Stanford University Press, 2009.

[192] Scarsini, M., *Dominance Conditions for Utility Functions with Multivariate Risk Aversion*, Technical Report N.233, Department of Statistics, Stanford University.

[193] Scarsini, M., *Multivariate Stochastic Dominance With Fixed Dependence Structure*, Technical Report No.255, Department of Statistics, Stanford Univesity, 1988.

[194] Schader, M. & Schmid F., *Fitting Parametric Lorenz Curves to Grouped Income Distributions. A Critical Note.*, Empirical Economics, Vol. 19, Number 3 (1994), 361-370.

[195] Schmid, P., *On the Kolmogorov and Smirnov Limit Theorems for Discontinuous Distribution Functions*, Ann. Math. Stat. Vol.29, No.24, pp-1011-1027, Dec.1958.

[196] Sen, A., *Desarrollo y Libertad*, Ed. Planeta, Buenos Aires 2000.

[197] Sen, A., *Inequality Reexamined*, Claredon Press, 2006.

[198] Sen, A., *Social Justice and Distribution of Income*, en Handbook of Income Distribution Vol.1, A.B. Atkinson & F. Bourguignon (eds), North Holland, 2007.

[199] Sen, A. & Foster, J., *On Economic Inequality*, Claredon Paperbacks, Oxford University Press, 1997.

[200] Sen, P.K., *Weak Convergence of Multidimensional Empirical Processes for Stationary $\phi -$ mixing Processes*, Ann. Prob. 1974, Vol. 2, No. 1, 147-154.

[201] Sen, P.K., *An Almost Sure Invariance Principle for Kolmogorov Smirnov Statistics*, The Annals of Probability, Vol.1, No.3, pp.488-496, Jun. 1973.

[202] Sheehy, A. & J. A. Wellner, *Uniform Donsker classes of functions*, Annals of Probability 20, 1983-2030, 1992.

[203] Shilov, G.E. & Gurevich, B.L., *Integral, Measure & Derivative: A Unified Approach*, Dover 1977.

[204] Shorrocks, A., *Ranking Income Distributions*, Economica, 50, 1-17, 1983.

[205] Shorrocks, A., *Approximating Unanimity Orderings: An Application to Lorenz Dominance*, Discussion Paper Series No. 443, Department of Economics, University of Essex, June 1995.

[206] Shorrocks, A. y Foster, J., *Transfer Sensitive Inequality Measurements*, RES, 54, 485-97, 1987.

[207] Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill, 1956.

[208] Simchi-Levi, D., Chen, X. & Bramel, J., *The Logic of Logistics. Theory Algorithms and Applications for Logistics and Supply Chain Management*, Springer 2005.

[209] Simon, C. & Blume, L., *Mathematics for Economists*, W. W. Norton & Co., 1994.

[210] Simpson, P., *Note on the Estimation of a Bivariate Distribution Function*, Ann. Math. Statist. Volume 22, No. 3 (1951), pp. 476-478.

[211] Slottje, D., *The Structure of Earnings and The Measurement of Income Inequality*, Contributions to Economics Analysis 184, Elsevier, 1989.

[212] Smirnov, N., *Table for Estimating the Goodness of Fit of Empirical Distributions*, Ann. Math. Statist., Volume 19, No. 2 (1948), 279-281.

[213] Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. 1, Liberty Classics, 1981.

[214] Steinhaus, H.*The problem of fair division*, Econometrica, Vol.16, No. 1., January 1948.

[215] Joseph Stiglitz, *La Economía del Sector Público*, 3^a Ed., Cap. 1. Antoni Bosch, 2000.

[216] Suresh, K.P. & Chandrashekara, S., *Sample size estimation and power analysis for clinical research studies*, J Hum Reprod Sci, 5(1), pp.7-13, 2012.

[217] Talagrand, M., *Donsker Classes and Random Geometry*, Annals of Prob. V.15, pp. 1327-1338, 1987.

[218] Tchen, A., *Inequalities for Distributions With Given Marginals*, The Annals of Probability, Vol.8, No.4, pp.814-827, 1980.

[219] Thomson, W. & Varian, H., *Theories of Justice Based on Symmetry*, en *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, Eds. Leonid Hurwicz, David Schmeidler, Hugo Sonnenschein, Cambridge University Press, 1985.

[220] Thorisson, H., *Coupling, Stationarity, and Regeneration*, Springer, Probability and its Applications, 2000.

[221] Topel, R., *Labor Markets and Economic Growth*, en *Handbook of Labor Economics, Volume 3C*, Edited by O. Ashenfelter and D. Card, North Holland, 1999.

[222] Topkis, D., *Supermodularity and Complementarity*, Princeton University Press, 1998.

[223] Tsui, K.Y., *Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach*, J. Ec. Theo. **67** (1995), 251-265.

[224] Tsui, K.Y., *Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation*, Soc Choice Welfare 16: 145-157, 1999.

[225] United Nations, *Household Sample Surveys in Developing and Transition Countries*, Department of Economic and Social Affairs Statistics Division, Studies in Methods Series F No. 96, New Yor, 2005.

[226] van de Geer, S., *Empirical Processes in M-Estimation*, Cambridge Series in Statistical and Probabilistic Mathematics, CUP, 2009.

[227] van de Geer, S., *On Hoeffding's Inequality for Dependent Random Variables*, en Dehling, H., Mikosch, T., and Sørensen, M., editors, *Empirical Process Techniques for Dependent Data*, Birkhäuser, Boston, 2002.

[228] van der Vaart W. & Wellner, J., *Weak Convergence and Empirical Processes. With Applications to Statistics*, Springer Series in Statistics, Springer NY, 2000.

[229] Weber, M., *Economía y Sociedad*, Fondo de Cultura Económica, Méx. DF, 1996.

[230] Weeden, K & Grusky, D., *Are There Any Big Classes at All?*, en Social Inequality Vol.22, Research in Social Stratification and Mobility, Elsevier, 2005.

[231] Weiss, Y., *The Determination of Life Cycle Earnings: A survey*, en *Handbook of Labor Economics, Volume I*, Edited by O. Ashenfelter and R. Layard, Elsevier Science Publishers B V, 1986.

[232] Wellner, J., *Empirical Processes in Action: A Review*, International Statistical Review, Vol. 60, No. 3, pp. 247-269, 1992.

[233] Weymark, J., *The Normative Approach to the Measurement of Multidimensional Inequality*, Working Paper No. 03-W14R, Vanderbilt University, 2004.

[234] Wheeden, R. & Zygmund, A., *Measure and Integral. An Introduction to Real Analysis*, Marcel Dekker, 1977.

[235] Halbert White, *Asymptotic Theory for Econometricians*, Academic Press, 2001.

[236] Willis, R.J., *Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions*, en en *Handbook of Labor Economics, Volume I*, Edited by O. Ashenfelter and R. Layard, Elsevier Science Publishers B V, 1986.

[237] Woelfel, J. & Murero, M., *Spaces and Networks. Concepts for Social Stratification*, en Social Inequality Vol.22, Research in Social Stratification and Mobility, Elsevier, 2005.

[238] Wößmann, L., *Specifying Human Capital: A Review, Some Extensions, and Development Effects*, Kiel Working Paper No. 1007, October 2000.

[239] Wooldridge, J.M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

[240] Adonis Yatchew, *Semiparametric Regression for the Applied Econometrician*, Themes in Modern Econometrics, Cambridge Univesity Press, 2003.

[241] Yohai, V., *Notas de Probabilidades y Estadística*, mimeo, 2006.

[242] Zheng, B., Formby, J., Smith, W. & Chow, K *Inequality Orderings, Normalized Stochastic Dominance, and Statistical Inference*, Journal of Business & Economic Statistics, Vol. 18, No. 4, pp. 479-488, Oct. 2000.

[243] Zinn, J., *A Note on the Central Limit Theorem in Banach spaces*, Annals of Prob. V.5, pp. 283-296, 1977.

[244] Zoli, C., *Intersecting generalized Lorenz curves and the Gini index*, Soc. Choice & Welfare (1999) 16: pp.183-196, 1999.