

Abril 2015

Modelando time to default sensible al contexto sistémico en carteras de consumo

Maestrando Act. Javier E. Barraza

Director Mg. Act. Darío Bacchini

Modelando time to default sensible al contexto sistémico en carteras de consumo

Javier E. Barraza

Agradecimientos

Quiero agradecer sinceramente a todas aquellas personas que de alguna manera contribuyeron para la realización de este trabajo.

A Darío Bacchini por su ayuda siempre dispuesta como Director.

A mis compañeros Melisa Prieto, Sabrina Roiter, Miguel Perucca, Juan Agustín Pollio, Nicolás Botbol, Martín Masci, y Ezequiel Cufone por haberme brindado ideas y recomendaciones respecto a esta Investigación.

A mi familia: Lorena, Bibiana, Luis, Francisco, Rosa y Alberto por el apoyo y el optimismo desde el inicio de esta Maestría.

A Banco Supervielle por brindarme la posibilidad de complementar mi labor con la Investigación.

A mis familiares y amigos de la vida por la confianza y el aliento.

Al Centro de Investigación en Métodos Cuantitativos Aplicados a la Economía y la Gestión de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires por promover el Grupo de Calibración en Ingeniería Financiera.

A todos ellos GRACIAS.

Tabla de contenido

Resumen	1
1. Introducción - Estado actual del conocimiento	2
1.1. Un poco de historia	2
1.2. Problema de Investigación	3
1.3. Modelos de Supervivencia	4
1.4. Objetivos del trabajo	4
2. Descripción de variables disponibles	5
2.1. Armado de Base de Datos	5
2.2. Variables Disponibles - Individuales	6
2.3. Variables Disponibles - Sistémicas.....	7
3. Análisis Bivariados	8
3.1. Flag de Campaña	8
3.2. Score de Bureau	9
4. Análisis Multivariados - Modelo Logístico	11
4.1. Definición del modelo de Regresión Logística	11
4.2. Resultados de la calibración	12
4.3. Ejemplos de aplicación	13
4.4. Medidas usuales de performance	14
4.5. Valuación económica de la performance	19
5. Análisis Multivariados - Modelo de Supervivencia	23
5.1. Definición del Modelo de Supervivencia	23
5.2. Resultados de la calibración	25
5.3. Ejemplos de aplicación	26
5.4. Medidas usuales de performance	28
5.5. Valuación económica de la performance	32
6. Análisis Multivariados - Modelo de Supervivencia sensible a la Mora Sistémica	33
6.1. Definición del Modelo de Supervivencia Dinámico	33
6.2. Tratamiento diferencial de la base	33
6.3. Indicador sistémico seleccionado.....	34
6.4. Resultados de la calibración	38
6.5. Ejemplos de aplicación	39
6.6. Medidas usuales de performance	41
6.7. Valuación económica de la performance	46
6.8. Sensibilidad de la PD al contexto sistémico.....	49
7. Conclusiones y futuras líneas de investigación	50
Referencias	52

Modelando time to default sensible al contexto sistémico en carteras de consumo

Javier E. Barraza

Resumen

En este trabajo se presentan, calibran y comparan tres modelos de scoring aplicados a la misma cartera de consumo, a saber: i) Modelo Logístico (es el modelo más utilizado en este tipo de carteras); ii) Modelo de Supervivencia Estático (Modelo Exponencial de Riesgos Proporcionales de Cox, el cual explica el tiempo que transcurre hasta que los individuos incurren en default); y iii) Modelo de Supervivencia Dinámico o sensible al contexto sistémico (se modifica el modelo anterior, incorporando un indicador de mora sistémica que modifica las estimaciones de las curvas de supervivencia a medida que pasa el tiempo).

Los análisis usuales de performance estadística muestran una superioridad marginal del Modelo de Supervivencia Estático frente al Modelo Logístico, ambos utilizando las mismas variables explicativas. La performance estadística del Modelo Dinámico medida de la misma manera es inferior a la de los otros modelos.

A su vez, cuando se realiza un análisis de Beneficios Económicos durante toda la vida del crédito, el Modelo Dinámico puede resultar dominante respecto de los otros dos, aunque su dominancia es sensible al nivel de pérdida o LGD.

1. Introducción - Estado actual del conocimiento

La estimación de probabilidades de eventos es un objetivo básico de la estadística. Las aplicaciones son utilizadas en diversos campos de la ciencia como la medicina, las ciencias actuariales, etc.

En el caso del estudio de un determinado evento binario se puede disponer de muchas variables que pueden contribuir a hacer más o menos frecuente su ocurrencia, pero estas variables no definen el evento de modo determinístico. Si además se dispone de masa crítica (es decir, cantidad suficiente de individuos expuestos a dicho evento), el problema se transforma en un asunto econométrico más que probabilístico. Esto es, la probabilidad del evento en cada grupo de individuos se puede calcular con definiciones clásicas de la Teoría de la Probabilidad, pero la medición de la influencia de las variables en el fenómeno se debe realizar mediante un modelo econométrico. El mismo permitiría medir la influencia de las variables explicativas en la probabilidad de evento ya calculada.

En particular, para el estudio del comportamiento de carteras de créditos, se dan muchas condiciones que hacen plausible el empleo de técnicas econométricas de explicación de dicho fenómeno.

En este trabajo se calibran modelos econométricos usuales para explicar el comportamiento de los individuos de una cartera de créditos, y se proponen otros que no son de usual utilización en el campo del riesgo crediticio, comparando los resultados de ambos modelos.

1.1. Un poco de historia

1.1.1. Modelos binarios

Las primeras metodologías de estimación para eventos binarios se desarrollaron con aproximaciones recursivas aplicadas al estudio de enfermedades coronarias (Walker & Duncan, 1967). Pocos años después se comenzó a hacer énfasis en los métodos de máxima verosimilitud en las estimaciones (Cox, 1970). Estos dos desarrollos fundaron bases para el estudio pormenorizado de este tipo de fenómenos, que llega aún hasta nuestros días. Asimismo, las necesidades específicas de cada campo de aplicación dieron retroalimentación a los modelos y generaron nuevos desarrollos.

Se desarrollaron métodos que permiten estimar la probabilidad del evento bajo un supuesto de distribución normal de la misma (modelos *probit*) o bien exponencial (modelos *logit* o de *regresión logística*, indistintamente). La diferencia entre los mismos radica en la mayor facilidad de interpretación de parámetros de los modelos *logísticos* (Scott Long & Freese, 2001). El modelado incluye *análisis multivariados*, donde todas las variables independientes explican la tasa de ocurrencia del evento, y *análisis bivariados* donde cada variable explicativa debe tener algún tipo de relación con la tasa de ocurrencia del evento en los grupos (Bessis, 2002). En general se eligen los parámetros de manera tal que se maximice una función de verosimilitud construida a partir de las probabilidades calculadas con el modelo.

1.1.2. Tiempo hasta la ocurrencia

Desarrollos posteriores al análisis de datos binarios (Cox & Oakes, 1984) llevaron a esbozar no sólo el modelado de una probabilidad de evento, sino también del tiempo que transcurre

hasta que dicho evento sucede. Este tipo de análisis es fundamental para el control de tratamientos médicos. También lo es para el objeto de estudio de este trabajo, que es la gestión de una cartera con riesgo crediticio. La construcción de este tipo de modelos sienta sus pilares fundamentales en los análisis bivariados y multivariados, pero no se analiza la tasa de ocurrencia sino el tiempo de permanencia de los individuos en la cartera sin incurrir en default, es decir el tiempo de “supervivencia”. Existirán entonces *observaciones censuradas*, ya que para una parte de la población en cuestión no ocurrirá el evento, o bien ya ha ocurrido una vez iniciado el período de observación. El concepto fundamental a modelar va a ser entonces el denominado *hazard rate* o *tasa instantánea de salida* de los individuos. El análisis propuesto por Cox & Oakes se denomina *verosimilitud parcial*, ya que no se maximiza la probabilidad de ocurrencia de la muestra, sino una función que puede interpretarse como una verosimilitud relativa de la intensidad de salida de un grupo respecto del grupo que sobrevivió hasta el instante anterior. Dicha función a maximizar se denomina *función de verosimilitud parcial*.

1.1.3. Carteras y Riesgo Crediticio

Una vez sentadas las bases del análisis de supervivencia, la aplicación a la medición del riesgo crediticio se fue dando de a poco y con un cierto retraso. Un trabajo de relevamiento de técnicas aplicadas en el ámbito financiero publicado hacia fines de la década del '90 (Altman & Saunders, 1998), no da aún testimonio explícito de la aplicación de modelos de *hazard rate* o *verosimilitud parcial*. Sí se deja constancia del uso de las técnicas *probit* y *logit* conjuntamente con técnicas de transición de estados y otras denominadas “*derivación de probabilidades de default de tipo actuarial*” asociadas al default pasado de bonos.

Los desarrollos computacionales no tardaron en llegar, y diversos softwares fueron lanzados y ampliamente utilizados para el modelado de tiempos de supervivencia (SAS Institute Inc., 1999) (Fox, 2002) (Smith & Smith).

Iniciado el nuevo milenio se fueron vislumbrando desarrollos específicos para la aplicación del análisis de supervivencia a la medición del riesgo de crédito (Saunders & Allen, 2002) (Duffie & Schaefer, 2003) (Roszbach, 2003) (Bluhm, Overbeck, & Wagner, 2003). Todos estos desarrollos están basados en las premisas fundamentales del modelo de Cox & Oakes y fueron iniciadores de muchos trabajos de aplicación.

Los desarrollos posteriores en el campo del riesgo crediticio llevaron naturalmente a la necesidad de incorporar variables macroeconómicas. Es decir, una vez evaluado un cliente mediante un *scoring*, pueden cambiar las condiciones macroeconómicas que se tuvieron en cuenta en el momento de evaluación. El análisis de supervivencia permite incorporarlas como variables que afectan la estimación con posterioridad a la evaluación. Esto no es fácilmente realizable con los modelos de regresión logísticos (Bellotti & Crook, 2007). Es por esto que los Modelos de Supervivencia se convierten en un pilar fundamental para la explicación del papel de las influencias de los fenómenos macroeconómicos en el riesgo de crédito.

1.2. Problema de Investigación

La práctica más usual a la hora de realizar modelos estimativos de probabilidad de default (o PD, indistintamente) es por medio del modelaje de una variable binaria a través de modelos *logísticos*. Esto es, se observa el comportamiento de pago del individuo y se lo clasifica con una marca (o *flag*, indistintamente) de default que puede tomar valor 0, si no hay default, o 1 en caso de default. En general, dichos modelos son calibrados en un momento, y se realizan

seguimientos periódicos. No habrá diferencia en la PD estimada de dos individuos idénticos en el inicio de la implementación o años después. Sin embargo, durante un auge económico o un mejoramiento en las condiciones sistémicas es de esperarse un mejoramiento generalizado de la capacidad de pago de los individuos. En épocas de recesión o de empeoramiento de las condiciones del sistema financiero, se espera que esa capacidad se deteriore. Surge naturalmente la siguiente pregunta: *¿En qué medida el ciclo económico o las condiciones sistémicas contribuirían con modificaciones en las probabilidades de default estimadas?*

1.3. Modelos de Supervivencia

En este sentido cabe preguntarse si las estimaciones realizadas con la información disponible en el momento de la evaluación deberían ser afectadas o no por acontecimientos posteriores a su estimación. Y aquí es donde surgen los modelos que intentan calibrar no una probabilidad de default, sino una intensidad del mismo. En términos más familiares, se modela la fuerza con la cual los individuos con el correr del tiempo son arrastrados hacia una situación de insolvencia. Denominaremos a estos modelos, *Modelos de Supervivencia*.

Nos enmarcaremos entonces en la Teoría de Supervivencia aplicada a la Gestión del Riesgo de Crédito, construyendo modelos para la dinámica estocástica del tiempo hasta el default de los créditos (Roszbach, 2003). A diferencia de los modelos de regresiones logísticas, el análisis de supervivencia crediticio permite incorporar más fácilmente variables macroeconómicas que sufran modificaciones con el transcurrir del tiempo, teniendo impacto en la intensidad del default (Bellotti & Crook, 2007).

1.4. Objetivos del trabajo

Con la información disponible, se construye y calibra un Modelo de Supervivencia de créditos que combina armoniosamente y con sólida lógica económica variables tanto de carácter individual de los deudores como de carácter sistémico.

Dentro de las variables de carácter individual, se incorporan variables demográficas y de bureau al momento de evaluación, calibrando las mismas mediante modelos logísticos usuales y por medio de Modelos de Supervivencia, realizando un análisis comparado.

Se estudia también la influencia del contexto sistémico en el comportamiento de pago de los individuos. Para ello se sensibiliza el Modelo de Supervivencia incluyendo algún indicador sistémico.

De esta manera, se explotan los modelos propuestos realizando estimaciones sensibilizadas para distintos contextos sistémicos. Se incluyen además, mediante simulaciones, distribuciones de los beneficios económicos en cada modelo.

2. Descripción de variables disponibles

La unidad de análisis para este trabajo está constituida por préstamos de consumo de una Entidad Bancaria Argentina. Por motivos de confidencialidad, algunos datos que se exponen se encuentran previamente multiplicados por un escalar.

2.1. Armado de Base de Datos

Se consideró la exposición real de los clientes siguiendo su comportamiento desde la fecha de originación de los productos hasta la última fecha disponible con datos. La situación de default se define como aquella en la cual el cliente registra más de 90 días de mora en cualquiera de sus productos, realizando las observaciones a final de cada mes. El cliente deja de observarse luego del primer mes donde registra situación de default.¹ En base a esto se construyó la variable dicotómica indicadora del default, y la variable entera que indica el mes en el que se produce el mismo desde el momento de la originación. Para el caso de los clientes que dejan de observarse sin haber entrado en default (ya sea porque llegaron hasta la última fecha de la ventana temporal de observación, o por haber dejado de ser clientes de la Entidad) se construye una variable dicotómica que representa la *censura* de los mismos. Es decir, que los clientes que fueron observados desde la fecha de su primer producto hasta el mes de default se consideran *clientes no censurados*. En caso que hasta la última fecha de observación disponible no registren default, se consideran *clientes censurados*.

La ventana temporal incluyó los clientes que tuvieron su primer producto de crédito entre enero 2010 y enero 2013. La última fecha de observación disponible fue enero 2014.

En el caso de Tarjetas de Crédito, se consideró producto de crédito sólo si se detectó un nivel de uso igual o superior a \$100 de saldo en los primeros 6 meses desde la activación. Luego se realizó el seguimiento mes a mes para detectar defaults.

La base final para el modelado constó de **35.598** clientes. Se reservó una muestra aleatoria del 20%, es decir **7.120** registros, como **Base de Validación** del Modelo. La **Base de Desarrollo** constó entonces de **28.478** registros.

La Tabla 1 muestra un cuadro descriptivo de las bases. Cabe aclarar nuevamente que, de aquí en adelante, los valores de defaults y otros reservados han sido multiplicados por una constante por motivos de confidencialidad de los datos.

En la Tabla 1 se observa que existe un desarrollo significativo de los defaults luego del mes 12. De los 2.955 defaults totales, 1.425 ocurren con posterioridad a los 12 meses (aproximadamente un 48% de los defaults totales), con lo cual a priori resultaría insuficiente estimar solamente la probabilidad de default a un año. En este trabajo se analizan ambos tipos de defaults (dentro del primer año, y con posterioridad) para realizar comparaciones.

¹ No se analiza el comportamiento del cliente luego de incurrir en default, ya que está dentro del estudio del *Loss Given Default* o LGD y no es objetivo de este trabajo. Es decir, puede que el cliente luego de incurrir en 90 días de mora pague toda su deuda poniéndose al día o bien que se recupere una porción de la deuda. El análisis de los recuperos con posterioridad al default no está dentro de los objetivos de este trabajo.

Concepto	Desarrollo	Validación	Total
Default a 12 meses (Casos)	1.227	303	1.530
Default a 12 meses (Tasa)	4,31%	4,25%	4,30%
Default luego de 12 meses (Casos)	1.140	285	1.425
Default luego de 12 meses (Tasa)	4,00%	4,00%	4,00%
Defaults totales (Casos)	2.367	588	2.955
Defaults totales (tasa)	8,31%	8,26%	8,30%
Casos Totales	28.478	7.120	35.598
% Casos Totales	80,00%	20,00%	100,00%

Tabla 1: Descripción de Base de Datos (Elaboración propia)

2.2. Variables Disponibles - Individuales

Respecto de las variables disponibles al momento de la admisión del cliente, es importante destacar que se trata de un segmento de Cartera General. Esto es, se trata de personas que no son clientes preexistentes de la Entidad y acuden a la misma por primera vez a solicitar un crédito. La Entidad no posee información previa del cliente ni tampoco posee algún mecanismo de cobro automático (como podría ser en el caso de que el individuo acredite salarios en la Entidad). Entonces, la voluntad y conducta de pago del cliente es muy relevante para definir si el mismo incurre o no en default. Teniendo en cuenta lo expuesto, las principales variables individuales disponibles son:

- **Edad**
- **Estado Civil**
- **Provincia de la Sucursal**
- **Flag de Tipo de Campaña:** indica si el cliente fue captado por una campaña masiva de convenios comerciales o no.
- **Score de Bureau:** puntaje del 0 al 1000 otorgado por bureau crediticio.
- **Nivel Socioeconómico:** Información proveniente de bureau. Se encuentra relacionado a un nivel de ingresos.
- **Tenencia de Casa y/o Vehículo:** variable declarativa que indica la tenencia de bienes.
- **Nivel de Estudios:** variable declarativa que indica el máximo nivel de estudios alcanzado (Primario, Secundario, Terciario o Universitario).
- **País de procedencia**
- **Sexo**
- **Cantidad de personas a cargo**

Cabe mencionar que por temas vinculados a la disponibilidad, veracidad y factibilidad de obtención de la información, se decidió no incluir las siguientes tipologías de variables:

- Aquéllas meramente declarativas y, por lo tanto, manipulables al momento de la originación.
- Las variables que no están 100% disponibles al momento de la originación del crédito.
- Otras cuya inclusión puedan dar cuenta de algún tipo de discriminación.

Por eso, se excluyeron las cinco últimas variables anteriormente expuestas.

2.3. Variables Disponibles - Sistémicas

Las principales variables macroeconómicas disponibles para la explicación del default son:

- **PBI real:** medido por la Consultora OJF (Orlando J. Ferreres & Asociados , 2014).
- **Nivel de Desempleo:** publicado por el Instituto Nacional de Estadísticas y Censos (INDEC).
- **Nivel de Mora Temprana en el Sistema Financiero:** publicación mensual del Banco Central de la República Argentina (o BCRA indistintamente), representando el porcentaje de préstamos informados en situación 2 en adelante (usualmente 30 días de mora o más en caso de Personas Físicas) sobre el total de Préstamos del Sistema Financiero.
- **Tasas de interés Activas:** Publicadas por el BCRA.

Se decidió incluir en el análisis únicamente el Nivel de Mora Temprana en el Sistema Financiero. Esta decisión se basa en tres pilares fundamentalmente.

- El primero, que el objetivo de este trabajo es encontrar **un indicador** sistémico simple para hacerlo influir en la capacidad de pago de los clientes.
- Además de ser simple, es muy **intuitivo** ya que ante un aumento en la mora temprana (30 días de atraso) en el sistema, se esperaría que en dos meses más pueda existir impacto en la mora a 90 días.
- Por último, se entiende que es práctica usual de las Entidades Financieras realizar estimaciones del nivel de mora en el sistema financiero en función de escenarios macroeconómicos. Por lo tanto, el escenario macroeconómico queda embebido en las proyecciones de la salud del sistema financiero realizadas por las Entidades. Se construye un modelo en función de algo simple y que cualquier Entidad podría estimar en función de lo que considere relevante.

Para la **recolección y tratamiento de los datos**, se utilizó el motor de bases de datos y cálculo SAS® Enterprise Guide.

3. Análisis Bivariados

Los análisis bivariados permiten detectar relaciones entre una variable explicativa (por ejemplo, el nivel de ingresos) y la variable explicada (por ejemplo, el default de un cliente). Son muy útiles para probar si existe la relación esperada entre la variable explicativa y la tasa de default. Permiten elegir de todas las variables explicativas disponibles, aquéllas que pueden ser candidatas para incluirse en un modelo.

Es preciso mencionar que para los modelos que se van a calibrar, las variables explicadas son diferentes. Mientras que para el Modelo Logístico, la variable explicada es la probabilidad de default (medida a través de la tasa de default), el Modelo de Supervivencia considera el tiempo que transcurre hasta dicha situación de insolvencia.

A continuación se muestran dos análisis bivariados que intentan explicar la tasa de default en los primeros 12 meses desde la originación (o corto plazo, indistintamente) y la tasa de default luego de los 12 meses (o largo plazo, indistintamente). Los análisis se realizaron con todas las variables disponibles y las decisiones de segmentación e inclusión se tomaron siguiendo los mismos criterios.

Respecto de las variables sistémicas, no se presentan en esta parte del trabajo porque las mismas aplican solamente en el Modelo de Supervivencia.

Para decidir si incluir una variable o no en el modelo, los análisis bivariados permiten mostrar si dicha variable:

- Puede ordenar a los clientes en dos o más categorías diferenciadas de riesgo.
- Cumple con el comportamiento esperado respecto a la tasa de default de cada categorización, es decir, guarda coherencia y lógica económica.

3.1. Flag de Campaña

La variable Flag de Campaña indica si el nuevo cliente se captó por campaña masiva o no. En el Gráfico 1 se observa que tanto la tasa de default a 12 meses como la de largo plazo no tienen diferencias significativas entre ambos grupos. Además, no existe gran cantidad de casos pertenecientes a campañas. Por estos motivos se decide no incluir la variable en el modelo final.

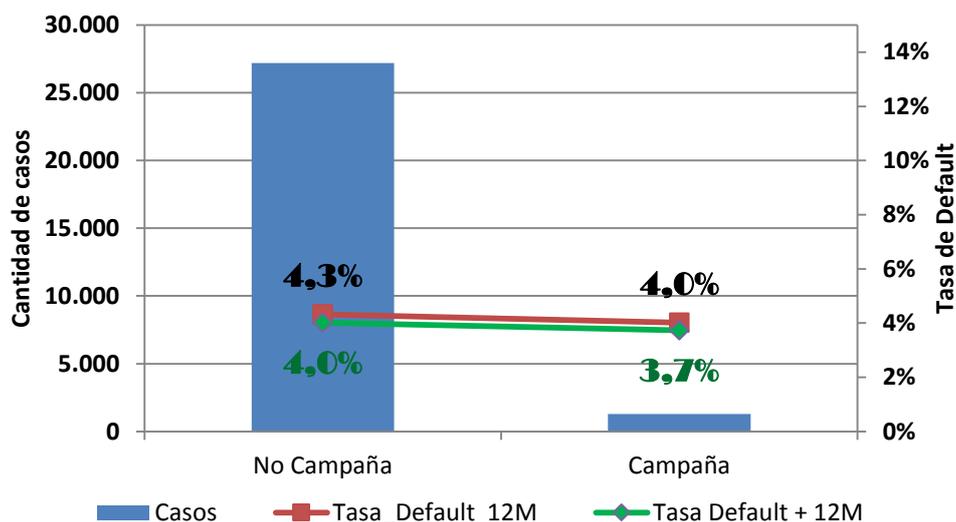


Gráfico 1: Análisis Bivariado - Flag de Campaña (Elaboración propia)

3.2. Score de Bureau

Dado que el Score de Bureau se presenta como un puntaje de 0 a 1000, resulta necesario realizar agrupaciones para determinar si existe una relación con el default. Las categorías se determinaron segmentando la variable en intervalos. Por motivos de confidencialidad, no se muestran los intervalos sino las categorías ordenadas en forma creciente (A es la categoría con puntajes más bajos, H corresponde a los puntajes mayores). En el Gráfico 2 se muestra la categorización realizada de dicha variable y las tasas de default.

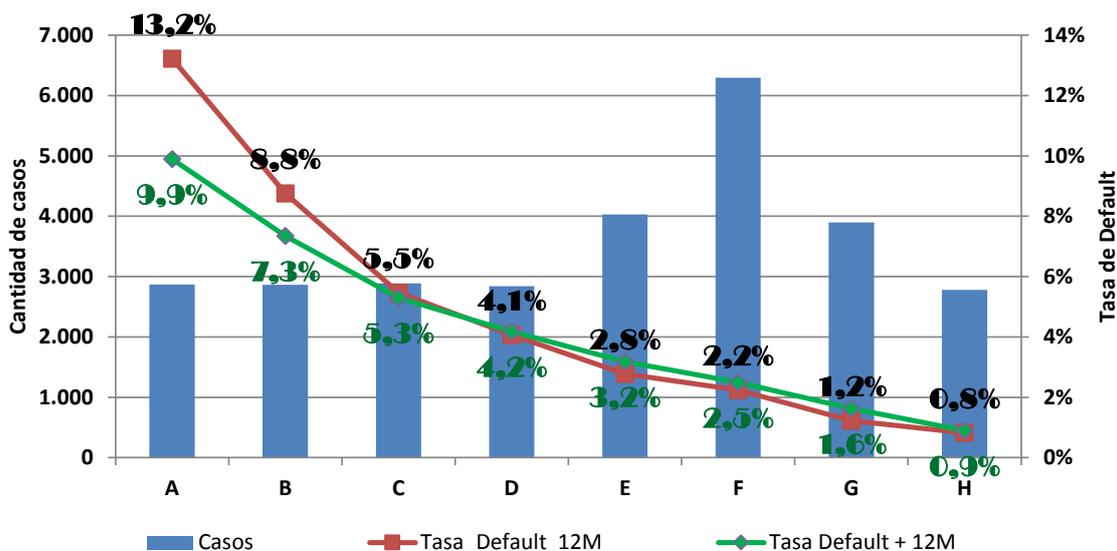


Gráfico 2: Análisis Bivariado - Score de Bureau (Elaboración propia)

Se puede observar que el tramado de la variable asigna tasas de default diferenciadas para todas las categorías, y que tiene sentido que a mayor puntaje de Bureau exista menor tasa de default tanto a corto como a largo plazo. Además, se observa también que en el caso de altos

puntajes, las tasas de default a corto y largo plazo se parecen entre sí, mientras que para puntajes bajos, la tasa de default a 12 meses es más alta que una vez transcurrido dicho plazo. Esto significa que los clientes de bajo puntaje que sobreviven a los 12 meses son notoriamente menos riesgosos que los que no sobrevivieron. Para puntajes bajos se desarrolla el 43% de los defaults luego de los 12 meses (Para la categoría A la tasa de Default total es 23,1% y la tasa de default luego de 12 meses 9,9%), mientras que para puntajes altos este porcentaje crece hasta un 53% (Para la categoría H la tasa de Default total es 1,7% y la tasa de default luego de 12 meses 0,9%). Incluso ocurre para las categorías de puntajes altos que la tasa de default de largo plazo supera a la tasa de default con anterioridad a los 12 meses.

Debido a que el Score de Bureau ordena a los clientes en categorías diferenciadas de riesgo, con cantidad razonable de observaciones en cada rango y cumpliendo el comportamiento esperado, se decide incluir esta variable en los análisis multivariados.

Los análisis bivariados se realizaron para las demás variables descritas en la Sección 2 para decidir si incluirlas o no en los análisis multivariados. Para no extender la exposición, los resultados de los demás análisis bivariados no se presentan aquí pero están disponibles mediante solicitud al autor.

4. Análisis Multivariados - Modelo Logístico

Los análisis multivariados permiten calibrar modelos guardando coherencia y lógica económica respecto del comportamiento esperado con cada variable. Además, se busca que las variables incluidas posean significatividad estadística, es decir, que la influencia sobre la tasa de default sea no nula con alta probabilidad.

En el caso de los Modelos Logísticos, la variable objetivo será la *probabilidad de default*, y en el caso de los Modelos de Supervivencia será la distribución del *time to default*.

4.1. Definición del Modelo de Regresión Logística

Para estimar la probabilidad de default se parte de un vector \mathbf{X} cuyos k componentes son las variables explicativas. Cada individuo tiene una realización particular de este vector según su rango de Score de Bureau, edad, etc. Existirá además un vector $\boldsymbol{\beta}$ de parámetros que afectarán a cada una de las variables explicativas, incluyendo un término de intercepto. Se dispondrá también de \mathbf{N} realizaciones de la variable a explicar \mathbf{Y} , siendo ésta el flag de default de cada uno de los \mathbf{N} individuos, que toma el valor uno en caso de default y cero en caso contrario.

4.1.1. Probabilidad de Default

Las probabilidades de default de cada individuo con variables explicativas representadas en el vector \mathbf{X} serán modeladas a partir de la siguiente Regresión Logística, donde el superíndice \mathbf{T} indica trasposición del vector y el símbolo $*$ indica multiplicación de vectores.

$$PD_{(X;\beta)} = \frac{1}{1 + e^{-X^T * \beta}}$$

Ecuación 1: Modelo de regresión logística

Dado que la determinación de la PD depende exclusivamente del vector de parámetros $\boldsymbol{\beta}$, una vez definido el valor observado del vector \mathbf{X} en cada individuo, se dice que el Modelo Logístico es un **modelo paramétrico**.

4.1.2. Score y ratio odds

El *score directo* está dado por la multiplicación vectorial que aparece en el denominador de la Ecuación 1, es decir:

$$\begin{aligned} \text{Score Directo} &= -X^T * \beta \\ &= -(\beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \dots + \beta_k \times X_k) \end{aligned}$$

Ecuación 2: Score directo

Despejando de la Ecuación 1, se tiene que:

$$\ln\left(\frac{1 - PD_{(X;\beta)}}{PD_{(X;\beta)}}\right) = -X^T * \beta$$

El ratio que aparece en el logaritmo natural del miembro izquierdo es conocido como ratio odds (o bien ODDS, indistintamente), y representa la probabilidad de no default sobre la probabilidad de default, es decir, cuántos clientes buenos (o no defaults) se presentan por

cada cliente malo (o que incurre en default). La ecuación anterior se puede reescribir entonces como:

$$\ln(ODDS) = \text{Score Directo}$$

Entonces, el score directo de la Ecuación 2 representa el modelado del logaritmo natural del ratio odds. Existe una relación directa el score directo y el ratio odds, e inversa entre el score directo y la probabilidad de default.

Asimismo, la interpretación de un coeficiente β_j es sencilla: indica el cambio en el logaritmo del ratio odds ante un cambio en la variable explicativa X_j , y tiene relación inversa con la probabilidad de default.

4.1.3. Funciones de Verosimilitud

Para realizar la estimación del vector de coeficientes β , es necesario plantear funciones de verosimilitud individuales, que dependen del default observado para cada individuo. Luego se plantea una función de verosimilitud total a maximizar.

La función de verosimilitud individual está dada por:

$$L_i(X_i; \beta) = PD_{(X_i; \beta)}^{y_i} \times (1 - PD_{(X_i; \beta)})^{1-y_i}$$

Ecuación 3: Función de Verosimilitud Individual

donde la variable y_i indica la ocurrencia de default cuando toma el valor 1 y la no ocurrencia de default si es 0. Esta variable y_i se denomina variable explicada.

Para ajustar los coeficientes β a los defaults observados y_i , y a los valores observados de las variables explicativas X_i , se realiza el producto de todas las verosimilitudes y luego se aplica el logaritmo para un mejor manejo. Siendo N el tamaño de la base de desarrollo del modelo, la función de verosimilitud total a maximizar es:

$$L(\beta) = \prod_{i=1}^N L_i(X_i; \beta)$$

$$\ln L(\beta) = \sum_{i=1}^N \ln L_i(X_i; \beta)$$

Ecuación 4: Función de Verosimilitud Total

4.2. Resultados de la calibración

Mediante la comparación de diversos modelos se seleccionó el siguiente:

$$\ln(ODDS) = -(\beta_0 + \beta_1 \times \text{ScoreBureau} + \beta_2 \times \text{Edad} + \beta_3 \times \text{EstCivil} + \beta_4 \times \text{Provincia})$$

Ecuación 5: Modelado del score directo

Los resultados de la estimación de los parámetros del modelo se muestran en la Tabla 2. Para consultar mayores detalles del proceso de estimación y de las propiedades de los estimadores, se puede consultar en el trabajo de Gourieroux & Jasiak, 2007.

Variable	Categoría	Estimador	Pr > ChiSq
Intercepto		-4,4837	<,0001
Score_Bureau	A	2,9083	<,0001
	B	2,3720	<,0001
	C	1,8425	<,0001
	D	1,5362	<,0001
	E	1,1473	<,0001
	F	0,9565	<,0001
	G	0,3741	0,0457
	H	0,0000	
Edad	A	0,3741	<,0001
	B	0,2882	0,0004
	C	0,2408	0,0015
	D	0,0989	0,2424
	E	0,0000	
Estado Civil	Resto	0,2906	<,0001
	Casado	0,0000	
Provincia	Resto	0,1680	0,003
	BsAs-Cap	0,0000	

Tabla 2: Parámetros del Modelo Logístico (Elaboración propia)

La columna “Estimador” representa los valores estimados de los coeficientes de β . Particularmente, la fila de intercepto representa la estimación de β_0 , que es la ordenada al origen del *score directo*. La columna “Pr > ChiSq” representa el p valor en el test individual de Wald.² Esto último es lo que se denomina *significatividad estadística* de cada estimación de β . Representa un proxy de la probabilidad de que el coeficiente sea nulo.

Se puede observar que todas las categorías, a excepción de la categoría D de Edad, tienen significatividad estadística menor al 5% (es decir, tienen probabilidad de coeficiente nulo muy baja). Cabe mencionar que las categorías de la variable Edad son crecientes, es decir, A es el menor rango de Edad y E es el intervalo de edades mayores. Además, los coeficientes son crecientes en función del comportamiento esperado, es decir, se cumple la lógica económica.

Una mención particular cabe hacerle a la decisión de no excluir la categoría de edad D. Dado que la significatividad estadística está medida por la probabilidad de que esa variable tenga coeficiente nulo, se espera que dicha probabilidad aumente a medida que el coeficiente estimado sea cercano a cero. No se excluyó el estimador de dicha variable ya que se considera aceptable que el coeficiente sea positivo, mayor que el de la categoría inmediata inferior y mayor que el de la inmediata superior. Es decir, la variable Edad sigue guardando lógica económica aunque una de sus categorías (la de coeficiente más cercano a cero), no tenga significatividad estadística.

4.3. Ejemplos de aplicación

Una vez estimados los coeficientes, se pueden obtener las estimaciones de la probabilidad de default de cada individuo. Por ejemplo, si un individuo tiene categoría de Score F, categoría de Edad B, es Soltero y reside en Capital Federal, el *score directo* (Ecuación 2) toma la forma de la Ecuación 6 (se omiten los términos de categorías nulas y se redondea al cuarto decimal).

² (SAS Institute Inc., 1999, pág. 1908)

$$-X^T * \beta = -(-4,4837 + 0,9565 + 0,2882 + 0,2906 + 0) = 2,9484$$

Ecuación 6: Score directo para un caso particular. Ver Ecuación 2.

Su probabilidad de default, en función de la Ecuación 1, es:

$$PD_{(X;\beta)} = \frac{1}{1 + e^{2,9484}} = 4,98\%$$

Ecuación 7: Probabilidad de default para un caso particular.

Cabe mencionar que de todos los individuos que cumplen con esta caracterización, se encontraron 783 casos, de los cuales 40 incurrieron en default. Por lo tanto, la tasa de default real resultante es de 5,11%, la cual es muy próxima a la estimada con el Modelo Logístico.

4.4. Medidas usuales de performance

Para evaluar si el modelo funciona correctamente, es necesario establecer uno o varios criterios, siempre y cuando no se pierda de vista el objetivo real que es establecer una relación cuantitativa lógica entre una serie de variables explicativas y el comportamiento a explicar. Para el caso particular del Modelo Logístico, se estima una función que transforma las variables explicativas en un número simple (el score directo, la PD, o un score mapeado que usualmente toma valores del 0 al 1000).

Es deseable que la función estimada tenga una relación con el grado de riesgo del individuo. Para evaluar si esta función es razonablemente un buen indicador del riesgo, se busca mostrar básicamente dos cuestiones.

1. Que el indicador logre **ordenar** a la población de menor a mayor riesgo.
2. Obtener una medida de probabilidad de default cercana a la tasa real de incumplimientos, es decir, que el modelo logre **predecir** qué porcentaje de los individuos de cada grupo entrará en insolvencia.

Esto es respecto de evaluar cada modelo particularmente. Luego, será de gran utilidad generar medidas o índices para comparar distintos modelos en ambas cuestiones.

Es importante destacar que todos los análisis deben realizarse no solamente en la base de desarrollo del modelo, sino también en la base de validación del mismo. Esto es importante porque en general los parámetros que se estiman funcionarán bien con la base de desarrollo ya que fueron ajustados a esos datos particulares. Entonces es necesario probar el funcionamiento en una muestra independiente de los datos con los cuales se ajustaron los parámetros.

4.4.1. Comparación con tasas de default reales

Como se mencionó anteriormente, la función ordena a la población en categorías de riesgo. Es conveniente en primer lugar realizar algún tipo de agrupación para generar categorías de riesgo similares. Usualmente se separa en deciles o bien alguna otra medida de uniformidad entre los grupos. La comparación se realiza en la base de desarrollo y en la base de validación

(quizá convenientemente reduciendo la cantidad de grupos en la última, ya que siempre tiene mucho menos observaciones que la primera).

4.4.1.1. Base de Desarrollo

El Gráfico 3 es un análisis bivariado del propio score directo en la Base de Desarrollo, para ver si existe relación de éste con las tasas de default reales, es decir para ver si el modelo **ordena** correctamente a la población. Se dividió a la base de desarrollo en 15 grupos de mayor a menor score directo. En el Gráfico 3 se muestra la relación entre el score directo asignado y las tasas de default reales (valores en negro) y las Probabilidades de Default estimadas con el modelo (valores en verde).

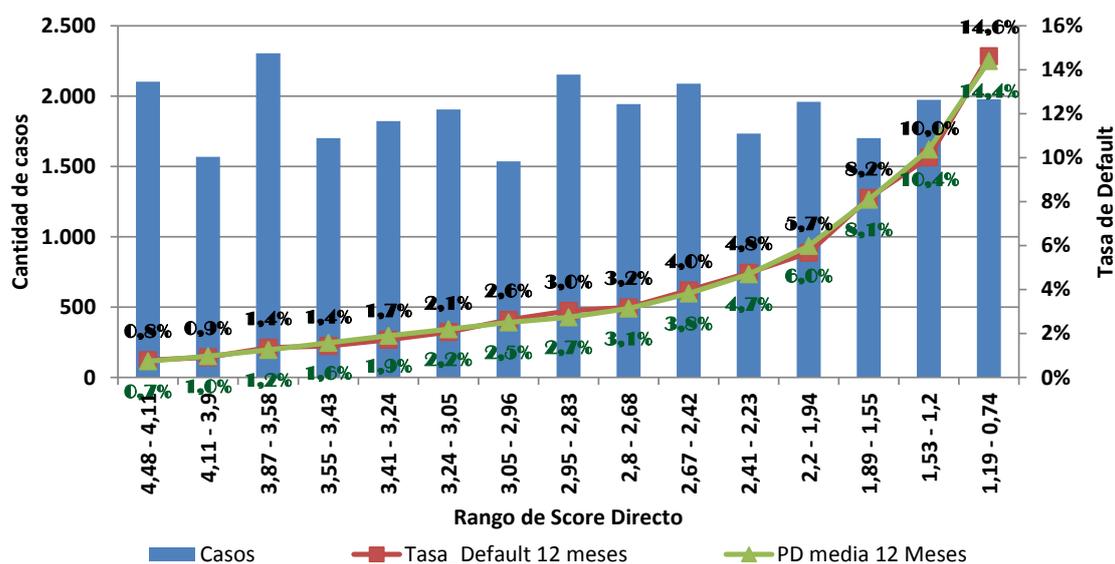


Gráfico 3: Análisis de ordenamiento y Predicción del Modelo Logístico (Elaboración Propia)

Se puede observar que fehacientemente el score directo **ordena** a los grupos en categorías de riesgo diferenciadas. Asimismo el modelo también **predice** la tasa de default real con bastante exactitud mediante la PD estimada.

El Gráfico 4 muestra la tasa de default a largo plazo, es decir, la tasa de default que contempla los clientes que se vuelven insolventes luego de los 12 meses de iniciado el crédito.

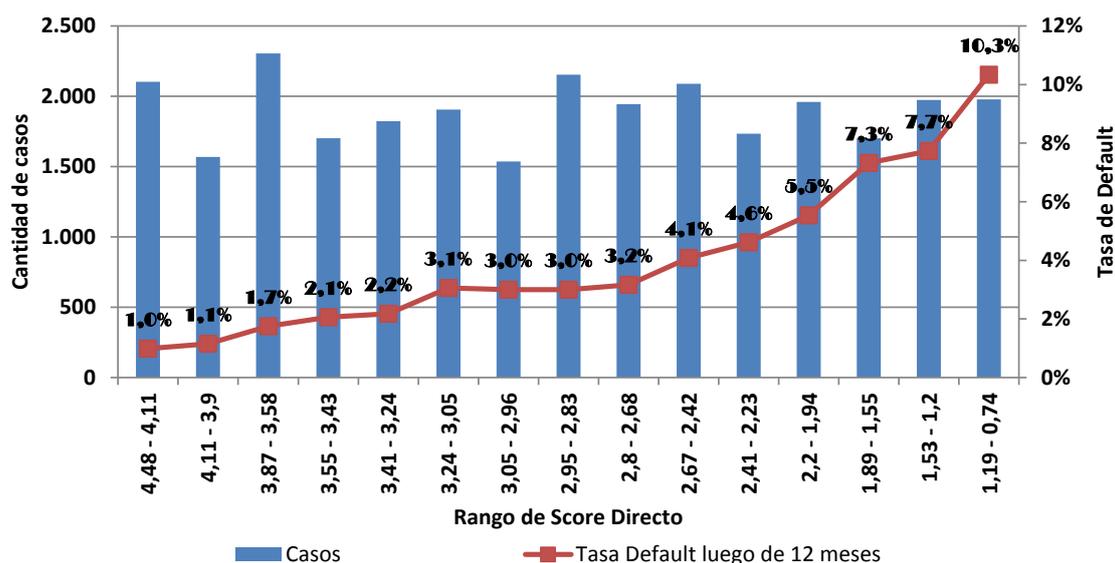


Gráfico 4: Análisis de ordenamiento y Predicción del modelo Logístico en Largo Plazo (Elaboración Propia)

Es destacable el hecho de que el modelo sigue sirviendo para explicar un cierto ordenamiento del comportamiento de los clientes a largo plazo, aunque para algunas categorías no se cumple crecimiento estricto de la tasa de mora a medida que disminuye el score directo asignado. Además, dado que el Modelo Logístico busca estimar solamente la PD a 12 meses, los clientes de mejor score directo (a la izquierda) tienen en general una tasa de default de corto plazo menor que la de largo plazo. Los clientes de peor score directo asignado por el modelo, tienen una tasa de default a largo plazo inferior a la de corto plazo, dando cuenta de que la permanencia por 12 meses en la cartera sin default no es tenida en cuenta por el Modelo Logístico.

En este gráfico, como en todos los que se muestre tasa de default a largo plazo, no se incluye una estimación de la probabilidad de default estimada ya que el Modelo Logístico estima solamente una PD a 12 meses.

4.4.1.2. Base de Validación

En la base de validación, para realizar los análisis será conveniente trabajar con menos grupos, ya que se trata de una base de una cantidad significativamente menor de registros. Se necesita un cierto volumen de casos por grupo para confiar estadísticamente en las conclusiones. Por eso se dividió a la base de validación en 8 grupos de menor a mayor riesgo.

Cabe destacar que es muy importante realizar los análisis también sobre la base de validación, ya que se trata de una base independiente de la cual se desarrolló el modelo, y siempre se espera que la calibración sirva para evaluar clientes nuevos.

En el Gráfico 5 se puede observar que fehacientemente el score directo **ordena** a los grupos en categorías de riesgo diferenciadas también en la base de validación. Además, se puede observar que la PD del modelo **predice** con bastante exactitud la tasa de default observada.

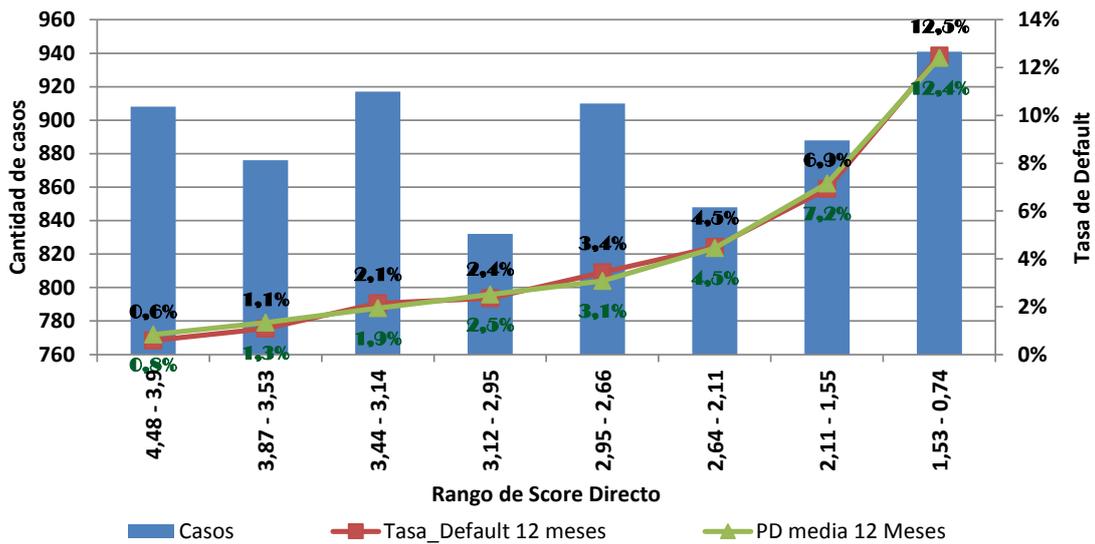


Gráfico 5: Análisis de ordenamiento y predicción del Modelo Logístico – Validación (Elaboración Propia)

El Gráfico 6 analiza la tasa de default a largo plazo, es decir, la tasa de default que contempla los clientes que se vuelven insolventes luego de los 12 meses de iniciado el crédito.

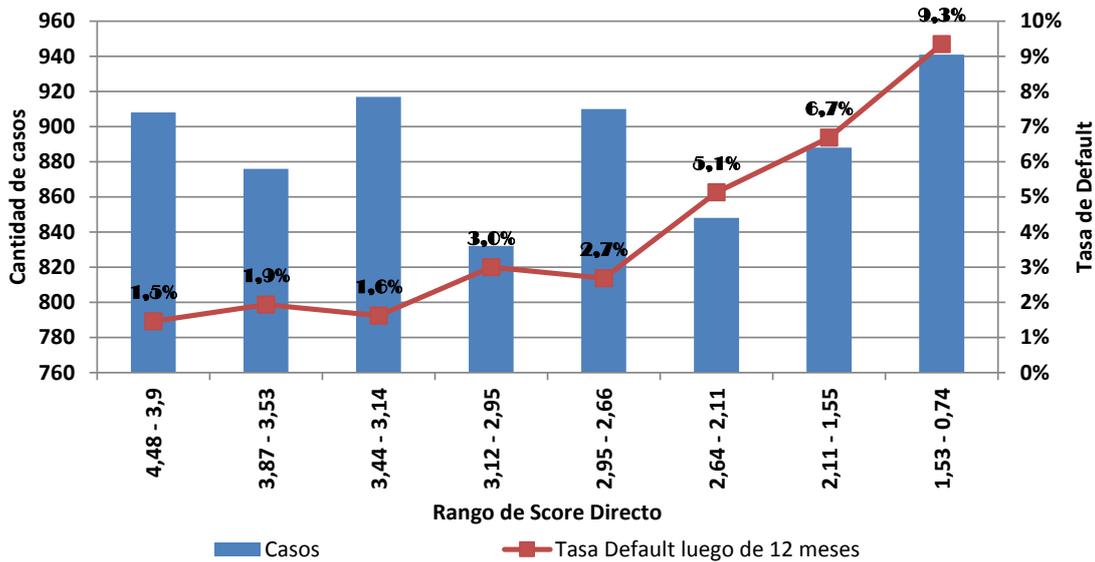


Gráfico 6: Análisis de orden y predicción del Modelo Logístico en Largo Plazo - Validación (Elaboración Propia)

Es destacable el hecho de que el modelo sigue sirviendo para explicar un cierto ordenamiento del comportamiento de los clientes a largo plazo, aunque para algunas categorías no se cumple crecimiento estricto de la tasa de mora a medida que disminuye el score directo asignado. Además, la performance en una base independiente como lo es la base de Validación se vuelve más pobre en este sentido.

4.4.2. Área ROC y Power Stat

Para evaluar la performance del modelo, además de entender gráficamente si ordena y predice los incumplimientos, es necesario generar indicadores que midan el funcionamiento general del mismo y que permitan compararlo con otros modelos.

Una de las maneras más usuales de analizar la performance del modelo es mediante el Power Stat o índice de poder asignado por el modelo, construyendo básicamente 3 curvas (Englemann, Hayden, & Tasche, 2003):

- **Curva del modelo:** Curva de frecuencia acumulada de las calificaciones del modelo a los registros que presentaron default. Se espera que esta curva crezca con mucha pendiente al inicio (a medida que la calidad crediticia aumenta), es decir, que el modelo prediga muchos defaults en bajas calidades crediticias.
- **Curva perfecta:** Curva de frecuencia acumulada de las calificaciones dadas a los registros que presentaron default por un “modelo perfecto”, es decir, aquel que asignaría a las “operaciones malas” las peores calificaciones.
- **Curva aleatoria:** Curva de frecuencia acumulada de las calificaciones dadas a los registros morosos por el “peor modelo posible”, es decir, el que carece de discriminación alguna entre operaciones que presentaron default y aquéllas que no lo presentaron.

El Gráfico 7 muestra las curvas para el Modelo Logístico sobre la base de validación, considerando el default a 12 meses.

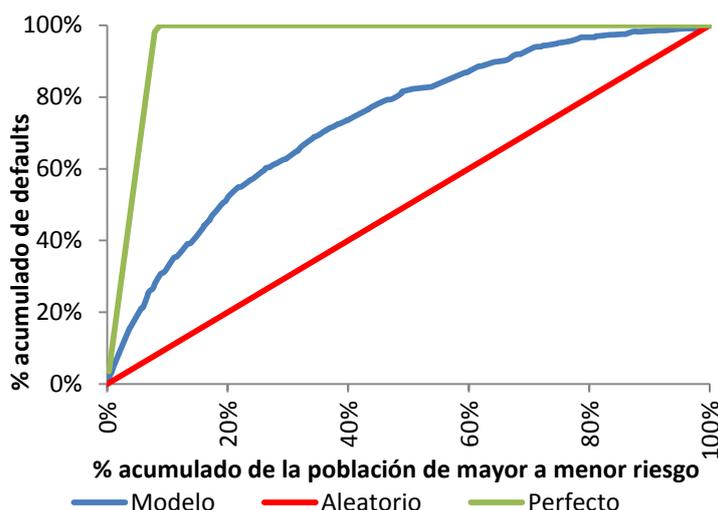


Gráfico 7: Curva del Modelo – Defaults a corto plazo – Validación (Elaboración Propia)

Se espera que en un buen modelo su curva se asemeje a la curva perfecta. Por eso se definen los siguientes indicadores:

- **Área CAP³:** es el área bajo la curva del modelo. Cuanto más área, mejor predicción. En el caso que el modelo sea perfecto, será igual al área debajo de la curva perfecta. No

³ Siglas de “Cumulative Accuracy Profile”

existe un parámetro estándar para decidir el uso de cierto modelo en función de un umbral mínimo de área CAP.

- **Power Stat o Accuracy Ratio:** es el cociente del área comprendida entre la curva del modelo y la curva aleatoria, dividida por el área comprendida entre la curva perfecta y la curva aleatoria. Cuanto más alto el cociente, con un valor máximo de 1, más semejanza entre el modelo calibrado y el modelo perfecto. El mismo es utilizado para comparar modelos, y no tiene un estándar de corte como umbral mínimo.

Muchos autores hacen referencia al área ROC⁴ como medición de la capacidad predictiva de un modelo. El desarrollo del mismo no se presentará aquí, pero cabe señalar que existe una relación directa con el Power Stat (Englemann, Hayden, & Tasche, 2003, pág. 12).

Este análisis puede realizarse tanto para los defaults en el primer año (como en el Gráfico 7), como para los defaults después de los 12 meses y también para los defaults totales. La Tabla 3 muestra los indicadores definidos anteriormente.

	Corto Plazo	Largo Plazo	Total
Área ROC	75,1%	69,6%	74,3%
Power Stat	50,2%	39,1%	48,6%

Tabla 3: Medidas de performance Modelo Logístico (Elaboración propia)

Se puede ver claramente cómo la potencia máxima del modelo medida en términos de estos indicadores se concentra en los defaults de corto plazo, mientras que para los de largo plazo existen medidas de performance un poco menores.

4.5. Valuación económica de la performance

Hasta el momento se realizó un análisis meramente estadístico respecto de la exactitud del ordenamiento y predictibilidad del modelo. Al momento de evaluar a un cliente, se debe establecer un **punto de corte**, que indica el nivel mínimo de calidad crediticia (o nivel máximo de riesgo) con el cual la Entidad decide aceptar a un cliente según su apetito a riesgo.

La ventaja de los modelos de scoring en general es que la calidad crediticia o nivel de riesgo se puede medir con alguno de los outputs. En particular, el Modelo Logístico arroja el score directo como nivel de calidad crediticia o bien la Probabilidad de Default como nivel de riesgo.

Cuando se define un punto de corte para el score directo (o para la PD estimada), el modelo divide a los clientes entre aquéllos que pueden ser incluidos dentro de la cartera crediticia según el apetito a riesgo de la Entidad, y aquéllos que no serían aceptados por la Entidad. Para aquéllos que cumplen las condiciones de aceptación, el modelo ayuda a que se maximice la cantidad de buenos clientes y que se minimice la cantidad de malos clientes. Para aquéllos que no cumplen las condiciones y por lo tanto serían rechazados, el modelo ayuda a que se minimice la cantidad de buenos clientes perdidos y que se maximice la cantidad de malos clientes bien detectados.

En otras palabras, al definir un punto de corte, el modelo tendrá:

⁴ Siglas de "Receiving Operating Characteristic"

- **Aciertos:** en los clientes buenos que fueron aceptados y los clientes malos que fueron rechazados.
- **Desaciertos:** en los clientes malos que fueron aceptados y los clientes buenos que fueron rechazados.

El Gráfico 8 muestra un ejemplo ilustrativo de las decisiones acertadas de los modelos según el nivel de exigencia. En el mismo se ilustra la decisión acertada de aceptar clientes buenos (como porcentaje de los clientes buenos totales, línea roja), y la otra decisión correcta que es rechazar clientes malos (como porcentaje de los clientes que incurrieron en default, línea azul). Asimismo se grafican los incrementos marginales como triángulos. La comparación de los últimos triángulos azules de mayor área que los rojos significa que si se relajan todas las exigencias, se pierde mucha más potencia en detectar clientes malos que en aceptar clientes buenos.

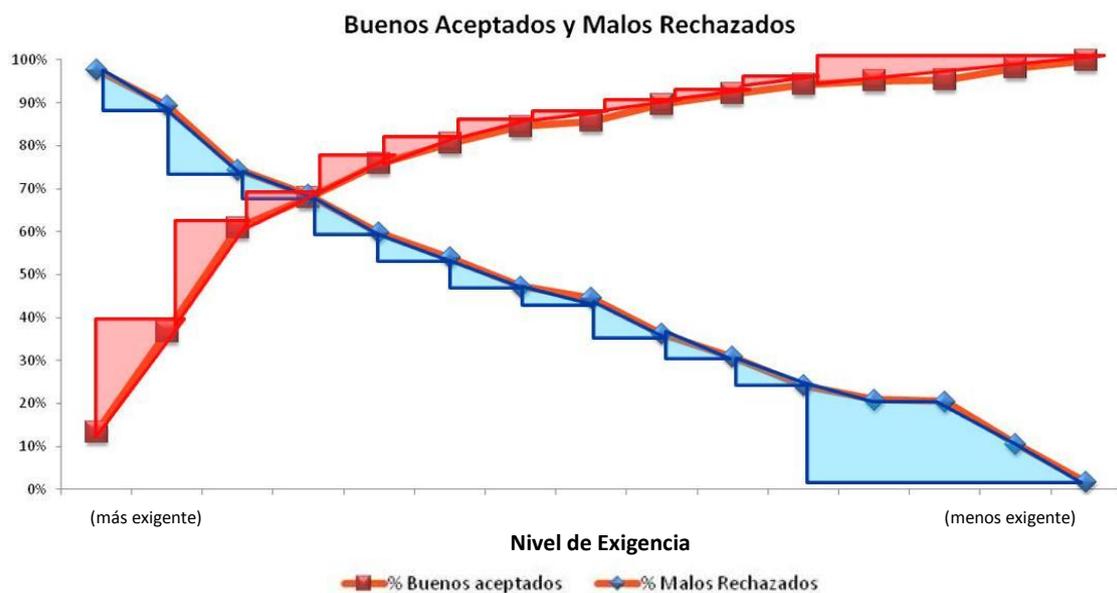


Gráfico 8: Decisiones acertadas en modelos de Scoring – Ejemplo (Elaboración Propia)

De aquí se desprenden conclusiones que pueden ser aplicables a cualquier modelo, a saber:

- En general, con altos niveles de exigencia, se tiene mucho beneficio por los clientes finalmente malos que el modelo rechaza, ya que es capaz de detectar a muchos de ellos. Pero el beneficio por buenos clientes aceptados es muy bajo ya que un bajo porcentaje de la población cumple con los niveles de exigencia.
- A medida que vamos siendo menos exigentes, damos la oportunidad para que entren más clientes buenos, pero al mismo tiempo perdemos potencia en detectar clientes malos. Es por eso que hay un *trade off* aquí entre los dos objetivos, es decir, detectar gran cantidad de clientes malos y al mismo tiempo aceptar gran cantidad de buenos.

Asimismo, los aciertos y desaciertos se pueden ubicar en dos grupos:

- **De la cartera aceptada**: son los clientes que se terminan aceptando, y el modelo acierta en los buenos y desacierta en los malos.
- **De la cartera rechazada**: son los clientes que se terminan rechazando, y el modelo acierta en los malos y desacierta en los buenos.

La clave para determinar un buen punto de corte consistirá entonces en evaluar adecuadamente los siguientes objetivos, teniendo en cuenta que existirá un *trade off* entre ellos:

- **La tasa de default final**: de la cartera aceptada, cuántos clientes terminan incurriendo en default. A medida que somos menos exigentes, la tasa de default será creciente.
- **El market share**: cuánto representa la cartera aceptada sobre el total de la cartera. A medida que somos menos exigentes, la tasa de aceptación será creciente.
- **El nivel de beneficios**: indica cuál es la valoración económica entre los aciertos y desaciertos del modelo, es decir, darle un peso monetario a los triángulos del Gráfico 8.

Para la evaluación del nivel de beneficios, se consideró a los clientes como candidatos a ofrecerles tarjetas de crédito, y adicionalmente:

- **Para los clientes buenos (que no incurrieron en default)**: la Entidad tiene como beneficio un margen mensual, y anualmente cobra el costo de renovación.
- **Para los clientes malos (los que sí incurrieron en default)**: la Entidad tiene como beneficio el margen mensual y costo de renovación hasta dos meses antes del default, y en el momento del default pierde la suma adeudada multiplicada por el porcentaje de no recupero o LGD (*Loss Given Default*).

La relación estimada entre la pérdida que tiene la Entidad ante un cliente que incumple y el beneficio anual por margen y costos de renovación es de 11. Es decir, un cliente que incurre en default hace perder a la Entidad el valor monetario que se obtuvo por 11 clientes buenos durante un año como margen y cargos.

El Gráfico 9 muestra el nivel de beneficios para distintos niveles de exigencia para el punto de corte (de mayor a menor exigencia), realizado sobre la base de validación del modelo y tomando la definición de 11 entre la pérdida en caso de default y el beneficio anual.

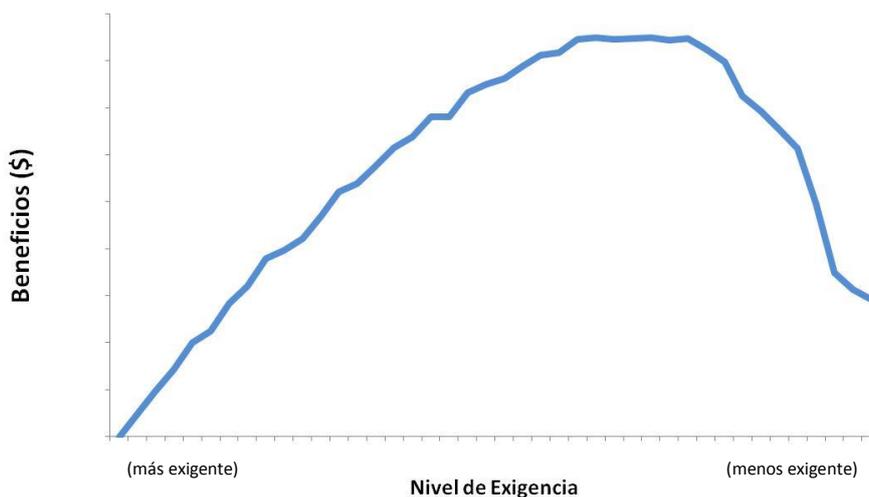


Gráfico 9: Gráfico de beneficios monetarios – Modelo logístico (Elaboración Propia)

Se puede apreciar que si la Entidad es demasiado exigente (a la izquierda del gráfico) no tiene beneficios ya que su cartera es muy pequeña o nula. Es 100% efectiva en no tener clientes malos, pero se pierde los beneficios de los buenos. A medida que es menos exigente, da oportunidad de ir aceptando más clientes buenos, pero también malos. Al principio el balance es positivo, es decir, entran muchos clientes buenos pero pocos malos, pero llega un punto en el cual si sigue relajando el nivel de exigencia y deja pasar clientes malos, estos empiezan a ser demasiados y los beneficios bajan. El último punto del gráfico representa el beneficio que hubiese obtenido de haber aprobado a todos los clientes sin exigirles una mínima calidad crediticia medida por el score.

Se puede concluir que el modelo ayuda a la Entidad a tomar mejores decisiones desde el punto de vista económico, siempre y cuando el último punto de beneficio esté por debajo del máximo, lo cual normalmente ocurre.

El punto de exigencia para el cual la Entidad maximiza los beneficios es aquél que denominaremos óptimo. Para el caso del Modelo Logístico, este nivel se encuentra en 5.080 veces el beneficio anual por margen y cargos.

5. Análisis Multivariados - Modelo de Supervivencia

5.1. Definición del Modelo de Supervivencia⁵

Se dispone de **N** realizaciones de la variable **Y**, siendo ésta el flag de default de cada uno de los **N** individuos. A diferencia del Modelo Logístico, esta no es la variable a explicar, sino que es una variable indicadora de **censura** de los datos. Si la variable **Y** toma valor 0, el individuo no pudo ser observado hasta su potencial fecha de default y por lo tanto se encuentra censurado. Si la variable **Y** toma el valor 1, se computa el tiempo en meses desde el inicio del crédito hasta la fecha de default. Así se construirá entonces las **N** realizaciones de la variable τ , siendo ésta el tiempo de observación de los individuos, o tiempo de supervivencia. En caso que el individuo se encuentre censurado, τ representa los meses de observación, y en caso contrario, representa el tiempo transcurrido hasta el default.

A diferencia del Modelo Logístico, se computan todas las situaciones de default, incluyendo aquellas que ocurren luego de los primeros 12 meses del crédito.

5.1.1. Función de Supervivencia

Indica la probabilidad de que cierto individuo sobreviva sin default hasta el momento **t**, o bien que el tiempo transcurrido hasta el default (variable τ) sea superior o igual a **t**. También representa el complemento de la probabilidad de default acumulada (**PDA**) hasta el momento **t**. En fórmula:

$$S_{(t)} = P[\tau \geq t] = 1 - PDA_{(t)}$$

Ecuación 8: Función de Supervivencia

5.1.2. Tasa Instantánea de Salida o Función Hazard Rate

Es una pieza fundamental de los Modelos de Supervivencia, como se mencionó al principio del trabajo. Indica la probabilidad de default inmediatamente después de alcanzar el momento **t** condicional a la supervivencia hasta el momento **t**. En fórmula:

$$h_{(t)} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq \tau < t + \Delta t | \tau \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq \tau) - P(t + \Delta t \leq \tau)}{P(t \leq \tau) \times \Delta t}$$

Ecuación 9: Definición matemática de Hazard Rate

La función de *Hazard Rate* puede interpretarse como la fuerza con la cual un individuo en un determinado momento es arrastrado a la situación de insolvencia. Reemplazando la Función de Supervivencia (Ecuación 8) en la Ecuación 9 se obtiene:

$$h_{(t)} = -\frac{1}{S_{(t)}} \frac{dS_{(t)}}{dt}$$

Ecuación 10: Hazard Rate en función de la Supervivencia

Entonces, la tasa instantánea de supervivencia caracteriza al tiempo de supervivencia. Se puede apreciar que al resolver la Ecuación 10 se obtiene la supervivencia al momento **t** como una función de todas las tasas instantáneas de salida vigentes hasta dicho instante:

⁵ Muchos aspectos técnicos serán explicados de modo general, ya que no se encuentra dentro de los objetivos del trabajo el desarrollar exhaustivamente todas las demostraciones pertinentes.

$$S(t) = e^{-\int_0^t h(u)du}$$

Ecuación 11: Supervivencia en función del Hazard Rate

Entonces, se puede caracterizar a la supervivencia hasta un cierto momento conociendo todas las tasas instantáneas desde 0 hasta ese momento. Resulta entonces primordial conocer esas tasas y cómo evolucionan con el tiempo. Se denomina **Duration Dependence** a la relación existente entre la tasa de salida y el tiempo que ha transcurrido en un determinado estado (Gourieroux & Jasiak, 2007) la cual puede ser positiva (a mayor tiempo sin default, mayor fuerza de arrastre), negativa o inexistente.

El modelado de la tasa instantánea de salida y su relación con el tiempo será crucial. El Modelo Exponencial asume que no existe relación de dependencia, o sea, que la tasa instantánea de salida es constante. Modelos más complejos como el de distribución Gamma, Weibull o Lognormal permitirán dotar de complejidad al análisis. Para los objetivos de este trabajo se utiliza el **Modelo Exponencial**, dado que es el más simple y adicionalmente se lo puede adaptar para modelar algún tipo de cambio en el *hazard* con posterioridad a la originación del préstamo.

5.1.3. Modelo Exponencial de Riesgos Proporcionales

El Modelo Exponencial considera a la tasa instantánea de salida constante en el tiempo. Sin embargo, cada individuo puede tener su propia *hazard rate* en función del riesgo inferido a través de su grupo de riesgo. Este grupo de riesgo está representado por la realización para el individuo del vector \mathbf{X} (también denominado vector de Covariantes).

El modelo de Riesgos Proporcionales de Cox (Cox & Oakes, 1984) propone modelar la tasa instantánea de salida de cada individuo de manera proporcional a una tasa base $h_{0(t)}$.

Se introduce entonces un vector \mathbf{X} cuyos k componentes serán las variables explicativas. Cada individuo tendrá una realización particular \mathbf{X}_i de este vector según su rango de score, edad, etc. Existirá además un vector $\boldsymbol{\theta}$ de parámetros que afectarán a cada una de las variables explicativas. Entonces, la *hazard rate* se define como:

$$h_{i(t)} = h_i = h_{0(t)} \times e^{\mathbf{X}_i^T * \boldsymbol{\theta}}$$

Ecuación 12: Modelo Exponencial de Riesgos Proporcionales

La tasa base $h_{0(t)}$ se puede modelar de distintas formas. La estimación del vector de parámetros $\boldsymbol{\theta}$ se podrá realizar independientemente de la forma de estimación de la tasa base. Para la estimación de los parámetros en este trabajo se utiliza el concepto de **Verosimilitud parcial**, ya que lo que se utiliza en la maximización no son probabilidades, sino un proxy de éstas en base a cada grupo de supervivencia.

Dado que la determinación de la PD dependerá del vector de parámetros $\boldsymbol{\theta}$, una vez definido \mathbf{X} , pero además se requiere una estimación para cada t de la tasa base $h_{0(t)}$, se dice que el modelo de Riesgos Proporcionales es un **modelo semi-paramétrico**. Remplazando el Modelo Exponencial de Riesgos Proporcionales (Ecuación 12) en la Ecuación 11, se obtiene:

$$S_i(t) = e^{-\int_0^t h_0(u)du} \times e^{X_i^T \theta} = \left(e^{-\int_0^t h_0(u)du} \right) e^{X_i^T \theta}$$

$$S_i(t) = S_0(t) e^{X_i^T \theta} = \text{Supervivencia Base}^{\text{Riesgo Relativo}}$$

Ecuación 13: Función de Supervivencia en Modelo Exponencial de Riesgo Proporcional

donde la expresión $S_0(t) = e^{-\int_0^t h_0(u)du}$ es el correlato de la tasa base en la función de supervivencia, y se la puede denominar entonces **Supervivencia Base**. La expresión $e^{X_i^T \theta}$ es también usualmente denominada **riesgo relativo**. Entonces, la función de supervivencia en el Modelo Exponencial de Riesgos Proporcionales se puede descomponer en una Supervivencia Base que aplica a todos los individuos en distintos momentos de tiempo, y el riesgo relativo de cada individuo que se mide a través de sus características **X**.

Las estimaciones de probabilidad de supervivencia para un cierto momento serán entonces definidas a través del riesgo relativo y la función de supervivencia base $S_0(t)$. Esta función, junto con los coeficientes θ son los outputs de la corrida del modelo.

Los detalles del proceso de estimación y de las propiedades de los estimadores están fuera del alcance de este trabajo. Los mismos se pueden consultar en Gourieroux & Jasiak (2007) y en SAS Institute Inc. (1999).

5.2. Resultados de la calibración

Uno de los propósitos de este trabajo es realizar comparaciones del Modelo de Supervivencia con el Modelo Logístico. Por eso, se estima directamente el Modelo de Supervivencia utilizando las mismas variables y categorías utilizadas en el Modelo Logístico. Esto se realiza para no dar ventaja a ninguno de los dos modelos por la utilización de mayor información.

Se seleccionó el siguiente modelo para el modelado del riesgo relativo:

$$\ln(RRelativo) = \theta_1 \times ScoreBureau + \theta_2 \times Edad + \theta_3 \times EstCivil + \theta_4 \times Prov$$

Ecuación 14: Modelado del riesgo relativo

En la Tabla 4 se muestran los resultados del proceso de estimación.

Variable	Categoría	Estimador	Pr > ChiSq	Riesgo Relativo
Score_Bureau	A	2,4975	<.0001	12,1520
	B	2,0658	<.0001	7,8910
	C	1,6096	<.0001	5,0010
	D	1,3612	<.0001	3,9010
	E	1,0599	<.0001	2,8860
	F	0,8832	<.0001	2,4190
	G	0,4199	0.0009	1,5220
	H	0,0000	.	.
Edad	A	0,3596	<.0001	1,4330
	B	0,2770	<.0001	1,3190
	C	0,2639	<.0001	1,3020
	D	0,1352	0.0146	1,1450
	E	0,0000	.	.
Provincia	Resto	0,0086	0.8301	1,0090
	BsAs-Cap	0,0000	.	.
Estado Civil	Resto	0,1940	<.0001	1,2140
	Casado	0,0000	.	.

Tabla 4: Parámetros del Modelo de Supervivencia (Elaboración propia)

La columna “Estimador” representa los valores estimados de los coeficientes θ_i . La columna “Pr > ChiSq” representa el p valor en el test individual de Wald⁶. Esto último es lo que se denomina *Significatividad Estadística* de θ_i y representa un proxy de la probabilidad de que el coeficiente sea nulo. La columna “Riesgo Relativo” indica el resultado de e^{θ_i} , y sirve para medir el impacto potencial que la variable tiene en la supervivencia (ver Ecuación 13).

Se puede observar que todas las categorías, a excepción de la de “Provincia Resto”, tienen significatividad estadística al 5%. Además, los coeficientes son crecientes en función del comportamiento esperado frente al riesgo.

La categoría de “Provincia Resto” no se excluyó ya que se considera aceptable que el coeficiente sea positivo y mayor que el de la categoría inmediata inferior (indicando que el resto de las provincias son levemente más riesgosas que Buenos Aires y Capital Federal). Es decir, la variable Provincia sigue guardando lógica económica aunque una de sus categorías no tenga significatividad estadística. Esto nos brinda la posibilidad de un ajuste (pequeño) en la categoría de riesgo del cliente según la provincia. Además, como se mencionó anteriormente, para realizar comparaciones con el Modelo Logístico se decidió tomar las mismas variables y categorías.

La función de supervivencia base $S_{0(t)}$ es otro de los outputs del modelo. Consiste en una tabla de probabilidades para cada tiempo de supervivencia. La misma se muestra en el Gráfico 10.

5.3. Ejemplos de aplicación

Como se mencionó anteriormente, el modelo calibrado se trata de un modelo semi-paramétrico. Eso significa que una vez estimados los coeficientes (parámetros), se debe

⁶ Este test evalúa la hipótesis nula de que cada uno de los coeficientes tiene valor cero. El p valor surge de comparar un estadígrafo de prueba con la distribución chi-cuadrado con un grado de libertad (SAS Institute Inc., 1999, pág. 2576)

obtener la función de Supervivencia Base $S_{0(t)}$ o bien la tasa base $h_{0(t)}$ para poder calcular las estimaciones de la probabilidad de default de cada individuo a distintos plazos.

Por ejemplo, si un individuo tiene categoría de Score F, Categoría de Edad B, es Soltero y Reside en Capital Federal, el exponente del riesgo relativo (Ecuación 14) es⁷:

$$\ln(RRelativo) = 0,8832 + 0,2770 + 0,1940 + 0 = 1,3542$$

Ecuación 15: Riesgo relativo para un caso particular

Adicionalmente, se toman los valores de la función de supervivencia entre los outputs del modelo $S_{0(12)} = 0,9853$ y $S_{0(24)} = 0,9724$.

La probabilidad de default a 12 meses para este individuo es $1 - S_{i(12)}$, que en función de la Ecuación 13 resulta ser:

$$PD_{(12)} = 1 - S_{i(12)} = 1 - 0,9853 e^{1,3542} = 1 - 94,42\% = 5,58\%$$

Ecuación 16: Probabilidad de Default a 12 meses para un caso particular

El valor se encuentra bastante cercano al 4,98% estimado por el Modelo Logístico para este mismo ejemplo, siendo la tasa de default real 5,11%.

Adicionalmente, la probabilidad de default a 24 meses para este individuo en función de la Ecuación 13 es:

$$PD_{(24)} = 1 - S_{i(24)} = 1 - 0,9724 e^{1,3542} = 1 - 89,72\% = 10,28\%$$

Ecuación 17: Probabilidad de Default a 24 meses para un caso particular

El Gráfico 10 muestra la curva de supervivencia base $S_{0(t)}$ (obtenida entre los outputs del modelo), la curva de supervivencia para el caso del ejemplo, y el punto de probabilidad de supervivencia estimado por el Modelo Logístico (complemento de la PD).

⁷ Se omiten los términos de categorías nulas y se redondea al cuarto decimal.

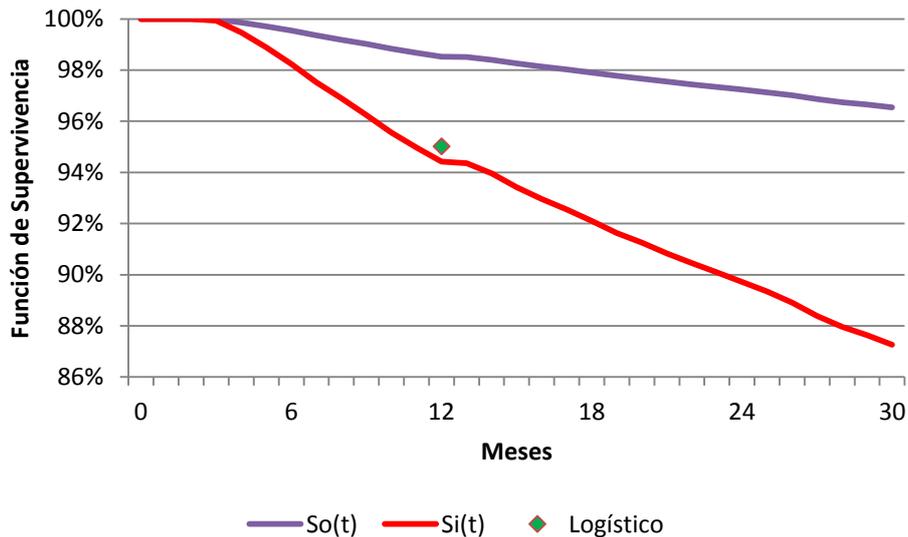


Gráfico 10: Curvas de Supervivencia (Elaboración Propia)

Se puede observar claramente que la curva roja, que representa la función de supervivencia del individuo, pasa muy cercana a la estimación dada por el Modelo Logístico (punto verde). Esto muestra la similitud entre ambos enfoques para el plazo de 12 meses, y pone en evidencia además que el Modelo de Supervivencia puede estimar toda la curva, mientras que el Modelo Logístico estima solo un punto de la misma.

Como el individuo tiene un riesgo relativo mayor que 1, su curva o función de supervivencia se encuentra por debajo de la supervivencia base. Además, existe un alejamiento a ritmo constante de la curva de supervivencia del individuo a la curva de supervivencia base, fiel reflejo del supuesto del modelo que indica que el *hazard rate* es constante (ver Sección 5.1.2., donde se explica el concepto de *duration dependence*). Esto también en parte muestra una limitación del modelo de riesgos proporcionales de Cox, ya que no existe posibilidad de algún cambio relativo en la forma de alejamiento de una función de supervivencia particular respecto de la función de supervivencia base.

5.4. Medidas usuales de performance

A diferencia del Modelo Logístico, para el Modelo de Supervivencia no se estima solo una función paramétrica que tiene relación con el grado de riesgo del individuo, sino que se estima una función que tiene una parte paramétrica (el riesgo relativo) y una parte no paramétrica (la curva de supervivencia base).

Al igual que en el Modelo Logístico, para evaluar si esta función es razonablemente una buena indicadora del riesgo, se buscarán mostrar básicamente dos cuestiones, a saber:

- La primera, es que la parte paramétrica (que es particular de cada individuo), logre **ordenar** a la población de menor a mayor riesgo.
- La segunda, es obtener una medida de probabilidad de supervivencia (o su complemento, la probabilidad de default) cercana a la real, es decir, que el modelo logre **predecir** qué porcentaje de los individuos de cada grupo entraría en insolvencia en cada plazo.

5.4.1. Comparación con tasas de default reales

La parte paramétrica de la función (riesgo relativo) ordena a la población en categorías de riesgo. Es conveniente en primer lugar realizar algún tipo de agrupación para generar categorías de riesgo similar. Usualmente se separa en deciles o bien alguna otra medida de uniformidad entre los grupos. La comparación se realiza en la base de desarrollo y en la base de validación.

5.4.1.1. Base de Desarrollo

Se dividió a la base de desarrollo en 15 grupos de menor a mayor riesgo relativo para comparar las tasas de default reales a 12 meses (en negro) con la probabilidad de default estimada para 12 meses con el Modelo de Supervivencia (en rojo en el Gráfico 11).

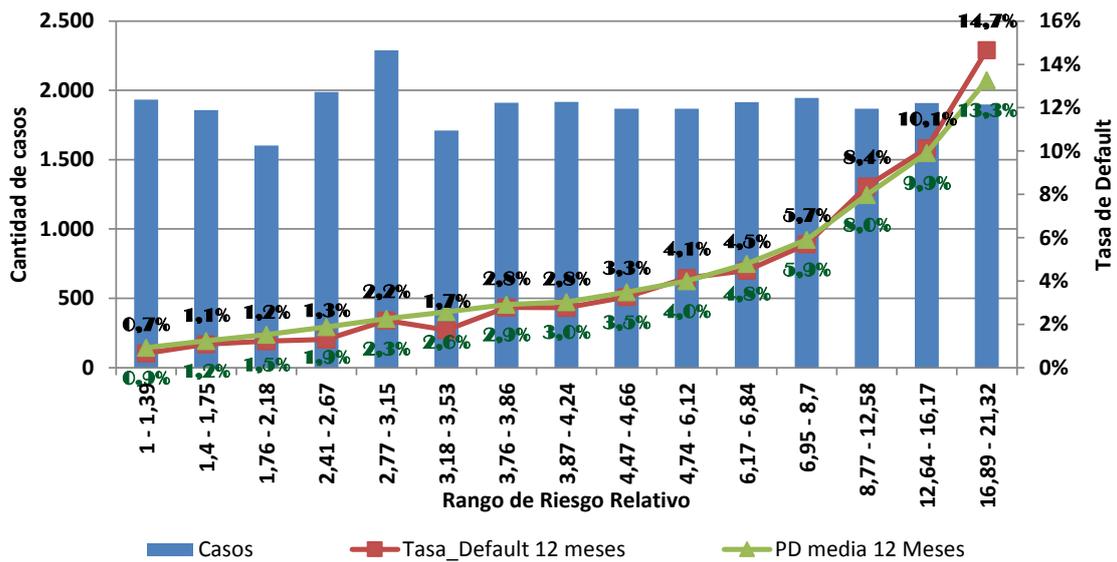


Gráfico 11: Análisis de ordenamiento y Predicción del Modelo de Supervivencia (Elaboración Propia)

Se puede observar que el riesgo relativo **ordena** a los grupos en categorías de riesgo diferenciadas para el análisis de default a un plazo 12 meses. Solamente se encuentra una pequeña inconsistencia entre el quinto y sexto grupo donde las tasas de default caen cuando tendrían que mantenerse o incrementarse. Además, si se compara este gráfico con el equivalente en el Modelo Logístico (ver Gráfico 3 del capítulo anterior), los valores de tasas de default son muy similares en cada grupo.

El Gráfico 12 analiza la tasa de default una vez transcurridos los 12 meses.

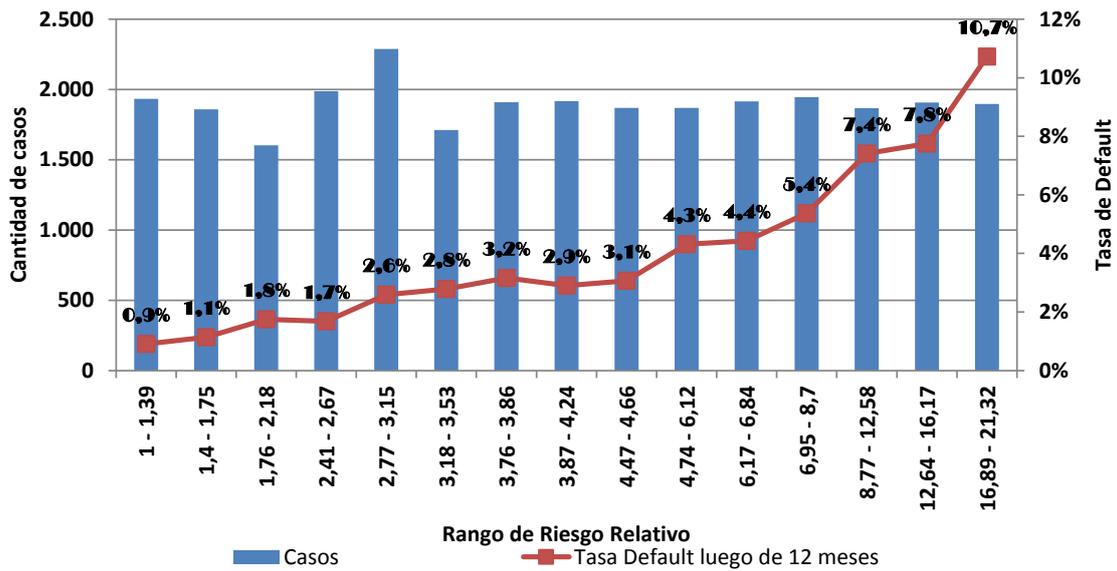


Gráfico 12: Análisis de ordenamiento y Predicción del Modelo de Supervivencia luego de 12 meses (Elaboración Propia)

Es destacable el hecho de que el modelo sigue sirviendo para explicar el ordenamiento del comportamiento de los clientes a largo plazo, aunque para algunas categorías no se cumple crecimiento estricto de la tasa de mora a medida que disminuye el riesgo relativo asignado.

5.4.1.2. Base de Validación

Se dividió a la base de validación en 8 grupos de menor a mayor riesgo relativo para comparar las tasas de default (ver Gráfico 13).

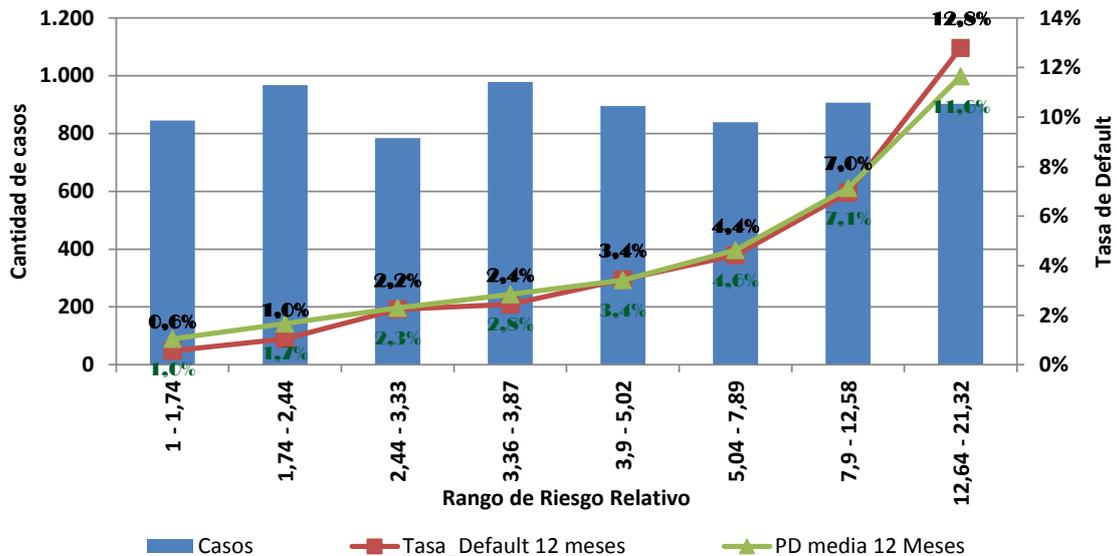


Gráfico 13: Análisis de ordenamiento y Predicción del modelo Supervivencia - Validación (Elaboración Propia)

Se puede observar que el riesgo relativo **ordena** a los grupos en categorías de riesgo diferenciadas también en la base de validación.

El Gráfico 14 muestra la tasa de default a largo plazo, es decir, la tasa de default que contempla los clientes que se vuelven insolventes luego de los 12 meses de iniciado el crédito.

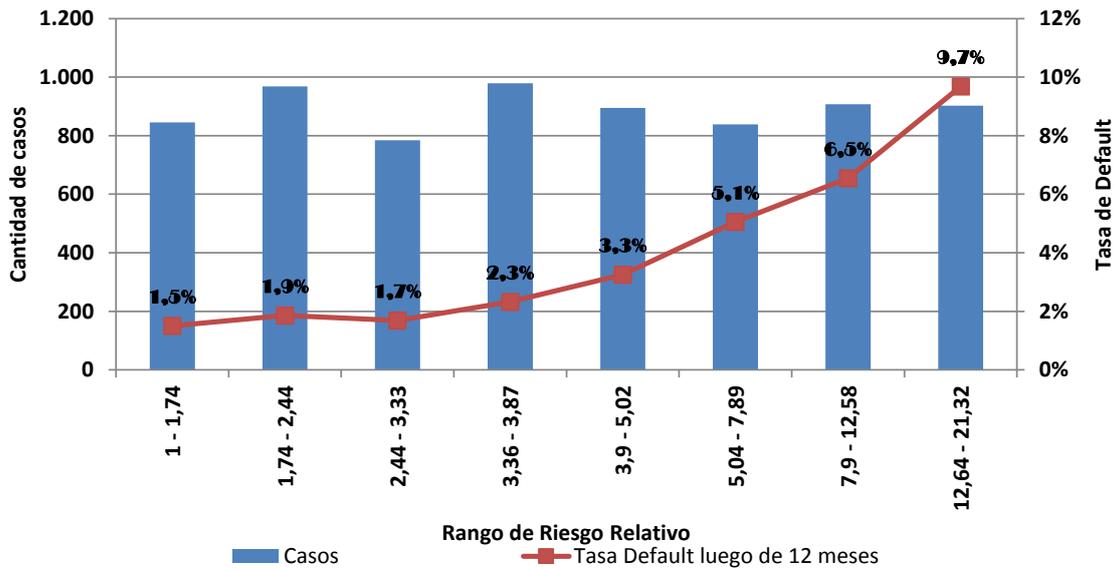


Gráfico 14: Análisis de orden y Predicción del modelo Supervivencia en Largo Plazo - Validación (Elaboración Propia)

Es destacable el hecho de que el modelo sigue sirviendo para explicar un ordenamiento del comportamiento de los clientes a largo plazo, aunque para una categoría no se cumple el crecimiento estricto de la tasa de mora a medida que aumenta el riesgo relativo asignado. La performance observada del modelo en una base de datos independiente, como lo es la base de validación, es aceptable.

5.4.2. Área ROC y Power Stat

El Gráfico 15 muestra comparativamente para el Modelo de Supervivencia y el Modelo Logístico las curvas del modelo (ver Página 18), considerando el default a 12 meses.

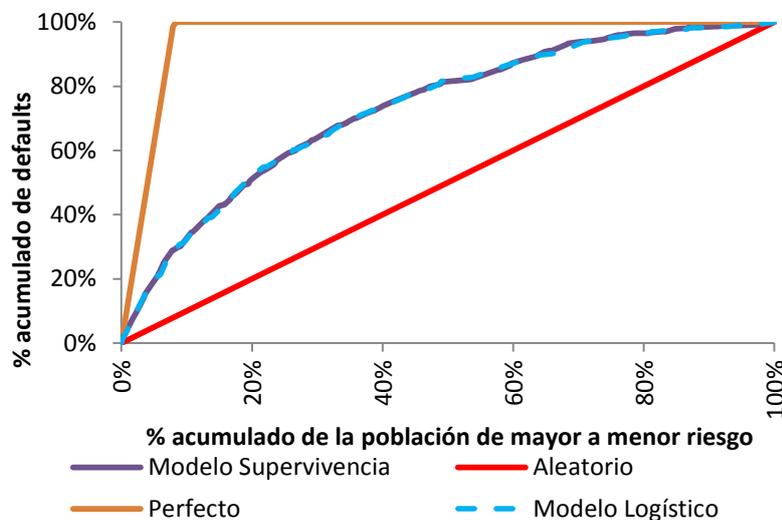


Gráfico 15: Curva del Modelo – Defaults a corto plazo – Modelo Supervivencia – Validación (Elaboración Propia)

La Tabla 5 muestra el resumen de indicadores.

	Corto plazo		Largo Plazo		Total	
	Supervivencia	Logístico	Supervivencia	Logístico	Supervivencia	Logístico
Área ROC	75,12%	75,08%	69,76%	69,55%	74,59%	74,32%
Power Stat	50,25%	50,16%	39,53%	39,10%	49,18%	48,64%

Tabla 5: Medidas de performance Modelo de Supervivencia (Elaboración propia)

Como se puede apreciar en la Tabla 5, las medidas de performance son similares en ambos modelos. En ambos plazos, existe una pequeña superioridad del Modelo de Supervivencia. Para apreciar visualmente la similitud, basta con observar las curvas azul continua y celeste discontinua en el Gráfico 15.

Si hablamos del fenómeno de la supervivencia, independientemente del plazo, se puede apreciar en la sección “**Total**” de la Tabla 5 que el riesgo relativo medido por el Modelo de Supervivencia explica un poco mejor el fenómeno que el Modelo Logístico.

5.5. Valuación económica de la performance

El análisis de Punto de Corte que se realizó para el Modelo de Supervivencia es el mismo que el expuesto en la Sección 4.5 para el Modelo Logístico. Se utilizó el riesgo relativo como ordenador de los clientes, y se considera hay un mayor grado de exigencia cuando se solicita un riesgo relativo bajo para aceptar a un cliente.

En el Gráfico 16 se muestran los beneficios para distintos niveles de exigencia (de mayor a menor exigencia), realizado sobre la base de validación del modelo y en forma comparativa con el Modelo Logístico. Se observa que ambos modelos compiten equitativamente en generar buenos beneficios, y parece apreciarse que el máximo del Modelo de Supervivencia es ligeramente superior al del Modelo Logístico. El punto de exigencia óptimo para el caso del Modelo de Supervivencia es de 5.168 veces el beneficio anual por margen y cargos, superando a los 5.080 del Modelo Logístico en 1,7%. Si bien esta diferencia parecería en principio despreciable, en la Sección 6.7.2. se analiza la distribución estadística de dicha diferencia porcentual y se muestra que resulta significativa.

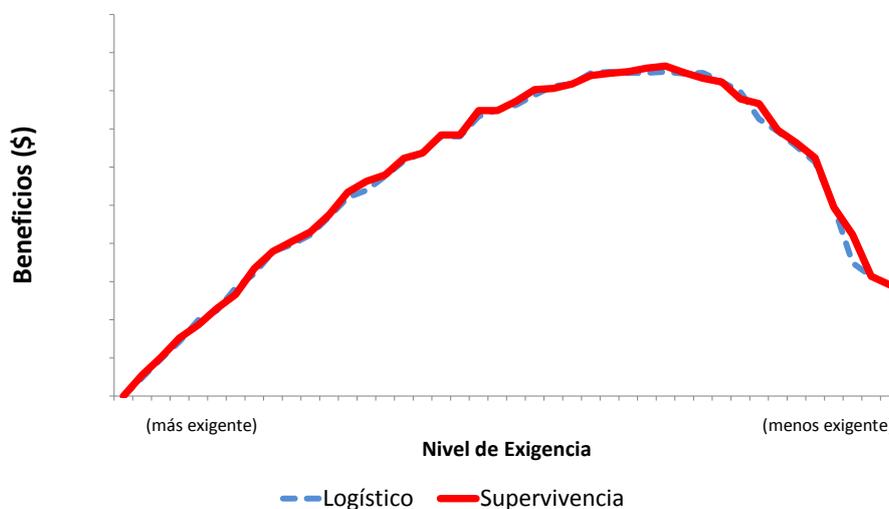


Gráfico 16: Gráfico de beneficios monetarios – Modelo de Supervivencia (Elaboración Propia)

6. Análisis Multivariados - Modelo de Supervivencia sensible a la Mora Sistémica

6.1. Definición del Modelo de Supervivencia Dinámico

El Modelo de Supervivencia Dinámico (o sensible a la Mora Sistémica, indiferentemente) tiene fundamentos similares al expuesto en el capítulo anterior y, por lo tanto, no se explicarán mayores detalles en este capítulo.

Como se señaló en la Ecuación 11, la función de supervivencia hasta el momento t , $S_{(t)}$, se definirá conociendo todas las tasas de salida $h_{(u)}$ desde el momento 0 hasta el momento t .

En el Modelo Exponencial de Riesgos Proporcionales se asume que las tasas de salida $h_{i(t)}$ son constantes para cada individuo. En particular, en el modelo que se calibró, se asumen constantes desde 0 hasta t . Este es el motivo por el cual cada curva de supervivencia individual $S_{(t)}$ suele alejarse a un ritmo constante de la curva de Supervivencia Base $S_{0(t)}$ (ver Gráfico 10) y no es posible revertir dicho alejamiento o cambiar su ritmo. Para generar algún tipo de comportamiento diferencial, como se comentó en el capítulo anterior, existen modelos más complejos como el de distribución Gamma, Weibull o Lognormal. También se comentó que el modelo exponencial puede adaptarse para modelar algún tipo de cambio en $h_{i(t)}$ con posterioridad a la originación.

En el Modelo Dinámico, se construye una función con las componentes de Supervivencia Base $S_{0(t)}$ y riesgo relativo $e^{Xi^T * \theta}$, donde el exponente del último tiene una componente estática inicial dependiente de las variables disponibles, y una componente dinámica que depende del desarrollo de un indicador seleccionado. Esto es, desarrollaremos el modelo de la Ecuación 13 con la siguiente particularidad:

$$S_{i(t)} = S_{0(t)} e^{Xi^T * \theta + IS(t)^T * \gamma}$$

Ecuación 18: Función de Supervivencia en Modelo Exponencial de Riesgo Proporcional con Indicador Sistémico

donde $IS(t)$ será el **Vector de Indicadores Sistémicos**, que depende de la fecha de supervivencia y es igual para todos los individuos que alcancen dicha fecha como supervivientes, es decir, afecta a todos los individuos de la misma cohorte por igual. El vector γ será el vector de parámetros de la componente **dinámica**, y θ es el vector de parámetros de la componente **estática**.

6.2. Tratamiento diferencial de la base

Como se comentó anteriormente, en la base disponible tanto para el modelo Logístico y el de Supervivencia, se contó con 35.598 individuos. Para el desarrollo de esos modelos, fue necesario un registro por cada individuo. En cambio, para el desarrollo presente es necesario contar por un registro por cada individuo, cada mes que fue observado.

Por ejemplo, si un individuo originó su crédito en el mes de Marzo 2010, e incumplió (o incurrió en default, indistintamente) o dejó de ser observado en Agosto del mismo año, se contarán con 6 observaciones de los meses de supervivencia 0 (marzo), 1(abril), 2(mayo),

3(junio), 4(julio) y 5(agosto). En caso que haya incurrido en default en el mes 5, será marcado como no censurado. En caso que simplemente haya dejado de ser observado, será marcado como censurado. En todos los meses intermedios, será marcado como censurado. La idea no es solamente obtener múltiples observaciones por cada cliente, sino rescatar los datos del conjunto de indicadores sistémicos vigentes en cada una de las fechas intermedias.

Resumiendo, se obtendrá una base por cada cliente y mes de observación, recabando los datos intermedios del indicador sistémico, y censurando a todas las observaciones excepto por aquel último mes de cada cliente que incumplió. De esta manera, se obtendrá una base con mucho más de 35.598 observaciones, pero con la misma cantidad de defaults. En el Modelo Logístico y en el de Supervivencia presentado en el Capítulo 5 (o Modelo de Supervivencia Estático, indiferentemente), a diferencia del Modelo Dinámico, se evalúa al cliente solamente al momento de la originación y, por lo tanto, existen tantas observaciones como individuos.

6.3. Indicador sistémico seleccionado

Se realizaron diversos ensayos para seleccionar un indicador sistémico. Por motivos explicados anteriormente, se decidió tomar la serie mensual de mora temprana en el sistema (que representa la cartera en situación 2 en adelante sobre la cartera total de créditos en el Sistema Financiero Argentino, publicada por el Banco Central de la República Argentina). A partir de la realización de diversos análisis sobre la serie de mora temprana, se decidió construir un indicador de riesgo siguiendo los siguientes pasos:

- Realizar las variaciones porcentuales mensuales de dicha serie, es decir:

$$\frac{Mora\ Temprana_{(t)}}{Mora\ Temprana_{(t-1)}} - 1$$

- Realizar el promedio de dichas variaciones de los últimos 3 meses anteriores al mes de análisis, con el objeto de quitarle volatilidad intermensual al índice.
- Desfasar la serie dos meses, ya que conceptualmente representa mora de 30 días y la definición de default adoptada es de atrasos mayores a 90 días.
- Evaluar si el índice así desfasado tiene una variación de 2 puntos porcentuales⁸ o más en un mes, construyendo una serie de indicadores de alerta de acuerdo al siguiente proceso:
 - Si la variación del índice desfasado es mayor que 2 puntos porcentuales entre el mes 0 y el tercer mes desde la originación, se crea un indicador llamado “**alerta_3**” que se mantendrá vigente desde el momento de activación hasta el final del crédito.
 - Si el evento se da entre el cuarto y sexto mes desde la originación, se crea un indicador denominado “**alerta_6**” que también se mantendrá vigente desde que se active hasta el final del crédito.
 - De igual modo se construyen los indicadores para los períodos comprendidos entre el séptimo y décimo segundo mes (“**alerta_12**”); entre el décimo tercer y décimo octavo mes (“**alerta_18**”); y entre el décimo noveno y trigésimo (“**alerta_30**”).
- Las **cinco alertas** así construidas irán colocando a los clientes con el correr del tiempo en categorías de riesgo más elevado en función de la tendencia de la mora sistémica temprana.

⁸ La decisión respecto de considerar una variación de 2 puntos porcentuales para activar la alerta se basó en el análisis de la cartera particular bajo estudio y debería recalibrarse al analizar otra cartera.

La idea detrás de este indicador es generar una variable de alerta, que permita explicar por qué una condición sistémica puede arrastrar a los clientes en un momento dado a la situación de insolvencia al mismo tiempo y con distintas intensidades según la antigüedad del crédito. Por su construcción mediante promedios de variaciones desfasados, en el momento de la originación del crédito se utiliza la información de la mora sistémica de los últimos 5 meses, quedando definidos 2 de los 3 meses en los cuales se determina la “alerta_3”. Para definir la PD a asignar en cada plazo, debe incluirse una proyección de la mora sistémica futura.

6.3.1. Análisis bivariado del indicador sistémico

A los 28.478 clientes de la base de desarrollo, se les ejecutó el procedimiento definido con anterioridad, resultando en 699.929 clientes-mes desde el mes de supervivencia 0 hasta el último mes de observación. Esto resulta en un plazo promedio de observación de 23.6 meses (excluyendo el mes 0).

6.3.1.1. Indicador “Alerta 3”

Como se indicó anteriormente, entre el mes 0 y el mes 3 de observación, a un cliente se le puede activar esta alerta cuando se cumplan las condiciones de deterioro sistémicas. Al mes 4, se detecta la cantidad de clientes que cumplen con las condiciones de deterioro, y se van contabilizando los defaults, creando un porcentaje de supervivencia en cada momento posterior. El mismo análisis se realiza con los clientes que no sufrieron dicha alerta. De la comparación de las curvas puede observarse si dicha alerta es un buen indicador de riesgo o no. El Gráfico 17 muestra el análisis comparativo.

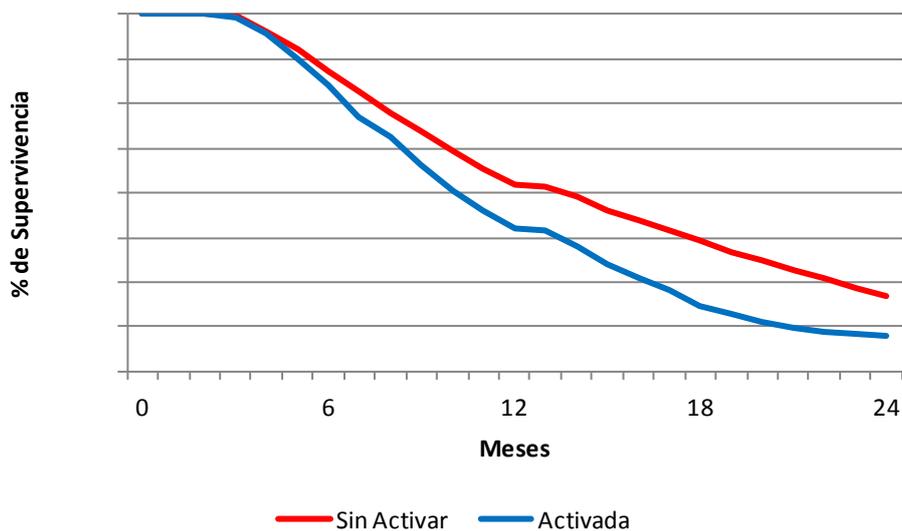


Gráfico 17: Análisis Bivariado de “Alerta 3” (Elaboración Propia)

Se observa claramente que aquellos clientes que ven activada la alerta, tienen en los meses subsiguientes al tercero una supervivencia menor que el resto. Las diferencias se van haciendo más evidentes con el transcurso de los meses e incluso se mantienen hasta el mes 24 de observación. Sin embargo, no existe un alejamiento a ritmo constante de la curva azul respecto de la roja, como lo indicaría el supuesto del Modelo Exponencial de Riesgos Proporcionales. Siendo conscientes de dicha limitación del modelado, se incluye la variable

sabiendo que luego de un cierto tiempo de permanencia de los individuos, su impacto debería aminorarse.

El análisis para el resto de los indicadores es similar, haciendo la diferencia en la cantidad de meses a considerar.

6.3.1.2. Indicador “Alerta 6”

Entre el mes 4 y el mes 6 de observación, a un cliente se le puede activar esta alerta cuando se cumplan las condiciones de deterioro sistémicas. Al mes 7, se detecta la cantidad de clientes que cumplen con las condiciones de deterioro, y se van contabilizando los defaults, creando un porcentaje de supervivencia en cada momento posterior. El mismo análisis se realiza con los clientes que no sufrieron dicha alerta. En el Gráfico 18 se puede advertir que aquellos clientes que ven activada la alerta hasta el mes 6 tienen supervivencias muy similares, pero luego se observa una supervivencia menor que el resto. Las diferencias se van haciendo más evidentes con el transcurso de los meses e incluso se mantienen hasta el mes 30 de observación. Sin embargo se observa, al igual que el indicador “Alerta 3” que no existe un alejamiento a ritmo constante entre ambas curvas.

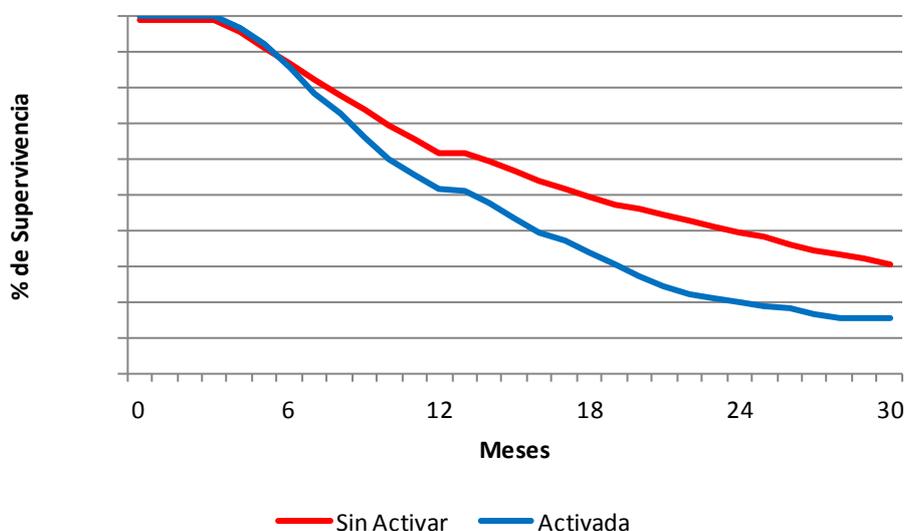


Gráfico 18: Análisis Bivariado de “Alerta 6” (Elaboración Propia)

6.3.1.3. Indicador “Alerta 12”

Entre el mes 7 y el mes 12 de observación, a un cliente se le puede activar esta alerta cuando se cumplan las condiciones de deterioro sistémicas (ver Gráfico 19). Es destacable que aquellos clientes que ven activada la alerta desde el mes 7 hasta el mes 12 tienen supervivencias muy similares, pero luego se observa una supervivencia menor que el resto. Las diferencias se van haciendo más evidentes con el transcurso de los meses e incluso se mantienen hasta el mes 36 de observación

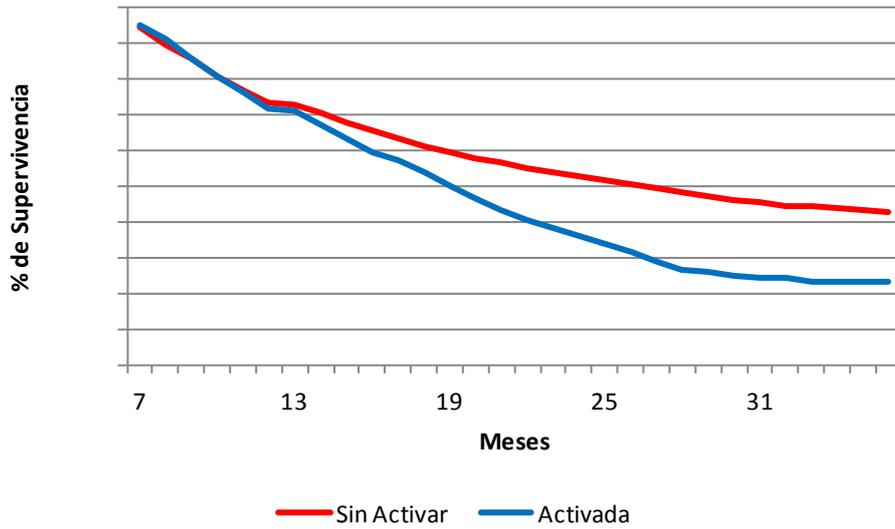


Gráfico 19: Análisis Bivariado de "Alerta 12" (Elaboración Propia)

El Gráfico 20 muestra el análisis bivariado para la "Alerta 18", mientras que el Gráfico 21 hace lo propio para la "Alerta 30". Las conclusiones de los gráficos son similares a las mencionadas para las alertas 3, 6 y 12.

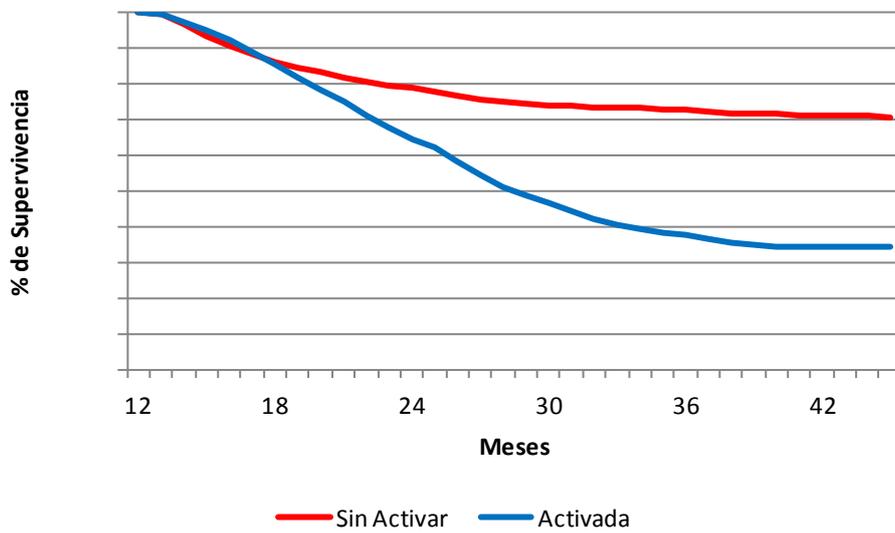


Gráfico 20: Análisis Bivariado de "Alerta 18" (Elaboración Propia)

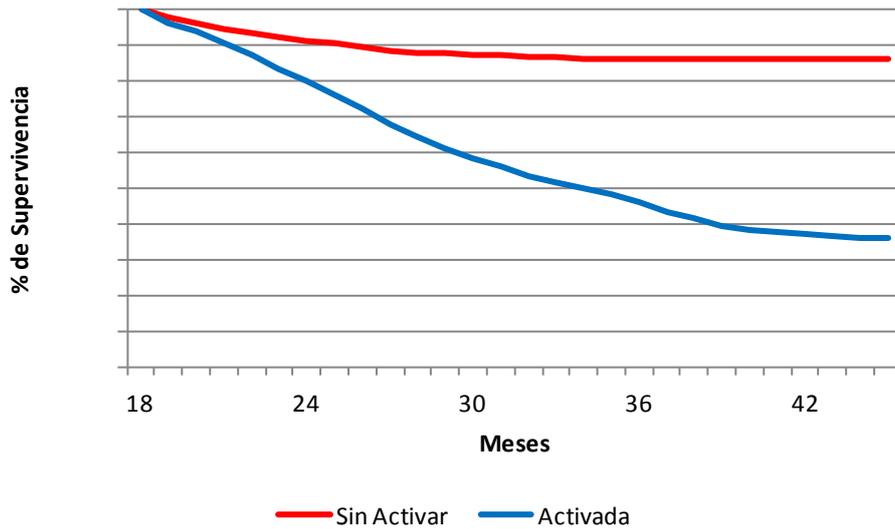


Gráfico 21: Análisis Bivariado de "Alerta 30" (Elaboración Propia)

Dado que resulta suficientemente sólida la comparación de todas las curvas presentadas, se decidió incluir las cinco variables en la parte dinámica del Modelo de Supervivencia.

6.4. Resultados de la calibración

Si bien en primer lugar se intentó calibrar el Modelo de Supervivencia sensible a la Mora Sistémica con las mismas variables explicativas utilizadas en los modelos anteriores, agregando lógicamente el indicador sistémico, se terminó descartando la variable de Provincia de la Sucursal, ya que la misma perdió lógica económica y significatividad estadística al competir con dicho indicador.

Finalmente, se seleccionó el siguiente modelo:

$$\ln(RRelativo) = \theta_1 \times ScoreBureau + \theta_2 \times Edad + \theta_3 \times EstCivil + \gamma \times IndMora$$

Ecuación 19: Modelo para el riesgo relativo con indicador sistémico

En la Tabla 6 se muestran los resultados de la calibración.

Variable	Categoría	Estimador	Pr > ChiSq	Riesgo Relativo
Score_Bureau	A	2,5330	<.0001	12,5910
	B	2,0748	<.0001	7,9630
	C	1,6290	<.0001	5,0990
	D	1,3714	<.0001	3,9410
	E	1,0633	<.0001	2,8960
	F	0,8810	<.0001	2,4130
	G	0,4157	0.0010	1,5150
	H	0,0000	.	0
Edad	A	0,3443	<.0001	1,4110
	B	0,2652	<.0001	1,3040
	C	0,2522	<.0001	1,2870
	D	0,1256	0.0234	1,1340
	E	0,0000	.	0
Estado Civil	Soltero	0,1997	<.0001	1,2210
	Resto_	0,0000	.	0
Indicador de Mora en Sistema Financiero	alerta_3	0,3056	<.0001	1,3570
	alerta_6	0,1082	0.0216	1,1140
	alerta_12	0,1726	0.0004	1,1880
	alerta_18	0,1817	0.0003	1,1990
	alerta_30	0,1160	0.1016	1,1230

Tabla 6: Parámetros del Modelo de Supervivencia sensible a la Mora Sistémica (Elaboración propia)

Se puede observar que todas las categorías, a excepción de “Alerta_30”, tienen significatividad estadística al 5%. Además, los coeficientes son crecientes en función del comportamiento esperado frente al riesgo, y en el caso del Indicador Sistémico todas las categorías tienen coeficiente positivo, indicando que todas las alertas guardan relación creciente con el riesgo. Cabe mencionar que no se requiere que los coeficientes de las alertas sean crecientes, ya que no se espera que el impacto del sistema financiero sea creciente, decreciente o uniforme respecto del momento en el cual se dispara el alerta.

La curva de supervivencia base, el otro output de la calibración, se muestra en forma gráfica en la Sección 6.5.

6.5. Ejemplos de aplicación

Siguiendo el ejemplo del resto de los modelos, si un individuo tiene categoría de Score F, Categoría de Edad B, es Soltero y Reside en Capital Federal, la parte estática del exponente del Riesgo Relativo (Ecuación 18) es:

$$\ln(RR_{\text{Relativo}}) = X_i^T * \theta = 0,8810 + 0,2652 + 0,1997 = 1,3459$$

Ecuación 20: Riesgo Relativo Estático para un caso particular

Adicionalmente, se contempla el valor de $S_{0(12)} = 0,9872$ y de $S_{0(24)} = 0,9772$.

Se puede observar que existe similitud entre los valores estimados bajo este modelo y el que no posee el ajuste sistémico. Sin embargo, en el modelo con ajuste sistémico, la estimación del Riesgo Relativo y la Supervivencia Base resultan insuficientes para estimar la probabilidad de default a 12 y 24 meses. La PD se podría obtener con dichas variables solamente si se supone que el cliente no pasará por ninguna situación de turbulencia sistémica que active alguna alerta. Es necesario conocer cuál es el impacto que potencialmente tendrá la mora sistémica en la estimación. Para esto, se deben realizar las estimaciones suponiendo algún patrón de evolución del indicador.

Para ser ilustrativos, supongamos que no se activa ninguna alerta en los primeros doce meses. La probabilidad de default a 12 meses, en función de la Ecuación 13 es:

$$1 - S_{1(12)} = 1 - 0,9872 e^{1,3459+0+0+0} = 1 - 95,19\% = 4,81\%$$

Ecuación 21: Probabilidad de Supervivencia Dinámica para un caso particular (no se activan alertas)

El valor se encuentra bastante cercano al 4,98% estimado por el Modelo Logístico para este mismo ejemplo y al 5,54% estimado por el Modelo de Supervivencia estático.

Sin embargo, suponiendo un escenario desfavorable en el que se activan las 3 alertas del primer año, la probabilidad de default es:

$$1 - S_{2(12)} = 1 - 0,9872 e^{1,3459+0,3056+0,1082+0,1726} = 1 - 91,52\% = 8,47\%$$

Ecuación 22: Probabilidad de Supervivencia Dinámica para un caso particular (se activan alertas)

Se puede apreciar que el impacto entre estas situaciones, que es el máximo que el modelo puede brindar a un año, es aumentar un 75% la probabilidad de default.

Para la estimación de la probabilidad de default a 24 meses también es necesario conocer la evolución del indicador sistémico. Considerando el primer caso donde no hay alertas en el primer año, pero suponiendo que al individuo se le activan las alertas de 13 a 18 meses y de 18 a 24, su probabilidad de default a 24 meses sería:

$$1 - S_{1(24)} = 1 - 0,9772 e^{1,3459+0+0+0+0,1817+0,1160} = 1 - 88,78\% = 11,22\%$$

Ecuación 23: Probabilidad de Supervivencia Dinámica para un caso particular (se activan alertas)

Si el individuo del segundo caso (con todas las alertas activadas en el primer año) tiene un alivio en las condiciones macro durante el segundo año, no presentando ninguna alerta, su probabilidad de default a 24 meses sería:

$$1 - S_{2(24)} = 1 - 0,9772 e^{1,3459+0,3056+0,1082+0,1726+0+0} = 1 - 85,31\% = 14,69\%$$

Ecuación 24: Probabilidad de Supervivencia Dinámica para un caso particular (se activan alertas)

Se puede apreciar numéricamente cómo ahora ambos individuos que pasaron por situaciones de turbulencia tienen diferencias menos significativas entre sus PD estimadas.

El Gráfico 22 muestra la curva de Supervivencia Base $S_{0(t)}$, la curva de supervivencia para los dos casos del ejemplo, la curva del Modelo Estático del capítulo anterior y el punto de Probabilidad de Supervivencia estimado por el Modelo Logístico (complemento de la PD).

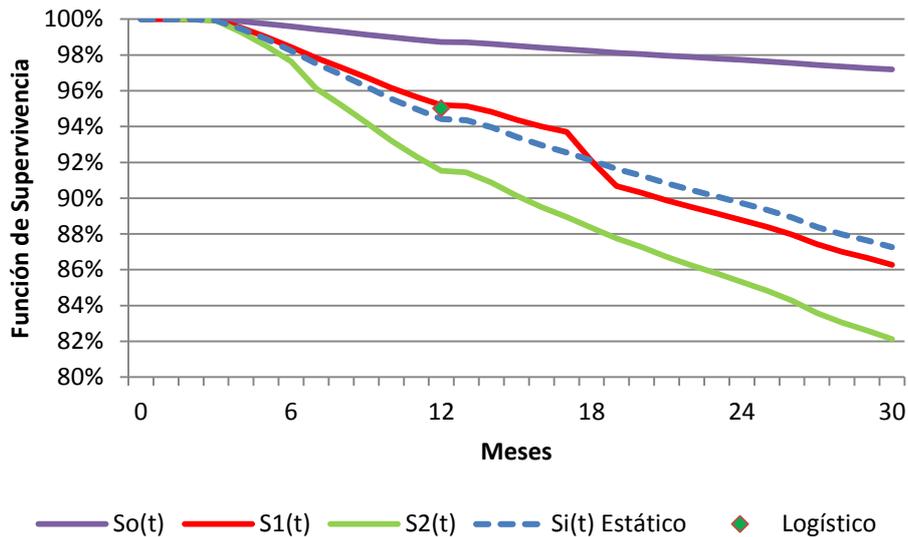


Gráfico 22: Curvas de Supervivencia (Elaboración Propia)

Se puede observar claramente que la curva roja, que representa la función de supervivencia estimada para el primer individuo (sin alertas en el primer año y con alertas en el segundo), pasa muy cercana a la estimación dada por el Modelo Logístico (punto verde), y además que en el primer año presenta una supervivencia mejor que la del Modelo Estático dado que fue un año sin alertas sistémicas. Sin embargo, en el segundo año se revierte la tendencia inicial porque aparecen algunos indicadores en alerta para este individuo. Esto muestra la similitud entre los tres enfoques para el plazo de 12 meses, y pone en evidencia además que el Modelo de Supervivencia Dinámico puede reconocer cómo empeora la supervivencia si se activa alguna alerta. En el caso del individuo 2 (con alertas en el primer año y sin alertas en el segundo), el hecho de que se hayan disparado las tres alertas el primer año lo condiciona durante toda su supervivencia, que se aleja de la supervivencia estática y mantiene una distancia que no recupera.

El hecho de que una vez disparada el alerta no exista posibilidad de revertir el riesgo es una de las limitaciones del Modelo de Riesgos Proporcionales, ya que una vez que el individuo tiene su riesgo relativo desmejorado no es posible que el efecto se suavice a lo largo del tiempo, sino que se mantiene durante toda la vida del crédito. Adicionalmente, en este tipo de modelos, si se mejorara repentinamente riesgo relativo, la curva de supervivencia podría ser creciente en algún intervalo. Es preciso remarcar que existen otros tipos de modelos para el *hazard*, como *Weibull* o *Log-normal*, que reconocen que los efectos sobre el *hazard* no son permanentes sino que cambian su intensidad con el transcurso del tiempo.

6.6. Medidas usuales de performance

El análisis se realiza por cliente y mes de observación. Esto implica que la base de desarrollo del modelo es de casi setecientos mil registros. A diferencia del Modelo de Supervivencia Estático, para el Modelo de Supervivencia Dinámico no se realiza una sola estimación de la Función de Supervivencia, sino que se realiza una estimación en cada mes de observación. En esta sección se pretende aplicar al Modelo de Supervivencia Dinámico los mismos análisis de performance que se realizaron para el Modelo de Supervivencia Estático y el Modelo Logístico.

Resulta entonces indispensable definir un único riesgo relativo dinámico a asignar a cada cliente. Para ello debe elegirse sólo uno de todos los plazos de estimación, y esto impacta en la cantidad de alertas que se considerarán. El criterio que se siguió para la comparación fue asignar el riesgo relativo dinámico hasta los 12 meses, entendiendo a este como el riesgo relativo dinámico asignado con las alertas que se dispararon hasta el año de iniciado cada crédito⁹. Pero también se podría haber elegido otro criterio diferente, como considerar el riesgo relativo dinámico hasta el segundo año o bien hasta cada fecha de censura. Por estos motivos, los análisis que se realizan aquí tienen un motivo meramente comparativo y no serán definitivos para medir la performance real del Modelo Dinámico.

6.6.1. Comparación con tasas de default reales

Se observó a cada individuo desde el inicio de su crédito hasta 12 meses después, y se otorgó el puntaje total observado a este último momento, ignorando todos los momentos intermedios, pero teniendo en cuenta la evolución del indicador sistémico que ha sido construido de modo acumulativo.

6.6.1.1. Base de Desarrollo

Se dividió a la base de desarrollo en 15 grupos de menor a mayor riesgo relativo dinámico hasta el mes 12 para cada cliente, y en base a dicho análisis se construyó el Gráfico 23. Se puede observar que el riesgo relativo dinámico **ordena** a los grupos en categorías de riesgo diferenciadas para el análisis de tasas de default a un plazo 12 meses, a excepción de algunas categorías intermedias como en la cuarta, quinta y octava. El nivel de predictibilidad no es tan bueno ya que hay desviaciones considerables entre lo predicho y lo real.

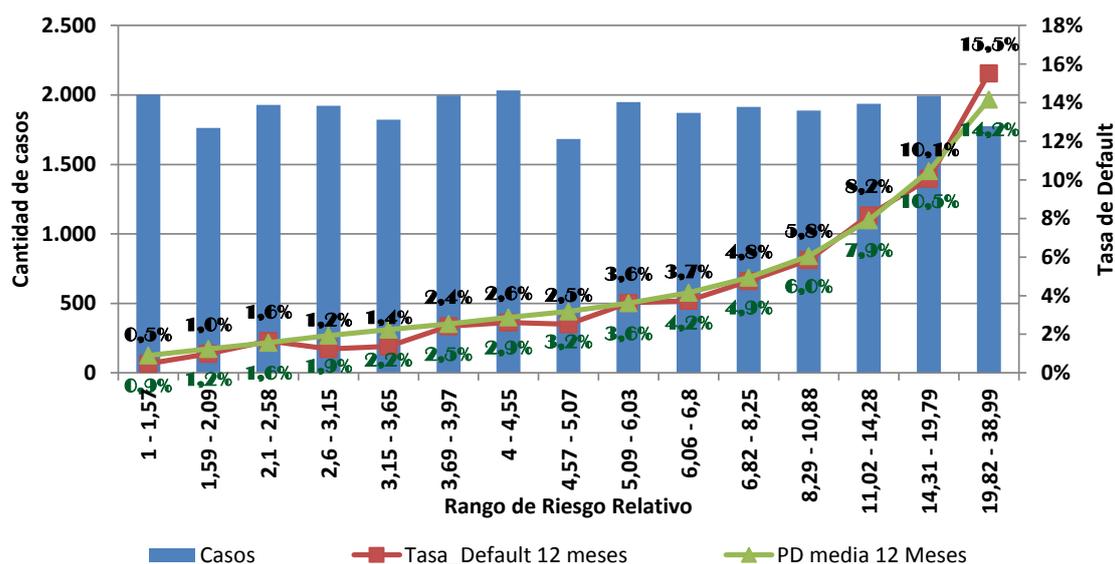


Gráfico 23: Análisis de ordenamiento y Predicción del Modelo de Supervivencia Dinámico en Corto Plazo (Elaboración Propia)

⁹ Notar que al considerar el riesgo relativo dinámico hasta los doce meses, no se activarán los indicadores “alerta_18” y “alerta_30”.

Cabe mencionar que la PD graficada corresponde al plazo de un año para todos los clientes, ignorando el momento de default del mismo. Es decir, como se mencionó anteriormente, se está forzando al Modelo de Supervivencia Dinámico a un plazo de 12 meses con el simple objeto de hacerlo comparable al Modelo Logístico.

El Gráfico 24 analiza la tasa de default luego de los 12 meses. Es destacable el hecho de que el modelo sigue sirviendo para explicar el ordenamiento del comportamiento de los clientes a largo plazo, aunque para algunas categorías no se cumple crecimiento estricto de la tasa de mora a medida que disminuye el riesgo relativo dinámico asignado.

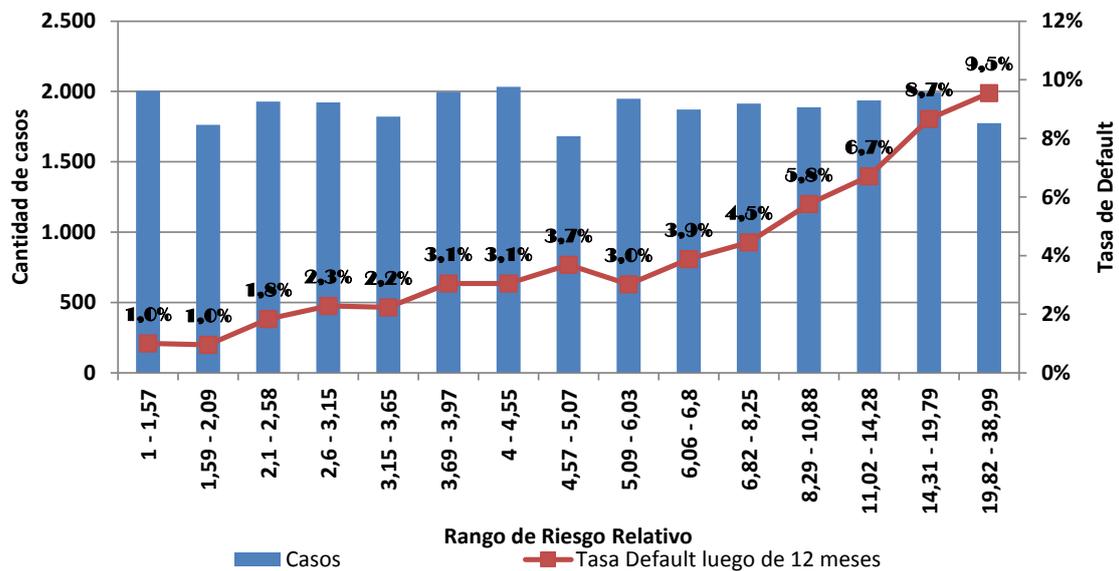


Gráfico 24: Análisis de ordenamiento y Predicción del Modelo de Supervivencia Dinámico en Largo Plazo (Elaboración Propia)

6.6.1.2. Base de Validación

Se dividió a la base de validación en 8 grupos de riesgo relativo dinámico hasta los 12 meses, y se analizó la tasa de default realizada y la probabilidad de default estimada con el modelo en cada grupo (ver Gráfico 25). Se puede observar que el riesgo relativo dinámico **ordena** a los grupos en categorías de riesgo diferenciadas también en la base de validación. Además, se puede observar que el modelo también predice aceptablemente mediante la PD estimada a 12 meses, cabiendo las mismas observaciones realizadas para la base de desarrollo.

En el Gráfico 26 se analiza la tasa de default a largo plazo, es decir, la tasa de default que contempla los clientes que se vuelven insolventes luego de los 12 meses de iniciado el crédito. Es destacable el hecho de que el modelo sigue sirviendo para explicar un ordenamiento del comportamiento de los clientes a largo plazo, aunque para la cuarta categoría no se cumple crecimiento estricto de la tasa de mora a medida que aumenta el riesgo relativo dinámico asignado.

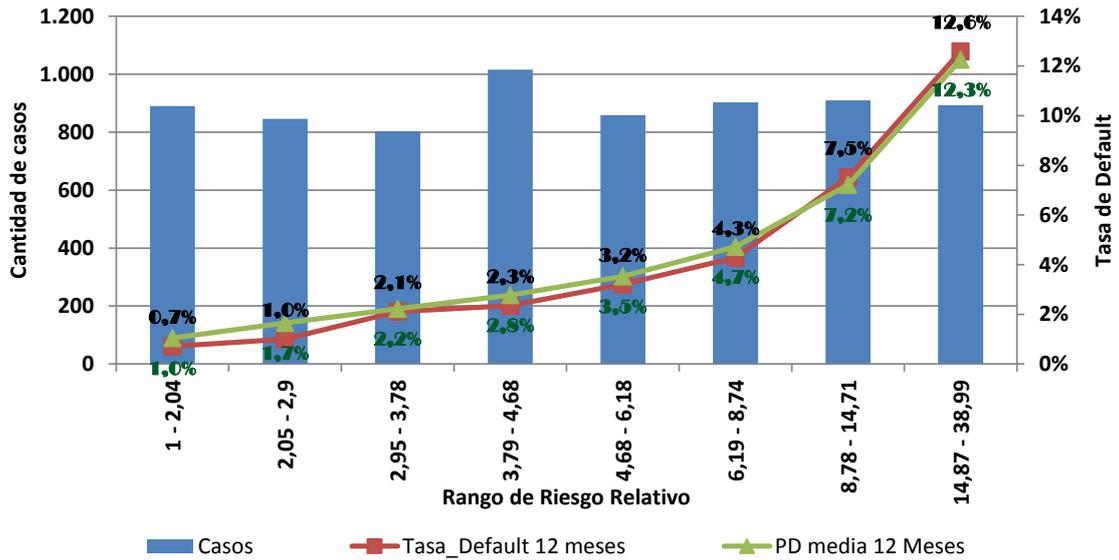


Gráfico 25: Análisis de ordenamiento y Predicción del modelo Supervivencia Dinámico – Validación (Elaboración Propia)

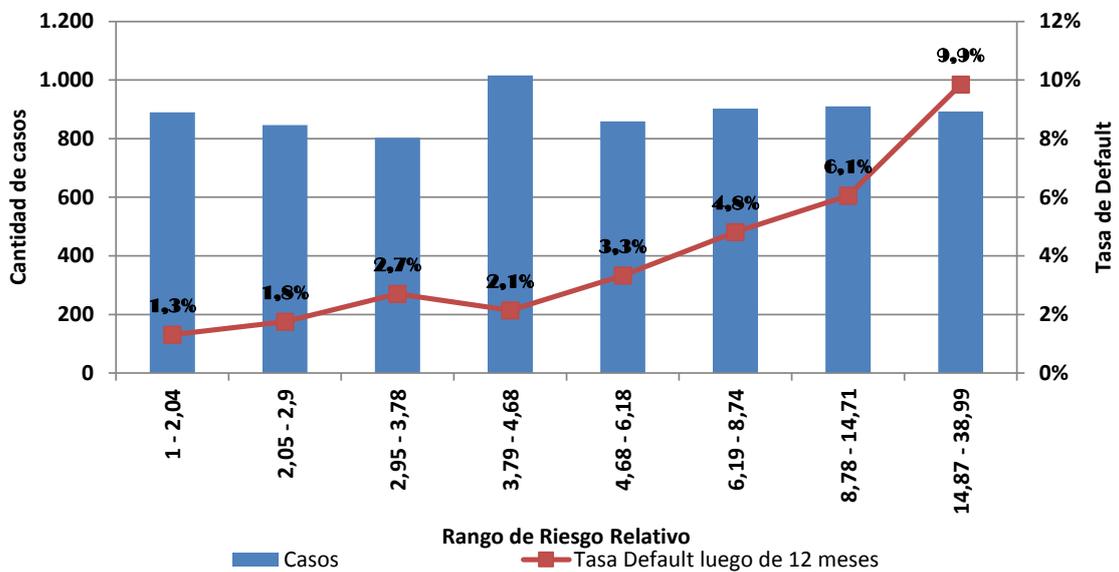


Gráfico 26: Análisis de orden y Predicción del modelo Supervivencia Dinámico en Largo Plazo - Validación (Elaboración Propia)

6.6.2. Área ROC y Power Stat

En el Gráfico 27 se muestran las curvas ROC para el Modelo de Supervivencia Dinámico comparativamente con la curva del Modelo Logístico, considerando el default a 12 meses.

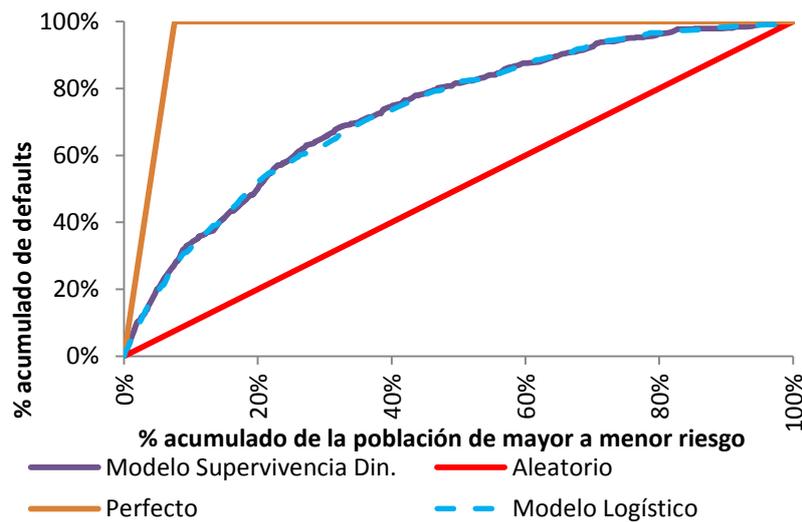


Gráfico 27: Curva del modelo – Defaults a corto plazo – Modelo Supervivencia Dinámico – Validación (Elaboración Propia)

El Gráfico 28 contiene las curvas considerando el default con posterioridad a 12 meses. Para la construcción de la curva del Modelo Logístico se consideró el score directo como ordenador para los defaults que ocurrieron con posterioridad a los 12 meses.

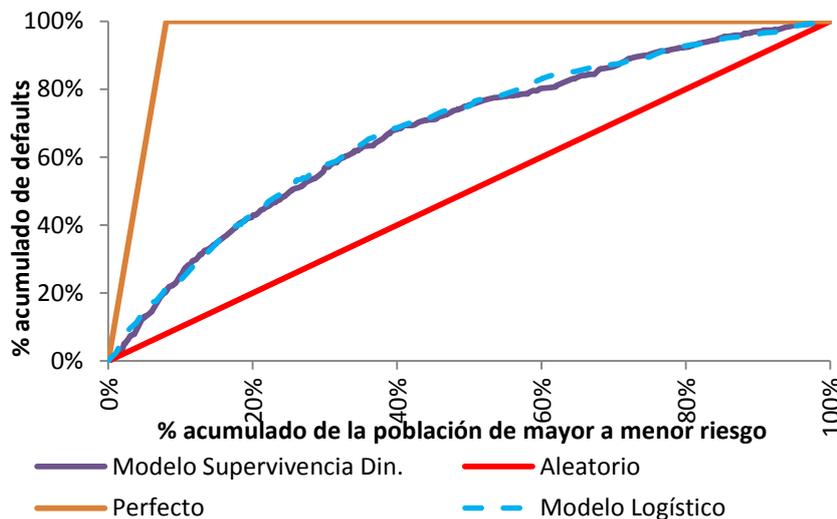


Gráfico 28: Curva del modelo – Defaults a largo plazo – Modelo Supervivencia Dinámico - Validación (Elaboración Propia)

La tabla 7 contiene las medidas de performance en forma comparativa.

	Corto plazo			Largo Plazo			Total		
	Supervivencia Dinámico	Supervivencia Estático	Logístico	Supervivencia Dinámico	Supervivencia Estático	Logístico	Supervivencia Dinámico	Supervivencia Estático	Logístico
Área ROC	75,39%	75,22%	75,17%	68,95%	69,82%	69,61%	74,31%	74,67%	74,54%
Power Stat	50,77%	50,43%	50,35%	37,89%	39,64%	39,21%	48,61%	49,35%	49,07%

Tabla 7: Medidas de performance comparadas (Elaboración propia)

Como se puede apreciar en la Tabla 7, las medidas de performance del Modelo de Supervivencia Dinámico para el corto plazo son muy similares e incluso supera marginalmente a los modelos anteriores, mientras que para el largo plazo están por debajo. Los motivos de esto tienen que ver con que no se siguió a los clientes con posterioridad a los 12 meses y también con la limitación del modelo para generar mejoras en el riesgo con la permanencia de los individuos sin incurrir en default. El análisis considerando los defaults a ambos plazos se aprecia en la parte “Total” de la Tabla 7, e indica que el Modelo Dinámico evaluado con este criterio queda un poco rezagado respecto de los otros modelos.

6.6.3. Limitaciones del análisis comparativo de performance

Hasta aquí se han realizado análisis comparativos para poder determinar si los Modelos de Supervivencia propuestos pueden competir con el Modelo Logístico usual, y se mostró que el análisis de supervivencia agrega información valiosa al considerar el estudio del fenómeno crediticio. Esto se ve claramente plasmado en las medidas de performance analizadas: todas indican que el Modelo de Supervivencia Estático compite muy bien e incluso supera al Modelo Logístico. Sin embargo, los mismos indicadores también señalan que el Modelo de Supervivencia Dinámico tiene un rezago respecto de los otros dos modelos. Esto se debe en parte a que los indicadores de performance están contruidos para hacer comparables los tres modelos, y en particular están obligando a considerar solo hasta los 12 meses la posibilidad de cambiar las estimaciones con posterioridad al otorgamiento ofrecida por el Modelo Dinámico.

6.7. Valuación económica de la performance

Para agregar más claridad al responder si un Modelo de Supervivencia es mejor que otro, es necesario también valorar económicamente las predicciones que arrojan.

Para ello, se utilizó el riesgo relativo dinámico como ordenador de los clientes y se consideró más exigente pedir un riesgo relativo dinámico bajo para aceptar un cliente. También se tomó el riesgo relativo dinámico inferido para los primeros 12 meses de vigencia del crédito como estimador.

El Gráfico 29 muestra en forma comparativa los beneficios para distintos niveles de exigencia (de mayor a menor exigencia), realizado sobre la base de validación del modelo. Se observa que los tres modelos compiten equitativamente en generar buenos beneficios, y parece apreciarse que el máximo del Modelo de Supervivencia Dinámico es ligeramente superior al del resto de los modelos. El punto de exigencia óptimo para el caso del Modelo de Supervivencia Dinámico es de 5.316 veces el beneficio anual por margen y cargos, superando a los 5.168 del modelo estático y a los 5.080 del Modelo Logístico en 2,9% y 4,6% respectivamente. Cabe preguntarse si estos mejores beneficios se dieron porque el Modelo de Supervivencia Dinámico es superador de los otros dos o bien puede haberse dado por casualidad. Por ello en las siguientes secciones se realizan algunos análisis de sensibilidad de dicho resultado.

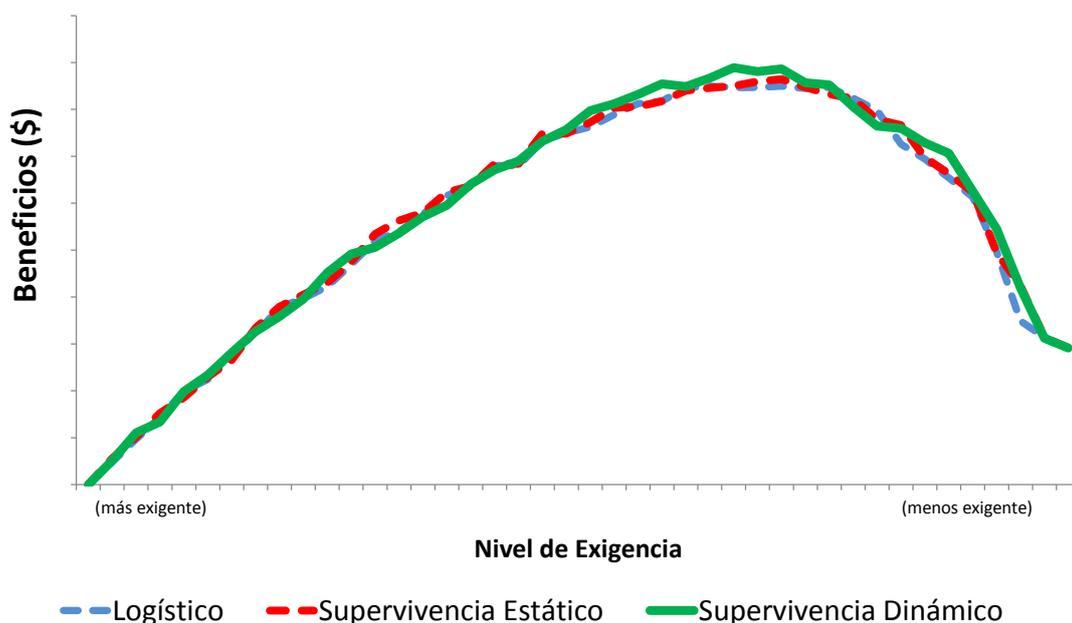


Gráfico 29: Gráfico de beneficios monetarios – Modelo de Supervivencia Dinámico (Elaboración Propia)

6.7.1. Sensibilidad del Máximo Beneficio respecto de cambios en el nivel de pérdida

Se realizó un análisis de sensibilidad considerando un ratio de pérdida en caso de default sobre beneficios anuales por margen y cargos de 13 y 9, en comparación con el ratio de 11 que se utilizó en los análisis previos. Los resultados indican lo siguiente:

- Cuando el ratio es de 13 (es decir, los clientes malos hacen perder más dinero), el Modelo de Supervivencia Dinámico se ubica en beneficios máximos un 0,9% por debajo del Modelo de Supervivencia Estático pero un 3,1% por encima del Modelo Logístico.
- Cuando el ratio es de 9 (es decir, los clientes malos hacen perder menos dinero), el Modelo de Supervivencia Dinámico se ubica en beneficios máximos un 0,4% por debajo del Modelo de Supervivencia Estático pero un 1,7% por encima del Modelo Logístico.

Se puede apreciar entonces que puede considerarse la superioridad del Modelo Dinámico sobre el Modelo Logístico, y un cierto emparejamiento de beneficios en el caso de ambos Modelos de Supervivencia.

6.7.2. Distribución de la diferencia porcentual en beneficios

Para comparar la performance del Modelo de Supervivencia Dinámico con los otros dos modelos analizados, se recalcularon los beneficios óptimos de cada modelo en diferentes muestras de tamaño equivalente a la base de validación, utilizando la técnica de *bootstrapping*. Se generaron re-muestréos de 7.120 casos con reemplazo. Luego se computó la diferencia porcentual en los beneficios máximos entre el Modelo Dinámico y el Modelo Estático por un lado, y entre el Modelo Dinámico y el Modelo Logístico por el otro.

En el Gráfico 30 se pueden apreciar las distribuciones obtenidas.

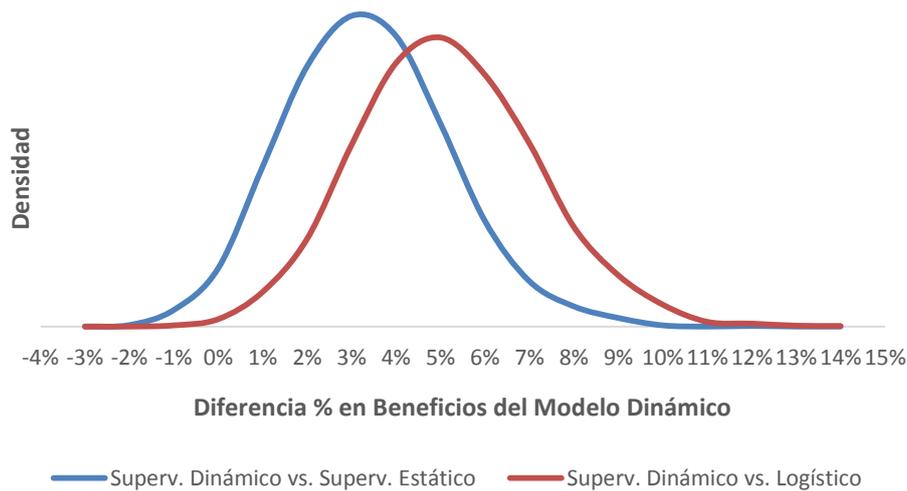


Gráfico 30: Distribución de la diferencia porcentual en beneficios (Elaboración Propia)

La variable de porcentaje adicional de beneficios con el Modelo de Supervivencia Dinámico respecto del Modelo Logístico es positiva casi en el 99% de los casos, y tiene un media de 5%. Si se lo compara con el Modelo de Supervivencia Estático, el porcentaje adicional de beneficios es positivo en el 96% de los casos, con una media de 3%. Es decir, el diferencial porcentual de los beneficios económicos que se menciona en la Sección 6.7. se da con alta probabilidad.

En el Gráfico 31 se puede apreciar la distribución de la frecuencia acumulada de los beneficios económicos en los tres modelos. Se puede ver claramente que la curva de frecuencias acumuladas del Modelo Dinámico está por debajo de la del Modelo Estático, y esta última a su vez por debajo del Modelo Logístico. Esto significa que para un nivel de Beneficio Económico dado, el Modelo Dinámico tiene mayor probabilidad de tener beneficios más altos que el resto de los modelos.

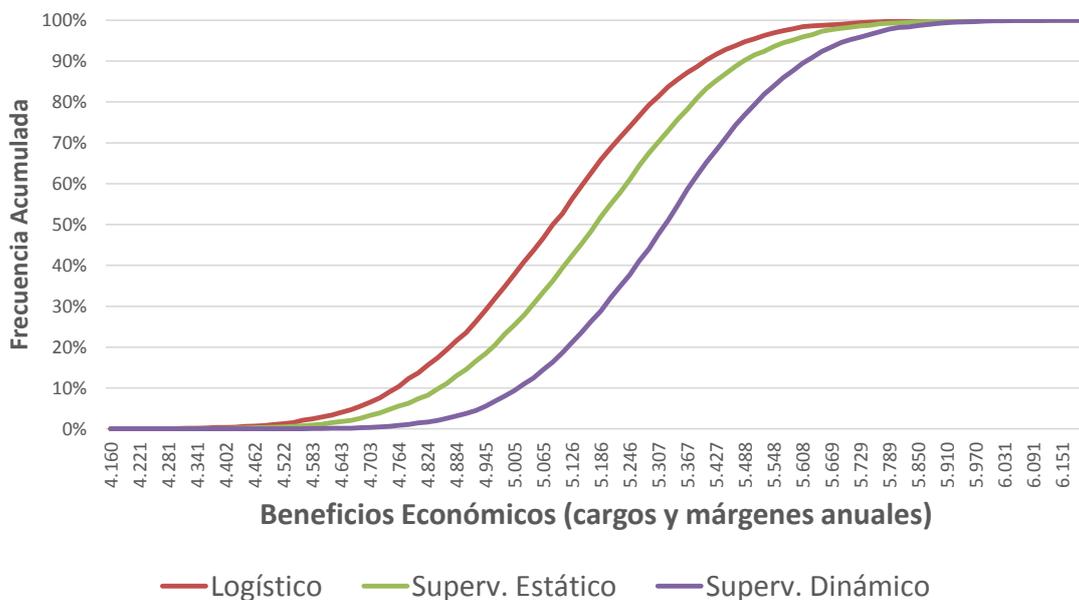


Gráfico 31: Distribución de la diferencia porcentual en beneficios (Elaboración Propia)

6.8. Sensibilidad de la PD al contexto sistémico

A continuación se analiza cómo cambiarían las estimaciones de la probabilidad de default a 12 meses, ante escenarios más o menos severos de mora sistémica.

Se analizó por un lado la frecuencia con que se observaron distintas situaciones de deterioro, y se evaluó toda la base de validación como clientes que entraban a la Entidad en el mismo momento. En la Tabla 8 se muestran los resultados.

La primera columna de la Tabla 8 representa una clave que describe al contexto, mientras que las siguientes tres columnas representan cuáles alertas se dispararían en cada uno de estos contextos. La quinta columna surge de aplicar los coeficientes de riesgo relativo para las alertas disparadas. Se puede observar que los escenarios están ordenados de menor a mayor grado de deterioro. La sexta columna representa la frecuencia con que se dio cada contexto en la base que corresponde a los meses desde Enero 2010 a Enero 2013. La séptima columna representa el promedio de las probabilidades de default para la base de validación, calculado según el Modelo Dinámico. La última columna (que es la más importante) indicaría que si se espera un deterioro en el contexto sistémico, debería ser necesario incrementar la probabilidad de default multiplicándola por un factor de 1,17 en el caso menos severo hasta un factor de 1,71 en el caso más severo. En los contextos de deterioro intermedios se guarda una relación coherente y lógica entre el riesgo relativo dinámico asignado y el factor de incremento en la PD.

Contexto primer año	alerta_3	alerta_6	alerta_12	Riesgo Relativo dinámico	Frecuencia del Contexto	PD Media	Factor Incremento
000	0	0	0	0,00000	51,4%	3,82%	
001	0	0	1	0,17258	16,2%	4,49%	1,17
011	0	1	1	0,28081	8,1%	4,96%	1,30
100	1	0	0	0,30560	10,8%	5,08%	1,33
110	1	1	0	0,41383	8,1%	5,60%	1,47
111	1	1	1	0,58641	5,4%	6,55%	1,71

Tabla 8: PD media por contexto (Elaboración propia)

7. Conclusiones y futuras líneas de investigación

En este trabajo se mostró que los Modelos de Supervivencia compiten eficientemente con el Modelo Logístico para la estimación de la Probabilidad de Default a 12 meses. Además, se pudo comprobar que los Modelos de Supervivencia pueden agregar valor en la maximización de los beneficios económicos de la cartera. Esta adición de valor no se debe a la utilización de más información para la toma de decisiones, sino a un cambio cualitativo en el modelo, ya que los Modelos de Supervivencia no hacen énfasis en un plazo particular para el modelado sino en toda la vida del crédito.

A su vez, cuando al Modelo de Supervivencia se le agrega un ajuste basado en la mora sistémica, creando de esta manera un Modelo de Supervivencia Dinámico, se observan dos cuestiones. La primera es que la performance estadística es marginalmente menor a los otros modelos. Esta merma en la performance puede explicarse en parte porque se está midiendo al Modelo Dinámico con las mismas herramientas de medición que al resto de los modelos, y estas herramientas no contemplan la posibilidad de cambios en las estimaciones del riesgo a diferentes plazos. Por otro lado, los beneficios económicos del Modelo Dinámico pueden ser superiores al de los otros modelos. Esto se debe a que hay una estimación del impacto futuro que el contexto sistémico tendrá en la cartera luego de otorgado el crédito. Si dicho indicador muestra que las condiciones sistémicas empeoran, se ajustan las estimaciones de Probabilidad de Default.

Según el Modelo Dinámico propuesto, al momento del otorgamiento del crédito el indicador debe considerar dos componentes: i) La tendencia actual en la mora sistémica, ya que al momento de otorgarse el crédito se conoce la evolución de la mora sistémica de los últimos meses y el indicador considera promedios de variaciones trimestrales desfasados; y ii) Una proyección de los indicadores futuros que luego irán ajustándose a medida que se conozca la evolución de la mora sistémica.

El Modelo de Supervivencia Dinámico permitió hacer ajustes de alrededor de un 70% a la probabilidad de default media de la cartera en función del contexto sistémico actual y proyectado. Además, los beneficios económicos del Modelo de Supervivencia Dinámico pueden ser superiores del Modelo Estático y del Modelo Logístico con un elevado nivel de confianza. Sin embargo, la dominancia de beneficios económicos del Modelo Dinámico es sensible a cambios en el nivel de pérdida ante default (LGD), y según este nivel de LGD los Modelos de Supervivencia analizados podrían emparejarse en su nivel de beneficios. Por lo tanto, no puede decirse que el Modelo Dinámico propuesto sea claramente superior al resto de los modelos, sino que puede competir eficientemente con ellos brindando la posibilidad de hacer estimaciones sensibles al escenario sistémico. También debería considerarse dentro de los costos el relacionado con la proyección del indicador sistémico y la validación y seguimiento de los Modelos de Supervivencia Dinámicos que serán más complejos que los Modelos Logísticos usuales.

Adicionalmente, existen ciertas limitaciones en los Modelos de Supervivencia utilizados que podrían superarse realizando cambios en el modelado.

En primer lugar se encuentra la forma de impacto del indicador sistémico. Según la Modelización Exponencial de Riesgos Proporcionales de Cox, un cambio en las condiciones sistémicas tiene impacto permanente en el *hazard*. Sin embargo, según los análisis bivariados

del indicador sistémico, algunas veces se puede dar un impacto inicial y luego el mismo se diluye en el tiempo con la permanencia de los individuos. Cabe preguntarse, entonces, si se puede incluir un impacto escalonado o suavizado. Surge también la factibilidad de considerar impactos o shocks positivos que mejoren la estimación de la Probabilidad de Default de un cliente (en el modelo presentado, los indicadores de alerta solamente permiten ajustar hacia arriba la PD, i.e. aumentar el riesgo). Para ello, es posible que el modelo exponencial se deba reemplazar por una distribución Weibull, Log-normal, o bien alguna otra que permita diferentes intensidades de impacto a través del tiempo, incluyendo indicadores de mejora en el *hazard*.

En segundo lugar, debe considerarse la situación en la cual el individuo deja de ser cliente de la Entidad, sin incurrir en default. En el modelado que se presentó, la variable explicada es el tiempo hasta el default. También se podría haber modelado el tiempo que transcurre hasta que el cliente abandona la Entidad pero sin incurrir en default. Si bien esta última situación no genera pérdidas crediticias a la Entidad, genera la pérdida de la rentabilidad de dichos clientes (costos de oportunidad). Si dicha situación se da mucho más frecuentemente que los defaults, puede ser muy perjudicial para los beneficios económicos de la Entidad. La cuantificación de las tasas de abandono permitiría establecer estrategias de comercialización para captar a los buenos pagadores que tienen más chances de continuar siendo clientes de la Entidad en una relación de largo plazo. El análisis de las tasas de abandono en conjunto con las tasas de default se podría realizar mediante un modelo de respuestas múltiples.

En tercer lugar, el Modelo Dinámico utilizado incluyó solo un indicador sistémico y descartó una variable que utilizaban el resto de los modelos. Esto fue coherente con el objetivo de no generar un beneficio por tomar mayor información que el resto de los modelos ya que todos los modelos utilizados contienen la misma cantidad de variables explicativas. Se espera que un Modelo Dinámico con más indicadores sistémicos contribuya a mejorar la performance tanto estadística como económica.

Finalmente deberían incorporarse shocks positivos de permanencia. Esto consiste en otorgar un beneficio en términos de riesgo a aquellos individuos que permanecen sanos más de un cierto tiempo, disminuyendo su *hazard*. Por ejemplo, un individuo que pudo superar un año de turbulencias en el sistema sin incurrir en default tiene un comportamiento de buen pagador y debería ser premiado por esa permanencia aunque sus variables de originación y el indicador sistémico no lo hayan beneficiado en el pasado. La cuantificación de este premio por permanencia no puede ser captada en el modelo propuesto en este trabajo, a menos que se incorpore la variable de permanencia como explicativa, lo cual generaría un bucle ya que esta variable sería explicativa y explicada a la vez. Este problema generalmente es mitigado por las Entidades Financieras con la utilización de modelos de comportamiento que hacen un seguimiento (usualmente luego del año de otorgado un crédito) de los clientes para cuantificar su riesgo y, adicionalmente, ofrecerles productos crediticios a aquéllos que tengan mejores puntajes.

Se espera que superar o mitigar algunas de estas limitaciones en desarrollos posteriores contribuirá a mejorar tanto la performance estadística como económica en los Modelos de Supervivencia Dinámicos.

Referencias

- Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance* 21, 1721-1742.
- Bellotti, T., & Crook, J. (2007). *Credit Scoring With Macroeconomic Variables Using Survival Analysis*. Management School and Economics, University of Edinburgh.
- Bessis, J. (2002). *Risk management in banking*. New York: John Wiley & Sons, Ltd.
- Bluhm, C., Overbeck, L., & Wagner, C. (2003). *An introduction to Credit Risk Modelling*. Munich: Capman & Hall.
- Cao, R., Vilar, J. M., & Devia, A. (2009). *Modelling consumer credit risk via survival*. Universidade da Coruña.
- Cox, D. R. (1970). *Analysis of Binary Data*. London: Methuen.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. Oxford: Chapman & Hall.
- Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*.
- Duffie, D., & Schaefer, S. (2003). *Credit Risk: Pricing, Measurement, and Management*. New Jersey: Princeton University Press.
- Engelmann, B., & Rauhmeier, R. (2006). *The Basel II parameters: Estimation, validation and Stress Testing*. Berlin: Springer.
- Englemann, B., Hayden, E., & Tasche, D. (2003). *Measuring the Discriminative Power of Rating Systems*. Frankfurt am Main: Deutsche Bundesbank.
- Fox, J. (2002). *Cox Proportional-Hazards Regression for Survival Data*.
- Gourieroux, C., & Jasiak, J. (2007). *The Econometrics of Individual Risk*. New Jersey: Princeton University Press.
- Löffler, G., & Posch, P. N. (2007). *Credit risk modeling using Excel and VBA*. West Sussex, England: John Wiley & Sons Ltd.
- Orlando J. Ferreres & Asociados . (2014). *Índice General de Actividad IGA*. Buenos Aires.
- Palmer Pol, A. L. (1993). Modelo de Regresión de Cox: Ejemplo numérico del proceso de estimación de parámetros. *Psicothema*, 5(2), 387-402.
- Pimbley, J. M. (2012). *The Hazard Rate Matrix Approach to Credit Transitions*. Risk Professional.
- Roszbach, K. (2003). *Bank Lending Policy, Credit Scoring and the Survival of loans*. Central Bank of Sweden, Stockholm: Sveriges Riksbank Working Paper Series.
- SAS Institute Inc. (1999). *SAS/STAT® User's Guide*. Cary, NC, USA: Version 8.
- Sauders, A., & Allen, L. (2002). *Credit Risk Measurement*. New York: John Wiley & Sons, Inc.
- Scott Long, J., & Fresee, J. (2001). *Regression models for categorical data dependent variables using Stata*. Texas: Stata Press.

- Smith, T., & Smith, B. (s.f.). *Survival Analysis And The Application Of Cox's Proportional Hazards Modeling Using SAS*. Naval Health Research Center, San Diego, CA.
- Therneau, T. M. (1996). Extending the Cox Model. *Technical Report Series of Mayo Clinic, Minnesota*.
- Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 57, 167-79.