

Balzarotti, Verónica; Castelpoggi, Fernando

Working Paper

Modelos de puntuación crediticia: La falta de información y el uso de datos de una central de riesgos

Working Paper, No. 2008/37

Provided in Cooperation with:

Economic Research Department (ie), Central Bank of Argentina

Suggested Citation: Balzarotti, Verónica; Castelpoggi, Fernando (2008) : Modelos de puntuación crediticia: La falta de información y el uso de datos de una central de riesgos, Working Paper, No. 2008/37, Banco Central de la República Argentina (BCRA), Investigaciones Económicas (ie), Buenos Aires

This Version is available at:

<https://hdl.handle.net/10419/86131>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Modelos de puntuación crediticia: la falta de información y el uso de datos de una central de riesgos

Verónica Balzarotti / Fernando Castelpoggi
BCRA

Agosto de 2008



ie | BCRA
INVESTIGACIONES ECONÓMICAS

Banco Central de la República Argentina
ie | Investigaciones Económicas

Agosto, 2008
ISSN 1850-3977
Edición Electrónica

Reconquista 266, C1003ABF
C.A. de Buenos Aires, Argentina
Tel: (5411) 4348-3719/21
Fax: (5411) 4000-1257
Email: investig@bcra.gov.ar
Pag.Web: www.bcra.gov.ar

Las opiniones vertidas en este trabajo son exclusiva responsabilidad de los autores y no reflejan necesariamente la posición del Banco Central de la República Argentina. La serie Documentos de Trabajo del BCRA está compuesta por material preliminar que se hace circular con el propósito de estimular el debate académico y recibir comentarios. Toda referencia que desee efectuarse a estos Documentos deberá contar con la autorización del o los autores.

Modelos de Puntuación Crediticia: la Falta de Información y el Uso de Datos de una Central de Riesgos

Verónica Balzarotti, Fernando Castelpoggi

Abstract

El principal objetivo de este trabajo es estudiar el problema introducido por la falta de información de comportamiento crediticio para algunos deudores en las bases utilizadas para desarrollar modelos de puntuación crediticia (*scoring*) y el uso de información de comportamiento contenida en una central de riesgos como una potencial solución. Se abordan dos problemáticas (i) la necesidad de proveer una estimación del riesgo crediticio de los deudores cuyo comportamiento se ignora (porque son dados de baja de las bases sin que se registre el motivo), y (ii) la estimación del impacto de no tener en cuenta la falta de estos datos en la evaluación del riesgo de las carteras de las entidades. La estrategia fundamental será utilizar el comportamiento crediticio de los deudores en otras entidades, registrados en la central de riesgo¹.

I. Introducción

Estudios previos han mostrado resultados interesantes, tanto para Argentina como para otros países, respecto del desarrollo de modelos de puntuación de calidad crediticia (*scoring*) para los deudores bancarios basándose en la información disponible en una central de información pública, a pesar de las limitaciones que impone este tipo de base². Dichos modelos pueden ser útiles como *benchmark* para comparar modelos internos, para evaluar el riesgo de las entidades o para tomar decisiones regulatorias³.

Una limitación de esas bases de datos que es de interés particular en este trabajo viene dada por la baja de la central de riesgos de un número significativo de deudores, en forma regular y generalizada entre entidades, sin que se registre los motivos de tal remoción. En consecuencia, es imposible saber con certeza si estos deudores fueron removidos porque cancelaron el total de sus deudas o porque no pagaron y el banco abandonó las acciones de cobro. Es decir, dos razones “opuestas”, entre otras menos representativas, pueden estar detrás de un deudor faltante. Si el número de deudores removidos es considerable y el proceso de exclusión no es aleatorio, puede introducirse un error en la evaluación del riesgo de la cartera según el modelo⁴.

¹ Este trabajo tiene como antecedente local el Documento de Trabajo de Balzarotti, Gutiérrez Girault y Vallés (2006). En este nuevo trabajo se aplican nuevas técnicas de inferencia, se trabaja con datos más recientes y se centra la atención en la incorporación del comportamiento en otras entidades crediticias.

² Por ejemplo Balzarotti, V., C. Castro y A. Powell (2004), Balzarotti, V., C. Castro y A. Powell (2002), Schechtman R, Salomão García et al (2004), Powell, Mylenko et al (2004).

³ En esa línea véase el trabajo de Carlos Trucharte Artigas (2004)

⁴ Una carencia típica de las bases de información de comportamiento crediticio deriva de que normalmente se registran los datos de deudores que han recibido crédito pero no los de los solicitantes rechazados. Un sistema de *scoring* sobre un conjunto de datos con esas características está sujeto a “truncamiento”, lo cual se conoce en la literatura específica como “sesgo de selección” (por ejemplo en Crook, J., 2002; Banasik, J. and Crook, J., 2004; Hand, D. J. and Henley, W. E., 1997; Verstraeten, G. and Van den Poel, D., 2004 y Greene, W., 1992). En este trabajo, ese sesgo en particular no es de mayor preocupación, en la medida en que el objetivo es estimar el riesgo de un portafolio bancario desde un punto de vista prudencial, en comparación con el caso en el cual el objetivo fuera desarrollar un sistema para la decisión crediticia de originación.

Para estudiar el problema y evaluar su impacto en el caso argentino, en primer lugar se intentará estimar el riesgo crediticio de los casos faltantes, poniendo especial atención en la mejor forma de incorporar los datos de comportamiento de los deudores en entidades diferentes de aquellas que los dan de baja. Se analizará luego el impacto de no tener en cuenta los faltantes en la evaluación de riesgo de las carteras de los bancos.

El problema de las bajas es bien frecuente en las bases de datos, no sólo en las de origen público. Este trabajo versará sobre el caso argentino y por ende los resultados dependerán fuertemente del proceso particular que genera las bajas en la base local, el cual podrá diferir en otras instituciones o países⁵. Se intentará poner en evidencia que los errores en la evaluación del riesgo pueden no ser despreciables, que las metodologías usuales de corrección son deficientes y costosas y que una mejor evaluación conlleva un costo y un nivel de imperfección tales como para hacer aconsejables dos acciones: (i) recomendar la modificación de las bases de datos de manera de descartar la posibilidad de que existan bajas no explicadas y (ii) prestar especial atención, por parte de los supervisores que estén desarrollando y/o supervisando modelos de riesgo crediticio, respecto de esta problemática.

En lo que sigue, la sección 2 resume el desarrollo de los modelos de scoring y la estimación de la probabilidad de default; la sección 3 resume la literatura sobre datos faltantes y sesgo de selección. La sección 4 describe la central de riesgos pública en Argentina y las definiciones que se utilizan en el trabajo (por ejemplo, la condición de default) y contiene un análisis a priori de los deudores removidos. La sección 5 describe las técnicas de estimación de distintos modelos ensayados mientras que la sección 6 presenta y compara sus resultados. Finalmente, la sección 7 concluye.

2. Modelos de Scoring Crediticio y Probabilidades de Default (PD)

Una variedad de modelos estadísticos de puntaje o “score” se han desarrollado para apoyar la decisión crediticia, desde larga data. Estos modelos pueden extenderse para estimar probabilidades de default de los deudores. Estas “PDs” son requeridas en el Marco Revisado de Convergencia Internacional de Medición y Estándares de Capital (Basilea II)⁶ para armar sistemas de rating y calcular capitales mínimos dentro del enfoque de Ratings Internos (IRB). Por este motivo los modelos de scoring crediticio han recibido mucha atención recientemente.

Existe un conjunto de métodos provenientes de diversas disciplinas como la estadística, el aprendizaje automático de patrones o inteligencia artificial, usualmente utilizados para construir scorings o reglas de clasificación crediticia: análisis discriminante, redes neuronales, árboles de decisión y métodos estadísticos tradicionales, como las regresiones Logit y Probit, los análisis de supervivencia, entre otros. Una cantidad considerable de estudios que comparan la performance de estos modelos concluyen que sus resultados son relativamente similares en este campo. En este trabajo no apuntamos a discutir los pros y contras de los distintos modelos, sino que preferimos concentrarnos en el análisis de la problemática de los deudores dados de baja y cómo aprovechar los datos de comportamiento de central de riesgos. Se aplicarán diferentes versiones de modelos logit, en los que la variable explicada es binaria e indica el default del deudor. Hemos seleccionado el modelo logit por su simplicidad y porque la influencia de las variables explicativas sobre la variable explicada puede obtenerse fácilmente. Adicionalmente, tiene la

⁵ Las características precisas del proceso que genera las exclusiones deben obtenerse a partir de una revisión de las definiciones, procesos y sistemas involucrados en cada entidad y de la normativa y procesos del BCRA.

⁶ BCBS (2004), actualizado a junio 2006.

ventaja (en forma similar a los modelos probit) de que los scores pueden ser trasladados fácilmente a las PDs de los deudores⁷.

El modelo logit se puede representar en términos de una variable observada binaria y una variable latente, de la siguiente manera⁸:

$$D_i = \begin{cases} 1 & \text{si } X_i' \beta + \varepsilon_i > 0 \\ 0 & \text{si } X_i' \beta + \varepsilon_i \leq 0 \end{cases}$$

Donde D_i (la variable de comportamiento observado) es igual a 1 cuando hay default y 0 en otro caso, X_i' es la matriz de variables explicativas de cada deudor i , β es el vector de parámetros a estimar y ε_i es el error de la regresión, en este caso con distribución logística. Los $X_i' \beta_i$ representan los scores de los deudores. Usualmente se estiman los parámetros maximizando la función de verosimilitud dada por⁹:

$$\text{Log}(L) = \sum_K y_K \ln(p_K) + (1 - y_K) \ln(1 - p_K) = \sum_{\{y=1\}}^{N_1} \ln(p_K) \sum_{\{y=0\}}^{N_0} \ln(1 - p_K)$$

donde $p_K = \frac{1}{1 + \exp(-X' \beta)}$, y N_1 y N_0 son la cantidad de defaults y no defaults respectivamente.

Las PD pueden calcularse a partir del score, mediante:

$$P(D_i = 1 / X_i) = 1 / (1 + \exp(X_i' \beta)) \quad (1)$$

Vale aclarar que, en el marco de Basilea II, la PD asociada a cada grado de un sistema de rating debería ser un promedio de largo plazo de las tasas de default anuales dentro de cada segmento, o de las PDs individuales de los deudores asignados a ese grado (aunque también pueden “mapearse” los ratings del banco a un rating externo y tomar sus PDs). En ese documento, se distinguen PDs estresadas y no estresadas y sistemas de rating Punto-en-el-Tiempo (conocidos en la literatura por sus siglas en inglés: *PIT*, por *Point-in-Time*) y A-Través-Del-Ciclo (*Through-the-cycle* o *TTC*). En este trabajo estimaremos PDs anuales no estresadas, las cuales son un input en la construcción de sistemas de rating y constituyen un punto de partida para estudios posteriores de PDs estresadas y sistemas de rating *TTC*. Los niveles de riesgo que resulten en el trabajo estarán probablemente por debajo de un promedio de largo plazo, por basarse en datos de un período de fuerte expansión económica.

3. Scoring, inferencia de rechazos y sesgo de selección. Repaso de la literatura.

La falta de información sobre el comportamiento de los casos cuyas solicitudes de crédito fueron rechazadas conduce usualmente a un sesgo en la muestra y, consecuentemente, en la predicción.

⁷ En Basilea II, se subraya la importancia de que se comprenda el uso de los modelos internos de rating como metodología para otorgar crédito. Los bancos deberían usar modelos simples y abarcativos para estimar las PDs, de forma tal que las personas en los departamentos comerciales puedan entender cómo trabajan estos modelos y cuál es la influencia de las variables explicativas sobre las PDs. Lo mismo sería aplicable al supervisor.

⁸ Véase por ejemplo, Johnston y Di Nardo (1997).

⁹ En esta función se asume la independencia entre los comportamientos de los deudores.

La literatura respecto de cómo inferir la información de los rechazos para evitar o corregir ese sesgo es amplia y se extiende por más de 30 años¹⁰. Como se explicará en el Recuadro 1 de la sección 5, el caso que nos ocupa no es estrictamente de sesgo de selección; sin embargo, se tomó esa literatura como antecedente para analizar el problema de los faltantes (véase el Anexo 1).

En general, existe consenso respecto a la gravedad de no corregir el sesgo de selección por la decisión de aceptación /rechazo. No obstante, parece no existir aún un procedimiento totalmente satisfactorio para controlar apropiadamente el problema. “Los procedimientos de inferencia de rechazos corrigen el sesgo menos que lo esperado. Dada la pérdida sustancial de información que se produce cuando los solicitantes son rechazados porque no cumplen con los umbrales de riesgo, un modelo confiable basado en inferencia estadística puede bien ser imposible”¹¹. La importancia de la corrección se incrementa cuando la proporción de rechazados es mayor y, adicionalmente, cuando aumenta la tasa de rechazados en términos de la tasa de default. No se ha establecido bajo qué condiciones puede considerarse no-significativo el problema.

Diferentes estudios muestran que los métodos en general son problemáticos por basarse en supuestos fuertes, que la inferencia nunca es suficientemente confiable y que el único enfoque robusto es aceptar un número de solicitantes que serían rechazados para observar su comportamiento. Especialmente, las técnicas tradicionales, como el procedimiento de *listwise deletion* y la corrección de Heckman, fueron analizados y su uso desalentado. En cambio, la literatura parecería poder concluir con cierto grado de confianza que los procedimientos de imputación tiene ventajas sobre esas técnicas. En especial las técnicas de imputación múltiple y en algunos casos con ponderadores fraccionales, parecen tener ciertas ventajas, aunque no estén totalmente libres de críticas. Últimamente, se ha subrayado la importancia de que las técnicas de imputación y/o ponderación hagan el mejor uso posible de los datos disponibles, incluyendo variables adicionales. Al usar estas variables debe tenerse en cuenta el grado de afinidad entre el registro “receptor” del dato y los registros “donantes”. Estos antecedentes en particular son relevantes para las técnicas que se usarán en este trabajo. Más detalles se dan en Anexo 1.

Si aún conociendo sus limitaciones se aplican métodos de inferencia de rechazo, surge la cuestión de cómo evaluarlos. Esta área de estudio tampoco es completamente satisfactoria. Típicamente no es posible evaluar las ganancias en términos de reducción del sesgo y aumento del poder predictivo de las técnicas de inferencia basadas en el supuesto de homogeneidad entre los comportamientos observados y los no observados, ya que no es contrastable en casos reales. Estudiando los diferentes métodos, se deduce que el beneficio de la inferencia de rechazos depende de los datos, del diseño de la muestra, de la distribución de la población y del grado en que se satisfacen los supuestos de cada técnica. “Los métodos de inferencia de rechazo deben ser evaluados sobre una base caso por caso, y puede ser necesario su uso en forma combinada”¹².

4. La Central de Riesgos Pública en Argentina: la “Central de Deudores”

La “Central de Deudores del Sistema Financiero” (CD) es una base de datos desarrollada por el Banco Central de la República Argentina (BCRA), en la que se registran las deudas de individuos y personas jurídicas de las entidades financieras, las emisoras de tarjetas de crédito y los fideicomisos financieros, en tanto el saldo sea superior a 50 pesos. La información de cualquier deudor es accesible en forma gratuita en la página web del BCRA, indicando el

¹⁰ Una bibliografía no exhaustiva se puede encontrar en Carpenter (2006).

¹¹ Puntos destacados de la conferencia “Credit Risk Modeling and Decisioning”, FRB of Philadelphia, 2002

¹² Idem nota al pie n° 11.

número de identificación tributaria del deudor. Las entidades envían la información al BCRA mensualmente. El diseño actual de la CD data de 1997, cuando se fusionaron dos bases parciales.

Los registros de la CD comprenden préstamos, créditos por intermediación financiera, leasing y otros créditos, así como créditos contingentes (garantías, acuerdos de sobregiro en cuentas corrientes y otros acuerdos de márgenes crediticios). Se registra tanto información positiva (deudores que cumplen) como negativa (incumplimientos). Para cada deudor en cada entidad, la CD registra el número de identificación tributaria, el nombre, si es un individuo o una persona jurídica, el saldo en cada tipo de asistencia y cuánto está cubierto por garantías reconocidas en la regulación del BCRA, la tasa de interés, el plazo, la calificación (conocida habitualmente como “situación”) y las previsiones. No hay datos demográficos ni de días de atraso. Algunas variables son altamente confiables, especialmente aquellas que se publican, mientras que otras muestran baja calidad, como son los casos de plazo y tasa de interés¹³.

Esta base de datos se creó persiguiendo varios objetivos. Brevemente ellos son: (i) fomentar la accesibilidad al crédito, (ii) reforzar la supervisión, (iii) promover la competencia, (iv) reforzar la voluntad de pago y (v) proveer información para investigación económico-financiera, para decisiones de política macroeconómica y para regulación bancaria.

A mediados de 2007 la CD tenía información de 7,2 millones de individuos y 113 mil personas jurídicas, los cuales en conjunto representan 10,7 millones de registros de deudas (las personas pueden tener más de un registro de deuda). Considerando sólo los deudores del sector privado, el portafolio de consumo o vivienda (incluyendo PyMes) sumaba \$53 mil millones y la cartera comercial (definida en el párrafo anterior) \$54 mil millones en 30 mil registros.

La normativa del BCRA contempla dos criterios, o grupos de pautas, a efectos de la clasificación crediticia de las carteras de las entidades: (i) el primer criterio se aplica a la cartera de “consumo o vivienda” la cual abarca préstamos destinados a financiar el consumo, hipotecas sobre la vivienda y, a opción de la entidad, a créditos para actividades comerciales de hasta 500.000¹⁴ pesos, éstos últimos denominados “cartera comercial asimilable a consumo” y (ii) el criterio aplicable a la “cartera comercial”, que incluye al resto.

Los deudores deben tener una calificación de “1” a “5” otorgada por la entidad (donde 1 es la mejor)¹⁵. La calificación representa la probabilidad de que los términos contractuales del crédito sean honrados, sobre la base de una evaluación individual de la condición financiera futura. Se establecen porcentajes de previsión para cada grado, los que se reducen a la mitad para deudas garantizadas. Los deudores cuyas deudas están totalmente cubiertas por las garantías “preferidas A” (básicamente autoliquidables) no se califican y se informan a la CD en la mejor calificación. En la práctica, los créditos se originan con calificación 1 y sólo migran cuando son revisados.

Para analizar la situación financiera de deudores comerciales la normativa enumera un conjunto de pautas que incluyen la presentación de información contable a tiempo, las perspectivas de sector económico, atrasos, etc. Por lo menos una vez al año debe revisarse la calificación de los deudores comerciales de acuerdo con su situación financiera proyectada, situación legal de insolvencia y, cuando el crédito ha sido reestructurado, según la cancelación progresiva del saldo de capital¹⁶. Por su parte, las calificaciones de los deudores de consumo y de la cartera comercial asimilable a consumo son revisadas sólo de acuerdo con los atrasos y la situación legal

¹³ La CD tiene un apartado con datos de deudores dados de baja contablemente. Un análisis de la información allí contenida indica que son muy pocos los casos informados y no corresponden a las exclusiones que nos ocupan. Además existe un apartado con datos de balance, disponible para los grandes deudores.

¹⁴ Este umbral se incrementó desde 200 mil pesos en marzo de 2005.

¹⁵ Texto ordenado de Clasificación de Deudores, BCRA. Existe una situación 6, “Irrecuperable por Disposición Técnica” que se aplica a los morosos de ex entidades financieras para impulsar su regularización, la cual no se va a tomar en este análisis porque no se vincula directamente con la situación actual ni la proyectada del cliente.

¹⁶ Estas reglas fueron cambiadas en varias ocasiones luego de la crisis del 2002.

de insolvencia y existe un esquema de mejora de la clasificación en función de la cancelación de saldos o cuotas para los deudores refinanciados.

Las normas del BCRA obligan a calificar en grado “3” o peor a un deudor de consumo con un atraso de 90 días o más, mientras que el mismo atraso *sugiere* tal rating en el caso de un deudor comercial (se pueden considerar también las otras pautas). La normativa también establece que el banco debe cambiar su calificación si existe una diferencia mayor a un grado entre su rating y aquéllos asignados, por lo menos, por otras dos entidades cuyos créditos para el deudor representen 40% o más de su crédito total en el sistema¹⁷.

4.1. Muestra de desarrollo y definición de default

Analizaremos el comportamiento de personas jurídicas en la CD, con un horizonte de predicción anual, utilizando datos del período 2004 - 2006. De la base total se han seleccionado las financiaciones que al momento inicial (que se definirá con más precisión en el próximo apartado) reunían las siguientes características: (i) superaban \$100.000 por entidad¹⁸, (ii) estaban en cabeza de personas jurídicas del Sector Privado no Financiero, cuya actividad informada no es Administración Pública¹⁹, y (iii) no estaban en default. La aplicación de estos criterios lleva a que, en principio y salvo errores o casos puntuales, la muestra contenga datos de cartera comercial y comercial asimilable a consumo

Como se señaló, trabajamos con los grados 1 a 5 del sistema de calificación. Para decidir cuáles se tomarán como default básicamente hemos considerado la definición “consensuada” de Basilea II. Así, el default tiene lugar cuando se observan uno o ambos de los siguientes eventos:

- El banco considera improbable que el deudor pague en forma completa sus obligaciones crediticias a la entidad, sin mediar acciones por parte del banco, como la venta de garantías.
- El deudor está atrasado más de 90 días en cualquier obligación significativa con el banco²⁰.

Por lo tanto, tomamos a las situaciones 3 a 5 como defaults. Nuevamente, un atraso de más de 90 días para un deudor de la cartera comercial sugiere situación 3, pero podría prevalecer la consideración de otras pautas, aunque ello no ocurre habitualmente. Se toman entonces en la muestra los deudores que al momento inicial están clasificados 1 ó 2 en la entidad. Como la tasa de default observada en 2005 y 2006 ha sido muy baja, los defaults por año son muy pocos a los fines de la estimación. Por eso, se trabajó con la base agregada de los dos años, esto es, los períodos dic-04 a dic-05 y dic-05 a dic-06, tratando los datos como una única muestra y sumando 489 defaults. Vale la pena remarcar que los registros en default en dic-05 no son tomados en cuenta en el siguiente período porque se trabaja con registros que no están en default en el momento inicial. Al trabajar a nivel de deudor-banco, un deudor puede tener diferentes calificaciones en dos entidades (sujeto a la restricción normativa de máxima diferencia).

A los efectos de poder testear la generalización de los modelos (validación fuera de la muestra), se seleccionaron 70% de los registros para el desarrollo y 30% para su validación.

La muestra tiene 35.559 deudores-banco, mientras que el número de deudores consolidado a través de los bancos es de 15.785 (la misma empresa puede ser deudora en más de una entidad: en promedio cada deudor tiene crédito en 2,25 entidades). Las entidades tienen en promedio 428

¹⁷ Esta definición generaría cierta correlación “artificial” entre las calificaciones de un deudor en distintos bancos. Se aplica tanto a los deudores de la cartera comercial como a los de la cartera de consumo o vivienda.

¹⁸ En el estudio anterior de Balzarotti et al se utilizó como umbral \$200.000. La escasez de defaults obligó a disminuir este umbral.

¹⁹ Ambos criterios deberían ser casi coincidentes, pero en la práctica son habituales los errores.

²⁰ Para las obligaciones minoristas y entidades descentralizadas del Sector Público, y siempre que el supervisor nacional lo considere apropiado a las condiciones locales, puede utilizarse 180 días como criterio en lugar de 90.

créditos, pero estos no se distribuyen equitativamente entre ellas: de 83 entidades, 48 tienen menos de 100 créditos cada una.

Las variables explicativas se construyeron a partir de la información de la CD, a la fecha de observación y de seis meses antes. Estas variables principalmente se refieren a la calificación del deudor en diferentes períodos, su saldo de deuda con el banco y con el sistema, garantías, obligaciones en default y aperturas por tipo de línea crediticia (es decir, composición de productos crediticios). Algunas no resultaron significativas en la estimación. Se han descartado variables cuya calidad es dudosa (para más detalles ver Anexo 3).

Por otra parte, ciertas variables habituales en modelos de comportamiento crediticio a partir de una central de riesgos, como el ratio “saldo del crédito/limite disponible” y su evolución, o la antigüedad del crédito, no pueden ser calculados con la información disponible. Algunas fueron reemplazadas por proxies, como es el caso de la antigüedad crediticia, que fue aproximada por la antigüedad de la empresa que se deriva de la clave de identificación tributaria²¹.

4.2 Registros con comportamiento faltante

Dentro del conjunto de deudores que al momento inicial no están en default, algunos tienen datos de comportamiento en la misma entidad (situación al final del período), y otros no. Se genera así una primera clasificación. Puede establecerse otra clasificación según existan datos de comportamiento provistos por otras entidades. Estos dos criterios permiten la identificación de cuatro sub-poblaciones: la población “R” (por *Response*), que se subdivide en R1 y R2, y la población M (por *Missing*) que se subdivide en M1, M2, según la siguiente lógica:

Cuadro 1: clasificación de la población muestral

		Existen datos de comportamiento en la misma entidad	
		Sí (Población R)	No (Población M)
Existen datos de comportamiento en otras entidades	Sí=1	R1 datos X datos Y^0, Y^j ($j=1, \dots, N_k$)	M1 datos X datos Y^j ($j=1, \dots, N_k$)
	No=2	R2 datos X datos Y^0	M2 datos X sin datos Y

Los datos X representan variables del deudor en el momento inicial (a la “fecha de observación”) y los datos Y son datos de comportamiento (default o no). El superíndice indica la entidad acreedora (Y^0 es el comportamiento en la misma entidad, Y^j , ($j=1, \dots, N_k$), son los comportamientos en las N_k entidades adicionales con las que opera el deudor k); en general se omite el sub-índice k que indicaría el individuo, por simplicidad.

Los siguientes cuadros describen los tamaños de las poblaciones R y M, considerando las fechas de observación inicial dic-04 y dic-05 y las tasas de default. La clasificación en default / no-default es función de la situación un año después (dic-05 y dic-06, respectivamente). El porcentaje de faltantes es de 10,7% (3.792 casos); esto no es un hecho aislado sino persistente²².

²¹ La Clave Única de Identificación Tributaria –CUIT- que se otorga a las empresas se forma por un prefijo y un número secuencial por orden de ingreso al sistema central de la Agencia Fiscal, AFIP. Sólo cuando comenzó a utilizarse la CUIT se asignaron números que ya poseían las empresas en la Dirección Nacional de Recaudación. Esta proxy se testeó contra una base de datos que contiene el CUIT y la antigüedad, observando una alta correlación.

²² En el caso de deudores en el segmento de consumo este porcentaje suele ser mayor. La exclusión tiene lugar desde todos los grados de rating y aún para horizontes temporales mensuales.

Cuadro 2: tamaño de las poblaciones R (1 y 2) y M (1 y 2)

		Existen datos de comportamiento en la misma entidad	
		Sí (Población R)	No (Población M)
Existen datos de comport. en otras ent.	Sí=1	R1 24.541	M1 2.957
	No=2	R2 7.226 (22,7%)	M2 835 (22,0%)
total		31.767	3.792

Cuadro 3: poblaciones R y M, tasas de default y % de faltantes

	Observación inicial		Total Muestra
	dic-04	dic-05	
R	12.013	19.754	31.767
No defaults	11.806	19.472	31.278
Defaults	207	282	489
% defaults (sobre R)	1,72%	1,43%	1,54%
M	1.885	1.907	3.792
% M sobre Total	13,6%	8,8%	10,7%
Total (R+M)	13.898	21.661	35.559

Cuanto más baja la tasa de default, relativamente más serio será el problema de la existencia de faltantes, dado que los resultados serán sensibles a la cantidad de defaults (en general, en un modelo binario los resultados son sensibles al evento menos frecuente)²³.

La razón de estas exclusiones no se registra en la CD y por lo tanto no se sabe si los deudores faltantes han sido o no defaults. A priori, hay dos motivos preponderantes y opuestos:

- 1) Cancelación total de los créditos. Estos deudores deberían clasificarse como no-defaults o bien “indeterminados”²⁴.
- 2) Deudores dados de baja por impago, los cuales deberían ser clasificados como defaults.

Que un crédito en situación 1 ó 2 desaparezca de la CD por impago en un plazo de 12 meses implicaría la aplicación de criterios de castigo (“*write off*”) mucho más estrictos que los regulatorios. Ello es así porque las normas para deudores comerciales indican situación 5 para atrasos de 365 días o más. Los créditos clasificados en situación 5 y totalmente provisionados deben ser transferidos a Cuentas de Orden (CO) luego de permanecer siete meses en esa situación. Si un crédito en situación 5 tiene garantías reconocidas en las normas, se provisiona el 50% durante dos años (en general), y luego se pasa a provisionar completamente. Un crédito permanece en estas cuentas si la entidad continúa ejerciendo acciones de cobro. En todo el proceso, los créditos deben ser informados a la CD, inclusive aquéllos en CO.

Una tercera razón posible para la exclusión de los deudores de la base es la venta de créditos a otra entidad o a un fideicomiso. Sin embargo, estos créditos continuarán siendo informados a la

²³ Por ejemplo, Balzarotti, Gutiérrez Girault et al (2006) encontraban para el período 1999-2000 un 16% de faltantes, pero entonces este porcentaje era sólo el doble de la tasa de default, que rondaba el 8%. Con los nuevos datos, los faltantes representan más de 6 veces los defaults observados.

²⁴ Cuando se conoce que la cancelación ha sido voluntaria en determinados casos, en especial para productos *revolving* de consumo, es usual clasificarlos como ‘indeterminados’ y excluirlos del desarrollo del modelo. Usualmente se los incorpora luego en los análisis de distribución de score de la población, además de verificar que su distribución se encuentra entre la de los ‘buenos’ y los ‘malos’.

CD si el comprador es un fideicomiso financiero (el tipo más común de comprador). Y aún cuando no sea uno de estos fideicomisos, si la entidad vendedora u otra entidad financiera provee servicios de cobro, entonces la institución debe informar los créditos a la CD. No obstante, como nuestro análisis es a nivel de deudor-banco, un cambio en el tenedor será considerado como un faltante. Sólo a nivel sistémico pueden inferirse las transferencias. Adicionalmente, existe un motivo operativo que puede llevar a una pérdida temporal de la rastro de un crédito transferido, originado en el hecho de que los fideicomisos nuevos pueden informar sólo cuando completan los procedimientos de autorización, los cuales pueden tomar un par de meses y este plazo puede coincidir con la fecha de la observación final.

Una cuarta y última razón para la exclusión de los deudores de la CD es un simple error de información, en cuyo caso podría suponerse aleatorio.

Mientras que las mencionadas cuatro causas pueden estar detrás de un registro removido, por lo dicho anteriormente puede argumentarse que, a priori, las cancelaciones, ventas y errores son más probables que el envío a pérdida.

Para tener más indicios, se puede comparar las características de los deudores removidos con aquéllas de los deudores que permanecen. A partir de una comparación extensiva de las distribuciones por variable, puede observarse que ambos grupos presentan alta similitud, a excepción de algunas variables particulares. La principal variable que distingue claramente ambos grupos es la cantidad de asistencias con el banco, aunque no así la cantidad de asistencias con todas las entidades con las que opera. La situación del deudor-banco seis meses atrás (en especial un faltante de información) es también un elemento distintivo, así como el financiamiento a través de bienes en locación financiera. El grupo de bajas presenta una densidad algo mayor en niveles de deuda altos. El Anexo 2 brinda más información.

Asimismo, resulta indicativo también observar que, de los 60 bancos con más de 20 créditos en la muestra, sólo 1 no presenta registros faltantes. Esto indica que el problema es generalizado.

Otro indicio útil puede encontrarse en el análisis de los saldos de deuda de los deudores que son removidos de algún banco, pero que aún tienen crédito en el sistema financiero (Cuadro 4). En general, la deuda total de estos deudores en el sistema ha decrecido.

Cuadro 4: Cambio en el saldo de deuda total en el sistema de los deudores M (obs= dic. 2004 y dic. 2005)

Cambio deuda total sistema	Frecuencia	%	Tasa de cambio promedio
Aumenta	543	18,5%	33%
Disminuye	1.222	41,7%	-38%
Pasa a 0	1.165	39,8%	
Total	2.930	100,0%	

5. Modelos de score y esquemas alternativos estudiados

Antes de entrar en detalle a los diferentes esquemas que se usarán para estimar los modelos y las inferencias del comportamiento de los registros faltantes, vale la pena remarcar la diferencia entre truncamiento, censura, sesgo de selección y el problema de los deudores faltantes. El Recuadro 1 examina esa diferencia.

En este trabajo no adoptamos un enfoque que intente modelar el proceso de faltantes. El foco está puesto en (i) estimar el riesgo de los deudores removidos y (ii) incorporar la información adicional de comportamiento en otras entidades acreedoras para la estimación del primer punto y

para mejorar la predicción del score. Aplicaremos tres técnicas para estimar el modelo de score, que se definen básicamente por la forma en que se incorpora dicha información.

(i) El procedimiento más habitual / tradicional, de *listwise deletion*, (en adelante Modelo LD). Se eliminan de la base los registros cuyo comportamiento en el banco no está registrado, es decir, se estima sobre la población R. No usa datos adicionales. Se estima este modelo a los efectos de comparar sus resultados con los que surjan de aplicar inferencia de comportamiento para los deudores excluidos.

(ii) El siguiente método incorpora el comportamiento del deudor con otras entidades por *imputación directa* (ID). Para la población M1 se imputa el comportamiento faltante en la entidad en función de la peor situación del deudor en otras entidades con las que opera (que se representa en la variable *comport_ob*). Existen las siguientes posibilidades para un deudor en M: (a) que se le impute comportamiento de default (*comport_ob*=3 o peor), (b) que se le impute no default (*comport_ob*=1 ó 2) o (c) si el deudor no opera con otro banco (*comport_ob*=-1), esto es, la sub-población M2, no se imputa ningún comportamiento y no se reincorpora a la muestra. Así, se estima el modelo sobre la población R y la población M1 reincorporada.

Es notorio (Cuadro 5) que M1 tiene una tasa de defaults imputados mucho mayor que los observados para R (8,3% versus 1,5%). Para buscar un indicio de si esa diferencia se debe al comportamiento real de los deudores M1 o a la imputación, se analiza qué pasaría si se imputara el comportamiento de los deudores sobre los cuales existe comportamiento observado (R1), en vez de tomar éste. Como se ve en el Cuadro 6, la tasa de default de la población R pasaría de 1,5% a 4,5%, lo cual constituye una señal de que ID tiende a sobreestimar el riesgo.

Cuadro 5: Defaults por Imputación Directa (ID) de la población M

	cantidad	porcentaje
Población M	3.792	100.0
M2 Sin otro banco	835	22,0% de M
M1 Con información en otra entidad	2.957	78,0% de M
Defaults por ID	245	8,3% de M1
No defaults por ID	2.712	
Población R	31.767	
Defaults de R	489	1,5% de R
Cantidad incremental de defaults	245	50,1%

Cuadro 6: Tasa de default de R si se imputara el comportamiento por ID

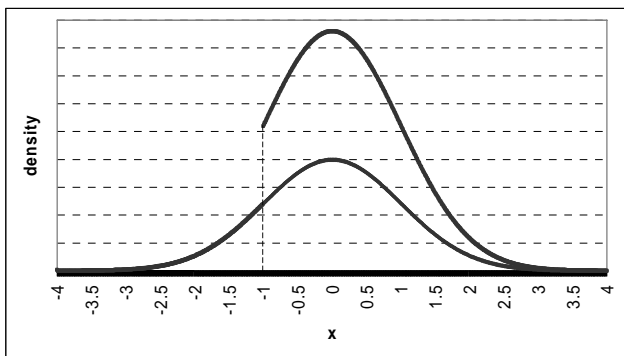
	cantidad	Porcentaje
Población R	31.767	100.0
R2 (sin información en otra entidad)	7.226	
Defaults	151	2,1% de R2
No defaults	7.075	
R1 (con información en otra entidad)	24.541	
Defaults por ID	1.264	5,2% de R1
No defaults por ID	23.277	
Total Defaults imputados	1.415	4,5% de R
Total Defaults observados	489	1,5% de R

Recuadro 1: Truncamiento, censura, sesgo de selección y deudores excluidos

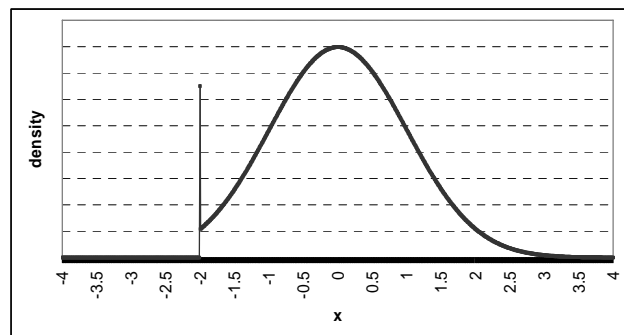
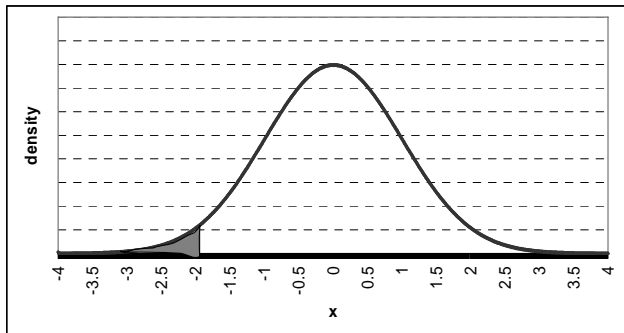
Hay truncamiento cuando los datos de una muestra se toman de un subconjunto de la población, donde el proceso de truncamiento no es aleatorio, mientras que la censura es un defecto en los datos muestrales: los valores que caen dentro de cierto rango de una variable censurada se informan como un único valor. El Sesgo de Selección (SS) es una forma de truncamiento.

Los gráficos siguientes ilustran el truncamiento y la censura para una distribución Normal.

Distribución Normal Truncada (en -1)



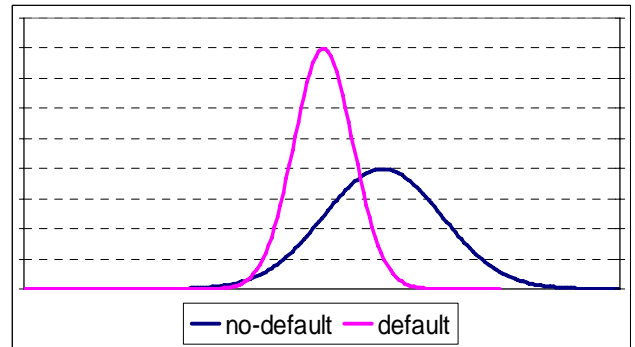
Distribución Normal Completa y Censurada (en -2)



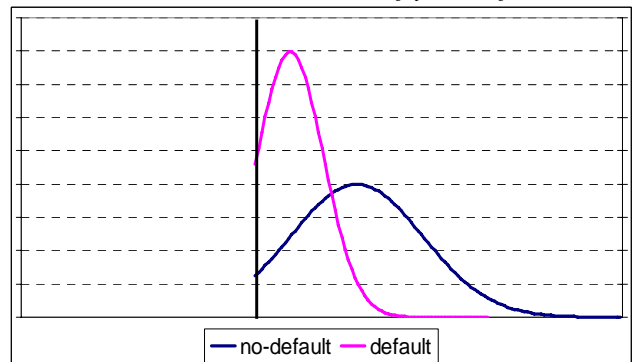
En caso de truncamiento, si el análisis se limita a la sub-población, entonces los coeficientes sesgados son válidos. Si el estudio intenta extender sus resultados a la población, lo que importa son los parámetros insesgados.

Cuando existe SS, el muestreo no es aleatorio. En el caso de entidades de crédito, el SS es un problema común: normalmente se mantienen registros de los créditos otorgados y no de los rechazados. En esos casos, el truncamiento tiene “una dirección”. La pre-selección de la muestra puede representarse, en forma simplificada, por una condición de aceptación de una solicitud igual a que un score (el modelo previamente utilizado para la aceptación/rechazo) sea mayor a cierto umbral (una simplificación de los criterios para otorgar crédito, intuitivamente atractiva). Las distribuciones de los defaults y no-defaults serán observables para scores censurados, como ilustran los gráficos: el primero muestra las densidades completas de los defaults y no-defaults, y el segundo, las densidades de los defaults y no-defaults con SS.

Densidades completas de deudores en def y no-def



Densidades de deudores en def y no-def con SS



Podría haber sesgo en las estimaciones de PD cuando se omite considerar el mecanismo de selección.

En el caso de este trabajo, falta el registro de algunos deudores en el punto final de observación (y la variable truncada es binaria), sin que se registre el motivo de la baja. No considerar este hecho puede conducir a estimaciones sesgadas. Pero este sesgo, a priori, no tiene una dirección, o tiene ambas direcciones como se explicó anteriormente.

(iii) Por último, el procedimiento más elaborado que analizaremos, asimilable a un *fractionally weighted imputation* (FWI). Este procedimiento también tiene dos etapas: una de imputación de comportamiento y otra de estimación del modelo de scoring. A diferencia del método anterior, la imputación también se hace mediante un modelo. En efecto, se estima un modelo logit sobre la población con comportamiento observado (R), en el cual la variable a explicar es el comportamiento observado (default o no default) y *las variables predictoras incluyen el comportamiento en otros bancos*²⁵.

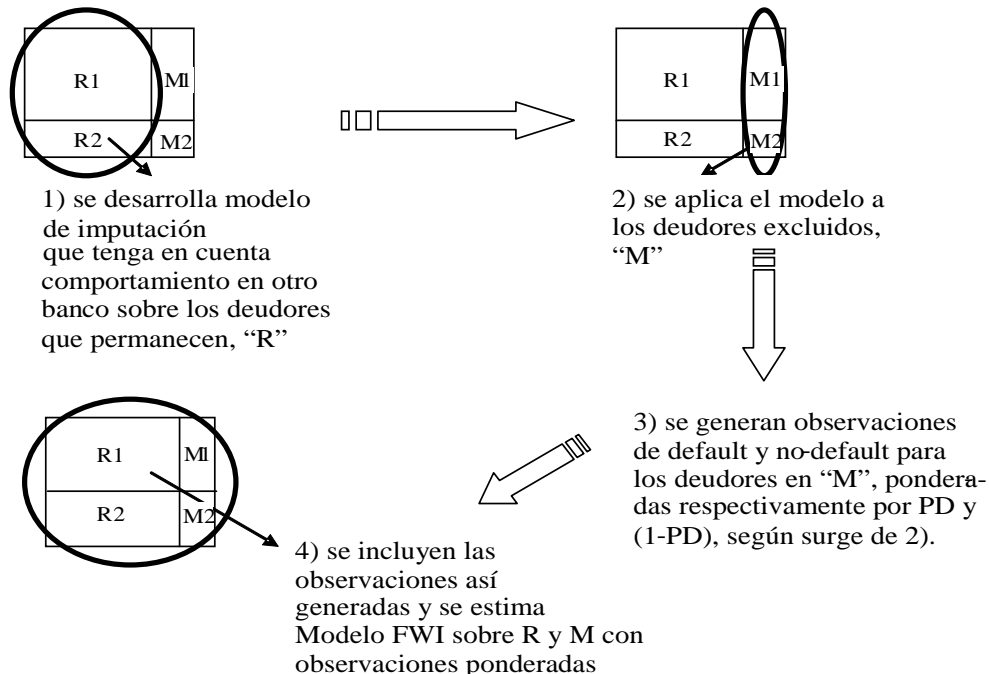
El modelo de imputación, estimado sobre R, se aplica sobre M para inferir el comportamiento en la entidad de los deudores-banco excluidos, aprovechando la información del comportamiento en otras entidades. Se calculan así las PDs para los individuos de M, lo cual permite imputar dos registros para cada faltante: en uno se imputa el valor default, con ponderador PD, y en el otro no-default, con ponderador (1-PD).

Finalmente, con la muestra ampliada, que ahora cuenta con datos de comportamiento real o inferido para toda la población, R y M (estos últimos con dos registros ponderados para cada caso), se construye un nuevo modelo de scoring, al que denominaremos Modelo FWI. El procedimiento se ilustra esquemáticamente en el Recuadro 2.

El método de imputación parte de la base de que el comportamiento de un deudor en distintas entidades tiene una correlación alta, aunque no perfecta, como asume el método ID²⁶. El supuesto implícito es que las relaciones estimadas para R son trasladables a M.

Para *imputar* comportamiento el modelo utiliza información de comportamiento futuro en otra entidad. Obviamente, para *predecir* comportamiento no sería válido usar información del futuro (el modelo de score crediticio no podría incluir variables explicativas referidas al futuro).

Recuadro 2: esquema de estimación por FWI (Imputación ponderada fraccional)



²⁵ Para la construcción de este modelo se utiliza el mismo conjunto de variables que para el modelo LD. La variable explicativa que mide comportamiento en otros bancos se construye de la misma manera que se explicó para ID.

²⁶ Como extensión a este trabajo, se estudiará más adelante el efecto de la correlación del comportamiento de un mismo deudor en distintas entidades en la estimación de los modelos finales (es decir que no se supondrá independencia en la construcción de la función de verosimilitud). La práctica usual en el armado de modelos de buró de crédito no toma en cuenta esta correlación, aún cuando no se trabaja en forma consolidada a nivel deudor.

Son varias las ventajas del método FWI respecto de la ID: (i) estima el verdadero efecto del comportamiento en otros bancos sobre el default del deudor en la entidad en cuestión, al medirlo para el caso de la población R, para la cual Y^0 e Y^j son observables; (ii) puede contemplar el caso en el que un deudor no opere con otra entidad, a través de un valor específico para la variable explicativa, y (iii) refleja la incertidumbre en la imputación del comportamiento a los faltantes, al trabajar con dos imputaciones ponderadas para cada registro de M, mientras que ID imputa un valor como si existiera certeza respecto de él.

6. Resultados

A continuación se muestran los resultados resumidos de las estimaciones de los modelos, las validaciones efectuadas sobre ellos y la comparación de los mismos. Los resultados del modelo utilizado en la imputación fraccional de comportamiento (IMP) se muestra en el Anexo 5. Los modelos se parametrizan de manera que a mayor puntaje obtenido, menor es la PD estimada.

6.1. Comparación de los modelos

El Cuadro 8 muestra los parámetros estimados por los tres modelos. No hay diferencias sustanciales en los valores estimados de los parámetros, especialmente comparando el modelo LD con el FWI (a excepción de la constante, debido a que la tasa de default difiere en las poblaciones de desarrollo: en el Cuadro 7, se puede ver que las mismas son: 1,54%, 2,11% y 1,67% para LD, ID y FWI respectivamente). Sólo un coeficiente referido al tipo de entidad cambia de signo en el modelo ID, pero en ese caso resulta no significativo al 10% de confianza.

Cuadro 7: Resumen del impacto en la muestra de los procesos de imputación

		Modelo LD	Modelo	
			ID	FWI
Población R	total	31.767	estimado = observado	
	defaults	489		
	no-defaults	31.278		
	faltantes	-		
	tasa de default	1,54%		
Población M	total	3.792	3.792	3.792
	defaults	-	245	105,1
	no-defaults	-	2.712	3.686,9
	faltantes	3.792	835	-
	tasa de default	-	8,29%	2,77%
Tasa de default total muestra		1,54%	2,11%	1,67%

En general, los coeficientes tienen los signos esperados. La variable “Cantidad de bancos en situación 3+, 6 meses atrás” puede llamar la atención por su coeficiente positivo; sin embargo, cabe recordar que la muestra sólo contiene deudores clasificados 1 ó 2 al momento inicial, por lo que esta variable indica una mejora respecto del semestre anterior. Es interesante ver que algunas variables que típicamente resultaban significativas en estudios anteriores, como el monto de la deuda, o el ratio de la deuda con la entidad en términos del total en el sistema, o el porcentaje de crédito cubierto con garantías preferidas, resultan no significativas y son desplazadas en términos de significación por otras variables, que dan cuenta principalmente del tipo de asistencia (adelantos, leasing, etc.) y del número de líneas de ciertos tipos de crédito. Estas variables, que describen la composición de tipos de productos financieros, están relacionadas con la escala a la cual opera el deudor. También es llamativo que el hecho de contar con crédito hipotecario aumenta el riesgo, si bien este resultado es consistente con la literatura que sostiene que los

bancos demandan garantía a los deudores que son a priori más riesgosos y que la provisión de la misma no llega a compensar totalmente el mayor riesgo²⁷.

Cuadro 8: Comparación de los parámetros estimados para los distintos modelos de score

Default es la variable explicada						
Variable	Valor / rango	Modelo LD		Modelo ID		Modelo FWI
Constante		-1,266	**	-0,827	*	-1,698 ***
Cantidad de bancos en situación 3+	[0-10]	-0,440	***	-0,470	***	-0,420 ***
Situación al momento de observación	1	1,083	***	0,768	***	1,162 ***
	2	0,000	***	0,000	***	0,000 ***
Dummy: mínima situación 6 meses atrás es 3+ (a nivel deudor)	No	1,129	***	0,855	***	1,273 ***
	Sí	0,000	***	0,000	***	0,000 ***
Ratio monto en default/total deuda	[0-1)	-1,502	***	-1,732	***	-1,487 ***
Dummy Inicio de actividades reciente (CUIT >=702)	No	0,938	***	0,822	***	0,890 ***
	sí	0,000	***	0,000	***	0,000 ***
Agrupación por tipo de banco ("Grupo homogéneo")	Resto	0,530	***	-0,037		0,390 ***
	Mín. Grnd.	1,098	***	0,631	***	1,056 ***
	Púb. Nac.	0,000	***	0,000	***	0,000 ***
Dummy: posee otros créditos intermediación financiera	no	0,622	***	0,473	***	0,695 ***
	sí	0,000	***	0,000	***	0,000 ***
Dummy: posee bienes en locación financiera (<i>leasing</i>)	no	-1,257	***	-1,230	***	-1,230 ***
	sí	0,000	***	0,000	***	0,000 ***
Dummy: posee asistencias de tipo adelanto	no	0,479	***	0,238	**	0,451 ***
	sí	0,000	***	0,000	***	0,000 ***
Dummy: posee asistencias de tipo hipotecario	no	0,421	***	0,328	**	0,398 ***
	sí	0,000	***	0,000	***	0,000 ***
Cantidad de líneas prendario (a nivel deudor)	[0-5]	0,338	***	0,356	***	0,432 ***
Posee garantías de tipo "A"	no	0,545	***	0,491	***	0,716 ***
	sí	0,000	***	0,000	***	0,000 ***
Peor situación 6 meses atrás (a nivel deudor)	0 (faltante)	1,659	***	2,313	***	1,733 ***
	1	0,787	***	1,393	***	0,907 ***
	2+	0,000	***	0,000	***	0,000 ***
Cantidad de bancos en situación 3+, 6 meses atrás	[0-14]	0,225	***	0,261	***	0,275 ***
Dummy: empeoró su situación desde hace 6 meses (a nivel deudor)	no	0,870	***	1,077	***	0,803 ***
	sí	0,000	***	0,000	***	0,000 ***
Cantidad de líneas tipo préstamos personales o tarjetas de crédito (a nivel deudor)	[0-5]	0,485	***	0,444	***	0,420 ***

Nota: ***, ** y * indican niveles de confianza al 99%, 95% y 90% respectivamente.

El coeficiente para la calificación inicial tiene el signo esperado: a mejor situación se reduce la probabilidad de default. Los deudores más antiguos (medidos según la variable derivada del

²⁷ Un resultado similar obtienen Schechtman, Salomão García et al (2004) al estimar PDs para deudores comerciales en Brasil, donde encuentran que la relación con las garantías no es estadísticamente significativa.

número de CUIT) son menos riesgosos, en cambio los que aportan garantías preferidas A (fundamentalmente descuento de documentos) son más riesgosos. Los créditos de bancos grandes minoristas resultan menos riesgosos, en comparación con los públicos. El Anexo 3 muestra las relaciones bivariadas entre algunas variables explicativas y las tasas de default.

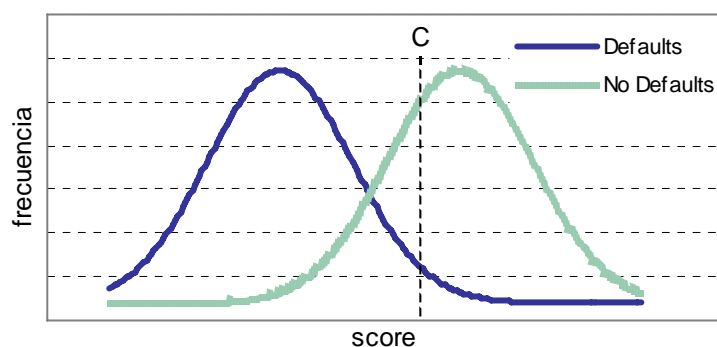
6.2.- Poder discriminatorio y calibración. Validaciones.

Como ya se dijo, los datos de la CD para estimar un score (y eventualmente un sistema de rating) son necesariamente incompletos, en la medida en que los bancos tienen acceso a información de características del deudor que no se registran en la CD. Sin embargo, es importante testear los méritos de los modelos de este trabajo por interés metodológico y para evaluar su utilidad para los supervisores, como benchmark de riesgo, o para calibrar la regulación prudencial.

Los scores crediticios son herramientas de clasificación que proveen indicadores de la situación futura del deudor. Hay una serie de aspectos a ser evaluados sobre ellos: analizaremos el poder discriminatorio y la calibración. La publicación del Comité de Basilea “*Studies on the Validation of Internal Rating Systems*” (2005)²⁸ sugiere algunas herramientas estadísticas para validar ambos aspectos de sistemas de rating. En esta sección mostraremos los resultados de aplicar algunas de las técnicas más ampliamente utilizadas a los modelos obtenidos. Se aplicarán validaciones dentro y fuera de la muestra (sobre las observaciones que se separaron a ese fin)²⁹.

Cuando un score es bueno, el conjunto de deudores que obtiene los mejores puntajes mostrará baja frecuencia de default y el grupo que resulta asignado a los peores grados mostrará mayor frecuencia de default. Por lo tanto, un score tiene más **poder discriminatorio** cuanto mayor es la diferencia entre las distribuciones de los scores para los deudores que hacen default y de los que no (Gráfico 1). Existen varias medidas estadísticas del poder discriminatorio de un sistema de rating, algunas de las cuales se describen en el Anexo 4. Ellas tienen un significado limitado como valores absolutos; se utilizan más bien para comparar entre sistemas de rating, para evaluar el poder predictivo relativo del modelo.

Gráfico 1: Distribución de defaults y no-defaults



Chequear la **calibración** de un sistema de rating apunta a evaluar si el nivel de riesgo que se deduce del sistema es correcto, lo cual cobra más relevancia en la medida en que el sistema se utilice para hacer “pricing” de créditos, calcular primas de riesgo o cargos de capital y, en ese sentido, la calidad de la calibración es un criterio prudencial importante en la evaluación de estos sistemas. Para evaluar la calibración, cada grado de un sistema de ratings o cada valor de score debe estar asociado a una PD. Un sistema con calibración correcta exhibiría PDs proyectadas

²⁸ BCBS (2005).

²⁹ Además de ser práctica usual, Basilea II también incluye esto como requerimiento. Ver BCBS (2004), pfo. 251.

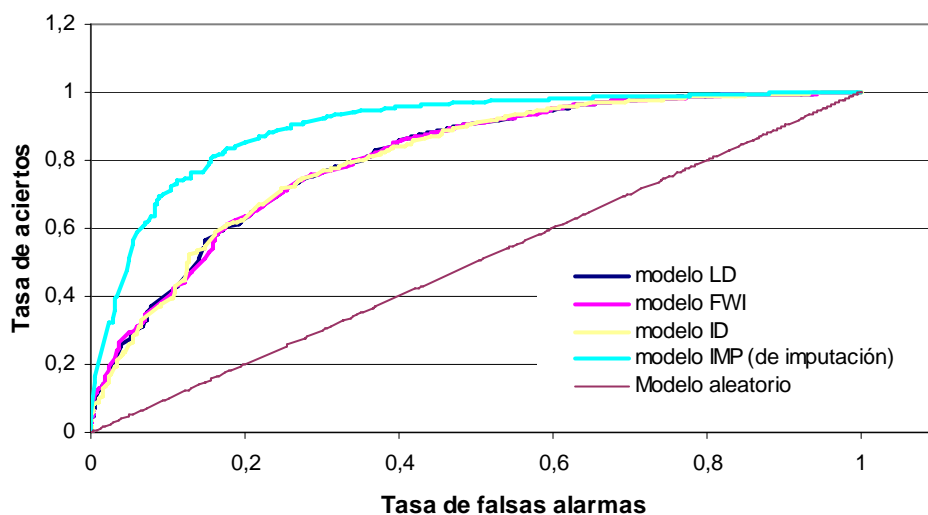
similares a la tasa de default de los deudores que pertenecen al mismo grado de rating. En la práctica, las PDs estimadas diferirán de las tasas de default observadas. Estas desviaciones pueden deberse a factores aleatorios o pueden ocurrir sistemáticamente, en cuyo caso el modelo puede necesitar ser recalibrado.

- **Estadísticos de Poder Discriminatorio**

- *Curva ROC (Receiver Operating Characteristics).*

Tomando un score C como punto de corte (*cut-off*), se clasifica a un deudor como default potencial si su score es inferior a C y como potencial no-default si es superior. El sistema de rating será mejor cuanto mayor sea la tasa de aciertos o “hit rate” (correcta clasificación de los deudores que incumplen como potenciales defaults) y al mismo tiempo menor sea la tasa de “falsa alarma” (incorrecta clasificación de un deudor cumplidor como un potencial default). Para construir la curva ROC, se calculan la tasa de aciertos y de falsa alarma para cada score (véase Anexo 4). La performance de un sistema de rating es mejor cuanto más empujada sea la curva ROC y cuanto más cerca se encuentre del punto (0;1). Las curvas ROC de los modelos se presentan en el Gráfico 2.

Gráfico 2: Curvas ROC



El área bajo la curva, llamada AUROC, se mide por el índice ROC. El valor de este índice va desde 0,5 para un modelo aleatorio (donde la curva ROC es la diagonal del gráfico) hasta 1 para el modelo ideal. Otras medidas de poder discriminatorio están relacionadas con la curva ROC, como la curva CAP o “lift” o curva de Lorenz, el Accuracy Ratio, el coeficiente Gini y el “K-S”.

Test K-S: Vale la pena incluir el estadístico de Kolmogorov Smirnov (K-S) por su amplia difusión. Este estadístico varía entre 0 y 1 y mide la distancia máxima entre la curva ROC y el eje x (que puede ocurrir en cualquier punto de la curva). También puede obtenerse como la máxima diferencia entre las distribuciones de score acumuladas de defaults y no-defaults. Dado que el K-S no contempla dónde se da esta máxima distancia, es preferible el estadístico de AUROC.

Cuadro 9 – Medidas de Poder Discriminatorio

Modelo	Muestra	AUROC	K-S
LD	Muestra de desarrollo (R)	0,809	0,456
	Validación fuera de la muestra (R)	0,799	0,434
ID	Muestra desarrollo (R+M1)	0,804	0,463
	Validación fuera de la muestra (R+M1)	0,797	0,455
FWI	Muestra desarrollo (R+M)	0,809	0,457
	Validación fuera de la muestra (R+M)	0,800	0,446

Los modelos estimados tienen estadísticos de poder discriminatorio altos, en todos los casos.

Como referencia, los estadísticos obtenidos para el modelo de imputación de comportamiento IMP (sobre R) fueron AUROC=0,90 y KS=0,648, lo que está indicando el valor incremental de la información del comportamiento futuro en otros bancos a los efectos de la imputación (aunque no predicción perfecta, en cuyo caso el modelo IMP hubiera sido análogo a la imputación directa).

- **Estadísticos de Calibración**

El documento sobre validación del Comité de Basilea³⁰ muestra distintos tests para evaluar calibración. Su utilización no se encuentra tan difundida como la de los estadísticos de poder discriminatorio, en parte porque no todas las aplicaciones de scoring requieren contar con estimaciones tan precisas de las PD, como si lo requiere el cálculo de requisitos de capital bajo el enfoque IRB.

Aquellos tests cuyo uso está más difundido, los tests Binomial y Hosmer-Lemeshow (H-L, el cual se usa también como test de bondad de ajuste), suponen que los eventos de default son independientes. Mientras que el test Binomial examina cada grado del sistema de rating en forma separada, el test H-L considera todos los grados al mismo tiempo.

La independencia de los eventos de default no es un supuesto demasiado fuerte ya que, empíricamente, es normal observar bajas correlaciones. Además, dicho supuesto es conservador ya que al no considerar que los eventos de default están correlacionados aumenta la probabilidad de rechazar la hipótesis nula (H_0 = las PDs estimadas son correctas). Luego, si el test indica que los grados del sistema de rating son correctos, este es un resultado robusto.

Hemos establecido, para el cálculo de este test, un sistema de ratings por deciles³¹ de score para cada modelo y para todo el sistema (en contraposición a los resultados de un banco en particular). Al usar el método logit, las PDs asociadas a cada deudor³² se calculan de manera directa mediante la relación (1) (página 3).

Se aplicó el test H-L para los tres modelos, resultando que no se puede rechazar la hipótesis nula de que las PDs promedio de los grados del sistema de rating son las verdaderas probabilidades³³.

³⁰ BCBS (2005).

³¹ Basilea II establece requerimientos mínimos para el diseño de estos sistemas. Entre los requerimientos para la estructura de un sistema de rating de exposiciones con empresas, soberanos y bancos se señala que las exposiciones no deben estar excesivamente concentradas en algún(os) grado(s) del sistema de ratings y que éste debe tener como mínimo siete grados para deudores que no están en default y uno para deudores en default.

³² La PD corresponde al deudor-banco. La dimensión de diferentes tipos de crédito o de deudor en el sistema será considerada en futuras investigaciones.

³³ La hipótesis nula se rechazaría, por ejemplo, a un nivel de confianza del 95%, es decir, un p-value de 0.05.

El test debió modificarse para el modelo FWI de forma de tener en cuenta que existen en la base de datos observaciones ponderadas (ver detalles de esta adaptación en Anexo 4).

Cuadro 10 – Test H-L: Validación de la calibración

Modelo	HL	Grados de libertad	p-value
LD	7,2385	8	0.5111
ID	9,8559	8	0.2753
FWI	9,2880	8	0.3186

6.3. Análisis de los resultados

Contamos con dos poblaciones relevantes (R y M), que a su vez pueden dividirse en cuatro subpoblaciones (R1, R2, M1 y M2), y 3 modelos alternativos. Además, se pueden analizar los resultados a nivel del sistema financiero agregado o por banco (más acorde a un análisis de supervisión). En vistas del número de dimensiones de análisis posibles, nos concentraremos en los aspectos en los cuales este trabajo hace mayor foco: (i) cómo medir el riesgo de la población con comportamiento faltante, M, especialmente utilizando la información sobre su comportamiento en otras entidades, y (ii) qué impacto puede tener el no considerar la población M sobre los modelos de scoring y sobre la evaluación del riesgo.

Antes de pasar a esos dos puntos, comenzaremos efectuando algunos controles cruzados sobre los modelos.

Como el modelo de imputación IMP se estima sin utilizar información de M, resulta informativo evaluar cómo predice el comportamiento sobre esa población (ya mencionamos que el modelo IMP tiene muy alta discriminación sobre R, mejor que cualquiera de los otros modelos sobre R). Para hacer este control es necesario suponer el comportamiento real de los registros en M. Ello se puede hacer descartando la población M2, sobre la cual no hay ninguna información de comportamiento, y suponiendo para los deudores de M1 que su comportamiento coincide con el peor en otra entidad. Si bien podemos esperar que la tasa de default se sobreestime mediante este mecanismo, se puede obtener un estadístico de poder discriminatorio orientativo. Recordando también que una variable explicativa del modelo IMP es el comportamiento en otros bancos, se obtiene como es esperable un poder predictivo muy alto: AUROC=0.981 y KS=0.923.

El segundo control consiste en revisar que el modelo final FWI (que se desarrolla sobre R más las imputaciones de M) no pierda poder discriminatorio sobre la población R respecto del modelo LD (que se desarrolla sólo sobre R). Efectivamente, el poder discriminatorio no se deteriora significativamente como puede observarse en el Cuadro 11 (AUROC 0,806 vs 0,809). El cuadro muestra otros controles cruzados relevantes, los que no indican que existan problemas.

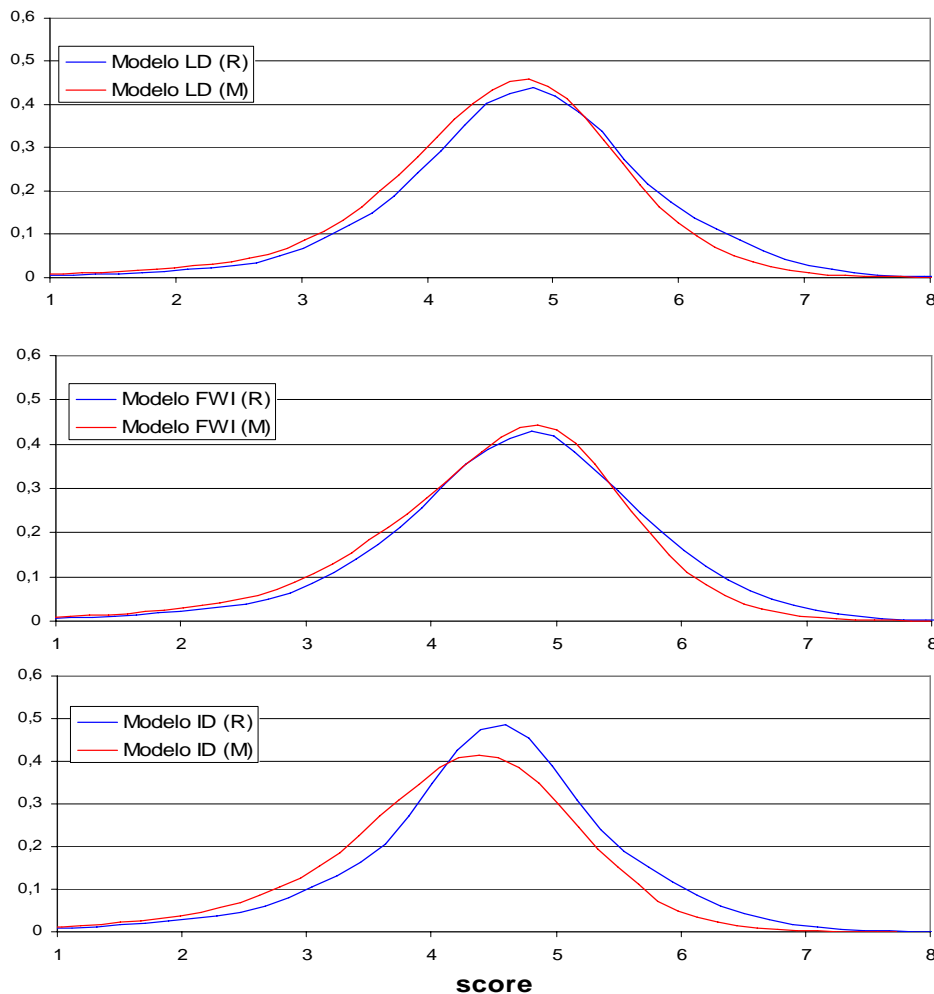
Cuadro 11: Controles cruzados de poder discriminatorio

Modelo	Muestra para validar	AUROC	KS
LD	Muestra de desarrollo (R)	0,809	0,456
	Valid. R+M (imputada)	0,807	0,454
ID	Muestra desarrollo (R+M1)	0,804	0,463
	Valid. R+M(imputada)	0,799	0,454
FWI	Muestra desarrollo (R+M)	0,809	0,457
	Valid. R	0,806	0,449

Pasando a analizar el riesgo crediticio de la población M, se debe tener presente que, a ciencia cierta, nunca se podrá medir la calidad de la estimación sobre esta población ya que su comportamiento real no es observable; sólo se analizan elementos que permitirían inferir su riesgo.

Un análisis a priori de las distribuciones de las características de M comparadas con R muestra que no hay diferencias apreciables entre las dos poblaciones (por lo menos en las variables que resultan predictoras de riesgo)³⁴. No obstante, con los 3 modelos estimados se detecta un mayor riesgo para M con respecto a R (distribuciones de score a la izquierda, con diferencias estadísticamente significativas), especialmente en el modelo ID (Gráfico 3).

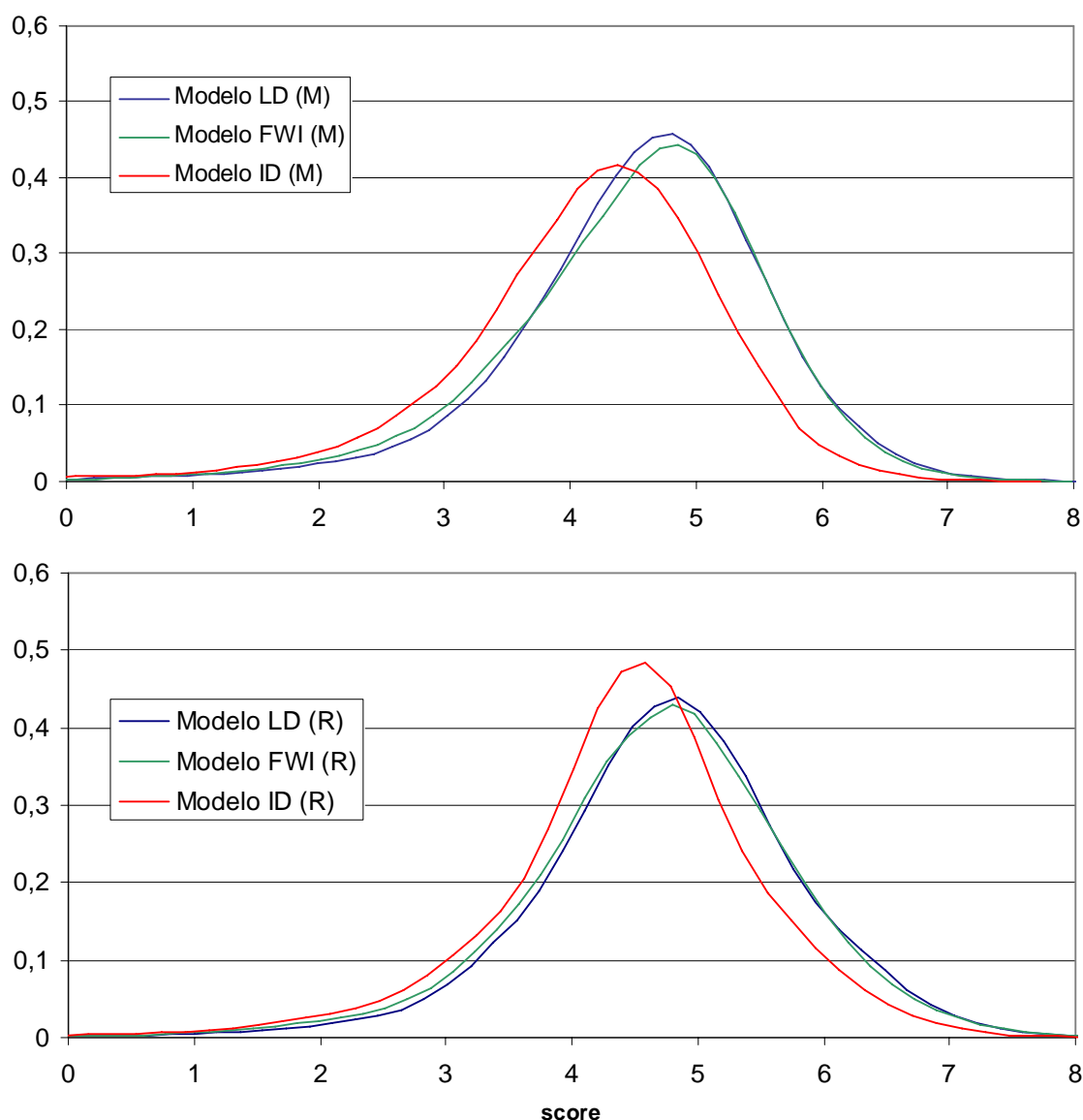
Gráfico 3: Distribuciones de score para M y R



Comparando ahora las distribuciones de score que estiman los 3 modelos para la población M, observamos que LD y FWI son más coincidentes y que el modelo ID parece exagerar el riesgo. (Gráfico 4). Lo mismo ocurre al aplicar los 3 modelos a la población R, cuyo comportamiento es conocido y por lo tanto no depende de los supuestos al respecto (Gráfico 4). Ya habíamos intuido previamente que la imputación directa podía sobre-estimar riesgo al mostrar que su aplicación a la población R llevaba a sobre-estimar la tasa de defaults.

³⁴ Modelar el proceso que genera los faltantes no es el enfoque que se adoptó en este trabajo y por eso no se profundiza.

Gráfico 4: Distribuciones de score para M y R según los 3 modelos



Las primeras conclusiones importantes son, entonces, que el riesgo de M es mayor que el de R y que la imputación del peor comportamiento en otra entidad para estimar el modelo de score conduce a una sobre-estimación del riesgo.

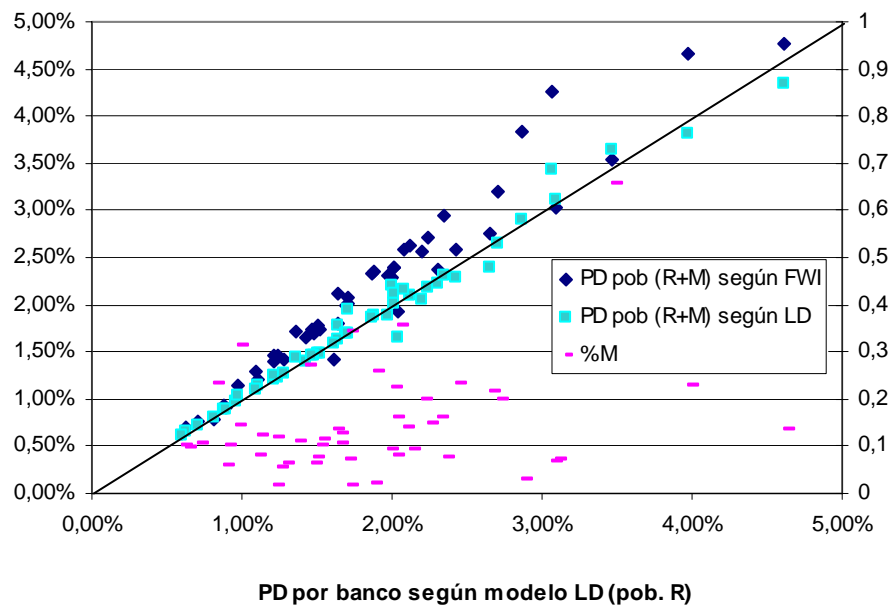
El impacto de la población M sobre la evaluación de riesgo se da a través de dos vías: que no se considere al estimar el modelo (éste se estima sólo sobre R, ignorando que existe M) y que se omita al aplicar un modelo estimado para cuantificar el riesgo (el modelo se aplica sobre R, por ej., para calcular la PD promedio de una cartera). Por ejemplo, un analista que desarrollara el modelo de scoring aplicando LD, probablemente calcularía el riesgo del portafolio sobre la misma población de desarrollo. En cambio, un analista que intentara estimar el riesgo de M, probablemente tendría en cuenta esta población para estimar el riesgo agregado.

Una medida simple para comparar el riesgo estimado de las carteras de las entidades es la PD estimada promedio por entidad, calculada como el promedio de las estimaciones de PD sobre todos los deudores k que operan con el banco B , según la siguiente expresión:

$$\overline{PD}_B = \text{prom}_{k \in B}(PD_k) = \text{prom}_{k \in B} \left(\frac{1}{1 + \exp(X_k' \beta)} \right)$$

Si se comparan las estimaciones del modelo LD sobre la población R –que llamamos “LD(R)”- con las de los modelos FWI y LD sobre R+M – FWI(R+M) y LD(R+M)- se obtiene una diferencia sistemática que indica un mayor riesgo sobre R+M, el cual sería subestimado si sólo se usan estimaciones sobre R. Esto se ilustra en el Gráfico 5: el eje horizontal mide las PDs promedio por entidad para el modelo LD(R) y el eje vertical las mide para el modelo FWI (R+M) y para el modelo LD(R+M). Si las estimaciones fueran similares, los puntos se ubicarían sobre una línea diagonal. Se observa que los puntos correspondientes a los dos modelos que se miden en el eje vertical, pero especialmente el FWI, tienden a estar por sobre la línea, es decir, estiman mayor riesgo que LD(R).

Gráfico 5: Comparación del riesgo promedio por entidad con distintos modelos



El mismo resultado cualitativo se obtiene si las PDs promedio se calculan como un promedio ponderado por los montos de exposición en vez de un promedio simple (las diferencias se agrandan en algunos casos como resultado de los diferentes tamaños de las exposiciones).

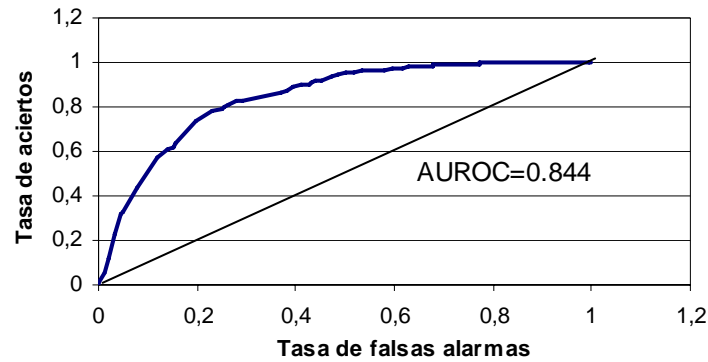
También se observa que el modelo FWI estima sistemáticamente mayor riesgo que el LD, aunque ambos se aplican a R+M, lo que implica que el hecho de omitir la información de M al estimar el modelo de scoring introduce más discrepancia (sobre la PD promedio) que el hecho de omitir la población M al aplicar un modelo ya estimado.

No se observa una regularidad entre las diferencias de PD estimada y el tamaño de la entidad o la proporción de M sobre la cartera total (los guiones del Gráfico 5 muestran dicha proporción). Ello está indicando el valor de la información procesada en el modelo FWI.

Queda entonces por evaluar si los costos de la estimación del modelo FWI (incluyendo el modelo de imputación intermedio) se justifican en vistas de la ganancia aparente en la estimación de riesgo, o bien si se puede introducir alguna corrección sobre el modelo LD, cuyo procedimiento sería más sencillo.

Si se pretendiera usar un modelo “LD ajustado” en lugar de un modelo FWI, debiera controlarse que funcionara bien, tanto en cuanto al poder discriminatorio como a la calibración. Aquí, resulta interesante verificar que LD (que se estima sobre R) discrimina sobre la población M (en particular sobre M1 considerando nuevamente la peor situación del deudor), tal como muestra el Gráfico 7. En efecto, se obtiene AUROC=0.844, KS= 0.538, y cabe remarcar que la población M1 no formó parte del desarrollo del modelo LD por lo que es un resultado destacable.

Gráfico 6: Curva ROC para modelo LD sobre población M1



En consecuencia, el ajuste sería necesario sólo desde el punto de vista de la calibración. Los Gráficos 7 y 8 ilustran la diferencia de tasa de defaults estimadas con el modelo LD para cada rango de score, versus la “observada” (en rigor, imputada por peor comportamiento para M1). El primer gráfico muestra las diferencias de las tasas de default en porcentajes y el segundo muestra las mismas sujetas a una transformación logit. Es interesante ver en éste último que, dado el patrón lineal observado, una transformación lineal del score podría resolver la cuestión de calibración, llevando la TD observada a la esperada. Este proceso se conoce en la jerga del scoring como realineación de un modelo.

Gráfico 7: tasa de default (TD) estimada por LD y “observada” – en %

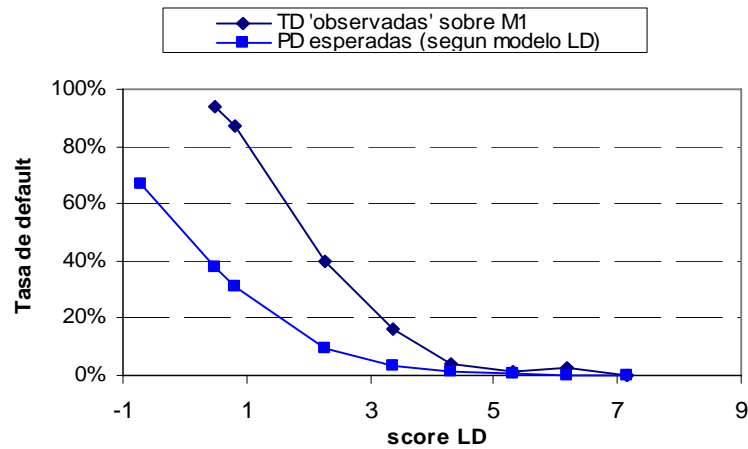
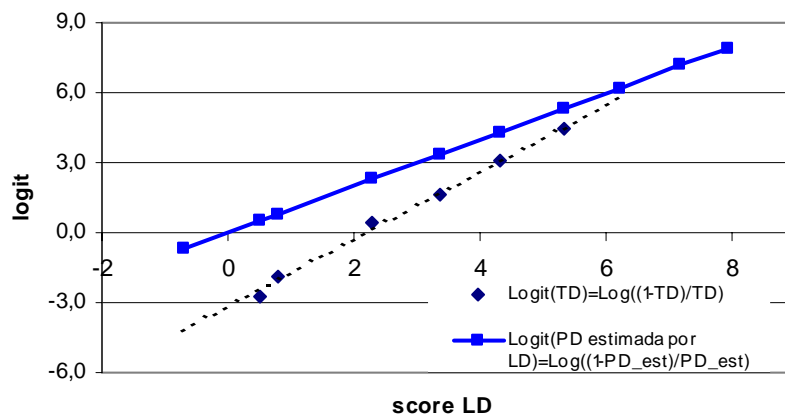


Gráfico 8: tasa de default (TD) estimada por LD y “observada” – en escala logit



Sólo a efectos ilustrativos, se calculó dicha realineación, la cual adoptaría la siguiente forma (ajustando sobre M1 un modelo logit entre el peor comportamiento en otros bancos –como variable dependiente- y el score LD como única variable independiente):

$$\text{Logit}' = -3.36 + 1.40 * \text{score_LD}$$

Alternativamente, si creemos que el modelo de imputación IMP es mejor estimador del verdadero riesgo, desearemos ajustar las tasas de default por rango de score a las que surgen de éste. El resultado sería más acorde al objetivo de no sobreestimar el riesgo. En ese caso, se obtendría:

$$\text{Logit}' = -0.32 + 0.96 * \text{score_LD}$$

No obstante que en este caso se llegue a la conclusión de que podría encontrarse un ajuste sencillo, debe recalarse que, para calibrarlo, es necesario haber analizado y estimado el sesgo, como se hace en este trabajo. Luego, es preciso suponer que el proceso de faltantes y el riesgo de los mismos en términos relativos se sostienen en períodos posteriores al ajuste, lo cual permitiría mantenerlo. Esta problemática es general para casos con faltantes de comportamiento significativos, pero los resultados dependerán del proceso subyacente, que es específico y local.

7. Conclusiones

Al igual que en otros trabajos, es interesante observar que los resultados de los modelos de scoring desarrollados sobre una base de información crediticia pública son muy buenos, a pesar de la selección limitada de variables explicativas. Por ejemplo, no se ha usado información sobre sector de actividad, ratios financieros u otra información de balance³⁵. Claramente, los bancos usan estas variables, entre otras, algunos con herramientas de scoring o bien con sistemas basados en el juicio de expertos. Sin embargo, modelar el riesgo de crédito con un conjunto de información de la CD, aún tan limitado, podría ser una herramienta poderosa especialmente para reguladores y supervisores, de acuerdo a los estadísticos de performance obtenidos en este estudio. En esa línea podemos prever distintos usos: (i) evaluar la regulación y tomar decisiones informadas sobre ella; (ii) usar los resultados como un punto de referencia contra el cual comparar los modelos desarrollados por los bancos y (iii) hacer ajustes a los central de riesgos de crédito, en particular, buscando evitar que no estén cubiertas todas las posibilidades de migración de un deudor y ello dificulte el seguimiento del comportamiento de los individuos.

En relación a este último punto, hemos mostrado que no es inocua la decisión de eliminar de la muestra aquellos deudores que no permanecen en la base de datos cuando las razones de su desaparición son desconocidas y no pueden ser modeladas. En consecuencia, los supervisores que estén desarrollando y/o supervisando modelos de riesgo crediticio deberían prestar especial atención a esta problemática. Hemos mostrado que el sesgo que puede introducirse es difícil de corregir y, aún intentando una corrección, no se podrá conocer la precisión de la misma con certeza. En consecuencia, debe subrayarse la importancia de asegurar que las bases de datos de riesgo de crédito eliminen cualquier “agujero” que dificulte el seguimiento en el comportamiento de los individuos. Este punto es de interés para la industria, para los supervisores y para los investigadores, ya que es frecuente encontrar modelos en los cuales un grupo de deudores ha sido eliminado de la muestra porque su información tiene algún problema, está incompleta, tiene mala calidad, o cualquier otro defecto.

³⁵ Hasta cierto punto las situaciones 1 y 2 capturan la información contenida en estas variables no utilizadas.

El trabajo logra inferir el riesgo de los deudores que desaparecen. Aparentemente, esta subpoblación tiene una tasa de default mayor que la población con comportamiento observado. Para medir este riesgo y para estimar correctamente un modelo de scoring sobre la población total, se utilizó, para los deudores cuyo comportamiento no es observable en una entidad, la información del comportamiento en otras entidades. En este sentido, la comparación de diferentes técnicas parece ser concluyente en indicar que la imputación directa de comportamiento como el peor estado observado en otras entidades sobreestima el riesgo.

El modelo que parece más preciso es el que se llamó “imputación con ponderación fraccionaria”, en el cual dichas imputaciones utilizan la información de comportamiento de central de riesgos en un modelo de regresión logit. Este enfoque es innovador en la literatura de scoring y es aparentemente preferible a la imputación directa. No obstante, no se puede ser concluyente respecto de la conveniencia de encararlo en todos los casos. El análisis particular de la CD en Argentina sugiere que posiblemente un ajuste en la calibración de un modelo más sencillo, que utiliza la técnica de listwise deletion, puede subsanar parte importante de las falencias. Que un ajuste así sea implementable dependerá, por supuesto, de que se mantenga en el tiempo el proceso de faltantes y el riesgo relativo de los mismos. Y dependerá también del caso específico bajo análisis.

Los datos utilizados en el estudio refieren a un período particularmente bueno en las tasas de incumplimiento locales, debiendo extenderse un estudio de un eventual ajuste a diferentes tipos de coyunturas en diferentes momentos del ciclo.

En este trabajo hemos hecho uso de la información de deudores-banco en otras entidades, lo cual se justifica por la correlación de comportamientos para una misma persona. Sin embargo, el efecto de esta correlación no se analizó en la estimación del modelo de scoring en sí, sino específicamente en la imputación de comportamiento. Este aspecto podrá ser encarado en futuras investigaciones.

Anexo 1: Scoring, inferencia de rechazos y sesgo de selección. Repaso de la literatura

La falta de información sobre el comportamiento de los casos cuyas solicitudes de crédito fueron rechazadas conduce usualmente a un sesgo en la muestra y, consecuentemente, en la predicción. La literatura respecto de cómo inferir la información de los rechazos para evitar o corregir ese sesgo es amplia y se extiende por más de 30 años³⁶. Parte de esa literatura trata la inferencia de rechazos como un caso especial del problema estadístico de datos faltantes (missing data - MD). Como se explica en el texto, el caso bajo estudio no es estrictamente de sesgo de selección; sin embargo, re tomó esta literatura como antecedente para analizar el problema de los faltantes.

Rubin (1976) desarrolló una topología de clasificación general de los métodos de MD que se cita frecuentemente y que puede aplicarse a la inferencia de rechazo. Según este enfoque, se denomina Y al conjunto total de datos, mientras que observados y faltantes se identifican como Y_{obs} e Y_{mis} , respectivamente. Una variable indicador R identifica datos observados y faltantes (R es una matriz con la misma dimensión que Y , con un 1 en cada celda cuyo elemento correspondiente de Y es observado y un cero en caso contrario). R puede ser tratada como un fenómeno probabilístico, con cierta distribución.

Se puede clasificar las distribuciones de R de acuerdo con la naturaleza de su relación con los datos. Se considera que los datos faltantes son MAR (*Missing At Random*, aleatoriamente faltantes) si Y_{obs} pero no Y_{mis} , ayudan a predecir R . Esto es, si $P(R|Y)=P(R|Y_{obs})$.

Dentro de este caso, si la distribución de R no depende tampoco de Y_{obs} , se denomina MCAR, por *Missing Completeley At Random*. En este caso $P(R|Y)=P(M)$. Un ejemplo de MCAR sería una encuesta en la cual se responde a “cara o seca”. MAR es un caso más realista que MCAR. Un ejemplo de MAR es una encuesta en la cual los sujetos con mayor ingreso son más proclives a no contestar una pregunta sobre ingresos. El proceso es MAR si los encuestados responden las otras variables y un modelo estadístico puede explicar la falta de respuesta, de forma tal que la probabilidad de falta de respuesta sea independiente de los valores no observados, controlando por el modelo. Por eso se dice en estos casos que la falta de datos es “ignorable”.

Si la distribución de R depende de Y_{mis} , se dice que la falta de datos es MNAR (*Missing Not At Random*) o “no-ignorable”. La probabilidad de faltante depende del valor de Y y tal dependencia subsiste aún controlando por Y_{obs} . El ejemplo típico es el de los pacientes bajo un tratamiento médico que dejan las pruebas por razones vinculadas con los resultados que se intenta medir.

La presencia o no de MNAR es, por definición, no demostrable usando los datos observados. Si se supone que existe MNAR, los datos faltantes no pueden predecirse sin sesgo y no hay un método general de corrección. Se hace necesario especificar un modelo de R que sea aproximadamente correcto (y que permita que las probabilidades de faltantes dependan de los mismos valores faltantes) y aún así la performance puede ser pobre.

Si agregamos a esta descripción teórica la consideración del conjunto de características que se utilizan para modelar el proceso de faltantes y/o estimar la calidad crediticia, tenemos una terminología más completa. En este sentido, se suele llamar X al conjunto de variables con datos observados para todos los registros e Y a las variables para las cuales faltan algunos registros. En inferencia de rechazos, puede pensarse X como las variables independientes disponibles, mientras que Y sería el comportamiento, el cual sólo es observable para los aceptados. Además se llama Z a un conjunto de características cuya definición es algo distinta para diferentes autores, pero cuya particularidad en todos los casos es que describe características adicionales a las disponibles para la decisión de aceptación/rechazo (o respuesta). Según Fogarty (2006), Z

³⁶ Una bibliografía no exhaustiva se puede encontrar en Carpenter (2006).

puede incluir o no al conjunto X. Según Schafer y Graham (2002), Z contiene variables que explican el proceso de faltantes que no están relacionadas con X ni con Y³⁷. Para un repaso del tema en la literatura véase Fogarty (2006).

Respecto de las técnicas para tratar con datos faltantes, la literatura floreció desde los 70s. Por ejemplo, los trabajos de Little (1992), Little y Rubin (1989), Rubin (1987) y Schafer (1997). Entre los métodos propuestos, posiblemente el procedimiento más sencillo sea el de “*listwise deletion*” o eliminación a lo ancho de la lista, en el que se eliminan las observaciones para las cuales falta algún dato. Una variante de este procedimiento consiste en descartar los registros que no cuentan con algún dato sólo en las muestras en las cuales la variable resultó ser necesaria (*pairwise deletion*). Estas técnicas pueden ser válidas sólo bajo MCAR o bajo MAR cuando los datos faltantes sean muy poco significativos. No obstante, es habitual su aplicación y ha sido por mucho tiempo la opción por default.

Otros métodos de inferencia de rechazos se denominan de imputación, parcelamiento y aumentación, e incluyen varias alternativas. Entre las alternativas de imputación, un enfoque habitual y sencillo consiste en imputar un único valor estimado a cada dato faltante. El valor estimado puede consistir en la media de los valores observados para la variable, el valor vecino, o un valor estimado en forma sistemática o estocástica. Se llama imputación “hot deck” cuando se imputan valores tomados de registros similares y completos del mismo conjunto de datos. Estos métodos de imputación simple fueron objeto de creciente crítica (Rubin y Schenker, 1986). Algunos autores los consideran aceptables cuando la proporción de valores faltantes es de hasta 5%³⁸.

Otra técnica sencilla de imputación consiste en buscar los rechazados con información negativa grave (aquella que haría que cualquier acreedor lo descartase) y reclasificarlos como aceptados con status de “default”, considerando no-defaults al resto. Así se reincorporan los rechazados a la muestra y se estima un nuevo modelo sobre la muestra ampliada. Alternativamente, se puede imputar como “default” a los solicitantes rechazados que obtuvieron los peores scores según un modelo de aceptación/rechazo, suponiendo que la proporción de malos entre los rechazados es la misma que entre los aceptados, y así reincorporar a la muestra a los rechazados para estimar el modelo. Este método intensifica la relación entre la variable indicadora de default y las variables X que se usaron en el modelo original y, por ende, distorsiona las covarianzas y sobreestima el poder predictivo del modelo. En general, las técnicas de inferencia de rechazo que tratan los valores imputados como si fueran valores observados subestiman la incertidumbre. Esta distorsión puede eliminarse si cada valor de Y se reemplaza no por una predicción en función de una regresión sino por la salida aleatoria de la distribución de Y dado X, sujeto a un error residual. Este método supone MAR y produce estimaciones insesgadas bajo MAR.

La técnica de parcelamiento segmenta a los deudores buenos, malos y rechazados por un score genérico o customizado y luego proyecta sobre los rechazados un ratio de default, extrapolado a partir de su comportamiento para los aceptados y suponiendo un ajuste que se estima estadísticamente o por expertos. Con esta clasificación se reincorporan los rechazados y se vuelve a correr la estimación. Esta técnica depende de la calidad del score utilizado en la extrapolación.

La técnica de aumentación consiste en dos pasos. En el primero, se reclasifica a los rechazados con información negativa grave como defaults. En el segundo paso se asignan ponderaciones a los solicitantes aceptados de tal manera que se sintetice una muestra que represente también a los solicitantes rechazados. Para derivar los ponderadores se estima un modelo de aceptación/rechazo. La inversa de la probabilidad estimada es igual al número de casos que representará cada caso aceptado en la muestra, esto es, se usa como ponderador en la estimación del modelo. Se supone que los solicitantes con menor probabilidad de ser aceptados tienen características

³⁷ Schafer y Graham (2002) y Fogarty (2006)

³⁸ Yarandi (2002)

más similares a los casos originalmente rechazados. Al asignárseles ponderadores altos a estos solicitantes, el modelo ponderado se acercaría al que hubiera sido estimado si los solicitantes rechazados se hubiesen incluido. A diferencia del parcelamiento, los rechazados que reingresan en la muestra son sólo los que entran como malos por tener información negativa grave.

Es muy importante revisar, una vez aplicado uno de estos métodos, cuantos “nuevos” malos³⁹ se han agregado a la muestra. La inferencia de rechazos trata de mejorar los datos conocidos (de malos y buenos) con los rechazados. Si muy pocos rechazados se reincorporan o es poco el impacto de la ponderación, básicamente se seguirá teniendo un modelo de buenos-malos. Por otra parte, si muchos rechazados se incorporan, la muestra puede sesgar el modelo hacia uno que distinga rechazos vs buenos, en lugar de malos vs buenos. Esto puede subestimar los verdaderos malos en la población, haciendo el modelo menos efectivo.

Uno de los métodos de corrección precursores más difundidos (que no necesita suponer MAR) es la corrección en dos etapas de Heckman (1979), la cual arroja estimaciones consistentes si la variable explicada es continua y los residuos tienen distribución normal. Sin embargo, hay evidencia de que este método puede incluso empeorar las estimaciones, en circunstancias ordinarias y aún cuando se satisfagan sus supuestos y requisitos formales (ver Stolzenberg y Relles 1997).

De ese modelo se desprendieron otros, que pueden aplicar a variables categóricas, y que típicamente involucran la estimación simultánea de dos regresiones probit. Así, trabajos como el de Meng y Schmidt (1985) dieron origen al modelo probit bivariado de selección muestral, el cual plantea un sistema de dos ecuaciones probit: una modela la decisión de selección y la otra el comportamiento crediticio⁴⁰. Ambas ecuaciones incluyen sendos errores, los cuales no serán independientes salvo que la selección fuese totalmente aleatoria. Por ende, la estimación de la ecuación de comportamiento sobre la muestra observable, por sí sola, será sesgada. El modelo supone que los errores tienen una distribución condicional normal bivariada; no supone que las distribuciones de aceptados y rechazados sean similares pero depende de que las ecuaciones de selección y de comportamiento estén plenamente especificadas.

Existen técnicas que utilizan Máxima Verosimilitud (ML, del inglés *Maximum Likelihood*) que tratan los datos faltantes como variables aleatorias a ser removidas de la función de verosimilitud como si no hubieran sido muestreadas, en lugar de eliminar o completar los casos faltantes. Bajo MAR tiene aceptación bastante generalizada las virtudes de obtener inferencias de una función de ML⁴¹. La distribución marginal de los datos observados provee la verosimilitud correcta de los parámetros desconocidos, si el modelo para los datos completos es suficientemente bueno. Excepto en casos especiales, el cómputo de las estimaciones de ML requiere iteración, por ej., el método de maximización de expectativas (EM). Este es un algoritmo general para resolver problemas de datos faltantes, formalizado por Dempster, Laird et al (1977). Los estimadores ML no tienen sesgo sustancial bajo MCAR y MAR, pero sí bajo MNAR. ML descansa en algunos supuestos cruciales. En primer lugar, supone una muestra suficientemente grande para que los estimadores sean aproximadamente insesgados y se distribuyan normalmente. Segundo, la función de distribución proviene de un modelo paramétrico para los datos completos. En algunos casos el método no es robusto a alejamientos de los supuestos. También se critica este método porque supone que la distribución de las variables observadas es idéntica para aceptados y rechazados, y que estas variables no están correlacionadas con otras características que son sólo observables para los aceptados⁴².

³⁹ Usualmente es el grupo menos frecuente, pero podrían llegar a serlo los buenos (por ejemplo es el caso de los scores de cobranzas utilizados en estadíos de mora avanzado)

⁴⁰ Este enfoque se exploró en Balzarotti, Gutiérrez Girault et al (2006).

⁴¹ Gold et. al (2003)

⁴² Schafer y Graham (2002).

En el método de imputación múltiple (“MI”, Rubin, 1987) cada valor faltante se reemplaza por $m > 1$ valores estimados para representar la incertidumbre sobre el valor correcto a imputar. Una vez completados los datos faltantes, cada uno de los m conjuntos de datos es analizado utilizando técnicas para datos completos y los resultados se combinan para producir estimaciones finales. Si bien los análisis MI en general se han aplicado bajo el supuesto de que los datos faltantes son MAR, también hay aplicaciones MNAR. Es una ventaja importante en MI la flexibilidad que se deriva de la posibilidad de usar diferentes modelos para la imputación y para la estimación del modelo (si bien esta flexibilidad puede introducir falta de robustez). Por ejemplo, con MI el proceso de imputación puede usar datos adicionales y así incrementar el poder predictivo. Las propiedades de MI cuando el modelo de imputación y el de análisis difieren fueron estudiadas académicamente por Meng (1994) y Rubin (1996) y desde el punto de vista práctico por Collins, Schafer y Kam (2001)⁴³. Numerosos autores⁴⁴ sostienen que MI es el “estado del arte” para el manejo de datos faltantes, tanto por sus propiedades teóricas como por haber sido implementada en aplicaciones informáticas de amplia difusión. En cambio, otros critican esta técnica en algunas situaciones, por producir estimaciones sesgadas y varianzas que no son consistentes con el diseño de la muestra, conduciendo a problemas de inferencia estadística. También se critica la imputación de valores continuos a variables discretas.

Para respetar la naturaleza discreta de la variable a imputar, se propone una forma de imputación con ponderadores fraccionales (FWI, por *fractionally weighted imputation*, Rubin y Fay 1996) en la que los valores imputados son elegidos al azar entre datos observados para registros que comparten ciertas características con los registros con datos faltantes. Para determinar los ponderadores o “fracciones de imputación” se proponen distintos mecanismos⁴⁵. La muestra con los valores imputados ponderados se trata en un único análisis, en lugar de m análisis. FWI produce estimaciones con menor varianza que MI, para el mismo número de imputaciones. También permite la construcción de intervalos de confianza basados en aproximaciones normales, en lugar de los procedimientos más complejos que requiere MI.

Autores como Cranmer (2007) proponen que el sistema de ponderaciones debe hacer el mejor uso posible de los datos disponibles, en especial de las variables adicionales de que se disponga, y que los ponderadores deben reflejar el grado de afinidad entre el registro “receptor” del dato y los registros “donantes”. Cuando los conjuntos de datos son grandes y los datos faltantes son muchos, el proceso se hace algo tedioso y computacionalmente intensivo. Para hacerlo más eficiente, se propone aumentar las exigencias de afinidad o bien seleccionar aleatoriamente un número de donantes mucho menor que el total, con muestreo sin reposición (modelo FEFI⁴⁶).

El uso de información adicional para el proceso de inferencia de rechazos ha sido estudiado por algún tiempo, no sólo en el marco de FWI. Hand y Henley (1997) proponen tres métodos para usar información adicional (que llaman “calibración de la muestra”). Collins et al (2001) estudian el problema de datos faltantes (no específicamente asociado al crédito) y argumentan a favor de estrategias inclusivas –que hagan uso de variables auxiliares-. Más recientemente, Ash y Meester (2002) discuten el problema de rechazados de crédito y también argumentan a favor del uso de data adicional, específicamente el comportamiento en una central de riesgos de aceptados y rechazados, al final del período de observación⁴⁷. El mayor obstáculo reside en conseguir esta información para una parte significativa de los rechazados y los aceptados. Normalmente habrá un costo de acceder a esa información y cierto rezago. Por otra parte, ésta es una información

⁴³ Schafer y Graham (2002).

⁴⁴ Por ejemplo Cranmer (2007)

⁴⁵ Por ejemplo Fuller y Kim (2005) introducen un mecanismo en gran parte intuitivo en el cual el ponderador es la inversa del número de “donantes”.

⁴⁶ *Fully Efficient Fractional Imputation*, imputación fraccional completamente eficiente), Kim y Fuller (1999)

⁴⁷ Estos autores llaman al uso de información crediticia con otros acreedores *cohort performance*.

muy poderosa sobre el deudor rechazado y permite la inferencia de comportamiento sobre la base de información de comportamiento con otros acreedores, en el mismo período de interés, por lo cual es menos subjetiva.

ML y MI se están convirtiendo en estándares gracias a los desarrollos tecnológicos y los métodos de simulación, incluso en implementaciones comerciales. Sin perjuicio de ello, se siguen ensayando enfoques alternativos. Un caso es el análisis bayesiano, en el cual toda la evidencia sobre los parámetros se resume en una función de verosimilitud. Al igual que con ML, la forma paramétrica supuesta para el modelo es crucial; si el modelo es inexacto, la distribución posterior puede brindar una imagen irreal del estado del conocimiento sobre los parámetros. El análisis bayesiano requiere una distribución a priori para los parámetros desconocidos que es subjetiva y a veces se critica por “a-científica”, disminuyendo su influencia al aumentar el tamaño muestral.

Anexo 2: Características de las poblaciones R y M

Esta sección apunta a caracterizar las poblaciones R y M mediante el análisis de las distribuciones de las principales variables, destacando aquellas que muestran una mayor diferencia entre ambas poblaciones.

En primer término se analizan las variables que fueron seleccionadas para los modelos de probabilidad de default, y luego se analizan otras que presentan mayor capacidad de discernir entre ambas poblaciones.

Variable: cant_sit3 (# bancos en situación 3 (a nivel deudor))

Población	0	1	2	3	4	5	6+
R	95,3%	3,4%	0,6%	0,2%	0,2%	0,1%	0,3%
M	93,2%	4,1%	1,1%	0,4%	0,3%	0,2%	0,7%

Variable: v3_rating (situación a la fecha de observación)

Población	1	2
R	97.58%	2.41%
M	96.04%	3.95%

Variable: D_deumin_sit0605eq3 (mínima situación 6 meses atrás es 3+ , a nivel deudor)

Población	no	si
R	99.29	0.71
M	99.24	0.76

Variable: D_NEW_cuit (Dummy Inicio de actividades reciente ,CUIT >=702)

Población	no	si
R	83.46	16.54
M	82.23	17.77

Variable: Grupo_h (Agrupación de bancos, reagrupado)

Población	Minoristas grandes	Publicos Nacionales	Resto
R	48.48	12.56	38.96
M	34.97	8.47	56.57

Variable: asist_14 (Dummy: posee otros créditos intermediación financiera)

Población	no	si
R	72.95	27.05
M	90.30	9.70

Variable: asist_15 (Dummy: posee bienes en locación financiera (leasing))

Población	no	si
R	85.38	14.62
M	96.86	3.14

Variable: asist_adel (Dummy: posee asistencias de tipo adelanto)

Población	no	si
R	49.45	50.55
M	60.63	39.37

Variable:asist_hip (Dummy: posee asistencias de tipo hipotecario)

Población	no	si
R	88.60	11.40
M	93.54	6.46

Variable:deu_asist_pren (Cantidad de líneas prendario (a nivel deudor)

Población	0	1	2	3	4	5+
R	74.31	19.99	4.45	0.92	0.23	0.10
M	84.28	12.32	2.58	0.61	0.18	0.03

Variable:d_garan_A (Posee garantías de tipo "A")

Población	no	si
R	94.67	5.33
M	92.99	7.01

Variable:deu_max_situ_6matras (Peor situación 6 meses atrás)

Población	0 (faltante)	1	2 ó peor
R	22.21	68.37	9.42
M	42.22	45.91	11.87

Variable:D_deudor_empeora_sit (Dummy: empeoró su situación desde hace 6 meses, a nivel deudor)

Población	no	si
R	75.56	24.44
M	55.85	44.15

Variable:deu_asist_perst (Cantidad de líneas tipo préstamos personales o tarjetas de crédito)

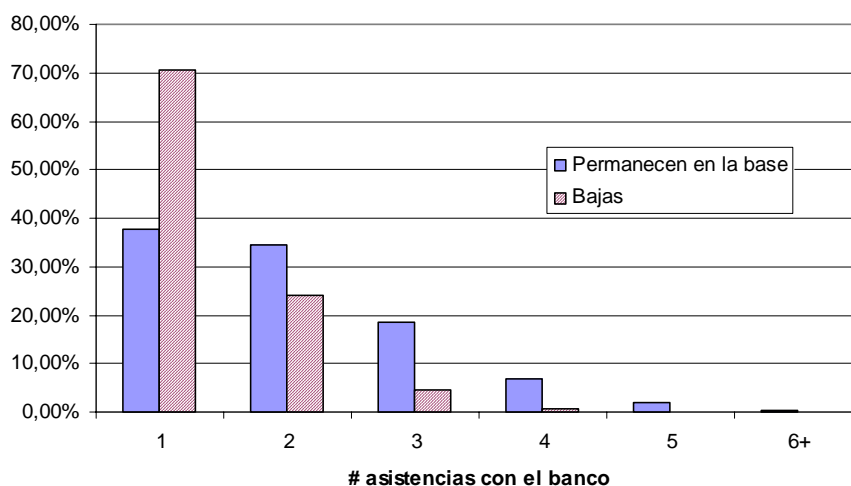
Población	0	1	2	3	4	5+
R	69.84	25.34	4.19	0.51	0.11	0.01
M	76.13	19.67	3.85	0.24	0.11	0.00

La variable que mayor diferencia muestra entre las poblaciones R y M es al cantidad total de asistencias (es decir distintos productos financieros) con el banco. Cuando el deudor tiene poca relación con el banco (dado por un bajo número de asistencias), la probabilidad de no ser informado se incrementa considerablemente.

Variable:# asistencias con el banco

Población	1	2	3	4	5	6+
R	37,7%	34,6%	18,6%	6,8%	1,9%	0,4%
M	70,6%	24,0%	4,7%	0,7%	0,1%	0,0%

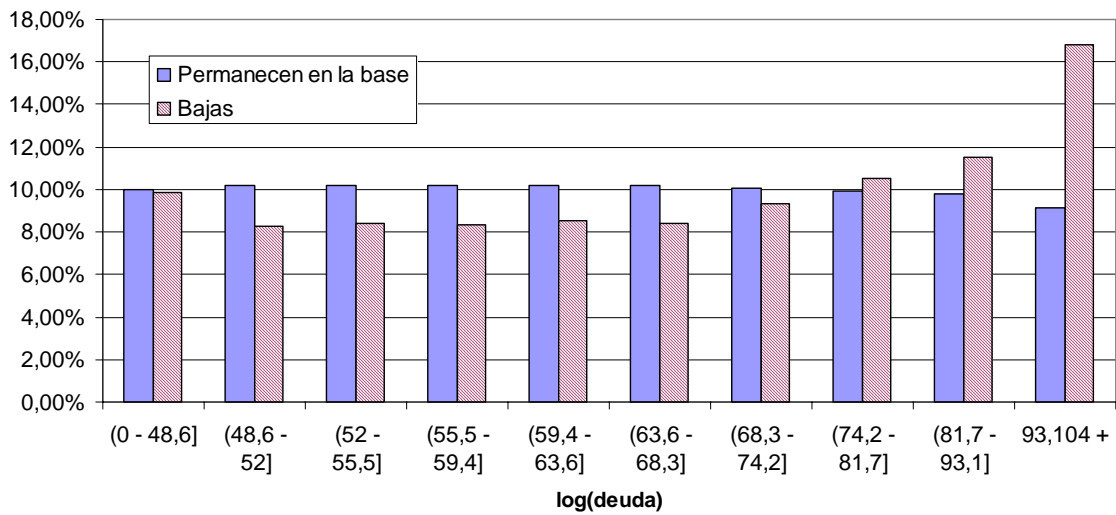
Distribución de # asistencias con el banco



Otra variable que muestra un comportamiento digno de ser interpretado es la deuda total con el sistema. En este caso, se observa que la población M presenta una concentración más alta de deuda total con el sistema financiero

Variable: Rango log(deuda sistema)

Población	(0 - 48,6]	(48,6 - 52]	(52 - 55,5]	(55,5 - 59,4]	(59,4 - 63,6]	(63,6 - 68,3]	(68,3 - 74,2]	(74,2 - 81,7]	(81,7 - 93,1]	93,1 +
R	10,0%	10,2%	10,2%	10,2%	10,2%	10,2%	10,1%	9,9%	9,8%	9,2%
M	9,9%	8,3%	8,4%	8,4%	8,5%	8,4%	9,3%	10,5%	11,5%	16,8%



Anexo 3: Variables explicativas y relaciones bivariadas con la variable default

Variable:	cant_sit3 (# bancos en situación 3 (a nivel deudor))						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	360	29911	30271	73,6%	95,6%	95,3%	1,19%
1	46	1021	1067	9,4%	3,3%	3,4%	4,31%
2-5	58	279	337	11,9%	0,9%	1,1%	17,21%
6+	25	67	92	5,1%	0,2%	0,3%	27,17%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	Situ (Situación BCRA a la observación)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
1	402	30598	31000	82,2%	97,8%	97,6%	1,30%
2	87	680	767	17,8%	2,2%	2,4%	11,34%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	D_deumin_sit_6m_atras_eq3 (Dummy Peor Situación BCRA 6 meses atrás es 3 ó peor)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	458	31082	31540	93,7%	99,4%	99,3%	1,45%
1	31	196	227	6,3%	0,6%	0,7%	13,66%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	D_NEW_cuit (Dummy Inicio de actividades reciente ,CUIT >=702)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	374	26138	26512	76,5%	83,6%	83,5%	1,41%
1	115	5140	5255	23,5%	16,4%	16,5%	2,19%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	asist_14 (Dummy: posee otros créditos intermediación financiera)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	323	22848	23171	66,1%	73,0%	72,9%	1,394%
1	166	8430	8596	33,9%	27,0%	27,1%	1,931%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,539%

Variable:	asist_15 (Dummy: posee bienes en locación financiera (leasing))						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	466	26658	27124	95,3%	85,2%	85,4%	1,72%
1	23	4620	4643	4,7%	14,8%	14,6%	0,50%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	asist_adel (Dummy: posee asistencias de tipo adelanto)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	234	15476	15710	47,9%	49,5%	49,5%	1,49%
1	255	15802	16057	52,1%	50,5%	50,5%	1,59%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	asist_hip (Dummy: posee asistencias de tipo hipotecario)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	380	27764	28144	77,7%	88,8%	88,6%	1,35%
1	109	3514	3623	22,3%	11,2%	11,4%	3,01%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	d_garan_A (Posee garantías de tipo "A")						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	427	29648	30075	87,3%	94,8%	94,7%	1,42%
1	62	1630	1692	12,7%	5,2%	5,3%	3,66%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	D_deudor_empeora_sit (Dummy: empeoró su sit. desde hace 6 meses, a nivel deudor)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	361	23643	24004	73,8%	75,6%	75,6%	1,50%
1	128	7635	7763	26,2%	24,4%	24,4%	1,65%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	deu_asist_perst (Cantidad de líneas tipo préstamos personales o tarjetas de crédito)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	409	21776	22185	83,6%	69,6%	69,8%	1,84%
1	62	7987	8049	12,7%	25,5%	25,3%	0,77%
2	18	1314	1332	3,7%	4,2%	4,2%	1,35%
3+	0	201	201	0,0%	0,6%	0,6%	0,00%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	nbancos_default_6m_atras (Cantidad de bancos en situación 3+, 6 meses atrás)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	370	29530	29900	75,7%	94,4%	94,1%	1,24%
1	48	1229	1277	9,8%	3,9%	4,0%	3,76%
2-3	42	350	392	8,6%	1,1%	1,2%	10,71%
4+	29	169	198	5,9%	0,5%	0,6%	14,65%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	deu_asist_pren (Cantidad de líneas prendario (a nivel deudor))						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	401	23204	23605	82,0%	74,2%	74,3%	1,70%
1	82	6268	6350	16,8%	20,0%	20,0%	1,29%
2	6	1409	1415	1,2%	4,5%	4,5%	0,42%
3-4	0	366	366	0,0%	1,2%	1,2%	0,00%
5+	0	31	31	0,0%	0,1%	0,1%	0,00%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	r_finan_default (ratio monto en default/total deuda)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0	360	29913	30273	73,6%	95,6%	95,3%	1,19%
<0.018	16	302	318	3,3%	1,0%	1,0%	5,03%
0.018-0.12	12	306	318	2,5%	1,0%	1,0%	3,77%
0.12-0.28	15	303	318	3,1%	1,0%	1,0%	4,72%
0.28-0.63	39	279	318	8,0%	0,9%	1,0%	12,26%
>0.63 (max=1)	47	175	222	9,6%	0,6%	0,7%	21,17%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Variable:	deu_max_situ_6matras (Peor situación 6 meses atrás)						
Valores - rango	#D	#ND	#Tot	%D	%ND	%Tot	TD(%)
0 (faltante)	84	6970	7054	17,2%	22,3%	22,2%	1,19%
1	244	21476	21720	49,9%	68,7%	68,4%	1,12%
2 ó peor	161	2832	2993	32,9%	9,1%	9,4%	5,38%
Total	489	31278	31767	100,0%	100,0%	100,0%	1,54%

Anexo 4: Medidas estadísticas de discriminación y calibración

Índice ROC: Dado C, la tasa de aciertos (*hit rate*) y tasa de falsa alarma (*false alarm rate*) se calculan en función de predicciones correctas e incorrectas de que los deudores hagan default.

La tasa de aciertos, HR (C), y la tasa de falsa alarma, FARC (C), son las siguientes:

$$FARC(C) = \frac{F(C)}{N_{ND}}$$
$$HR(C) = \frac{H(C)}{N_D}$$

Donde, H (C) es el número de deudores que hacen default predichos correctamente a partir del punto de corte C y N_D el número total de deudores que en la muestra hacen default. F(C) representa el número de falsas alarmas, es decir el número de deudores que no hicieron default clasificados incorrectamente como deudores que harían default al usar C como punto de corte y N_{ND} es el número total de deudores que no hacen default.

Gráficamente, la curva ROC muestra en una línea los puntos que corresponden a computar la tasa de aciertos y la tasa de falsa alarma para los valores de corte contenidos en el rango del score de rating estimado.

Estadístico de Kolmogorov-Smirnov (K-S)

El estadístico de Kolmogorov-Smirnov (KS) mide la máxima diferencia absoluta entre dos distribuciones acumuladas.

Sean $F_D(s)$ y $F_{ND}(s)$ las distribuciones acumuladas de score de los defaults y de los no-defaults respectivamente. Entonces

$$KS = \max_s \{ |F_D(s) - F_{ND}(s)| \}$$

Test de Hosmer-Lemeshow

Una condición para la validez de la mayoría de los tests de bondad de ajuste para el tipo de modelos considerados (Pearson y Deviance) es que haya suficiente replicación dentro de las subpoblaciones definidas por las combinaciones de valores de las variables independientes. En el caso en que se utilizan variables continuas como predictores, esta condición claramente no puede cumplirse.

En estos casos suele ser utilizado el Test de Hosmer-Lemeshow (Hosmer and Lemeshow, 1989). El mismo consiste en el siguiente procedimiento: se ordenan las observaciones en forma creciente por la probabilidad estimada del evento p. Las observaciones son entonces agrupadas usualmente en diez rangos aproximadamente equi-poblados (se tiene en cuenta si existen *ties*, es decir valores repetidos -para lo cual existen leves variantes de implementaciones según el paquete de software utilizado-).

Habiendo clasificado las observaciones en estos grupos (usualmente $g=10$) se calcula el siguiente estadístico Chi-cuadrado:

$$\chi_{HL}^2 = \sum_1^g \frac{(O_i - N_i \bar{\pi}_i)^2}{N_i \bar{\pi}_i (1 - \bar{\pi}_i)}$$

donde:

- N_i es la frecuencia total de registros en el i-ésimo grupo,
- O_i es la frecuencia de eventos observados en el i-ésimo grupo, y
- $\bar{\pi}_i$ es el promedio de la probabilidad de evento estimada en el i-ésimo grupo,

Mediante simulación, HL demuestran que este estadístico presenta una distribución Chi-cuadrado de $g-2 = 8$ grados de libertad. Un valor grande del estadístico (es decir un p-value bajo) es indicio de falta de ajuste.

Este test asume que todos los registros tienen ponderación uno. En nuestro caso, esto no se cumple para el modelo FWI dado que los registros que fueron inferidos por falta de información de comportamiento presentan ponderaciones fraccionales.

La adaptación del test (y del cálculo del estadístico) se llevó a cabo de la siguiente manera: por un lado, los percentiles fueron calculados tomando en cuenta el acumulado de registros ponderados (pero forzando que los dos registros generados a partir de la inferencia de un faltante caigan en el mismo rango); las magnitudes N_i y O_i fueron calculadas como sumas ponderadas y $\bar{\pi}_i$ fue calculado como un promedio ponderado.

Dado que la distribución utilizada en el test mencionado (y especialmente los grados de libertad) provienen de análisis de simulación, es aconsejable llevar a cabo análisis de simulación para verificar que, bajo la modificación mencionada al cálculo del estadístico, se mantienen las hipótesis acerca de su distribución, lo cual formará parte de futuras líneas de trabajo.

Anexo 5: Modelo de imputación (IMP)

Variable	Valor / rango	Estimación
Constante		-3,556 ***
Cantidad de bancos en situación 3+	[0-10]	-0,163 ***
	1	1,911 ***
Situación al momento de observación	2	0,000 ***
Dummy: mínima situación 6 meses atrás es 3+ (a nivel deudor)	no	1,437 ***
	sí	0,000 ***
Dummy Inicio de actividades reciente (CUIT >=702)	no	0,734 ***
	sí	0,000 ***
Agrupación por tipo de banco ("Grupo homogéneo")	Resto	0,531 ***
	Minoristas grandes	1,081 ***
	Públicos nacionales	0,000 ***
Dummy: posee otros créditos intermediación financiera	no	0,679 ***
	sí	0,000 ***
Dummy: posee bienes en locación financiera	no	-1,134 ***
	sí	0,000 ***
	no	0,515 ***
Dummy: posee asistencias de tipo adelanto	sí	0,000 ***
Dummy: posee asistencias de tipo hipotecario	no	0,409 **
	sí	0,000 **
Posee garantías de tipo "A"	no	0,727 ***
	sí	0,000 ***
Cantidad de bancos en situación 3+, 6 meses atrás	[0-14]	0,153 **
Cantidad de bancos en situación 2 en período de performance.	[0-9]	0,958 ***
	0 (sin otro banco)	2,315 ***
Peor situación en otros bancos, en período de performance.	1	3,938 ***
	2	1,422 ***
	3+	0,000 ***

Bibliografía

Balzarotti, V.; Castro, C. y Powell, A. (2002). "Reforming capital Requirements in Emerging Countries". *Business School Working Papers XIX*. Universidad Torcuato Di Tella.

Balzarotti, V.; Castro, C. y Powell, A. (2004). "Capital Requirements in Emerging countries: Calibrating Basel II using Historical Argentine Credit Bureau Data and Credit Risk+". *Business School Working Papers*. Universidad Torcuato Di Tella.

Balzarotti, V., M. A. Gutiérrez Girault y V. A. Vallés (2006). "Modelos de Scoring Crediticio con Muestras Truncadas y su Validación", Banco Central de la República Argentina. Documento de Trabajo 2006/3.

BCBS (Junio, 2004), Basel Committee on Banking Supervision, BIS "International Convergence of Capital Measurement and Capital Standards. A Revised Framework". Actualizado junio 2006.

BCBS (Mayo, 2005), Basel Committee on Banking Supervision, Bank of International Settlements. "Studies on the Validation of Internal Rating Systems". *Working Paper 14*.

Carpenter, J. (2006). "Annotated bibliography on missing data". En: <http://www.lshtm.ac.uk/msu/missingdata/biblio.html>.

Collins, L. M.; Schafer, J. L. y Kam, C. M. (2001). "A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures". *Psychological Methods* Vol. 6, No. 4: 330-351.

Cranmer, S. J. (2007). "Hot Deck Imputation for Discrete Data, Working Draft". Invited talk given at the Research Workshop in Applied Statistics in the Institute for Quantitative Social Science, Harvard University. En: <http://courses.gov.harvard.edu/gov3009/spring07/HotDeck.pdf>

Crook, J. (2002). "Adverse Selection And Search In The Bank Credit Card Market". Credit Research Centre, University of Edinburgh.

Fogarty, D. J. (2006). "Multiple imputation as a missing data approach to reject inference on consumer credit scoring". *Interstat*. <http://interstat.statjournals.net/YEAR/2006/articles/0609001.pdf>

Fuller, W. y Kim, J. K. (2005). "Hot deck imputation for the response model". *Statistics Canada* 31(2): 139-149.

Gold, M. S.; Bentler, P. M. y Kim, K. H. (2003). "A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data". *Structural Equation Modeling* 10(1): 47-79.

Hand, D. J. y Henley, W. E. (1997). "Statistical Classification Methods in Consumer Credit Scoring: A Review". *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 160,3: 523-541.

Johnston, J y DiNardo, J (1997) *Econometric methods*, McGraw-Hill.

Kim, J.K. y Fuller, W. A. (1999). "Jackknife variance estimation after hot deck imputation". *ASA Proceedings of the Section on Survey Research Methods*, 825-830. American Statistical Association, Alexandria, VA.

- Little, R. J. y Rubin, D. B. (1989). "Statistical Analysis with Missing Data". New York, Wiley.
- Powell, A., N. Mylenko, M. Miller, and G. Majnoni (2004). "Improving Credit Information, Bank Regulation and Supervision: On the Role and Design of Public Credit Registries." World Bank Policy Research Working Paper 3443.
- Rubin, D. y R. Fay (1996). "On Variance Estimation With Imputed Survey Data: Rejoinder" *Journal of the American Statistical Association*. Vol. 91, No. 434. (Jun., 1996), pp. 515-519.
- Rubin, D. B. (1976). "Inference and Missing Data". *Biometrika* 63: 581-92.
- Rubin, D. y Schenker, N. (1986). "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse". *Journal of the American Statistical Association* 81(394): 366-374.
- Rubin, D.B. (1987). "Multiple Imputation for Nonresponse in Surveys". J. Wiley & Sons, New York.
- Schafer, J. L. y Graham, J. W. (2002). "Missing Data: Our View of the State of the Art". *Psychological Methods* Vol. 7, No. 2: 147-177.
- Schechtman, R.; Salomão García, V.; Mikio Koyama, S. y Cronemberger Parente, G. (2004). "Credit Risk Measurement and the Regulation of Bank Capital and Provision Requirements in Brazil-A Corporate Analysis". *Working Paper Series* 91. Banco Central Do Brasil.
- Stolzenberg, R. M. y D. Relles (1997). "Tools for Intuition about Sample Selection Bias and its Correction". *American Sociological Review* 62(3): 494-507.
- Trucharte Artigas , C. (2004), A Review of Credit Registers and their Use for Basel II, Bank for International Settlements, Financial Stability Institute Occasional Papers, Basel, September.
- Verstraeten, G. y Van den Poel, D. (2004). "The Impact of Sample Bias on Consumer Credit Scoring: Performance and Profitability". *Working Paper Universiteit Gent*.
- Yarandi, H.N. (2002). "Handling Missing Data with Multiple Imputation using PROC MI in SAS". *Users Group Annual Meetings*, ST14. Institute for Advanced Analytics, North Carolina State University.