

Modelos de scoring crediticio con
muestras truncadas y su validación

Verónica Balzarotti / Matías Gutiérrez Girault
Verónica Vallés
BCRA

Mayo de 2006



ie | BCRA

Investigaciones Económicas
Banco Central
de la República Argentina

Banco Central de la República Argentina
ie | Investigaciones Económicas

Mayo, 2006
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 deseé efectuarse a estos Documentos deberá contar con la autorización del o los autores.

Modelos de scoring crediticio con muestras truncadas y su validación

Verónica Balzarotti, Matías Gutiérrez Girault y Verónica Vallés

El principal objetivo de este trabajo es desarrollar una metodología de scoring crediticio para los deudores bancarios comerciales en Argentina, basándose en información disponible en la Central de Deudores pública del Banco Central de la República Argentina (“CD”), como herramienta referencial para evaluar el riesgo crediticio en los bancos locales. La experiencia previa en este campo ha mostrado resultados prometedores¹. En este trabajo nos concentramos en dos aspectos innovadores: en primer lugar, el potencial sesgo introducido por el hecho de que un número considerable de deudores son removidos de la base de datos sin que se pueda conocer el motivo, y en segundo término, la aplicación de técnicas de validación a los modelos obtenidos, siguiendo las propuestas de un documento publicado recientemente por el BCBS².

La CD es una fuente valiosa de información para evaluar riesgo crediticio, aunque impone ciertas limitaciones. En particular, la base de datos comprende deudores que ya son clientes bancarios, mientras que no se registran los datos de potenciales clientes. Por lo tanto, un sistema de scoring crediticio desarrollado usando la base de datos de la CD está sujeto a los efectos de este truncamiento. El sesgo resultante es bien conocido en la literatura como “sesgo de selección”³. En el contexto de este trabajo, no obstante, dicho sesgo no es de mayor preocupación, en la medida en que el objetivo es estimar el riesgo crediticio de un portafolio bancario desde un punto de vista regulatorio, en comparación con el caso en el cual el objetivo fuera desarrollar un sistema para tomar decisiones de otorgamiento de crédito. Existe otra limitación en la base de datos de interés particular en este trabajo, y que viene dada por el hecho de que un grupo de deudores regularmente desaparece de la CD por motivos que no quedan registrados. 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 pueden estar detrás de la desaparición de un deudor de la base de datos, además de la posibilidad de que se trate simplemente de un error de la información. Si el monto de deudores removidos es considerable y el proceso de exclusión no es aleatorio, puede existir un sesgo en las estimaciones del modelo. Trataremos de probar si este sesgo existe y su impacto en el caso argentino. En función de ello formularemos una propuesta para introducir algunos cambios en la base de datos argentina, al tiempo que exploramos metodologías para corregir el sesgo.

Adicionalmente, ha existido cierta falta de consenso en la literatura sobre las metodologías de validación a aplicar a los modelos de scoring crediticio y sistemas de rating, hasta que esta brecha fue zanjada en buena medida por la publicación del documento de trabajo del Comité de Basilea *Studies on the Validation of Internal Rating Systems*⁴. En este trabajo aplicaremos a los modelos

¹ Balzarotti, V., C. Castro y A. Powell (2004) y Balzarotti, V., C. Castro y A. Powell (2002).

² Basel Committee on Banking Supervision, “Studies on the Validation of Internal Rating Systems”, Working Paper 14, May 2005.

³ 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) and Greene, W. (1992).

⁴ BCBS Publications No. 14, May 2005.

Email: vbalzarotti@bcra.gov.ar

algunas de las metodologías propuestas en ese documento para examinar el poder discriminatorio y la calibración.

Una variedad de modelos estadísticos se han desarrollado para apoyar la decisión crediticia. Estos modelos pueden utilizarse no sólo con dicho objetivo sino también para estimar las 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) para calcular requisitos de capital dentro del enfoque de Ratings Internos, o Internal Ratings Based (IRB). Por este motivo los modelos de scoring crediticio han recibido mucha atención recientemente.

En este trabajo no intentamos construir un modelo de scoring crediticio como herramienta de administración para un banco, ni apuntamos a discutir los pros y contras de los distintos modelos. Aplicaremos directamente un modelo probit, el cual ha probado dar buenos resultados en estudios previos⁵, y preferimos concentrarnos en los aspectos novedosos del trabajo, es decir, la corrección del sesgo antes mencionado y la investigación de técnicas de validación.

En lo que sigue, la sección II describe la CD y explica la definición de default utilizada en la estimación del modelo probit, mientras la sección III describe el modelo teórico, explica la diferencia entre truncamiento, censura, sesgo de selección y el problema de los deudores removidos. La sección IV describe las estimaciones de los distintos modelos que intentan corregir este último problema y la sección V presenta su validación (poder discriminatorio y calibración). Finalmente, la sección VI contiene las conclusiones.

II. La Base de Datos Crediticia Pública en Argentina: la Central de Deudores

El BCRA ha desarrollado una base de datos llamada “Central de Deudores del Sistema Financiero”(CD) en la cual se registran las deudas de consumo y comerciales de las entidades financieras, emisoras de tarjetas de crédito y fideicomisos financieros, en tanto el saldo sea superior a 50 pesos (este monto era equivalente a USD50 antes de la devaluación de 2002 y ahora representa aproximadamente USD17). La información de cualquier deudor es accesible en forma gratuita en la página web del BCRA, indicando el número de identificación tributaria del deudor (CUIT o CUIL). Las entidades envían la información al BCRA mensualmente.

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 acuerdo de márgenes crediticios). Se registra tanto información positiva (deudores que cumplen) como negativa (deudores con incumplimientos). Para cada deudor en cada entidad, la CD registra el número de identificación tributaria, el nombre, si el deudor es un individuo o una persona jurídica, el saldo de las facilidades en cada líneas de crédito, la tasa de interés, el plazo, las garantías (aquéllas reconocidas por el BCRA), la calificación o rating (conocida normalmente como “situación”) y las provisiones. No hay datos demográficos ni información de atrasos. La información para algunas variables muestra baja calidad, como son los casos de plazo y tasa de interés.

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

⁵ Balzarotti, V., C. Castro y A. Powell (2004) y Balzarotti, V., C. Castro y A. Powell (2002).

voluntad de pago y (v) proveer información para investigación económico-financiera, para decisiones de política macroeconómica y para regulación bancaria.

A los efectos de su clasificación crediticia, la cartera de créditos de las entidades se divide en dos grupos: el primer grupo se denomina “consumo y vivienda” y abarca préstamos destinados a financiar el consumo, hipotecas sobre la vivienda y, a opción de la entidad, créditos que financian actividades comerciales de hasta 500.000⁶ pesos. La “cartera comercial” incluye al resto.

El diseño actual de la CD data de 1997, cuando el BCRA fusionó en una única base de datos las dos bases anteriores, que eran parciales. Actualmente la CD contiene información de 5,36 millones de individuos y 96.500 deudores comerciales, los cuales en conjunto representan 7,4 millones de créditos en el sector bancario. A fines de 2005, y considerando solo los deudores del sector privado, el portafolio de consumo (incluyendo PyMes) sumaba \$35,8 mil millones (7,5 millones de deudores) y la cartera comercial \$43,6 mil millones (19.600 deudores).

Al momento de originación de un crédito, todos los deudores deben obtener una calificación que va de “1” a “5” (donde 1 es el mejor), de acuerdo con la normativa del BCRA, la cual fundamentalmente apunta a considerar 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 cuales se reducen a la mitad para las deudas garantizadas. Los deudores cuyas deudas están totalmente cubiertas por el mejor tipo de garantías (tales como efectivo, depósitos a plazo, etc.) no se califican y se informan a la CD en la mejor situación⁷.

La calificación de los deudores comerciales debe ser revisada 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 capital⁸. La normativa enumera un conjunto de pautas para analizar la situación financiera, incluyendo la presentación de información contable a tiempo, las perspectivas de sector económico, atrasos, etc. Por su parte, las calificaciones de los deudores de consumo son revisadas sólo de acuerdo con los atrasos y la situación legal de insolvencia. Las reglas 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 (dado que también deben considerarse las otras pautas). Otro párrafo de la regulación establece que un 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 con el deudor representen 40% o más de sus créditos totales en el sistema.

Definición de Default

Como se señaló, existen 5 grados en el sistema de calificación⁹. Para decidir cuáles de ellos se tomarán como “default” principalmente hemos considerado la definición “consensuada” que establece Basilea II. En ese marco, el default tiene lugar cuando se observan uno o ambos de los siguientes eventos:

⁶ Este umbral se incrementó de \$200 mil en marzo de 2005.

⁷ Texto ordenado de Clasificación de Deudores, BCRA.

⁸ Estas reglas fueron cambiadas en varias ocasiones luego de la crisis del 2002 y este es uno de los motivos para que en este trabajo se eligiera el año 2000 como año base.

⁹ Existe una categoría “6” que incluye a un pequeño número de deudores que no han pagado sus créditos con bancos liquidados. Decidimos no considerar estos créditos en el análisis.

- El banco considera que es improbable que el deudor pague de manera 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 grupo bancario.¹⁰

Por lo tanto, clasificamos a las situaciones 3 a 5 como defaults. Nuevamente, el grado 3 significa que un deudor de la cartera de consumo está atrasado más de 90 días, lo mismo que un deudor de la cartera comercial, excepto por el hecho de que en este caso el criterio es sólo indicativo y podría prevalecer la consideración de otras pautas (aunque esto no ocurre habitualmente).

Deudores que desaparecen

Puede observarse de manera regular que grupos de deudores dejan de ser registrados en la CD. Por ejemplo, 16% de la muestra inicial de los deudores comerciales utilizados en este estudio (ver características de la muestra más adelante) no estaban presente en la CD un año después, y lo mismo ocurre con 24% de los deudores en el segmento de deudores de consumo (Cuadro 1) Más aún, los exclusiones tiene lugar desde todos los grados de rating, como se muestra en el Anexo 1¹¹. Aún para horizontes temporales mensuales, se observan deudores que son removidos de todos los grados de rating.

Cuadro 1: Deudores removidos en las carteras comercial y de consumo

	Default	Frecuencia	Porcentaje
Comercial	No-Default (0)	12.632	76%
	Default (1)	1.352	8%
	Faltante (.)	2.668	16%
	Total	16.652	100%
Consumo	No-Default (0)	3.352.717	68%
	Default (1)	364.956	7%
	Faltante (.)	1.188.773	24%
	Total	4.906.446	100%
	Total muestra	4.923.098	
	Total población	6.153.873	

La razón de estas exclusiones no se registra, de allí que no se pueda clasificar a los deudores faltantes como default o no-default al final del período¹². A priori, hay dos motivos opuestos para una exclusión:

- 1) Total cancelación de los créditos. Estos deudores deberían ser clasificados como no-defaults.
- 2) Deudores dados de baja, lo cuales deberían ser clasificados como defaults.

¹⁰ 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.

¹¹ Los modelos de scoring trabajan con deudores calificados 1 ó 2 en el momento inicial, no obstante, el hecho de que se remuevan deudores de todas las situaciones puede ser informativo y por eso se muestran en el cuadro.

¹² Existe una base de datos separada donde deben registrarse los deudores dados de baja contablemente, pero la información disponible no tiene suficiente calidad.

La normativa del BCRA establece que para un deudor comercial un atraso de 365 días o más es indicativo de “pérdida” es decir, situación 5, y debe ser provisionado completamente (excepto cuando existe una garantía reconocida, en cuyo caso los niveles de provisionamiento son menores). Los créditos clasificados en situación 5 y totalmente provisionados deben ser transferidos a Cuentas de Orden luego de permanecer siete meses en esta situación. Los créditos garantizados pierden el beneficio de poder constituir menores provisiones luego de permanecer dos años en los grados 4 ó 5 (en general), por lo cual la existencia de garantías puede retrasar la migración final a Cuentas de Orden. Un crédito puede permanecer en estas cuentas en la medida en que la entidad continúa ejerciendo acciones de cobro.

Los créditos en Cuentas de Orden también deben ser informados a la CD. Por lo tanto, si un deudor comercial originalmente clasificado “1” o “2” (indicativo de un máximo de 89 días de atraso) no aparece en la CD un año después, y suponiendo que se debe a que el banco a enviado a pérdida un default, ello implicaría que la entidad ha aplicado procedimientos de castigo muy estrictos, mucho más que los requeridos por el BCRA. Esto no es muy probable, aunque no es imposible, considerando que el monto de cualquier deuda comercial es relativamente importante para los bancos locales. En esta línea, puede observarse que algunos deudores clasificados como “1” o “2” en diciembre de 1999 están clasificados en “5” en diciembre de 2000 en nuestra muestra (ver Cuadro 2).

Cuadro 2: Rating de deudores comerciales en Diciembre 1999 y 2000

Rating en Diciembre 1999	Rating en Diciembre 2000						Total
	0	1	2	3	4	5	
1	2.533 16%	11.785 75%	449 3%	282 2%	572 4%	143 1%	15.764
2	135 15%	137 15%	261 29%	104 12%	179 20%	72 8%	888

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. Estos créditos continuarán siendo informados a la CD a menos que el comprador no sea un fideicomiso obligado a informar (los fideicomisos financieros constituyen el tipo más común de comprador y ellos deben informar a la CD). Aún cuando el comprador no sea un fideicomiso financiero, si la entidad vendedora u otra entidad financiera, provee servicios de cobro, entonces esa 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 una exclusión. Sólo a nivel sistémico pueden inferirse las transferencias. Adicionalmente, existe un motivo operativo que puede llevar a una pérdida temporaria del rastro de un crédito transferido, el cual se origina en el hecho de que los nuevos fideicomisos pueden informar sólo cuando hayan completado ciertos procedimientos de autorización, los cuales pueden tomar un par de meses. Si este retraso coincide con la observación al final del período, algunos créditos pueden estar faltando por esta razón.

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

Mientras que las mencionadas cuatro causas pueden estar detrás de un deudor que desaparece, por lo dicho anteriormente puede argumentarse que las cancelaciones, ventas y errores son más

probables que el envío a pérdida. Para tener más indicios, se puede analizar las características del conjunto de deudores que desaparecen y compararlos con aquéllos que permanecen en la base de datos. Puede observarse (Anexo 1) que ambos grupos son altamente similares. Sólo existe una clara discrepancia en el porcentaje de deudores extranjeros, el cual es mayor entre los que son removidos. Esto también apunta en la dirección de cancelaciones, ventas o errores, más que defaults y total cese de acciones de cobro.

Otro indicio útil puede encontrarse en el análisis de los saldos de deuda de los deudores que desaparecen de un banco (recuérdese que los datos entran en el modelo a nivel deudor-banco) pero que aún tienen crédito (comercial o de consumo) en el sistema financiero. Tanto el monto como la calificación pueden ser informativos y se presentan en el Cuadro 3. En general, la deuda total de estos deudores en el sistema ha decrecido debido a la desaparición de algunos créditos en ciertas entidades.

Cuadro 3: Cambio en el saldo de deuda total en el sistema, dic. 1999 – dic. 2000, y peor calificación de los deudores que desaparecen (dic. 2000)

Cambio deuda total sistema	Frecuencia	%	Promedio tasa de cambio	Promedio peor calificación
Aumento	537	20	118%	1,30
Disminución	1.531	57	-52%	1,68
Desaparecidos	600	22		
Total	2.668	100		

III El Modelo de Scoring Crediticio y Cálculo de las PDs

Existe un conjunto de métodos estadísticos usuales para construir un scoring crediticio. Los más utilizados son análisis discriminante, redes neuronales, árboles de decisión y métodos estadísticos tradicionales como las regresiones Logit y Probit. Una cantidad considerable de estudios que comparan la performance de estos modelos de scoring crediticio arrojan como principal conclusión que sus resultados son relativamente similares. La técnica utilizada en este documento es una regression Probit con una variable explicada binaria para el default del deudor. Hemos seleccionado este modelo por su simplicidad y porque la influencia de las variables explicativas sobre la probabilidad de default se puede obtener fácilmente.¹³ Adicionalmente, se ha probado en estudios previos que este modelo funciona bien y tiene la ventaja (lo mismo que los modelos logit) de que los scores pueden ser trasladados fácilmente a las PDs de los deudores.

En el modelo Probit, la variable binaria explicada es la situación de default, lo cual se puede representar de la siguiente manera:

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

¹³ Entre los requerimientos de 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.

Donde X_i' es la matriz de variables explicativas de cada deudor i y ε_i es el error de la regresión. Los $X_i'\beta_i$ de estas estimaciones son los scores de los deudores. Las probabilidades de default pueden calcularse a partir de la Distribución Normal Acumulada para modelos Probit junto con el score. En consecuencia, la PD se puede estimar como

$$P(D_i = 1 / X_i) = \Phi(X_i' \beta)$$

Se trabajó con deudores del sector privado no financiero¹⁴ y se distinguió entre créditos comerciales y de consumo¹⁵. La estimación del modelo de crédito para deudores de consumo necesitaría otros estudios en los cuales sería más preciso distinguir por tipo de crédito. Esto estaría más en línea con Basilea II, donde los bancos que aplican el enfoque IRB debe agrupar a sus deudores minoristas de acuerdo con sus características (tales como tipo de crédito, días de atraso, etc.) y estimar las PDs agregadas que comparten todos los créditos dentro de un segmento.

Basilea II también ha establecido que las PDs agregadas de cada grado en un sistema de rating debe 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. En este trabajo hemos restringido el análisis a la estimación de PDs anuales no estresadas, las cuales son un input en la construcción de sistemas de rating Punto-en-el-Tiempo (*Point-in-Time - PIT*), y constituyen un punto de partida para estudios posteriores que consideren las PDs estresadas y sistemas de rating a través de ciclo (*Through-the-Cycle - TTC*).

El horizonte temporal del modelo es un año, para obtener PDs anuales, y hemos elegido trabajar con el período diciembre 1999 – diciembre 2000. Las PDs estimadas representan la probabilidad de que un deudor calificado como no-default (situaciones “1” o “2”) en diciembre de 1999 pueda estar calificado como default (situaciones “3”, “4” o “5”) en diciembre 2000. Si bien la economía estaba en recesión en ese año, el período es previo a la crisis extrema del sector bancario que se aceleró a fines del 2001. De manera que, si bien en nuestra opinión los niveles de riesgo resultantes del modelo probablemente estarán por encima de los que surjan de un promedio de largo plazo, aún así las conclusiones que se pueden obtener en temas como la existencia del sesgo por los deudores removidos o la aplicación de herramientas de validación, son relevantes y de interés.

El modelo trabaja a nivel de deudor-banco, es decir, un deudor podría obtener diferentes ratings en dos bancos (sujeto a la restricción normativa de máxima diferencia entre deudas significativas, que se explicó arriba). Mientras que la CD contiene todo el universo de deudores, se tomó una muestra para poder testear la performance del modelo de scoring crediticio dentro y fuera de la muestra. Se utilizó un método de muestreo secuencial para obtener un 75% de las observaciones, controlando por las variables principales (situación de los deudores en diciembre del 2000 y saldo de la deuda en cada entidad). Se estimó un modelo Probit con corrección de outliers considerando los *dfbetas* o *dcooks* de las observaciones.

La muestra de los deudores comerciales utilizada en las estimaciones tiene 16.652 deudores-banco, mientras que el número de deudores consolidado a través de los bancos es de 10.343 (la misma

¹⁴ Incluyendo deudores nacionales y extranjeros.

¹⁵ Esta distinción no se pudo obtener directamente de la base de datos para los meses bajo estudio debido a que la variable categórica que indica si un crédito es comercial o de consumo no está disponible. Por eso, se construyó la variable aplicando el máximo monto que puede tener un crédito comercial para ser tratado como crédito de consumo (a opción de la entidad) según la normativa, y que es una deuda de \$200 mil.

empresa puede ser deudora en más de una entidad: en promedio cada deudor tiene crédito en 1,6 entidades). Las entidades tienen en promedio 152 créditos comerciales, pero estos no se distribuyen equitativamente entre ellas: de 109 entidades, 81 tienen menos de 100 créditos cada una.

Las variables explicativas usadas en los modelos se construyeron con la información limitada de la CD, cuyas variables tienen una calidad desigual. Por ese motivo se han utilizado aquellas que muestran relativamente buena calidad. Estas variables principalmente se refieren a la situación del deudor en diferentes períodos, la peor calificación registrada recientemente, saldo de deuda del deudor con el banco y con el sistema, garantías y obligaciones en default en términos del total de obligaciones (para más detalles véase el Anexo 2).

Tratamiento tentativo de las exclusiones

El tratamiento de las exclusiones (deudores que desaparecen de la muestra entre diciembre de 1999 y diciembre de 2000) va a determinar diferentes versiones del modelo de scoring.

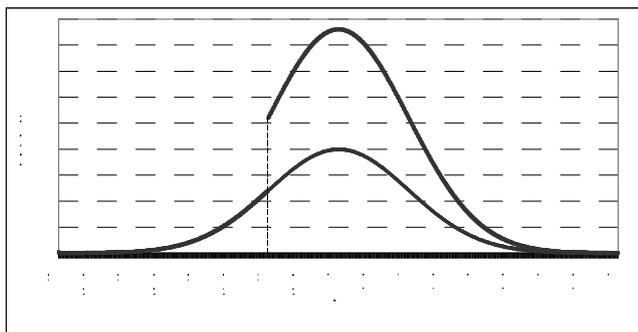
Antes de entrar en detalle a los modelos, vale la pena dar algunas precisiones sobre la diferencia entre truncamiento, censura, sesgo de selección y el problema de los deudores que desaparecen. El recuadro adjunto examina esa diferencia.

Truncamiento, censura, sesgo de selección y el problema de los deudores que desaparecen

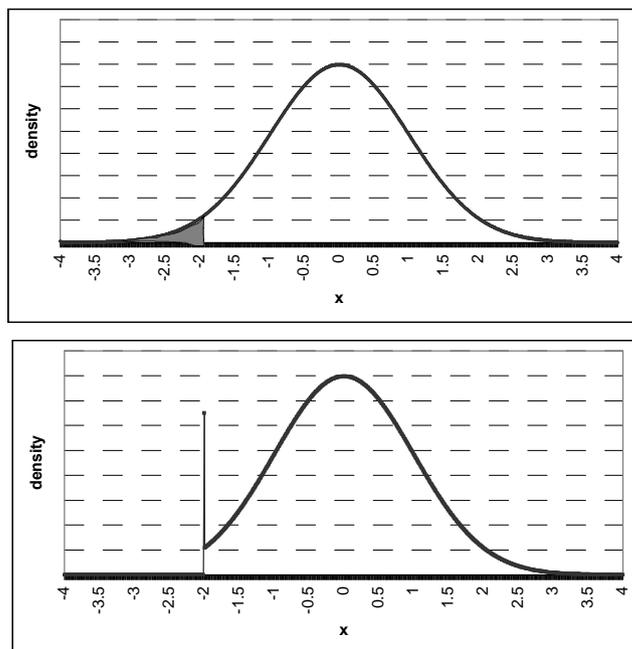
Los efectos del truncamiento tienen lugar cuando los datos de una muestra se toman de un subconjunto de una población de interés más amplia. La censura es esencialmente un defecto en los datos muestrales. Cuando una variable dependiente está censurada, los valores que pertenecen a cierto rango se informan como un único valor. Ello introduce una distorsión en los resultados estadísticos. El sesgo de selección es una forma de truncamiento, pero sobreviene cuando el proceso de truncamiento no es aleatorio.

Los gráficos a continuación ilustran los problemas de truncamiento y censura para una distribución Normal.

Distribución Normal Truncada (en -1)



Distribuciones Normales Completa y Censurada (en -2)



Que el efecto del truncamiento sobre los coeficientes sea de interés depende de las inferencias que se busquen en el estudio en cuestión. Si el análisis se limita a la subpoblación, entonces los coeficientes

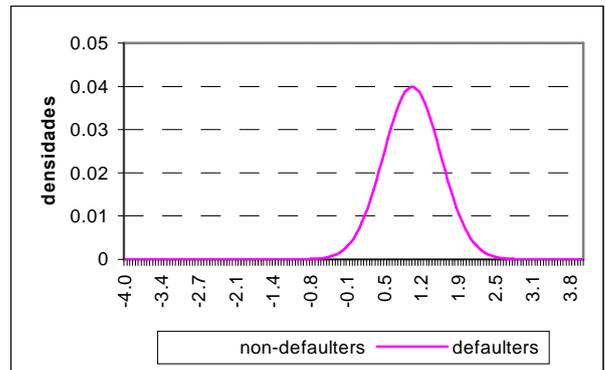
sesgados son válidos. Si el estudio intenta extender sus resultados a la población total, entonces lo que realmente importa son los coeficientes insesgados.

Cuando existe sesgo de selección muestral, el muestreo no es aleatorio. La literatura ha estudiado las consecuencias de este defecto sobre las propiedades de los estimadores convencionales.

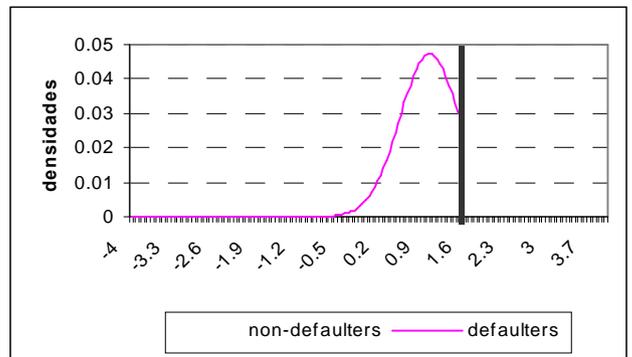
En el caso de los bancos que evalúan riesgo crediticio, el sesgo de selección es un problema común. Normalmente se mantienen registros sólo de los solicitantes que obtienen crédito y no de los que son rechazados. En esos casos, en comparación con el sesgo que estudiamos en este trabajo, el truncamiento tiene una dirección y se puede encontrar un número de propuestas en la literatura para corregir sus efectos. La pre-selección de la muestra puede representarse por una condición de aceptación de una solicitud igual a que el score sea menor que un cierto umbral. Si bien esto es una simplificación de los criterios para otorgar crédito que aplica un banco, es bastante intuitiva. En consecuencia, las distribuciones de los defaults y no-defaults serán observables para los scores censurados, como lo ilustran los siguientes gráficos. El primer gráfico muestra las densidades completas de los defaults y no-defaults, según los scores, y el segundo, las densidades de los defaults y no-defaults con sesgo de selección, donde la cola derecha de las densidades no se observan (las áreas bajo las dos curvas truncadas suman 1).

El eventual sesgo de los coeficientes aparece cuando los proxies del mecanismo de selección se omiten en la ecuación de las PD. Se han estudiado formas de corregirlo en la literatura que se aplican en la práctica. En el caso de este trabajo, aparece una forma de truncamiento por el hecho de que algunos deudores que están presentes en la CD en dic. 1999 son excluidos por un proceso no aleatorio en dic. 2000. No considerar este grupo también puede conducir a estimaciones sesgadas.

Densidades completas de deudores en default y no-default



Densidades de deudores en default y no-default con sesgo de selección



Este sesgo por la desaparición de deudores, sin embargo, a priori no tiene una dirección o tiene ambas direcciones. Las exclusiones pueden originarse en la cancelación total de la deuda, su venta o, en el otro extremo, la baja contable definitiva, así como a un error de información.

La existencia (o no existencia) de deudores que desaparecen no sólo afecta la estimación del modelo, sino también tiene efectos sobre la evaluación de la performance de los modelos de rating.

IV - Los 5 modelos

Modelo 1 : “Muestra Incompleta – Modelo sin Corrección”. Un primer enfoque es proceder con la estimación del modelo restringiendo la base de datos al grupo de deudores que están presentes tanto en dic. 1999 como en dic. 2000, es decir, sin intentar corregir el potencial sesgo. Este va a ser el modelo “benchmark” y también es el enfoque utilizado en estudios previos en Argentina y otros bancos centrales.

Modelo 2: “Muestra incompleta – Corrección tipo Heckman”. La misma muestra truncada del Modelo 1 se utiliza para calibrar un modelo probit por máxima verosimilitud con selección muestral. El método utilizado para corregir el sesgo de selección

en este modelo es entonces similar al método de Heckman, pero con una variables explicada binaria (default), como sigue:

Sea X_t la matriz de variables explicativas (información financiera sobre los deudores). Sea Y_t una variable dummy que indica si un deudor específico permanece en el sistema o no (esta es la variable que considera el proceso de selección). Esta variable toma el valor 1 si un deudor está presente al final del período o 0 en otro caso. D_t es una variable dummy que indica default. Así, D_t es observable sólo en los casos en que $Y_t = 1$, o sea, cuando el deudor permanece en la CD, pero es desconocida para aquéllos que desaparecen ($Y_t = 0$). La matriz de información financiera, X_t , combina la probabilidad de permanecer en la CD y la probabilidad de hacer default. La cadena que determina los eventos es:

$$X_i \rightarrow Y_i (X_i) \rightarrow D_i(Y_i, X_i)$$

Si Y_t se estableciera en forma aleatoria, el segundo paso no agregaría información adicional y no habría necesidad de considerar el proceso de selección. En ese caso el Modelo 1 sería apropiado para estimar las PDs. Pero si Y_t contiene información adicional, el proceso de selección que determina la permanencia en la CD debería considerarse para evitar el sesgo. El Modelo 2 supone que Y_t agrega información y por lo tanto trata de modelar el proceso de selección.

Modelos 3 a 5: Otro enfoque que se va a explorar es la reincorporación de los deudores que desaparecen, suponiendo su comportamiento (default o no) al final del período. Esta metodología es similar a la de Hand and Henley¹⁶. Para proceder de esta manera, es necesario una regla de decisión para marcar como default o no-default a los deudores antes de reincorporarlos a la muestra. La regla de clasificación de defaults de los deudores removidos define los modelos 3 a 5.

Modelo 3: “Reincorporación Global como Cancelaciones”. Siguiendo el razonamiento apriorístico que se propuso antes, existen varios argumentos que sostienen la idea de que la mayoría de los deudores removidos son en realidad cancelaciones o ventas (o tal vez errores). Desde el punto de vista del riesgo, las ventas representan realizaciones de los créditos y son similares a obtener el repago de los mismos. Entonces, un supuesto extremo sería que todos los deudores que desaparecen en verdad han cancelado totalmente. En consecuencia, bajo esta regla todos ellos serán reincorporados como no-defaults. Este enfoque parece ser un tanto excesivo pero en la dirección correcta.

Modelo 4: “Reincorporación de acuerdo con el score promedio de los que hicieron default”. Teniendo en cuenta que las características del conjunto de deudores removidos son por cierto muy similares a las de los que permanecen (lo cual más adelante se ve confirmado por los scores estimados para ambos grupos), se puede hacer un ejercicio en el cual, en primer lugar, se estima un scoring usando la muestra truncada de deudores que están presentes en ambos puntos del tiempo, dic. 1999 y dic. 2000 (similar al Modelo 1) y luego los deudores que desaparecen obtienen un score usando los parámetros obtenidos de este modelo. A

¹⁶ Hand D.J. and Henley. W.E. (1993) y Hand D.J. and Henley. W.E. (1994).

continuación, se puede pensar en dos reglas para discriminar a los que hacen default de los que no, usando esos scores estimados: la primera regla es marcar como defaults a los casos cuyos scores son menores que el score promedio de los deudores que permanecen que hacen default. Finalmente, luego de clasificar a los deudores desaparecidos se integran en la muestra y se estiman nuevamente las PDs, usando un modelo Probit, esta vez con la muestra incrementada.

Este enfoque es similar al de Hand y Henley. Ellos analizan el problema de selección en el proceso de decisión de otorgamiento de crédito. En su caso, no pueden seguir a los solicitantes en la región de “rechazos” porque no obtuvieron crédito. Con un proceso similar, re-incluyen a los casos rechazados en la muestra, aunque ellos encuentran que esto parece tener poca información adicional.

La segunda regla sería marcar como defaults a un porcentaje de los deudores que desaparecen que obtienen los scores más altos, usando el Modelo 1, tal que este porcentaje es igual a la tasa de default obtenida del conjunto de los deudores que permanecen. En la práctica, esta regla es casi exactamente igual a la regla previa (es decir, siguiendo al score promedio de los que hacen default) y los parámetros estimados son también casi idénticos. Por esta razón, no se presenta como otro modelo.

Modelo 5: “Deudores ‘limpios’ han cancelado, el resto han sido dados de baja”. Se les asigna una marca de no-default a los deudores removidos sólo si tienen una historia crediticia “limpia” en términos de rating en el banco y en el sistema. Esto significa que los deudores con una situación mayor o igual a “3” como máximo rating en el sistema, ya sea en diciembre de 1999 o seis meses antes (junio 1999) son considerados defaults. Otras variables como saldo de deuda, garantías, etc., no son consideradas para distinguir a quienes se supone hacen default. El modelo se corre sobre la base así incrementada.

Cuadro 4: Estadísticas de los 5 modelos

Modelo	Default o no	Frecuencia	Tasa de Default
Modelos 1 y 2	No-default	12.626	10,69%
	Default	1.350	
Modelo 3	No-default	15.294	8,83%
	Default	1.350	
Modelo 4	No-default	15.079	10,38%
	Default	1.565	
Modelo 5	No-default	15.090	10,0%
	Default	1.554	

Resultados de los modelos

El Cuadro 5 muestra los parámetros estimados por los cinco modelos. Existen diferencias en los valores estimados de los parámetros y hasta hay coeficientes de diferente signo. Los coeficientes en los Modelos 1 y 3 a 5 son los más similares mientras que los resultados del Modelo 2 son marcadamente diferentes. En general, los coeficientes tienen los signos esperados. El Anexo 2 muestra las relaciones univariadas “a priori” entre las variables explicativas y las tasas de default.

Los coeficientes para las situaciones de calificación tienen los signos esperados: situaciones inferiores reducen la probabilidad de default. Hay una relación positiva entre el número de bancos acreedores que tiene un deudor y la probabilidad de default, aunque los deudores con una institución acreedora son más riesgosos que aquéllos con dos.

Existe una diferencia en el nivel de riesgo de los deudores de distintos grupos de bancos, clasificados por titularidad del capital (por ej. bancos de capital extranjero, bancos domésticos, sucursales de instituciones financieras del exterior y bancos privados nacionales); los bancos privados locales obtienen los coeficientes más favorables.

El porcentaje de crédito garantizado (hipotecas sobre la vivienda y prendas) tiene un coeficiente positivo y significativo en dos de los modelos, indicando que más respaldo conduce a mayor probabilidad de default. Aspectos tales como relaciones de clientela pueden estar presentes¹⁷, o bien los bancos pueden estar pidiendo garantías para contrarrestar un nivel de riesgo percibido mayor, de manera de incrementar los montos de recupero, aunque la existencia de garantías puede estar reduciendo el comportamiento de default sólo parcialmente¹⁸. La relación a priori entre la frecuencia de defaults y garantías es débil (véase Anexo 2)¹⁹.

La probabilidad estimada de default decrece a mayor monto de saldo de deuda en el banco, como también a mayor saldo de deuda en el sistema. En cambio, el riesgo estimado se incrementa para los deudores que exhiben un porcentaje mayor de deudas en default sobre su deuda total.

La Bondad del Ajuste se mide por el estadístico Pseudo R^2 (multiplicador de Lagrange) ajustado por el número de variables explicativas. El Modelo 4 muestra el mejor ajuste.

¹⁷ véase Berger A.N and G.F. Udell (1995) y Berglof E. and Gollier, C. (1997)

¹⁸ Un resultado similar obtienen Schechtman et al (2004) al estimar PDs para deudores comerciales en Brasil, donde encuentran que la relación con las garantías no es estadísticamente significativa.

¹⁹ En el caso de los créditos al consumo la relación entre el porcentaje garantizado del crédito y la frecuencia de defaults es claramente negativa.

Cuadro 5: Default es la variables explicada

<i>Variabes explicativas</i>	<i>Modelo 1</i>	<i>Modelo 2</i>	<i>Modelo 3</i>	<i>Modelo 4</i>	<i>Modelo 5</i>
<i>Constante</i>	1.76***	0.35	2.36***	2.15***	2.1***
<i>Rating_previo 0</i>	-0.44**	-0.04	-0.8***	-0.45***	-0.57***
<i>Rating_previo 1</i>	-0.29*	-0.19	-0.6***	-0.23	-0.42***
<i>Rating_previo 2</i>	-0.26	-0.13	-0.57***	-0.2	-0.42**
<i>Rating_previo 3, 4 y 5</i>					
<i>Peor_rating 1</i>	-0.49***	-0.25***	-0.8***	-0.59***	-0.93***
<i>Peor_rating 2</i>	0.14	0.13	-0.18**	0.07	-0.31***
<i>Peor_rating 3, 4 y 5</i>					
<i>Rating 1</i>	-0.66***	-0.63***	-0.56***	-0.83***	-0.59***
<i>Rating 2</i>					
<i>Bancarización 1</i>	-0.56***	0.14*	-0.74***	-0.65***	-0.61***
<i>Bancarización 3</i>	-0.73***	-0.11	-0.81***	-0.8***	-0.72***
<i>Bancarización 5</i>	-0.62***	-0.2***	-0.64***	-0.66***	-0.58***
<i>Bancarización 7</i>	-0.49***	-0.16***	-0.49***	-0.52***	-0.47***
<i>Bancarización 9</i>	-0.1	0.01	-0.12*	-0.1*	-0.1
<i>Bancarización 10</i>					
<i>Tipo_institución: Bancos públicos nacionales</i>	-0.29**	-0.23**	-0.22***	-0.37***	-0.27**
<i>Tipo_institución: Bancos locales de capital extranjero</i>	-0.38***	0.24***	-0.48***	-0.46***	-0.38***
<i>Tipo_institución: Bancos privados cooperativos</i>	-0.2	-0.21**	-0.15	-0.23*	-0.14
<i>Tipo_institución: Bancos privados S.A. de capital nacional</i>	-0.47***	-0.24***	-0.45***	-0.51***	-0.43
<i>Tipo_institución: Bancos públicos provinciales y municipales</i>	-0.09	-0.09	-0.04	-0.14	-0.04
<i>Tipo_institución: Bancos sucursales de ent. fin. del exterior</i>	-0.34***	-0.02	-0.32***	-0.43***	-0.32***
<i>Tipo_institución: Otras instituciones financieras</i>					
<i>Garantías</i>	0.16***	-0.04	0.18***	0.2***	0.19***
<i>Ldeuda_banco</i>	-0.05*	-0.11***	-0.02	-0.05*	-0.05*
<i>Ldeuda_sistema</i>	-0.12**	0.08***	-0.16***	-0.14**	-0.11***
<i>Porcentaje_default</i>	0.51**	0.61***	0.48**	0.63***	0.47**
<i>Nlineas_default</i>	0.28***	0.27***	0.02	0.28***	0.23***
<i>Nlineas_previas_default</i>	0.43***	0.37***	0.12**	0.47***	0.4***
<i>Pseudo R²</i>	0.21	0.072	0.194	0.2684	0.2562
<i>PD Media Estimada</i>	10.64%	22.86%	8.79%	10.31%	10.25%
<i>% de defaults en la muestra</i>	10.69%	10.69%	8.83%	10.38%	10.30%

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

Los valores faltantes de las variables dummy son las variables base

V- Validación: poder discriminatorio y calibración

La publicación del Comité de Basilea “*Studies on the Validation of Internal Rating Systems*” (2005) sugiere algunas herramientas estadísticas para validar sistemas de rating. En esta sección mostraremos los resultados de aplicar las técnicas más ampliamente utilizadas que se describen en ese documento para evaluar diferentes aspectos de los modelos de scoring estimados.

Como se indicó, los datos provistos por la CD para estimar un sistema de rating son necesariamente deficientes, en la medida que los bancos tienen acceso a información de las características del deudor que no se registran en la CD. Sin embargo, es importante testear los méritos de nuestros modelos por interés metodológico así como para evaluar si este tipo de modelo podría ser desarrollado por los supervisores para ser utilizado como una especie de benchmark de riesgo o para calibrar la regulación prudencial.

Los sistemas de rating y las funciones de score pueden ser vistos como herramientas de clasificación que proveen indicadores de la situación futura del deudor. Hay dos aspectos a ser evaluados sobre ellos: el poder discriminatorio y la calibración.

Cuando un sistema de rating es bueno, el conjunto de deudores que obtiene los mejores scores mostrará bajas frecuencias de default y el grupo que resulta asignado a los peores grados mostrará las mayores frecuencias de default. Por lo tanto, un sistema de rating tiene más poder discriminatorio cuanto mayor es la diferencia entre las distribuciones de los scores para los deudores que hacen default y para los que no (habilidad para discriminar los malos de los buenos deudores).

Presentamos algunas medidas estadísticas de poder discriminatorio. Las medidas absolutas de poder discriminatorio de un sistema de rating tienen un significado limitado. En cambio, se utilizan para comparar entre sistemas de rating para obtener una idea realista del ajuste del modelo y, por ejemplo, para detectar “ruido” estadístico.

Las medidas utilizadas en el presente documento, el ratio AR (Accuracy Ratio) y los índices ROC (Receiver Operating Characteristics) y Pietra, explícitamente toman en cuenta el tamaño de la muestra en default. Estas medidas no dependen de la tasa de default del portafolio, por lo cual pueden ser usadas correctamente para comparar sistemas de rating con diferentes tasas de deudores que hacen default.

Chequear la **calibración** de un sistema de rating es distinto de ver su poder discriminatorio. Para evaluar la calibración, los deudores deben ser agrupados en grados de rating considerando sus scores. Un sistema con calibración correcta exhibiría PDs proyectadas 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. El segundo caso indicaría que el modelo necesita ser mejorado. En el presente documento se consideran los tests Binominal y Chi-cuadrado para evaluar la calidad de las PDs estimadas por el sistema de rating.

Como hemos desarrollado 5 modelos, en primer lugar seleccionaremos un modelo en función de su poder discriminatorio y luego testaremos la calibración sólo para ese modelo.

Poder Discriminatorio

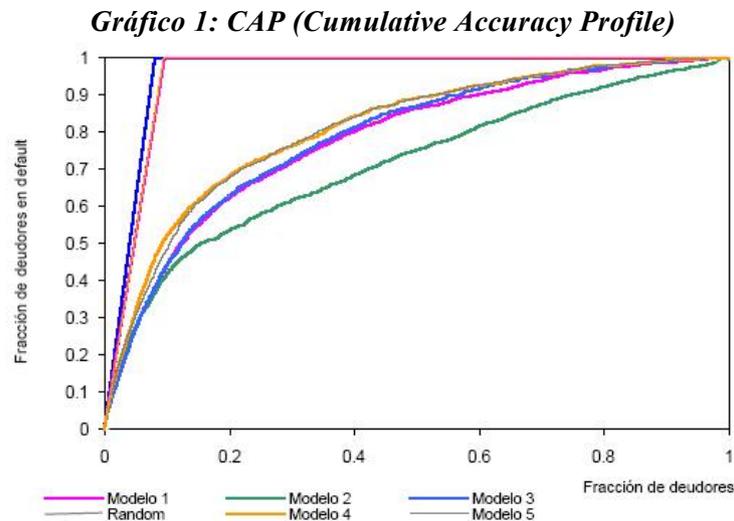
- **CAP (Cumulative Accuracy Profile) y AR (Accuracy Ratio):** El CAP se conoce también como curva de poder o curva de Lorenz. Muestra el porcentaje acumulado observado de deudores en default atribuibles a un ranking de observaciones ordenadas por sus scores. Visualmente, la curva CAP se determina graficando el porcentaje acumulado de deudores en el eje horizontal, del más riesgoso al menos riesgoso según su score, y los correspondientes porcentajes acumulados de defaults en el eje vertical.

Si fuera perfecto, el proceso de rating le asignaría a los que hacen default los menores scores y en consecuencia la curva CAP subiría linealmente desde el punto 0 hasta reflejar todos los defaults y luego pasaría a ser horizontal. De allí que, a más empinada la curva CAP en el origen, más precisión tiene el proceso de rating.

En el otro extremo se encontraría un modelo puramente aleatorio, sin ningún poder discriminatorio. La curva esperada del CAP en este caso sería la diagonal, dado que una fracción X de deudores contendría X% de defaults. Cuanto más cóncava sea la curva CAP, será mejor el poder discriminatorio del modelo de rating, dado que una curva más cóncava estaría más cerca del modelo ideal.

El índice resumido del CAP es el AR (Accuracy Ratio), el cual se basa en el coeficiente Gini del CAP. En este ratio el numerador es el área entre la curva CAP y la diagonal (el modelo aleatorio) y el denominador es el área entre el modelo ideal y la diagonal (véase el Anexo 2). Un sistema de rating es más preciso cuanto más cerca de uno esté el AR.

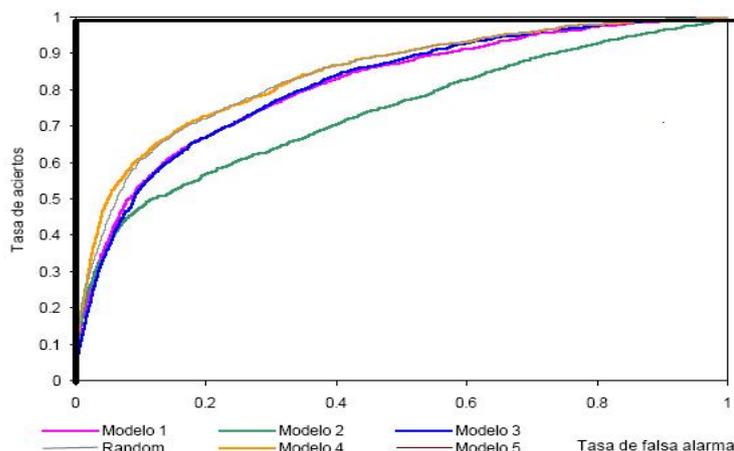
Comparando los modelos de rating estimados en el presente documento, la curva CAP del sistema de rating derivado del Modelo 4 es la más cóncava y tiene mayor AR. Haremos algunas calificaciones sobre estos resultados más adelante.



▪ **ROC (Receiver Operating Characteristics).** Una de las características de un buen sistema de rating es que tenga la mayor tasa de aciertos (“hit rate”) como sea posible (correcta clasificación de los deudores que impagan como potenciales defaults) y al mismo tiempo la más baja tasa de “falsa alarma” que sea posible (incorrecta clasificación de un deudor cumplidor como un potencial default). La curva ROC es un concepto que se relaciona con estas dos tasas y también con la curva CAP. Para construir la curva ROC, se calculan la tasa de aciertos y de falsa alarma para cada score, tomando cada nivel de score como un punto de corte (cut off) para otorgar crédito (véase el Anexo 2). La performance de un sistema de rating es mejor cuanto más empinada 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. Usando este criterio, el sistema de rating derivado del Modelo 4 tiene la mejor performance. Haremos algunas observaciones sobre este resultado más abajo.

El área bajo la curva ROC se mide por el índice ROC. El valor de este índice va desde 0,5 para un modelo aleatorio (la curva ROC es la diagonal) hasta 1 para el modelo ideal.

Gráfico 2: ROC (Receiver Operating Characteristics)



▪ **Índice Pietra.** El índice Pietra considera el mayor área triangular que puede obtenerse entre la curva ROC y la diagonal (véase Anexo 2).

Estas medidas para el poder discriminatorio de los modelos estimados se presentan en el Cuadro 6. El sistema de rating estimado por el Modelo 4 muestra mejor performance considerando estos tres índices. En particular, es claramente superior al Modelo 2, el cual también trata de corregir el sesgo proveniente de la exclusión de deudores. Haremos algunos comentarios sobre estos resultados a continuación.

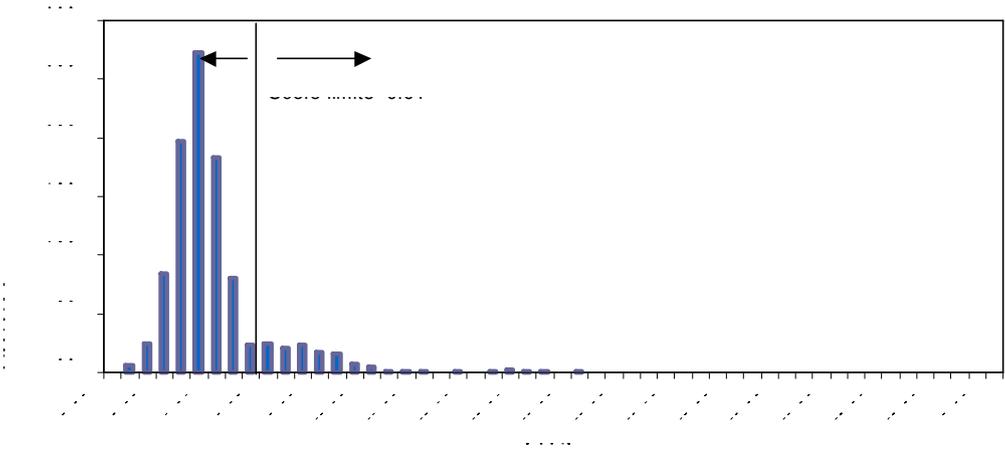
Cuadro 6: Medidas estadísticas de poder discriminatorio considerando a todos los deudores

<i>Deudores Comerciales</i>	<i>Accuracy Ratio (AR)</i>	<i>Indice ROC</i>	<i>Indice Pietra</i>
Modelo 1	62.3%	81.2%	0.168
Modelo 2	47.4%	73.7%	0.136
Modelo 3	62.7%	81.4%	0.169
Modelo 4	68.8%	84.4%	0.190
Modelo 5	68.1%	84.0%	0.188
Max (mayor poder discriminatorio)	100%	100%	>
Min (menor poder discriminatorio)	0%	50%	<
Referencias			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitative Default Risk Models: A validation methodology (Moody's)	50-75%		

Sobre los resultados del Modelo 4

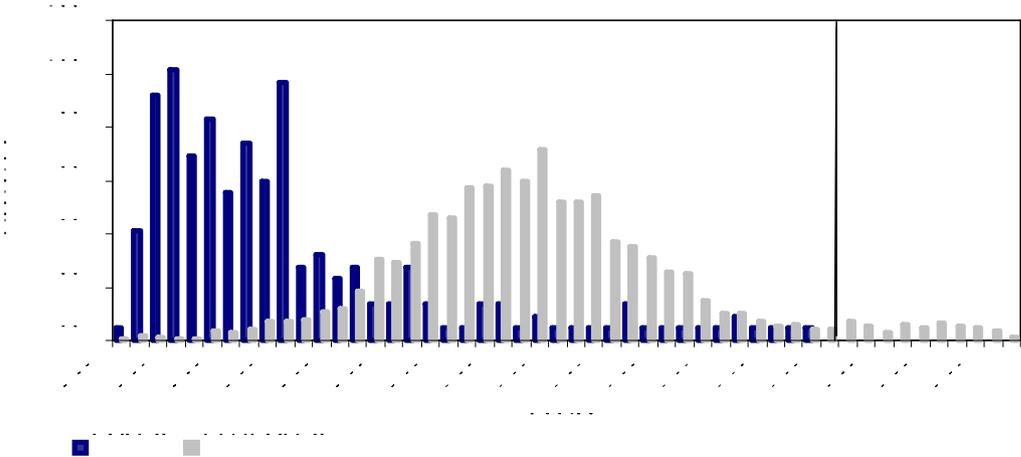
Para estimar los Modelos 4 y 5 hemos reincorporado a los deudores desaparecidos, asignándoles una clasificación de default o no-default siguiendo distintas reglas. Al hacer esto, en realidad estamos afectando los indicadores de performance. En el Modelo 4, los indicadores de performance se ven mejorados como resultado de marcar como default a los deudores que obtienen los scores más bajos con el Modelo 1. El Gráfico 3 muestra la división hecha con el Modelo 1 (los deudores que se marcan como default son los que obtienen un score con el Modelo 1 que es más bajo que el score promedio de los deudores que permanecen y que hacen default).

Gráfico 3: Regla de asignación de clasificación default o no--default a los deudores removidos según los scores del Modelo 1



Como los parámetros estimados del Modelo 4 son diferentes de los del Modelo 1, luego de correr el Modelo 4 la distribución de los scores para los defaults y no-defaults en el grupo de deudores removidos, también cambia. Esto se ilustra en el Gráfico 4.

Gráfico 4: Distribuciones de los Defaults y No-Defaults en el grupo de deudores removidos, de acuerdo con los scores del Modelo 4



La separación entre ambas distribuciones ahora no es exacta, como consecuencia de la diferencia entre el Modelo 1 y el Modelo 4. También puede observarse que los deudores reincorporados tienden a obtener mejores scores (indicando menor riesgo) en el Modelo 4 que en el Modelo 1. Esto, sostenemos, es el resultado de que la regla sea demasiado conservadora. Como explicamos antes, hay motivos para creer que los deudores excluidos son mayormente cancelaciones, ventas o errores, más que pérdidas. En otros términos, creemos que la regla supone una tasa de default más alta que la real para los deudores excluidos, y por lo tanto una regla así significa mayor probabilidad de asignar una calificación de default a un no-default que lo contrario (marcar como no-default a un default), lo cual sería aceptable desde un punto de vista del supervisor aunque no sería tan bueno en términos del modelo.

Como se describió, los deudores que participan y los que desaparecen tienen características similares (Anexo 1) En consecuencia los deudores reincorporados en el Modelo 4 tenderán a estar rankeados casi en el mismo orden que en el Modelo 1. De allí que esta técnica implica incorporar un sub-conjunto de observaciones que, por construcción, tienen un Accuracy Ratio casi perfecto, aún luego de estimar el Modelo 4. Por lo tanto, estamos muy probablemente incrementando el AR y otros estadísticos de poder discriminatorio para toda la muestra. Esto no está para nada claro con otras reglas supuestas.

Intentar tomar en cuenta este efecto es un desafío. En el caso tradicional de sesgo de selección puede demostrarse que (i) el sesgo es considerable cuando se testea poder discriminatorio y (ii) el sesgo puede ser positivo o negativo, es decir, no hay una regla simple para corregir este sesgo²⁰. En nuestro caso, cuando el sesgo no tiene una dirección sino que puede tener ambas direcciones, sería aún más difícil encontrar una técnica correctiva, especialmente tomando en cuenta la baja calidad de las variables explicativas para modelar el proceso de selección.

En consecuencia, hemos calculado estadísticos “duros” de poder discriminatorio para diferentes modelos usando sólo el conjunto de observaciones de deudores que están presentes en ambos puntos del tiempo. Si estos estadísticos mejoran, muy probablemente el poder total del modelo ha mejorado. Si los estadísticos no muestran una mejora, se debería medir la diferencia entre los parámetros estimados para evaluar si vale la pena reincorporar a los deudores que desaparecen.

Cuadro 7: Medidas estadísticas “duras” de poder discriminatorio considerando los deudores presentes en Dic-99 y Dic-00

<i>Deudores Comerciales</i>	<i>Accuracy Ratio (AR)</i>	<i>Indice ROC</i>	<i>Indice Pietra</i>
Modelo 1	62.3%	81.2%	0.168
Modelo 2	47.4%	73.7%	0.136
Modelo 3	61.7%	80.8%	0.167
Modelo 4	62.2%	81.1%	0.168
Modelo 5	62.0%	81.0%	0.167
Max (mayor poder discriminatorio)	100.0%	100.0%	>
Min (menor poder discriminatorio)	0.0%	50.0%	<
Referencias			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitative Default Risk Models: A validation methodology (Moody's)	50-75%		

²⁰ Kraft, Kroisandt and Müller (2004)

Nuevamente, los estadísticos de poder discriminatorio muestran que el Modelo 2 luce como una solución pobre para tomar en cuenta el sesgo debido a las exclusiones, dado que los valores se deterioran considerablemente. Entre otros modelos, los estadísticos “duros” (calculados usando sólo deudores que permanecen pero con los modelos estimados con las muestras incrementadas) muestran que el Modelo 4 puede conservar los valores del Modelo 1, mientras que el resto muestra un deterioro, aunque menor. Como se vio en el Cuadro 6 y como es de esperar, los estadísticos “suaves” (medidos sobre la muestra incrementada) son claramente mejores para los Modelos 4 y 5, siempre en comparación con el Modelo 1.

Como resultado, el Modelo 4 se selecciona para clasificar a los deudores en grados y testear la calibración, así como también para desarrollar tests fuera de la muestra.

Si los estadísticos “duros” no mostraran ninguna mejora en ninguno de los modelos, debería medirse la diferencia entre los parámetros estimados para evaluar si vale la pena reincorporar a los deudores que desaparecen. En el caso de nuestro Modelo 4 seleccionado, aún cuando la versión “dura” de los estadísticos de poder discriminatorio son muy similares a los del Modelo 1, los parámetros muestran que las características del deudor reciben diferente consideración bajo los dos modelos. Por ejemplo, un deudor en una sucursal de un banco extranjero cuya situación previa era 3 (seis meses antes), pero está en situación 1 al comienzo del período y también está en situación 1 en otros bancos (dos bancos más tienen acreencias con este deudor), no tiene garantías, su saldo de deuda con el banco es de \$500.000 y \$700.000 con el total del sistema, obtendrá una probabilidad del default estimada de 0,3% con el Modelo 1, y de un tercio de ese valor, 0,1% con el Modelo 4. Por el otro lado, consideremos un deudor en un banco cooperativo, cuya situación seis meses antes era 2 con el banco, lo mismo que en el momento inicial, mientras que su peor rating en el sistema es 4, tiene otros ocho bancos acreedores y una línea informada en default en el sistema (en el momento inicial y también seis meses antes). Se han dado garantías que cubren el 100% del crédito, el saldo de deuda con el banco es de \$50 millones y con el sistema \$75 millones. Este deudor obtendrá una PD estimada de 23% con el Modelo 1, mientras que el Modelo 4 indicará 30% (30% de incremento).

Calibración

El sistema de rating estimado a partir del Modelo 4 se seleccionó por su ajuste econométrico y su poder discriminatorio. La calibración debe chequearse para tener una completa validación del modelo. Un sistema de ratings está calibrado cuando sus PDs pronosticadas son similares a las tasas de default observadas en cada grado del sistema de ratings.

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 entidades deben distribuir sus exposiciones en las calificaciones sin concentraciones excesivas. Para cumplir con este requerimiento debe tener como mínimo ocho calificaciones: siete para deudores que no están en default y uno para deudores en default.

En consecuencia, hemos establecido un sistema de rating de ocho grados de calificaciones en base a los resultados del Modelo 4. Como este modelo usa un método estadístico Probit, las PDs asociadas a cada deudor se pueden calcular de manera directa. En este punto, no consideramos distintos tipos de crédito, luego las PDs representan el riesgo del deudor²¹. Hemos construido los

²¹ La dimensión de diferentes tipos de crédito será considerada en futuras investigaciones.

grados del sistema de rating a partir de los resultados para los deudores comerciales de todo el sistema financiero, no con los deudores de un banco en particular.

Así, deudores comerciales con similares probabilidades de default (o scores) fueron agrupados en los ocho grados del sistema de rating. La técnica para determinar los rangos de cada grado se basó principalmente en la comparación gráfica entre los promedios móviles de las tasas de default y las PDs estimadas.

Tests de Calibración

El documento *Studies on the Validation of Internal Rating Systems* (2005)²² muestra distintos tests para evaluar la calibración de un sistema de ratings. Aquellos cuyo uso está más difundido, los tests Binomial y Hosmer-Lemeshow (H-L), asumen que los eventos de default son independientes. Mientras que el test Binomial examina cada grado del sistema de rating de manera separada, el test H-L considera la calidad de todas las categorías del sistema de rating al mismo tiempo (Anexo 2).

El mencionado documento indica que el test Binomial es, a un nivel fijo, el más potente para testear el ajuste de las PDs. También se señala que la independencia de los eventos de default es un supuesto demasiado fuerte y que, empíricamente, es normal observar bajas correlaciones. Para este estudio hemos decidido tomar un enfoque conservador y aplicar el test Binomial. Al no considerar que los eventos de default están correlacionados estamos aumentando la probabilidad de rechazar, de manera injustificada, la hipótesis nula (H_0 = las PDs estimadas son correctas). Luego, si el test Binomial indica que los grados del sistema de rating son correctos, este es un resultado robusto.

En el test Binomial, el número observado de deudores que hacen default en cada grado del sistema de rating, N_D , debe compararse con el estadístico k^* para un determinado nivel de confianza. Si $N_D < k^*$ la hipótesis nula de PDs correctas no puede ser rechazada a dicho nivel de confianza.

El Cuadro 8 muestra que las categorías del sistema de rating construidas con las PDs estimadas son correctas, de acuerdo al test Binomial.

Cuadro 8: Medidas de Calibración del Modelo 4

Categorías		Deudores		Defaults				Resultados de Default
Nº	Rango de PD (%)	N	(%)	N_D	N_D/N (%)	Estimada de PD (%)	k^* (99% nivel de confianza)	
1	0-1.5	1686	10.13	10	0.59	1.01	27	Correcto
2	1.5-2.7	3101	18.63	55	1.77	2.12	85	Correcto
3	2.7-3.75	2618	15.73	75	2.86	3.19	105	Correcto
4	3.75-4.75	1815	10.90	64	3.53	4.24	97	Correcto
5	4.75-5.6	1254	7.53	78	6.22	5.16	83	Correcto
6	5.6-6.31	859	5.16	64	7.45	5.94	67	Correcto
7	6.31-18.3	3241	19.47	322	9.94	9.47	346	Correcto
8	18.3-100	2070	12.44	897	43.33	42.96	942	Correcto

²² Publicación N°14 del Comité de Basilea de Supervisión Bancaria, mayo de 2005.

Suponiendo de nuevo independencia en los eventos de default, se calcula el estadístico Hosmer-Lemeshow. Como resultado no se rechaza la hipótesis nula de que las PDs promedio de los grados del sistema de rating son las verdaderas probabilidades.²³

Entonces, la calibración de los grados del sistema de rating construidos a partir del Modelo 4 parece ser aceptable, considerando cada categoría de manera individual y todas ellas de manera simultánea.

Desempeño del Modelo 4, test “fuera de la muestra”

Para validar el uso de un modelo estadístico en un proceso de rating, es necesario chequear su desempeño incluyendo tests fuera del período (“out-of-time”) y fuera de la muestra (“out-of-sample”). Por ese motivo, y siempre cuidando que la muestra que se utilice para estimar los componentes de riesgo tenga suficiente cantidad de exposiciones como para que las estimaciones sean confiables y robustas, también es deseable separar algunas observaciones de la muestra empleada en la estimación para realizar estos tests.²⁴

Para estimar las PDs se tomó una muestra secuencial del 75% de los deudores-banco en la CD. El test “fuera de la muestra” se computaría sobre el resto. Este test consiste en calcular las PDs usando los coeficientes estimados del Modelo 4 y las variables explicativas de los deudores-banco “fuera de la muestra”. Luego, se construye el sistema de ratings de los deudores “fuera de la muestra”.

Cuadro 9: tests de poder discriminatorio “fuera de la muestra”

<i>Deudores Comerciales</i>	<i>Accuracy Ratio (AR)</i>	<i>Indice ROC</i>	<i>Indice Pietra</i>
Modelo 4	62.9%	81.4%	0.172
Referencias			
Studies on the Validation of IRB (Basel Committee)	50-80%		
Benchmarking Quantitative Default Risk Models: A validation methodology (Moody’s)	50-75%		

Cuadro 10: tests de calibración “fuera de la muestra” para el Modelo 4

<i>Categorías</i>		<i>Deudores</i>		<i>Defaults</i>				<i>Resultados de Calibración</i>
<i>Nº</i>	<i>Rango de PD (%)</i>	<i>N</i>	<i>(%)</i>	<i>N_D</i>	<i>N_D/N (%)</i>	<i>Media estimada PD (%)</i>	<i>k* (99% nivel de confianza)</i>	
1	0-1.5	230	4.13	1	0.43	1.09	6	Correcto
2	1.5-2.7	777	13.94	16	2.06	2.14	26	Correcto
3	2.7-3.75	926	16.62	29	3.13	3.21	42	Correcto
4	3.75-4.75	719	12.90	18	2.50	4.19	43	Correcto
5	4.75-5.5	477	8.56	19	3.98	5.13	36	Correcto
6	5.5-6.32	413	7.41	29	7.02	5.91	36	Correcto
7	6.32-18.3	1368	24.55	124	9.06	9.33	153	Correcto
8	18.3-100	662	11.88	270	40.79	39.14	288	Correcto

²³ La hipótesis nula no se rechaza a un nivel de confianza del 95%, usando el estadístico de Hosmer-Lemeshow con un valor de 15.36 ($\Pr(\chi^2_9) = 0.0815$).

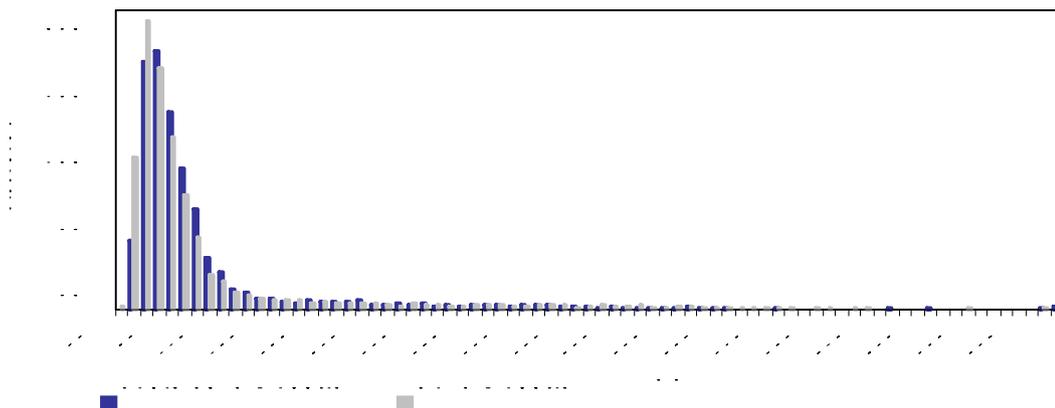
²⁴ Basilea II también incluye este requerimiento. Ver *International Convergence of Capital Measurement and Capital Standards*, Comité de Basilea para la Supervisión Bancaria (2004), párrafo 251.

El Cuadro 9 presenta los resultados de los tests de poder discriminatorio del sistema de rating en el ejercicio “fuera de la muestra”. En este caso los deudores “fuera de la muestra” tienen una peor performance en términos de poder discriminatorio en comparación con los deudores usados en la estimación del modelo (“in-sample”). El AR es 62% aproximadamente para los deudores “fuera de la muestra” mientras que 68% para los “in-sample”. Sin embargo, esta performance es igualmente aceptable teniendo en consideración la práctica habitual y la literatura. La calibración “fuera de la muestra” se exhibe en el Cuadro 10, donde se ve que los grados de riesgo se comportan bien también para los deudores “fuera de la muestra”.

Finalmente, el Gráfico 5 muestra la distribución de las PDs de las observaciones de la muestra y fuera de ella. Las distribuciones de las PDs son similares aunque la distribución de los deudores “in-sample” con bajas PDs está levemente más concentrada en los valores inferiores (a la izquierda).

En conclusión, el sistema de rating tiene una performance aceptable, considerando los tests “fuera de la muestra”.

Gráfico 5: comparación de distribuciones de PD dentro y fuera de la muestra



VI - Conclusiones

Las conclusiones que surgen de este estudio pueden agruparse en torno a tres temas: (i) el uso de un modelo como el desarrollado aquí, lo cual es de interés fundamentalmente para reguladores y supervisores; (ii) conclusiones técnicas en relación al problema de los deudores que desaparecen o problemas similares con las bases de datos, que son de interés para reguladores y la industria; y (iii) una aplicación práctica de las metodologías de validación.

Los resultados del modelo son muy buenos a pesar de la selección limitada de variables explicativas, que se refieren básicamente al saldo de deuda y su distribución (en la institución y en el sistema), garantías y calificaciones de riesgo (actuales, en otras instituciones o en el pasado).

No hemos usado información sobre sector de actividad, ratios financieros u otra información de balance, así como tampoco otras variables que están en la CD pero que desafortunadamente presentan mala calidad. Claramente, un banco podría usar un conjunto de variables mucho más amplio sobre sus deudores; un modelo como el desarrollado en este estudio no es un arquetipo a ser usado por los bancos.

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) Para evaluar la regulación y tomar decisiones informadas sobre ella. Por ejemplo, el nivel de provisiones regulatorias podría compararse con las tasas de default esperadas (que se pueden obtener a partir de las PDs, y de las tasas de recupo); podría evaluarse la exigencia de capital por riesgo de crédito desarrollando un modelo de portafolio de riesgo de crédito a partir de las PDs obtenidas. En países como Argentina donde existe un sistema de clasificación estándar implementado por la regulación del Banco Central, un modelo como el de este estudio también podría usarse para evaluar cambios en dicho sistema. En nuestro caso, por ejemplo, puede concluirse que la primer situación del sistema de calificación debería abrirse en calificaciones más precisas y pequeñas, para evitar la concentración de las deudas y para separar el amplio rango de riesgos que son asignados al mismo rating.
- (ii) Para usar los resultados como un punto de referencia contra el cual comparar los modelos desarrollados por los bancos.
- (iii) Para hacer ajustes a los burós 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 analizado la aparentemente inocua 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. Hemos mostrado que esto puede introducir un sesgo que no debería ser ignorado y que es difícil de corregir sin distorsionar los estadísticos de performance. En consecuencia, es de importancia primordial asegurar que las bases de datos de riesgo de crédito eliminen cualquier “agujero” que dificulte el seguimiento en el comportamiento de los individuos.

Este último 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, como estar incompleta, tener mala calidad, o cualquier otro defecto. Esto podría conducir a un sesgo no deseado o resultar en la selección deliberada de los deudores para mejorar las medidas de validación de los modelos de scoring crediticios usados por los bancos. En consecuencia, esta práctica debería ser evitada en virtud de la precisión del modelo y no debería ser aceptada por los supervisores.

Adicionalmente, este estudio aplica las técnicas de validación propuestas por el Comité de Basilea referentes a la calibración y el poder discriminatorio. Los estadísticos de performance que se obtienen son muy buenos, a pesar de las limitaciones de las variables explicativas. Será necesario que los supervisores que evalúen los modelos de scoring crediticios de los bancos entiendan las herramientas de validación, como las presentadas en este documento.

Anexo 1: Características de los participantes y deudores eliminados de la muestra

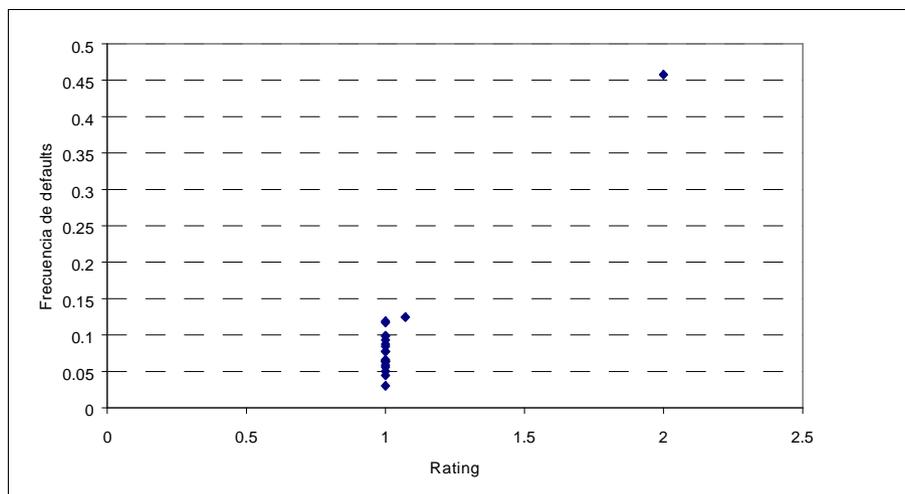
Número de obs/porcentaje	Participantes	No Participantes
Rating previo		
0	1,652 11.8%	590 22.1%
1	11,706 83.7%	1,990 74.6%
2	502 3.6%	81 3.0%
3	58 0.4%	5 0.2%
4	30 0.2%	1 0.04%
5	36 0.3%	1 0.04%
Rating		
1	13,231 94.6%	2,533 94.9%
2	753 5.4%	135 5.1%
Peor rating		
1	11,624 83.1%	2,260 84.7%
2	1,377 9.8%	209 7.8%
3	414 3.0%	83 3.1%
4	338 2.4%	74 2.8%
5	231 1.7%	42 1.6%
Bancarización		
1	1,652 11.8%	579 21.7%
3	3,903 27.9%	541 20.3%
5	2,996 21.4%	401 15%
7	1,837 13.1%	315 11.8%
9	1,298 9.3%	265 9.9%
10	2,298 16.4%	567 21.3%
Sector		
Sector privado-nacionales	13,979 99.96%	2,391 89.6%
Sector Privado-residentes en el extranjero	5 0.04%	277 10.4%
Total	13,984 100%	2,668 100%
Medias		
	Participantes	No Participantes
Garantía	33%	19%
Ldeuda_banco	6.39	6.65
Ldeuda_sistema	7.48	7.97
Porcentaje_default	1.3%	1.7%
Nlineas_default	7.6%	10.6%
Nlineas_previas_default	5.6%	5.6%

Anexo 2: Variables explicativas y relaciones univariadas a priori con las tasas de default

Las definiciones de las variables explicativas se incluyen abajo, en algunos casos junto con gráficos de dispersión que ilustran la relación con la frecuencia de defaults. Estos gráficos muestran relaciones univariadas a priori, i.e., no se controla por otras variables. Para construir cada gráfico, todas las observaciones son primero ordenadas por la variable en el eje x. Luego, las observaciones se dividen en 20 cuantiles y se calculan y grafican el valor promedio de la variable en el eje x junto con la tasa de default para cada cuantil. En otras palabras, cada punto en el gráfico representa 5% de las observaciones. De esta manera, los gráficos muestran no sólo la relación a priori entre las variables y la tasa de default, sino también la distribución de las observaciones en distintos valores de la variable x.

Default: variable categórica para identificar los deudores que hacen default. Esta variable toma un valor 0 si el deudor es calificado en las situaciones 1 o 2 y toma el valor 1 si el deudor está en las situaciones 3, 4 o 5 en diciembre de 2000.

Rating: variable categórica que registra la calificación del deudor o “situación” (a nivel de deudor-banco). Toma valores 1 o 2 si el deudor está clasificado en situaciones 1 o 2, respectivamente, en diciembre de 1999.

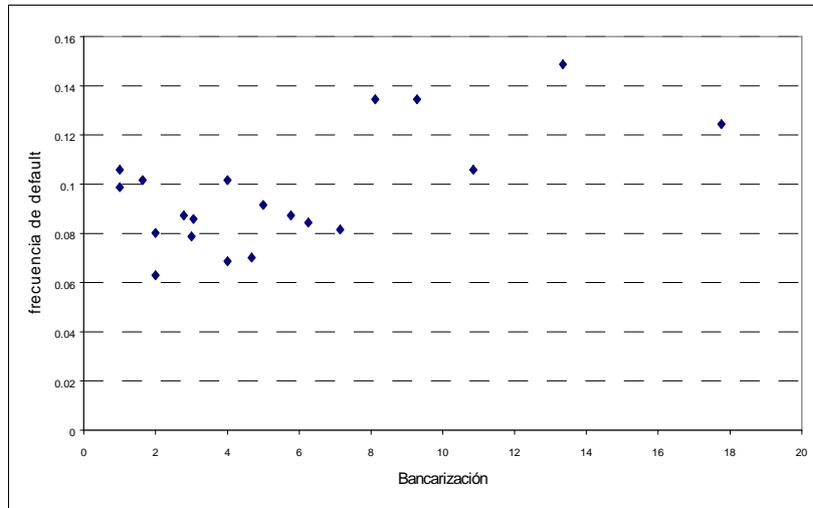


Rating_previo: variable categórica que considera la calificación a nivel de deudor-banco. Esta variable toma valores entre 0 y 5 representando la calificación del deudor en junio de 1999. Toma el valor 0 si el deudor no estaba registrado en junio de 1999.

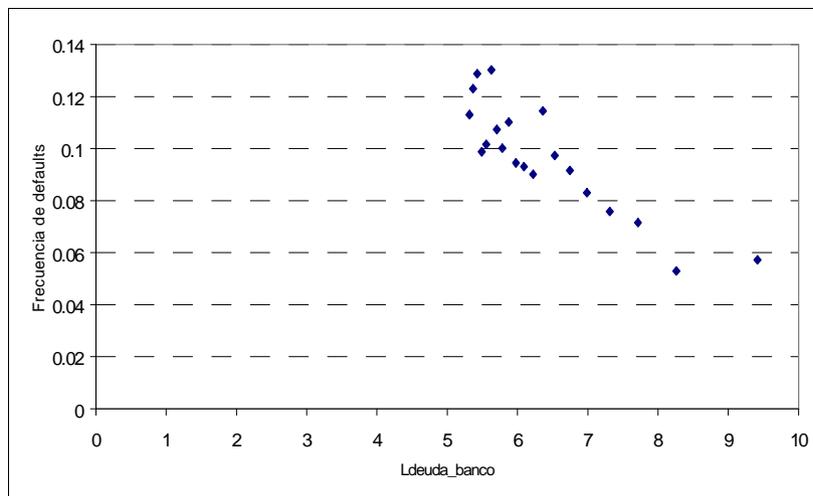
Peor_rating: variable categórica que toma valores de 1 a 3 representando la peor (mayor) calificación de riesgo del deudor-banco en diciembre de 1999. Un valor de 3 identifica a los deudores clasificados de 3 a 5.

Tipo_institución: Variable categórica que clasifica a las instituciones financieras de acuerdo a su capital de origen. Esta variable está compuesta por 7 *dummies* excluyentes: bancos públicos nacionales, bancos locales de capital extranjero, bancos privados cooperativos, bancos privados nacionales, bancos públicos provinciales y municipales, sucursales de instituciones financieras extranjeras y otras instituciones financieras.

Bancarización: variable categórica que considera el número de instituciones financieras acreedoras del deudor. Esta variable toma el valor 1 si una institución tiene acreencias con el deudor, 3 si ello es cierto para dos o tres instituciones, 5 para cuatro o cinco (el deudor-banco es deudor de cuatro o cinco bancos), 7 para seis o siete, 9 para ocho o nueve y 10 para deudores-banco que tienen deudas con diez o más entidades financieras.



Ldeuda_banco: el logaritmo del monto de deuda pendiente de pago con una institución financiera específica de un deudor particular en diciembre de 1999.



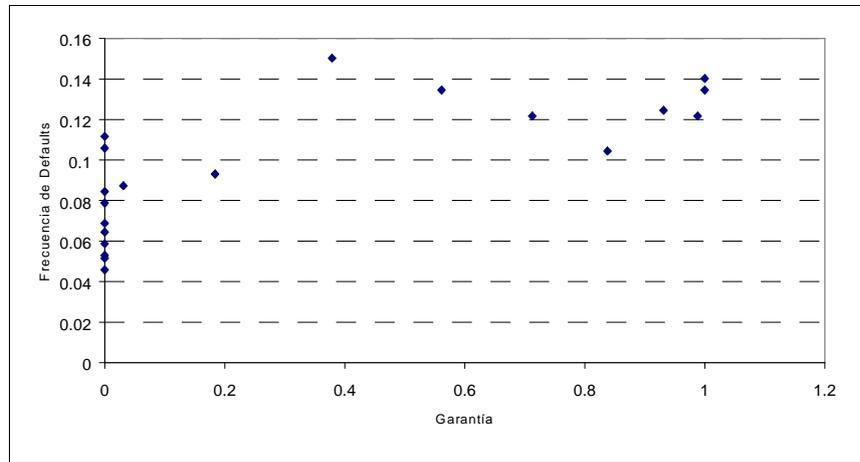
Ldeuda_sistema: el logaritmo del monto de deuda pendiente de pago con el sistema bancario de un deudor particular en diciembre de 1999.

Porcentaje_default: proporción de las exposiciones del sistema bancario con un deudor particular que es informada en default en diciembre de 1999.

Nlineas_default: número de líneas crediticias del deudor clasificadas en default en el sistema bancario en diciembre de 1999.

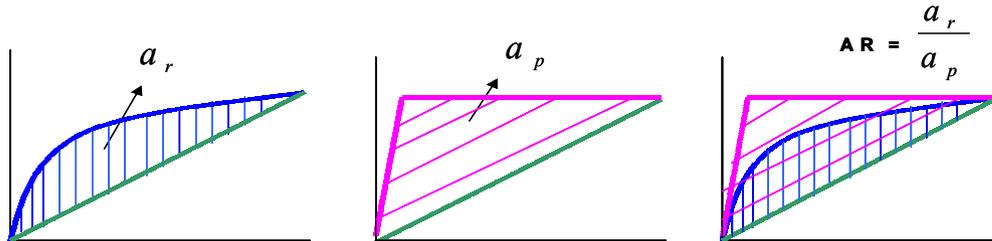
Nlineas_previas_default: número de líneas crediticias del deudor clasificadas en default en el sistema bancario en junio de 1999.

Garantías: porcentaje de las exposiciones que están garantizadas en diciembre de 1999.

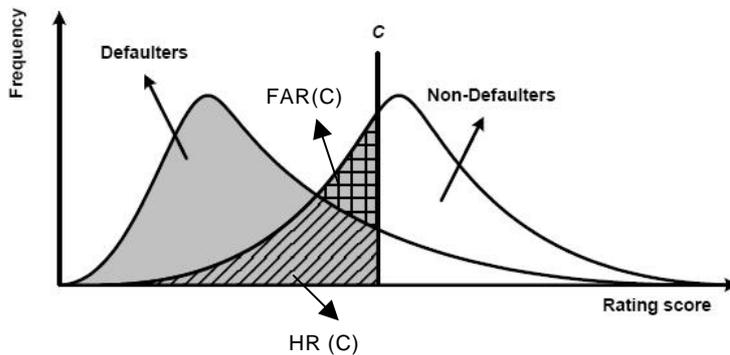


Anexo 3: Medidas estadísticas de discriminación y calibración

Accuracy Ratio (AR): se define como el ratio del área a_r comprendida entre la curva CAP del modelo de rating que está siendo validado y la curva CAP del modelo aleatorio, y el área a_p comprendida entre la CAP del modelo de rating perfecto y la CAP del modelo aleatorio. Este índice toma valores entre 0 y 1, cuanto mejor es el modelo de rating más cerca está de 1.



Índice ROC: la construcción de la curva ROC considera las distribuciones de los scores de rating de los deudores que hacen default y de los que no. Las distribuciones que se muestran en el siguiente gráfico deberían estar lo más separadas posible:



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 su score es superior a C . 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:

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

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.

Índice Pietra: geométricamente, este índice puede estimarse como la máxima área de un triángulo ubicado entre la curva ROC y la diagonal. De manera equivalente, el Índice Pietra puede calcularse como la máxima distancia entre la curva ROC y la diagonal. En el caso de una curva cóncava puede calcularse como:

$$\text{Índice Pietra} = \sqrt{2}/4 \max_c |\text{HR}(C) - \text{FARC}(C)|$$

Test Binomial: este es un test que considera la calibración de cada categoría de un sistema de ratings. Este test está construido bajo el supuesto que los eventos de default son independientes, lo que permite emplear la distribución Binomial.

La hipótesis nula (H_0) es que las PDs del sistema de ratings son correctas. Para un nivel de confianza q (e.g. 99%), la hipótesis nula se rechaza si el número N_D de deudores que hacen default en una categoría del sistema de ratings es mayor que el valor crítico k^* .

$$k^* = \Phi^{-1}(q) \sqrt{nPD(1 - PD)} + nPD$$

En consecuencia, el test Binomial es simple pero está basado en el supuesto de independencia de los eventos de default. Empíricamente, se sabe que los defaults están correlacionados con un bajo coeficiente de correlación, por lo que el estadístico cambia si ello se considera.

La presencia de correlación implica que son probables grandes desviaciones entre la PD estimada y las tasas de default. De esta manera, aún para muestras con un gran número de deudores, la correlación implica que no funcionará la ley de los grandes números y las PD estimadas no se aproximarán a las tasas reales de default.

Sin embargo, como los valores críticos de los tests de PD que incorporan la correlación tienden a ser mayores que los valores críticos del test Binomial (con el supuesto de independencia), una aplicación de este último sería conservadora en presencia de correlación. Así, el test Binomial lleva a un rechazo temprano de H_0 en un contexto de correlación. La verdadera magnitud del Error Tipo 1 (rechazo de H_0 cuando es verdadera) será mayor que el nivel nominal indicado por el test. En consecuencia, testear la calibración de cada categoría del sistema de rating con el test Binomial, como se hace en este documento, implica punto de vista conservador, en el sentido de que si hay una pequeña correlación en los eventos de default podríamos estar rechazando H_0 cuando es verdadera, diciendo que las PDs estimadas no son correctas cuando en realidad lo son.

Test Chi-cuadrado (o Hosmer-Lemeshow)

Este test asume también independencia de los eventos de default, pero testea todas las categorías del sistema de rating de manera simultánea en contraposición al test Binomial que es un test para cada categoría.

Si las probabilidades de default pronosticadas fueran: p_0, \dots, p_k para k categorías de rating, n_i el número de deudores en la categoría i y θ_i el número de deudores que hacen default en la categoría i , el estadístico de este test se define como:

$$T_k = \sum_{i=0}^k \frac{(n_i p_i - \theta_i)^2}{n_i p_i (1 - p_i)} \sim \chi_{k+1}^2$$

Cuando $n_i \rightarrow \infty$, por el teorema central del límite, la distribución de T_k convergerá a una distribución χ^2_{k+1} si todas las p_i son las verdaderas probabilidades de default.

Bibliografía

- Kiefer N. M. and Larson E. C., *"Specification and Informational Issues in Credit Scoring"*, Office Comptroller of the Currency, E&PA Working Paper 2004-5, Diciembre 2004.
- Hand D.J. and Henley W.E., *"Statistical methods in Consumer Scoring: A Review, Journal of the Royal Statistical Society. Series A (Statistics in Society)"*, Vol. 160, N°3 , 1997, 523-541.
- Basel Committee on Banking Supervision, *"Studies on the Validation of Internal Rating Systems"*, Working Paper N° 14, Mayo 2005.
- Berger A.N and G.F. Udell, *"Relationship Lending and Lines of Credit in Small Firm Finance"*, Journal of Business, 68(3), July 1995, pag. 351-81.
- Berglof E. and Von Thadden, *"Short Term Versus Long Term Interest: Capital Structure with Multiple Investor"s*, Quarterly Journal of Economics, 109, 1994, pag. 1055-1084.
- Basel Committee on Banking Supervision, Bank of International Settlements, *"International Convergence of Capital Measurement and Capital Standards. A Revised Framework"*, Junio 2004.
- Basel Committee on Banking Supervision, Bank of International Settlements, *"Studies on the Validation of Internal Rating Systems"*, Working Paper N° 14, Mayo 2005.
- Balzarotti V., Castro C. and Powel A., *"Capital Requirements in Emerging countries: Calibrating Base II using Historical Argentine Credit Bureau Data and Credit Risk+"*, 2004, Business School Working Papers, Universidad Torcuato Di Tella.
- Balzarotti V., Castro C. and Powell A., *"Reforming capital Requirements in Emerging Countries"*, 2002 , Business School Working Papers, XIX, Universidad Torcuato Di Tella.
- Moody's Inverstors Service, *"Benchmarking Quantitative Default Risk Models: A Validation Methodology"*, Marzo 2000.
- Schechtman, R., Salomão García, V., Mikio Koyama, S. and Cronemberger Parente, G., *"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, Diciembre 2004.
- Verstraeten, G. and Van den Poel, D., *"The Impact of Sample Bias on Consumer Credit Scoring: Performance and Profitability"*, Working Paper Universiteit Gent, Marzo 2004.
- Dubin, J. and Rivers, D., *"Selection Bias in Linear Regression Logit and Probit Models"*, Sociological Methods and Research, Vol. 18, 2&3, Noviembre 1989/Febrero1990.
- Barron, J.M. and Staten M., *"The Value of Comprehensive Credit Reports: Lesson for the U.S. Experience Summary"*, Word Bank, Mimeo, 2000.
- Crook, John, *"Adverse Selection And Search In The Bank Credit Card Market"*, Credit Research Centre, University of Edinburgh, June 2002.