Tabla de Contenidos

1. Introducción 1

2. Estado de la tecnología 3

2.1 Prácticas tradicionales para medir el riesgo crediticio 3

3. Solución Propuesta 5

3.1 Definiciones Básicas 5

3.1.1 Situación en el BCRA 5

3.1.2 Default 5

3.1.3 Horizonte de Predicción 5

3.1.4 Período de recolección de datos 6

3.1.5 Base de Datos 6

3.2 Desarrollo del Modelo 7

3.2.1 Muestras 7

3.2.3 Variables Independientes 7

3.2.3 Metodología 7

4. Validación del modelo 11

4.1 Validación in-sample: Indicadores 11

4.2 Validación out-of-sample 12

4.3 Poder Discriminatorio 16

4.3.1 Test de Kolmogorov-Smirnov 16

4.3.2 Curva CAP (Cumulative Accuracy Profile) y Ratio de Precisión 19

5. Aplicación práctica del modelo 21

5.1 Determinación del corte de score en función del riesgo de incobrabilidad junto con otras variables 21

5.2 Simulación de la utilidad en función de las tasas pasiva y activa 25

6. Conclusiones 29

7. Bibliografía..... 31

8. Anexos..... 33

1. Introducción

Los bancos, financieras y cooperativas de crédito necesitan evaluar el riesgo crediticio con el fin de preservar el capital y generar utilidades. Es por esto que a la hora de dar un préstamo le exigen a las empresas información detallada acerca de su situación financiera. Esta información es utilizada por los bancos para determinar la calificación crediticia de la empresa con el fin de decidir si la solicitud de crédito será aceptada o rechazada. Los solicitantes que alcancen la calificación mínima verán aceptada su solicitud mientras que aquellos que no la alcancen tendrán denegada su solicitud de crédito o serán sujetos a un análisis más detallado, el cual puede implicar un ajuste de la tasa de interés.

El objetivo de este escrito es presentar un modelo estadístico para establecer la probabilidad de insolvencia de las empresas y así facilitar la decisión de los prestadores. A su vez se busca identificar las variables que mejor predicen la insolvencia o el default, entendidos como el atraso en los pagos de la deuda. En segunda instancia se buscará validar el modelo a través de diferentes criterios y métodos estadísticos. Finalmente, se realizará una aplicación práctica del modelo en la cual se planteará un criterio para aceptar o rechazar solicitudes de crédito de manera tal que se maximice la utilidad de la entidad prestadora.

¿Por qué surge interés por estudiar la medición del riesgo crediticio?

"Las grandes agencias de calificación crediticia se encuentran de nuevo en el punto de mira. Su fracaso en detectar señales tempranas de los problemas de Enron, WorldCom y otras empresas de alto vuelo han hecho que los reguladores se pregunten de dónde obtienen la información las agencias, cómo la analizan y si sus opiniones valen el papel en el que han sido escritas." [Brealey & Myres & Allen, 2006]

Saunders, explica las razones por las cuales surge el reciente interés por este tema:

1) Se percibió un fuerte aumento en el número de bancarrotas en los últimos años.
2) Hubo un significativo incremento en la cantidad de empresas que reciben crédito, inclusive aquellas con niveles de riesgo considerables.
3) Los márgenes de los préstamos son cada vez más competitivos.
4) La crisis bancaria en países desarrollados como Suiza y Japón han demostrado lo complejo que es predecir el valor de propiedades y activos reales.
5) El avance tecnológico de los últimos años permite el desarrollo de modelos predictivos más complejos y poderosos.
6) Las exigencias impuestas por las Normas de Basilea sobre las políticas de regulación crediticia de los Bancos Centrales de todo el mundo.

Por las razones expuestas y dado la dinámica del tema es necesario mejorar continuamente el sistema de *scoring* de riesgo crediticio.

Es interesante hacer un análisis de este tema desde el punto de vista de la ingeniería ya que se requiere un enfoque sistémico. El modelo de *scoring* se desarrolla realizando sucesivas pruebas como por ejemplo la inclusión de diferentes variables en el modelo hasta alcanzar un nivel óptimo de predicción. Se realiza un proceso de mejora continua donde se realizan propuestas cuyo resultado se mide y si este es positivo se adopta lo sugerido.

¿Cuál será la contribución a la solución del problema?

En este escrito se propone una solución original al problema de gestionar el riesgo crediticio mediante el desarrollo de un modelo, la validación del mismo y su aplicación práctica. Como aporte a destacar, se provee una herramienta que permite a la entidad prestadora determinar el otorgamiento de crédito de manera tal que maximice su utilidad. La entidad debe ingresar parámetros tales como la tasa pasiva, la tasa activa, el monto de los préstamos que pretende otorgar y el sistema le indica el *score* a partir del cual puede aceptar las solicitudes y por debajo del cual debe rechazarlas para optimizar su resultado.

El modelo de *credit-scoring* permite que se tomen decisiones sobre las solicitudes de crédito de manera automática, homogénea, objetiva y más rápida que en un sistema experto, donde el oficial de crédito toma la decisión en función de su experiencia.

Asimismo, se plantea como objetivo identificar a las empresas con una buena situación crediticia para que reciban una tasa de interés menor y que no deban pagar un excedente que compense el riesgo de las empresas con situación crediticia comprometida. Esto trae aparejado para la economía en general una mejor asignación de recursos, un menor costo del dinero y consecuentemente una reducción en los índices de precios.

2. Estado de la tecnología

2.1 Prácticas tradicionales para medir el riesgo crediticio

Saunders plantea tres sistemas para abordar esta medición: los "*expert systems*", los "rating *systems*" y los "*credit-scoring systems*".

En un "*expert system*" el oficial de créditos, que pertenece a la compañía prestamista, basándose en su experiencia y juicio subjetivo, pondera ciertos factores claves para luego decidir acerca del otorgamiento del crédito. Los factores primordiales a analizar son cinco: el carácter del tomador del préstamo, es decir su reputación y su historial de repago; su capital en relación con el endeudamiento (coeficiente de endeudamiento); la capacidad de repago, determinada por la generación de utilidades para atender el pago de la deuda y sus intereses; las garantías, tanto las constituidas sobre determinados activos a favor de algunos acreedores o garantías de terceros; y el ciclo económico, es decir la etapa en el ciclo del negocio en el que se encuentra la compañía, en especial para industrias sujetas a ciclos.

Uno de los "*rating systems*" más antiguos es el elaborado por la U.S. Office of the Comptroller of the Currency (OCC). Este sistema es utilizado por los bancos y otros entes reguladores para evaluar la suficiencia de sus reservas para préstamos incobrables. El sistema en cuestión ubica un préstamo en una de las cinco categorías siguientes, a las que les corresponde un porcentaje de reserva específico: "*Other assets especially mentioned*", "*Substandard assets*", "*Doubtful assets*", "*Loss assets*", "*Pass/performing*".

La mayoría de los análisis de riesgo crediticio, tanto para consumidores como para préstamos comerciales, se basa en "*credit-scoring systems*". La idea es identificar previamente algunos factores clave que determinan la probabilidad de default y ponderarlos según su importancia. En algunos casos esta ponderación puede interpretarse como la probabilidad de default, en otros, como una forma de clasificación: ubica a un potencial tomador de crédito en una clase según su puntaje y los puntos de corte de ésta.

Resulta útil plantear un ejemplo de este tipo de modelos ya que el objetivo primordial de este estudio es hallar un algoritmo que ayude a los dadores de crédito a aceptar o rechazar una solicitud de crédito, dependiendo de la probabilidad de default del tomador.

El "*Z-score model*" de Altman (1968) se utiliza para clasificar a los prestatarios. A partir de una muestra que contiene empresas solventes y otras quebradas y utilizando regresiones lineales, el modelo que plantea Altman para clasificar préstamos comerciales, es el siguiente:

$$(1)\quad Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$$

Donde: X_1 = capital de trabajo/total activo

X_2 = resultados no asignados/total activo

X_3 = resultados antes de intereses e impuestos/total activo

X_4 = valor de mercado de acciones/total pasivo

X_5 = ventas/total activo

Si al ingresar los ratios contables de un prestatario (las X_i), ponderados por los coeficientes de la ecuación, ésta resulta en un valor de Z menor a un valor crítico predeterminado (1,81), el oficial de crédito clasificará al tomador como "malo" y rechazará su solicitud de crédito.

Acerca de las limitaciones de este modelo Alan Elizondo (2003) escribe en su libro "Medición Integral del riesgo de crédito" que un punto en contra del *Z-score* es que utiliza razones financieras, y éstas tienen un efecto de subestimación, es decir, en un análisis realizado a través de razones financieras, no se pueden identificar datos relevantes en el otorgamiento de crédito, como puede ser el tamaño de las empresas.

Además de lo expresado por Elizondo, el *Z-score* tiene dos tipos de limitaciones: limitaciones específicas, que surgen de contener únicamente variables contables, y limitaciones que son comunes a todo modelo lineal cuando se utiliza para predecir probabilidades.

Otro problema que presenta el *Z-score* es que los datos contables no se actualizan tan seguido como los datos de comportamiento crediticio cuyo atraso suele ser de unos pocos días.

El problema más grave del modelo lineal es que predice valores negativos o mayores a la unidad. Esto no tiene sentido si se tiene en cuenta que la probabilidad de un evento toma valores en un rango entre cero y uno. Así lo establecía el Ing. Roberto Mariano García: "a no dudarlo, el inconveniente más serio que presenta este modelo es el ya mencionado de arrojar frecuentemente valores absurdos para la probabilidad" [Modelo de Regresión Logística, 2005].

Para subsanar ambos tipos de limitaciones, en este estudio se proponen modelos de tipo logit con variables de diversa índole y no únicamente contable.

Entre las diferentes ventajas del modelo Logit sobre el resto de las metodologías existentes, Alan Elizondo menciona que "estos modelos permiten, además de obtener estimaciones consistentes de la probabilidad de incumplimiento, identificar los factores de riesgo que determinan dichas probabilidades, así como la influencia o peso relativo de éstos sobre las mismas" [Medición Integral del riesgo de crédito, 2003].

3. Solución Propuesta

3.1 Definiciones Básicas

3.1.1 Situación en el BCRA

El Banco Central de la República Argentina (BCRA) clasifica a las empresas según el atraso en el pago de sus deudas. Les asigna situaciones que van de la 1 a la 6, con el significado que se consigna en la siguiente tabla.

Tabla 1: Clasificación de deudores del BCRA		
Situación		Comprende los clientes que:
1	Normal	Atienden en forma puntual el pago de sus obligaciones o con atrasos que no superan los 31 días
2	Riesgo bajo	Registran incumplimientos ocasionales en la atención de sus obligaciones, con atrasos de más de 31 hasta 90 días
3	Riesgo medio	Muestran alguna incapacidad para cancelar sus obligaciones, con atrasos de más de 90 hasta 180 días
4	Riesgo alto	Con atrasos de más de 180 días hasta un año o que se encuentran en gestión judicial de cobro, en tanto no registren más de un año de mora
5	Irrecuperable	Insolventes, en gestión judicial o en quiebra con nula o escasa posibilidad de recuperación del crédito, o con atrasos superiores al año
6	Irrecuperable por disposición técnica	Deudores irrecuperables de entidades financieras liquidadas

3.1.2 Default

Para este análisis se considera que una empresa está en default cuando el BCRA la clasifica en situación 3, 4, 5 o 6 o tiene algún concurso o quiebra. Es decir cuando la empresa incurre en atrasos mayores a 90 días o se presenta en concurso o quiebra durante el año siguiente a la fecha de cálculo de su *score*.

Se llama entonces P_t a la probabilidad de default, la cual toma valores entre 0 y 1, donde 0 significa que la empresa t no caerá en default y 1 implica que la empresa caerá en default. En otras palabras, esta variable mide la probabilidad de que una empresa tenga un atraso en el pago de sus deudas mayor a 90 días o que tenga un concurso o quiebra durante los próximos 12 meses.

En este escrito se hace referencia al “default observado” que se refiere a un hecho real y al “default pronosticado” que es una predicción en base a una probabilidad calculada P_t.

3.1.3 Horizonte de Predicción

El Horizonte es el período en el cual se estima la probabilidad de default. Para este escrito se toma un año como horizonte. Es decir, el objetivo es determinar la probabilidad de default de las empresas un año después del momento en que son evaluadas.

3.1.4 Período de recolección de datos

Los datos que se utilizan como variables independientes se recaban desde cinco años antes de la fecha en la cual se mide el default que es la variable dependiente.

3.1.5 Base de Datos

Para realizar el modelo se utiliza una base de datos de empresas que fue cedida por una empresa de riesgo crediticio. La misma contiene datos demográficos y sobre comportamiento crediticio de todas las empresas de la Argentina. La base de datos se nutre de información de la Administración Federal de Ingresos Públicos (AFIP), la Administración Nacional de la Seguridad Social (ANSES), el Banco Central de la República Argentina (BCRA), entre otras. Además cuenta con datos propios del servicio de consultas de la empresa de donde se obtiene información de pedidos de quiebra, cheques rechazados, etc. Estas últimas variables propias de la base de datos hacen que el modelo desarrollado sea único.

3.2 Desarrollo del Modelo

3.2.1 Muestras

Para este trabajo se utilizaron tres muestras tomadas de la base de datos. La primera llamada muestra modelo con la cual se construyó el algoritmo, la segunda llamada muestra control con la que se validó el modelo de *score* y la tercera denominada muestra aplicación con la cual se muestra el uso práctico del modelo.

3.2.3 Variables Independientes

Hay dos tipos de variables independientes en el modelo. Las variables características y las de comportamiento crediticio. Las primeras incluyen la antigüedad de la empresa y las exportaciones o importaciones en pesos. Las de comportamiento crediticio incluyen el monto que se adeuda a las entidades financieras, la situación según el BCRA, la cantidad de bancos con los que la empresa mantiene una deuda, la cantidad de meses desde que opera con bancos, cantidad de cheques rechazados, los pedidos de quiebra y los concursos y quiebras.

3.2.3 Metodología

Los modelos de respuesta binaria se utilizan para estimar la probabilidad de que ocurra un suceso en función de un conjunto de variables explicativas que se consideran causales del mismo.

La tabla de la muestra modelo comprende 50.000 empresas de las cuales están registrados los valores de las variables independientes X y la variable dependiente Y. El default es la variable dependiente. Dicha variable toma en cada individuo el valor 0 o 1, según haya presentado o no el atributo cuya probabilidad se desea estimar en otros individuos. La base tiene la siguiente estructura:

Y	X_1	X_2	...	X_K
0	X_{11}	X_{21}	...	X_{K1}
1	X_{12}	X_{22}	...	X_{K2}
...	...	...	...	...
1	X_{1n}	X_{2n}	...	X_{Kn}

Siendo X_{it} el valor que toma la variable independiente X sub i para la empresa t.

Una vez definidas las variables se justifica la metodología a aplicar.

Si se hubiera utilizado un modelo lineal como

$$(2) \quad Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \delta_t$$

se obtendrían resultados absurdos, como por ejemplo valores de Y fuera del intervalo (0,1), o sea probabilidades negativas o mayores que la unidad.

Es por eso que para resolver el problema se propuso el modelo Logit o de Regresión Logística.

En este modelo, la probabilidad P_t de que el individuo t presente el atributo está dada por la

expresión (3), esto es:

$$(3) \qquad P_t = \frac{1}{1+e^{-w_t}} = \frac{1}{1+Exp\left[-(\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + ...)\right]}$$

Para estimar los parámetros del modelo se debe minimizar la expresión:

$$(4) \qquad Q = \sum_{t=1}^{n} (Y_t - P_t)^2 = \sum_{t=1}^{n} \left\{ Y_t - \frac{1}{1+Exp\left[-(\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + ...)\right]} \right\}^2$$

Esta expresión mide el error de predicción, ya que toma el valor absoluto de la diferencia entre el valor real de la variable dependiente y el valor estimado. En este caso el valor real es el default observado (Y_t) y el valor estimado es el default pronosticado (P_t). Es por eso que al minimizar la expresión se obtiene un modelo que estima la probabilidad de default.

Utilizando una planilla de cálculo se analiza la correlación entre las variables independientes y la dependiente para determinar cuales se deben incluir en el modelo. Además de utilizar las variables con el valor con el que están en la base de datos, también se calcula el logaritmo de las mismas o se las expresa en función binaria. Por ejemplo, para el "Monto adeudado en bancos", se calcula el logaritmo en base diez de ésta y por ejemplo para la "Cantidad de pedidos de quiebra" se expresa de forma tal que si la empresa tuvo un pedido de quiebra o más se le da valor 1 mientras que si no tuvo ninguno se le da valor 0. De esta manera queda expresada en forma binaria. Esto se hace porque para algunas variables la correlación con el default se incrementa si se la transforma.

Hecho esto la base de datos queda definida y se importa al software SPSS donde se procesa para obtener un algoritmo que minimice la expresión (3).

Al correr el modelo de regresión logística se obtuvo la significatividad y los coeficientes β de las variables independientes del modelo. Con esta información se evalúa nuevamente el nivel de significatividad de cada variable.

Tabla 2: Resultados del SPSS			
Número	Variable Independiente	Beta	Significatividad
1	CI_Inicio_Sit0	-1,71E+00	4,97E-193
2	CI_Inicio_Sit1	-2,71E+00	7,85E-131
3	AntiguedadCI_I	-4,20E-03	1,20E-113
4	SRT_Estado_I	0,543	1,06E-112
5	CI_Inicio_Sit5	3,36	2,01E-111
6	L_HC_NoPagados_Cant_I_1	2,754	1,05E-45
7	L_Edad	-3,67E-01	4,50E-17
8	CI_Inicio_Sit4	1,427	2,59E-11
9	L_PQ_I_1	2,026	5,77E-11
10	CI_Inicio_Sit6	3,335	2,43E-08
11	CI_Inicio_Sit2	-8,75E-01	3,49E-08
12	TotImpor	-7,59E-08	0
13	L_CI_Inicio_CantBancos_1	1,067	0
14	CI_Inicio_Sit3	0,469	0,002
15	TotExpor	1,70E-08	0,041
16	L_HC_NoPagoMulta_Cant_I_1	-3,41E-01	0,163
17	L_CI_Inicio_SumMonto_1	0,017	0,522

La descripción de cada variable independiente se encuentra en el Anexo.

Si el nivel de significatividad (columna 4 de la tabla) es menor que 0,05 entonces el parámetro es útil para el modelo. Por lo tanto quedan excluidas del modelo las variables 16 y 17 que son: el logaritmo de la cantidad de multas impagas por cheques rechazados y el logaritmo de la suma adeudada.

A continuación se presenta un análisis de las variables y sus ponderaciones.

A mayor antigüedad de la empresa y a mayor antigüedad en el sistema bancario menor es la probabilidad de default de la empresa. Esto parece lógico ya que para que una empresa alcance cierta antigüedad ésta debe ser financieramente estable de lo contrario quedaría fuera de actividad en el corto plazo.

Las empresas con situación 1 en el BCRA (es decir con atrasos en los pagos de sus deudas dentro de 30 días) tienen menor probabilidad de caer en default. Estar en situación 2 también tiene una relación inversa con el default pero de menor magnitud (-0,88 versus -2,71 de la situación 1). Por el otro lado, las situaciones de 3 a 6 tienen un beta positivo. Se puede concluir que a mayor situación en el BCRA, es decir, a mayor atraso en el pago de las deudas, mayor es la probabilidad de que la empresa entre en default.

A mayor cantidad de bancos acreedores mayor es la probabilidad de default. Las empresas con cheques rechazados, pedidos de quiebra y concursos y quiebra tienen una alta probabilidad de default.

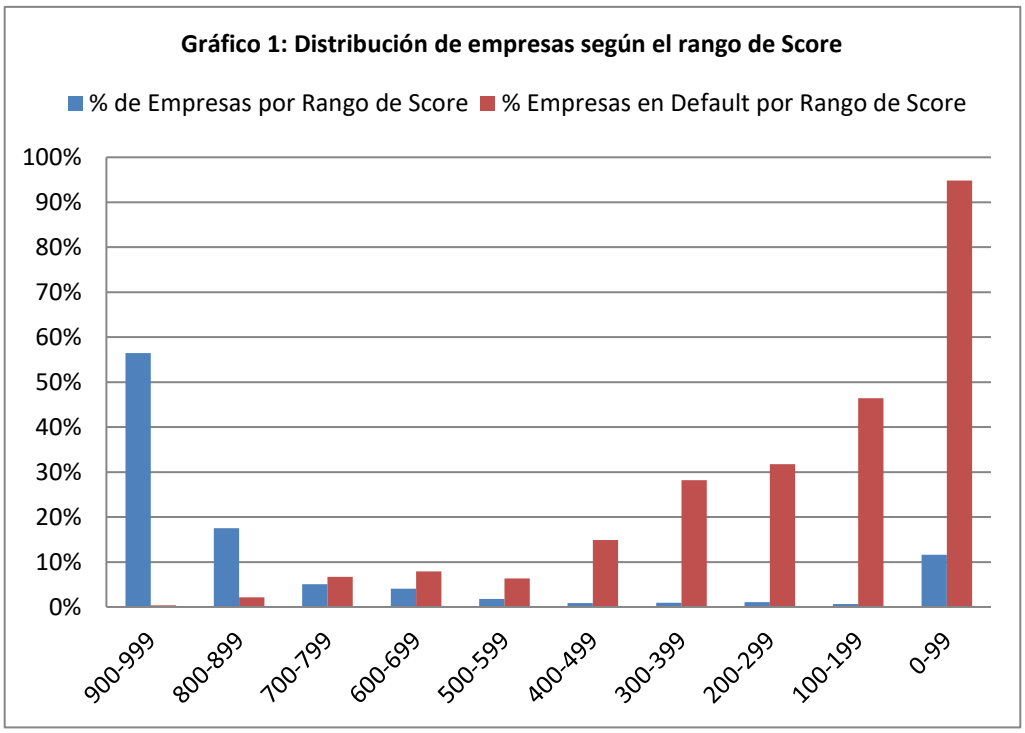
El hecho de tener un atraso en los pagos a la Superintendencia de Riesgo de Trabajo implica una mayor probabilidad de default.

Luego, con el algoritmo previamente elaborado se calcularon las probabilidades de default de

las empresas en la muestra control. Para esto se regresó el modelo obtenido de la muestra modelo sobre los datos de las empresas de la muestra de control. A continuación se calculó el *score* de cada empresa.

El *score* se deduce de la probabilidad de default estimada y es igual a $(1-P_t)*1000$. Su valor puede tomar valores de 0 a 1000, aunque el rango que se utiliza es 1 a 999 para permitir un margen de error.

A continuación se puede ver como quedaron distribuidas las 50.000 empresas de la muestra control por rango de *score* y cómo quedan distribuidas las empresas en default.



En el gráfico 1 se puede ver como el porcentaje de empresas en default aumenta a medida que baja el *score* a tal punto que de las empresas en el rango de *score* más bajo (0-99), el 94,8% cae en default. Esto indica que el modelo está otorgando malas calificaciones a las empresas incumplidoras y buenas calificaciones a las cumplidoras.

En cuanto a la distribución se pude ver que hay una acumulación del 56% de las empresas en el rango superior de *score* (900-999) y otra acumulación un poco menor en el rango inferior de *score* (0-99). Esto es evidencia de cómo el modelo extrapola a las empresas discriminando incumplidoras por un lado y cumplidoras por el otro.

4. Validación del modelo

La validación de los modelos de *scoring* es de suma importancia para las entidades financieras. El Comité de Supervisión Bancaria de Basilea. [2004, Art. 500] establece que: "Los bancos deberán contar con un buen sistema para validar la precisión y coherencia de los sistemas de calificación, los procesos y la estimación de todos los componentes de riesgo pertinentes. Asimismo, deberán demostrar a sus supervisores que su proceso de validación interna les permite evaluar, de forma consistente y significativa, el funcionamiento de los sistemas de calificación interna y de estimación de riesgos".

Con el fin de validar el modelo estimado en la sección anterior, a continuación se calculan indicadores de bondad de ajuste *in-sample* (sobre la muestra utilizada para estimar el modelo, es decir sobre la muestra modelo). Además se testea el modelo sobre una segunda muestra, la muestra control, y se analizan sus resultados *out-of-sample*.

4.1 Validación *in-sample*: Indicadores

Para los modelos de regresión logística existen varias medidas de bondad de ajuste:

Indicador	Valor
$R^2_{clásico}$	0,6459
$R^2_{McFadden}$	0,7738
$R^2_{Cox\text{-}Snell}$	0,6579
$R^2_{Nagelkerke}$	0,8772

Los coeficientes de determinación son bastante significativos indicando que una alta proporción de la variación de la variable dependiente (default) puede ser explicada por las variables independientes incluidas en el modelo. En el anexo se detallan las fórmulas para calcular estos coeficientes.

4.2 Validación out-of-sample

Se calculó la probabilidad de default para la muestra control para unas 50.000 empresas diferentes a las que se usaron para el desarrollo del modelo. Se determinó que aquellas empresas que tuvieran una probabilidad de default P_t mayor o igual a 0,5 el default pronosticado es 1 mientras que aquellas cuya probabilidad de default sea menor a 0,5 el default pronosticado es igual a 0.

Supongamos que las 50.000 empresas de la muestra control son empresas que solicitaron crédito a un banco y que éste les aplicó el modelo de *score* para aceptar o rechazar dichas solicitudes. De esta manera podemos calcular los aciertos y errores del modelo.

Default Observado	Default Pronosticado		Porcentaje Correcto
	0	1	
0	41.765	1.529	96,47%
1	683	6.023	89,82%
Porcentaje Global			95,58%

Fuente: SPSS

Como se puede ver en la tabla, de las empresas que no cayeron en default, es decir con default observado igual a 0, el modelo acertó 41.765 casos y se erró 1.529 casos resultando en un porcentaje correcto de 96,47%.

Mientras que de las empresas que cayeron en default (default observado igual a 1) el modelo acertó 6.023 casos y desacertó en 683 obteniendo un porcentaje correcto de 89,82%.

Tomando en cuenta todos los individuos el porcentaje global de aciertos fue 95,58%.

En este caso se aceptaron solicitudes de crédito de empresas cuyo *score* era mayor o igual a 500 pero que pasaría si se usara otro corte de *score*. A continuación se detallan los aciertos y errores para diferentes cortes.

A modo de referencia se establece:

(0-0) Buenos pagadores pronosticados correctamente

(0-1) Buenos pagadores pronosticados erróneamente

(1-1) Malos pagadores pronosticados correctamente

(1-0) Malos pagadores pronosticados erróneamente

Tabla 3: Muestra control ordenada por corte de probabilidad												
1	2	3	4	5	6	7	8	9	10	11	12	13
Rango de *Score*	Empresas	Empresas aprobados acumulados	Malos por Rango	Malos acum	Buenos acum	Aciertos s/ buenos (0-0)	Errores s/ buenos (0-1)	% de Aciertos en la aprobación de buenos	Aciertos s/ malos (1-1)	Errores s/ malos (1-0)	% de Aciertos en el rechazo de malos	Porcentaje correcto s/ el total de la muestra
900-999	28.242	28.242	109	109	28.133	28.133	15.161	64,98%	6.597	109	98,37%	69,46%
800-900	8.766	37.008	189	298	36.710	36.710	6.584	84,79%	6.408	298	95,56%	86,24%
700-800	2.514	39.522	168	466	39.056	39.056	4.238	90,21%	6.240	466	93,05%	90,59%
600-700	2.016	41.538	159	625	40.913	40.913	2.381	94,50%	6.081	625	90,68%	93,99%
500-600	910	42.448	58	683	41.765	41.765	1.529	96,47%	6.023	683	89,82%	95,58%
400-500	429	42.877	64	747	42.130	42.130	1.164	97,31%	5.959	747	88,86%	96.18%
300-400	475	43.352	134	881	42.471	42.471	823	98,10%	5.825	881	86,86%	96.59%
200-300	523	43.875	166	1.047	42.828	42.828	466	98,92%	5.659	1.047	84,39%	96,97%
100-200	308	44.183	143	1.190	42.993	42.993	301	99,30%	5.516	1.190	82,25%	97,02%
1-100	5.817	50.000	5.516	6.706	43.294	43.294	-	100,00%	-	6.706	0,00%	86,59%
	50.000		6.706									

El corte resaltado en la tabla es el caso descripto anteriormente donde se hizo el corte de *score* en 500.

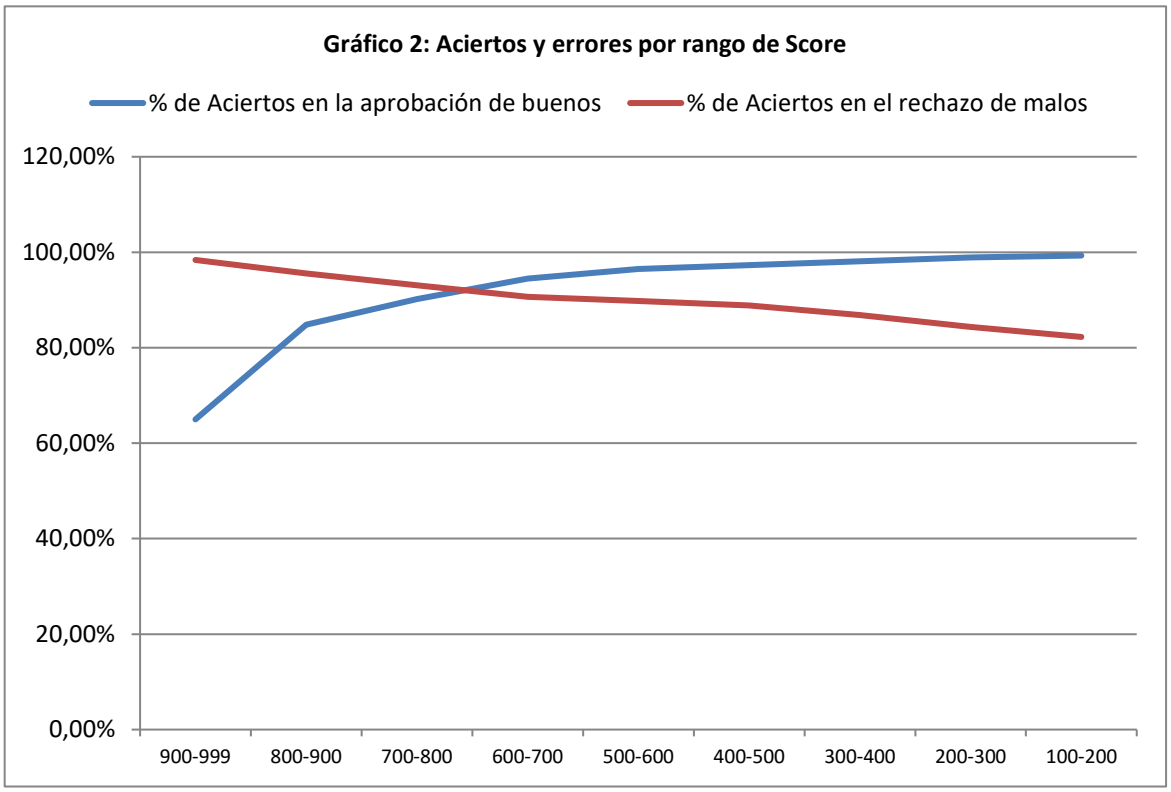
Si se aprueba la solicitud de crédito a las empresas con *score* de 900 a 999 se estarían aprobando 28.242 empresas dentro de las cuales 109 resultarían incumplidores y 28.133 buenos pagadores. Al tomar este corte se estaría cometiendo dos tipos de error:

Error tipo I: Denegar un crédito a un buen pagador.

Error tipo II: Otorgar un crédito a un mal pagador.

Al rechazar de 900 de *score* para abajo se les está negando crédito a 15.161 buenos pagadores de los 43.294 buenos pagadores que hay en la muestra control. Es decir que se estaría acertando en la aceptación de buenos pagadores en un 64,98% y desacertando en la negación de crédito al 35,02% de los buenos pagadores (Error tipo I).

Por el otro lado se les está ofreciendo crédito a 109 malos pagadores de los 6.706 malos pagadores que hay en la muestra control. Es decir que se estaría acertando en el rechazo de malos pagadores en un 98,37% y aceptando el 1,63% de malos pagadores (Error tipo II).



Del gráfico se aprecia que a medida que se exige un *score* menor disminuye el porcentaje de aciertos en el rechazo de malos pagadores y aumenta el acierto en la aceptación de buenos. Entonces surge la pregunta ¿en dónde debe hacerse el corte de *score* a partir del cual se aceptan solicitudes de crédito? Para responder a esta pregunta se calcula el porcentaje de incobrabilidad generado por los malos pagadores de cada rango de *score*. De esta manera se podrá determinar el porcentaje de incobrabilidad marginal que aporta cada rango de *score* (columna 5 de la siguiente tabla).

Tabla 4: Porcentaje de Default por Rango de *Score*

1	2	3	4	5
Rango de *Score*	Individuos	Malos pagadores por Rango	Buenos pagadores por Rango	% de Default por Rango
900-999	28.242	109	28.133	0,39%
800-900	8.766	189	8.577	2,16%
700-800	2.514	168	2.346	6,68%
600-700	2.016	159	1.857	7,89%
500-600	910	58	852	6,37%
400-500	429	64	365	14,92%
300-400	475	134	341	28,21%
200-300	523	166	357	31,74%
100-200	308	143	165	46,43%
1-100	5.817	5.516	301	94,83%
	50.000	6.706		

A medida que se les exige un *score* menor a las empresas aumenta el porcentaje de defaults. De las empresas con un rango de *score* entre 900 y 999 hay un 0.39% de empresas que

defaultean mientras que del rango de *score* entre 1 y 100 el 94,83% de las empresas cae en default. El banco debe determinar el porcentaje de incobrabilidad que está dispuesto a aceptar. Es decir que si un banco puede hacer frente a un nivel de incobrabilidad del 5% debe aprobar a partir de 800 de *score* para arriba. En cuyo caso estaria otorgando crédito a 37.008 empresas que representa el 74% de las 50.000 en cuestión. Si aceptara a las empresas que están en el rango de *score* de 700 a 800, este grupo le estaría aportando un default de 6,68%, el cual supera el nivel de incobrabilidad que soporta el banco.

Más adelante se plantea una aplicación del modelo más abarcativa.

4.3 Poder Discriminatorio

El poder discriminatorio de un modelo se refiere a la capacidad de clasificar y ordenar a los individuos (las empresas) en función del riesgo crediticio de manera tal que los mejores pagadores tengan los *scores* más altos y los incumplidores tengan los *scores* más bajos. Para determinar el poder discriminatorio del modelo se utiliza el test de Kolmogorov-Smirnov que compara las distribuciones de los buenos y malos pagadores.

4.3.1 Test de Kolmogorov-Smirnov

Este estadístico toma la mayor diferencia en términos absolutos entre dos distribuciones: la distribución acumulada de los buenos pagadores y la distribución acumulada de los malos pagadores. En base al estadístico de Kolmogorov-Smirnov (KS) se realiza un test de hipótesis para determinar cuan significativa es la dferencia entre ambas distribuciones. A mayor KS, mayor es la distancia entre las distribuciones y en consecuencia mayor es el poder discriminatorio del modelo.

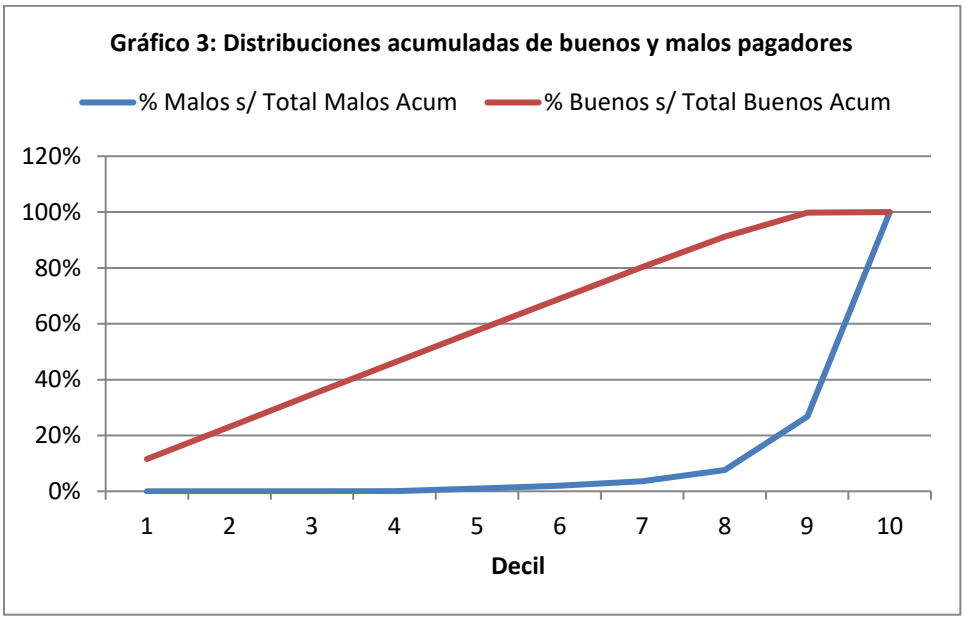
En la siguiente tabla se ordenaron las empresas de la muestra control de mayor a menor *score*. Luego se las agrupa en deciles, es decir, el 10% de las empresas que tuvieron mayor *score* se encuentra en el primer decil, el siguiente 10% en el segundo decil y asi sucesivamente hasta ordenar las 50.000 empresas.

Tabla 5: Muestra control ordenada por deciles											
1	2	3	4	5	6	7	8	9	10	11	12
Deciles	*Score* Min	*Score* Max	Cant. Indiv	% de Indiv	Malos	Buenos	% Malos s/ Total Malos	% Buenos s/ Total Buenos	% Malos s/ Total Malos Acum	% Buenos s/ Total Buenos Acum	Diferencia
1	999	999	5.000	10%	-	5.000	0,00%	11,55%	0,00%	11,55%	11,55%
2	999	999	5.000	10%	-	5.000	0,00%	11,55%	0,00%	23,10%	23,10%
3	998	999	5.000	10%	-	5.000	0,00%	11,55%	0,00%	34,65%	34,65%
4	947	998	5.000	10%	3	4.997	0,04%	11,54%	0,04%	46,19%	46,14%
5	917	947	5.000	10%	62	4.938	0,92%	11,41%	0,97%	57,59%	56,63%
6	888	917	5.000	10%	66	4.934	0,98%	11,40%	1,95%	68,99%	67,04%
7	834	888	5.000	10%	108	4.892	1,61%	11,30%	3,56%	80,29%	76,73%
8	692	834	5.000	10%	268	4.732	4,00%	10,93%	7,56%	91,22%	83,66%
9	15	692	5.000	10%	1.286	3.714	19,18%	8,58%	26,74%	99,80%	73,06%
10	1	15	5.000	10%	4.913	87	73,26%	0,20%	100,00%	100,00%	0,00%
Total	1	998	50.000	100%	6.706	43.294	100,00%	100,00%			

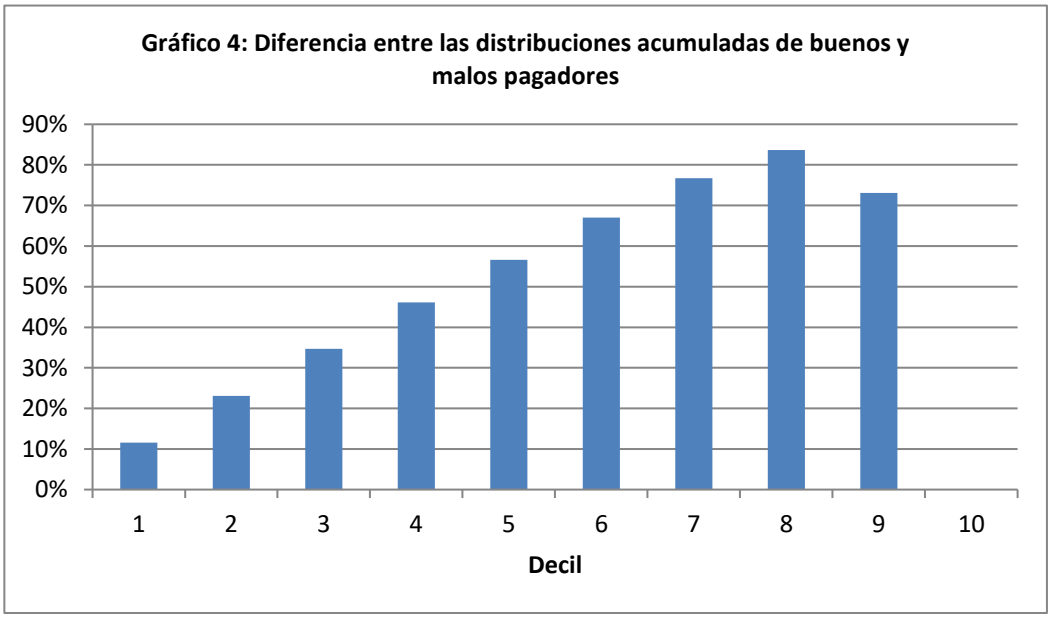
"% de Malos sobre total de malos acumulado" (columna 10 de la Tabla 4) es la frecuencia acumulada de malos pagadores a medida que se suma el "% de Malos sobre el Total de Malos" (columna 8) de cada decil. Analógamente se hace este calculo para los buenos pagadores en la columna 11. En la columna 12 se calcula la diferencia entre estas frecuencias acumuladas y para determinar el KS se toma la máxima de estas diferencias que puede tomar valores entre 0% y 100%. En este caso el KS es 83,66% lo que deja en evidencia que la

diferencia entre las distribuciones de buenos y malos pagadores es significativa.

Los siguientes gráficos ilustran la diferencia entre las frecuencias acumuladas.



La distancia entre las curvas muestra el poder discriminatorio de malos y buenos pagadores del modelo.



Para evaluar qué tan significativa es la diferencia entre ambas distribuciones se calcula el estadístico Z de Kolmogorov-Smirnov en función de la distancia máxima.

$$(5) \qquad Z_{K\text{-}S} = \text{máx}_i \, |D_i| \, \sqrt{((n_1 * n_2)/(n_1+n_2))}$$

Donde $\text{máx}_i \, |D_i|$ es la distancia máxima entre la distribución acumulada de buenos y malos pagadores; n_1 es la cantidad de buenos y n_2 es la cantidad de malos.

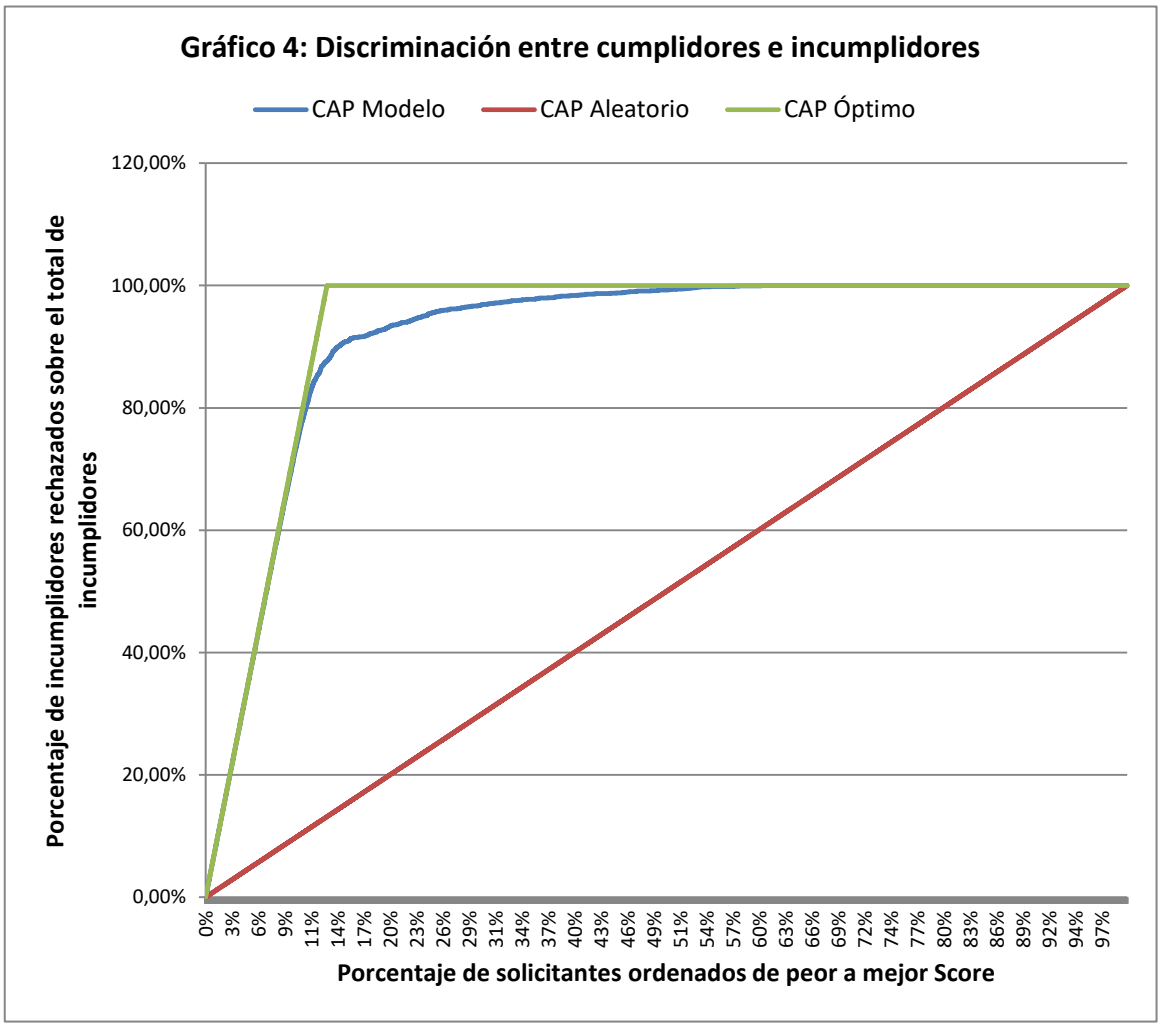
Tabla 6: Test K-S	
Diferencia más extrema ($\text{máx}_i\lvert D_i\rvert$)	0,8366
Z de Kolmogorov-Smirnov	63,75
Probabilidad de Z	0,000

Fuente: SPSS

La probabilidad del estadístico Z de Kolmogorov-Smirnov es menor que 0,005 por lo que las distribuciones acumuladas de buenos y malos pagadores resultan significativamente diferentes. Esto es prueba de que el algoritmo posee un alto poder discriminatorio entre pagadores e incumplidores. Más aún, se puede concluir que el modelo de *scoring* de riesgo crediticio tiene un alto poder de predicción.

4.3.2 Curva CAP (Cumulative Accuracy Profile) y Ratio de Precisión

Otro método para validar el modelo es graficar la Curva CAP y calcular el ratio de precisión.



El eje horizontal representa a los solicitantes de crédito ordenados por *score* en sentido ascendente, es decir de menor a mayor *score*. El eje vertical representa el porcentaje de incumplidores rechazados sobre el total de incumplidores.

Un buen modelo de riesgo crediticio concentra a los incumplidores en el rango de menor *score*. Por lo tanto el porcentaje de incumplidores detectados sobre el eje vertical aumenta rápidamente a medida que se pasa de menor a mayor *score* de las empresas en la muestra (eje horizontal).

La curva CAP óptimo detecta inmediatamente al 100% de los incumplidores que existen en la muestra modelo. Es decir que acumula a todos los incumplidores en el rango de menor *score*.

La curva CAP aleatorio representa de qué manera se rechazan las solicitudes de incumplidores. La curva resulta ser una linea recta de pendiente uno ya que no distingue entre cumplidores e incumplidores por lo tanto no hay una rápida acumulación de los malos pagadores sino que los detecta en la proporción en que existen incumplidores sobre el total de la muestra. Por ejemplo, al recorrer el 10% de la muestra control se detectó el 10% de los incumplidores mientras que para el CAP óptimo al recorrer el 10% de la muestra ya se

acumularon el 75,89% de los incumplidores.

Para determinar que tan efectiva es la curva CAP del modelo se debe comparar el área comprendida por la curva CAP modelo y la CAP aleatoria sobre el area comprendida por la curva CAP óptimo y la curva CAP aleatoria.

La fórmula descripta es el ratio de precisión:

$$(6)\ \text{Ratio de Precisión} = \frac{\text{CAP del modelo estimado} - \text{CAP del modelo aleatorio}}{\text{CAP del modelo óptimo} - \text{CAP del modelo aleatorio}} = 95{,}8\%$$

Dado que el valor del ratio de precisión alcanza el 95,8% del nivel óptimo se establece que el modelo tiene un alto poder de discriminación.

5. Aplicación práctica del modelo

En este capítulo se muestra cómo aplicar el modelo de *scoring*. Para esto se va a suponer que un banco tiene una base de datos de empresas a las cuales les ofrecerá crédito, y debe decidir a cuáles le hará el ofrecimiento y cuáles dejará de lado. El objetivo es determinar el corte de *score* que separe las aceptadas de las rechazadas tal que se maximice el resultado.

La fijación del corte de *score* no depende únicamente del riesgo de incobrabilidad. También depende de la tasa activa y la tasa pasiva, los costos fijos y variables y el monto de los préstamos otorgados por la entidad financiera. Con la inclusión de estos conceptos, en este capítulo se plantea una solución a un problema más complejo que el anterior donde sólo se tenía en cuenta la incobrabilidad que podía costear el banco. Además el objetivo es romper con el concepto de que el corte de *score* es único para todas las entidades financieras y demostrar que éste depende de las variables particulares de cada entidad.

5.1 Determinación del corte de *score* en función del riesgo de incobrabilidad junto con otras variables

El objetivo del banco es maximizar los beneficios derivados de la intermediación financiera. Cumplir este objetivo no implica solamente otorgar crédito a la franja de empresas con mayor *score*. Un conjunto de solicitantes de crédito con un alto riesgo de incobrabilidad puede ser aceptable si se ajusta con una alta tasa de interés que compense el nivel de riesgo además de compensar el costo de la tasa pasiva, los costos fijos y los costos variables. Es por esto que el cálculo del corte de aprobación es complejo y no se basa únicamente en el riesgo de incobrabilidad.

Con el fin de ejemplificar cómo se determina el corte que maximiza el resultado se utiliza un esquema simplificado de ingresos y egresos. En este supuesto la utilidad a maximizar responde a la siguiente ecuación:

$$(7)\quad \text{Utilidad} = \text{IF} - \text{CFin} - \text{Cfijo} - \text{GA} - \text{PInc}$$

Donde:

IF: Ingresos financieros = Monto total de préstamos otorgados x Tasa activa

CFin: Costos Financieros = Monto total de préstamos otorgados x Tasa pasiva

CFijo: Costo Fijo = una constante

GA: Gastos Administrativos = Cantidad de préstamos otorgados x Gastos administrativos por crédito

PInc: Pérdida por Incobrabilidad

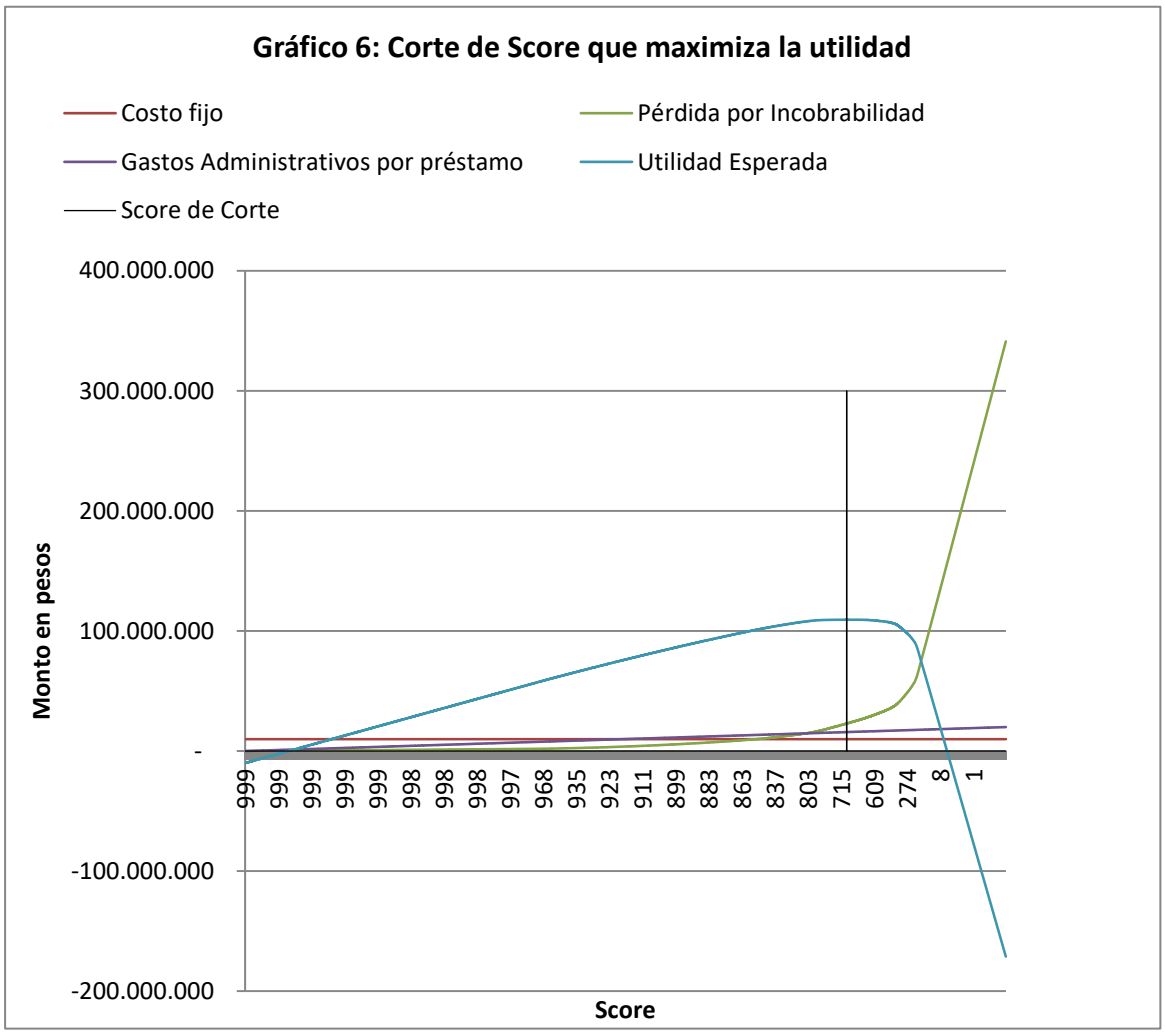
$$= \sum_{i=1}^{n} (\text{Monto del préstamo a la empresa}_{\,i} \text{ x Probabilidad de Default de la empresa}_{\,i})$$

A continuación se plantea una campaña de marketing en la que se ofrecen créditos por un valor de 100.000 pesos cada uno a una tasa activa de 23% con un costo fijo de 10.000.000 de pesos y gastos administrativos por préstamo de 1.000 pesos. El costo del dinero para la entidad financiera es de 13% (tasa pasiva). Estos supuestos se detallan en la tabla. Con la

campaña se supone que hay 20.000 empresas en la base de prospectos, los cuales deben evaluarse.

Tabla 7: Supuestos	
Monto de cada préstamo	100.000
Costo Fijo	10.000.000
Gastos Administrativos por préstamo	1.000
Tasa Pasiva	13%
Tasa Activa	23%

Seguido de esto se grafica el corte de *score* que maximiza la utilidad para estos supuestos.



En el eje horizontal se encuentran ordenadas las empresas de mayor a menor *score*. Posteriormente se calculan los costos e ingresos correspondientes a la cantidad de empresas aprobadas hasta cada nivel de *score*. Luego se calcula la utilidad para cada corte de *score* y se determina el corte que maximiza la utilidad. El *score* de corte implica que aquellas empresas con un *score* superior a éste verán su solicitud de crédito aprobada y aquellas con un *score* inferior verán denegado el crédito o pendiente de un análisis posterior. En el gráfico se puede apreciar que el corte que optimiza la utilidad es 698.

En la siguiente tabla se detallan los valores que toman las variables para este nivel de *score*.

Tabla 8: Calculo del Resultado para el corte óptimo	
Score de Corte	**698**
Empresas Aprobadas	15.816
Empresas Rechazadas	4.184
Total de empresas analizadas	20.000
Capital Prestado	1.581.600.000
Ingreso financiero	363.768.000
Costo financiero	205.608.000
Costo Fijo	10.000.000
Gastos Administrativos	15.816.000
Pérdida por Incobrabilidad	22.933.518
Utilidad Máxima	109.410.482

Dados los supuestos descriptos anteriormente la máxima utilidad que se puede obtener es $ 109.410.482 aceptando solicitudes de crédito a partir de 698 puntos de *score*. Esto es aceptando 15.816 empresas y rechazando 4.184.

A continuación se muestra como varía el corte y por lo tanto la cantidad de solicitudes aceptadas cuando se modifican diferentes variables. En la siguiente tabla se puede apreciar como fluctúa el corte que maximiza el resultado para diferentes montos por préstamo.

Tabla 9: Fluctuación del *score* de corte al variar el monto de los préstamos					
Costo Fijo	10.000.000				
Gastos Adm por préstamo	1.000				
Tasa Pasiva	13%				
Tasa Activa	23%				
	Monto de cada préstamo	*Score* de corte	Empresas Aceptadas	Empresas Rechazadas	Utilidad Máxima
	200.000	664	16.177	3.823	244.888.824
	175.000	669	16.147	3.853	211.007.087
	150.000	675	16.122	3.878	177.129.916
	125.000	684	16.089	3.911	143.257.409
	100.000	698	15.816	4.184	109.410.482
	75.000	720	15.578	4.422	75.625.188
	50.000	754	15.203	4.797	41.971.196
	25.000	790	14.942	5.058	8.454.764

En la siguiente tabla se muestra cómo varía la tasa activa y consecuentemente el *score* de corte:

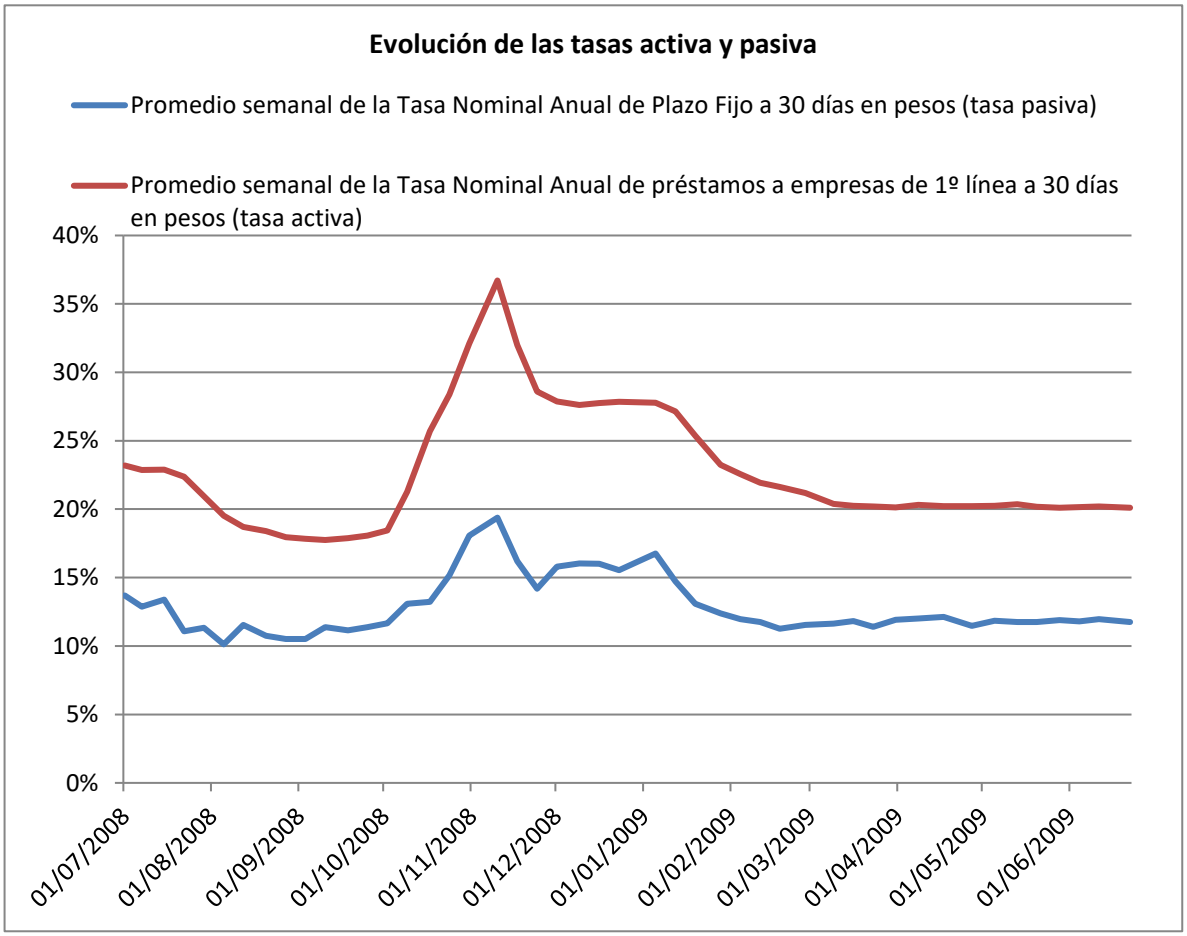
Tabla 10: Fluctuación del *score* de corte al variar la tasa activa					
Monto de cada préstamo	100.000				
Costo Fijo	10.000.000				
Gastos Adm por préstamo	1.000				
Tasa Pasiva	13%				
	Tasa activa	*Score* de corte	Empresas Aceptadas	Empresas Rechazadas	Utilidad Máxima
	31%	506	16.941	3.059	240.459.019
	29%	544	16.894	3.106	207.186.395
	27%	584	16.766	3.234	174.063.577
	25%	625	16.381	3.619	141.436.785
	23%	698	15.816	4.184	109.410.482
	21%	770	15.085	4.915	79.087.505
	19%	805	14.778	5.222	49.469.760
	17%	841	13.776	6.224	20.923.428
	15%	923	9.625	10.375	-3.514.429

Es así como esta herramienta permite determinar el corte que maximiza la utilidad en función de parámetros propios de cada entidad financiera.

5.2 Simulación de la utilidad en función de las tasas pasiva y activa

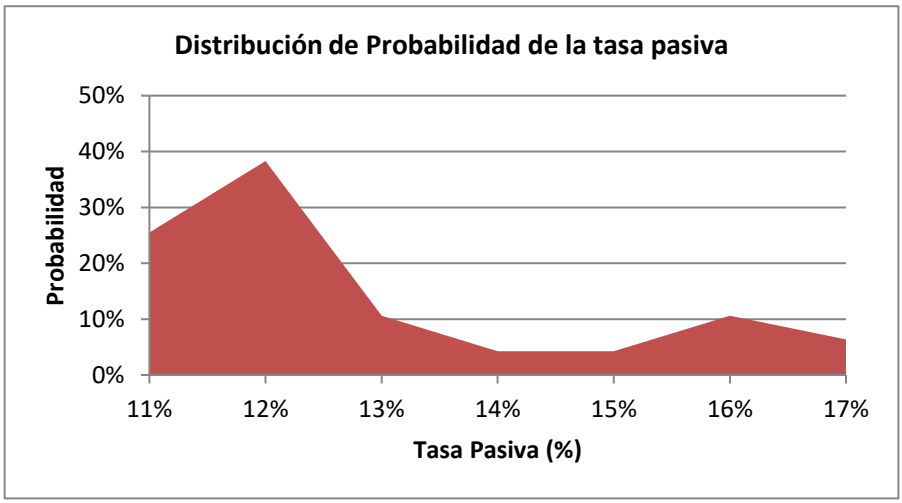
La utilidad que tendrá la entidad bancaria dependerá, entre otras variables, de la tasa pasiva y de la tasa activa. A continuación se estudiará cómo varía la utilidad en función de la tasa activa y del spread, que es la diferencia entre la tasa activa y la pasiva. Para simular los diferentes escenarios posibles, para cada variable se deben definir los valores que pueden tomar y su distribución de probabilidad.

Para la distribución de probabilidad de la tasa pasiva se utilizaron los datos de la Tasa Nominal Anual de Plazo Fijo a 30 días en pesos. A su vez, para la tasa activa se utilizaron los datos de la Tasa Nominal Anual de préstamos a empresas de 1º línea a 30 días en pesos. Se tomaron ambas series de datos de Julio de 2008 a Junio de 2009. La evolución de estas variables puede verse en el siguiente gráfico.



Fuente: BCRA

A partir de estos datos, se grafica la distribución de probabilidad de la tasa pasiva.



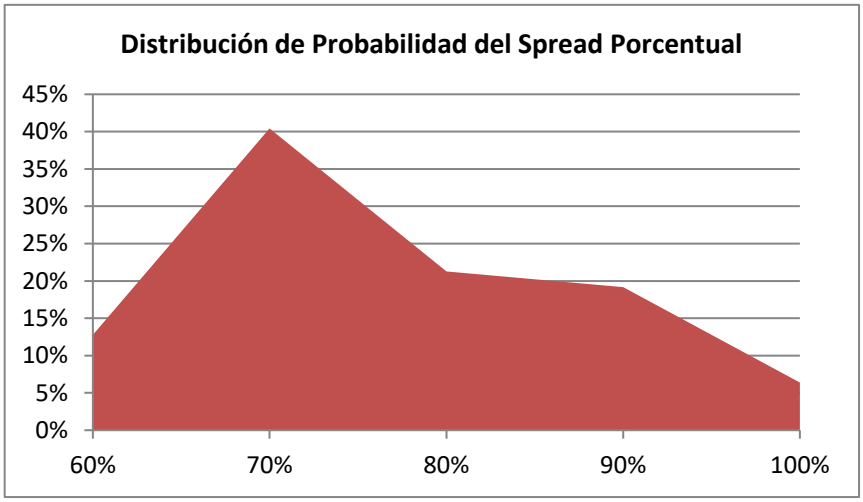
Estos son los datos de entrada que se utilizarán para simular la tasa pasiva que deberá pagar el banco para financiarse durante los próximos doce meses.

Para determinar el spread que cobrará el banco por su intermediación se definió la siguiente variable:

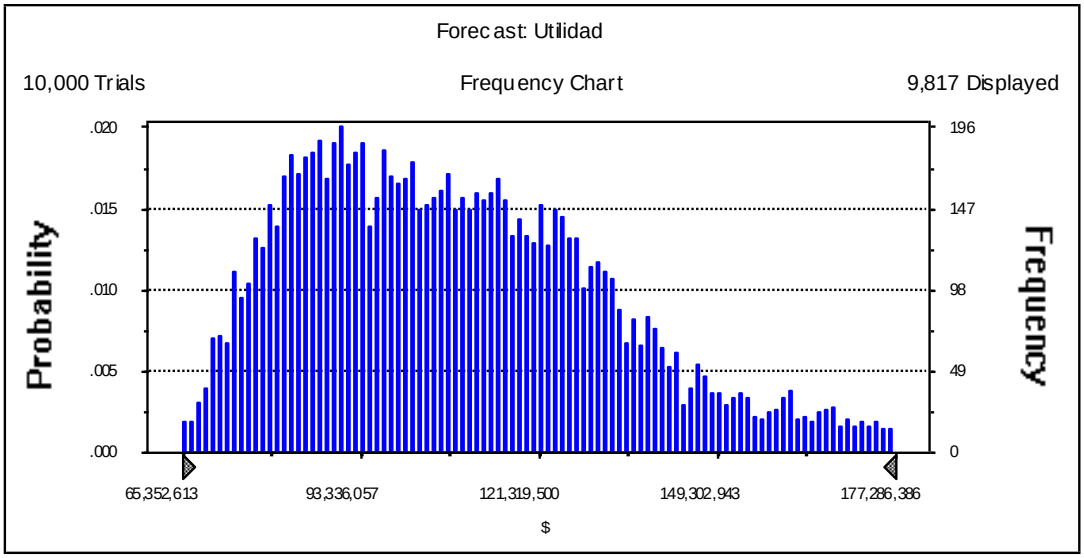
$$(8)\ \text{Spread Porcentual} = \frac{\text{Tasa Activa} - \text{Tasa Pasiva}}{\text{Tasa Pasiva}}$$

Tabla 11: Tasas y Spread Porcentual			
Semana	Promedio semanal de la Tasa Pasiva	Promedio semanal de la Tasa Activa	Spread Porcentual
01/07/2008	14%	23%	70%
07/07/2008	13%	23%	78%
15/07/2008	13%	23%	71%
22/07/2008	11%	22%	102%
29/07/2008	11%	21%	85%
05/08/2008	10%	20%	93%
12/08/2008	12%	19%	62%
20/08/2008	11%	18%	71%
27/08/2008	11%	18%	71%
03/09/2008	11%	18%	70%
10/09/2008	11%	18%	56%
18/09/2008	11%	18%	60%
25/09/2008	11%	18%	59%
02/10/2008	12%	18%	58%
09/10/2008	13%	21%	63%
17/10/2008	13%	26%	94%
24/10/2008	15%	28%	87%
31/10/2008	18%	32%	78%
10/11/2008	19%	37%	89%
17/11/2008	16%	32%	97%
24/11/2008	14%	29%	102%
01/12/2008	16%	28%	77%
09/12/2008	16%	28%	72%
16/12/2008	16%	28%	73%
23/12/2008	16%	28%	79%
05/01/2009	17%	28%	66%
12/01/2009	15%	27%	84%
19/01/2009	13%	25%	94%
28/01/2009	12%	23%	88%
04/02/2009	12%	23%	89%
11/02/2009	12%	22%	87%
18/02/2009	11%	22%	92%
27/02/2009	12%	21%	84%
09/03/2009	12%	20%	75%
16/03/2009	12%	20%	71%
23/03/2009	11%	20%	77%
31/03/2009	12%	20%	69%
08/04/2009	12%	20%	69%
17/04/2009	12%	20%	67%
27/04/2009	11%	20%	76%
05/05/2009	12%	20%	71%
13/05/2009	12%	20%	73%
20/05/2009	12%	20%	72%
28/05/2009	12%	20%	69%
04/06/2009	12%	20%	71%
11/06/2009	12%	20%	69%
22/06/2009	12%	20%	71%

Tomando los valores del spread porcentual, la distribución de probabilidad de esta variable durante el período 2008-07 a 2009-06 es:



Haciendo el supuesto de que la tasa pasiva y el spread porcentual mantendrán la mismas distribuciones de probabilidad para el próximo período, se grafica la utilidad esperada. Para esto se utiliza el software de simulación *Crystal Ball*.



En el gráfico se pueden ver los valores que toma la utilidad así como la probabilidad con que toma cada valor teniendo en cuenta la fluctuación de la tasa pasiva y el spread. La utilidad fluctúa entre $ 62.629.693 y $ 204.942.441 con un valor promedio de $ 109.578.385. En cuanto a medidas de dispersión, el gráfico tiene un rango de $ 142.312.747 y un desvío estándar de $ 25.751.899.

Más allá de que tan buenos o malos sean estos valores, lo que se destaca son las facilidades que permite esta herramienta para gestionar la incertidumbre y obtener un resultado acotado de la utilidad que se espera obtener.

6. Conclusiones

En mercados cada vez más competitivos donde las tasas de interés están establecidas por la oferta y la demanda de crédito, los bancos deben hacer foco en minimizar sus costos. La incobrabilidad es un costo importante, es por eso que los bancos deben realizar una precisa gestión de riesgo crediticio. En este escrito se presentó una solución que comprende el desarrollo estadístico del modelo, su validación y su aplicación práctica.

En el desarrollo del modelo se identificaron variables independientes con una significativa correlación con la variable dependiente. Entre las más importantes se destacan la antigüedad de la empresa, la cantidad de bancos acreedores, el monto adeudado en bancos, el atraso en el pago de las deudas (representado por las situaciones del BCRA), la antigüedad con que opera en el sistema financiero y la cantidad de cheques impagos. Además, las correlaciones de éstas con el default son razonables. Por ejemplo, es lógico pensar que a mayor antigüedad de una empresa menor es la probabilidad de default y que a mayor cantidad de cheques rechazados mayor es la probabilidad de default.

Más allá de la lógica que se espera que siga el modelo que hasta cierto punto podría ser subjetiva, la manera objetiva de evaluarlo es realizar diversos métodos de validación que se fundamentan en la estadística.

Para la validación del modelo se analizaron dos tipos de indicadores: los que surgen de la muestra utilizada para desarrollar el modelo denominados *in-sample* y los que se obtienen de aplicar el modelo a una muestra control llamados *out-of-sample*. De los primeros se puede destacar que los coeficientes de correlación dan valores altos con lo cual se puede establecer que las variables independientes explican en un alto porcentaje la variación de la variable dependiente. De los segundos se destaca tanto los valores del test de Kolmogorov-Smirnov como el Ratio de Precisión de la curva CAP que resultaron muy satisfactorios.

Finalmente se realizó una aplicación práctica del modelo de *scoring* donde se determinó con precisión el corte de *score* que se debe realizar para maximizar la utilidad. Además se pudo apreciar cómo fluctúa el corte en función del monto por préstamo y en función de las tasas activa y pasiva. Con esto se le provee una herramienta que permita a las entidades financieras determinar el corte que maximiza su resultado en función de sus propias variables. Además se presentó una simulación para gestionar la incertidumbre de algunas variables para estimar la utilidad esperada.

En miras de futuras líneas de investigación se puede calcular la tasa activa a ofrecer a cada empresa en función de su probabilidad de default. Con esto se logrará un mercado financiero más justo donde cada empresa paga intereses en función de su riesgo de incobrabilidad. De esta manera se evita que los buenos pagadores cubran la pérdida por incobrabilidad de los malos pagadores.

Otra aplicación de este modelo de *scoring* es para auxiliar a los departamentos de cobranzas que se dedican a recuperar préstamos en mora contactando a los deudores con atrasos. En este caso se calcula la probabilidad de que una empresa en mora pague su deuda con la entidad financiera y en función de esto se determina la acción a seguir y la prioridad que se le asigna a cada deudor en mora, comenzando la tarea de cobranza con los de mayor probabilidad de recupero.

7. Bibliografía

- Alan Elizondo. 2003. Medición Integral del riesgo de crédito.
- Saunders, A. 1999. Credit Risk Measurement. New Approaches to Value at Risk and Other Paradigms, John Wiley & Sons Inc., New York.
- BCRA. .2008. Clasificación de deudores. www.bcra.gov.ar
- Comité de Supervisión Bancaria de Basilea. 2004. Convergencia internacional de medidas y normas de capital, Marco revisado, Banco de Pagos Internacionales. www.bis.org
- García, R. 2004. Inferencia estadística y diseño de experimentos, Eudeba, Buenos Aires
- Gross, D y Souleles, N. 2002. An empirical analysis of personal bankruptcy and delinquency, The Society for Financial Studies.
- Software SPSS 15

8. Anexos

Variables independientes

Tabla 12: Descripción de las variables	
Variable Independiente	**Descripción**
AntigüedadCI_I	Cantidad de meses desde que opera con bancos.
CI_Inicio_Sit0	Toma valor 1 si la empresa no tiene deuda en un banco, en caso contrario toma valor 0.
CI_Inicio_Sit1	Toma valor 1 si la empresa se encuentra en situación 1 en el BCRA, en caso contrario toma valor 0.
CI_Inicio_Sit2	Toma valor 1 si la empresa se encuentra en situación 2 en el BCRA, en caso contrario toma valor 0.
CI_Inicio_Sit3	Toma valor 1 si la empresa se encuentra en situación 3 en el BCRA, en caso contrario toma valor 0.
CI_Inicio_Sit4	Toma valor 1 si la empresa se encuentra en situación 4 en el BCRA, en caso contrario toma valor 0.
CI_Inicio_Sit5	Toma valor 1 si la empresa se encuentra en situación 5 en el BCRA, en caso contrario toma valor 0.
CI_Inicio_Sit6	Toma valor 1 si la empresa se encuentra en situación 6 en el BCRA, en caso contrario toma valor 0.
CQ_I_BIN	Toma valor 1 si la empresa tuvo un concurso o quiebra, en caso contrario toma valor 0.
L_CI_Inicio_CantBancos_1	Logaritmo en base 10 de la cantidad de bancos en los que adeuda.
L_CI_Inicio_SumMonto_1	Logaritmo en base 10 de la suma que adeuda en bancos.
L_Edad	Logaritmo en base 10 de la antigüedad en meses de la empresa.
L_HC_NoPagados_Cant_I_1	Logaritmo en base 10 de la cantidad de cheques impagos.
L_HC_NoPagoMulta_Cant_I_1	Logaritmo en base 10 de la cantidad de multas impagas por cheques rechazados.
L_PQ_I_1	Logaritmo en base 10 de la cantidad de pedidos de quiebra.
SRT_Estado_I	Toma valor 1 si la empresa tiene deuda con la SRT, en caso contrario toma valor 0.
TotExpor	Valor en pesos de exportaciones.
TotImpor	Valor en pesos de importaciones.

Validación del Modelo

1. Medidas de bondad de ajuste. Los diversos R^2

En esta cuestión hay fuertes discrepancias entre los investigadores, pero estas medidas tienen en común que valores cercanos a la unidad son deseables y valores cercanos a 0 indican un ajuste pobre.

1.1. R^2 clásico

Llamando

$$(9)\quad Q=\sum_{t=1}^{n}(Y_t-P_t)^2\ ;\quad T=\sum_{t=1}^{n}(Y_t-\overline{Y})^2=\sum_{t=1}^{n}Y_t^{\,2}-\frac{1}{n}\left(\sum_{t=1}^{n}Y_t\right)^2$$

Ahora, dado que

$$\sum_{t=1}^{n}Y_t=r\ ;\qquad \sum_{t=1}^{n}Y_t^{\,2}=r$$

siendo r el número de individuos para los cuales Y_t=1, se tiene

$$(10)\quad T=r-\frac{r^2}{n}=r\left(1-\frac{r}{n}\right)$$

Entonces, el R^2 clásico es:

$$(11)\quad R^2_{clásico}=1-\frac{Q}{T}=1-\frac{Q}{r\left(1-\frac{r}{n}\right)}$$

1.2. R^2 de McFadden (Pseudo R^2)

$$(12)\quad R^2_{McFaddem}=1-\frac{Ln\,L}{Ln\,L_0}$$

En esta expresión, Ln L es el logaritmo de la verosimilitud máxima y Ln L_0 el logaritmo de la verosimilitud calculada sólo con el término b_0 y maximizada, o sea anulando todos los coeficientes de las variables explicativas. Se puede demostrar que la verosimilitud L_0 es igual a

$$(13)\quad L_0=\left(\frac{r}{n}\right)^r\left(1-\frac{r}{n}\right)^{n-r}\ ;\quad Ln\,L_0=rLn\frac{r}{n}+(n-r)Ln\left(1-\frac{r}{n}\right)$$

siendo r el número de individuos para los que es Y=1.

1.3. R^2 de ALDRICH-NELSON (Pseudo R^2)

$$(14)\quad R^2_{Aldrich-Nelson} = \frac{G}{n+G} = \frac{2(Ln\,L - Ln\,L_0)}{n+2(Ln\,L - Ln\,L_0)}$$

1.4. R^2 de COX-SNELL

$$(15)\quad R^2_{Cox-Snell} = 1 - Exp\left[-\frac{2}{n}(Ln\,L - Ln\,L_0)\right] = 1 - e^{-G/n}$$

1.5. R^2 de NAGELKERKE

$$(16)\quad R^2_{Nagelkerke} = \frac{1 - Exp\left[-\frac{2}{n}(Ln\,L - Ln\,L_0)\right]}{1 - Exp\left(\frac{2}{n}Ln\,L_0\right)} = \frac{R^2_{Cox-Snell}}{1 - Exp\left(\frac{2}{n}Ln\,L_0\right)}$$

Estos indicadores son medidas diferentes de la ganancia en la información que aportan las variables explicativas sobre la verosimilitud (probabilidad) con respecto a un cálculo que no las tuviera en cuenta.