

***Modelo de Comportamiento y Predicción
de Default***

El caso de empresas Pymes en Argentina

Autor: Bruno Plotnicki

Tutor: José P. Dapena

1. Introducción

Los instrumentos financieros en los mercados globalizados se han vuelto cada vez más complejos, creando incertidumbre y exposición dinámica entre contrapartes por lo cual el desarrollo de mecanismos de mejora del análisis de crédito para tomar riesgos de forma más sofisticada y anticiparse al default de las empresas ha sido una constante en los últimos años.

Los inconvenientes surgen dado que el proceso de evaluación de riesgo y asignación de crédito es un ejemplo de decisión hecha en un marco de incertidumbre, que incorpora información asimétrica entre el proveedor de fondos y el cliente.

La información asimétrica introduce dos cuestiones básicas que afectan a los proveedores de fondos: selección adversa (efectuar errores en el proceso de decisión) y riesgo moral (relacionado con el monitoreo de los deudores).

Cuando a una firma se le dificulta el cumplimiento en el pago del servicio de su deuda, existe lo que se denomina riesgo de default. Antes de que esto suceda, no existe forma de discriminar sin ambigüedades entre aquellas compañías que entrarán en default y cuales no. En el mejor de los casos, sólo se pueden efectuar análisis probabilísticos para calcular la probabilidad de default. Como resultado de ello, los deudores deben pagar a los primeros un spread que incluye una sobretasa de interés proporcional a la probabilidad de incumplimiento para compensar a los acreedores por esta incertidumbre.

Según evidencia empírica, el default es un evento poco común. La probabilidad de incumplimiento en una compañía tipo es de alrededor del 2% en cualquier año. Sin embargo, existe una variación considerable en la probabilidad, si la medición se efectúa entre empresas. Por ejemplo, las posibilidades para una sociedad con una calificación crediticia de AAA, el riesgo de no cumplir con sus obligaciones y entrar en default, no superan el 2 por 10000 (base anual).¹

¹ Ver “Modeling Default Risk”, Investigación de P. Crosbie y J. Bohn donde utilizan la metodología denominada “Expected Default Frequency™”, para definir la probabilidad de que una empresa entre en default en un período dado de tiempo. Para ello recaban información sobre trabajos anteriores en cuanto a probabilidades de incumplimiento en distintas compañías, utilizando para ello el valor de mercado de la firma, el nivel de su endeudamiento y la volatilidad del valor de la firma.

En cambio, en una empresa categorizada como CCC, las probabilidades se multiplican exponencialmente, alcanzando un 4 por 100, es decir 200 veces más que las chances de una firma con un rating altamente satisfactorio.

La pérdida sufrida por la contraparte en caso de incumplimiento es generalmente significativa y está altamente relacionada con las particularidades de los convenios y/o arreglos firmados. Las cláusulas de Cross Default en contratos de deuda usualmente aseguran que las probabilidades de default para cada una de las clases de deuda son las mismas. No obstante, la pérdida dependerá de la naturaleza de las obligaciones (garantías, plazo, privilegios, etc.).

Si bien todos los ítems resultan críticos para poder lograr un buen manejo de la administración de cartera de créditos, ninguno de ellos es tan importante o difícil de estimar, como el de probabilidad de default.

El objetivo de este trabajo se centrará en elaborar un modelo que permita predecir la probabilidad de default, utilizando información pública.

Se pueden evaluar dos componentes críticos para medir la probabilidad de default: información y modelos.

La información surge de tres fuentes principales:

- los estados financieros de la empresa (datos ex post)
- los precios de mercado de la Deuda y Patrimonio (datos ex ante).
- apreciaciones subjetivas sobre el riesgo y perspectivas de la compañía (datos ex ante)

Los modelos son los medios que permiten transformar los datos en probabilidad de incumplimiento.

Por otra parte, los modelos de comportamiento más precisos de medición derivarían de emplear todas las fuentes cuantitativas y cualitativas disponibles en el mercado.

Para estimar las señales de default para compañías, el modelo requiere operar el perfil de comportamiento normal de las empresas y calcular el desvío de los parámetros diseñados. Por ejemplo, al examinar el posible impacto de un cambio de directorio, perfil de bancos y

deuda, entre otros, se intentará obtener un estimador lo más exacto posible respecto estas señales y a las nuevas señales de default.

Estos requisitos llevan a un modelo basado en la correlación más que en la causalidad.

El problema para los financiadores, es que buscan la resolución de dos cuestiones fundamentales:

- Cuál es la probabilidad de incumplimiento.
- Cuál es la pérdida estimada en caso de default.

La “pérdida dado el incumplimiento” depende, básicamente, de cuan rápido se actúe, de las garantías y de la antigüedad en la industria.

El otorgamiento de facilidades crediticias conlleva una porción importante de pérdidas potenciales.

La expectativa de pérdida dado el incumplimiento es, consecuentemente, altamente dependiente de las facilidades otorgadas. El origen de ésta es la incertidumbre mientras que el riesgo es la probabilidad de default misma.

El desafío principal para los acreedores es caracterizar anticipadamente la distribución en las probabilidades de incumplimiento de cada deudor, y luego, monitorear la evolución en la calidad de los parámetros. Uno de ellos, los Estados Contables, no han evolucionado de manera tal que permita predecir la amenaza de problemas financieros de la compañía puesto que como se expuso anteriormente, se trata de información ex post.

Los otros inputs del proceso, es decir, los precios de mercado de Deuda y Patrimonio del deudor y las apreciaciones subjetivas, cuentan con una medida de incertidumbre inserta en él. No obstante, de éstos se puede generar información sumamente útil para estimaciones, dado que se trata de datos ex ante.

Es así como en un análisis de portafolio, la incorporación de todos los inputs presenta muchas ventajas, pues permite cuantificar costos de concentración para toda la cartera y las transacciones marginales, creando un marco donde considerar todas las agrupaciones por nombre, industria, país o producto.²

² En el trabajo de “Credit Risk Modeling and Credit Derivatives” P. Schönbucher aclara que tanto como el valor del activo de la firma, el valor de mercado del Patrimonio deben ser analizados en el contexto del riesgo de negocios de la empresa. Por ejemplo, compañías en el ramo alimenticio, pueden alcanzar niveles más altos de endeudamiento (y consecuentemente, menor valor de mercado del equity) que las compañías de tecnología pues sus negocios son más estables y menos inciertos.

2. Descripción del modelo

2.1 Herramienta a utilizar

El rol de la herramienta para armar un modelo es el de transformar el seguimiento de un grupo de parámetros externos en predictivos para la revisión de potenciales alertas.

Se trata de un *modelo de probabilidad lineal*.

El análisis de regresión múltiple como herramienta, se utiliza para demostrar el poder explicatorio y la significatividad estadística de las variables que resultan altamente correlacionadas.

La anticipación del modelo estará expresada en la distribución de los casos y en las medidas que el acreedor tome una vez detectada la desviación a los parámetros fijados por él.

El otro logro en el desarrollo de un modelo basado en un análisis de regresión múltiple, está caracterizado por la posibilidad de cambio de ponderaciones y ajustes de variables a la dinámica evolución de los mercados.

Como resultado del modelo que se origina en la utilización de esta herramienta, se puede detectar oportunidades de cross-selling, cuantificar en una rápida revisión la calidad del portafolio, mejorar la productividad, reducir costos y manejar indirectamente las relaciones con los clientes (al evaluar la performance de una cartera de clientes, actual o potencial).

2.2. Ecuación

Para determinar la relación entre las variables exógenas y la dependiente, se planteó la siguiente ecuación:

$$Y_i = \alpha + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \gamma_4 D_{4i} + \gamma_5 D_{5i} + \gamma_6 D_{6i} + \varepsilon_i, \text{ donde}$$

Y = variable endógena o dependiente, está dada por el concurso (que asume dos valores posibles, 0 cuando no presentó concurso, y 1 cuando sí se convocó).

D = variables exógenas o regresoras.

Ordenada al origen: α

Pendientes de la recta:

γ_1 = efectos de inhabilitación en la cuenta corriente

D_1 = 1 Inhabilitación por rechazo de más de cinco cheques (causal sin fondos) y sin posibilidad de rescatarlos.

= 0 Sin inhabilitación, o con menos de cinco cheques rechazados por sin fondos o rescatados.

γ_2 = efectos de Ingreso financieras/bancos categoría C y/o salida bancos categoría A últimos 18 meses.

D_2 = 1 Si se verifica movimiento de bancos.

= 0 Mantiene igual categoría de bancos.

γ_3 = efectos de Aumento / Disminución en la deuda bancaria, últimos 18 meses.

D_3 = 1 Si aumentó/disminuyó en un 20% los compromisos financieros.

= 0 Si aumentó/disminuyó en menos de un 20% los compromisos financieros.

γ_4 = efecto de demandas judiciales

D_4 = 1 Juicio/s como demandado.

= 0 Sin juicios.

γ_5 = efecto de pago de la ART

D_5 = 1 Rescisión por falta de pago.

= 0 Afiliación vigente.

γ_6 = efecto de inhibición por embargos de AFIP

D_6 = 1 Con embargos vigentes.

= 0 Sin embargos vigentes o ya levantados.

Término de error: ε_i

Como se desprende de la valoración de cada una de las pendientes, se espera que en todos los casos las mismas sean cercanas a 1, a medida que se observa una señal desfavorable (inhabilitación/rechazo de cheques, ingreso - egreso de bancos, demandas judiciales, falta de pago de ART, etc.) y haya mayor probabilidad de default previo al concurso.

Tipo de variables utilizadas: dicotómicas, las cuales se pueden utilizar en los modelos de regresión con la misma facilidad que las variables cuantitativas.

2.3. Selección de la muestra

Para el diseño del modelo se eligió una muestra inicial de 100 empresas Pymes, y luego para su validación una nueva muestra de 75 empresas, de acuerdo a los criterios que se detallan a continuación:

<i>Cuadro 1</i>		
	Muestra para el diseño	Muestra para la validación
Período muestreo (casos en default)	09/1999 - 12/2000	09/1999 - 12/2000
Período muestreo (casos no default)	09/1999 - 09/2002	09/1999 - 02/2003
Tipo de industria	Todas las ramas	Todas las ramas
Número de casos	100	75
Número de casos en default	30	20
Tasa de default	30.00%	26.67%

- ✓ El universo de empresas concursadas se tomó entre fines 1999 y fines de 2000, para evitar la información sesgada que se generó a partir de 2001 con el flujo negativo de capitales y la declaración argentina de default a principios de 2002.
- ✓ Fuente de datos: información proporcionada de las bases de empresas de Credit Bureau en los períodos comprendidos.

3. Aplicación y resultados

3.1. Regresión

A los efectos de conocer el impacto de cada uno de los factores que participan en la determinación estadística del modelo, se detalla a continuación un resumen, incluyendo el Análisis de Varianza:

Cuadro 2

Resumen

Estadísticas de la regresión	
Coef correlación múltiple	0.671184166
Coef determinación R ²	0.450488184
R ² ajustado	0.415035809
Error típico	0.352254614
Observaciones	100

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Prom de cuadrados	F	Valor crítico de F
Regresión	6	9.46025187	1.576708645	12.70685481	1.99521E-10
Residuos	93	11.53974813	0.124083313		
Total	99	21			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	0.084221942	0.045276679	1.860161649	0.066023156	-0.005688479	0.174132363
Inhab	0.38195028	0.127954711	2.985042736	0.003622868	0.127857838	0.636042723
E/S	0.21994419	0.106765114	2.06007544	0.042185176	0.007930044	0.431958336
Variac	0.092132899	0.089265754	1.032119215	0.304692906	-0.085131019	0.269396818
Juicios	0.320146039	0.104040139	3.077139666	0.002745676	0.113543149	0.52674893
ART	0.085301446	0.12251692	0.696242168	0.488013631	-0.157992633	0.328595526
Oficios	0.493599454	0.118048868	4.181314599	6.55189E-05	0.259178033	0.728020875

3.2. Interpretación de los resultados

La ordenada al origen (α) representa la probabilidad de que una empresa con ART al día, sin oficios por embargo, con deuda y cantidad de bancos sin cambios, sin juicios como demandado ni inhabilitación se haya concursado. Si bien el coeficiente no es comparativamente alto frente al resto de las variables, es estadísticamente significativo al 7%, por lo cual se incluyó en el análisis.

En el caso de las pendientes, por ejemplo, el 0.3819 asociado a D_1 significa que, si las demás variables permanecen constantes, se espera que la probabilidad de que una empresa haya solicitado su concurso sea más alto en un nivel de alrededor del 38.19% en el caso de que haya sido inhabilitada que para aquellas compañías que no registraban inhabilitación en sus cuentas corrientes.

Todos estos factores corresponden al componente sistemático y determinístico de la variable que tratamos de interpretar (Y). Por ello, ϵ_1 sustituye a todas las variables ignoradas u omitidas que pueden afectar a Y.

3.2.1. Conclusiones

- El coeficiente R^2 permite evaluar la bondad del ajuste de la línea de regresión ajustada a un conjunto de datos y da una idea del poder explicatorio de la regresión. El análisis empírico de R^2 , expresa la proporción de la variación en la variable dependiente explicada por las variables regresoras, y alcanza el 41.50%.

- Del mismo análisis se han encontrado cuatro coeficientes de regresión que son estadísticamente significativos al 5%: inhabilitación en cuenta corriente, entrada/salida de bancos, demandas judiciales y oficios por embargos.

- Una propiedad importante del R^2 es que es una función no decreciente del número de variables explicativas o de regresores presentes en el modelo. Ello se observa en los anexos, en donde al agregar nuevas variables el coeficiente de determinación ajustado pasa de 35.33% (5 variables) al 41.50% (6 variables).

- Asimismo, se puede observar que las variables “aumento/Disminución 20% en deuda” e “incumplimiento en ART” no son estadísticamente significativas, y superan el límite para determinar la razonabilidad y confiabilidad de la variable (entre 1% y 10%), pero dado que los coeficientes resultantes son inferiores al resto, ello no perjudica a las otras variables y a la vez potencia las posibilidades de explicación del modelo.

3.2.2 Prueba F:

- Esta prueba se utiliza para evaluar la significancia estadística conjunta de todos los coeficientes utilizados en el modelo, dado que no se puede utilizar la prueba t usual.

- Si el valor calculado excede el valor F crítico de la tabla estadística respectiva al nivel de significancia α , se rechaza la hipótesis nula (es, decir, que todos los coeficientes de pendientes parciales son simultáneamente iguales a cero). De la información de la tabla se evidencia que para la cantidad de observaciones muestreadas y el nivel de significancia elegido (99%) el valor de F crítico es de 2.96, ampliamente superado por el 12.71 que surge del análisis.

- Por lo tanto, se comprueba que la hipótesis nula es falsa, es decir que las variables explicatorias definitivamente ejercen influencia sobre Y.

4. Validación del modelo

4.1. Determinación del límite

Una medida para evaluar la habilidad del modelo para segmentar entre empresas buenas y malas (con posibilidad de ingresar en default) es el Estadístico Kolmogorov-Smirnoff (K-S) o de Separación.

Esta herramienta permite verificar la máxima diferencia entre las distribuciones acumuladas de las cuentas que no entraron en default (ND) comparadas contra las que sí se concursaron (D). El Estadístico se muestra como porcentajes, siendo 100% el mejor desempeño, pues da una idea de completa separación, mientras que 0% evidencia la no-discriminación entre los dos grupos.

De acuerdo a la aplicación en modelos de scoring para el área de banca minorista³, la siguiente tabla exhibe las diferentes puntuaciones para diversas industrias:

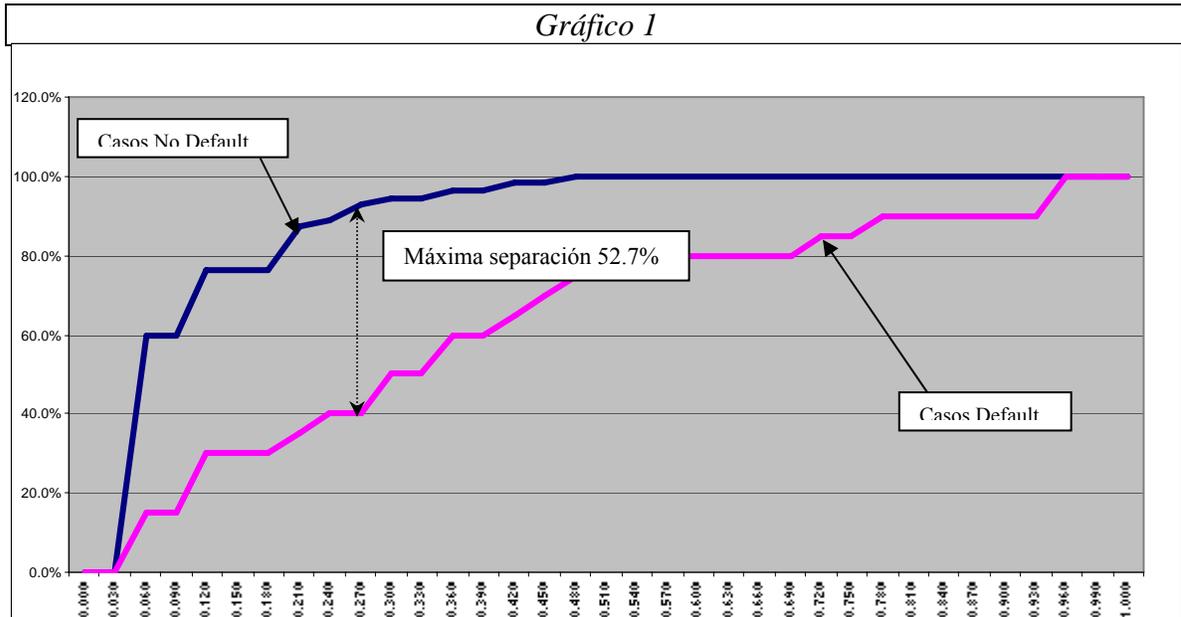
<i>Cuadro 3</i>	
Categoría	Puntuación (expresado en %)
Mínimo	20%
Bueno	30%
Muy bueno	40%
Excelente	> 45%

Según lo descripto en el apartado 3.1, se tomó una muestra de validación para verificar la efectividad sobre una nueva cartera de empresas y determinar si es posible realizar una segmentación.

Gráficamente, del Estadístico de Separación se observa que el 92.7% de los casos en default presentan una calificación de 0.27 o menos y sólo el 40.0% de las empresas no concursadas tiene el mismo puntaje. El estadístico K-S (máxima separación) es de 52.7% (92.7%-40.0%).

De lo anterior se verifica que el modelo diseñado para la validación está bien por encima de los niveles de aceptación mínimos.

³ Uno de los modelos de validación utilizados mediante la técnica del Estadístico de Separación, ha sido elaborado por la consultora Statistical Decisions, quien prepara informes para bancos y entidades financieras, básicamente con el propósito de Due Dilligence.



Otro de los tests comúnmente utilizados para estimar la probabilidad de default es el gráfico de Perfil de Certeza Acumulada (Cumulative Accuracy Profile - CAP) y la Tasa de Certeza (Accuracy Ratio - AR).

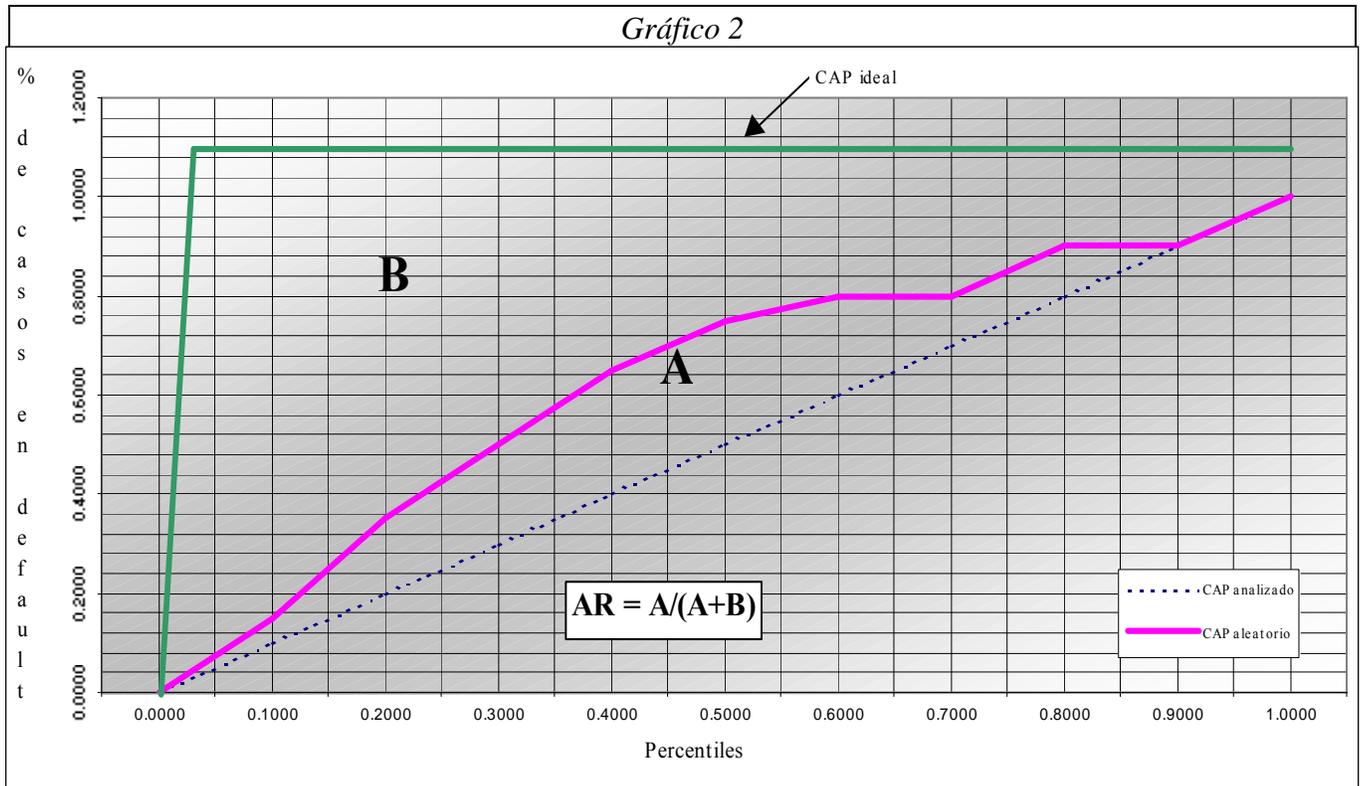
La curva CAP representa la distribución de probabilidad acumulada de eventos de default para diferentes percentiles en la escala de un parámetro que se desea medir.

Mientras que el CAP es una forma conveniente de visualizar el desempeño del modelo, se hace necesario contar con otra medida que totalice la precisión predictiva de los factores a evaluar. La AR se construye con este propósito. Esta métrica se obtiene al comparar el CAP con el ideal (predicción perfecta) y una línea de 45° llamada caso naíve (o aleatorio), pues no se obtiene información valiosa de la misma.

El AR es la fracción entre 0 y 1, siendo que cuando el factor que se desea testear es cercano a 0 muestra un bajo poder predictivo, mientras que valores próximos a 1 exhiben una estimación casi perfecta, similar a la información que se obtiene del R^2 en una regresión.

Del gráfico que se expone a continuación, se observa un AR de 0.268. Si se tiene en cuenta que los AR's de los principales modelos de riesgo crediticio (p.e. Z-Altman, Moody's Risk

Calc, etc.) se encuentran en un rango de entre 0.50 y 0.75, entonces un ratio como el obtenido resulta interesante y sólido⁴.



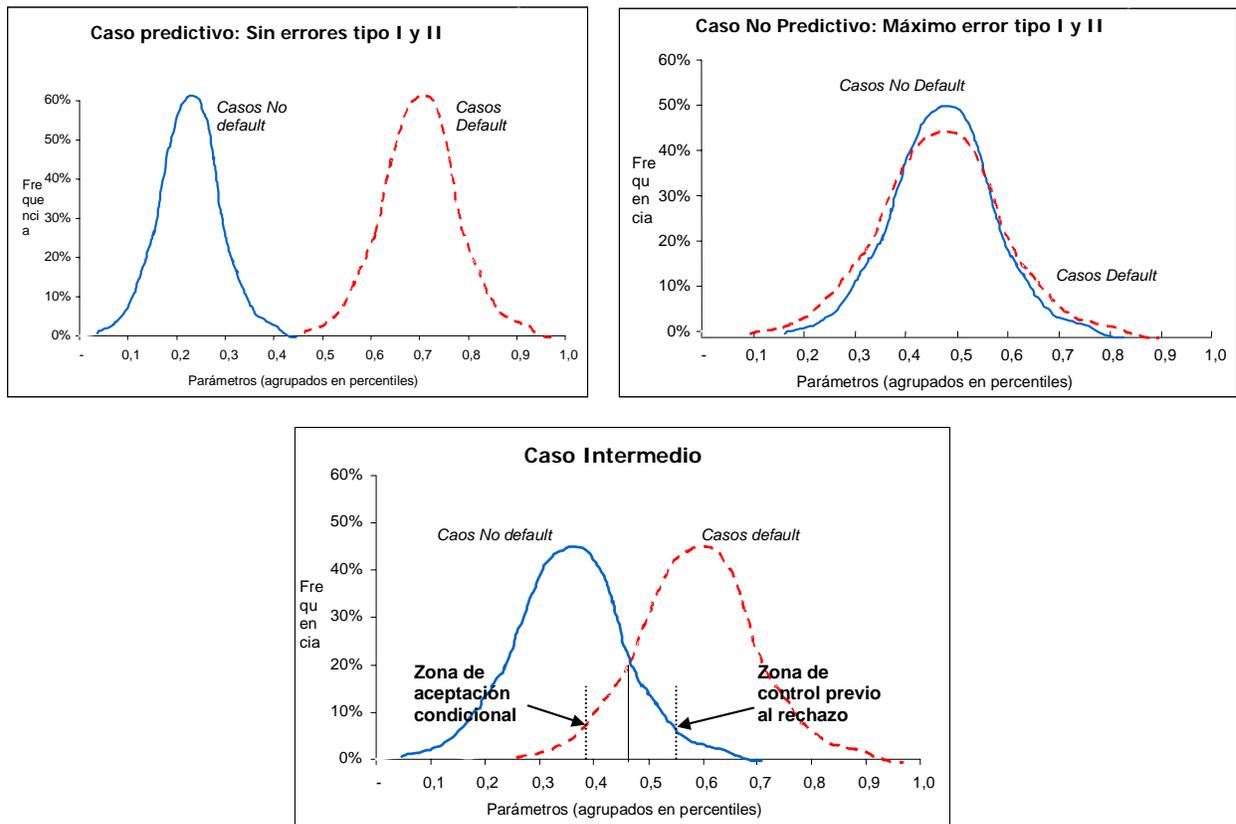
El último método para estimar y validar la predicción de default es el “Poder Discriminatorio Univariado” (UDP), que consiste en comparar dos histogramas: uno para las empresas en default y otro para las empresas que no entraron en default. La potencia predictiva se observa a través de la distancia entre ambas distribuciones.

Cuanto más separadas estén las distribuciones de probabilidad, más eficaz es el método para predecir el concurso o default.

Asimismo, esta herramienta permite observar claramente las dos formas en que los resultados de la aplicación de las variables estudiadas en el modelo del punto 2.2, pueden fallar. En principio, dado que no existe ningún factor perfectamente anticipatorio, cierto nivel en el resultado puede indicar bajo riesgo, cuando en realidad, el riesgo es alto. Esto corresponde a las empresas con baja calificación en el modelo diseñado, que sin embargo

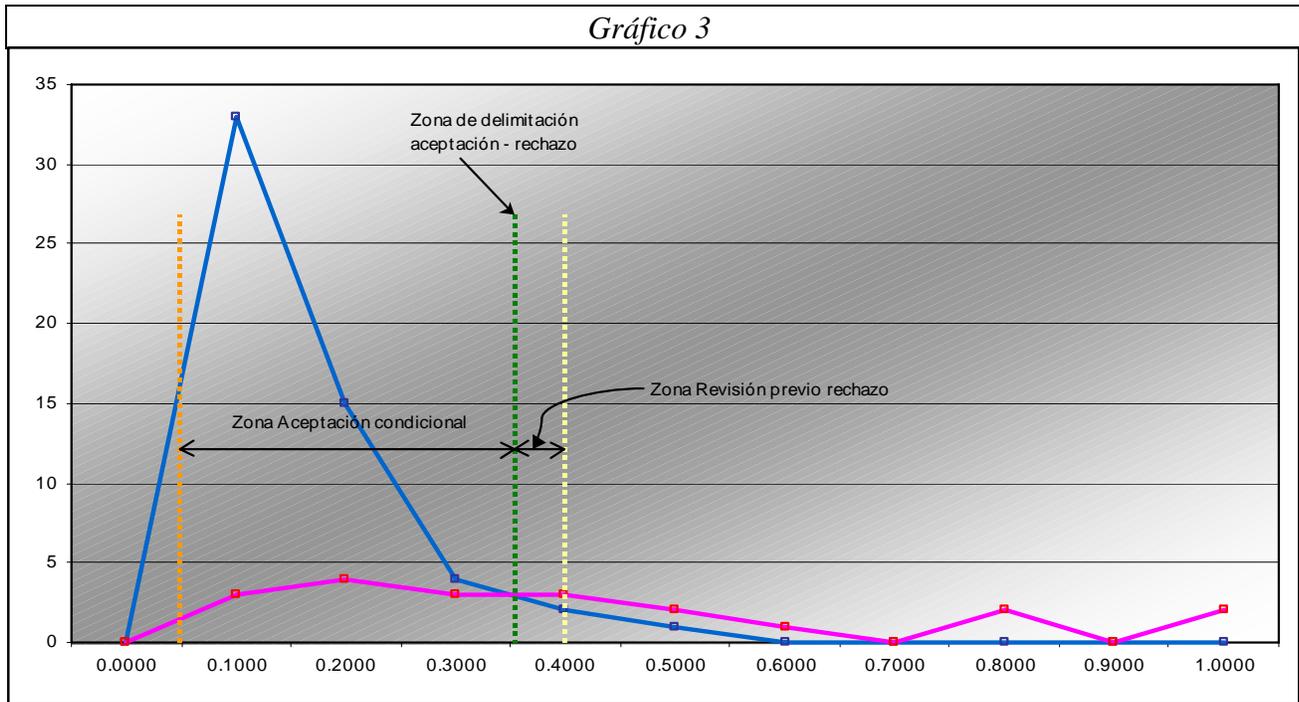
⁴ Ver “Benchmarking Quantitative Default Risk Models: A Validation Methodology”, elaborado por J.Sobehart, S. Keenan y R. Stein para Moody’s Investors Service.

se concursan. Se trata de un fenómeno conocido como “Error de tipo I”. Por otro lado, están los casos de firmas que tienen un alto puntaje en la ponderación de parámetros (riesgoso) pero que sin embargo, no han entrado en default. Este es el “Error de tipo II”. Los ejemplos ilustrados a continuación, describen en el ámbito teórico ambos tipos de error y destacan el hecho de que al intentar minimizar un tipo de error, necesariamente es a expensas de incrementar el otro tipo de error.



Al aplicarlo sobre el modelo construido, el resultado es el siguiente:

Gráfico 3



De acuerdo con este gráfico, las reglas de decisión para fijar los criterios de aceptación o rechazo están dadas por las siguientes categorías:

- ✓ *Aceptar*: cuando el límite es igual o inferior a 5%.
- ✓ *Aceptación condicional*: entre 5% y 35%.
- ✓ *Revisión previa al rechazo*: entre 35% y 40%.
- ✓ *Rechazar*: superior al 40%.

Bajo este esquema, se pueden determinar los porcentajes de distribución de las 75 observaciones, como se muestra en el siguiente cuadro:

<i>Cuadro 4</i>		
Zona Aceptación	60.0%	Empresas no default
	15.0%	Empresas en default
Zona Aceptación condicional	34.5%	Empresas no default
	35.0%	Empresas en default
Zona Revisión previo rechazo	3.6%	Empresas no default
	10.0%	Empresas en default
Zona Rechazo	1.8%	Empresas no default
	40.0%	Empresas en default

Conclusiones

Tal como se estableció al comienzo de este trabajo, el objetivo fue elaborar un modelo que permita predecir la probabilidad de default, utilizando información pública y hallar cuáles son los principales factores que afectan el desempeño de una empresa y que pueden comprometer seriamente su capacidad para cumplir obligaciones, llegando inclusive a la convocatoria de acreedores y/o quiebra.

Las conclusiones principales y sus implicancias, se pueden resumir en lo siguiente:

- Una de las principales variables que explican el modelo es la “Inhabilitación en cuenta corriente”, la cual es un indicador de falta de liquidez (pues la misma surge del rechazo de cheques por causa sin fondos, y que no han podido ser rescatados) y de mayor desconfianza en el circuito financiero, por lo cual la empresa debe recurrir a entidades financieras de “tercera línea”, cuando las de primer orden retiran su apoyo.
- Tal como se expone en Anexo 1, las correlaciones del default son positivas, independientemente de la industria en la que se encuentren las empresas muestreadas.
- Con la utilización de técnicas como K-S y AR aplicadas a una nueva muestra - denominada de validación - para generar robustez en el alcance del trabajo y fundamentar que los parámetros elegidos y sus pesos relativos (coeficientes) podían ser aplicados a nuevos casos, se demostró que el modelo aún mantiene su poder predictivo y validez.
- Mediante la determinación de reglas de decisión, se observa que aproximadamente el 95% de las empresas no concursadas son aceptadas por el modelo (incluye 35% de aceptación condicional, que podría requerir de - por ejemplo - referencias bancarias y/o comerciales para definir su aceptación definitiva) versus el 50% de las concursadas (en donde sólo 15% corresponde a aceptación sin restricciones). En cambio, el 50% de las empresas en default son rechazadas por el modelo (incluye un 10% de empresas que necesitan ser controladas), frente al 5% de las no concursadas.

Si bien el modelo no resuelve todos los problemas crediticios, puede ser complementado con un adecuado análisis del Sistema de Información Gerencial interno, para retroalimentar y fortalecer las conclusiones del modelo.

Bibliografía:

- 📖 “Modeling Default Risk” (1993), Moody’s KMV. Peter J. Crosbie y Jeffrey R. Bohn.
- 📖 “Credit Risk Modeling and Credit Derivatives” (1999), Tesis doctoral. Philipp J. Schönbucher.
- 📖 “Input Modeling Tools for Complex Problems” (1998), Northwestern University. Michael Yamnitsky y Barry Nelson.
- 📖 “Benchmarking Quantitative Default Risk Models: A Validation Methodology”, (2000) elaborado por J.Sobehart, S. Keenan y R. Stein para Moody’s Investors Service.
- 📖 “Correlated Default Risk” (2002), trabajo a cargo de Sanjiv Das, Laurence Freed, Gary Geng y Nikunj Kapadia.
- 📖 “Estadística para las Ciencias Administrativas”, Lincoln Chao (edición 1993).
- 📖 Información pública, provenientes de la empresa de Credit Bureau Nosis (www.nosis.com.ar)

Anexo 1

Correlación entre variables del modelo

	Conc	Inhab	E/S	Variac	Juicios	ART	Oficios
Concurso	1.00						
Inhabilitación	0.47	1.00					
Entrada Bcos C/Salida Bcos A	0.31	0.37	1.00				
Variación deuda +/- 20%	0.20	0.05	0.18	1.00			
Juicios como demandado	0.30	0.13	0.06	0.15	1.00		
Incumplimiento en ART	0.30	0.26	0.17	0.04	-0.06	1.00	
Oficios por embargos	0.46	0.24	-0.01	0.09	0.02	0.41	1.00

La tabla reporta que, salvo dos excepciones, las correlaciones entre distintas variables que permiten explicar el default son positivas.

Correlación entre inhabilitación y cheques rechazados rescatados

	Inhab	Ch rech resc
Inhabilitación	1.00	
Ch rech resc	0.53	1.00

En el caso particular de la inhabilitación y el evento de rechazo de más de cinco cheques por causal sin fondos girados en cantidades exactas que no pudieron ser rescatados, se quiso demostrar la alta correlación entre ambos sucesos, a efectos de justificar la utilización de la primera, como variable explicatoria del concurso. El resultado, positivo y superior a 0.50, es satisfactorio y significativo a los fines propuestos.

Anexo 2

Efecto en el agregado de variables en el análisis de regresión múltiple

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correl	0.621318173
Coefficiente de deterr	0.386036271
R ² ajustado	0.353378626
Error típico	0.370354033
Observaciones	100

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>medio de los cuadra</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	5	8.106761701	1.62135234	11.8207014	7.00873E-09
Residuos	94	12.8932383	0.13716211		
Total	99		21		

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
Intercepción	0.125631446	0.045915603	2.736138451	0.007432713	0.034464987	0.216797905
Inhab	0.410514295	0.135598196	3.027431823	0.003182875	0.141281034	0.679747555
E/S	0.17819101	0.110365345	1.614555819	0.109758837	-0.040941858	0.397323878
Variac	0.072062442	0.095225214	0.75675799	0.451087466	-0.117009357	0.26113424
Juicios	0.365337974	0.110647661	3.301813803	0.001359053	0.145644561	0.585031386
ART * Oficio	0.573683183	0.169689902	3.380773837	0.001054279	0.236760075	0.91060629

En el modelo construido y validado con 6 variables, la proporción de la variación en la variable dependiente explicada por las variables regresoras (R² ajustado) alcanza el 41.50%. En cambio en esta simulación con 5 variables, la proporción desciende al 35.33%.