



Revista Ingenierías Universidad de Medellín

ISSN: 1692-3324

revistaingenierias@udem.edu.co

Universidad de Medellín

Colombia

Fernández Castaño, Horacio; Pérez Ramírez, Fredy Ocaris

El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito

Revista Ingenierías Universidad de Medellín, vol. 4, núm. 6, enero-junio, 2005, pp. 55-75

Universidad de Medellín

Medellín, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=75040605>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



El modelo logístico: Una herramienta estadística para evaluar el riesgo de crédito

HORACIO FERNÁNDEZ CASTAÑO:

Magíster en Matemáticas Aplicadas de la Universidad EAFIT, Ingeniero Civil de la Escuela de Ingeniería de Antioquia, Especialista en Sistemas de Administración de la Calidad ISO 9000, Especialista en Gerencia de Construcciones y Licenciado en Matemáticas de la Universidad de Medellín. Profesor de tiempo completo, Facultad de Ingenierías. Programa de Ingeniería Financiera. Universidad de Medellín. e-mail: hfernandez@udem.edu.co

FREDY OCARIS PÉREZ RAMÍREZ:

Matemático de la Universidad de Antioquia, Magíster en Matemáticas Aplicadas de la Universidad EAFIT y Estudios de Especialización en Estadística de la Universidad Nacional, sede Medellín. Profesor de tiempo completo, Facultad de Ingenierías. Programa de Ingeniería Financiera. Universidad de Medellín. e-mail: foperez@udem.edu.co

RESUMEN

Las propuestas contenidas en el documento del nuevo Acuerdo de Capital de Basilea representan un sólido esfuerzo por incorporar los desarrollos recientes de la teoría financiera de cuantificación del riesgo en la determinación del capital regulatorio mínimo que debe ser exigido a las entidades financieras. Es propósito entonces no incrementar los requerimientos agregados de capital para la banca que utilice el sistema estandarizado de medición del riesgo de crédito, y de estimular la migración hacia la utilización del sistema basado en los rating internos.

En la práctica diaria se presentan situaciones en las que es necesario tomar decisiones. Es frecuente que estas elecciones se hagan de manera automática, y haciendo solo uso de un razonamiento lógico. Sin embargo, en otras situaciones la toma de alguna decisión, con el objeto de determinar una estrategia óptima, se requiere de un análisis más profundo y para lo cual la mera intuición es insuficiente.

Por ejemplo, para una entidad financiera es de gran utilidad disponer de un modelo, alejado de la mera intuición, que le ayude a decidir sobre otorgar o no un crédito. La entidad debe evaluar la probabilidad de que el cliente devuelva el dinero, lo cual será un hecho positivo para la entidad, o bien que el cliente llegue a ser moroso, y esto será un hecho negativo para la entidad.

Ante esta situación, de elección binaria, es necesario hacer uso de modelos dicotómicos que permitan evaluar la probabilidad asociada a cada alternativa de decisión. Es de gran ayuda utilizar un modelo que, de acuerdo a lo ocurrido con otros préstamos, con diferentes estratos económi-

cos, diferentes salarios, diferentes niveles de escolaridad, entre otros, permita calcular la probabilidad de que el cliente cancele el préstamo, y con base en los resultados tomar una decisión después de una reflexión más profunda, de manera que se calcule la provisión necesaria para cubrir eventualidades de morosidad.

En este artículo se presenta un planteamiento no lineal de los modelos de elección dicotómica que, sin duda, es una buena elección para evaluar el riesgo de crédito de una cartera comercial, que se incluye aquí como una aplicación.

Palabras clave: Basilea, elección binaria, modelos dicotómicos, probabilidad

ABSTRACT

The proposal contained in the document of the new Basel's agreement represents a solid effort to incorporate the results of the development of the financial theory of risk quantification in the determination of minimum regulatory capital that must be requested in the financial entities. Then, the purpose is not to increment the added requirements of capital for the banking house that uses standardized system of the risk credit measurement and the stimulation upon the use of the system based in the internal ratings.

In the daily practice, it presents situations in which, it's necessary to take decisions. It is frequently that these choices are made in automatic way and just making use of a logic reasoning. However in the other situations taking some decisions with the purpose of determinations a very good strategy, it requires a deeper analysis in which the simple intuition is not enough.

For example, for a financial entity is very useful to have a model away of just intuition which have for to take a decision whether they give or not a credit. The financial entity must evaluate the probability the customer gives back the money which will be a positive fact for the entity, now then that the customer becomes debtor, and this will be a negative fact for the entity.

Before this situation, of binary choice, it is necessary to make used dichotomy models that allow evaluating the associated probability to every alternative. It's a great help to use a model that, concerning about that happened with other loans. With different economic levels, different salaries, and different academic levels, among others, it permits to calculate the probability that the customer pays for the loan and based on these results they can take a decision after a deeper meditation.

Key words: Basel, binary choice, dichotomies models, probability.

INTRODUCCIÓN

Las finanzas, como disciplina científica, estudian la manera como se asignan recursos escasos a lo largo del tiempo en condiciones de incertidumbre. Existen tres pilares analíticos de las finanzas: la optimización en el tiempo, la valuación de activos y la evaluación y administración del riesgo. Este último de gran

importancia en la actualidad, debido al control que están ejerciendo las agencias internacionales sobre los países en desarrollo, y por esto se ha intensificado su estudio e investigación.

La incertidumbre es una de las características principales con las cuales debe convivir una institución financiera. Una amplia serie de fe-

nómenos, cuyo comportamiento es impredecible, tiene un impacto directo en el desempeño de dichas instituciones. En el caso de las compañías de seguros, éstas tienen que realizar erogaciones por concepto de accidentes u otros eventos asegurados. En el caso de las instituciones bancarias, ellas tienen que hacer provisiones de capital para hacer frente a pérdidas originadas tanto por la calidad crediticia de sus deudores como por los cambios en los factores de mercado que afectan sus portafolios.

El análisis de las variaciones de los factores cuyo comportamiento es impredecible puede ser realizado por medio de diversas herramientas estadísticas, lo cual, en el caso particular de las compañías de seguros, ha dado lugar a la teoría del riesgo. El desarrollo de la teoría de riesgos ha permitido a las compañías de seguros conocer mejor la exposición de sus portafolios y establecer las pérdidas a las cuales se exponen. Sólo en tiempos recientes se ha explotado esta herramienta en el ámbito financiero, debido a la similitud existente con el principal riesgo que las instituciones financieras enfrentan: el riesgo crediticio.

La administración del riesgo, como puede verse, ha evolucionado notablemente en los últimos años y si se hace un análisis resumido de esta evolución, no se podrían dejar de repasar los diversos estudios realizados por el comité de Basilea.

El comité de Basilea (ciudad del norte de Suiza), se creó en 1974 por los gobernantes del G-10 con el propósito de coordinar la supervisión de los bancos internacionales. El comité es un foro de debate sobre la supervisión bancaria cuyo fin es ser garantes de una supervisión eficaz de las actividades bancarias en todo el mundo.

Recientemente se ha convocado al análisis de propuestas metodológicas para el Nuevo Acuerdo de Basilea sobre Capitales en la

Banca, en el entendido de que en el ámbito de una mayor globalización, las instituciones financieras requieren ampliar sus alternativas para representar el riesgo y mantener adecuados índices de capitalización. Uno de los aspectos centrales de las propuestas del Comité es el manejo del riesgo de crédito.

El riesgo, como la probabilidad de obtener un resultado no esperado, hace necesario que en su estudio se deban tener en cuenta las matemáticas para modelar los procesos de optimización, dado que las características de las variables se ven reflejadas mediante procesos estocásticos. Es necesario que en el planteamiento se incluya esta situación y para ello se deben considerar como pilares: La Estadística, los Procesos Estocásticos, la Simulación, las Series de Tiempo y la Econometría.

En este contexto, se considera importante que en las diferentes regiones latinoamericanas se fomenta una actitud proactiva en cuanto a las metodologías de administración del riesgo crediticio y los requerimientos de capital para la banca. En ese sentido, se ha desarrollado una serie de propuestas para adaptarse a las metodologías requeridas, que tienen como finalidad fortalecer la gestión de los diferentes riesgos de mercado [11], de crédito, de liquidez, contraparte y operativo, y otros inherentes a la Banca.

El pilar fundamental de la supervisión bancaria está en la gestión interna del riesgo en las entidades financieras y, como tal, está orientado a que se fortalezca esta gestión en riesgos de crédito y de mercado, generando cultura y prácticas de alto nivel técnico en la administración del mismo.

Se deben establecer, así mismo, los principios y criterios generales que las entidades deben adoptar para mantener adecuadamente evaluados los riesgos crediticios implícitos en los activos de crédito, y definir las modalidades

de crédito, determinar las calificaciones que se deben otorgar a tales operaciones según la percepción de riesgo que de las mismas se tenga, establecer la periodicidad con que se deben efectuar tales calificaciones y consagrar los mecanismos de recalificación.

El valor en riesgo crediticio se encuentra todavía en su infancia y está creciendo, en parte, como consecuencia de las instituciones reguladoras; es por esto que en Colombia se ha creado la necesidad de que en las entidades financieras se conformen grupos de trabajo en esta temática, con una rigurosa formación Matemática, Estadística Multivariada, la Simulación, los Procesos Estocásticos y la Econometría, que les permita hacer las valoraciones del riesgo de manera científica y así puedan cumplir con los requerimientos internacionales para ser garantes de un buen manejo de los elementos financieros de alto riesgo. Es necesario modelar, desarrollar y analizar sistemas de información que permitan evaluar el riesgo de crédito, y hacer las provisiones de cartera.

El incremento en la varianza de las principales variables financieras ha creado un nuevo campo, en la Ingeniería Financiera, cuyo objetivo es proporcionar alternativas creativas para protegerse contra los riesgos financieros o para especular con ellos.

El riesgo puede ser definido como la volatilidad de los flujos financieros no esperados, generalmente derivada del valor de los activos o los pasivos. Las empresas están expuestas a tres tipos de riesgo, que son:

- Los riesgos de negocios son aquellos que la empresa está dispuesta a asumir para crear ventajas competitivas y agregar valor para los accionistas.
- Los riesgos estratégicos son los resultantes de cambios fundamentales en la economía o en el entorno político.
- Los riesgos financieros están relacionados con las posibles pérdidas en los mercados

financieros. La extensa literatura sobre crisis financieras internacionales y el análisis de los agregados macroeconómicos y bancarios para varios países, entre ellos Colombia, han resultado interesantes a la hora de ahondar en el estudio del desempeño del sistema bancario y la economía durante el período de transición de una crisis financiera; el colapso crediticio es una característica de una crisis bancaria y un período en el cual los prestatarios y los deudores ajustan sus balances.

Toda empresa está expuesta a riesgos que pueden causar graves problemas financieros, poner en peligro la operación e, incluso, llegar al cierre definitivo de la empresa. La administración del riesgo ha evolucionado notablemente en los últimos años y si se hace un análisis resumido de esta evolución, no se podrían dejar de repasar los diversos estudios realizados por el Comité de Basilea para la Supervisión Bancaria.

Un modelo econométrico que es de gran utilidad, en la evaluación del riesgo de crédito el modelo de regresión logística [6, 17] que modela la toma de decisiones cuando se está enfrente de un proceso de elección binaria, y el proceso de decisión de la probabilidad asociada a cada alternativa posible que puede tener un cliente.

MODELO LOGIT

La regresión Logit [6] se utiliza cuando queremos predecir un resultado binario, por ejemplo, quiebra vs. no quiebra y sabemos que existen varios factores que pueden incidir sobre tal resultado. Esta regresión binaria es un tipo de análisis de regresión donde la variable dependiente es una variable dummy: código 0 (Buen Cliente) o 1 (Mal Cliente).

La regresión logística se basa en la denominada función logística, donde se relaciona la variable dependiente con las vari-

ables independientes $x_1, x_2, \dots, x_i, \dots, x_k$ a través de la siguiente ecuación:

$$Y_i = \frac{1}{1 + \exp(-z)} + u_i$$

Donde,

Y_i : Variable dependiente. Puede tomar valores de cero o uno.

z : Scoring Logístico.

En el cual $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$, o de forma matricial sería $\mathbf{Z} = \beta \mathbf{X}$

u : Es una variable aleatoria que se distribuye normalmente $N(0, \sigma^2)$

$x_1, x_2, \dots, x_i, \dots, x_k$ Las variables independientes son fijas en el muestreo.

Si denotamos por Y a la variable a predecir, y por $x_1, x_2, \dots, x_i, \dots, x_k$ a las k variables predictoras, la regresión logística se expresa de la manera siguiente:

$$P(Y=1|X) = P(Y=1|X_1=x_1, \dots, X_k=x_k) = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}$$

Donde X representa un patrón a clasificar, y $\beta_0, \beta_1, \dots, \beta_k$ son los parámetros, que deben ser estimados a partir de los datos, a fijar para tener determinado un modelo concreto de regresión logística.

Si consideramos que la variable a predecir Y es binaria, podemos calcular $P(Y=0|X)$ de la siguiente manera:

$$P(Y=0|X) = 1 - P(Y=1|X) = 1 - \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]} = \frac{\exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}$$

Otro concepto de interés es el de odds ratio (OR) de un determinado patrón x el cual se denota por $OR(x)$ y se define como el cociente entre la probabilidad de que el patrón pertenezca a la clase 1 entre la probabilidad de que el patrón pertenezca a la clase 0. Es decir:

$$OR(X) = \frac{P(Y=1|X)}{1 - P(Y=1|X)}$$

Para trabajar con $OR(x)$ de manera ágil, es conveniente expresar el modelo de regresión logística en la manera logit. Para ello se efectúa una transformación del modelo, de la manera siguiente:

$$\text{logit}(P(Y=1|X)) = \ln OR(X) = \ln \left[\frac{P(Y=1|X)}{1 - P(Y=1|X)} \right]$$

Sustituyendo en la fórmula anterior las expresiones correspondientes al modelo logístico obtenemos:

$$\begin{aligned} \text{logit}(P(Y=1|X)) &= \ln \left[\frac{\frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}}{\frac{\exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right]}} \right] \\ &= \ln \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right) = \beta_0 + \sum_{i=1}^k \beta_i x_i \end{aligned}$$

Tal y como se ha comentado anteriormente, el odds ratio (OR) de un individuo con patrón x se define como el cociente entre la probabilidad de que $Y=1$ dado dicho patrón X y la probabilidad de que $Y=0$ dado X . Así un odds ratio de $1/3$ para un patrón X se interpreta diciendo que para dicho patrón la probabilidad de que se de $Y=1$ es una tercera parte de la probabilidad de que $Y=0$.

Además se tiene que

$$\ln(OR(0)) = \ln \left[\frac{P(Y=1|X)}{1-P(Y=1|X)} \right] = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

Y por lo tanto si: $x = 0, 0, 0, \dots, 0$ entonces, $\ln(OR(x)) = \beta_0$. Ahora bien, en un modelo de regresión logística, el coeficiente β_i ($i = 1, 2, \dots, k$) representa el cambio en el logit resultante al aumentar una unidad en la i -ésima variable X_i ($i = 1, 2, \dots, k$)

Otro concepto que resulta de interés es el de risk odds ratio (ROR) de x_0 frente a x el cual mide el riesgo del odds ratio de x_0 frente al odds ratio de x ($OR(x)$) es decir:

$$ROR_{x',x} = \frac{OR(x')}{OR(x)} = \frac{\exp \left(\beta_0 + \sum_{i=1}^k \beta_i x'_i \right)}{\exp \left(\beta_0 + \sum_{i=1}^k \beta_i x_i \right)} = \exp \left(\sum_{i=1}^k \beta_i (x'_i - x_i) \right)$$

Obviamente $ROR_{x',x}$ se puede expresar de manera alternativa como:

$$ROR_{x',x} = \prod_{i=1}^k \exp[\beta_i (x'_i - x_i)] = \exp[\beta_1 (x'_1 - x_1)] \dots \exp[\beta_k (x'_k - x_k)]$$

La estimación de los parámetros del modelo logit, se realiza por máxima verosimilitud (apéndice).

Para la aplicación del modelo Logístico se requiere que las variables exógenas sean linealmente independientes y esto obliga a hacer uso del análisis de componentes principales [8], que a continuación se presentan.

ANÁLISIS DE COMPONENTES PRINCIPALES

El Análisis de Componentes Principales es un método estadístico, muy útil en la evaluación del riesgo crediticio y fue propuesto por Pearson (1901), y de forma independiente tam-

bién por Hotelling (1933), y que consiste en describir la variación producida por la observación de p variables aleatorias x_i en términos de un conjunto de nuevas variables incorrelacionadas entre sí (denominadas Componentes Principales), cada una de las cuales sea combinación lineal de las variables originales.

Estas nuevas variables son obtenidas en orden de importancia, de manera que la primera componente principal incorpora la mayor cantidad posible de variación debida a las variables originales; la segunda componente principal se elige de forma que explique la mayor cantidad posible de variación que resta sin explicar por la primera componente principal, sujeta a la condición de ser incorrelacionada con la primera componente principal, y así sucesivamente.

El propósito del Análisis de Componentes Principales es ver si las dos o tres primeras componentes principales reúnen ya la mayor parte de la variación producida por las p variables originales puesto que, de ser así, considerando sólo estas dos o tres primeras, reduciremos la dimensionalidad de los datos al considerar únicamente dos o tres variables en lugar de p , y apenas perderemos información relevante.

Un ejemplo muy simple puede ser el de querer resumir la información obtenida al observar en los individuos de la muestra las variables Sueldo Anual, Rendimientos Patrimoniales, etc.; las primeras componentes principales nos darán un resumen de las variables observadas (en forma de combinaciones lineales de variables originales) y, al ser incorreladas, discriminarán mejor a los individuos de la muestra en razón de sus características económicas.

En algunos casos, la obtención de las componentes principales es el propósito del estudio, pero en muchas ocasiones suele ser el medio

de reducir el número de variables originales para realizar después un Análisis Estadístico determinado, como por ejemplo una Regresión Lineal Múltiple en la que originalmente habrían sido consideradas muchas covariables; esta segunda posibilidad es especialmente útil si las variables originales están fuertemente correladas. (De hecho, si éstas fueran incorreladas ellas serían ya las componentes principales.) No obstante, hacemos la observación de que, en Regresión, unas variables son independientes y otra (u otras) dependientes; en el Análisis de Componentes Principales todas las variables tienen la misma consideración.

PRUEBAS DE BONDAD DE AJUSTE [8]

Prueba de esfericidad de Bartlett

Esta prueba se utiliza para realizar el siguiente contraste

$$H_0 : \Sigma = \sigma^2 I$$

$$H_1 : \Sigma \neq \sigma^2 I$$

Donde la hipótesis nula afirma que las variables no están correlacionadas en la población, es decir, que la matriz de correlación de la población es una matriz diagonal.

Tiene sentido el modelo si podemos rechazar la hipótesis nula, lo cual sería indicativo de que existen correlaciones entre las variables. En caso de no poder rechazar la hipótesis nula, se pondrá en duda lo adecuado del modelo.

El determinante de la matriz de correlaciones muestral es un estimador del determinante de la matriz de correlaciones poblacional. A partir del valor del determinante muestral, se puede calcular un estadístico que se distribuye según la *chi-cuadrada* con 2 grados de libertad igual a $\frac{1}{2} (k^2 - k)$, donde k es el número de variables de la matriz de correlaciones. El valor del

estadístico que va a servir para contrastar las hipótesis es la prueba de Bartlett,

$$B = \left[n - 1 - \frac{1}{6(2k+5)} \right] \ln |R|$$

La condición de aplicabilidad de la prueba de Bartlett, es que las variables procedan de una población con una distribución normal multivariante.

Medida de adecuación de la muestra (kaiser-meyer-olkin) kmo

Este índice permite comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial. El índice KMO se calcula:

$$KMO = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n s_{ij}^2}$$

Donde:

r_{ij} : es el coeficiente de correlación entre las variables i -ésima y j -ésima; se excluyen de las sumatorias los coeficientes de correlación de una variable consigo, por lo tanto, el campo de aplicación de las sumatorias no es aplicable en los casos $i = j$.

s_{ij} : es el coeficiente de correlación parcial entre las variables i -ésima y j -ésima, también se excluyen los casos $i = j$.

Interpretación

Un índice **KMO** bajo indica que la intercorrelación entre las variables no es grande y, por tanto, la aplicación del modelo no será práctica, ya que necesitaríamos casi tantos factores como variables para incluir un porcentaje de la información aceptable.

- Un KMO con *valores mayores de 0.7* indican alta intercorrelación, por tanto, es indicativo de que éste análisis es apropiado.
- Los *valores entre 0.5 y 0.6* indican que el grado de intercorrelación es medio, pero aplicable.
- Cuando KMO *es menor que 0.5* indicaría que el análisis factorial no resultaría una técnica útil.

Medida de adecuación de la muestra (MSA)

Este índice se calcula para cada variable, de forma similar al índice KMO.

Su fórmula es $MSA = \frac{\sum_{i=1}^n r_{ij}^2}{\sum_{i=1}^n r_{ij}^2 + \sum_{j=1}^n s_{ij}^2}$ se excluye el caso $i = j$.

Interpretación

- Un MSA con *valores mayores de 0.7* indica alta intercorrelación, por tanto, es indicativo de que este análisis es apropiado.
- Los *valores entre 0.35 y 0.6* indican que el grado de intercorrelación es medio, pero aplicable.
- Un MSA *menor que 0.5* indica que no se aconseja este tipo de análisis.

Correlación múltiple

Este coeficiente indica el grado de asociación entre una variable y todas las otras que intervienen en el análisis. Si hay muchas variables con un coeficiente de correlación múltiple muy alto, el análisis puede utilizarse. Las variables con un coeficiente de correlación múltiple bajo podrían eliminarse.

La comunalidad

Es un valor que se obtiene para cada una de las variables originales, sumando los cuadrados de las correlaciones o cargas de los factores retenidos con la variable para la que se calcula y que expresa la proporción de varianza de la variable extraída o explicada con m factores, donde m es el número de factores retenidos. Si m es igual al número total de variables la comunalidad será igual a 1.

APLICACIÓN

Para presentar las metodologías de medición de riesgo de crédito, basados en modelos Logit y Probit y aplicarlas a la base de datos, que permitan mejorar el control, la toma de decisiones de la administración financiera y la gestión de los riesgos, es necesario realizar un análisis de comportamiento para cada una de las variables y la correlación entre ellas. Esto se hace con el fin de determinar algunas relaciones para grupos determinados de la población de acuerdo con sus características particulares.

Por tanto, se cruzan variables de cliente, del crédito y de comportamiento contra la definición de fallidos y no fallidos. Definición que es resultado del default para establecer un procedimiento de clasificación que permita determinar las ponderaciones necesarias y establecer la probabilidad de fallido. Dicho procedimiento estará basado en las técnicas del Análisis Estadístico Multivariado.

Un aspecto muy importante sobre el cual enfatiza el sistema SARC¹ es el seguimiento y control de procesos que tengan relación directa con el riesgo crediticio. Por lo tanto se hace necesario el monitoreo de procesos tales como: Otorgamientos, Comportamiento y Provisiones. Cada uno de estos procesos sintetiza las diferentes etapas de la vida de una obligación, razón por la cual las variables

contempladas en cada proceso deben tener relación directa con el objeto del mismo del crédito, así como el análisis y seguimiento a las mismas.

Desde hace más de dos años, en el país se esta observando cómo realizar el SARC (Sistema de Análisis de Riesgo de Crédito), el cual consiste en menor regulación y mayor supervisión; éste incluye un aumento en la capacitación tanto de las entidades vigiladas como de la propia Superintendencia Bancaria para aumentar su capacidad de evaluación. El objetivo principal de las entidades financieras es lograr por medio del Sistema de Administración de Riesgo Crediticio (SARC) estimar el nivel de provisiones adecuado para el crédito otorgado.

Existen 2 tipos de provisión:

a) Provisión Individual

Es la pérdida esperada de cada uno de los créditos, que se define como:

$$PE_t = P_t \times SE_t \times (1 - TR)$$

Donde:

P_t : Es la probabilidad de incumplimiento o probabilidad de Default, este es obtenido por medio de un modelo logístico:

$$P_t = \frac{e^{X\beta}}{1 + e^{X\beta}} \quad \text{donde} \quad X\beta = \beta_0 + \sum_{i=1}^m \beta_i X_i$$

SE_t Es el saldo expuesto del crédito t.

TR, Es la tasa de recuperación, la cual se obtiene a partir de los créditos que entran en cobro jurídico.

Para determinar en nivel de provisiones individuales anticipadas es necesario emplear el modelo de otorgamiento del crédito, en el cual el analista debe estimar la probabilidad de Default, según las características del crédito.

El Saldo

La probabilidad de incumplimiento

El plazo

Estableciendo
cual es la
provisión óptima
de cada crédito.

Cada entidad decide hasta que punto asume el nivel de riesgo, estableciendo rangos de probabilidad para calificar a cada cliente, de la siguiente manera:

Rango de probabilidad	Calificación
1	AAA
2	AA
3	A
4	BBB
5	BB
6	C
7	D
8	E

b) Provisión General

Se realiza por medio de modelos VAR o modelos ADL (modelos autorregresivos de rezagos distribuidos), tomando la cartera general y las variables internas de la entidad, realizando un detallado seguimiento, observando cómo las variables macroeconómicas afectan el cálculo en la provisión.

Se ha observado, a través de la historia, que en épocas de crecimiento económico se disminuyen las provisiones afectando positivamente las utilidades de las instituciones financieras. El problema es que en el momento de correr la serie con los datos históricos, ésta puede quedar sobrestimada por el pasado.

Para evitar esto se debe tener presente la Prociclicidad, la cual considera 2 factores:

1. En épocas de crisis se da un exceso de provisiones (exceso de morosidad), golpeando directamente las utilidades de

las instituciones (disminuye la utilidad neta), empeorando, así, la crisis más de lo que estaba (pues las instituciones financieras están castigando fuertemente sus utilidades, mostrando a la economía malos resultados).

2. Épocas de bonanza: Se dará un déficit (la entidad no realiza la cantidad necesaria de provisiones, pues según la teoría “cuando hay bonanza la morosidad disminuye”, pero en la práctica puede que no se haya reflejado realmente lo que proponía la teoría, ocasionándole grandes pérdidas a la institución, porque no estaba preparada para el no pago de sus clientes) y en el momento en que llega la crisis la organización se acaba, pues no la soporta.

Variables macroeconómicas que se deben considerar

En el momento de escoger las variables, un buen parámetro es la correlación que hay entre ellas; las más importantes son:

El PIB (producto interno bruto)

Entre el PIB vs. la calidad de la cartera hay una correlación perfecta (como se enunció anteriormente), pero el PIB posee una característica, pues éste es una variable que no la tenemos en tiempo real, pues esta se da trimestralmente, por lo tanto es buena para el pasado, mas no para el futuro. Pero a esta variable se le puede realizar un tratamiento, el cual consiste en el filtrado baso permitiendo que la serie no quede desfasada.

EL DESEMPLEO

Tasa de desempleo vs. provisiones: Según la historia se ha observado que a mayor tasa de empleo menor es la morosidad, por lo tanto, están correlacionadas negativamente.

CONSUMO

Consumo vs. provisiones

EL IPC (índice de precios al consumidor)

IPC vs. cartera vencida: Se ha observado que a menor inflación mayor nivel de morosidad, es decir, tiene una correlación negativa.

LA DTF (tasa de deposito a termino fijo)

DTF ajustado por inflación vs. cartera vencida: Se observa que a menor DTF menor nivel de morosidad.

Es necesario realizar un modelo de seguimiento en el cual se consideran variables dinámicas, permitiendo la actualización del modelo, debido a que la probabilidad de fallo cambia a través del tiempo, cambiando, a su vez, la provisión general, de la siguiente manera:

- Buen comportamiento: la provisión baja.
- Mal comportamiento: la provisión sube.

En este artículo se presenta el procedimiento de cálculo de la provisión individual y no el estudio para hallar la provisión general.

Procedimiento para hallar la provisión individual

Para determinar la provisión es necesario determinar el riesgo de crédito y clasificar la base de datos en 2 tipos de cartera:

- La cartera comercial
- La cartera de consumo

Los modelos econométricos mas usados para calcular la probabilidad de Default son:

- El Logit
- El Probit
- El tobit

El modelo econométrico que se presenta en este artículo es el logístico.

Análisis de la cartera comercial

Para realizar dicho análisis se debe realizar el siguiente procedimiento:

1. La unidad objeto de análisis son las empresas; está compuesto por:

A. Variables cuantitativas: Indicadores financieros. Las variables son:

Año Información, Crecimiento en Ventas, Rentabilidad Bruta, Utilidad operativa, Utilidad neta / Ventas, Crecimiento en Utilidades, Ciclo Operativo (rot. cart. + rot. inv.), Utilidad neta / Patrimonio, Crecimiento del Patrimonio, Endeudamiento sin Valz (Pas.T/Act. T), Endeudamiento Total / Ventas, Endeudamiento (Pas.T/Act.T), Endeudamiento Financiero / Ventas, Apalancamiento Financiero, Pasivo Corriente/Pasivo Total, Crecimiento del Activo, Rentabilidad del Activo, Prueba Acida (razón), Relación Corriente, Rotación de Cartera (Comercial), Rotación de Proveedores, Rotación de Inventarios, Otros Ingresos / Utilidad Neta, Corrección Monetaria / Utilidad Neta, Utilidad Operativa / Gastos Financiero, Flujo de Caja Libre / Gastos Financieros, Calificación Interna, Default, Calificación Pagos Externos, Cupo en Millones de Pesos, Año Calificación, Saldo Capital Promedio por Año y Máximos días vencido año; los cuales son establecidos por el PYG y el balance General.

Es necesario observar el nivel de correlación que existe entre dichas variables, debido a que

pueden existir problemas de multicolinealidad; para evitar estos problemas, realizaremos el análisis de componentes principales que nos permite pasar de un espacio linealmente dependiente (LD) a un espacio linealmente independiente (LI), permitiendo estimar el modelo logístico.

A. Variables cualitativas o categóricas: las cuales se caracterizan por poseer diferentes estados. Dichas variables son:

Antigüedad de la Empresa, Antigüedad como Cliente, Socios, Referencias de Socios / Administradores, Administración – Experiencia, Administración – Estructura, Estados Financieros – Auditor, Estados Financieros – Disponibilidad, Competencia, Productos, Número de Clientes / Estabilidad, Proveedores, Barreras de entrada, Barreras de salida, Instalaciones / Activos, Recursos Humanos, Nivel de regulación, Nivel de riesgo ambiental, Fuentes Financieras y Alternativa más probable de Financiación.

Se realizan pruebas de asociación entre las variables categóricas y el Default, por medio de tablas de contingencia.

Ajuste del modelo de regresión logístico

A continuación se realiza el scoring logístico, donde se consideró como variable endógena es el Default y como variables exógenas las 21 componentes principales y la variable cuantitativa calificación pagos externos; y que se expresa como una combinación lineal de 5 variables dummies

$$X\beta = \beta_0 + \beta_1cc_1 + \dots + \beta_{21}cc_{21} + \alpha_1Dx_{37}a + \alpha_2Dx_{37}b + \dots + \alpha_4Dx_{37}e$$

Variables in the Equation					
	B	S.E.	Wald	df	Sig.
CC1	0,781	0,329	5,622	1,000	0,018
CC2	0,400	0,437	0,837	1,000	0,360
CC3	-0,348	0,475	0,536	1,000	0,464
CC4	-1,462	2,455	0,355	1,000	0,551
CC5	-4,247	6,536	0,422	1,000	0,516
CC6	1,435	1,398	1,054	1,000	0,304
CC7	-0,848	2,024	0,176	1,000	0,675
CC8	0,058	0,425	0,019	1,000	0,891
CC9	-0,239	0,369	0,418	1,000	0,518
CC10	0,620	1,200	0,267	1,000	0,605
CC11	0,338	0,553	0,373	1,000	0,541
CC12	-0,470	0,364	1,666	1,000	0,197
CC13	-0,889	0,578	2,362	1,000	0,124
CC14	0,666	0,392	2,879	1,000	0,090
CC15	2,426	0,844	8,268	1,000	0,004
CC16	0,612	0,541	1,280	1,000	0,258
CC17	2,110	1,089	3,753	1,000	0,053
CC18	-0,656	0,249	6,940	1,000	0,008
CC19	-0,315	0,482	0,428	1,000	0,513
CC20	0,068	0,504	0,018	1,000	0,892
CC21	0,129	0,238	0,295	1,000	0,587
DX37_A	-182,679	1101,743	0,027	1,000	0,868
DX37_B	-147,075	1057,346	0,019	1,000	0,889
DX37_C	-143,210	1057,331	0,018	1,000	0,892
DX37_E	-104,843	21087,777	0,000	1,000	0,996
Constant	141,696	1057,324	0,018	1,000	0,893

Se debe eliminar del modelo la variable menos significativa, que en este caso es Dx37_E y se vuelve a correr el modelo, se procede así hasta que queden las variables más significativas. El modelo final es:

Variables in the Equation					
	B	S.E.	Wald	df	Sig.
CC1	0,9131	0,1954	21,8312	1,0000	0,0000
CC3	-0,7507	0,2201	11,6299	1,0000	0,0006
CC4	-4,8795	1,2269	15,8169	1,0000	0,0001
CC7	-0,3867	0,1528	6,4027	1,0000	0,0114
CC12	-0,4097	0,1647	6,1865	1,0000	0,0129
CC15	0,7978	0,1717	21,5886	1,0000	0,0000
CC18	-0,3942	0,1320	8,9200	1,0000	0,0028
CC19	-0,5767	0,1672	11,8900	1,0000	0,0006
DX37_B	-5,0232	0,4285	137,3922	1,0000	0,0000
DX37_A	-38,9191	326,7941	0,0142	1,0000	0,9052
Constant	0,1099	0,2370	0,2151	1,0000	0,6428

OBSERVACIÓN

La variable Dx37 A no fue sacada del modelo ya que a pesar de ser una variable no significativa, al sacarla del modelo ésta afecta el scoring de otorgamiento.

Y la clasificación del modelo es:

Classification Table(a)

Observed			Predicted		Percentage Correct
			Default	Fallido	
Step 1	Default	No Fallido	4005	70	98,3
		Fallido	7	243	97,2
	Overall Percentage				98,2

aThe cut value is ,100

Se observa que el modelo clasificó 4.005 clientes buenos como buenos clientes y 70 buenos como malos clientes, mostrándonos un buen ajuste.

Validación del modelo

Para validar el modelo, es necesario realizar diferentes pruebas estadísticas como:

1. Prueba Ómnibus

Esta prueba es útil para analizar la significancia de las componentes conjuntas del modelo, donde:

$$H_0: \beta_1 = \beta_3 = \beta_4 = \beta_7 = \beta_{12} = \beta_{15} = \beta_{18} = \beta_{19} = \alpha_1 = \alpha_2 = 0$$

H₂: Al menos un parámetro diferente de cero

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	1589,262	10	,000
Block	1589,262	10	,000
Model	1589,262	10	,000

Como el valor $p < 0.05$, se rechaza la hipótesis nula, significa que al menos una de las com-

ponentes es significativa en el modelo.

2. Prueba de Hosmer- Lemeshow

Es una prueba de bondad de ajuste, donde:

H₀: El modelo esta bien ajustado

H₁: Falta ajuste en el modelo

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	,396	8	1,000

El valor de Hosmer Lemeshow fue de 0.396 demostrando una buena calibración del modelo, además se observa que el valor $p > 0.05$, lo que lleva a aceptar la hipótesis nula.

3. Prueba KS (Kolmogorov- Smirnov)

Es un histograma acumulado de los buenos y malos, donde se muestra la máxima diferencia absoluta entre éstos, para determinar si el modelo global es significativo, donde:

H₀: La distribución de los buenos es equivalente a la distribución de los malos

H_1 : La distribución de los buenos no es equivalente a la distribución de los malos

Esta prueba utiliza los valores de cada distribución.

		Predicted probability
Most Extreme Differences	Absolute	,963
	Positive	,963
	Negative	,000
Colmogorov-Smirnov Z		14,773
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: Deafult

Se puede observar que se rechaza la hipótesis nula, debido a que el valor $p=0$, por lo tanto la distribución de los buenos difiere de la distribución de los malos, lo que nos demuestra que no hay errores en el modelo.¹

4. test Mann-Whitney

Este test consiste en sacar rangos en las muestras sucesivamente de menor a mayor a los datos de cada distribución.

H_0 : La distribución de los buenos es equivalente a la distribución de los malos

H_1 : La distribución de los buenos no es equivalente a la distribución de los malos

Mann-Whitney Test

Ranks				
Predicted probability	Deafult	N	Mean Rank	Sum of Ranks
	No Fallido	4075	2039,03	8309037,00
	Fallido	250	4183,75	1045938,00
	Total	4325		

Test Statistics ^a		Predicted probability
Mann-Whitney U		4187,000
Wilcoxon W		8309037,0
Z		-41,633
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: Deafult

Se puede observar que se rechaza la hipótesis nula, debido a que el valor $p=0$, por lo tanto la distribución de los buenos difiere significativamente de la distribución de los malos.

5. Prueba ROC

Esta prueba permite ver el criterio de especificidad y sensibilidad, para obtener el proceso de discriminación.

Para interpretar la curva ROC se analiza el área bajo la curva (AUC) que para una buena discriminación debe ser superior a 0.80.

Definiciones de sensibilidad y especificidad

Sensibilidad: es la probabilidad de que a los clientes malos, el modelo los clasifique como malos, bajo un punto de corte determinado.

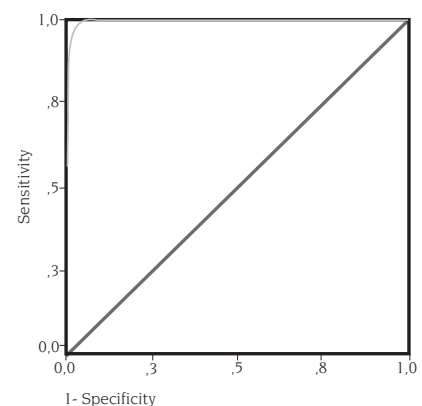
Especificidad: Es la probabilidad de que a los clientes buenos, el modelo los clasifique como malos, bajo un punto de corte determinado.

Criterio para clasificar los clientes

Si $P < \text{punto de corte}$, entonces el cliente es malo = 1

Si $P > \text{punto de corte}$, entonces el cliente es bueno = 0

$$P(M / B) \rightarrow 1 - \text{Especificidad}$$



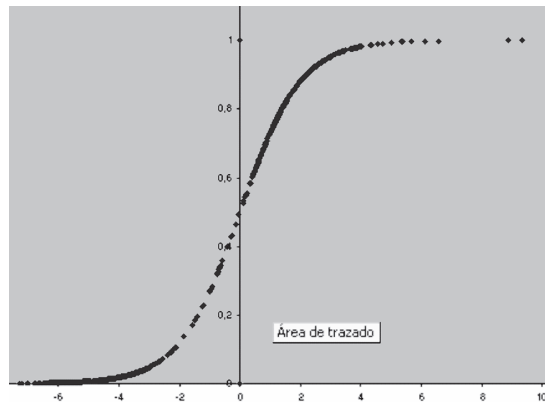
$$X\beta = \ln \left[\frac{P_t}{1 - P_t} \right]$$

Si $X\beta < 0$, entonces el cliente es bueno

Si $X\beta > 0$, entonces el cliente es malo

Para el análisis se realiza el gráfico entre

$X\beta$ vs P_t



Pérdida esperada o provisión individual

Como ya se explicó al inicio la provisión individual es calculada como la pérdida esperada¹ de cada uno de los créditos, que se define como

$$PE_t = P_t \times SE_t \times (1 - TR)$$

Donde:

P_t : Es la probabilidad de incumplimiento o probabilidad de Default; éste es obtenido por medio del modelo logístico,

$$P_t = \frac{e^{X\beta}}{1 + e^{X\beta}} \quad \text{donde} \quad X\beta = \sum_{i=1}^m \beta_i X_i$$

SE_t : es el saldo expuesto del crédito t.

TR : es la tasa de recuperación, la cual se obtiene a partir de los créditos que entran en cobro jurídico.

TABLA DE PROVISIONES

		Count	Maximum	Mean	Sum
Año	1994	6	66,28	11,05	66,28
Informacion	1995	45	83,79	10,23	460,28
- Empresa	1996	105	148,46	4,05	425,03
	1997	218	235,93	7,03	1531,70
	1998	831	1307,18	5,53	4596,83
	1999	855	472,97	4,75	4056,99
	2000	821	427,28	9,69	7923,56
	2001	687	384,14	5,25	3594,99
	2002	763	336,12	1,48	1132,63

Como se observa en la tabla de provisiones, el mayor nivel de provisión media se observa en el año 2000; a pesar de que los clientes disminuyeron el nivel de provisión se incrementó posiblemente a la crisis financiera del país.

En 1994 la mayor provisión fue de 66.28 debido a la alta probabilidad de incumplimiento de pago, de dicho cliente, lo que nos muestra la importancia de que la institución financiera establezca límites de otorgamientos de créditos, ya que no es rentable que la entidad tenga tanto capital improductivo.

CONCLUSIONES

Las propuestas contenidas en el documento de nuevo acuerdo de capital representan un sólido esfuerzo por incorporar los desarrollos recientes de la teoría financiera de cuantificación del riesgo en la determinación del capital regulatorio mínimo que debe ser exigido a las entidades financieras.

El propósito en el nuevo acuerdo de Basilea es el de no incrementar los requerimientos agregados de capital para la banca que utilice el sistema estandarizado de medición del riesgo de crédito, y de estimular la migración hacia la utilización del sistema basado en los rating internos. Frente a lo anterior, varios analistas han señalado que los requerimientos de capital van a aumentar sustancialmente para la generalidad de los bancos, aún en

países de altos ingresos. Los bancos que están en capacidad en el corto plazo de presentar sistemas de rating internos que satisfagan las exigencias de los reguladores, y que podrían por lo tanto beneficiarse de las reducciones en los requerimientos de capital, conforman un grupo relativamente pequeño de entidades de países de altos ingresos.

Por otra parte, la complejidad de los sistemas de medición de riesgo, aún para el caso del enfoque estandarizado, exigirá un esfuerzo notable de los entes supervisores, por mejorar sus capacidades técnicas. Estos esfuerzos serán considerablemente superiores cuando se trate de estudiar la consistencia de los sistemas de rating internos de los bancos.

Dado lo anterior, la generalidad de los bancos de estos países va a verse obligada a utilizar el enfoque estandarizado, que depende de la calificación del riesgo emitida por las agencias especializadas. Sin embargo, estas calificaciones son también menos comunes en los países de bajos ingresos. Se corre el riesgo de que los esfuerzos por mejorar la calidad de los activos no se traduzca en menores requerimientos de capital para sus bancos, contrario a lo sucedería en los países de altos ingresos, donde está más generalizada la práctica de la calificación.

Particular preocupación se deriva de la situación que se puede presentar en relación con los requerimientos de capital de la deuda soberana de los países en vía de desarrollo. Frente a una ponderación actual de estos papeles del 0%, se pasaría a una ponderación por riesgo del 100% (caso de la deuda externa de la República de Colombia). El efecto que ello tendría en el encarecimiento de la deuda soberana podría traducirse en una disminución del flujo de capitales hacia estos países, con las obvias consecuencias desde el punto de vista del crecimiento y del empleo.

Las anteriores consideraciones hacen aconsejable que se evalúen cuidadosamente

las condiciones de aplicación del Nuevo Acuerdo de Capital, a fin de que no vaya él a provocar mayores dificultades en la situación de las naciones, los bancos y las empresas de los países en vía de desarrollo. La meta propuesta de implementación del nuevo acuerdo no parece realista, tanto desde el punto de vista de la preparación técnica de las entidades bancarias y de los entes supervisores, como de la posibilidad de atender los importantes requerimientos de capital que se derivarían de esa implementación.

El riesgo, como la probabilidad de obtener un resultado no esperado, hace necesario que en su estudio se incluyan las matemáticas como herramienta para modelar los procesos de optimización, dado que las características de las variables se ven reflejadas mediante procesos estocásticos; y para ello se deben considerar como pilares: Estadística multivariada, Procesos Estocásticos, Simulación, Series de Tiempo y Econometría.

El pilar fundamental de la supervisión bancaria está en la gestión interna del riesgo en las entidades financieras y, como tal, está orientado a que se fortalezca esta gestión en riesgos de crédito y de mercado, generando una cultura y prácticas de alto nivel técnico en la evaluación, administración y control del mismo.

Se deben establecer los principios y criterios generales que las entidades adoptarán para mantener adecuadamente evaluados los riesgos crediticios implícitos en los activos de crédito, se deben definir las modalidades de crédito, determinar las calificaciones otorgadas a tales operaciones según la percepción de riesgo que de las mismas se tenga, establecer la periodicidad con que se deben efectuar tales calificaciones y consagrar los mecanismos de recalificación.

Los score de créditos son construidos para predecir el comportamiento de pagos de un

deudor usando diversa información (centrales de riesgo) tal como la historia de pagos del deudor, nivel de endeudamiento o utilización, así como de acuerdo con el apetito de exposición de la entidad. Algunos score incorporan los ingresos, capacidad de pago y alguna otra información relacionada con los productos ofrecidos.

Al emplear este método es necesario establecer ciertos elementos que son importantes para determinar la probabilidad de no pago de un cliente.

- Identificar las variables con las cuales se seleccionan los sujetos de crédito de cada portafolio.
- Peso de las variables dentro de la metodología de calificación.
- Ordenamiento de los sujetos de crédito dentro del rango de calificación. Calificación en categorías de riesgo.
- Selección del sujeto de crédito conforme al nivel de riesgo aceptable para la entidad.

La probabilidad que estima el Scoring usualmente no está medida a un horizonte temporal fijo. Éstas se ajustan a la definición de **mal pagador** establecida en el desarrollo del modelo y que puede incluir no sólo mora sino también retrasos en el pago de cuotas y otros aspectos que definen al cliente como **mal pagador**.

En los modelos de Scoring generalmente la variable respuesta o dependiente es de carácter dicotómico, mientras que las demás variables explicativas pueden ser continuas, categóricas (convertidas a dummies) y/o dicotómicas. Para ajustar este tipo de modelos, los principales métodos estadísticos usados son las regresiones lineales, Logit o Probit.

El modelo de probabilidad es atractivo porque es mas fácil de entender, pues la relación entre las variables explicativas y la variable estimada es directa (lineal), pero su desventaja consiste en que puede arrojar probabilidades negativas y/o mayores a 1 (las cuales conceptualmente no tiene sentido).

El modelo Logit se basa en la función logística para obtener la probabilidad, mientras que el Probit lo hace por medio de la distribución normal acumulada. La regresión logística puede ser una herramienta estadística más apropiada para estimar la probabilidad de ocurrencia de un evento en nuestro caso el default del cliente, dado que se han definido dos clases en la variable respuesta.

Aunque el documento de Basilea no obliga a las entidades financieras a usar un modelo específico en la evaluación del riesgo crediticio, establece algunos parámetros que sirven como guía para la construcción del modelo de evaluación y control del riesgo de crédito.

En Colombia, la Asobancaria propone algunas metodologías que se están implementando en algunas entidades financieras, pero en el nivel de cooperativas de ahorro y crédito las metodologías econométricas son las más aplicadas; es posible que se requiera más tiempo para poder evaluar las bondades de dichas metodologías, dadas las diferencias que se presentan en algunas de ellas, tales como Logit y Probit.

Las metodologías seleccionadas mostraron sus fortalezas y debilidades al evaluar el riesgo crediticio de la base de datos, pero, dados los resultados, consideramos necesario seguir investigando y comparando con otras metodologías, con el fin obtener parámetros de medición de eficiencia que arrojen de manera más segura niveles comparación para el Logit y el Probit.

BIBLIOGRAFIA

1. ARAGONÉS, José Ramón BLANCO, Carlos. Valor en Riesgo. Aplicación a l Gestión empresarial. Ediciones Pirámide. Madrid. 2000.
2. BEST, P. Implementing Value at Risk. January. 1999.
3. BÜHLMANN, Hans. *Mathematical Methods in Risk Theory*. Springer – Verlag. 1970.
4. DOWD, Kevind. Beyond Value at Risk: The new Science of Risk Management. 2002.
5. D, HENDRICKS. Evaluation of VAR models using historical data, Economy Policy Review, Federal Reserve Bank of New York, Abril, Vol.2, número1.
6. GREENE, William H. Análisis econométrico. Prentice Hall. Madrid.1999.
7. GUJARATI, Damodar N. Econometría. McGraw-Hill.3ª Edición. Bogotá.1997.
8. HAIR, Joseph F., ANDERSON, Rolph E., TATHAM, Ronald L. y BLACK, William C. Análisis multivariante. Prentice Hall. Madrid. 5ª Edición. 1999.
9. HILLIER, F. and LIEBERMAN, G. F. (1986). *Introduction to Operations Research*. 4th ed. San Francisco, Holden-Day.
10. HOLTON, Glyn A. Value at Risk: Theory and practice. March. 2003.
11. JORION, Philippe. Valor en Riesgo. El nuevo paradigma para el control de riesgos con derivados. Editorial Limusa. 2002.
12. LAMBERTON, Damien. Introduction to Stochastic Calculus Applied to Finance. CRC Press. 2003.
13. LATORRE LLORENS, Luis. Teoría del Riesgo y sus Aplicaciones a la Empresa Aseguradora. Editorial Mapfre. 1992.
14. LEFÈVRE, Claude; UTEV, Sergey. Comparison of individual risk models. *Insurance: Mathematics & Economics* 28 (2001) 21-30.
15. MADDALA, G.S. Introduction to Econometrics. Macmillan. Nueva York. 1988.
16. MITTELHAMMER, Ron C., JUDGE, George G. y MILLER, Douglas J. Econometric Foundations. Cambridge University Press. New York. 2000.
17. MONTGOMERY, Douglas C., PECK, Elizabeth A. y VINNING, G. Geoffrey. Introducción al análisis de regresión Lineal. 3ª Edición. CECSA. 2002.
18. PENZA, Pietro anf BANSAL, Vipul K. Measuring Market Risk with Value at Risk. October 2000.



19. PINDYCK, Robert S. y RUBINFELD, Daniel L. Econometría, modelos y pronósticos. McGraw-Hill. 4ª Edición. 1998.
20. RIPLEY, B. D. *Stochastic Simulation*. Wiley, New York. 1987

APÉNDICE

Estimación máximo verosímil de los parámetros del modelo logit

La estimación de los parámetros $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ de un modelo de regresión logística se efectúa por medio del método de estimación por máxima verosimilitud. Según dicho método se obtienen los estimadores máximo verosímiles como funciones de la muestra que hacen que se maximice la función de verosimilitud asociada a la muestra. Denotando por

$L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k)$ a la función de verosimilitud asociada a una muestra de tamaño N , para un modelo de regresión logística con parámetros $\beta_0, \beta_1, \dots, \beta_k$, con una variable Y dicotómica, se tiene que:

$$L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k) = \prod_{i=1}^N P(Y=1|x^{(i)})^{y^{(i)}} (1-P(Y=1|x^{(i)}))^{1-y^{(i)}}$$

Por otra parte, teniendo en cuenta que $\ln z$ es una función creciente estrictamente, y por tanto el valor de los parámetros $\beta_0, \beta_1, \dots, \beta_k$ maximizando:

$$L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k)$$

$$\ln L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k)$$

Desarrollando el logaritmo natural de la función de verosimilitud obtenemos:

$$\begin{aligned} & \ln L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k) \\ &= \sum_{j=1}^N y^{(j)} \ln P(Y=1|x^{(j)}) + \sum_{j=1}^N (1-y^{(j)}) \ln (1-P(Y=1|x^{(j)})) \end{aligned}$$

$$\begin{aligned} &= \sum_{j=1}^N y^{(j)} [\ln P(Y=1|x^{(j)}) - \ln (1-P(Y=1|x^{(j)}))] + \sum_{j=1}^N \ln (1-P(Y=1|x^{(j)})) \\ &= \sum_{j=1}^N y^{(j)} \ln \frac{P(Y=1|x^{(j)})}{1-P(Y=1|x^{(j)})} + \sum_{j=1}^N \ln (1-P(Y=1|x^{(j)})) \end{aligned}$$

Teniendo en cuenta que

$$\ln \frac{P(Y=1|x^{(j)})}{1-P(Y=1|x^{(j)})} = \beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}$$

Y que

$$(1-P(Y=1|x^{(j)})) = \frac{\exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)\right]}{1 + \exp\left[-\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)\right]} = \frac{1}{1 + \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)}$$

Obtenemos

$$\begin{aligned} & \ln L((x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)}), \beta_0, \beta_1, \dots, \beta_k) \\ &= \sum_{j=1}^N y^{(j)} \left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)} \right) - \sum_{j=1}^N \ln \left(1 + \exp \left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)} \right) \right) \end{aligned}$$

Los estimadores máximos verosímiles $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ para los parámetros $\beta_0, \beta_1, \dots, \beta_k$ se van a obtener al resolver el siguiente sistema de $n+1$ ecuaciones y $n+1$ incógnitas:

$$\frac{\partial \ln L}{\partial \beta_0} = \sum_{j=1}^N y^{(j)} - \sum_{j=1}^N \frac{\exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)} = 0$$

$$\frac{\partial \ln L}{\partial \beta_1} = \sum_{j=1}^N y^{(j)} x_1^{(j)} - \sum_{j=1}^N x_1^{(j)} \frac{\exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)} = 0$$

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{j=1}^N y^{(j)} x_k^{(j)} - \sum_{j=1}^N x_k^{(j)} \frac{\exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i^{(j)}\right)} = 0$$

En el anterior sistema de $n+1$ ecuaciones y $n+1$ incógnitas no es posible obtener una fórmula cerrada para los estimadores de los parámetros $\beta_0, \beta_1, \dots, \beta_k$, de ahí que lo habitual sea utilizar técnicas iterativas para llevar a cabo dichas estimaciones. Al utilizar el método de Newton-Raphson para llevar a cabo dichas

iteraciones, se obtiene la siguiente fórmula de actualización de los parámetros:

$$\hat{\beta}^{\text{nuevo}} = \hat{\beta}^{\text{viejo}} + (X'WX)^{-1} X' (c - \hat{p})$$

Donde:

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$$

X: Matriz cuyas filas son $x^{(j)}, j=1,2,\dots,N$. Es decir $X \in M(N, n)$

W: Matriz diagonal con elementos

$$p^{(j)}(1-p^{(j)}), j=1,2,\dots,N. W \in M(N, n)$$

$$W = \begin{pmatrix} p^{(1)}(1-p^{(1)}) & \dots & \dots & 0 \\ 0 & p^{(2)}(1-p^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & p^{(N)}(1-p^{(N)}) \end{pmatrix}$$

\hat{p} : Vector cuya componente j-ésima indica la probabilidad estimada en esa iteración.

$$\text{Es decir, } \hat{p} \in M(N, 1), \text{ con } p^{(j)} = \frac{\exp(x^{(j)}\hat{\beta}^{\text{viejo}})}{1 + \exp(x^{(j)}\hat{\beta}^{\text{viejo}})}$$

y: Vector de componentes $y^{(j)}, j=1,2,\dots,N$. Por tanto $c \in M(N, 1)$

Los criterios de convergencia del método iterativo utilizado para estimar los parámetros pueden ser varios, pero en todos ellos la idea subyacente es que bien

$$\hat{\beta}^{\text{nuevo}} \approx \hat{\beta}^{\text{viejo}} \text{ o de igual manera } \ln L(\hat{\beta}^{\text{nuevo}}) \approx \ln L(\hat{\beta}^{\text{viejo}})$$

Recibido: 05/04/2005

Aceptado: 12/05/2005