

Aplicación de árboles de decisión en modelos de riesgo crediticio

PAOLA ANDREA CARDONA HERNÁNDEZ *

Resumen

En este artículo se presentan algunos puntos generales del marco teórico de los riesgos a los que se enfrenta una institución financiera, su clasificación y definición, centrándose específicamente en el riesgo crediticio, para el que se presenta el marco legal: los enunciados básicos del Acuerdo de Basilea II y la reglamentación del sistema de administración de riesgo crediticio de la Superintendencia Bancaria en Colombia. Dentro de este marco, se ilustrará cómo la estadística juega un papel importante en el cumplimiento de esta normatividad. Específicamente se presenta la utilización de los árboles de decisión como herramienta para el cálculo de probabilidades de incumplimiento en crédito, mostrando sus ventajas y desventajas.

Palabras Claves: Teoría de riesgo, Riesgo crediticio, árboles de decisión.

Abstract

This paper presents some general concepts in risk theory, and especially in credit risk, where statistics plays a central role. The text emphasizes on decision trees as a very usefull tool in the calculus of probabilities related to credit risk.

Keywords: Risk theory, Credit risk, Decision trees.

*Estadística. Profesional de Modelos, Análisis de Portafolio, Banco Colpatria Red Multi-banca. Email: cardonp@banco.colpatria.com.co.

1. Introducción

Por esencia, la actividad de una entidad financiera es la toma de riesgos, cada una de sus operaciones contiene implícita o explícitamente la incertidumbre. Desde las distintas dimensiones de su operación está expuesta a diferentes tipos de riesgo que deben ser identificados, medidos y controlados, como base para fijar las estrategias de mercadeo y especialmente de precios, que resulten en una ecuación favorable entre el riesgo asumido y la recompensa obtenida, medida como la rentabilidad neta del negocio.

Esta importante relación ha hecho que en los últimos años las instituciones financieras, hayan dado un vuelco en su forma de gestionar el riesgo, al evolucionar de procesos empíricos a metodologías apoyadas en procesos estadísticos.

Técnicamente la valoración del riesgo mide el grado de variación de los resultados financieros de una empresa frente a los estimados. Sin importar que sean a favor o en contra, entre más volátiles sean los resultados, más riesgo verá un accionista, pues no tiene la certeza de obtener los resultados presupuestados.

Muchos elementos pueden incidir en los resultados de una empresa, derivados de su gestión de riesgo: los criterios de aceptación de los créditos, la posterior administración, el estudio de la tendencia de los portafolios y los virajes de política con base en ellos, entre muchos otros. El resultado estará influenciado tanto por factores internos de la empresa, como por factores externos que pueden provenir de los mercados financieros, como tasas de interés, tasas de cambio o precios, o de variables macroeconómicas como inflación, tasa de desempleo, etc.

Un buen entendimiento de los conceptos de riesgo, permite a los administradores planear adecuadamente la forma de anticiparse a posibles resultados adversos y sus consecuencias y, de este modo, prepararse para enfrentar la incertidumbre sobre las variables que puedan afectar dichos resultados.

Dentro del análisis de riesgo se deben identificar las principales fuentes de exposición, desde el enfoque de los activos financieros se tiene la información presentada en la Figura 1:

2. Definiciones

Se definen los tipos de riesgo mencionados anteriormente¹:

¹ Capítulo VI de la Circular Externa 100 de 1995 - Circular Básica Financiera, Contable y de Remisión de información de la Superintendencia Bancaria. Circular Externa 011 de 2002.



Gráfica 1: Riesgo en las operaciones activas del balance.

Riesgo de liquidez: es la posibilidad de que una institución financiera no pueda cumplir un compromiso financiero con un cliente o mercado en algún lugar, moneda o momento determinado. Las exposiciones de liquidez pueden surgir de las actividades de “trading” relacionados con productos o mercados específicos; y de las actividades relacionadas con la obtención de fondos para la entidad. El primero es un riesgo en el cual no se puede deshacer una posición determinada al precio o cerca del precio del mercado, debido a la poca profundidad o desórdenes del mismo. El segundo es el riesgo de no poder cumplir con las obligaciones de pago en la fecha de cancelación y surgen principalmente del desfase de activos, pasivos, y madurez de contratos de los compromisos adquiridos.

Riesgo legal: es la contingencia de pérdida derivada de situaciones de orden legal, que puedan afectar la titularidad de las inversiones o la efectiva recuperación de su valor, que los contratos no sean legalmente aplicables o no estén correctamente documentados, ó variaciones de orden normativo que impliquen cambio en las reglas de juego impuestas por el regulador.

Modifica el Capítulo II de la Circular Externa No. 100 de 1995, Referente a la gestión del riesgo de crédito. Resolución 200/1995 - Capítulo V, Evaluación de los riesgos de solvencia y jurídicos. Circular Externa 100/1995, Capítulo XIII, Superintendencia Bancaria de Colombia.

Riesgo operativo: es la posibilidad de pérdida como resultado de deficiencias a causa de fallas en los sistemas de información, fallas en procesos, control de atribuciones y cupos en tiempo real, errores humanos, mala fe de los funcionarios y fallas en el control gerencial.

Riesgo de mercado: el riesgo de mercado o precio es la contingencia de pérdida o ganancia de una posición de la entidad financiera, como resultado de un cambio en el nivel o la volatilidad de las tasas de interés, tasas de cambio o precios.

Riesgo de contraparte: es la posibilidad de incumplimiento de las obligaciones contractuales entre la entidad financiera y el sector real o financiero. Se genera entre otros por incumplimiento en la entrega de documentos, ignorancia de los acuerdos de neteo para la liquidación de los contratos de derivados y problemas operativos o resolución de los mismos.

Riesgo de crédito: es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos como consecuencia de que sus deudores fallen en el cumplimiento oportuno o cumplan imperfectamente los términos contractuales acordados.

Cada uno de estos tipos de riesgo tiene sus métodos y formas de medición, algunos estadísticos, como el caso de riesgo de mercado en el que se usa la metodología *value at risk-var* (Jorion 1997), o riesgo de crédito con el uso de modelos estadísticos para predecir probabilidades de incumplimiento. En otros casos como el riesgo legal, la forma de controlarlos es más de procedimientos y gestión gerencial y el riesgo de contraparte en el que se hace un exhaustivo análisis financiero.

Este documento se centra en los modelos para la medición y control del riesgo de crédito.

3. Riesgo de crédito

La necesidad de medir el riesgo y promover que las instituciones financieras hagan una correcta evaluación de ellos ha sido un esfuerzo de todos los bancos a nivel mundial. El comité de supervisión bancaria de Basilea², ha sido pre-

² Establecido por los bancos centrales del grupo de los 10, a finales de 1974 cuyos miembros son: Bélgica, Canadá, Francia, Alemania, Italia, Japón, Luxemburgo, Holanda, España,

cursor de la reglamentación de la medición integral de riesgos y el adecuado provisionamiento de capitales, para sobrellevar los posibles riesgos incurridos y evitar la quiebra de las instituciones financieras. Este comité no posee una autoridad de supervisión supranacional y sus conclusiones no son mandatorias, sin embargo formula estándares de supervisión internacional y da guías de las mejores prácticas del mercado. Se aspira que todos los bancos activos internacionalmente apliquen las nuevas recomendaciones consignadas en el acuerdo de capitales (Basilea I – 1988) y el nuevo acuerdo de capitales (Basilea II–1999) que definen el capital mínimo que deben tener las entidades financieras para poder operar, relacionando el riesgo de los activos con el nivel y calidad del patrimonio, además, determina el capital mínimo necesario para que un banco haga frente a posibles quebrantos debido a los riesgos que maneja.

Así mismo, siguiendo la tendencia mundial, en Colombia la Superintendencia Bancaria reglamenta la creación del sistema de administración de riesgo crediticio SARC, para las entidades financieras bajo su vigilancia. Mediante la Carta Circular 31 de 2002 y la Circular externa 11 de 2002, se definen los lineamientos básicos que las entidades colombianas deben implementar para la medición de riesgos. Así, el sistema de administración del riesgo crediticio (SARC) en una entidad financiera debe contar al menos con los siguientes componentes básicos: unas políticas claras de administración de riesgos, una estructura organizacional adecuada, unas metodologías y procesos apropiados para la gestión de riesgos, una infraestructura y capital humano idóneos, así como un proceso de auditoría general (Carta circular 31 de 2002).

Dentro de este proceso de medición de riesgo crediticio, se contempla la utilización de una metodología interna para el cálculo de la probabilidad de deterioro de cada crédito y la estimación o cuantificación de la pérdida esperada si el deudor llegara a incumplimiento.

4. Pérdida esperada y probabilidad de incumplimiento

Ya se definió el riesgo de crédito como la pérdida potencial debida a la incapacidad de la contraparte de cumplir sus obligaciones, esto lleva a la necesidad de cuantificar dicha pérdida. La pérdida esperada puede definirse entonces como el monto de capital que podría perder una institución como resultado de la exposición crediticia en un horizonte de tiempo dado (Wilson & Press 1978). De este modo, una definición importante para calcular la pérdida esperada es la de

Suecia, Suiza, Reino Unido y Estados Unidos.

incumplimiento y basada en ésta, calcular la probabilidad que un cliente tenga esta condición. Se define el *incumplimiento* de un cliente cuando este alcanza una altura de mora m , en la que el banco asume la pérdida del capital.

Así, la pérdida esperada se calcula como:

$$PE = PI * S * E, \quad \text{donde,}$$

- PI : Probabilidad de incumplimiento para el período dado.
- S : Severidad: es el porcentaje de la cantidad expuesta que la entidad pierde si los deudores incumplen sus obligaciones.
- E : Exposición: es el monto de los activos expuestos al incumplimiento en el período definido.

Usando esta definición, las pérdidas esperadas aumentan en función del monto del crédito o exposición crediticia y de la probabilidad de deterioro asignada a cada activo. Las pérdidas serán menores cuanto más baja sea la severidad. Las provisiones de capital se basan en el resultado obtenido de esta ecuación y son precisamente las reglamentadas por la Superintendencia Bancaria en las circulares mencionadas anteriormente.

Para el cálculo de la probabilidad de incumplimiento, existen diferentes metodologías estadísticas con las que es posible predecir, según las características del cliente, la probabilidad de llegar a la altura de mora definida previamente como incumplimiento, en un período dado. Entre estas metodologías están análisis discriminante, regresión logística, análisis probit, redes neuronales, matrices de transición y árboles de decisión, que son las más comúnmente usadas en el sector financiero.

En este artículo se presenta específicamente el uso de árboles de decisión como herramienta para el cálculo de probabilidades de incumplimiento, por cuanto representa un método efectivo para la estimación, al igual que la mayoría de los métodos alternativos, pero ofrece una ventaja fundamental al ser un método de fácil entendimiento para personas que no cuentan con conocimientos avanzados de estadística. Un mismo modelo permite hacer diferentes usos, como mantenimiento de clientes considerados como buenos (probabilidades bajas de incumplimiento), cobranza proactiva y discriminada por nivel de riesgo para los clientes considerados como malos o con probabilidades altas de llegar a incumplimiento.

5. Modelos estadísticos para predecir probabilidades de incumplimiento

Se definen dos tipos de modelos para predecir la probabilidad de incumplimiento: el primero de iniciación, con el que se controla el primer ciclo de crédito, el de otorgamiento. Así, usando las variables de la solicitud de crédito se trata de predecir, para cada perfil o segmento identificado, la probabilidad de que un año después del otorgamiento del crédito, el cliente alcance la altura de mora definida como incumplimiento.

El segundo tipo de modelo es el de comportamiento, con el que se controla el proceso durante la maduración del crédito. Este tipo de modelo se calcula en el mes t y se basa en el comportamiento de pagos del cliente x meses atrás desde el tiempo t , ($t - x$, usualmente $x = 12$) para determinar su probabilidad de deterioro y que llegue a incumplimiento un año adelante ($t + 12$).

Se define entonces la variable respuesta Y para cualquier tipo de modelo:

- 1 Si el cliente ha alcanzado altura de mora en el período $t + 12$,
- 0 Si el cliente no ha alcanzado altura de mora M en el período $t + 12$.

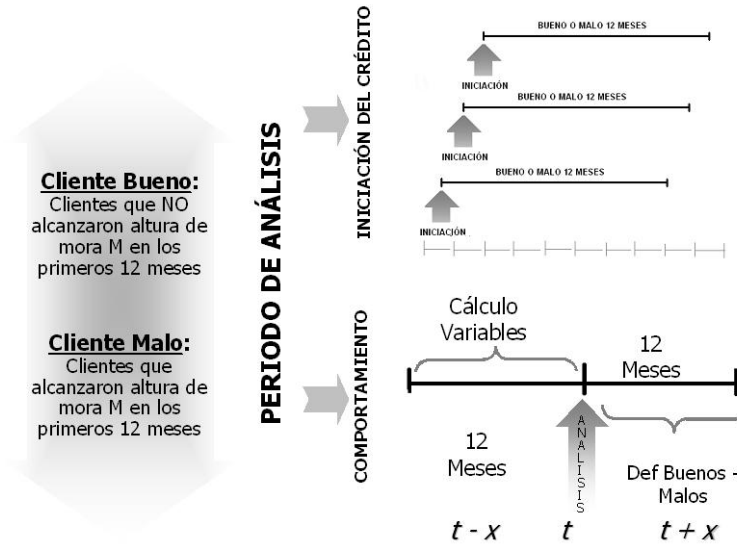
en el caso de iniciación $t = 0$.

Se debe aclarar que la definición de los períodos de tiempo evaluados no es estándar pero sí es la más comúnmente utilizada.

Los modelos son calculados por tipo de producto dentro del portafolio de crédito de la entidad financiera, supeditado a la información histórica disponible y a su calidad.

Así, se usan árboles de decisión binarios, método no paramétrico que no requiere supuestos distribucionales, permite detectar interacciones, modela relaciones no lineales y no es sensible a la presencia de datos faltantes y outliers ((Breiman, Friedman, Olshen & Stone 1984), (Kass 1980)). Su principio básico es generar particiones recursivas por reglas de clasificación hasta llegar a una clasificación final, tal que es posible identificar perfiles (nodos terminales) en los que la proporción de clientes malos es muy alta (o baja) y de esta forma asignar su probabilidad.

Para los árboles de decisión, existen diferentes metodologías, entre ellas Cart, Chaid, Chaid exhaustivo, Quest, C4.5, que difieren en la forma de asignación, las reglas de partición y los criterios de parada. Cualquiera de ellos generará n nodos terminales y una escala de probabilidades con n posibles valores, que es el resultado final y la probabilidad empleada tanto para validar



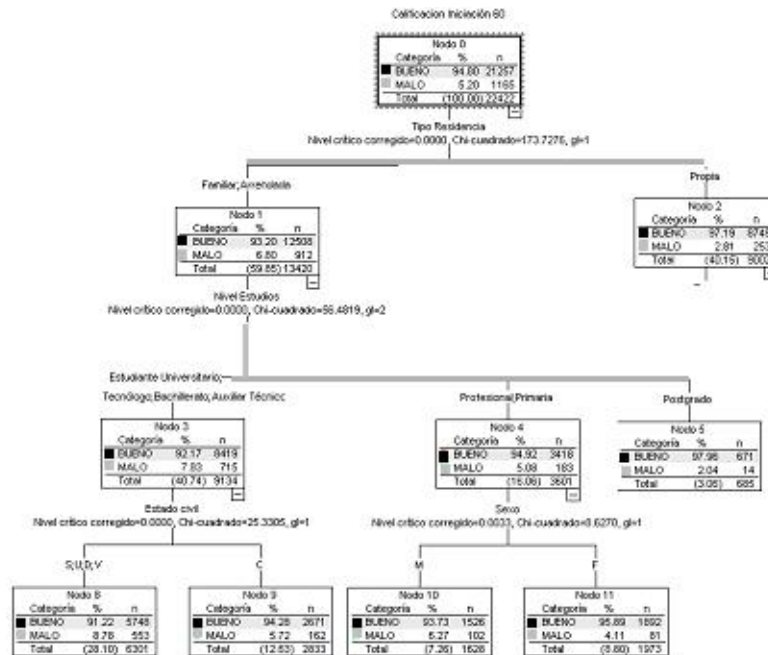
Gráfica 2: Esquema de la definición de variables.

el modelo, como para el cálculo de la pérdida esperada, mencionada anteriormente.

El siguiente gráfico muestra un ejemplo de árbol de decisión para el caso de modelos de iniciación, con 6 nodos terminales, es decir una escala de probabilidades de 6 categorías (cada probabilidad asociada a un nodo), que finalmente permite identificar 6 perfiles de riesgo, con los que se tomarían decisiones de otorgamiento de crédito.

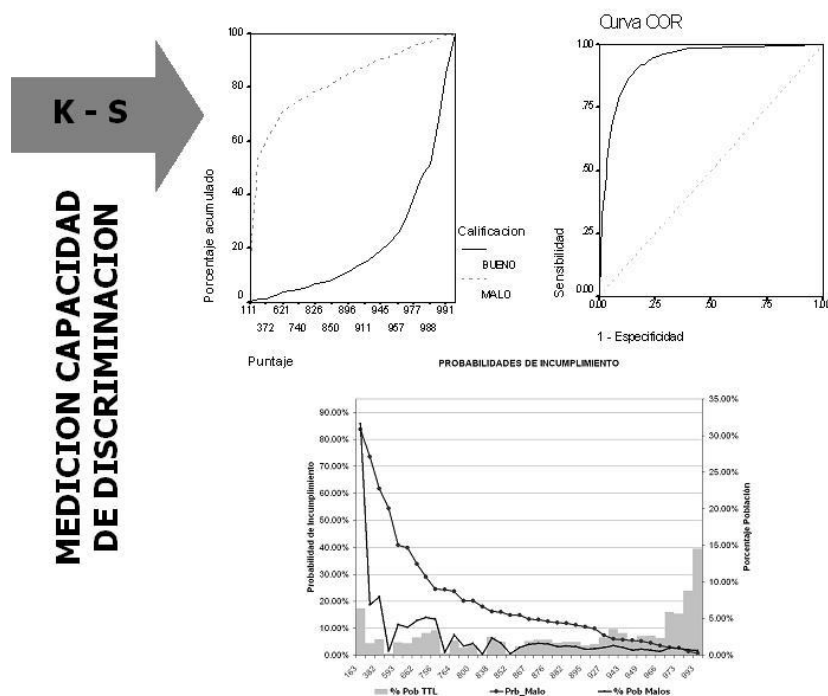
Las premisas básicas para la construcción del modelo son la simplicidad, potencia y estabilidad. La simplicidad es un factor relevante, en cuánto es muy importante para una entidad que cualquier persona de diferentes instancias, sea capaz de comprender y entender por qué el modelo funciona y qué es lo que está prediciendo. La potencia, medida en su capacidad de discriminar correctamente los clientes buenos, de los clientes malos. Y estabilidad, que el modelo a través del tiempo conserve su capacidad de discriminación y pueda detectar los cambios en la calidad de la cartera.

Algunas medidas de capacidad de discriminación empleadas son la prueba de bondad de ajuste de Kolmogorov-Smirnov para 2 muestras ($K - S$), la curva ROC (Receive Operative Curve), el coeficiente Gini y la prueba F . De igual



Gráfica 3: Ejemplo de árbol de decisión para iniciación.

manera se evalúa la distribución de la población, la proporción de malos en cada nodo y el número de malos por cada bueno, mejor conocido como “odds to be good”, como se presenta en la siguiente gráfica: La capacidad de predicción



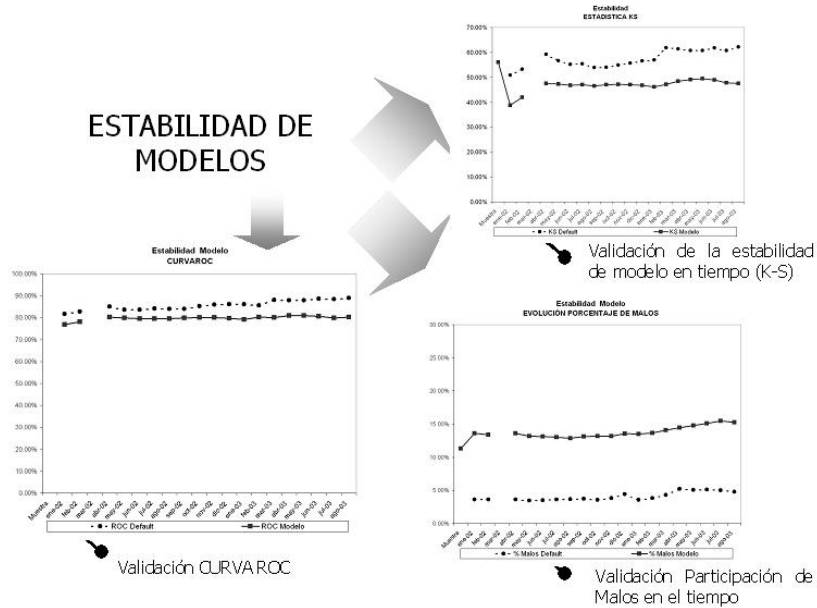
Gráfica 4: Ejemplos de medición de potencia de un modelo.

($K - S$ y área bajo la curva ROC), de los árboles de decisión es comparable con la obtenida en una regresión logística³ con la ventaja de la facilidad de entendimiento. Sin embargo un factor que puede estar en contra de los árboles de decisión es que no es posible cuantificar la magnitud con la que una variable aporta para la predicción del incumplimiento.

En cuanto a la estabilidad de los modelos, cada una de las medidas mencionadas anteriormente es calculada con una periodicidad mensual y de esta manera es posible monitorear el desempeño del modelo medido en términos de capacidad de discriminación. Adicionalmente se verifica que no haya cambios

³ Ver “Comparación entre regresión logística y árboles de decisión en modelos de riesgo crediticio para el Banco Colpatria”. Almanza D., Cardona P., Rodríguez N. (2004). Trabajo de Grado para optar al título de estadística, Universidad Nacional de Colombia, Sede Bogotá.

dramáticos en las probabilidades asignadas a cada nodo terminal. Cabe resal-



Gráfica 5: Ejemplos de medición de estabilidad de un modelo.

tar que otro de los cálculos indispensables para la pérdida esperada es el de la severidad, en el que también es necesario emplear herramientas estadísticas para estudiar el comportamiento histórico de las pérdidas netas de los clientes que se han siniestrado en la cartera estudiada y asumir un comportamiento futuro. Así, empleando las probabilidades obtenidas de los árboles de decisión y la severidad, se calcula el valor de la pérdida esperada con la cuál se realiza la provisión tal como lo reglamenta la Superintendencia Bancaria. Históricamente en Colombia se hacía provisión únicamente sobre los clientes que ya tenían algún grado de morosidad, pero con la aplicación de las nuevas técnicas estadísticas se está empezando a hacer una provisión sobre clientes que estando al día en sus pagos pueden tener una probabilidad considerable de incurrir en moras hacia futuro.

6. Conclusión

Los árboles de decisión se presentan como una herramienta efectiva para la predicción de probabilidades de incumplimiento, no solo a nivel de capacidad de discriminación (potencia), estabilidad a través del tiempo, sino como una herramienta de fácil entendimiento que permite potencializar sus usos y servir además de la predicción, para la planeación de estrategias comerciales de venta de servicios, estrategias de cobranza entre muchas otras.

La importancia de tener un modelo de cálculo de probabilidad de incumplimiento confiable y con una alta capacidad de discriminación radica en que esta impacta considerablemente el cálculo de provisiones, afectando directamente el balance y las utilidades que podría llegar a tener la entidad. Adicionalmente como los modelos son empleados para otorgamiento de créditos, mantenimiento de cuentas, hacen parte fundamental de la gestión integral de riesgo, por tanto un cálculo u operación inapropiada podría llevar a una institución financiera a situaciones de insolvencia.

Nota: Las opiniones y los errores eventuales son responsabilidad exclusiva del autor y sus contenidos no comprometen al Banco Colpatría Red Multibanca.

Bibliografía

- Araya, R. (1994), *Induction of Decision Trees when Examples are Described with Noisy Measurements and with Fuzzy Class Membership*.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984), *Classification and Regression Trees*, Wadsworth. Belmont.
- Comite, B. (2001a), Carta circular 31 de 2001, Technical report, Superintendencia Bancaria de Colombia.
- Comite, B. (2001b), Circular externa 011 de 2001, Technical report, Superintendencia Bancaria de Colombia.
- Committee, B. (1999), Credit risk modeling current practices and applications, Technical report, Basle Committee on Banking Supervision. Basle Switzerland.
- Committee, B. (2000), Principles for the management of credit risk, Technical report, Basle Committee on Banking Supervision. Basle Switzerland.

- Elizondo, A. (2003), *Medición integral del riesgo*, Limusa Noriega Editores.
- Joos, P., Vanhoof, K., Ooghe, H. & Sierens, N. (1998), *Credit classification: a comparison of logit models and decision trees*. Faculteit Economie en Bedrijfskunde.
- Kass, G. V. (1980), 'An exploratory technique for investigating large quantities of categorical data', *Applied Statistics* **29**.
- Mays, E. (2001), *Handbook of Credit Scoring*, The Glenlake Publishing Company Ltda.
- Ong, M. (1999), *Internal Credit Risk Models*, Risk Books.
- Wilson, S. & Press, S. J. (1978), 'Choosing between logistic regression and discriminant analysis', *American Statistical Association* **73**, 699–705.