

**LOS MODELOS INTERNOS (IRB) EN BASILEA II:
LA METODOLOGÍA ROUGH SET EN UNA APROXIMACION A LA DETERMINACIÓN DE
LA PROBABILIDAD DE IMPAGO**

Reyes Samaniego Medina
Profesor Colaborador
Universidad Pablo de Olavide
Correo electrónico: rsammed@upo.es

María José Vázquez Cueto
Catedrática
Universidad de Sevilla
Correo electrónico: pepi@us.es

Datos de contacto:
Reyes Samaniego Medina
Universidad Pablo de Olavide
Área Finanzas
Departamento Dirección de Empresa.
Carretera de Utrera, Km. 1
41013 Sevilla
Teléfono: 954 34 98 45
Fax: 954 34 83 53
Email: rsammed@upo.es

Área Temática: Valoración y Finanzas

Palabras claves: Calificación de préstamos, Riesgo de crédito, Basilea II, Rough Sets

**LOS MODELOS INTERNOS (IRB) EN BASILEA II:
LA METODOLOGÍA ROUGH SET EN UNA APROXIMACION A LA DETERMINACIÓN DE
LA PROBABILIDAD DE IMPAGO**

RESUMEN

Basilea II da cabida e incentiva la implantación de modelos propios para la medición de los riesgos financieros en las entidades de crédito. En el trabajo que presentamos nos centramos en los modelos internos para la valoración del riesgo de crédito (IRB) y concretamente en la aproximación a uno de sus componentes: la probabilidad de impago (PD).

Nuestro trabajo se estructura en tres bloques, en el primero mostramos los aspectos más sobresalientes del tratamiento del riesgo de crédito en Basilea II, para centrarnos en las necesidades de desarrollar modelos matemáticos que nos permitan valorar el riesgo de crédito tal cual exige Basilea. En el segundo exponemos teóricamente el modelo matemático objeto de nuestro trabajo: rough set. A continuación realizamos una aplicación empírica de dicha metodología a una base de datos de una entidad financiera andaluza, con el objetivo de aproximarnos al cálculo de una de las variables en la valoración del riesgo de crédito, la probabilidad de impago

I. INTRODUCCIÓN

El Comité de Basilea, consciente de los cambios que se han ido produciendo en el mundo bancario en los últimos años, se planteó en 1999 una reestructuración del Acuerdo de 1988 sobre la medición y control de los riesgos de las entidades financieras. Así, en junio de 2004, se aprueba el documento definitivo del nuevo Acuerdo: Basilea II. Acuerdo que marcará un punto de partida tanto en la gestión de los riesgos como en las relaciones que habrán de mantener entidades financieras y supervisores.

Basilea II da cabida e incentiva la implantación de modelos propios, llamados internos (IRB), para la medición de los riesgos financieros en las entidades de crédito. Estos modelos están basados en estimaciones propias de la entidad financiera y tienen como objetivo el cálculo de la pérdida inesperada (IL) de las exposiciones crediticias. Esta pérdida depende de un conjunto de factores: probabilidad de impago (PD), pérdida en caso de impago (LGD), exposición en el momento del impago o severidad (EAD), vencimiento (M) y granularidad (G). El método IRB se plantea en el Acuerdo en dos versiones: básico y avanzado. Ambos tienen en común la necesidad de estimar la probabilidad de impago (PD). Estas estimaciones deberán estar basadas en datos históricos y mostrar una visión conservadora a medio y largo plazo. El IRB avanzado requiere además la estimación del resto de los componentes de la pérdida inesperada (IL).

Este trabajo está organizado como sigue. En la sección dos se exponen los aspectos más sobresalientes del tratamiento del riesgo de crédito en Basilea II. En la sección tres hacemos un breve repaso a la modelización del riesgo de crédito en la literatura financiera. En la sección cuatro introducimos la teoría rough set. Continuamos, en la sección cinco, con una descripción de la muestra de empresas que se ha utilizado. La aplicación empírica realizada se muestra en la sección sexta, donde, en primer lugar, a través de la metodología rough set intentamos determinar cuáles son las variables que pueden actuar como explicativas de la morosidad y posteriormente, enfrentamos los resultados de esta metodología con los obtenidos con el análisis discriminante clásico. Finalmente obtenemos una serie de conclusiones que exponemos en la sección siete.

II. EL RIESGO DE CRÉDITO EN EL NUEVO ACUERDO DE BASILEA

El tratamiento de los riesgos financieros se ha convertido en un factor estratégico, ya no sólo para las entidades financieras, sino para cualquier organización, con independencia

del tamaño que posea y del sector en que realice su actividad; un factor que puede marcar el futuro de cualquier entidad. Si nos centramos en las entidades financieras, el incremento de la competencia, así como los avances producidos en la diversificación y en la liquidez y los cambios tan significativos en la regulación, tales como los requerimientos de capital basados en los riesgos financieros, han llevado a las entidades a buscar formas innovadoras que ayuden a la medición y gestión de todos los riesgos que pueden influir en su desarrollo.

El objetivo básico del nuevo Acuerdo de Capital, Basilea II, es proporcionar una estimación del capital regulatorio más sensible a los riesgos financieros y para ello propone la utilización de metodologías internas elaboradas por los propios bancos. Para conseguir este objetivo, el Comité incluye en el nuevo Acuerdo partes o técnicas hasta ahora no tenidas en cuenta. Así aparecen elementos novedosos tales como:

- Técnicas internas, con distinto grado de valoración, del riesgo de crédito.
- Técnicas de cobertura de riesgo de crédito.
- Nuevos riesgos como el operacional y el de tipo de interés en carteras de inversión.
- Dos nuevos pilares complementarios: de supervisión y de disciplina de mercado.

Basilea II establece dos métodos para la estimación del capital regulatorio que cubra el riesgo de crédito. Por un lado, el método *estándar*, que es una mejora del utilizado en el acuerdo anterior, y que establece nuevas categorías de riesgo, agrupando a cada tipo de empresa dentro de éstas. Por otro lado, el *Método Interno o IRB (Internal Rating Based)* basado en calificaciones internas, distinguiéndose dentro de éste dos niveles: *Método IRB Básico* donde el banco calcula parte de las variables, pero no todas, siendo las restantes proporcionadas por el supervisor; y el *Método IRB Avanzado*, donde el banco calcula todas las variables que conforman el modelo. El Comité espera que todas las entidades, con el tiempo, desarrollen el método IRB avanzado.

Antes de explicar las diferencias entre las dos variantes del método IRB sería interesante recordar los distintos componentes del riesgo de crédito, distinguiendo entre pérdida esperada e inesperada.

La *pérdida esperada (EL)* depende de un conjunto de factores:

- *PD*: Probabilidad de impago
- *LGD*: Pérdida en caso de impago
- *EAD*: Exposición en el momento de impago

- *M*: vencimiento
- *G*: Granularidad (se aplica a todas las carteras excepto a las de carácter minorista).

Y se obtendría de la siguiente formula:

$$EL = PD \times LGD \times EAD \quad [1]$$

O bien,

$$EL = \text{Prob. de incumpliendo} \times (1 - \text{tasa de recuperación}) \times \text{cuantía}$$

La pérdida esperada (*EL*) es una pérdida media mientras que la pérdida inesperada (*IL*) es una pérdida máxima con un determinado nivel de confianza:

$$IL = \text{Desviación típica de la probabilidad de incumplimiento} \times (1 - \text{tasa de recuperación}) \times \text{cuantía del préstamo} \times C \quad [2]$$

Donde *C* es un parámetro a estimar, que será mayor cuanto mayor sea el nivel de confianza deseado.

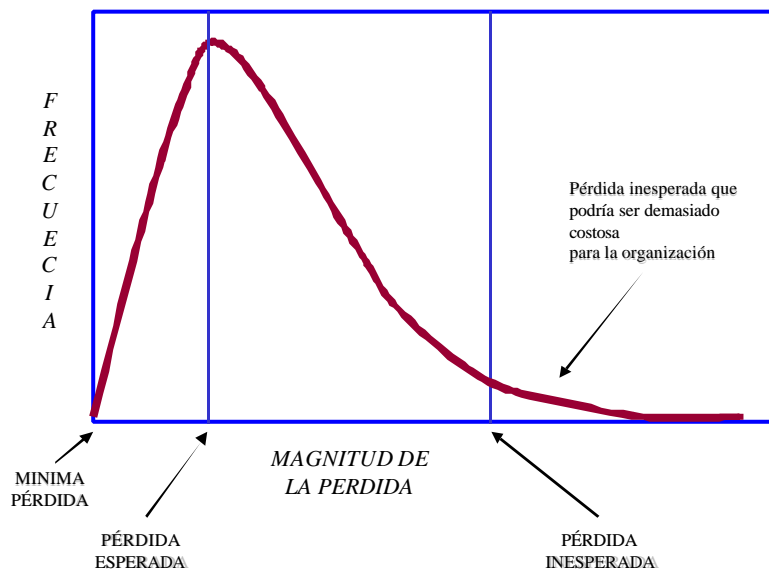


Figura 1. Pérdida esperada e inesperada en función del nivel de confianza.

En el método IRB básico, la entidad sólo estima sus probabilidades de impago (*PD*) siendo el supervisor el que proporciona los valores del resto de las variables. En la metodología avanzada, será el banco el que estime cada una de las variables, así como el tratamiento de las garantías y derivados de crédito.

Para poder aplicar alguno de estos métodos, el banco debe cumplir una serie de requisitos. Así, tendrá que cumplir unos mínimos para poder estimar PD y unos adicionales para estimar cada una de las restantes variables.

El Comité de Basilea permite que las entidades financieras estimen estas variables con sus propios modelos, pero no que determinen todos los elementos necesarios para calcular sus propios requerimientos de capital ya que la ponderación de riesgo y, por tanto, las exigencias de capital, se establecen combinando los valores proporcionados por el banco y la formulación específica dada por el Comité.

Por tanto, para utilizar el método IRB, en cualquiera de sus dos versiones, es necesario poder calcular la probabilidad de incumplimiento (PD). Para ello se tiene que calcular la PD media de todos los prestatarios incluidos en un grupo o grado de clasificación. Estas estimaciones deberán estar basadas en sus datos históricos y mostrar una visión conservadora a medio y largo plazo.

III. MODELIZACIÓN DEL RIESGO DE CRÉDITO EN LA LITERATURA FINANCIERA

Mucho se ha escrito sobre los modelos de riesgo de crédito en los últimos años. De hecho, Basilea establece su formulación a través de un estudio realizado en las principales entidades financieras sobre los modelos seguidos por ella. Así, los modelos tradicionales utilizados son los denominados “Modelos de forma reducida”¹, en los que se considera la insolvencia como un fenómeno exógeno. Todo el esfuerzo teórico se centra en la modelización del proceso estocástico asociado a la misma y en la determinación de las variables que habrán de ser introducidas en dichos modelos. En este sentido podemos distinguir entre los modelos univariantes, que estudian el comportamiento de cada una de las variables potencialmente explicativas de la insolvencia por separado², y los modelos multivariantes, que combinan la información proporcionada por un conjunto de variables, siendo el trabajo pionero en este sentido, el realizado por Altman en 1968³. Más

¹ Desarrollado, entre otros, por Jarrow y Turnbull (1995) y Duffie y Singleton (1999).

² Uno de los estudios clásicos a este respecto fue realizado por Beaver (1966), el cual encontró que un número de ratios financieros podría discriminar entre empresas sanas y fallidas, en un periodo de 5 años antes de que se produjera el impago.

³ Altman (1968) realizó un análisis discriminante combinando la información proporcionada por 25 ratios financieros.

recientemente, Cabedo, Reverte y Tirado (2004) hacen una adaptación de este modelo para el cálculo de las probabilidades de impago en corporaciones, sobre una cartera hipotética perteneciente a un mismo sector, con el objeto de calcular el capital regulatorio según el método IRB básico del Acuerdo de Basilea. En las últimas décadas, sin embargo, se ha optado fundamentalmente por el análisis logit y probit⁴. Así, Wilson (1997) desarrolló el modelo CreditPortfolio View para Mckinsey, estableciendo un modelo discreto de periodos múltiples. Con esta metodología obtiene la probabilidad de impago como funciones logit de índices de variables macroeconómicas que, de alguna manera, representen el funcionamiento de la economía. Más recientemente, Fernandez (2005) utiliza la misma metodología realizando previamente un análisis univariante con el fin de elegir, de entre los 23 ratios de partida, aquellos con mayor poder discriminante, dentro de cada una de las categorías de ratio establecidas⁵. Posteriormente efectúa un análisis multivariante logit/probit con los que se obtiene puntuaciones para cada empresa, lo que le permite establecer un sistema de rating y asignar probabilidades de impago.

En la última década del pasado siglo, y en un intento de superar las hipótesis tan restrictivas que imponen los modelos de inferencia estadística a las variables, hipótesis que no se suelen cumplir en la realidad y que distorsionan los resultados de las técnicas⁶, se empiezan a aplicar técnicas procedentes del campo de la Inteligencia Artificial, elaborando programas capaces de generar conocimiento a través de datos empíricos y, posteriormente, usar ese conocimiento para realizar inferencias sobre nuevos datos. Dentro de este enfoque podemos distinguir las técnicas que buscan el conocimiento a través de anticipar patrones en los datos, entre ellas las diversas redes neuronales, y las consistentes en inferir reglas de decisión a partir de los datos de la base. A este último grupo pertenece la metodología de los Rough Sets.⁷

IV. MARCO TEORICO DE LA METODOLOGÍA ROUGH SETS

La teoría Rough Set fue propuesta por Z. Pawlak en 1982 y se ha afianzado como una buena herramienta para tratar con la vaguedad y la imprecisión. Más concretamente, el método es eficiente para tratar con los problemas de clasificación multidimensionales (Pawlak y Slowinski, 1994). La idea básica descansa en la relación de indiscernibilidad que

⁴ El primero en usar este tipo de técnicas aplicada a la predicción empresarial fue Ohlson (1980).

⁵ Liquidez, apalancamiento, actividad, cobertura de deuda y productividad.

⁶ Eisembeis (1977), Ohlson (1980) y Zavgren (1983) cuestionan por estas razones la validez del método.

⁷ Daubie et al (2002) han aplicado dicha técnica a la clasificación de préstamos comerciales.

describe elementos indistinguibles. Su objetivo básico es encontrar reglas de decisión básicas que permitan adquirir nuevo conocimiento. Sus conceptos básicos son 1) la discernibilidad, 2) la aproximación, 3) los reductos y 4) las reglas de decisión. El punto de partida del método es la existencia de una tabla de información/decisión, donde cada elemento viene caracterizado por un conjunto de variables y una variable de decisión, que clasifica al elemento en una de dos ó más categorías. La indiscernibilidad se produce cuando dos elementos vienen caracterizados por los mismos valores de las variables y, sin embargo, no coinciden las categorías en las que se clasifican. Esta es la base de los conjuntos rough set. En tal caso, y para cada clase de decisión o categoría X , y cada subconjunto B de variables, se construyen dos conjuntos, denominados respectivamente aproximaciones por debajo y aproximaciones por encima de la clase de decisión. El conjunto aproximación por debajo de la clase de decisión X respecto a las variables B , $\underline{B}X$ viene dado por el conjunto de todos los elementos que, caracterizados por B , con toda seguridad pertenecen a X . El conjunto aproximación por encima de la clase X , $\overline{B}X$, viene dado por el conjunto de elementos que, en base a la información B que poseemos, no podemos asegurar que no estén en X . La diferencia entre los dos conjuntos la forman los elementos “dudosos”, es decir, aquellos que no sabemos con toda certeza, usando sólo la información contenida en B , si pertenecen o no a X . Cuando esta diferencia es no vacía se dice que X es un rough set respecto a B . Este conjunto puede ser caracterizado numéricamente por el cociente entre el cardinal del conjunto aproximación por debajo y el cardinal del conjunto aproximación por encima. A este cociente se le denomina precisión. Si existen varias clases de decisión, a la suma de los cardinales de todas las aproximaciones por debajo dividida por el total de elementos, se denomina calidad de la clasificación por medio del conjunto B . Es el porcentaje de elementos correctamente clasificados.

Otro aspecto importante de esta técnica es la reducción de la tabla inicial de datos, eliminando aquella información redundante. Este proceso se realiza a través de los reductos. Un reducto es un conjunto mínimo de variables que conserva la misma habilidad de clasificación de los elementos que la tabla completa de información. Un reducto es pues la parte esencial del conocimiento y constituyen la manera más concisa mediante la cual podemos distinguir las clases de decisión⁸.

La etapa final del análisis rough set es la creación de reglas de decisión, es decir, reglas que nos permitan decir si un elemento dado pertenece a determinada clases de

⁸ La obtención de los reductos está basada en las clases de equivalencia que define, sobre el conjunto de las observaciones, la relación de indiscernibilidad.

decisión. Estas reglas representan el conocimiento y se generan combinando los reductos con los valores de los datos analizados. Una regla de decisión es una sentencia lógica del tipo “SI(se cumplen determinadas condiciones) ENTONCES(el elemento pertenece a una determinada clase de decisión)”. Estas reglas nos permitirán, de manera fácil, clasificar nuevos elementos⁹.

V. DATOS Y VARIABLES UTILIZADAS

Tanto en la selección de la muestra como en la elección de las variables independientes utilizadas en nuestro estudio empírico se han seguido las siguientes pautas:

- Siguiendo a Altman (1968) hemos emparejado, bajo criterios de tamaño y sector, empresas sanas y fallidas, situando el porcentaje de empresas fallidas en un 50% del total.
- En principio, parece de especial relevancia a la hora de seleccionar la muestra que los datos considerados se obtengan para el mismo periodo de tiempo en empresas sanas y fallidas. Sin embargo, las empresas en quiebra o suspensión de pagos tienden a retrasar la presentación de sus datos contables en el periodo de tiempo previo a la declaración concursal. Para subsanar este inconveniente, hemos considerado, siguiendo a Trujillo (2002), como datos contables del último año previo al fallido, los datos más recientes disponibles.

Selección de la Muestra.

En el desarrollo de nuestro trabajo, hemos empleado una base de datos suministrada por una caja de ahorros española, que contiene información de empresas que solicitaron y obtuvieron un crédito en dicha entidad. Estas compañías fueron divididas en dos grupos: sanas y fallidas. En concreto, la muestra de empresas fallidas utilizada para el análisis ha sido obtenida de aquellas compañías, prestatarias de la entidad financiera, cuya deuda resultó impagada, de intereses o principal, en un porcentaje superior al 10% del riesgo concedido. El momento de cómputo ha sido el 31 de diciembre de 2003.

El grupo de empresas sanas, es decir aquellas que no generaron situaciones de impago en el horizonte considerado, se ha seleccionado mediante la técnica del emparejamiento individual, controlado por aquellas características que podrían incidir en las relaciones entre ratios financieros y fracaso. A cada empresa del grupo de fallidas, se ha

⁹ Para más detalle sobre los aspectos matemáticos formales de la metodología puede consultarse Komorowski et al (1999)

unido una empresa sana de la misma industria y de aproximado tamaño. En relación al sector, el emparejamiento se ha realizado a nivel de cuatro dígitos del C.N.A.E. de 1993. El criterio seguido para emparejar por tamaño ha sido el activo total.

Como factor homogeneizador a todas las empresas, hemos controlado que el total de operaciones del cliente con la entidad financiera, o riesgo vivo, fuese superior a 60.120 euros, y que se tratase de sociedades anónimas o limitadas, lo que facilitaría el acceso a sus estados contables.

En total, la muestra se compone de 106 empresas, 53 fallidas y 53 sanas de las más diversas actividades económicas¹⁰.

Selección de las variables independientes de los modelos

Las variables independientes escogidas para la elaboración de los modelos fueron seleccionadas de los estados financieros, fundamentalmente balance y cuenta de pérdidas y ganancias, de las empresas que componen la muestra. Dichos estados contables fueron extraídos de la base de datos SABI de la compañía INFORMA, S.A., que incluye más del 95% de las compañías que presentan sus cuentas en el Registro Mercantil en España. Dado que la mayoría de las empresas que resultaron fallidas no presentaron los estados financieros el año anterior o incluso en los dos años previos a la fecha de impago, hemos considerado los últimos datos disponibles como los del año previo al fallo empresarial. Así, el año t-1 corresponde al último de los estados contables disponibles.

La información contable derivada de la muestra seleccionada fue objeto de un minucioso estudio con el fin de detectar y subsanar, en su caso, posibles anomalías o incidencias significativas, que pudieran distorsionar el análisis final. Aquellas empresas atípicas por claras e insalvables anomalías contables quedaron excluidas de la muestra. En este sentido, fueron eliminadas, por ejemplo, aquellas entidades que, estando en situación de impago, presentaban beneficios.

La selección de ratios se ha hecho escogiendo un conjunto amplio de variables, 25, potencialmente explicativas del fracaso empresarial sobre la base de la frecuencia y eficacia

¹⁰ Hemos dejado fuera del análisis las sociedades de promoción inmobiliaria, por ser éstas de unas características muy peculiares y diferentes al resto de compañías, y porque en la evaluación de la solicitud de un préstamo a este tipo de empresas el factor decisivo para su concesión es la viabilidad del proyecto concreto para el que se solicita, la cual no aparece reflejada en los estados contables pertinentes.

con que han sido utilizadas en otros modelos de predicción de la insolvencia empresarial o en el análisis de riesgos bancarios, siguiendo la clasificación realizada por Trujillo (2002).

Las variables utilizadas incluyen ratios de liquidez, endeudamiento, estructura, rotación, generación de recursos y rentabilidad. En concreto, los ratios considerados en el análisis se recogen en el cuadro 1.

Cuadro 1. *Ratios considerados en el análisis*

RATIOS DE LIQUIDEZ	Grado en que los activos liquidables a corto plazo de la empresa son suficientes para hacer frente a los pagos provenientes de las deudas contraídas a corto plazo.	$R1 = \text{Activo Circulante} / \text{Pasivo Circulante}$
		$R2 = (\text{Exigible} + \text{Disponible}) / \text{Pasivo Circulante}$
		$R3 = \text{Disponible} / \text{Pasivo Circulante}$
		$R4 = (\text{Exigible} + \text{Disponible} - \text{Pasivo Circulante}) / (\text{Consumos de Explotación} + \text{Gastos de Personal} + \text{Variación Provisiones} + \text{Otros Gastos de Explotación})$
RATIOS DE ENDEUDAMIENTO	Relación entre los distintos componentes del pasivo, a corto y largo plazo, y los recursos propios; así como entre el coste de la deuda y el pasivo o los beneficios y recursos generados	$R5 = \text{Pasivo Fijo} / \text{Neto}$
		$R6 = \text{Neto} / \text{Pasivo Total}$
		$R7 = \text{Pasivo Fijo} / (\text{Pasivo Fijo} + \text{Pasivo Circulante})$
		$R8 = \text{Costes Financieros} / (\text{Pasivo Fijo} + \text{Pasivo Circulante})$
		$R9 = \text{Costes Financieros} / (\text{BAIT} + \text{Dotación Amortización})$
		$R10 = \text{Costes Financieros} / \text{BAIT}$
		$R11 = \text{Pasivo Ajeno} / \text{Pasivo Total}$
RATIOS DE ESTRUCTURA	Proporcionalidad entre las masas patrimoniales del activo y del pasivo, así como en la composición de las mismas.	$R12 = (\text{Activo Circulante} - \text{Pasivo Circulante}) / \text{Activo Total}$
		$R13 = \text{Activo Circulante} / \text{Activo Total}$

RATIO DE ROTACIÓN	Medida del dinamismo de la actividad empresarial en relación a la estructura de la empresa.	$R14 = \frac{\text{Activo Circulante} - \text{Pasivo Circulante}}{\text{Importe Neto Cifra de Negocio} + \text{Otros Ingresos de Explotación}}$
RATIOS DE GENERACIÓN DE RECURSOS	Relación de la capacidad de autofinanciación de la empresa con diferentes magnitudes.	$R15 = \frac{\text{Resultado Ejercicio} + \text{Dotación Amortización}}{\text{Importe Neto Cifra Negocios} + \text{Otros Ingresos de Explotación}}$
		$R16 = \frac{\text{Resultado Ejercicio} + \text{Dotación Amortización}}{\text{Pasivo Circulante}}$
		$R17 = \frac{\text{Resultado Ejercicio} + \text{Dotación Amortización}}{\text{Pasivo Fijo} + \text{Pasivo Circulante}}$
		$R18 = \frac{\text{Resultado Ejercicio} + \text{Dotación Amortización}}{\text{Pasivo Total}}$
		$R19 = \frac{\text{BAIT} + \text{Dotación Amortización}}{\text{Pasivo Circulante}}$
RATIOS DE RENTABILIDAD	Comparación del beneficio obtenido en sus distintos niveles, con los recursos invertidos	$R20 = \frac{\text{Resultado Explotación} + \text{Ingresos Financieros} + \text{Beneficios Inversiones Financieras} + \text{Diferencias Positivas de Cambio}}{\text{Activo Total}}$
		$R21 = \frac{\text{Resultado de Actividades Ordinarias}}{\text{Pasivo Total}}$
		$R22 = \frac{\text{Resultado antes de Impuestos}}{\text{Neto}}$
		$R23 = \frac{\text{Resultado antes de Impuestos}}{\text{Pasivo Total}}$
		$R24 = \frac{\text{Resultado del Ejercicio}}{\text{Neto}}$
		$R25 = \frac{\text{BAIT}}{\text{Activo Total}}$

Fuente: Trujillo (2002)

VI. APLICACIÓN EMPÍRICA.

Para llevar a cabo la aplicación empírica desarrollada en el presente trabajo, se han calculado los valores de los 25 ratios económico – financiero mostrados en el cuadro 1, para cada una de las 53 empresas fallidas, un periodo antes de la entrada en mora, y se ha

procedido de igual forma para cada “pareja” sana. Se obtiene así una tabla de información de 106x25 datos. A esta tabla se le ha unido una columna indicativa de la situación de fracasada o saneada de la empresa en cuestión. Así hemos asignado el valor 0 a la empresa fallida y 1 a la correspondiente pareja sana. Obtenemos, de esta forma, una tabla de información-decisión de 106x26 datos.

Con esta tabla tratamos de determinar cuál o cuáles de estos 25 ratios pueden actuar como explicativos de la morosidad de una empresa, como primer paso para calcular la probabilidad de impago.

A continuación, y dada la naturaleza de las variables consideradas, hemos procedido a la discretización de los valores. Esto no es requisito imprescindible para la aplicación de la técnica, pero facilita la interpretación de los resultados y es más coherente identificar empresas fallidas o sanas, no cuando coincidan exactamente los valores de las variables consideradas, sino cuando éstos se encuentren dentro de un mismo intervalo. Para ello hemos utilizado la codificación que se recoge en el cuadro dos¹¹.

Cuadro 2. Intervalos de codificación de las variables

Varia ble	VALOR CODIFICADO			
	0	1	2	3
R1	(-inf, 0.00434058)	(0.00434058, 0.00437793)	(0.00437793, 0.00467188)	(0.00467188, +inf)
R2	(-inf, 0.00131066)	(0.00131066, 0.00215777)	(0.00215777, 0.00653424)	(0.00653424, +inf)
R3	(-inf, 3.01154e-005)	(3.01154e-005, 4.19928e-005)	(4.19928e-005, 0.000818944)	(0.000818944, +inf)
R4	(-inf, -0.00126886)	(-0.00126886, -0.000469693)	(-0.000469693, -0.000412292)	(-0.000412292, +inf)
R5	(-inf, -0.000651859)	(-0.000651859, 0.000583821)	(0.000583821, 0.00176257)	(0.00176257, +inf)
R6	(-inf, 1.38138e-005)	(1.38138e-005, 0.00072438)	(0.00072438, 0.00077608)	(0.00077608, +inf)
R7	(-inf, 1.46802e-005)	(1.46802e-005, 6.811e-005)	(6.811e-005, 0.000260608)	(0.000260608, +inf)
R8	(-inf, 4.45802e-005)	(4.45802e-005, 6.1332e-005)	(6.1332e-005, 0.000253627)	(0.000253627, +inf)
R9	(-inf, -0.00032564)	(-0.00032564, 0.00129406)	(0.00129406, 0.00412803)	(0.00412803, +inf)
R10	(-inf, -0.00032564)	(-0.00032564, 0.00170486)	(0.00170486, 0.00355543)	(0.00355543, +inf)
R11	(-inf, 0.00308829)	(0.00308829, 0.00355161)	(0.00355161, 0.00430769)	(0.00430769, +inf)
R12	(-inf, -0.00169722)	(-0.00169722, 3.00724e-005)	(3.00724e-005, 0.000163319)	(0.000163319, +inf)
R13	(-inf, 0.00131974)	(0.00131974, 0.00255195)	(0.00255195, 0.00265304)	(0.00265304, +inf)
R14	(-inf, -0.000937993)	(-0.000937993, 0.000271131)	(0.000271131, 0.00183654)	(0.00183654, +inf)
R15	(-inf, 9.4422e-006)	(9.44223e-006, 7.94461e-005)	(7.94461e-005, 0.000744896)	(0.000744896, +inf)

¹¹ Hemos discretizado las variables agrupándolas en cuatro intervalos en base al número de observaciones pertenecientes a cada uno. Para ello se ha utilizado el software ROSE proporcionado por el Institute of Computing Science of Poznan University of Technology, al que agradecemos su disponibilidad.

R16 : (-inf, 3.03552e-005) (3.03552e-005, 0.000232274) (0.000232274, 0.000290301) (0.000290301, +inf)
R17 : (-inf, 6.49981e-006) (6.49981e-006, 0.000152804) (0.000152804, 0.000620383) (0.000620383, +inf)
R18 : (-inf, 6.4581e-006) (6.4581e-006, 1.90195e-005) (1.90195e-005, 0.000153841) (0.000153841, +inf)
R19 : (-inf, 5.4507e-005) (5.4507e-005, 5.56532e-005) (5.56532e-005, 0.000473715) (0.000473715, +inf)
R20 : (-inf, -0.000115164) (-0.000115164, 5.14362e-005) (5.14362e-005, 0.00121164) (0.00121164, +inf)
R21 : (-inf, -0.000148011) (-0.000148011, 4.68013e-005) (4.68013e-005, 0.00120903) (0.00120903, +inf)
R22 : (-inf, 1.16457e-005) (1.16457e-005, 0.000203684) (0.000203684, 0.00417805) (0.00417805, +inf)
R23 : (-inf, 7.0378e-007) (7.0378e-007, 1.22549e-005) (1.22549e-005, 0.000117764) (0.000117764, +inf)
R24 : (-inf, 1.10402e-005) (1.10402e-005, 0.000426896) (0.000426896, 0.00416212) (0.00416212, +inf)
R25 : (-inf, 1.2036e-005) (1.2036e-005, 2.53061e-005) (2.53061e-005, 0.000161758) (0.000161758, +inf)

Fuente: Elaboración propia.

El siguiente paso es la construcción de los reductos. Al existir correlaciones entre las variables explicativas introducidas en el análisis, el número de reductos que ofrecía el programa era muy elevado. Optamos por establecer como criterio de parada los cuatrocientos doce primeros, número que consideramos suficientemente significativo, así, aparecieron reductos que contenían entre siete y once variables, lo que ya supone una reducción con respecto a las veinticinco originales. En el cuadro 3 ofrecemos, a modo ilustrativo, los primeros cinco reductos encontrados. La frecuencia con la que aparece cada variable en los reductos viene dada en el cuadro 4.

Cuadro 3. *Reductos.*

1	{R8, R11, R13, R15, R16, R18, R20, R24, R25}
2	{R10, R11, R13, R16, R22, R24, R25}
3	{R10, R11, R13, R17, R19, R22, R24, R25}
4	{R10, R11, R13, R18, R19, R22, R24, R25}
5	{R10, R11, R13, R19, R20, R22, R24, R25}

Cuadro 4. *Aparición de las variables en los reductos*

Variable	Frecuencia	Variable	Frecuencia
R1	0%	R7	16.5%
R2	0.72%	R8	16.75%
R3	3.6%	R9	46.36%
R4	4.36%	R10	44.66%
R5	5.33%	R11	49.75%
R6	5.82%	R12	21.12%

R13	59.71%	R20	29.85%
R14	29.13%	R21	39.56%
R15	47.33%	R22	72.08%
R16	62.37%	R23	55.09%
R17	26.46%	R24	56.31%
R18	26.46%	R25	54.85%
R19	27.91%		

Fuente: Elaboración propia.

Esto nos indica que las variables más discriminatorias en la muestra son R22 y R16, ratios de rentabilidad y generación de recursos respectivamente, aparecen en un 72.08% y 62.37% respectivamente, de los reductos, lo que significa que no podemos prescindir de ellos sin renunciar a parte de información contenida en la muestra. Les sigue la variable R13, ratio de estructura, que aparece en los reductos con una frecuencia cercana al 60%. Llama la atención el hecho de que los ocho primeros ratios aparecen con una frecuencia inferior al 6%, por lo que parece que se puede prescindir de ellos sin perder información. Como los cuatro primeros ratios se corresponden con los de liquidez, parece que la mayor o menor liquidez de las empresas no es determinante de la morosidad de la misma. De entre todos los reductos que ha proporcionado el software nos hemos quedado con los noventa y cuatro que contenían conjuntamente a las variables que presentan mayor frecuencia de aparición. De ellos hemos seleccionado el reducto 161, {R10, R11, R13, R14, R16, R22, R25}. Los criterios seguidos para la selección han sido: primero, que contuviese el menor número de variables posibles, segundo que las variables presentasen una frecuencia elevada de aparición en los reductos, y tercero que contuviese los ratios económicos y financieros que aparecen con mayor frecuencia¹²

Con estas variables hemos reducido la tabla de información/decisión a una tabla 106x8 datos. A continuación se han generado reglas de decisión mediante el procedimiento lem2¹³ y hemos obtenido las veinticuatro reglas de decisión que figuran en el cuadro 5 .

¹² Estas condiciones las cumplen también los reductos 116, 141, 201, 373 y 392, que sustituyen el ratio R25 por otros ratios de rentabilidad (R21, R23 o R24) y/o prescinden del ratio R14. Nos ha parecido más conveniente mantener R14 porque es el único ratio de rotación que intervendrá en la clasificación, mientras que otros ratios de rentabilidad ya están introducidos.

¹³ Chan and Grzymala-Busse, 1994

Cuadro 5. Reglas de decisión.

- regla 1. Si $(R10 = 2)$ y $(R11 = 2)$ y $(R16 = 1)$ entonces $(D1 = 0)$;
- regla 2. Si $(R22 = 0)$ entonces $(D1 = 0)$;
- regla 3. Si $(R10 = 3)$ y $(R13 = 3)$ y $(R14 = 1)$ y $(R16 = 1)$ entonces $(D1 = 0)$;
- regla 4. Si $(R10 = 3)$ y $(R16 = 3)$ y $(R22 = 2)$ entonces $(D1 = 0)$;
- regla 5. Si $(R11 = 2)$ y $(R16 = 3)$ y $(R25 = 2)$ entonces $(D1 = 0)$;
- regla 6. Si $(R16 = 0)$ entonces $(D1 = 0)$;
- regla 7. Si $(R14 = 3)$ y $(R22 = 1)$ entonces $(D1 = 0)$;
- regla 8. Si $(R13 = 0)$ y $(R14 = 1)$ y $(R22 = 1)$ entonces $(D1 = 0)$;
- regla 9. Si $(R11 = 0)$ y $(R14 = 0)$ entonces $(D1 = 0)$;
- regla 10. Si $(R14 = 3)$ y $(R25 = 2)$ entonces $(D1 = 0)$
- regla 11. Si $(R10 = 1)$ y $(R13 = 3)$ y $(R22 = 3)$ entonces $(D1 = 0)$;
- regla 12. Si $(R13 = 1)$ y $(R16 = 1)$ $(R25 = 2)$ entonces $(D1 = 0)$;
- regla 13. Si $(R10 = 1)$ y $(R22 = 2)$ y $(R25 = 3)$ entonces $(D1 = 1)$;
- regla 14. Si $(R11 = 1)$ y $(R25 = 3)$ entonces $(D1 = 1)$;
- regla 15. Si $(R10 = 2)$ y $(R11 = 2)$ y $(R16 = 3)$ y $(R25 = 3)$ entonces $(D1 = 1)$;
- regla 16. Si $(R10 = 2)$ y $(R16 = 1)$ y $(R22 = 2)$ y $(R25 = 2)$ entonces $(D1 = 1)$;
- regla 17. Si $(R13 = 2)$ entonces $(D1 = 1)$;
- regla 18. Si $(R16 = 2)$ entonces $(D1 = 1)$;
- regla 19. Si $(R10 = 3)$ y $(R11 = 2)$ y $(R16 = 1)$ y $(R22 = 2)$ y $(R25 = 3)$ entonces $(D1 = 1)$;
- regla 20. Si $(R10 = 3)$ y $(R16 = 1)$ entonces $(D1 = 1)$;
- regla 21. Si $(R11 = 1)$ y $(R22 = 1)$ entonces $(D1 = 1)$;
- regla 22. Si $(R13 = 1)$ y $(R16 = 1)$ y $(R22 = 2)$ entonces $(D1 = 1)$;
- regla 23. Si $(R11 = 2)$ y $(R14 = 1)$ y $(R16 = 3)$ y $(R22 = 1)$ entonces $(D1 = 1)$;
- regla 24. Si $(R11 = 0)$ y $(R14 = 2)$ entonces $(D1 = 1)$

Fuente: Elaboración propia.

Las doce primeras reglas clasifican a las empresas morosas, y las doce siguientes a las restantes, es decir, a las empresas sanas. Así, por ejemplo, si el ratio de endeudamiento $R11$ (Pasivo ajeno/ Pasivo total) es igual a 0 y el ratio de rotación $R14$ (Activo Circulante – Pasivo Circulante) / (Importe Neto Cifra de Negocio + Otros Ingresos de Explotación) es igual a 0, o lo que es lo mismo, según los criterios de discretización, si el ratio 11 es menor que 0.7117 (exp -0.34) y el ratio 14 es menor que 0.8607 (exp -0.15), la empresa será clasificada como fallida ($D_1 = 0$), (Regla 9). Por el contrario, si el ratio de endeudamiento $R11$ (Pasivo Ajeno/Pasivo Total) es igual a 1, o lo que es lo mismo, está comprendido entre 0.7117 y 0.97044 (exp -0.34 y exp -0.03 respectivamente), y el ratio de rentabilidad $R25$ (BAIT/Activo Total), toma valores superiores a 0.8179 (exp -0.296), entonces la empresa será clasificada

como no fallida, ($D1 = 1$), (Regla 14). Estas veinticuatro reglas clasifican con un 100% de aciertos a los elementos utilizados. Además si realizamos una validación de los resultados mediante el procedimiento de “dejar uno fuera”, seguimos obteniendo un porcentaje aceptable de aciertos, en concreto un 79.25%, coincidiendo este porcentaje tanto para las empresas morosas como sanas.

Así, aplicando estas reglas a un nuevo préstamo, podríamos clasificarlo, detectando las posibles operaciones potencialmente fallidas.

Con idea de analizar la efectividad de esta técnica aplicada a la clasificación de operaciones crediticias, hemos comparado los resultados con los obtenidos si realizamos un análisis discriminante al mismo conjunto de datos, sin discretizar. Hemos construidos dos funciones discriminantes, una utilizando los ratios seleccionados por el propio modelo discriminante, usando como criterio de entrada $F\text{-Snedecor} > 1,5$ y como criterio de salida $F\text{-Snedecor} < 1$ y otra escogiendo sólo las variables del reducto seleccionado con el método rough set. Los coeficientes de estas funciones discriminantes se muestran en la cuadro 6. Por otro lado, en el cuadro 7 se presentan los resultados de la reclasificación con ambas funciones discriminantes, aplicadas a la muestra original y validándolas mediante el procedimiento de validación cruzada.

Cuadro 6. *Funciones discriminante y estadístico de los modelos*

Funciones	Lambda de Wilks	P-valor
-0.571 R4-0.687 R6+0.312 R7+0.559 R8+0.274 R9 +1.074 R15-0.354 R19-.713 R23	0.572	0.0
0.250 R10+0.779 R11-0.223 R13-0.295 R14-0.224 R16+0.177 R22-0.149 R25	0.749	0.0

Fuente: Elaboración propia

Cuadro 7. Porcentajes de clasificación correcta.

	Análisis Discriminante (ocho variables)		Análisis Discriminante (variables rough set)	
	Muestra de elaboración	Global	82.1	Global
Sanas		86.8	Sanas	75.5
Fallidas		77.4	Fallidas	71.7
Muestra de validación	Global	71.7	Global	70.8
	Sanas	77.4	Sanas	75.5
	Fallidas	66	Fallidas	66

Fuente: Elaboración propia.

De esta tabla se infiere que la metodología rough set se ha mostrado como una herramienta más útil que el análisis discriminante. Utilizando menor número de variables, e, incluso, con las mismas variables, no sólo muestra porcentajes globales de clasificación correcta mejores sino que también clasifica mejor dentro de cada una de las clases de decisión. Estos porcentajes, aunque disminuyen en la muestra considerada para la validación de los resultados, siguen manteniéndose superiores a nivel global y en cada grupo considerado separadamente. Cuando combinamos ambas técnicas utilizando las variables que previamente ha seleccionado la metodología rough set y repitiendo con ellas el análisis discriminante, los resultados obtenidos en la clasificación con el análisis discriminante son peores que si aplicamos las 24 reglas que nos proporciona la primera de las metodologías.

VII. CONCLUSIONES

El objetivo del nuevo Acuerdo de Capital, Basilea II, es proporcionar un capital regulatorio más sensible a los riesgos en general y al de crédito en particular. Para ello, en lo referente al riesgo de crédito, ha propuesto la utilización de metodologías internas de medición elaboradas por los propios bancos.

En nuestro trabajo, que se enmarca en una línea de investigación más amplia, encaminada a la determinación de una de las variables fundamentales de los modelos

internos: la probabilidad de impago (PD), presentamos una metodología alternativa al clásico análisis discriminante para determinar cuál o cuáles son las variables que pueden actuar como explicativas de la morosidad de una empresa.

Apoyándonos en una muestra de empresas sanas y fallidas y en un conjunto de 25 ratios financieros potencialmente explicativos de la morosidad de las mismas, hemos obtenido que la metodología Rough set puede ser una alternativa válida al análisis discriminante cuando se trata de problemas de clasificación de objetos en dos clases de decisión. Junto con la obtención de porcentajes de clasificación correcta aceptables se une el hecho de que no exige ningún tipo de comportamiento previo de las variables que intervienen en la clasificación, a diferencia del análisis discriminante que exige normalidad de las distribuciones e igualdad de las matrices de varianza-covarianza. Entre sus ventajas podemos destacar que elimina la información redundante y expresa las dependencias entre las variables consideradas y el resultado de la clasificación mediante unas reglas de decisión cuyo lenguaje está más próximo al lenguaje natural de un experto.

Bibliografía:

ALTMAN,E.I. (1968): "Financial Ratios, Discriminant Análisis and the Predictions of Corporate Bankruptcy" *Journal of Finance* (September) 589-609.

ALTMAN, E.I. (1998): *Managing Credit Risk. The New Great Financial Challenge.* John Wiley & Sons, Inc.

BASLE COMMITTEE ON BANKING SUPERVISION (2001): "The Internal Ratings-Based Approach". Consultive Document. Supporting Document to the New Basle Capital Accord, enero.

BEAVER, W. H. (1966): "Financial Ratios as Predictors of Failure". *Empirical Research in Accounting: Selected Studies*, (Suplemento de *The Accounting Review*).

CAREY, M., HRYCAY, M. (2001): "Parameterizing Credit Risk Models with Rating Data". *Journal of Banking & Finance*, 25.

COMITÉ DE SUPERVISIÓN BANCARIA DE BASILEA (2004): "Convergencia Internacional de Medidas y Normas de Capital". junio.

CABEDO,J., REVERTE, J.A., TIRADO, J.M. (2004): "Riesgo de Crédito y Recursos Propios Mínimos en Entidades Financieras". *Revista Europea de Dirección y Economía de la Empresa*, volumen 13, núm. 2.

CROUHY, M., GALAI, D., MARK, R. (2001): "Prototype Risk Rating System". *Journal of Banking & Finance*, 25.

DAUBIE,M.;LEVECK,P.;MESKENS,N.(2002):"A comparison of the rough sets and recursive partitioning induction approaches: An application to commercial loans" *International Transactions in Operational research* 9, 681-694.

DUFFIE,D.;SINGLETON,K.(1999): "Modeling Term Structures of Defaultable Bonds" *Review of Financial Studies*, 4.

EISEMBEIS,R.A.(1977):"Pitfalls in the applicationof discriminant anlysis in business and economics" *Journal of Finance*, 32, 875-900.

FERNANDEZ, J.E. (2005): "Corporate Credit Risk Modeling: Quantitative Rating System and Probability of Default Estimation". *DefaultRisk.com*, april.

JARROW, R. A.; TURNBULL,S.M.(1995): "Pricing Derivatives on Financial Securities Subject to Credit Risk", *Journal of Finance*, Vol. L, 1, 53-85.

KOMOROWSKI,Z.(1999):"Rough Sets: A Tutoria"l, Rough-Fuzzy Hybridization - A New Trend in Decision Making, (S.K. Pal and A. Skowron, Eds.), 3-98, Springer-Verlag Singapore Pte Ltd.

OHLSON, J.A. (1980): "Financial ratios and the probabilistic prediction of bankruptcy" Journal Of Accounting Research (Spring) 109-131.

PAWLAK Z., Rough sets. Int. J. Computer and Information Sci., 11, 1982, 341--356

PAWLAK,Z.; SLOWINSKI,R.(1994): "Decision analysis using rough sets. International Transactions in Operational Research 1, 107-114.

TRUJILLO, A. (2002): Gestión del Riesgo de Crédito en Préstamos Comerciales. Instituto Superior de Técnicas y Prácticas Bancarias. Madrid.

WILSON, T.(1997): "Portfolio Credit Risk", Risk magazine, Sept., 111-117.

ZAVGREN,C.V.(1983):"The prediction of corporate failure: the state of the art" Journal of Financial Lietrature 2, 1-37.