

## Diseño de un modelo específico para la predicción de la quiebra de *micro-entities*

BLANCO OLIVER, ANTONIO J.

Departamento de Economía Financiera y Dirección de Operaciones  
Universidad de Sevilla (España)  
Correo electrónico: aj\_blanco@us.es

IRIMIA DIÉGUEZ, ANA I.

Departamento de Economía Financiera y Dirección de Operaciones  
Universidad de Sevilla (España)  
Correo electrónico: anairimia@us.es

VÁZQUEZ CUETO, MARÍA JOSÉ

Departamento de Economía Aplicada III  
Universidad de Sevilla (España)  
Correo-e: pepi@us.es

### RESUMEN

La importancia de las *micro-entities* como generadoras de empleo y propulsores de la actividad económica conlleva, unida a sus mayores tasas de quiebra y a su dificultad para acceder a las fuentes de financiación, la necesidad de diseñar métodos apropiados que anticipen sus quiebras. Con este fin, en este trabajo se desarrolla un modelo híbrido mediante la combinación de enfoques paramétricos y no paramétricos para la detección de sus quiebras. Para ello, se seleccionan las variables con mayor poder predictivo para detectar la quiebra mediante un modelo híbrido de regresión logística (LR) y árboles de regresión y clasificación (CART). Nuestros resultados muestran que este modelo híbrido obtiene una mejor *performance* que aquellos modelos implementados de forma aislada, además de tener una más fácil interpretación y una convergencia más rápida. Por otra parte, se constata la conveniencia de la introducción de variables no financieras y macroeconómicas que complementen a la información proporcionada por los ratios financieros para la predicción de la quiebra de las *micro-entities*, lo cual está en línea con las características propias e idiosincrasia de este tamaño empresarial recientemente definido por la Comisión Europea.

**Palabras claves:** modelos de quiebra; métodos híbridos; métodos no paramétricos; árboles de decisión; *micro-entities*.

**Clasificación JEL:** G33; G41; G21; G32.

**MSC2010:** 20E08.

# Design of a Specific Model for Predicting Micro-Entities Failure

## ABSTRACT

The importance of micro-entities due to their generation of employment and propelling economic activity, together with the fact of their particularities, implies the need to design appropriate methods that anticipate their bankruptcies. For that purpose, a hybrid model by combining parametric and nonparametric approaches is developed in this paper. First, the variables with the highest predictive power to detect bankruptcy are selected using logistic regression (LR). Subsequently, a non-parametric method, namely regression trees and classification (CART), is then applied to companies classified as “bankruptcy” or “non-bankruptcy”. Our results show that this model provides a better result than when it is implemented in isolation, which joins its easier interpretation and faster convergence. Moreover, we demonstrate that the introduction of non-financial and macroeconomic variables complement the financial ratios for bankruptcy prediction. Findings are based on a data set of micro-entities (MEs), as recently defined by the European Union.

**Keywords:** bankruptcy models; hybrid methods; nonparametric methods; decision trees; micro-entities.

**JEL classification:** G33; G41; G21; G32.

**MSC2010:** 20E08.



## 1. Introducción

La reciente crisis financiera ha reactivado el interés por las investigaciones acerca de los modelos de predicción de quiebra. Por ello, en la literatura especializada se han desarrollado modelos muy variados, cada uno de ellos con sus debilidades y fortalezas, con la finalidad de incrementar la capacidad predictiva y la interpretabilidad de los actuales modelos. De hecho los métodos tradicionales, tales como el análisis discriminante o el *logit*, han sido ampliamente superados por la implementación de algoritmos de Minería de Datos (*Data Mining*), no sólo a nivel teórico (en cuanto a que las suposiciones tan restrictivas que exigían los primeros ya no son necesarias), sino también en el plano empírico.

Ahora bien, ante la resistencia a desaparecer de las técnicas estadísticas tradicionales (Caro *et al.*, 2013) y a la consideración de los métodos no paramétricos como algo ideal que nos proporciona resultados con escasa interpretabilidad (modelos de caja negra o *black-box*) empezaron a surgir los denominados métodos híbridos que, combinando lo mejor de cada uno de estos enfoques, construyen modelos paramétricos-no paramétricos de manera simultánea y/o correlativa y que obtienen mejor capacidad predictiva a la vez que proporcionan más información sobre cómo influye cada una de las variables consideradas en el resultado final. Es decir, mediante la implementación de modelos híbridos se pretende minimizar las debilidades de cada enfoque, aprovechando las sinergias entre ambos.

Este trabajo se alinea en esta dirección, utilizando un modelo híbrido que explota las ventajas del análisis *logit* y la teoría de los árboles de decisión (CART), creando sinergias y minimizando el coste asociado a la implementación de cada uno de los métodos separadamente. Por otra parte, creemos que es de destacar la aplicación que se realiza a un segmento homogéneo de empresas que recientemente han merecido –por sus características empresariales diferenciadas derivadas de su tamaño– una definición específica, como es el de las *micro-entities*, que representan el mayor porcentaje de la actividad empresarial en la Unión Europea y responsables del 75% de la actividad empresarial y aproximadamente del 30% del empleo en la Unión Europea. El presente trabajo constituye una de las escasas aplicaciones de un modelo diseñado específicamente para este tamaño empresarial y en el que se recoge la idiosincrasia de este segmento empresarial. Por último, y de acuerdo con las más recientes investigaciones y con las particularidades contables de las *micro-entities*, contrastamos la mejora del poder predictivo introduciendo variables no financieras y macroeconómicas como predictores de la quiebra en este tamaño de empresas.

Tras esta introducción, en la Sección 2, explicamos la metodología que utilizaremos con una muestra extraída de una población: las *micro-entities*, cuyas características expondremos en el apartado tercero. En la cuarta sección se definirán las variables utilizadas, y en la quinta se elaborará el modelo y se expondrán los resultados; terminando con las conclusiones en la sexta y última sección.

## 2. Metodología

Este estudio construye el modelo de quiebra en dos pasos. Primero, empleamos el método *logit*<sup>1</sup> para seleccionar las variables más relevantes, lo que nos permite también establecer las relaciones empíricas entre estos predictores y la quiebra de la microempresa (a través de los

---

<sup>1</sup> La principal razón para continuar utilizando *logit* en lugar de otros métodos de estimación es que proporciona unos resultados de precisión, eficiencia e interpretabilidad, apropiados (Crone y Finlay, 2012).

signos de sus coeficientes). A continuación, introduciendo solo dichas variables, implementamos la técnica no paramétrica CART. Desde un punto de vista teórico, este procedimiento nos permitirá reducir la dimensionalidad (principio de parsimonia) y acelerar la convergencia del árbol, así como mejorar la interpretabilidad y la precisión del modelo de quiebra resultante.

Para evaluar la mejora de nuestro modelo, utilizaremos el área bajo la curva ROC (AUC) que es uno de los indicadores más empleados en los problemas de clasificación. Siguiendo a West (2000), consideramos también dos medidas de *performance* adicionales: las probabilidades *a priori* de quiebra y los costes de clasificación (MC). El coste asociado al error tipo I (una empresa sin problemas financieros es clasificada erróneamente como quebrada) es obviamente inferior al coste del error tipo II (una empresa con problemas es mal clasificada como sana), pudiendo ser esta relación de 1:5.<sup>2</sup> Es por ello por lo que para evaluar la capacidad predictiva de un modelo debe prestarse especial atención al error tipo II. En los modelos de clasificación binarios, como es nuestro caso, el coste total de clasificación errónea viene dado por la siguiente expresión:

$$\text{Misclassification Cost (MC)} = C_{21}P_{21}\pi_1 + C_{12}P_{12}\pi_2 \quad (1)$$

donde  $\pi_1$  y  $\pi_2$  son, respectivamente, las probabilidades *a priori* de empresas sin y con problemas financieros;  $P_{21}$  y  $P_{12}$  miden la probabilidad de cometer los errores tipos I y II, respectivamente; y donde  $C_{21}$  y  $C_{12}$  son los correspondientes costes de clasificación erróneas de los tipos I y II. Para poder calcular el coste de clasificación errónea según se expresa en (1), utilizamos la estimación más común, que es la de la fracción de empresas sin problemas financieros clasificadas erróneamente para  $P_{21}$  y la fracción de empresas con problemas mal clasificadas para  $P_{12}$ .

El modelo *logit* se ha ajustado con la función *glm* en R, que calcula los estimadores de máxima verosimilitud de los  $n + 1$  parámetros a través de un algoritmo iterativo de mínimos cuadrados ponderados.

Para seleccionar los ratios más relevantes, seguimos los pasos establecidos por Altman y Sabato (2007). En primer lugar, una vez definidos y calculados los potenciales predictores, se calcula para cada uno el ratio de precisión (AR).<sup>3</sup> Para evitar los problemas de multicolinealidad entre las variables independientes, previamente se establecieron cinco categorías teóricas de ratios financieros (Apalancamiento, Liquidez, Rentabilidad, Actividad y Tamaño) y se consideró exclusivamente el ratio financiero con mayor AR de cada categoría. En segundo lugar, incorporamos junto a estos cinco ratios financieros más relevantes la información no financiera y macroeconómica. Para ello, utilizamos un procedimiento de selección *forward* en el modelo *logit*, resultando un conjunto de variables al que, de acuerdo con los objetivos del presente estudio, aplicamos la metodología de los árboles de decisión.

Un árbol de decisión (clasificación o regresión) es un conjunto de condiciones lógicas del tipo “si...entonces...” organizadas en un gráfico (Breiman *et al.*, 1984). De entre ellos, el modelo CART es el método más flexible para especificar la distribución condicional de la variable Y, dado un vector X de variables predictoras. Una de sus importantes ventajas en la predicción de quiebra es la habilidad para generar reglas de decisión fácilmente entendibles,

<sup>2</sup> Los costes asociados a los errores tipos I y II dependen de la decisión individual de los autores del modelo (Jones, 1987).

<sup>3</sup> Siguiendo a Engelmann *et al.*(2003), el ratio de precisión (AR) se calcula como  $2 \cdot (\text{AUC} - 0,5)$ .

siendo un método no paramétrico capaz de detectar relaciones complejas entre la variable dependiente y las predictoras. Esta característica no la comparten la mayoría de los métodos extraídos de los algoritmos de Minería de Datos y es una de las razones por la que nos hemos decantado por dicha técnica en la búsqueda de un modelo híbrido (*logit-CART*).

Bajo el método CART, comenzando por un nodo raíz al que pertenecen todas las observaciones, va conduciéndose cada una de ellas hacia un nodo terminal en el que cada observación es clasificada en una determinada categoría (empresa sana o fracasada). El árbol va creciendo mediante un procedimiento de partición binaria recursiva, seleccionando para dividir el nodo aquella variable predictora que maximiza la reducción en la suma residual de cuadrados para la variable de respuesta. Así, el conjunto de datos primitivo se divide en dos partes. El proceso se repite en cada uno de los nuevos nodos obtenidos (nodos hijos) hasta que estos son demasiado pequeños para ser divididos, llegando de este modo a los nodos terminales. Matemáticamente podemos decir que, siendo  $t$  un nodo interno, este se divide en dos nodos secundarios,  $t_d$  y  $t_i$  basándose en una característica  $X$  o una combinación lineal de ellas  $C(X_1, X_2, \dots, X_n)$  y un valor  $s$ . La característica  $X$ , o su combinación lineal, se selecciona de entre todas las existentes y el valor  $s$  se toma de tal manera que se minimice la heterogeneidad dentro de las dos submuestras resultantes.

Mediante el algoritmo CART, una vez seleccionadas las características que deben incluirse en el árbol, su estructura es automática. En cada nodo interno se busca el mejor valor para cada característica  $X_s$  y se queda con los que producen la menor cantidad de diversidad. La diversidad de un nodo está relacionada con el valor de la función de impureza del mismo (Breiman, 1988). Las funciones de impureza, como la función de Gini, se definen como  $g(t) = p_{1/t} p_{2/t}$ , donde  $p_{j/t}$  representa la proporción de casos que perteneciendo a la clase  $j$  ( $j= 1$  o  $2$ ) ha sido asignado al nodo  $t$ . La reducción del nivel de diversidad se mide por  $g(t) - p_d g(t_d) - p_i g(t_i)$  donde  $p_k$  es la proporción de casos que llegan al nuevo nodo  $t_k$  ( $k = d, i$ ). Los nodos se subdividen mientras que hay observaciones pertenecientes a varias clases y pueda reducirse el grado de diversidad.

Finalmente, cuando ya hemos construido el árbol  $T$  con  $t'$  nodos terminales, establecemos una regla para la asignación de cada nodo terminal a una clase y una estructura de coste de errores de clasificación para evaluar los resultados del árbol. El coste esperado de clasificación errónea para el árbol  $T$  se define como:

$$R(T) = \sum_{j=1}^2 \sum_{\substack{i=1 \\ i \neq j}}^2 c(i/j) \cdot q(i/j) \cdot p(j)$$

donde  $c(i/j)$  es el coste de clasificar una observación de la clase  $j$  en la clase  $i$ ,  $q(i/j)$  es la proporción de casos de la clase  $j$  mal clasificados en  $i$  y  $p(j)$  es una probabilidad *a priori* de pertenencia a la clase  $j$ . Esta es una estimación mínima del coste de clasificación errónea del árbol  $T$ .

Si el árbol resulta demasiado grande, se procede a su “poda”. La variable  $Rcv(T)^4$  se utiliza como un criterio de parada en la “poda” del árbol máximo. Para esto, se calcula el error estándar

---

<sup>4</sup>  $Rcv(T)$  es  $R(T)$  calculado en la muestra de validación.

de la variable ( $SE(T)$ ) y el árbol se va reduciendo (podando) hasta que se llega a un subárbol en el que  $Rcv$  tiene una desviación estándar próxima a  $SE(T)$ .<sup>5</sup>

### 3. Datos utilizados en la elaboración del modelo

#### 3.1. Descripción de la población: *micro-entities*

En 2012, el Parlamento de la Unión Europea (Directiva 2012/6/EU) redefine la clasificación de empresas según su tamaño, incluyendo a las *micro-entities* como una nueva categoría empresarial. Por ello, desde 2012 en la Unión Europea las PYMEs pueden ser clasificadas en cuatro grupos: *micro-entities*, *micro-enterprises*, empresas pequeñas y empresas medianas.<sup>6</sup>

Las características más importantes, y en consecuencia los problemas más relevantes asociados a este nuevo tamaño empresarial (*micro-entities*) pueden ser agrupados en las siguientes categorías:

1. De tipo financiero y legal. Las *micro-entities* han experimentado tradicionalmente fuertes dificultades de financiación, que han estado motivadas por el hecho de su opacidad respecto a la rendición de información financiera debido a que solo están obligadas a publicar cuentas anuales abreviadas y no auditadas (Berger y Frame, 2007). Otros factores que influyen en esta restricción financiera de las *micro-entities* son la incapacidad para acceder a los mercados de capitales, la ausencia de clasificación crediticias públicas y sus altas tasas de quiebra (Ciampi y Gordini, 2013).
2. Relacionados con su estructura de propiedad y tipo de empresa. Dado que la gestión es ejercida con frecuencia por los accionistas, las figuras de propietarios y gestores recaen en la misma persona. Este hecho implica que la fiabilidad de sus ratios financieros esté más en entredicho ya que sin los mecanismos de control pertinentes, la manipulación contable llega a ser usual (Claessens *et al.*, 2000).
3. Relacionados con su capacidad organizacional. Los recursos limitados para cumplir con los exigentes requisitos reglamentarios, el sometimiento a las mismas normas de información financiera que las grandes empresas y la concentración del negocio a nivel local o regional, sin o con actividad transfronteriza limitada, también son peculiaridades intrínsecas de las *micro-entities* que las diferencian del resto de tamaños empresariales.

Estos argumentos apoyan la consideración de las *micro-entities* como un nuevo tipo de empresa que requiere un tratamiento diferenciado. En respuesta a estas necesidades, la Directiva Europea 2012/6/UE no sólo creó este nuevo tamaño empresarial, microentidad, sino lo más importante, estableció un nuevo régimen de información financiera simplificada. Esta acción trató de reducir la carga administrativa motivada por la remisión de la información legal pertinente y de adecuar los requisitos de dicha información a las necesidades reales de los usuarios. Este nuevo régimen contable introdujo una serie de exenciones, de aplicación en el Reino Unido y en el resto de la Unión Europea, que favorecen a las *micro-entities* ya que solo necesitan presentar en el Registro Mercantil un sencillo balance con una pequeña información desglosada al pie. El beneficio económico de esta medida se estima en un ahorro de costes para

---

<sup>5</sup> El algoritmo nos permite elegir la regla xSERULE, donde x toma valores entre 0 y 1, indicativa de la desviación permitida al subárbol.

<sup>6</sup> Su importancia cuantitativa (BIS, 2013) y sus características homogéneas y específicas, justifican la creación de esta nueva clasificación que distingue a las *micro-entities* del grupo de las *micro-enterprises*.

las *micro-entities* de entre 5,9 y 6,9 billones de euros, lo que contribuye claramente a la mejora de la eficiencia, la competitividad y el crecimiento de las empresas de este tamaño.

En este contexto, se plantea la necesidad del desarrollo de modelos de fracaso diseñados específicamente para las características intrínsecas (sólo se utilizan datos financieros limitados contenidos en el nuevo régimen de información financiera en virtud de la Directiva antes mencionada) de las *micro-entities*. Las empresas muy pequeñas representan la mayor parte de la actividad económica en todo el mundo y han experimentado tradicionalmente una mayor probabilidad de fracaso que las grandes corporaciones (Carter y Van Auken, 2006). El modelo desarrollado en este trabajo intenta reducir el alto déficit informativo de que adolecen los accionistas, prestamistas y proveedores de las *micro-entities*, y por lo tanto, mejorar el proceso de toma de decisiones en el entorno de las empresas más pequeñas.

### **3.2. Determinación de la muestra**

Utilizamos un conjunto de datos proporcionado por una agencia de crédito del Reino Unido. Después de la depuración de datos perdidos y eliminación de los *outliers* (valores por encima y/o por debajo del 1% de cada ratio financiero), seleccionamos una muestra aleatoria de 39.710 *micro-entities* (50% quebradas y 50% no-quebradas) para el período 1999-2008.

Para estimar el error de predicción (error generalizado) de los modelos desarrollados seguimos la metodología de Hastie *et al.* (2009) y, al azar, particionamos el conjunto de datos en tres subconjuntos: un conjunto de entrenamiento, un conjunto de validación, y un conjunto de pruebas. El conjunto de entrenamiento se utiliza para la modelización, el conjunto de validación se utiliza para estimar el error de predicción del modelo y el conjunto de prueba para obtener el error generalizado del modelo finalmente seleccionado. Así, nuestro conjunto de datos fue aleatoriamente dividido en tres subconjuntos<sup>7</sup>: (i) conjunto de entrenamiento con el 60% de las observaciones, conjunto de validación<sup>8</sup> con el 20% de los casos y conjunto de test con el 20% restante (García-Gallego *et al.*, 2013).

## **4. Variables utilizadas en la elaboración del modelo**

### **4.1. Variable dependiente: Definición de fracaso empresarial**

En línea con otros estudios, se define el fracaso empresarial como entrada en liquidación, administración o administración judicial en el período analizado. Las cuentas analizadas para las empresas fallidas son las presentadas en el año anterior a la insolvencia. Para cada caso, la variable dependiente (fracaso empresarial) toma el valor 1 cuando la microentidad fracasa y 0 en caso contrario.

### **4.2. Variables explicativas**

La Tabla 1 describe las variables consideradas y la relación teórica con la situación de empresa quebrada.

**Tabla 1.** Variables financieras, no financieras y macroeconómicas.

<sup>7</sup> Otro argumento que justifica esta división es el gran número de datos (Hastie *et al.*, 2009).

<sup>8</sup> En el caso de la regresión logística el punto de corte óptimo se obtiene de la submuestra de validación.

Variable	Nomenclatura	Categoría	Relación teórica con la quiebra
<b>Ratios financieros</b>			
Capital empleado / Pasivo total	Celt	Apalancamiento	-
Pasivo a corto plazo / Activo total	Stlta	Apalancamiento	+
Pasivo total / Activo corriente	Tlca	Apalancamiento	+
Patrimonio neto / Activo total	Nwta	Apalancamiento	-
Activo disponible / Activo corriente	Qaca	Liquidez	-
Efectivo / Patrimonio neto	Cashnt	Liquidez	-
Activo corriente / Pasivo corriente	Cacl	Liquidez	-
Efectivo / Activo total	Cashta	Liquidez	-
Ganancias retenidas / Activo total	Rpta	Rentabilidad	-
Acreedores comerciales / Deudores comerciales	Tctd	Actividad	+
Acreedores comerciales / Pasivo total	Tctl	Actividad	+
Deudores / Activo total	Tdta	Actividad	+
Logaritmo neperiano del activo total	Ln_asset	Tamaño	+/-
Activo total	T_asset	Tamaño	+/-
<b>Variables no financieras y macroeconómicas</b>			
Cuentas auditadas	Audited	No (0)	+
		Sí (1)	-
Informe de auditoría favorable	Aq_clean	No (0)	+
		Sí (1)	-
Cambio auditor	Change_auditor	No (0)	-
		Sí (1)	+
Número de reclamaciones legales	Number_LCs		+
Valor de las reclamaciones legales	Value_LCs		+
Retraso en la presentación de las cuentas	Late_filing_day		+
Logaritmo neperiano de la edad	Ln_age		-
Activos pignorados	Charge_asset	No (0)	-
		Sí (1)	+
Empresa familiar	Family_firm	No (0)	-
		Sí (1)	+
Solvencia del sector	Industry_solvency		-

Dado que todos los ratios financieros utilizados en el presente estudio están disponibles públicamente para las *micro-entities* en sus cuentas anuales abreviadas, el modelo de quiebra resultante está adaptado íntegramente a la idiosincrasia de las mismas. Sin embargo, y dadas las razones intrínsecas ya comentadas de este tipo de empresas y su incidencia en la confiabilidad, parece razonable asumir que un modelo de quiebra no debe basarse únicamente en los ratios financieros. Por lo tanto, y como se sugiere en la literatura previa (véase Grunert *et al.*, 2005; Altman *et al.*, 2010; Wilson y Altanlar, 2013), también se incluyen variables explicativas no financieras sobre la presunción de que el uso combinado de ambas informaciones aumenta la precisión de los modelos construidos. Por último, también consideramos como variable independiente una variable macroeconómica –solvencia del sector– que mide la salud financiera del sector en el que opera la empresa y tiene una relación teórica inversa con la probabilidad de

quiebra. Varios estudios apoyan dicha relación demostrando una relación positiva entre un ciclo económico adverso y el número de quiebras empresariales (Moon y Sohn, 2010).<sup>9</sup>

En el Anexo A están recogidos los estadísticos descriptivos de estas variables para las empresas quebradas y no quebradas.

## 5. Elaboración del modelo y resultados

Como se ha comentado anteriormente, en primer lugar, aplicamos el análisis *logit* para elegir el ratio financiero de cada categoría teórica con mayor poder predictivo. Los resultados se presentan en la Tabla 2.

**Tabla 2.** Ratios financieros seleccionados.

Variable examinada	Categoría contable del ratio	AUC (%)	AR (%)	Variable seleccionada
Capital empleado / pasivo total	Apalancamiento	69,10	38,20	X
Pasivo a corto plazo / activo total	Apalancamiento	57,60	15,20	
Pasivo total / activo corriente	Apalancamiento	67,10	34,20	
Patrimonio neto / activo total	Apalancamiento	69,00	38,00	
Activo disponible / activo corriente	Liquidez	59,80	19,60	
Efectivo / patrimonio neto	Liquidez	51,10	2,20	
Activo corriente / pasivo corriente	Liquidez	66,70	33,40	
Efectivo / activo total	Liquidez	69,40	38,80	X
Ganancias retenidas / activo total	Rentabilidad	70,00	40,00	X
Acreedores comerciales / deudores comerciales	Actividad	57,20	14,40	
Acreedores comerciales / pasivo total	Actividad	54,40	8,80	
Deudores / activo total	Actividad	61,60	23,20	X
Logaritmo neperiano activo total	Tamaño	63,50	27,00	X
Activo total	Tamaño	63,20	26,40	

A continuación, construimos un primer modelo *logit* con estos cinco ratios financieros (modelo LR1) y, posteriormente, otro modelo (LR2), en el que se van incluyendo mediante el procedimiento *forward*, las variables no financieras y macroeconómica. Los resultados se presentan en la Tabla 3.

**Tabla 3.** Modelos logísticos de predicción de quiebra para *micro-entities*.

<sup>9</sup> Para un análisis detallado de las variables utilizadas, véase Altman *et al.* (2010).

Variable	Categoría	(LR1)			(LR2)		
		Coef.	Wald	Sig.	Coef.	Wald	Sig.
Capital empleado / pasivo total	Financiero	-0,054	179,9	0,00	-0,031	59,4	0,00
Efectivo / activo total	Financiero	-1,929	1477,	0,00	-1,504	781,3	0,00
Ganancias retenidas / activo total	Financiero	-0,385	834,9	0,00	-0,374	771,6	0,00
Deudores / activo total	Financiero	0,420	94,90	0,00	0,551	144,0	0,00
Logaritmo neperiano activo total	Financiero	0,804	1317,0	0,00	0,808	1175,4	0,00
Número de reclamaciones legales	NoFinanciero				1,681	695,2	0,00
Retraso en presentación de las cuentas	NoFinanciero				0,006	439,3	0,00
Ln Edad	NoFinanciero				-0,298	242,9	0,00
Empresa familiar	NoFinanciero				0,266	98,5	0,00
Solvencia industrial	NoFinanciero (Macroeconómico)				-0,626	508,4	0,00
Constante		-7,955	1183,3	0,00	-6,298	538,0	0,00
		AUC=0,770 MC=0,851			AUC=0,806 MC=0,843		

Nuestros resultados muestran que el AUC del modelo que incluye las variables no financieras (LR2) es del 80,6%, superior al que sólo contiene los ratios financieros (LR1) como variables predictoras (77,0%). Se obtienen resultados similares cuando se analizan los costes esperados de clasificación errónea. En este caso, nuestros resultados demuestran que el uso combinado de las variables financieras y no financieras (LR2) reduce el coste estimado de error un 0,8% (de 0,851 hasta 0,843) en comparación con el uso de sólo los ratios financieros (LR1). Por tanto, en consonancia con otros autores (Whittred y Zimmer, 1984; Peel *et al.*, 1986; Altman *et al.*, 2010), sugerimos que la información no financiera aporta un valor añadido al modelo con una mejora de más del 3,5% en términos de la AUC y una reducción del 0,8% del coste de error estimado.

Estos resultados confirman nuestra presunción teórica que afirma que parece razonable asumir que un modelo de quiebra adecuado y diseñado específicamente para las *micro-entities* no debe basarse únicamente en indicadores financieros, y que las variables no financieras y macroeconómicas desempeñan un importante papel a la hora de evaluar la capacidad predictiva de los modelos de quiebra para las *micro-entities*, lo cual está en línea con sus características contables y administrativas analizadas con anterioridad.

Por último, y partiendo de las variables que ha introducido el modelo con mayor poder predictivo (LR2), aplicamos el algoritmo CART.

Mediante el uso de la función de la impureza de Gini, las probabilidades previas observadas en la muestra, igual coste de clasificación errónea para ambos grupos, y la regla OSERULE, obtenemos doce árboles con sus costes de validación y de sustitución asociados. El mejor árbol contiene veintiocho nodos, y costes de validación y sustitución de 0,54868 (+/- 0,00587) y 0,47748, respectivamente (Anexo B). Además, el software empleado para construir este modelo determina la relevancia relativa de cada variable dentro de la construcción del árbol, y de esta manera, determina las variables de alerta temprana en la que las empresas deben actuar para evitar la quiebra. Nuestro modelo ofrece la siguiente clasificación: Rpta (100,00%),

Celt (94,14%), Cashta (79,64%), Late\_filing\_days (47,68%), Number\_LCs (38,48%), Industry\_solveny (31,65%), Tdta (27,96%), Ln\_asset (16,20%), Ln\_age (1,31%), y Family\_firm (0,85%).

En la muestra de entrenamiento, CART obtuvo una tasa de clasificación correcta promedio (CCR) del 76,18%, y errores Tipo I-II de 24,67% y 22,97%, respectivamente. El AUC es 0,816. En la muestra de test, el CCR es 72,63%, y los errores tipo I y II son 26,65% y 28,09%, respectivamente, y las AUC iguales a 0,771.

En la Tabla 4, se presentan los resultados obtenidos en términos de AUC, clasificación correcta (CCR), los errores de Tipo I y II, y los costes de clasificación errónea (MC) de los modelos logísticos elaborados y del modelo híbrido construido con CART, tanto para la muestra de entrenamiento como en la muestra de test.

**Tabla 4.** AUC, tasa de clasificación correcta, errores tipo I y II, y costes de clasificación errónea.

Técnica estadística	Modelo	Muestra de entrenamiento					Muestra de test				
		AUC	CCR (%)	Tipo I (%)	Tipo II (%)	MC	AUC	CCR (%)	Tipo I (%)	Tipo II (%)	MC
Regresión logística	LR1	0,736	70,22	31,49	29,05	0,885	0,770	70,74	30,97	27,77	0,851
	LR2	0,809	74,08	24,54	29,54	0,863	0,806	72,99	24,83	28,69	0,843
Árbol de clasificación / regresión	Logit + CART	0,816	76,18	24,67	22,97	0,707	0,771	72,63	26,65	28,09	0,781

En base a estos resultados, sugerimos que el modelo híbrido Logit+CART proporciona una precisión similar, considerando las diferentes medidas en su conjunto, al enfoque LR solo, encontrando, además, que disminuyen los errores Tipo II y los costes de clasificación erróneas en el modelo híbrido.

## 6. Conclusiones

Desde una perspectiva metodológica, nuestros resultados apoyan el desarrollo de modelos de quiebra utilizando de forma conjunta variables financieras, no financieras y macroeconómicas, además de la implementación de técnicas estadísticas híbridas. En primer lugar, el valor añadido de la información no financiera y macroeconómica es importante dada la escasez de datos financieros a disposición del público para las *micro-entities* en el marco del nuevo régimen de información financiera establecido en la reciente Directiva 2012/06 de la UE. En segundo lugar, y como consecuencia de la inclusión en el método no paramétrico de sólo las variables de entrada más relevantes del modelo paramétrico, se produce una aceleración de la tasa de convergencia de las técnicas estadísticas no paramétricas que justifican claramente la aplicación de estos modelos híbridos para predecir el fracaso. Por último, consideramos especialmente relevante la posibilidad que ofrece el desarrollo del modelo de distinguir las variables que explican el fracaso, lo cual es una importante ventaja desde el punto de vista técnico. Este conocimiento previo nos permite anticipar y tomar las medidas adecuadas para mejorar la situación financiera de las *micro-entities*.

Esto tiene implicaciones económicas relevantes para los prestamistas, las *micro-entities* y los supervisores bancarios, entre otros. Para los prestamistas porque representa una reducción de

la asimetría de la información en sus relaciones con las *micro-entities* y pueden controlar más eficazmente el riesgo de crédito en este sector (uno de sus principales clientes). Desde el punto de vista de las *micro-entities*, los modelos de quiebra proporcionan información crucial sobre la salud financiera de una empresa para los inversores, gerentes y auditores, y presentan una ayuda muy útil al tomar la decisión de invertir, la detección de problemas internos, y la clasificación de la empresa en términos de riesgo de solvencia. Para los supervisores bancarios, la aplicación de un modelo de fracaso que incluye las variables financieras y no financieras es una fuente de apoyo para determinar los requerimientos de capital.

## Referencias

- Altman E.I. & Sabato, G. (2007). Modeling credit risk for SMEs: Evidence from US market. *A Journal of Accounting, Finance and Business Studies (ABACUS)*, 43(3), 332-357.
- Altman, E.I., Sabato, G. & Wilson, N. (2010). The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management. *Journal of Credit Risk*, 6(2), 95-127.
- BIS (2013). Department for Business, Innovations and Skills. Simpler financial reporting for micro-entities: The UK's proposal to implement the 'micros directive', [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/86259/13-626-simpler-financial-reporting-for-micro-entities-consultation.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/86259/13-626-simpler-financial-reporting-for-micro-entities-consultation.pdf). Accessed on 24 September 2014.
- Breiman, L. (1988). Submodel selection and evaluation in regression: the x-fixed case and little bootstrap. Technical Report No. 169, Statistics Department, U.C. Berkeley.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). Classification and regression trees. Wadsworth, Belmont, CA.
- Berger A.N. & Frame, S.W. (2007). Small business credit scoring and credit availability. *Journal of Small Business Management*, 45(1), 5-22.
- Caro, N, Díaz, M. & Porporato, M. (2013). Predicción de quiebras empresariales en economías emergentes: uso de un modelo logístico mixto. *Journal of Quantitative Methods for Economics and Business Administration*, 16, 200-215.
- Carter, R. & Van Auken, H. (2006). Small firm bankruptcy. *Journal of Small Business Management*, 44(4), 493-512.
- Ciampi, F. & Gordini, N. (2013). Small enterprise default prediction modeling through artificial neural networks: An empirical analysis of Italian small enterprises. *Journal of Small Business Management*, 51(1), 23-45.
- Claessens, S., Djankov, S. & Lang, L.H. (2000). The separation of ownership and control in East Asian corporations. *Journal of Financial Economics*, 58, 81-112.
- Crone, S.F. & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, 28(1), 224-238.
- Engelmann, B., Hayden, L.E. & Tasche, D. (2003). Testing rating accuracy. *Risk*, 16, 82-86.

- García-Gallego, A. & Mures-Quintana, M.J. (2013). The Sample of Firms in Business Failure Prediction Models: Influence on Classification Results. *Journal of Quantitative Methods for Economics and Business Administration*, 14, 131-150
- Grunert, J., Norden, L. & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, 29(2), 509-531.
- Hastie, T., Tibshirani, R. & Friedman, J.H. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer Series in Statistics. New York.
- Wilson, N., Wright, M. & Scholes, L. (2013). Family business survival and the role of boards. *Entrepreneurship Theory and Practice*, 37(6), 1369-1389.

## ANEXO A

**Tabla A.1.** Estadísticos descriptivos de las variables explicativas cuantitativas

Variables	Quebrada		No quebrada	
	Media	Desv. típica	Media	Desv. típica
Capital empleado / pasivo total	0,45	1,32	1,77	4,95
Pasivo a corto plazo / activo total	0,12	0,23	0,06	0,17
Pasivo total / activo corriente	3,24	5,55	2,45	5,40
Patrimonio neto / activo total	-0,70	1,81	0,41	1,19
Activo disponible / activo corriente	0,81	0,29	0,88	0,26
Efectivo / patrimonio neto	4,55	5,66	2,91	4,59
Activo corriente / pasivo corriente	1,17	2,56	2,35	4,31
Efectivo / activo total	0,15	0,22	0,37	0,34
Ganancias retenidas /activo total	-0,56	1,43	0,01	0,03
Acreedores comerciales / deudores comerciales	6,67	17,25	12,66	22,84
Acreedores comerciales / pasivo total	0,84	0,27	0,85	0,30
Deudores / activo total	0,43	0,31	0,31	0,31
Logaritmo neperiano activo total	10,36	0,55	10,08	0,61
Activo total	36.312,26	16.952,53	28.585,85	16.637,25
Número de reclamaciones legales	0,31	0,85	0,03	0,09
Valor de las reclamaciones legales	1.519,40	4.756,56	64,76	214,70
Retraso en la presentación de las cuentas	32,59	82,89	18,92	69,01
Logaritmo neperiano de la edad	7,48	0,68	7,51	1,08
Solvencia industrial	-0,07	0,24	0,18	0,52

**Tabla A.2.** Estadísticos descriptivos de las variables explicativas cualitativas

<b>Variable</b>	<b>Categoría</b>	<b>Status</b>	<b>Frecuencia (%)</b>
Activos pignorados	No (0)	Quebrada	46,75
		No quebrada	49,40
	Sí (1)	Quebrada	3,25
		No quebrada	0,60
Empresa familiar	No (0)	Quebrada	28,30
		No quebrada	25,56
	Sí (1)	Quebrada	21,70
		No quebrada	24,44
Cuentas auditadas	No (0)	Quebrada	47,26
		No quebrada	48,14
	Sí (1)	Quebrada	2,74
		No quebrada	1,86
Informe favorable del auditor	No (0)	Quebrada	48,10
		No quebrada	48,32
	Sí (1)	Quebrada	1,90
		No quebrada	1,68
Cambio de auditor	No (0)	Quebrada	47,40
		No quebrada	47,63
	Sí (1)	Quebrada	2,60
		No quebrada	2,37

