

RIESGO CREDITICIO EN LAS PYME

Desarrollo de un modelo de predicción de default

Aldo H. Alonso

*Universidad Nacional de La Plata
Universidad del Centro Provincia de Buenos Aires*

Santiago Barraza

Gustavo Guglielmetti

Ana Legato

Universidad del Centro Provincia de Buenos Aires

SUMARIO: 1. Introducción; 2. Elección de la metodología de medición de riesgo; 3. Selección de las variables explicativas; 4. El modelo logit o logístico; 5. Resultados de la regresión; 6. Conclusiones.

Para comentarios: aldohalonso@yahoo.com.ar
santiago_barraza@hotmail.com
gustavoguglielmetti@hotmail.com
legato@econ.unicen.edu.ar

Resumen. El presente trabajo tiene por objeto desarrollar un modelo de predicción de default para las PYME del centro de la Provincia de Buenos Aires. A partir de metodologías de análisis de riesgo ampliamente conocidas y validadas en empresas de capital abierto, se determina un modelo de predicción adaptado a las PYME., utilizando información disponible en la región.

La investigación utiliza una base de datos creada ad hoc formada en parte por empresas concursadas y/o quebradas. A partir de esta información, se desarrolla un modelo que proporciona un *score* que puede asociarse con la probabilidad de default de la empresa en el año próximo al análisis.

La metodología consiste en la propuesta de variables explicativas del modelo, la eliminación sistemática de aquellas que no tengan un poder explicativo adecuado y la regresión logística como modo de determinar el modelo.

Los conceptos vertidos en la primera etapa de estudio pretenden ser innovadores en cuanto al ámbito de aplicación. Dicha metodología es similar a la ya utilizada en otros países para pequeñas y medianas empresas.

El modelo logrado tiene un poder de predicción del 71,4% en la muestra de desarrollo e indica que las variables más relevantes en la predicción de default de las PYME. son los ratios de Créditos por Ventas/Deuda, Disponibilidades/Activo y Patrimonio Neto/Deuda.

El impacto deseado del modelo es la reducción de los costos de asimetría de información en el proceso de financiación a empresas.

1. INTRODUCCIÓN

El objetivo del presente trabajo es desarrollar un modelo de predicción de default para las pequeñas y medianas empresas del centro de la Provincia de Buenos Aires. El interés surge por el rol protagónico que cumple este tipo de empresas en la estructura económica de la región y sus dificultades reveladas en el acceso a financiamiento en condiciones competitivas¹. Parte de estos problemas son explicados por las altas tasas requeridas por los proveedores de fondos, las cuales, a su vez, se relacionan con importantes asimetrías de información.

Existe la necesidad de una adecuación de las metodologías de análisis de riesgo en empresas abiertas -las cuales han sido típicamente objeto de estudio- a las PYME. Ello en reconocimiento de las naturales diferencias entre ambos tipos de empresas y en virtud de investigaciones recientes² que señalan la diferente capacidad de predicción de los modelos de análisis de riesgo crediticio, según el segmento de empresas donde se aplique.

La investigación identifica las variables más relevantes en la predicción de concurso de acreedores y/o quiebra de las empresas y desarrolla un modelo que proporciona un *score* asociado con la probabilidad de ocurrencia de tal evento en el año próximo al análisis. La investigación utiliza una base de datos creada ad hoc conteniendo más de cincuenta empresas de la región, parte de ellas concursadas y/o quebradas (en adelante, eventualmente las llamaremos ‘fallidas’).

Las PYME bajo análisis se corresponden a un rango de ventas dentro de los topes fijados por la Resolución N° 675/02 de la Secretaría de Pequeña y Mediana Empresa y Desarrollo Regional³ para el período 1996 – 2005.

La metodología consiste en la propuesta y prueba de las variables explicativas más utilizadas en la literatura de modelos de scoring crediticio, la eliminación sistemática de aquellas variables que no tengan un poder explicativo adecuado y la regresión logística como modo de determinar el modelo de predicción definitivo.

Los conceptos vertidos en la primera etapa de estudio pretenden ser innovadores en cuanto al ámbito de aplicación. Dicha metodología es similar a la ya utilizada en otros países en pequeñas y medianas empresas.

El impacto deseado del modelo es la reducción de los costos originados por la asimetría de información en el proceso de financiación a empresas.

2. ELECCIÓN DE LA METODOLOGÍA DE MEDICIÓN DE RIESGO

Tradicionalmente se han desarrollado distintos sistemas de medición del riesgo crediticio, todos ellos basados en conjuntos de variables consideradas relevantes para la predicción de default. Estos sistemas han sido más o menos técnicos y pueden ser comprendidos en tres clases: sistemas expertos, de rating y de *credit-scoring*⁴.

Los sistemas expertos basan sus criterios de decisión en el ‘expertise’, los conocimientos y los juicios del analista de crédito. Uno de los más conocidos es el de las cinco ‘C’ del crédito,

¹ En el trabajo Finanzas aplicadas en Tandil: evidencias empíricas de los mismos autores, el 76% de las empresas de la región manifiesta tener problemas en el acceso a la financiación. Encuentran también inconvenientes para generar y presentar la información solicitada por los proveedores de fondos, como así también para conseguir las garantías solicitadas y ajustar los plazos requeridos al ofrecido por las entidades bancarias.

² El estudio Modeling credit risk for SME: Evidence from the US market de E. Altman y G. Sabato es un buen ejemplo al respecto.

³ Según definición de Pequeña y Mediana Empresa (PYME.) las ventas anuales no pueden superar \$ 43.200.000 para las empresas del sector de industria y minería, \$ 86.400.000 para comercio, \$ 10.800.000 para el sector agropecuario y \$ 21.600.000 para servicios.

⁴ Para una más amplia reseña de estos modelos puede verse Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms de Anthony Saunders, Wiley, 1999.

las cuales representan Carácter, Capital, Capacidad, Colateral y Ciclo económico del sujeto evaluado. Estos sistemas presentan, básicamente, dos problemas: (1) falta de consistencia en el tratamiento de distintos prestatarios potenciales y (2) subjetividad en el peso relativo otorgado a las variables analizadas.

Los sistemas de rating se usan para calificar distintos instrumentos de deuda utilizando categorías asociadas a niveles de calidad crediticia y son muy difundidos para la colocación de instrumentos de deuda de los estados y las empresas abiertas.

Los sistemas de scoring tienen como objeto identificar factores claves en la determinación de la probabilidad de default y los ponderan en un *score* cuantitativo. El grado de éxito esperado del *score* es validado con información histórica y resulta totalmente objetivo. El modelo generado en el presente trabajo se encuadra dentro de estos sistemas.

3. SELECCIÓN DE LAS VARIABLES EXPLICATIVAS

El objetivo de la investigación es construir una ecuación con aquel subconjunto de las variables explicativas que más información aporte sobre las probabilidades de pertenecer a un grupo de empresas que entrará en default o a un grupo de empresas que no lo hará.

Previo a la aplicación de la metodología seleccionada se realiza la ‘depuración’ de las variables explicativas candidatas mediante el uso de la matriz de correlaciones y el índice de Gini. Posteriormente se ingresa en la etapa de obtención del modelo donde la herramienta de selección definitiva de variables es el método Forward y los estadísticos a utilizar son Puntuación eficiente de Rao y estadístico de Wald.

En el caso en estudio la variable a pronosticar es la variable dicotómica Fallida/No Fallida (Y) y el subconjunto de variables explicativas candidatas está formado por quince razones financieras (X) representadas en el cuadro 1.

Cuadro 1. Razones financieras candidatas a ingresar en el modelo, agrupadas por grupo de afinidad

RAZÓN FINANCIERA	GRUPO DE PERTENENCIA
Patrimonio Neto / Deuda Patrimonio Neto / Activo Deuda / Activo	Leverage
Disponibilidades / Activo Activo Corriente / Activo Capital de Trabajo Neto / Activo Disponibilidades / Ventas Intangibles / Activo	Liquidez
GAI / Ventas GAI / Activo Ingreso Neto / Activo Ingreso Neto / Ventas	Rentabilidad
Ventas / Activo Pasivo Corriente / Ventas Crédito por Ventas / Pasivo	Actividad

En un paso preliminar se elimina la razón INTANG/A por tener valor distinto de 0 en solamente uno de los individuos de la muestra.

Luego, se hace el análisis de la matriz de correlaciones de los quince ratios a efectos de detectar posibles colinealidades⁵. Así, resultan eliminadas en esta instancia las razones de DISP/VTAS y GAI/A. La primera de ellas, por estar altamente correlacionada con DISP/A y PCTE/VTAS; la segunda, por estarlo con INGNETO/A y GAI/VTAS.

A su vez, se encuentra una alta correlación entre los ratios D/A, PN/A y PN/D, lo cual es predecible. Para determinar cual de estos ratios puede resultar más explicativo en el modelo se construye el índice de Gini, el cual reporta el poder explicativo univariado de cada uno de ellos⁶. Así se decide la eliminación de las razones D/A y PN/A de la lista de potenciales variables explicativas.

Al final de esta etapa de depuración cinco variables explicativas candidatas resultan eliminadas y se da ingreso al modelo de regresión de diez covariables candidatas.

4. EL MODELO LOGIT O LOGÍSTICO

En general el *modelo logit* relaciona a la variable Y_i con las variables explicativas X_1, \dots, X_p a través de la siguiente ecuación

$$Y_i = \frac{1}{1 + e^{-z}} + e_i \quad \text{Ec 1}$$

donde Y_i es una variable dicotómica que puede tomar valores 0 ó 1.

Es decir que a partir de $(X_{i1}, \dots, X_{ip}), i = 1, \dots, n$, (muestra de n observaciones de las variables independientes X_1, \dots, X_p) en los dos grupos de individuos establecidos por los dos valores de la variable dependiente Y , se trata de obtener una combinación lineal de las variables independientes que permita estimar las probabilidades de que un individuo pertenezca a cada una de las dos sub-poblaciones o grupos. La probabilidad de que un individuo pertenezca a la segunda sub-población, p , vendrá dada por:

$$p = \frac{e^z}{1 + e^z} \quad \text{Ec 2}$$

o, equivalentemente,

$$p = \frac{1}{1 + e^{-z}} \quad \text{Ec 3}$$

siendo Z la combinación lineal

$$Z = \beta_1 X_1 + \dots + \beta_p X_p + \beta_0 \quad \text{Ec 4}$$

donde $\beta_0, \beta_1, \dots, \beta_p$ son parámetros desconocidos a estimar. En particular, la probabilidad de que el i -ésimo individuo de la muestra pertenezca a la segunda sub-población será:

$$p_i = \frac{1}{1 + e^{-(\beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_0)}} \quad \text{Ec 5}$$

⁵ Para mayor detalle, ver anexo I.

⁶ Para mayor detalle, ver anexo II.

Si dicha probabilidad es superior o igual a 0,5, el individuo será clasificado en la segunda sub-población; en caso contrario, será clasificado en la primera.⁷

A partir de b_0, b_1, \dots, b_p , las probabilidades estimadas de que un individuo pertenezca a las sub-poblaciones segunda y primera son, respectivamente:

$$\hat{p} = \frac{1}{1 + e^{-z}} \quad \text{Ec 6}$$

y

$$\hat{q} = 1 - \hat{p} \quad \text{Ec 7}$$

donde:

$$\hat{Z} = \beta_1 X_1 + \dots + \beta_p X_p + \beta_0 \quad \text{Ec 8}$$

Teniendo en cuenta que:

$$\ln \left(\frac{\hat{p}}{\hat{q}} \right) = Z \quad \text{Ec 9}$$

Una expresión alternativa para el modelo de regresión logística es:

$$\frac{\hat{p}}{\hat{q}} = e^{\beta_0} (e^{\beta_1})^{X_1} \dots (e^{\beta_p})^{X_p} \quad \text{Ec 10}$$

Este cociente se conoce como *odd ratio*, y compara la probabilidad de que ocurra un suceso respecto de la probabilidad de que no ocurra. Concepto a utilizar en el momento de interpretar los coeficientes.

Los coeficientes β son en realidad las medidas del cambio en el ratio de probabilidad. Cuanto mayor sea el coeficiente β_i mayor será el cociente entre las probabilidades y, en consecuencia, mayor será la probabilidad de pertenecer al segundo grupo.

5. RESULTADOS DE LA REGRESIÓN

En el cuadro 2 se presentan los resultados de la regresión logística por el método Forward y las variables incluidas en la ecuación.

Al final del proceso, tres variables explicativas han resultado seleccionadas para ingresar al modelo, ellas son: Créditos por Ventas/Deuda, Disponibilidades/Activo y Patrimonio Neto/Deuda.

⁷ Estadísticamente, la lectura a hacer para un valor $p_i > 0,5$ es que la probabilidad de la que la empresa i pertenezca al segundo grupo -por caso, empresas no fallidas- es mayor a la probabilidad de que pertenezca al primero. Nótese que, en el ejercicio de la calificación crediticia, una entidad de crédito tiene poder de decisión en base al valor que tome esta variable, lo cual se expresa en su política de otorgamiento de crédito. La firma puede, por ejemplo, decidir prestar sólo a aquellos solicitantes con un *score* mayor a 0,70. Este valor es llamado *cut-off point* y es una herramienta fundamental en la gestión del crédito cuando se utiliza modelos de scoring. Con esto se incrementan las probabilidades de no financiar a empresas potencialmente fallidas.

Cuadro 2. Variables incluidas en la ecuación

		B	E.T.	Wald	gl	Sig	Exp(B)
Paso 1(a)	CREDVTAS/D	3,491	1,509	5,356	1	,021	32,835
	Constante	-1,056	,565	3,491	1	,062	,348
Paso 2(b)	DISP/A	16,781	8,887	3,566	1	,059	19399314,301
	CREDVTAS/D	3,932	1,698	5,363	1	,021	51,006
	Constante	-1,997	,763	6,845	1	,009	,136
Paso 3(c)	PN/D	,693	,520	1,774	1	,183	2,000
	DISP/A	18,347	8,968	4,185	1	,041	92870051,297
	CREDVTAS/D	2,951	1,856	2,526	1	,112	19,118
	Constante	-2,310	,819	7,949	1	,005	,099

a Variable(s) introducida(s) en el paso 1: CREDVTAS/D.

b Variable(s) introducida(s) en el paso 2: DISP/A.

c Variable(s) introducida(s) en el paso 3: PN/D.

d Se ha detenido un procedimiento por pasos ya que al eliminar la variable menos significativa se obtuvo un modelo previamente ajustado.

Reemplazando en la ecuación 10, la estimación del odd ratio adopta la siguiente expresión:

$$\frac{\hat{p}}{\hat{q}} = e^{-2,310} (e^{0,693})^{PN/D} (e^{18,347})^{DISP/A} (e^{2,951})^{CREDVTAS/D}$$

$$\frac{\hat{p}}{\hat{q}} = 0,099 * 2,000^{PN/D} * 92896822,52^{DISP/A} * 19,125^{CREDVTAS/D} \quad Ec 11$$

La ecuación 11 es el modelo estimado. A continuación se analiza con el mismo un ejemplo de una empresa con los siguientes ratios: PN/D = 0,83; DISP/A = 0,07 y CREDVTAS/D = 0,20.

A tal empresa le corresponderá un odd ratio como el siguiente:

$$\frac{\hat{p}}{\hat{q}} = 0,099 * 2,000^{0,83} * 92896822,52^{0,07} * 19,125^{0,20}$$

$$\frac{\hat{p}}{\hat{q}} = 0,099 * 1,778 * 3,612 * 1,804$$

$$\frac{\hat{p}}{\hat{q}} = 1,1472$$

Para la empresa, la probabilidad de no fallar durante el próximo período es 1,1472 veces la probabilidad de fallar.

Operando con la ecuación 7, tenemos que:

$$\frac{\hat{p}}{(1 - \hat{p})} = 1,1472$$

$$\hat{p} = 1,1472 - 1,1472 \hat{p}$$

$$\hat{p} = 0,5343$$

Este es el *score* asignado e implica una probabilidad de 53,43% de pertenecer al grupo de no fallidas.

Los errores del modelo. El porcentaje de casos correctamente clasificados es un índice de la efectividad del modelo. Si el modelo es efectivo sobre la muestra observada, es de esperar que también lo sea cuando se trate de clasificar a un individuo para el que se desconoce a cuál de los dos grupos pertenece.

En el cuadro 3 puede observarse como mejora la clasificación a través de los pasos.

Cuadro 3. Tabla de clasificación

Observado			Pronosticado		
			FALL/NO FALL		Porcentaje correcto
			1	2	
Paso 1	FALL/NO FALL	1	16	7	69,6
		2	8	18	69,2
	Porcentaje global				69,4
Paso 2	FALL/NO FALL	1	17	6	73,9
		2	9	17	65,4
	Porcentaje global				69,4
Paso 3	FALL/NO FALL	1	17	6	73,9
		2	8	18	69,2
	Porcentaje global				71,4

a El valor de corte es ,500.

Los modelos de predicción pueden generar dos tipos de errores conocidos como errores de Tipo I y II⁸. Los errores de Tipo I ocurren cuando el modelo asigna baja probabilidad de default (alta calificación crediticia) a un prestatario que en realidad luego incumplirá con sus obligaciones financieras. Los errores de Tipo II, en cambio, consisten en que el modelo asigna alta probabilidad de default (baja calificación crediticia) a empresas que luego cumplen con sus obligaciones.

En el caso bajo estudio, el modelo genera en la muestra de desarrollo 14 errores⁹: 6 de Tipo I y 8 de Tipo II. El modelo identifica correctamente al 73,9% de las fallidas¹⁰ (comete error del Tipo I en el 26,1% de los casos) y al 69,2% de las empresas no fallidas (comete error del Tipo II en el 30,8% de los casos¹¹). Esta diferenciación resulta relevante dado que el costo de cometer un error de Tipo I implica la pérdida del capital prestado (o parte de este); en cambio, los errores de Tipo II sólo perjudican al prestamista en la pérdida del negocio crediticio.

⁸ Esto es una convención en las pruebas de modelos de scoring. Para más detalle ver J. Sobehart y S. Keenan 2001, *Measuring default accurately*.

⁹ En el Anexo III se observa el resumen de clasificación.

¹⁰ Este es el denominado 'Hit Rate' –o HR(C)– del modelo el cual se estima como $HR(C) = \frac{H(C)}{H(C) + M(C)}$,

donde HR(C): hit rate; H(C): número de hits, o la cantidad de casos en que el modelo clasifica correctamente a un incumplidor como tal; M(C) número de misses, o los casos en que el modelo asigna bajo riesgo a un incumplidor. H + M es el total de incumplidores.

¹¹ Este es el denominado *False Alarm Rate* o FAR(C) del modelo que se determina como

$FAR(C) = \frac{F(C)}{F(C) + R(C)}$, donde F(C): número de falsas alarmas, o los casos en que se señala como incumplidores a clientes de bajo riesgo y R(C): número de rechazos correctos, o casos en que el modelo correctamente asigna bajo riesgo a clientes de alta calidad crediticia. F + R es el total de cumplidores.

Bondad de ajuste del modelo. Comprobar la bondad de ajuste del modelo es analizar cuán probables son los resultados muestrales a partir del modelo ajustado. La probabilidad de los resultados observados se denomina verosimilitud y se basa en comparar el número de individuos observado en la segunda subpoblación con el número esperado si el modelo fuese válido.

Para comprobar que la verosimilitud no difiere de 1 (que el modelo se ajusta a los datos) se utiliza el estadístico $-2LL = -2 \times \text{Logaritmo de la verosimilitud}$ que, bajo la hipótesis nula de que el modelo se ajusta perfectamente a los datos, sigue una χ^2 con $n-2$ grados de libertad.

El valor del estadístico para el modelo se muestra en el cuadro 4.

Cuadro 4. Pruebas ómnibus sobre los coeficientes del modelo

		Chi-cuadrado	Gl	Sig.
Paso 1	Paso	7,603	1	,006
	Bloque	7,603	1	,006
	Modelo	7,603	1	,006
Paso 2	Paso	7,320	1	,007
	Bloque	14,922	2	,001
	Modelo	14,922	2	,001
Paso 3	Paso	1,939	1	,164
	Bloque	16,862	3	,001
	Modelo	16,862	3	,001

Y es significativo al nivel del 5%. Ello significa que no hay razones suficientes para rechazar la hipótesis nula de que el modelo es adecuado.

Acompañamos a los resultados anteriores con estadísticos que hablan del poder explicativo del modelo, incluido el R cuadrado de Nagelkerke, con lo cual se puede decir que el modelo explica el comportamiento de la variable Fall/NO FALL al 40%.

Cuadro 5. Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Nagelkerke
1	60,142(a)	,192
2	52,822(b)	,350
3	50,883(b)	,389

a La estimación ha finalizado en el número de iteración 5 porque las estimaciones de los parámetros han cambiado en menos de ,001.

b La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de ,001.

6. CONCLUSIONES

En el ámbito de trabajo de la investigación (centro de la Provincia de Buenos Aires) se reconoce a las PYME estudiadas con características propias y claramente diferentes de aquellas de capital abierto y de otras pequeñas y medianas de países desarrollados, donde se ha elaborado gran parte de los modelos de scoring divulgados.

Existen suficientes evidencias empíricas de limitaciones y dificultades en el acceso al crédito para las empresas bajo estudio, problema que en muchos casos se origina en la asimetría de información entre prestador y prestatario, así como en la falta de herramientas para la evaluación consistente de la calidad crediticia del solicitante. Las entidades financieras resuelven este últi-

mo punto aplicando metodologías genéricas que pueden no ajustarse a las características propias de las PYME de la región.

En este contexto, resulta relevante el modelo de predicción de default obtenido. No obstante, varias opciones fueron descartadas a medida que se avanzaba con las pruebas de eficiencia.

El trabajo desarrollado permite contar con un modelo de predicción de default para empresas, el cual resulta efectivo en el 71,4% de los casos dentro de la muestra de desarrollo. Los estadísticos son significativos.

Las metodologías estadísticas utilizadas son ampliamente reconocidas y su uso se ajusta a las necesidades del caso.

Por último, el poder predictivo del modelo combinado con un *cut-off point* eficiente puede resultar en una herramienta valiosa en el análisis crediticio, incrementando así la calidad de la gestión del crédito.

REFERENCIAS

- Altman, E. I. and Sabato, G., *Modeling Credit Risk for SME: Evidence from the US Market*, Enero 2006.
- Medina, F., *Consideraciones sobre el índice de Gini para medir la concentración del ingreso*, CEPAL, marzo 2001
- Pulido, A. y López, A. M., *Predicción y simulación aplicada a la economía y gestión de empresas*, Pirámide, 1999
- Saunders, A., *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*, Wiley, 1999
- Sobehart, J. and Keenan, S., Measuring default accurately, *Credit Risk Special Report*, 2001
- Sulla Misura della Concentrazione e della Variabilità dei Caratteri. Tai del R. Istituto Veneto di Scienza, *Lettere ed Arti*, pág. 1203 –1248, 1914.

ANEXO I

ANÁLISIS DE CORRELACIONES DE LAS VARIABLES EXPLICATIVAS CANDIDATAS

En el cuadro siguiente se muestran las correlaciones entre las variables analizadas.

	PN/D	PN/A	D/A	Disp/A	AC/A	NWC/A	Disp/Vtas	UAIT/Vtas	UAIT/A	IngNet/A	IngNet/Vtas	Vtas/A	PCte/Vtas	CRED VTAS/D
PN/D	1	,895(**)	-,895(**)	-,223	,011	,566(**)	-,139	,051	,005	,121	,057	-,145	-,155	,580(**)
PN/A	,895(**)	1	-1,000(**)	-,316(*)	-,073	,597(**)	-,231	,084	,114	,248	,078	-,098	-,246	,453(**)
D/A	-,895(**)	-1,000(**)	1	,316(*)	,073	-,597(**)	,231	-,084	-,114	-,248	-,078	,098	,246	-,453(**)
Disp/A	-,223	-,316(*)	,316(*)	1	,333(*)	-,008	,712(**)	,208	,205	,263	,244	,215	,621(**)	-,054
AC/A	,011	-,073	,073	,333(*)	1	,662(**)	,175	,457(**)	,439(**)	,564(**)	,576(**)	,297(*)	,067	,408(**)
NWC/A	,566(**)	,597(**)	-,597(**)	-,008	,662(**)	1	-,058	,457(**)	,448(**)	,621(**)	,510(**)	,086	-,173	,560(**)
Disp/Vtas	-,139	-,231	,231	,712(**)	,175	-,058	1	,145	-,041	,024	,136	-,141	,961(**)	-,198
UAIT/Vtas	,051	,084	-,084	,208	,457(**)	,457(**)	,145	1	,723(**)	,658(**)	,939(**)	,059	-,020	,206
UAIT/A	,005	,114	-,114	,205	,439(**)	,448(**)	-,041	,723(**)	1	,861(**)	,653(**)	,406(**)	-,150	,249
IngNet/A	,121	,248	-,248	,263	,564(**)	,621(**)	,024	,658(**)	,861(**)	1	,694(**)	,358(*)	-,078	,355(*)
IngNet/Vtas	,057	,078	-,078	,244	,576(**)	,510(**)	,136	,939(**)	,653(**)	,694(**)	1	,177	-,057	,284(*)
Vtas/A	-,145	-,098	,098	,215	,297(*)	,086	-,141	,059	,406(**)	,358(*)	,177	1	-,245	,198
PCte/Vtas	-,155	-,246	,246	,621(**)	,067	-,173	,961(**)	-,020	-,150	-,078	-,057	-,245	1	-,282(*)
CREDVTAS/D	,580(**)	,453(**)	-,453(**)	-,054	,408(**)	,560(**)	-,198	,206	,249	,355(*)	,284(*)	,198	-,282(*)	1

* La correlación es significativa al nivel 0,05 (bilateral).

** La correlación es significativa al nivel 0,01 (bilateral).

ANEXO II

ÍNDICE DE GINI

El índice de Gini es una medida de concentración que tiene como objetivo fundamental cuantificar el grado de desigualdad en el reparto o distribución de una magnitud económica. La aplicación en este trabajo busca identificar las razones financieras que resumen una mayor concentración en el conjunto total de las empresas a efectos de rescatar los indicadores que mejor desagregan al conjunto de empresas en fallidas y no fallidas.

El índice de concentración de Gini se define como:

$$I_G = \frac{\sum_{i=1}^{N-1} (p_i - q_i)}{\sum_{i=1}^{N-1} p_i} = 1 - \frac{\sum_{i=1}^{N-1} q_i}{\sum_{i=1}^{N-1} p_i}$$

cumpliendo que:

- Caso de **mínima concentración o máxima igualdad** ($p_i = q_i \quad \forall i = 1, \dots, N-1$): $I_G = 0$
- Caso de **máxima concentración o mínima igualdad** ($q_i = 0 \quad \forall i = 1, \dots, N-1$): $I_G = 1$
- Casos intermedios: cuanta mayor concentración se detecte, más cerca de 1 se situará el valor del índice, y cuanta mayor igualdad haya en el reparto más cerca de 0 estará.

Por tanto, $0 \leq I_G \leq 1$.

En el cuadro siguiente se observan los índices de Gini para las tres razones de endeudamiento, donde se observa que la variable que presenta mayor concentración es la de PN/D.

Razón de endeudamiento	Índice de Gini
PN/D	0,530
PN/A	0,348
D/A	0,210

ANEXO III

RESUMEN DE LA CLASIFICACIÓN

En el cuadro siguiente se muestran los resultados de la clasificación.

Elemento	Grupo observado	Probabilidad pronosticada	Grupo pronosticado
1	1	,37185	1
2	1	,20971	1
3	1	,38430	1
4	1	,11533	1
5	1	,73624	2
6	1	,14540	1
7	1	,39282	1
8	1	,61400	2
9	1	,18038	1
10	1	,19438	1
11	1	,38792	1
12	1	,42003	1
13	1	,30826	1
14	1	,51974	2
15	1	,60526	2
16	1	,33405	1
17	1	,11743	1
18	1	,37902	1
19	2	,97588	2
20	1	,81339	2
21	1	,29782	1
22	1	,30736	1
23	1	,26638	1
24	1	,60221	2
25	2	,22132	1
26	2	,99085	2
27	2	,41899	1
28	2	,51650	2
29	2	,87761	2
30	2	,45218	1
31	2	,53532	2
32	2	,76804	2
33	2	,42773	1
34	2	,27350	1
35	2	,79895	2
36	2	,70908	2
37	2	,31924	1
38	2	,51959	2
39	2	,58495	2
40	2	,97446	2
41	2	,99396	2
42	2	,96899	2
43	2	,88988	2
44	2	,98076	2
45	2	,94728	2
46	2	,72309	2
47	2	,79768	2
48	2	,26316	1
49	2	,36773	1