

Francisco Javier Castaño Gutiérrez



**VARIABLES Y MODELOS PARA LA EVALUACIÓN
DEL FRACASO EMPRESARIAL. PROPUESTA
DE UNA METODOLOGÍA DE FRONTERAS BASADA
EN PERCENTILES**

X Premio «ASEPUC» de Tesis Doctorales

**VARIABLES Y MODELOS PARA LA EVALUACIÓN
DEL FRACASO EMPRESARIAL. PROPUESTA
DE UNA METODOLOGÍA DE FRONTERAS BASADA
EN PERCENTILES**

X Premio «ASEPUC» de Tesis Doctorales (2014)

Francisco Javier Castaño Gutiérrez
Universidad de León

VARIABLES Y MODELOS PARA LA EVALUACIÓN DEL FRACASO EMPRESARIAL. PROPUESTA
DE UNA METODOLOGÍA DE FRONTERAS BASADA EN PERCENTILES

Edita: © Instituto de Contabilidad y Auditoría de Cuentas
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Edición en línea: Año 2015
NIPO en línea: 724-15-006-3
ISBN: 978-84-606-8058-1

*A todos aquellos que me han apoyado, porque me han ayudado a terminar
A todos aquellos que me han criticado porque me han obligado a mejorar
A todos aquellos que me dieron su aliento porque me empujaron para seguir
A todos los que me plantearon dudas, porque me facilitaron las decisiones
A todos, muchas gracias...
...pero sobre todo gracias a Pilar*

AGRADECIMIENTOS

La primera persona a la que quiero agradecer su apoyo, esfuerzo, tiempo, conocimientos, paciencia y amistad es, sin duda, a la directora de la tesis y amiga María Teresa Tascón Fernández. Soy consciente del enorme número de horas que me ha dedicado y por ello nunca podré agradecerle lo suficiente ese esfuerzo. Muchísimas gracias, Mayte.

Esta tesis es el resultado del trabajo en equipo entre doctorando y directora, pero no habría sido posible sin la contribución de una serie de personas e instituciones:

A la empresa *Iberinform*, que confió en nuestra capacidad de desarrollar un modelo de fronteras para evaluar el fracaso cuando solo era un proyecto y firmó con nosotros el contrato de investigación que nos ha proporcionado los datos necesarios para llevar a cabo todos los estudios empíricos que se incluyen en la tesis.

A la *Fundación Centro de Supercomputación de Castilla y León (FCSCL)*, porque nos permitió cientos de horas de uso en uno de sus ordenadores. Gracias a ello, podemos presentar resultados con la metodología de análisis envolvente de datos (DEA).

A la *Biblioteca de la Universidad de León* y, en especial, a los profesionales de la *Biblioteca de la Facultad de Económicas*. Las bases de datos bibliográficas y el apoyo eficaz para conseguir lo que faltaba en ellas, han facilitado mucho el trabajo.

A los compañeros del Área de Economía Financiera y Contabilidad que nos han ayudado a definir los problemas y a encontrar las soluciones, Carmen Fernández Cuesta y Borja Amor Tapia.

A Lubor Homolka, que revisó los planteamientos y razonamientos matemáticos de la segunda parte de la tesis e inspiró la formulación inicial de nuestro propio modelo.

A los dos evaluadores anónimos y el editor asociado de la *Revista de Contabilidad (RC-SAR)*, así como a los participantes en el *XV Congreso AECA*, donde se

presentó una versión previa de la primera parte de esta tesis, por sus valiosos comentarios y sugerencias.

A nuestra contraponente en el *IX Workshop on Empirical Resesarch in Financial Accounting*, Anne-Mie Reheul; a nuestros dos evaluadores anónimos en el *35th Annual Congress* de la *European Accounting Association (EAA)*; así como a los asistentes a ambas sesiones de presentación por los comentarios y sugerencias aportados a nuestras comunicaciones, que eran versiones anteriores de la segunda parte de la tesis. Destacamos entre las aportaciones de los asistentes, los comentarios de Joachim Gassen en el *Workshop*.

Finalmente, tenemos que agradecer el apoyo financiero prestado a este proyecto por la *Universidad de León (Proyecto de investigación ULE-2010-9)*.

*El camino es la meta, si sabes plantear correctamente
una pregunta entonces también sabrás encontrar tú mismo la respuesta
Alejando Magno (Gisbert Haefs)*

ÍNDICE

RESUMEN.....	19
INTRODUCCIÓN.....	21
1. VARIABLES Y MODELOS PARA LA IDENTIFICACIÓN Y PREDICCIÓN DEL FRACASO EMPRESARIAL: REVISIÓN DE LA INVESTIGACIÓN EMPÍRICA RECIENTE	27
1.1. Introducción a la primera parte	27
1.2. Fracaso empresarial, modelos y variables explicativas.....	29
1.2.1. De la teoría a la evidencia empírica: el concepto de fracaso y el planteamiento de modelos	29
1.2.2. Variables explicativas	36
1.3. Revisión de metodologías	46
1.3.1. Análisis univariante	47
1.3.2. Análisis discriminante	48
1.3.3. Regresión logística: logit y probit.....	50
1.3.4. Inteligencia artificial.....	52
1.3.4.1. Redes neuronales.....	52
1.3.4.2. Algoritmo de particiones recursivas	55
1.3.4.3. Técnicas de escalamiento multidimensional	56
1.3.4.4. Otros métodos de inteligencia artificial.....	56
1.3.5. Análisis envolvente de datos (DEA).....	57
1.3.6. Otros modelos.....	59
1.4. Conclusiones a la primera parte	62
2. SELECCIÓN DE VARIABLES EN EL ANÁLISIS DEL FRACASO EMPRESARIAL: SELECCIÓN DE MEDIAS VS. SELECCIÓN DE MEDIANAS	67
2.1. Introducción a la segunda parte	67
2.2. Variables en la literatura previa sobre pymes	69
2.3. Metodologías que se van a aplicar.....	72
2.3.1. Análisis discriminante lineal (LDA)	73
2.3.2. Análisis discriminante cuadrático (QDA).....	75
2.3.3. Análisis Discriminante Logístico (LogDA)	76

2.3.4. Análisis discriminante del K-ésimo vecino más próximo.....	77
2.3.5. Logit.....	78
2.3.6. Probit.....	80
2.3.7. Análisis Envolvente de Datos (DEA)	81
2.4. Los datos y la selección estadística de variables	86
2.5. Evaluación del fracaso empresarial.....	99
2.6. Análisis de datos winsorizados.....	104
2.7. Conclusiones a la segunda parte	113
3. DESCRIPCIÓN DEL FRACASO EMPRESARIAL MEDIANTE FRONTERAS DE RATIOS	
CONTABLES SELECCIONADAS CON DIFERENCIAS DE PERCENTILES	119
3.1. Introducción	119
3.2. Revisión del fracaso empresarial en pymes	122
3.3. Propuesta metodológica	129
3.4. Estudio empírico: una aplicación al sector de la construcción.....	139
3.4.1. La elección de los ratios más discriminantes y la frontera gráfica de quiebra	141
3.4.2. Distancia a una frontera de quiebra calculada con scores de ratios financieros	144
3.4.3. Situación relativa de empresas concretas.....	147
3.4.4. Evolución de los indicadores financieros de empresas que fracasan	154
3.4.5. Evolución de empresas en riesgo de fracaso que se recuperan	162
3.5. Conclusiones a la tercera parte.....	166
4. CONCLUSIONES FINALES.....	171
BIBLIOGRAFÍA.....	177
ANEXOS.....	197
Anexo I Variables explicativas	199
Anexo II Trabajos y metodologías	205
Anexo III Variables explicativas pymes.....	211
Anexo IV Instrucciones programa STATA	217
Anexo V Salida de resultados del programa STATA	229
Anexo VI Histogramas de frecuencias.....	341

Anexo VII Selección de ratios para DEA e identificación de inputs-outputs	345
Anexo VIII Representación gráfica de medianas de ratios de empresas españolas fracasadas en 2010. Sector de la construcción. Periodo 2008-2010	347
Anexo IX Representación de los valores de los ratios, de 2008 a 2010, para las 28 empresas fracasadas en 2010	349

TABLAS

Tabla 1. Ejemplos de criterios declarados para la selección de ratios financieros	37
Tabla 2. Ratios más frecuentes	40
Tabla 3. Factores económicos subyacentes a las variables.....	40
Tabla 4. Trabajos pioneros en distintas metodologías.....	47
Tabla 5. Ratios más frecuentes en pymes.....	70
Tabla 6. Factores económicos subyacentes a las variables en pymes	71
Tabla 7. Panel A. Matriz de correlaciones.....	91
Panel B. Matriz de correlaciones con ratios winsorizados	91
Tabla 8. Panel A. Estadísticos descriptivos.....	92
Panel B. Estadísticos descriptivos, ratios winsorizados	93
Tabla 9. Estadísticos descriptivos, fracasadas/no fracasadas, diferencias en medias	95
Tabla 10. Estadísticos descriptivos, fracasadas/no fracasadas, Rank sum test	98
Tabla 11. Análisis Discriminante Lineal (LDA)	99
Tabla 12. Análisis Discriminante Cuadrático (QDA)	100
Tabla 13. Análisis Discriminante Logístico (LogDA).....	100
Tabla 14. Análisis Discriminante del Knn vecino más próximo (KnnDA).....	101
Tabla 15. Análisis logit y probit	102
Tabla 16. Estadísticos descriptivos, fracasadas/no fracasadas diferencias en medias, ratios winsorizados	105
Tabla 17. Estadísticos descriptivos fracasadas/no fracasadas, Rank sum test, ratios winsorizados	106
Tabla 18. Análisis Discriminante Lineal (LDA), ratios winsorizados	107
Tabla 19. Análisis Discriminante Cuadrático (QDA), ratios winsorizados	108
Tabla 20. Análisis Discriminante Logístico (LogDA), ratios winsorizados.....	108
Tabla 21. Análisis Discriminante del K-ésimo vecino más próximo (KnnDA), ratios winsorizados	109
Tabla 22. Análisis logit y probit, ratios winsorizados	110
Tabla 23. DEA. Rendimiento constante a escala (CRS), ratios winsorizados.....	111
Tabla 24. Scores del grupo de empresas fracasadas con 5 y 10 ratios en 2006, 2007 y 2008.	147
Tabla 25. Scores Gqy ejemplos fracasadas y no fracasadas, 2006.....	149
Tabla 26. Scores Gqy ejemplos fracasadas y no fracasadas, 2007.....	152

Tabla 27. Medianas de ratios para empresas fracasadas y para empresas no fracasadas.	155
Tabla 28. Percentiles de ratios para empresas fracasadas.....	156
Tabla 29. Ratios con mayor poder predictivo (diferencias de percentiles >0,25).....	158
Tabla 30. Scores del grupo de empresas fracasadas en 2010 respecto a un grupo de no fracasadas estable, con 5 y 10 ratios en 2008, 2009 y 2010.	159
Tabla 31. Tabla empresas con igual o superior puntuación cada año que las fracasadas en 2010.....	160
Tabla 32. Tabla empresas con igual o superior puntuación que S(q) en 2010.....	161
Tabla 33. Scores de fracaso en 2010 y scores y distancias del grupo a en 2008, 2009 y 2010.	164

GRÁFICOS

Gráfico 1. Fronteras DEA.....	82
Gráfico 2. Segmentos de frontera.....	84
Gráfico 3. Niveles de scores	136
Gráfico 4. Niveles de distancias al fracaso	139
Gráfico 5. Fracaso empresarial. Empresas constructoras de Castilla y León. 2006	142
Gráfico 6. Fracaso empresarial. Empresas constructoras de Castilla y León. 2007	142
Gráfico 7. Fracaso empresarial. Empresas constructoras de Castilla y León. 2008	143
Gráfico 8. Fracaso empresarial. Empresas constructoras de Castilla y León. 2006. Caracterización de empresas concretas	148
Gráfico 9. Fracaso empresarial. Empresas constructoras de Castilla y León. 2007. Caracterización de empresas concretas	152
Gráfico 10. Percentiles de ratios. Evolución 2008-10	157
Gráfico 11. Fracaso empresarial. Grupo a. Empresas constructoras. España. 2008.....	165

RESUMEN

Este trabajo se ocupa de la selección de variables y de la aplicación de modelos para la identificación y predicción del fracaso empresarial en empresas pequeñas. La primera parte analiza la evolución en el tiempo de los estudios sobre fracaso empresarial. Con carácter general, partimos de la revisión crítica realizada en la literatura previa, y aportamos un análisis de la evidencia empírica adicional, con especial atención a la obtenida durante la última década. Pero además, para subsanar algunas deficiencias detectadas en las revisiones anteriores, nos ocupamos de tres aspectos, que pueden considerarse la principal contribución de esta parte: primero, analizamos la evolución en las últimas décadas del concepto de fracaso empresarial o fallido, detectando cierta evolución desde la identificación hacia la predicción; segundo, analizamos las variables empleadas en los modelos, aportando un estudio de los rasgos empresariales que se representan con las variables (frente al tradicional análisis de frecuencia de las propias variables individuales), siendo los resultados más acordes con los planteamientos y desarrollos teóricos clásicos sobre el fracaso empresarial; y, finalmente, destacamos los puntos fuertes y débiles de las metodologías que, por su reciente aparición, no habían sido analizadas o muy poco por revisiones anteriores: las técnicas de inteligencia artificial y el análisis envolvente de datos (DEA). Adicionalmente, integramos en la revisión el numeroso grupo de trabajos empíricos publicados en España sobre la cuestión, y que no aparecían en ninguna de las revisiones previas analizadas.

La segunda parte profundiza en la selección de variables, que constituye uno de los procesos determinantes en la evaluación del fracaso empresarial. Después de comprobar que el primer nivel de selección de variables, a partir de la literatura previa, es adecuado al grupo de estudio, las empresas pequeñas, se lleva a cabo un segundo nivel de selección mediante contrastes estadísticos, usando tanto las diferencias en medias como las diferencias en medianas. Dado que las variables resultantes de cada contraste difieren notablemente, se aplica un variado grupo de métodos de evaluación del fracaso (LDA, QDA, LogDA, KNNDA, logit, probit, y DEA) con el fin de identificar las implicaciones de usar uno u otro contraste.

Nuestros resultados indican que la naturaleza de la muestra determina no solo el contraste de selección de variables sino también los métodos de evaluación de fracaso más apropiados, lo que constituye nuestra principal contribución en esta parte. Además, aportamos nuevos resultados empíricos sobre el papel de la información cualitativa en este tipo de modelos, línea donde la evidencia previa sobre pymes es aún escasa.

La tercera parte propone una metodología novedosa de cálculo de *scores* y distancias al fracaso para el análisis en pymes. La metodología propuesta se basa en diferencias estadísticas entre el grupo de empresas fracasadas y el conjunto de la población al que pertenece (sector, periodo y zona geográfica determinados). Los resultados de selección de las variables más discriminantes, y de predicción de fracaso son consistentes con la literatura previa. Además, la metodología permite calcular distancias a la quiebra de empresas particulares o grupos de empresas, así como identificar los indicadores financieros que constituyen puntos fuertes o débiles de la empresa (o grupo de empresas) estudiada a efectos de su potencial reorganización.

Palabras clave

Fracaso empresarial, quiebra, análisis de variables, ratios financieros, pymes, análisis discriminante, probit, logit, DEA, información cualitativa, fronteras con diferencias de percentiles.

Clasificación JEL

G17, G33, G34, L25, M41.

INTRODUCCIÓN

En la actual situación financiera de incertidumbre, donde el riesgo de crédito parece seguir una escalada creciente, el aumento de los derechos de cobro impagados es causa y efecto del fracaso empresarial. La crisis ha renovado el interés por estos aspectos del riesgo de crédito que en años de bonanza económica parecían superados por los modelos cuantitativos sofisticados para el cálculo de probabilidades de fallido y exposiciones al fallido. En este momento de cambios, el orden de prioridades es distinto y de nuevo tiene más sentido investigar sobre el fracaso empresarial: cómo detectarlo, para poder anticiparlo, y, en su caso, poner las medidas correctoras necesarias para evitar o minimizar sus efectos.

El cambio de normativa contable que se ha producido en 2008 en España se enmarca en la adopción de las normas internacionales de contabilidad por parte de la Unión Europea, lo cual se suma a la creciente utilización de este grupo de normas por países de todo el mundo. Las normas internacionales de contabilidad se están convirtiendo en un potente instrumento de armonización internacional¹, con las ventajas que ello supone a efectos de la comparabilidad de la información. Un marco común de normas contables para empresas cotizadas posibilita la extensión de las muestras a zonas geográficas amplias, permitiendo no sólo la comparación de empresas de varios países; y los estudios de verificación de diferencias entre países detectadas cuando las variables contables no eran homogéneas²; sino también los estudios por sectores y otro tipo de grupos de empresas, que no podían hacerse por países por el reducido número de empresas en ese sector o grupo concreto.

¹ En la cumbre de Londres del G-20 celebrada en abril de 2009, se aprobó el apartado núm. 15 que solicita a los encargados de establecer normas contables que colaboren urgentemente con supervisores y reguladores para mejorar los criterios de valoración y provisión, y alcanzar un conjunto de normas contables mundiales de alta calidad.

² La línea de investigación contable sobre diferencias internacionales en las normas y en la aplicación de las mismas se ocupará de determinar si las normas internacionales comunes se aplican de formas distintas por áreas de influencia. Pero sin duda habrá un mayor grado de armonización.

En el caso del riesgo de crédito, y concretamente del fracaso empresarial, el cambio de normas puede modificar la importancia de las variables contables dentro de los modelos, lo que requeriría como mínimo una serie de trabajos empíricos que validen o modifiquen las referencias anteriores. Pero la convergencia internacional de la normativa tiene un efecto aún más relevante. La posibilidad de realizar estudios para zonas supranacionales con normas contables homogéneas proporciona a los investigadores mejores condiciones para el análisis específico de sectores y subsectores concretos: entidades de depósito, empresas eléctricas, grandes superficies comerciales, empresas constructoras, etc. cuya población en muchos países no tiene el tamaño crítico suficiente para aplicar las metodologías cuantitativas y que los resultados sean estadísticamente significativos. El efecto de este nuevo entorno normativo debería ser contrastado con nueva evidencia empírica.

Una cuestión subyacente en esta línea de investigación es la ausencia de una teoría firme sobre el fracaso empresarial. A pesar del profuso número de trabajos empíricos y teóricos sobre la cuestión, y de la creciente sofisticación de los métodos de medición, no se ha conseguido desarrollar el mapa conceptual de relaciones explicativas entre el comportamiento económico de las empresas, medido con variables internas, fundamentalmente contables, el efecto de los factores externos, tanto sectoriales como macroeconómicos, medidos todos ellos a través de las variables disponibles, y la quiebra efectiva o la posibilidad de quiebra de esas empresas. Esto tiene repercusiones sobre la implementación práctica de los modelos de evaluación del fracaso. Por ejemplo, son cuestiones determinantes qué concepto de fracaso debe considerarse, cuáles son las variables explicativas que se deben incluir o si el modelo debe ser de identificación o de predicción.

Esta tesis está dividida en tres partes, cada una de las cuales se construye sobre los hallazgos y contribuciones de la anterior. En la primera parte, se hace un análisis de la evolución en el tiempo del concepto de fracaso, de los modelos propuestos para evaluarlo y de las variables encontradas discriminantes por esos modelos, con los siguientes objetivos:

- Integrar los pasos que se han dado para avanzar en esa teoría sobre el fracaso, aún en proceso de construcción, prestando especial atención a las metodologías de más reciente introducción en esta línea de investigación, como son la inteligencia artificial y el análisis envolvente de datos, e incluyendo una exhaustiva revisión de los trabajos publicados en España sobre la cuestión;
- Seleccionar las variables que con mayor frecuencia han resultado significativas en la literatura previa; y
- Detectar las ventajas e inconvenientes de las principales metodologías empleadas hasta el momento, para identificar aquellas más adecuadas a nuestro estudio empírico posterior.

Si, pese a los cambios de escenario en el tiempo y en el espacio, se pueden detectar elementos comunes y regularidades sobre qué origina, cómo y cuándo el fracaso, es posible situar algunos elementos en la base del mapa conceptual.

Como resultado del análisis teórico y comparativo de los trabajos previos, se han encontrado diferencias en el concepto de fracaso, que comportan importantes ajustes en el planteamiento de los modelos y en la interpretación de los resultados; se ha encontrado también cierta convergencia en las variables con poder explicativo, en base a lo cual realizamos dos estudios de frecuencia, uno de variables individuales y otro de rasgos económicos subyacentes a esas variables; la matización que este estudio sobre rasgos introduce en el *ranking* de las variables más significativas puede considerarse la primera contribución de esta tesis; se ha encontrado además que, para evitar las carencias de las metodologías iniciales, se han ido proponiendo modelos revisados o metodologías nuevas, cada una con sus propias limitaciones. Dos carencias comunes a todas esas metodologías son: la utilización de resultados dicotómicos y la escasa consideración de factores ajenos a la propia empresa evaluada.

En la segunda parte, se complementa el análisis previo de selección de variables, realizado a partir de los resultados obtenidos por la literatura empírica, con un estudio sobre la selección estadística de variables. Nuestro objetivo es escoger,

entre el contraste de diferencias en medias y el contraste de diferencias en medianas, cuál es el método más adecuado para la evaluación de fracaso en empresas pequeñas. Para ello, los diferentes grupos de variables seleccionadas se incorporan a modelos de amplia difusión, tales como el análisis discriminante lineal, el análisis discriminante cuadrático, el logit y el probit, pero también a otros no utilizados previamente en esta línea de investigación, tales como el análisis discriminante logístico y el análisis discriminante del K-ésimo vecino más próximo, o aplicados solo recientemente, como el análisis envolvente de datos. Un segundo objetivo de esta parte es comprobar cuáles de estas metodologías consiguen mejor desempeño en la evaluación de empresas pequeñas. El último objetivo de la segunda parte es comprobar la utilidad de una variable cualitativa, incidencias de pago, en la evaluación del fracaso de las empresas de pequeño tamaño.

Nuestros resultados aconsejan descartar las diferencias en medias y utilizar las diferencias en medianas para seleccionar estadísticamente las variables, dada la gran dispersión de las variables en la muestra. En cuanto a las metodologías, parece que la distribución de los datos sesga los modelos paramétricos de análisis discriminante y los modelos binomiales, de manera que la selección de medianas consigue mejores resultados con metodologías no paramétricas o semi-paramétricas. Los resultados con DEA están en la línea de los escasos trabajos previos. Y se confirma de manera inequívoca la contribución de la variable cualitativa, que resulta significativa en ambos contrastes de diferencias y obtiene mejoras en todos los modelos utilizados.

En la tercera parte, nuestro objetivo es desarrollar una metodología nueva que permita soslayar las dos limitaciones generales identificadas en las metodologías previas: la simple dicotomía de las clasificaciones (fracasada / no fracasada) y el aislamiento de los modelos respecto al entorno. Proponemos un modelo que calcula distancias en percentiles entre los ratios financieros y permite seleccionar las variables más discriminantes para la población objeto de estudio; asigna puntuaciones (*scores*) por empresa (o por grupo de empresas) que nos permiten cuantificar tanto su distancia a la quiebra como su distancia al grupo de referencia de empresas más saludables; determina los indicadores financieros (ratios)

concretos que la empresa debe mejorar, y en cuánto, para alcanzar una situación financiera normal o, mejor aún, una situación financiera saludable. Los resultados obtenidos con esta metodología completamente novedosa muestran una capacidad discriminante muy buena en comparación con las aplicadas en la segunda parte de la tesis, pero además esta metodología ofrece posibilidades que no tienen las anteriores, constituyendo la principal contribución de la tesis.

1. VARIABLES Y MODELOS PARA LA IDENTIFICACIÓN Y PREDICCIÓN DEL FRACASO EMPRESARIAL: REVISIÓN DE LA INVESTIGACIÓN EMPÍRICA RECIENTE

1.1. Introducción a la primera parte

En nuestra opinión, la crisis que ha afectado a un elevado número de países, sin que muchos de ellos hayan conseguido todavía superarla, junto con el cambio de normas contables que mencionábamos en la introducción general de la tesis, va a dar lugar en los próximos años a una nueva etapa de trabajos sobre fracaso empresarial. Para afrontarla, parece necesario delimitar primero el estado de la cuestión. Aunque existen excelentes trabajos previos de revisión, tales como Zavgren (1983), Jones (1987), Keasey y Watson (1991), Altman (1993), Balcaen y Ooghe (2006), Ravi Kumar y Ravi (2007), Bellovary *et al.* (2007) y Aziz y Dar (2006), los cuatro primeros resultan parciales por no recoger las últimas dos (o incluso tres) décadas de evidencia, mientras que los tres siguientes se ocupan de parcelas concretas de esa evidencia previa y el último no tiene en cuenta las variables utilizadas en los trabajos previos. En esta parte de la tesis, vamos a tratar de complementar los estudios anteriores sobre la revisión de la literatura que existe en esta línea de investigación.

El primer objetivo es analizar la evolución en el tiempo de los estudios sobre fracaso empresarial. Para ello, partimos de la revisión crítica de la literatura que hacen otros autores, para poner de manifiesto los puntos fuertes y débiles de los modelos planteados. En este sentido, nuestra aportación consiste en la revisión de evidencia empírica adicional, con especial atención a la obtenida durante la última década. Además, intentamos subsanar algunas deficiencias detectadas en las revisiones anteriores. Así, nos ocupamos de tres aspectos, que pueden considerarse la principal aportación de esta parte: primero, analizamos la evolución en las últimas décadas del concepto de fracaso empresarial o fallido;

segundo, analizamos las variables empleadas en los modelos, lo que implica el cuestionamiento de su relación con una teoría cada vez más positiva y menos normativa del fracaso empresarial y, finalmente, prestamos especial atención a las metodologías que por su reciente aparición no habían sido analizadas o muy poco por revisiones anteriores, tales como las técnicas integradas dentro de la inteligencia artificial y el análisis envolvente de datos (DEA). Nuestro tercer objetivo es integrar en la revisión el numeroso grupo de trabajos empíricos publicados en España sobre la cuestión, y que no aparecen en ninguna de las revisiones citadas.

A continuación, hemos realizado un proceso de selección de variables. Partimos de una amplia selección de trabajos empíricos anteriores, donde identificamos las variables que han resultado significativas con más frecuencia, siendo nuestros resultados consistentes con los encontrados en los escasos trabajos similares. Adicionalmente, para la misma muestra de trabajos empíricos previos, contribuimos con una identificación de los rasgos económicos aproximados por esas variables, lo que proporciona información adicional relevante sobre la significatividad e importancia de los factores. Así, la rentabilidad, el endeudamiento y el equilibrio financiero son los rasgos que se han encontrado más discriminantes, y deberían ser considerados cuando se seleccionan ratios individuales a partir del primer tipo de análisis.

Como resultado del análisis, queda patente un elevado grado de dispersión en los tres apartados en los que se ha dividido nuestro objetivo de estudio, concepto de fracaso, variables y metodologías. Detectamos una cierta convergencia en las variables con poder explicativo, en línea con trabajos anteriores: los ratios sobre rentabilidad seguidos de los ratios de endeudamiento. En el caso de las metodologías, se percibe una evolución dentro de cada una, pero no se puede hablar de abandono de unas a favor de otras, porque las metodologías que han ido apareciendo compensan carencias de las anteriores pero aportan nuevos problemas. Además, identificamos dos carencias comunes a todos los trabajos: por un lado, la utilización de un resultado dicotómico: empresa fallida, empresa no

fallida; y por otro lado, la escasa consideración de factores ajenos a la propia empresa.

1.2. Fracaso empresarial, modelos y variables explicativas

El concepto de fracaso, el planteamiento del modelo y las variables explicativas son elementos clave en la teoría sobre el fracaso empresarial. En los siguientes apartados tratamos de identificar los avances de la literatura previa en estos tres aspectos y las causas por las que no se ha conseguido construir un marco conceptual sobre la cuestión.

1.2.1. De la teoría a la evidencia empírica: el concepto de fracaso y el planteamiento de modelos

La definición de empresa fracasada y empresa sana determina a qué empresas se puede aplicar un modelo y qué papel pueden desempeñar en él. Haciendo una revisión de las distintas acepciones que los autores plantean para el fracaso empresarial, la quiebra o la empresa fallida, vemos idéntica diversidad en los trabajos españoles y extranjeros. En un trabajo reciente, Graveline y Kokalari (2008) mencionan tres grupos de conceptos: dejar de pagar una deuda; reunir las condiciones previstas en la normativa vigente sobre quiebra; o tener una situación patrimonial precursora del fracaso futuro; si bien ellos se inclinan por cualquiera de los dos primeros.

Como ejemplo del primer grupo podemos citar a Beaver (1966), Marais *et al.* (1984), Gabás (1990), García, Arqués y Calvo-Flores, (1995) o Westgaard y Wijst (2001), que definen el fracaso como la incapacidad de atender obligaciones financieras a su vencimiento.

En el segundo grupo se puede clasificar a Deakin (1972), que considera fracaso empresarial a las empresas en quiebra, suspensión de pagos o concurso de

acreedores, al igual que hacen en España Lizarraga (1997), Ferrando y Blanco (1998), López *et al.* (1998), Román *et al.* (2001), Gómez *et al.* (2008). Zmijewski (1984) toma sólo aquellas que presentan quiebra; Taffler (1982) entiende la liquidación voluntaria, por orden legal o con intervención estatal como fracaso. Enlazando con este último, podemos incluir en este apartado el cese de actividad genérico que recogen Becchetti y Sierra (2003). Los autores que han estudiado el sector bancario estiman el fracaso por la intervención o no del Fondo de Garantía de Depósitos (Laffarga *et al.*, 1985 y Pina, 1989). De forma similar, para el sector asegurador, el trabajo de Mora (1994b) y de Sanchis Arellano *et al.* (2003) consideran una compañía fracasada por la intervención o no de la Comisión Liquidadora de Entidades Aseguradoras.

Un buen representante del tercer grupo sería Altman (1981), que define el fracaso como insolvencia técnica o en el sentido de capital consistente en falta de liquidez; un grupo de autores se refiere al mantenimiento de pérdidas continuadas (Gilbert *et al.*, 1990; Hill *et al.*, 1996); Gazengel y Thomas (1992) consideran empresa fracasada aquella que genera cíclicamente más carga financiera que ingresos; y más recientemente Davydenko (2007) plantea que cuando la situación patrimonial refleja un valor reducido en los activos o una escasez de tesorería se puede desencadenar el fracaso empresarial. Rubio (2008) y Correa *et al.* (2003) hablan de entidad fracasada cuando ésta incurre en quiebra técnica, entendiendo como tal el patrimonio neto contable negativo.

En conjunto, se aprecia que en unos casos el fracaso es definitivo, mientras que en otros es temporal u ocasional, o bien solo relativo³. Pero ¿es muy diferente una empresa declarada en quiebra de una no declarada como tal que se encuentra en serias dificultades? Aunque Keasey y Watson (1991) afirman que la mayoría de los modelos predictivos consideran fracasada a la empresa en liquidación o quiebra, algunos trabajos empíricos demuestran estar usando esta definición de forma bastante más flexible, porque si bien seleccionan empresas fracasadas que

³ Abad *et al.* (2004) denominan a estos niveles de fracaso: agujero negro, revés y crecimiento fracasado. Manzanque *et al.* (2010) mediante análisis clúster establece cuatro niveles de fracaso: económico, financiero I, financiero II y económico-financiero.

cumplan esa condición para una parte de su muestra, evitan la selección de empresas en dificultades⁴ para la parte de la muestra que debe contener empresas saludables y para la muestra de comprobación (Shrieves y Stevens, 1979; Taffler, 1982). Precisamente, para saber si un modelo es útil cuando se trata de distinguir empresas saludables de fracasadas, debería aplicarse a empresas no claramente saludables ni claramente fracasadas, tal como señala Jones (1987). Pero cuando se ha aplicado esto en la práctica los modelos han resultado ser muy poco precisos en la identificación (Gilbert *et al.*, 1990).

Así, tomar como empresas fracasadas aquellas que han sido declaradas oficialmente como tales tiene ventajas pero también inconvenientes. Las ventajas serían: poder identificar de forma objetiva el fracaso (es irreversible y hay un registro de empresas en esta situación); y situarlo en el tiempo de forma precisa. Los inconvenientes serían: que no haya diferencias claras con el grupo de empresas no declaradas todavía fallidas, pero en proceso de llegar a serlo⁵; y la posibilidad de que las normas recojan distintos criterios⁶ en el tiempo, o distintos criterios por país, para el mismo tipo de situación de una empresa.

Mientras los dos primeros grupos de conceptos de fracaso hablan de la situación de la empresa a la fecha de estudio, el tercero, que aparece cronológicamente más tarde que los anteriores, trataría de detectar en la situación patrimonial actual

⁴ Por ejemplo, la gran precisión del modelo de Wilcox (1976) debe tomarse con cautela porque no usó muestra de comprobación y muchas empresas tenían probabilidades de quiebra que estaban en los extremos de la distribución, esto es, cerca de cero o cerca del 100%. El modelo no parecía ser sensible al desempeño de la empresa fuera de los extremos.

⁵ Que sean empresas en proceso de llegar a ser fallidas, no implica necesariamente que acaben siendo liquidadas o declaradas en quiebra. Cuanto más largo sea el proceso de deterioro más posibilidades hay de detectarlo y poner remedio, ya sea actuando sobre la parte operativa (cambios en el proceso productivo, restructuración de activos,...), sobre la parte financiera (refinanciación, financiación adicional propia o ajena,...) o entrando en procesos de concentración con empresas en mejor situación (como las fusiones). De hecho, tal como apuntan Keasey y Watson (1991), son precisamente las empresas que muestran señales de deterioro con suficiente antelación las que pueden beneficiarse de la aplicación de este tipo de modelos de predicción. Barniv *et al.* (1997) distinguen en su modelo empresas que tras la resolución del proceso de quiebra acaban siendo absorbidas, de otras que siguen funcionando como empresas independientes y de las que se liquidan.

⁶ Por ejemplo, de acuerdo con la normativa actual en España, una empresa puede solicitar voluntariamente la declaración de concurso de acreedores para beneficiarse de las ventajas que la normativa le ofrece para salir de una situación transitoria de dificultades, o bien, para evitar las consecuencias negativas de no haberlo hecho, si posteriormente llega a quebrar y se la declara en quiebra culpable.

el origen del futuro fracaso. Se percibe, pues, una evolución de los trabajos desde la identificación hacia la predicción del fracaso empresarial. Y esto es un cambio relevante desde el punto de vista teórico, que debe tenerse en cuenta en la aplicación de los modelos y en la interpretación de los resultados.

En primer lugar, un modelo de identificación compara empresas fracasadas (y no fracasadas), las que sirven para calibrar el modelo, con empresas que reúnen las mismas condiciones, en la parte que las variables seleccionadas son capaces de captar. Cuanta mayor proporción de estas condiciones se reúna, más cerca estará la entidad analizada del fracaso. Mientras que un modelo de predicción no debería comparar las empresas que se quiere evaluar con las empresas fracasadas una vez concluido el proceso, sino con las fases anteriores de deterioro. Algunos trabajos⁷ han identificado la evolución de las variables en los años anteriores al fracaso, los problemas son: distinguir fases con diferentes denominaciones, que pueden estar fuertemente condicionadas por la normativa vigente; y que la velocidad de deterioro de las empresas no es homogénea, depende tanto de factores externos (restricciones al crédito, evolución negativa de la demanda, duración de los ciclos económicos,..), como internos (problemas en la sucesión de la propiedad, en la dirección, conflictividad laboral,..). Algunos autores creen que la definición de fracaso condicionará la elección de las variables (Balcaen y Ooghe, 2006), si bien otros encuentran una selección de variables similar en modelos formulados a partir de distintas definiciones (Hayden, 2003).

⁷ Shumway (2001) ataca los modelos estáticos atribuyéndoles estimaciones sesgadas, inconsistentes e ineficientes, por lo que propone el uso de modelos de riesgo para periodos discretos. Balcaen y Ooghe (2006) analizan el problema de no considerar la dimensión temporal del fracaso, porque la falta de consistencia de las señales en el tiempo, unida a la aplicación mayoritaria de modelos estáticos, genera muchos errores de clasificación y predicción. Laitinen (1993) pone de manifiesto cómo los coeficientes de los modelos van cambiando a medida que se toman las variables contables en periodos más alejados del fracaso. Jones y Hensher (2004) modelizan tres categorías diferentes de fracaso de la empresa (no fallidas, insolventes y en quiebra). Y en su trabajo de 2008 incorporan también empresas en dificultades en proceso de fusión o toma de control. Beaver *et al.* (2005; 2009) utilizan una aproximación logística multiperiodo que, a diferencia de los modelos estáticos, considera el riesgo de quiebra en varios años, incluidos los años en los que la empresa quebrada aún no lo estaba. Labatut *et al.* (2009) incorporan el tiempo como una variable fundamental en la detección del fracaso empresarial, para lo cual modelizan, con probabilidades el comportamiento de un sistema de control aplicado a ratios contables.

Otra cuestión importante, relacionada con la definición de empresa fracasada y no fracasada, atañe a la selección de la muestra. La necesidad en muchos modelos de formar dos sub-muestras, una de empresas fracasadas y otra de empresas sanas, plantea otro problema: que no se cumpla la teórica aleatoriedad de la muestra característica de un modelo estadístico. En la gran mayoría de trabajos empíricos sobre fracaso empresarial las muestras no son aleatorias. Esto se debe, en unos casos, a que el modelo se aplica emparejando cada empresa fallida con una no fallida, y suelen ser contemporáneas y similares en rasgos tales como el tamaño, el sector o la antigüedad (Altman, 1968; Blum, 1974; Ohlson, 1980; Zavgren, 1985; Tsukuda y Baba, 1994; Leshno y Spector, 1996; Piramuthu *et al.* 1998; Platt y Platt, 2002; Lee *et al.*, 2005). Otra causa habitual es la diferencia en la población de empresas fracasadas respecto a la de empresas sanas, lo que hace que la población menos frecuente quede sobrerrepresentada en el modelo (Zmijewski, 1984; Sueyoshi y Goto, 2009c). También se ha apuntado el sesgo de la muestra para dejar fuera a empresas que por su tamaño o juventud presentan datos incompletos (Taffler, 1982). El hecho de no seleccionar una muestra de forma aleatoria presenta los siguientes inconvenientes (Palepu, 1986): se sobreestima la capacidad predictiva del modelo; difícilmente se puede generalizar el modelo al resto de la población; y puede dificultar la interpretación económica de los resultados. En particular, si las muestras no son aleatorias en cuanto a características generales tales como el sector, el tamaño o la edad de la empresa, el modelo no será aplicable a las entidades infrarrepresentadas en la muestra (Balcaen y Ooghe, 2006).

En cuanto a la teoría que subyace a los modelos de fracaso empresarial, muy pocos autores que escriben sobre la cuestión, antes de proponer modelos, muestran su interés por identificar las causas y los síntomas de una empresa que fracasa. Para el planteamiento de los modelos, la mayoría de los autores parece basarse en alguna variante del modelo de flujos de efectivo de Beaver (1966) donde la empresa mantiene un remanente de efectivo como resultado de sus cobros y pagos por actividades operativas, y la empresa fracasa cuando ese remanente desaparece (Keasey y Watson, 1991).

Argenti (1976) identifica los defectos de las empresas que les llevan a cometer una serie de errores que desencadenan el fracaso. Estos defectos se producen por una mala gestión, deficiencias en el sistema contable, o una mala adaptación al cambio. En la segunda fase, los errores se producen por dejar aumentar en exceso el apalancamiento, seguir con negocios que no generan beneficios, o iniciar un gran negocio que una mala racha convierte en una gran carga. En la tercera fase aparecen los síntomas, siendo los más importantes los financieros (deterioro de liquidez, aumento de apalancamiento, reducción de ventas sobre activos, reducción de tesorería sobre deuda, reducción de la cotización sobre beneficios, etc.), pero también el empleo de contabilidad creativa para ocultar la realidad, señales no financieras sobre el empeoramiento de la calidad, del mantenimiento de las instalaciones o de retrasos en el suministro de los proveedores. Otras causas apuntan a la mala suerte, el fraude o las restricciones impuestas desde diversos organismos públicos o privados, (sindicatos, gobierno, grupos de protesta,...). Otro hallazgo importante de Argenti (1976) es que hay tres clases bien distintas de empresas que fracasan: las que nunca llegan a despegar después de su creación; las individuales que crecen demasiado y resultan inabarcables para su creador; y las empresas maduras que, sin una buena estructura de gestión, acaban quebrando tras varios años de deterioro gradual. Para estos tipos distintos de empresas los síntomas aparecen en diferente orden y el proceso de deterioro tiene una duración muy diferente.

Keasey y Watson (1991) se apoyan en el trabajo de Argenti (1976) y sostienen que la quiebra de una empresa depende no solo de sus deudas, sino también de los intereses económicos y el poder de sus partícipes (*stakeholders*), así como de la capacidad de los gestores para superar las dificultades, para hacer frente a las deudas y para actuar con independencia de los intereses de los partícipes.

Wilcox (1971) se muestra crítico con autores como Beaver o Altman por su falta de marco conceptual. Su intención es desarrollar un modelo teórico para explicar mejor los resultados de Beaver y para generar hipótesis que llevaran a predictores de fallido potencialmente mejores. Para ello, adapta el clásico problema de la ruina del jugador fijándose en el valor de liquidación neto y los factores que causan su

fluctuación. El primero es simplemente un nivel de dinero determinado por la entrada de liquidez y la tasa de salida. Asumiendo un “proceso estable”, Wilcox postula que la probabilidad de que el valor de liquidación neto se reduzca a cero (interpretado como quiebra o fallido definitivo) es una función de tres elementos: el valor de liquidación neto actual o riqueza actual, la media ajustada del flujo de efectivo, y la variabilidad del flujo de efectivo ajustado, medido por su varianza. El propio autor reconoce que el modelo es una gran simplificación, y sus desalentadores resultados empíricos le muestran el escaso significado de su probabilidad de fallido. Más tarde Wilcox (1976) abandona la estructura de la forma funcional y construye un modelo de predicción basado en las variables que sugería su primer modelo.

En cambio, la inmensa mayoría de autores parecen compartir la opinión de Altman, que desde sus primeros trabajos (en los años 60) admite que quizá hay diferentes causas para el fallido, pero que identificar las vinculaciones causales puede no ser muy útil para prevenir o predecir el fracaso. En esta línea, Scott (1981), tras un análisis comparativo de los modelos empíricos propuestos hasta ese momento, concluye que aunque los modelos no se basan en teoría explícita, su éxito sugiere la existencia de una regularidad subyacente fuerte.

Y es en la búsqueda de esa regularidad subyacente donde los modelos estadísticos desempeñan su papel. Mediante una agregación estadística se resume la información contenida en los estados financieros de la empresa para determinar si el perfil financiero de la empresa se asemeja más a los perfiles financieros de las empresas que previamente han resultado fallidas o han estado en dificultades, o bien, a las que no han resultado fallidas ni han estado en dificultades. Por tanto, los modelos estadísticos no constituyen una teoría explicativa del fracaso, sino herramientas descriptivas para el reconocimiento de pautas, tal como señalan Keasey y Watson (1991).

Los escasos intentos de formular modelos desde el razonamiento económico se han encontrado con resultados desalentadores en la fase de contraste empírico, a

causa de diversos inconvenientes para la generalización: hay distintos tipos de fracaso empresarial, con diferente velocidad de deterioro y con síntomas distintos; no hay una forma genérica buena de distinguir empresas fracasadas de empresas no fracasadas, salvo quizá en los extremos; las variables accesibles para el investigador reflejan fundamentalmente la situación económica y financiera de las empresas, pero difícilmente consiguen captar las características y el desempeño de los propietarios y gestores, que tanto tienen que ver con la forma de afrontar las crisis y la posibilidad de salir de ellas. Finalmente, para el planteamiento de los modelos es muy importante la finalidad que se les vaya a dar, a qué empresas se van a aplicar y con qué costes, lo cual depende subjetivamente del analista.

De manera que la incapacidad del marco conceptual (o la ausencia de éste, en opinión de otros autores) para identificar variables, ya sea con carácter general, para cualquier tipo de fracaso, o con carácter específico, para clases de empresas o de situaciones económicas concretas, ha llevado a los autores a sustituir la teoría, sobre qué variables debe contener un modelo, por la regularidad estadística de las variables explicativas para el grupo específico objeto de análisis en el periodo de estudio y con los condicionantes que el analista incorpore. Así, el punto de partida para seleccionar las variables es una especie de fondo común de variables elegibles que, en unos casos, respetan el razonamiento económico y, en otros, el respaldo empírico de trabajos anteriores que, a su vez, se apoyan en el razonamiento económico o en resultados de trabajos anteriores.

1.2.2. Variables explicativas

Como quedaba de manifiesto en el apartado anterior, la ausencia de un modelo teórico formal de las relaciones entre el fracaso de la empresa, los factores económico-financieros internos o externos a la empresa, los intereses económicos de los distintos agentes relacionados y las formas de actuar de los gestores, ha sido la causa de que las variables que se incluyen en los modelos se seleccionen cada vez menos a partir del razonamiento económico y cada vez más a partir de la experiencia empírica previa en trabajos sobre la cuestión y a partir del uso de procedimientos estadísticos que seleccionen las variables más explicativas de

entre una amplia variedad inicial (Scott, 1981). Esta es la norma actual en trabajos de acumulación de evidencia mediante el uso de una metodología previamente desarrollada y aplicada al fracaso empresarial.

En cambio, los trabajos que proponen nuevas metodologías, o innovaciones relevantes en alguna metodología ya empleada, tienden a seleccionar las mismas variables que se usan en los trabajos empíricos con los que quieren comparar el desempeño de su modelo. En la Tabla 1, se ofrecen varios ejemplos de criterios utilizados en la selección de ratios financieros, principal grupo de variables que se utilizan en el contraste de modelos de fracaso empresarial.

Tabla 1. Ejemplos de criterios declarados para la selección de ratios financieros

CRITERIO DE SELECCIÓN DE RATIOS	AUTORES
-Popularidad en la literatura. -Funcionamiento en estudios previos. -Relación del ratio con el concepto de <i>cash flow</i> .	Beaver (1966)
-Ratios del estudio de Beaver (1966).	Deakin (1972)
-Marco general para selección de variables, basado en el modelo de <i>cash flow</i> .	Blum (1974)
-Parte de ratios usados por Deakin (1972). -Análisis de componentes principales.	Libby (1975)
- Entre los cinco grandes bloques de información financiera (beneficio, <i>cash flow</i> , activo, pasivo patrimonio neto y fondo de maniobra), selecciona variables contables independientes y escoge los ratios formados con ellas. - Toma otras variables que la evidencia previa encuentra significativos	Lincoln (1984)
- Empleo frecuente en el análisis financiero. - Ratios computables según datos. - Criterio de investigadores y grado de significación estadística (test F).	Laffarga, Martín y Vázquez. (1985, 1986a, 1986b, 1991)
Juicio investigadores	Laffarga, Martín y Vázquez (1987)
- Los más comúnmente utilizados. - Emplea un modelo general de regresión y desecha los datos con un valor F^2 poco significativo	Pina Martínez (1989).
Índices contables frecuentes.	Rodríguez Fernández (1986; 1987, 1989a, 1989b)
Ratios contables diversos y año de fundación	Martínez, Navarro y Sanz. (1989)
Ratios utilizados por Pina (1989), previa comprobación de su poder discriminatorio	Serrano Cinca (1994, 1997); Serrano y Martín (1993); Mar Molinero y Serrano Cinca (2001); Fernández y Olmeda (1995); Serrano Cinca y

CRITERIO DE SELECCIÓN DE RATIOS	AUTORES
	Martín del Brío (1993); Bonsón, Escobar y Martín. (1997a; 1997b); De Andrés Sánchez (2005)
Ratios utilizados en literatura previa. Selección a partir de los criterios CAMEL (<i>Capital, Asset, Management, Equity and Liquidity</i>), muy utilizado en la banca en USA.	Tam (1991); Tam y Kiang (1992); Swicegood y Clark (2001)
Ratios efectivos para la predicción en literatura previa. Aplicación de contrastes de significación estadística y análisis <i>cluster</i> .	Tsukuda y Baba (1994)
Variables utilizadas en Altman (1968)	Serrano-Cinca (1996); Zhang <i>et al</i> (1999); Lee <i>et al</i> . (2005)
Ratios utilizados por Ohlson (1980)	Barniv <i>et al</i> . (1997)
Diversidad de ratios utilizados en la literatura y en estudios teóricos	López González y Flórez López (2000)
Variables que captan la fortaleza de una empresa y apoyan la evidencia previa	Cielen <i>et al</i> . (2004) Beaver, McNichols y Rhie (2005); Gómez, De la Torre y Román (2008)
Ratios utilizados en Beaver, McNichols y Rhie (2005)	Beaver, Correia y McNichols (2009)

Cuando se seleccionan las variables estadísticamente, el inconveniente apuntado por varios autores (Edmister, 1972; Keasey y Watson, 1991; Zavgren, 1983) es que los resultados no se puedan extrapolar a otras muestras de la población. Por otro lado, una selección estadística de variables puede incluir factores interrelacionados, generando problemas de correlación que explicarían por qué en algunos modelos los signos de una parte de las variables son contrarios al razonamiento económico (Balcaen y Ooghe, 2006). También puede ocurrir que queden fuera del modelo factores que a priori parecerían relevantes, tales como liquidez, endeudamiento, rentabilidad o actividad (Dambolena y Khoury, 1980).

Para analizar las variables relevantes en los estudios sobre fracaso empresarial y los principales factores económicos que subyacen a tales variables, hemos seleccionado una muestra de 40 trabajos empíricos que abarcan desde 1966 hasta 2009 y utilizan una amplia variedad de las metodologías mencionadas en esta revisión. Las variables que se consideran significativas para incluir en los modelos se recogen en una tabla en el Anexo I al final del trabajo. En ella, las variables pueden buscarse por orden alfabético y los autores por orden

cronológico. La última columna representa el número de trabajos de la muestra que considera explicativa cada variable, y la última fila indica el número de variables consideradas explicativas en cada uno de los trabajos analizados. Ambas líneas de totales presentan gran dispersión: los trabajos por variables varían entre 1 y 18; en tanto que las variables por trabajo varían entre 1 y 34. La Tabla 2 muestra los 8 ratios que se repiten con mayor frecuencia en la muestra de trabajos seleccionada. Revisiones similares sobre frecuencia de ratios se han hecho en Dimitras *et al.* (1996), en Daubie y Meskens (2002) y en Bellovary *et al.* (2007). Los ratios más frecuentes para Dimitras *et al.* (1996) son: fondo de maniobra/activo total, deuda total/activo total, activo circulante/pasivo circulante, BAIT/activo total y beneficio neto/activo total. Daubie y Meskens (2002) obtienen similares resultados en diferente orden: activo circulante/pasivo circulante, fondo de maniobra/activo total, BAIT/activo total, activos líquidos/pasivo circulante y beneficio neto/activo total. Por su parte, Bellovary *et al.* (2007) obtiene que los cinco más frecuentes son: beneficio neto/activo total, activo circulante/pasivo circulante, fondo de maniobra/ activo total, beneficio retenido/activo total y BAIT/activo total. Si comparamos nuestros ratios más frecuentes con los identificados en estos tres trabajos, vemos que entre nuestros 5 primeros ratios cuatro son los mismos (y en el mismo orden) que en Dimitras *et al.*; y tres coinciden con Daubie y Meskens (también en el mismo orden) y otros tres con Bellovary *et al.* Tres ratios aparecen en común en todos los estudios, dos de rentabilidad económica y uno de equilibrio económico-financiero. Aunque los otros ratios que aparecen como más frecuentes a estos autores aparecen en nuestro estudio, las frecuencias encontradas son mucho menores. Sin duda, la muestra de trabajos empleada en cada caso está sesgando las frecuencias⁸.

⁸ Por ejemplo, el ratio activo circulante/activo fijo que se sitúa en quinto lugar en nuestro estudio, y que ya aparecía en los trabajos pioneros de Beaver o Altman, es muy utilizado en trabajos publicados en España, que no se recogen en las otras dos revisiones.

Tabla 2. Ratios más frecuentes

VARIABLES EXPLICATIVAS	Nº de trabajos
Deuda Total/Activo total	18
Activo Circulante/Pasivo Circulante	14
BAIT/Activo Total	14
Beneficio Neto/Activo Total	14
Activo Circulante/Activo Total	10
Beneficios No Distribuidos/Activo Total	7
Gastos Financieros/Pasivo Exigible	7
Recursos Generados/Pasivo Exigible	7

Pero las variadas alternativas para formular ratios que miden un mismo aspecto de la empresa (como liquidez, endeudamiento, rentabilidad, etc.) hacen que los resultados de la tabla comprensiva recogida en el Anexo I, e incluso los resumidos en la Tabla 2 sobre los ocho ratios más repetidos, resulten difíciles de interpretar. Por ello, hemos agrupado las variables en bloques homogéneos por su significado económico y presentamos un resumen de los resultados en la Tabla 3. Así, podemos identificar qué aspectos de la empresa han resultado relevantes con mayor frecuencia en la muestra para la identificación o predicción del fracaso empresarial. En la Anexo I también se puede apreciar si los factores analizados han variado o no con el tiempo.

Tabla 3. Factores económicos subyacentes a las variables

Categorías de ratios	Items	% s/ Total	Ratios distintos
Rentabilidad	64	17,44%	16
Endeudamiento	55	14,99%	20
Equilibrio Económico - Financiero	42	11,44%	12
Estructura Económica	38	10,35%	15
Margen	35	9,54%	16
Rotaciones	32	8,72%	17
Otros ratios	76	20,71%	37
Variables	25	6,81%	21
Totales	367	100%	154

La Tabla 3 muestra que el rasgo de la empresa más repetido en los estudios de fracaso seleccionados para la muestra es la rentabilidad (64 items) y que ese rasgo se ha medido con 16 ratios diferentes. Consideramos ratios de rentabilidad aquellos que relacionan partidas de resultado con partidas de balance, ya sean económicas o financieras. En segundo lugar se sitúan los ratios de endeudamiento, donde hemos recogido todos aquellos que relacionan dos masas patrimoniales de financiación (al menos una de ellas de deuda). El tercer rasgo es el equilibrio económico-financiero, donde incluimos ratios que relacionen masas de activo con masas de financiación. El cuarto es la estructura económica, donde los ratios comparan dos masas patrimoniales del activo. El quinto es el margen, donde incluimos la relación entre partidas de resultados y otros componentes de la cuenta de pérdidas y ganancias. El sexto es la rotación donde se compara una partida de la cuenta de pérdidas y ganancias con una masa patrimonial del balance. Al final de la tabla se puede ver que los estudios han utilizado otros 37 ratios que no encajan en las categorías anteriores, y para los que cualquier grupo homogéneo alcanzaría mucha menor frecuencia que los cinco anteriores. También se han utilizado 21 variables sin forma de ratio. Como norma general, cada variable de estas solamente aparece en un trabajo.

La primera conclusión es que los trabajos empíricos han combinado todos los ratios posibles que pudieran aportar información adicional sobre el buen o mal funcionamiento de la empresa. La segunda conclusión es que los factores que la teoría apunta como factores fundamentales, y que se pueden aproximar con datos contables, son los que subyacen a los ratios más utilizados: rentabilidad, endeudamiento y equilibrio económico-financiero. Tanto en la Tabla 2 como en la recogida en el Anexo I, se puede apreciar que la inmensa mayoría son ratios calculados a partir de información contable de la empresa que se quiere analizar; y casi todas son variables referidas a la empresa en cuestión. Utilizar datos contables tiene dos inconvenientes básicos: la influencia que sobre las variables ejercen las políticas contables de la empresa (problema más acusado con empresas pequeñas y medianas) (Balcaen y Ooghe, 2006); y las diferencias en la normativa contable entre distintos países, o bien en el mismo país entre distintos periodos, cuando se ha producido un cambio de normas.

De forma complementaria a los ratios contables, y con escasa frecuencia, se han utilizado datos no procedentes de los estados financieros, aunque parte de esa información se calcule a partir de los anteriores o proceda de los informes anuales de las empresas. Por ejemplo, Dambolena y Khouri (1980) incluyen medidas de estabilidad de los ratios, tales como la desviación típica, el error estándar de la media y el coeficiente de variación de la función discriminante, obteniendo una gran mejora⁹ en la clasificación de empresas fracasadas y no fracasadas. Barniv *et al.* (1997) y Mateev *et al.* (2013) incorporan la complejidad de la estructura de capital medida por el número de tipos principales de deuda. Algunas variables no financieras con capacidad explicativa incremental han resultado ser, por ejemplo, la existencia de salvedades en el informe de auditoría (Peel y Peel, 1987) sobre todo si en empresas pequeñas varían de unos ejercicios a otros (Keasey y Watson, 1987), el retraso en la publicación de las cuentas (Ohlson, 1980; Whittred y Zimmer, 1984; Peel *et al.*, 1986; Peel y Peel, 1987; Keasey y Watson, 1988; Somoza, 2002), la diferencia de tiempo entre el final del ejercicio y la fecha de publicación de los estados contables (Peel *et al.*, 1986), la existencia de incidencias de pago (Román *et al.* 2001), problemas de liquidez temporales (Behr y Güttler, 2007) o sector industrial (Grunet *et al.*, 2005; Behr and Güttler, 2007).

Otro grupo de trabajos incorpora información de mercado para complementar la información contable. Marais *et al.* (1984) encuentran que una serie de variables de mercado, tales como las cotizaciones de las acciones o las calificaciones crediticias de los bonos, incorpora tanta información sobre el fracaso como los ratios contables. Barniv *et al.* (1997) incluyen los resultados anormales acumulados, ponderados por el precio de mercado, de los 60 días anteriores a la declaración de quiebra. Atiya (2001) utiliza la tasa de cambio de los precios de mercado, el precio sobre el *cash flow* y la volatilidad de los precios de mercado. Zavgren (1988) y Hillegeist *et al.* (2004) argumentan que las variables de mercado contribuyen a los modelos mejorando la oportunidad de la información.

⁹ Según Keasey y Watson (1991) estos resultados deberían interpretarse con cautela, ya que la mejora podría deberse a que las medidas de estabilidad incorporan más años de información para su cálculo.

Otros estudios destacan la utilidad de incluir variables cualitativas ajenas a los informes anuales de las empresas. El trabajo pionero de Argenti (1976), si bien no puede considerarse un trabajo estrictamente empírico, propone incluir en los modelos de fracaso cuestiones como la estructura de gestión, la inadecuación del sistema de información contable o la manipulación de los estados financieros publicados, todas ellas pertenecientes al ámbito interno, sobre el que la dirección puede influir. Stein y Ziegler (1984) comprueban el efecto de la gestión sobre el fracaso de empresas de tamaño mediano, usando variables tales como: el despido de empleados, el cambio de forma legal, algunos detalles de la producción y el perfil de los gestores, además de variables externas como las reducciones de la demanda y el aumento de la competencia. Keasey y Watson (1987) se basan en el modelo de Argenti y encuentran significativo para pequeñas empresas incluir variables como el número de directivos o los cambios en las salvedades de los auditores, esta última para aproximar la posible "contabilidad creativa" de las empresas en dificultades. Para McGahan y Porter (1997) y Rumelt (1997) los factores externos son los que mejor explican el fracaso empresarial, si bien la empresa puede ejercer escaso control sobre ellos. Entre estos factores estarían las variaciones de la estructura de la demanda, los cambios en los gustos de los consumidores, la rivalidad entre los competidores, tanto actuales como futuros, el declive de la demanda y la incertidumbre tecnológica. Barniv *et al.* (1997), tratando de identificar distintas formas de resolver la declaración de quiebra, incorporan factores como: el grado de competitividad, el comportamiento fraudulento, los cambios en la dirección, la concentración de la propiedad o la edad de la empresa. Becchetti y Sierra (2003), consideran conveniente incluir variables relacionadas tanto con la estrategia de la organización (la posición competitiva o el nivel de exportación) como con las características del sector (grado de concentración del mercado). Grunert *et al.* (2005) consideran factores como la edad, el tipo de negocio o el sector en combinación con los ratios financieros. Madrid y García (2006) utilizan factores internos: la posición tecnológica débil, las escasas actividades de innovación y la calidad. En el caso de Japón (Xu y Zhang, 2009), la capacidad predictiva de los modelos mejora al incorporar la dependencia bancaria y la dependencia de los grupos de empresas (*Keiretsu*), dos variables que representan rasgos institucionales del país.

Altman (1983) y Mensah (1984) muestran que diferentes entornos macroeconómicos afectan a la precisión de los modelos de predicción del fracaso empresarial. Somoza (2001) incluye como variables independientes externas algunas macroeconómicas y sectoriales llegando a resultados decepcionantes, ya que las mismas no parecen incrementar la capacidad predictiva de los modelos obtenidos. Esas variables son: el tamaño de la empresa deflactado por el PIB, la variación de la producción sectorial y la ponderación de cada ratio por la media de la muestra. Jones y Hensher (2008) no encuentran significativo ninguno de los factores macroeconómicos considerados para una muestra de empresas australianas (evolución del mercado bursátil, variación de tipos de interés y variación del producto interior bruto). El estudio de Jacobson *et al.* (2008) con empresas suecas sí muestra que la información macroeconómica puede resultar útil para complementar la información de las empresas tanto cotizadas como no cotizadas. En tanto que para empresas del Reino Unido, la evidencia muestra un poder explicativo relativamente bajo en el trabajo de Liou y Smith (2006).

En síntesis, se utilizan una gran cantidad de variables, aunque las más frecuentes son ratios contables que aproximan unos pocos aspectos del desempeño de la empresa: rentabilidad, endeudamiento, equilibrio económico-financiero y estructura económica. Destaca la escasa utilización de variables no contables, aunque los resultados apuntan en general a una mejora en la capacidad explicativa de los modelos cuando algunas de estas variables se incorporan. Por un lado, esta mejora puede deberse al hecho de que los gestores no puedan o no tengan interés en ajustar sus valores, como sucede con determinados ratios¹⁰. Por otra parte, se confirma el razonamiento sobre las causas del fracaso, que (además de los factores económicos financieros de la propia empresa) apunta a factores del entorno sectorial y macroeconómico, así como a factores internos relativos al desempeño del elemento humano de la empresa, fundamentalmente los gestores. Ninguno de estos elementos se refleja directamente en los estados financieros, sino que su efecto va apareciendo con el tiempo. Por ello, incluir variables no procedentes de los informes anuales que capten esos factores debería

¹⁰ En referencia al efecto que la propia información contable puede tener sobre la capacidad predictiva de los modelos, Beaver *et al.* (2009) miden las diferencias en los atributos de la información financiera y Abad *et al.* (2004) la calidad de la información contable, encontrándose, en ambos casos, un efecto significativo sobre los resultados empíricos.

proporcionar una mejora en la oportunidad de la información recogida por el modelo.

En cuanto a las variables macroeconómicas, destacamos su ausencia casi completa en los estudios analizados y que la evidencia encontrada varía por países, por lo que no resulta concluyente con carácter general. En nuestra opinión, para que las variables macroeconómicas puedan desempeñar su papel en los modelos de fracaso hay que tener en cuenta los ciclos económicos, lo cual requiere aplicar los modelos a series temporales amplias. En cambio, gran parte de los trabajos sobre fracaso no dispone de series temporales superiores a cinco años. Por otra parte, Keasey y Watson (1991) aconsejan diferenciar los modelos de identificación de los modelos de predicción, porque en estos últimos también sería necesario estimar los valores futuros (para el horizonte temporal de predicción) de las variables macroeconómicas que incluya el modelo.

Finalmente, en este apartado dedicado a las variables, hemos de señalar que para modelos de predicción no solo las variables macroeconómicas plantean problemas, todas las variables independientes deberían mantener una relación estable entre ellas y con la variable dependiente (Edmister, 1972; Zavgren, 1983; Mensah, 1984; Jones, 1987), especialmente si los modelos se aplican a un *pool* de datos (Zmijewski, 1984). Pero el entorno económico no permite que esas condiciones de estabilidad y no estacionalidad se cumplan. Los ciclos económico-financieros que vienen acompañados de cambios en la inflación, la demanda y los tipos de interés, cambios en la competitividad o en la tecnología del sector y en la estrategia competitiva de la empresa contribuyen a esa falta de estabilidad (Balcaen y Ooghe, 2006). Y el problema se agudiza precisamente en las empresas con dificultades, próximas al fracaso (Dambolena y Khoury, 1980), lo que se traduciría en estimaciones de coeficientes inconsistentes y en niveles de precisión bajos en los modelos (Platt *et al.*, 1994).

Estas dificultades para la generalización de las variables apoyan el establecimiento de modelos *ad hoc* para las condiciones que requiera el analista,

de acuerdo con su población objeto de estudio. Si hay unas variables básicas en las que los investigadores convergen y que son las que el razonamiento económico aconseja, a partir de ahí se añadirán unas u otras variables, las que funcionen mejor en cada caso.

1.3. Revisión de metodologías

Dado que varios trabajos de revisión bien conocidos (Zavgren, 1983; Jones, 1987; Altman, 1993; Balcaen y Ooghe, 2006, Demyanyk y Hasan, 2010) y muchos de los trabajos empíricos han comparado las distintas metodologías, para identificar sus ventajas e inconvenientes, no vamos a hacer aquí un análisis exhaustivo de los fundamentos y la capacidad explicativa o predictiva de aquellas metodologías que consideramos ampliamente utilizadas y suficientemente revisadas. En este apartado solamente recogemos las principales conclusiones de esos análisis, con indicación de algunas referencias donde el lector puede acudir para ampliar la información. En cambio, realizamos una revisión más profunda de dos metodologías de aparición reciente, que apenas se mencionaban o todavía no habían empezado a publicarse al realizarse muchos de los trabajos de revisión mencionados, nos referimos a los modelos de inteligencia artificial y a los que utilizan fronteras DEA (*Data Envelopment Analysis*).

En la revisión de la literatura sobre este tema, que viene interesando a analistas e investigadores desde los inicios del siglo XX, hemos encontrado que los modelos utilizados para el análisis del fracaso empresarial están dominados por el análisis discriminante y los enfoques econométricos de la regresión logística, como ya han señalado Aziz y Dar (2006) y Premachandra *et al.* (2009). La siguiente tabla recoge las fechas de publicación de trabajos de referencia que utilizan distintas metodologías sobre el fracaso empresarial.

Tabla 4. Trabajos pioneros en distintas metodologías

Fechas	Metodologías
Años 30 del siglo XX	Modelos univariantes básicos
1966	Beaver. Análisis univariante. Análisis de la varianza y el test de clasificación dicotómica
1968	Altman. Análisis Discriminante Múltiple. Modelo <i>Z-score</i>
1977	Martin. Regresión logística. Logit y Probit
1984	Marais <i>et al.</i> Algoritmo de particiones recursivas o iterativas
1990	Bell <i>et al.</i> Inteligencia Artificial. Redes neuronales
1991	Mar Molinero y Ezzamel. Técnicas de Escalamiento Multidimensional
1996	Serrano-Cinca. Inteligencia Artificial: Mapas autoorganizativos
2002	Park y Han. Análisis multicriterio
2002	Shin y Lee. Inteligencia Artificial. Algoritmos genéticos
2004	Paradi <i>et al.</i> Análisis Envolvente de Datos (DEA)

1.3.1. Análisis univariante

Los investigadores parecen coincidir en un primer¹¹ trabajo de referencia, fiable y contrastable, realizado con técnicas estadísticas por Beaver en 1966. Este primer trabajo utiliza como metodología un análisis univariante sobre seis categorías de ratios, calculando en primer lugar 30 ratios y quedándose posteriormente con un ratio por cada categoría. Aplica los cálculos sobre una muestra de 158 empresas; la mitad fracasadas y la mitad sanas, para un periodo de cinco años, llegando a la conclusión de que el análisis de perfiles diferencia los dos grupos de empresas (sanas y fracasadas), pero no permite medir dicha diferencia.

Otros autores utilizaron posteriormente el análisis univariante para sus trabajos, como Lev (1971) que introdujo el análisis por descomposición en la predicción del fracaso empresarial, Wilcox (1971, 1973, 1976) aplicó un modelo de “ruina del jugador” basado en el *cash-flow* o Viscione (1985) que llevó a cabo, sin mucho éxito, una combinación de ratios individuales.

¹¹ En Bellovary *et al.* (2007) se recoge un resumen histórico de los estudios sobre predicción de quiebras llevados a cabo entre 1930 y 1965.

El análisis univariante, que ha de estar presente en la fase inicial de cualquier investigación, es excesivamente simple. Estas técnicas tratan de determinar la significatividad de cada ratio para explicar el fracaso; y la posible predicción de los ratios más significativos. Aunque el trabajo de Beaver (1966) indicaba claramente que su principal motivación no era predecir el fracaso, sino mostrar el potencial informativo de los datos contables. Las técnicas fundamentales de análisis univariante, según Mora (1994a), son el análisis de la varianza y el test de clasificación dicotómica. Su principal limitación es la restrictiva premisa de que existe una relación lineal entre todas las variables explicativas y la situación de fracaso de la empresa.

1.3.2. Análisis discriminante

El siguiente trabajo de amplio reconocimiento sobre fracaso empresarial es el de Altman (1968). Su modelo *Z-score* formaliza el tipo de análisis más cualitativos de riesgo de fallido que estaban utilizando agencias de *rating* como *Standard&Poor's* o *Moody's*. Altman desarrolla una función discriminante resultado de la combinación de cinco ratios: Capital Circulante / Activo Total, Beneficios No Distribuidos / Activo Total, Resultados Antes de Intereses e Impuestos / Activo Total, Valor de Mercado de los Fondos Propios / Deudas y Ventas / Activo Total. La aplicación de los coeficientes hallados a las variables sugeridas le da a Altman una puntuación Z, a partir de la cual se diferencian las empresas sanas de las fracasadas. El punto de corte lo estableció el autor en $Z=2,67$. En este caso, la metodología utilizada para seleccionar los ratios y la ponderación adecuada es multivariante.

El análisis discriminante es una técnica cuyo objetivo es obtener una función que permita la clasificación de individuos, en clases definidas a priori, a partir de la puntuación obtenida con un conjunto de variables cuantitativas o cualitativas y de un nivel de corte. Variables que podían resultar insignificantes en un análisis univariante pueden aportar información significativa en este tipo de modelos al combinarse con otras variables (Altman, 1968). No obstante, los coeficientes no pueden interpretarse del modo que se hace en una regresión, por lo que algunos

autores sostienen que dichos coeficientes no permiten medir la importancia relativa de las variables para el fracaso de la empresa (Altman, 1968; Blum, 1974; Taffler, 1983; Zavgren, 1985; Balcaen y Ooghe, 2006).

Las premisas para la aplicación del análisis discriminante, cuando las variables independientes son cuantitativas, especialmente en el caso inferencial, son las siguientes:

1. Las variables independientes se distribuyen con normalidad multivariante.
2. Homoscedasticidad: las matrices de varianzas-covarianzas son iguales para ambos grupos de empresas (fallidas-no fallidas)
3. Los tamaños muestrales son representativos.
4. Los dos grupos de empresas son discretos, identificables y no se solapan.
5. Las probabilidades de fallido y los costes de clasificar mal a las empresas tienen que estar especificados

Si se cumplen las premisas, el modelo resulta óptimo en la precisión de las clasificaciones, por lo que el análisis previo de los datos es determinante. El requerimiento de la normalidad en la distribución ha puesto en duda la validez de algunos de los modelos que utilizaban esta técnica. Los resultados de varios estudios empíricos (Deakin, 1976; Eisenbeis, 1977; Collins y Green, 1982; García Ayuso, 1995) muestran que solo una parte de los ratios contables cumple la hipótesis de normalidad. En los trabajos que tratan de solventar esta limitación, ya sea transformando las variables o eliminando valores extremos, el modelo se distorsiona (Balcaen y Ooghe, 2006). Cuando las matrices de dispersión difieren, los investigadores tienden a transformar los datos para aproximar los valores de dispersión (Taffler, 1982), en lugar de utilizar el modelo cuadrático, que permitiría resolver el problema usando los datos originales (Zavgren 1983), ya que el modelo cuadrático resulta complejo¹² y solo en condiciones muy concretas consigue mejores resultados que el modelo lineal (Balcaen y Ooghe, 2006).

¹² El número de parámetros a estimar en el caso cuadrático es mucho mayor que el lineal. Esto hace que, excepto en muestras muy grandes, la discriminación cuadrática resulte bastante inestable y, aunque las matrices de covarianzas sean muy diferentes, con frecuencia se obtengan mejores resultados con la función lineal. La función discriminante cuadrática también resulta muy sensible a

Mientras el análisis discriminante lineal domina la literatura, se han desarrollado una serie de técnicas recientes para resolver cada uno de los inconvenientes que describe esta metodología. Entre estas técnicas recientes se encuentra el análisis discriminante cuadrático, el logístico y el del K-ésimo vecino más próximo, siendo estos dos últimos parcialmente paramétrico y no paramétrico respectivamente, a diferencia del lineal y el cuadrático. En la segunda parte de la tesis aplicamos los cuatro modelos.

1.3.3. Regresión logística: logit y probit

En 1977 Martin aplica al estudio de esta cuestión la regresión logística con el fin de evitar los inconvenientes metodológicos presentes en el análisis discriminante, especialmente la falta de normalidad de algunas variables. La regresión logística binaria o logit binario es un análisis de regresión en el que la variable independiente toma valores en el intervalo $[0,1]$ e indica la probabilidad de pertenencia a un grupo, el de empresas sanas o el de empresas fracasadas. Puede interpretarse que los coeficientes de las variables independientes miden la importancia relativa de los factores seleccionados para explicar la probabilidad de fallido que genera el modelo (Ohlson, 1980; Zavgren, 1985, Laitinen y Kankaanpää, 1999).

La técnica de regresión logística cumple objetivos similares al análisis discriminante, pero con la ventaja de ser más flexible, ya que permite trabajar con muestras no proporcionales (Hair *et al.*, 1999). Sus premisas son menos estrictas: no requiere que las variables se distribuyan con normalidad multivariante ni que las matrices de dispersión sean iguales (Ohlson, 1980; Zavgren, 1983). Se mantienen, en cambio, algunas limitaciones en común con la metodología anterior. Así, se asume que la variable dependiente es dicotómica, siendo los dos grupos identificables, discretos y sin solapamiento, lo que es de difícil cumplimiento para

desviaciones de la normalidad de datos, aunque se utilice en el análisis de determinadas poblaciones no normales (Lachembruch, 1975). La evidencia disponible no aplicada a fracaso empresarial parece indicar que la clasificación lineal es en general más robusta (Peña, 2002).

la población en estudio, como se ha explicado en el apartado 1.1. Por otro lado, hay que tener en cuenta el coste de los errores de mala clasificación para establecer el punto de corte del modelo entre empresas sanas y fracasadas, pero estos costes tienen un elevado componente subjetivo. Balcaen y Ooghe (2006) señalan que los modelos de análisis logístico son muy sensibles a la multicolinealidad, a la existencia de valores extremos y a la falta de datos omitidos.

Jones y Hensher (2004; 2008) introducen el *mixed logit*¹³, en el que se incluyen una serie de parámetros adicionales para capturar la heterogeneidad de una empresa en el tiempo y respecto a otras empresas. Además de los parámetros fijos, estos modelos estiman la desviación típica y la media de los parámetros aleatorios, así como la heterogeneidad de las medias. Así, la mejor información sobre el comportamiento de los datos de la muestra mejora la capacidad explicativa y predictiva respecto al análisis logístico clásico.

Los modelos probit utilizan una distribución normal acumulativa¹⁴ en lugar de una distribución logística, pero en todo lo demás los comentarios anteriores resultan aplicables. En la comparación que algunos autores hacen de modelos probit y logit frente al análisis discriminante, no hay resultados concluyentes (Lo, 1986; Laitinen y Kankaanpää, 1999), aunque Lennox (1999) sí obtiene mejores resultados con probit y logit, posiblemente por una mejor especificación de las variables. Baum (2006) realiza una comparación entre logit y probit encontrando que las muestras donde la proporción de probabilidades es diferente serán sensibles a la elección de la función de distribución acumulativa. Por tanto, como la distribución logística (logit) tiene colas más gruesas, dará lugar a un mejor ajuste del modelo.

¹³ También recibe los nombres de *logit* de parámetros aleatorios (RPL), *mixed multinomial logit* (MMNL), *logit* híbrido o *logit* de componentes del error.

¹⁴ La forma de esta función hace que las probabilidades intermedias (entre 0,3 y 0,7, por ejemplo) sean más sensibles al cambio de valor de las variables independientes que las probabilidades cercanas a los extremos (Laitinen y Kankaanpää. 1999).

Para evaluar el funcionamiento de los modelos de regresión logística, Altman *et al.* (2008) construyen curvas ROC (*Receiver Operating Characteristics*), que representan la tasa de positivos verdaderos frente a falsos positivos como umbral de discriminación entre cambios de empresas fracasadas y no fracasadas. El área bajo la curva ROC es una medida de precisión en la predicción, donde el valor 1 representa un modelo perfecto. Tanto el Coeficiente de Gini como el estadístico Kolmogorov-Smirnov (K-S) normalmente utilizados por los analistas de *scoring* derivan de esta medida.

1.3.4. Inteligencia artificial

A partir de la década de los '90 del pasado siglo XX, se introduce en el estudio del fracaso empresarial la aplicación de técnicas de inteligencia artificial, sobre todo redes neuronales y técnicas de inducción de reglas y árboles de decisión. Un trabajo pionero en esta técnica es Bell *et al.* (1990). La inteligencia artificial aplicada al fracaso empresarial consiste en elaborar programas de ordenador capaces de generar conocimiento a través fundamentalmente del análisis de los ratios, utilizando posteriormente ese conocimiento para realizar inferencias sobre nuevos datos. Ravi Kumar y Ravi (2007) analizan de forma minuciosa los trabajos relacionados con el fracaso empresarial, que utilizan diversos métodos de inteligencia artificial, publicados desde 1990 hasta 2005. Estos autores diferencian varias técnicas: la teoría de los conjuntos borrosos, las redes neuronales, el aprendizaje máquina (que incluye los árboles de decisión y el razonamiento basado en casos), la teoría de los conjuntos aproximados (*rough sets*), las máquinas de vectores de apoyo, y la inteligencia computacional o agrupación de varias de estas técnicas de inteligencia artificial en un mismo sistema de decisión. La mayoría de los trabajos publicados en relación con el fracaso utilizan redes neuronales, aprendizaje máquina o conjuntos aproximados, por lo que vamos a centrarnos en estas técnicas.

1.3.4.1. Redes neuronales

Se trata de un sistema de unidades de cálculo interconectadas, llamadas neuronas, dispuestas en niveles. Tres niveles de neuronas –de entrada, ocultas y

de salida- permiten realizar unas u otras tareas según el número de neuronas y la arquitectura de las conexiones. Las arquitecturas más utilizadas son el perceptrón multicapa (MLP), la red de funciones de base radial (RBFN), la red neural probabilística (PNN) y el mapa de rasgos autoorganizativos (SOM). El grupo más amplio de trabajos utiliza redes neuronales de aprendizaje con retropropagación (BPNN) [Tam (1991); Tam y Kiang (1992); Fletcher y Goss (1993); Altman *et al.* (1994); Wilson y Sharda (1994); Tsukuda y Baba (1994); Leshno y Spector (1996); Barniv *et al.* (1997); Bell (1997); Piramuthu (1998); Zhang *et al.* (1999); Charalambous *et al.* (2000), Atiya (2001); Swicegood y Clark (2001); Lee *et al.*¹⁵ (2005)] o bien, mapas de rasgos autoorganizativos (SOM) [Serrano-Cinca (1996); Kiviluoto (1998); Kaski *et al.* (2001); Lee *et al.* (2005)].

De acuerdo con Sarle (1994), las redes neuronales más utilizadas (perceptrón multicapa o BPNN) no son más que regresiones no lineales y modelos discriminantes que se pueden aplicar con programas estadísticos estándar. Se trata simplemente de modelos estadísticos no paramétricos. De hecho, el análisis discriminante logístico puede verse como un caso particular de red neuronal sencilla. En comparación con técnicas estadísticas tales como el análisis discriminante lineal, el análisis discriminante cuadrático, la regresión logística o el análisis probit, las redes neuronales se muestran significativamente mejores¹⁶, tanto en la predicción como en la estimación de la tasa de clasificación (Zhang *et al.*, 1999), especialmente cuando los datos disponibles son limitados (Fletcher y Gross, 1993). Esto no es sorprendente si tenemos en cuenta que el BPNN con función de activación logística se puede ver como un conjunto de regresiones logísticas ajustadas en paralelo, de manera que el método puede captar mejor las relaciones no lineales entre los datos (Ravi Kumar y Ravi, 2007). Las redes neuronales presentan la ventaja adicional de su robustez: no se asume ninguna distribución de probabilidad o iguales dispersiones para los grupos, las funciones *input* y *output* no están sujetas a restricciones salvo que sean continuas y diferenciables (Tam, 1991; Tam y Kiang, 1992; Wilson y Sharda, 1994). Por otra

¹⁵ De acuerdo con Lee *et al.* (2005), las redes con retropropagación funcionan mejor cuando se puede usar un vector objetivo. En el caso de predicción de fracaso empresarial, el vector objetivo es "si la empresa ha fracasado o no".

¹⁶ En algunos trabajos no se obtienen mejores resultados con redes neuronales que con los modelos lineales (Altman *et al.*, 1994).

parte, hay que destacar su capacidad de adaptación a cambios en la realidad que analizan: los nuevos ejemplos van ajustando el modelo, de manera que una proporción de la muestra anterior se mantiene en función de la estabilidad de la distribución y el nivel de ruido de dicha muestra (Tam, 1991; Tam y Kiang, 1992). Zhang *et al.* (1999) demuestran además que el modelo de redes neuronales que utilizan es robusto a cambios en la muestra.

A cambio, presentan algunas limitaciones propias de este tipo de métodos. Al no haber un método formal para derivar la configuración de la red a partir de una determinada tarea de clasificación, las neuronas ocultas pueden ser arbitrariamente numerosas (Tam, 1991; Tam y Kiang, 1992). Otro gran inconveniente que presenta este tipo de técnica es la gran cantidad de tiempo necesario para el desarrollo de la aplicación (Altman *et al.*, 1994; Piramuthu *et al.*¹⁷, 1998; De Andrés, 2000), aunque se ha detectado que algunos modelos – como los que utilizan retropropagación– resultan más eficientes computacionalmente (Tam, 1991; Tam y Kiang, 1992). En cuanto a la capacidad explicativa de los modelos, los coeficientes de las variables resultan difíciles de interpretar (Altman *et al.* 1994), especialmente cuando el método incorpora un mecanismo para reducir la dimensión del espacio de variables (Tam y Kiang, 1992). Esto se debe a la falta de transparencia en el uso de las variables para las conexiones neuronales (Altman *et al.*, 1994). Como en los métodos tratados antes, permiten una predicción (binaria) del fracaso pero no permiten calcular una probabilidad de fallido¹⁸ (Atiya, 2001). La composición del grupo de empresas que sirven para entrenar al modelo es determinante en la precisión final de las predicciones: parece que debería haber una muestra equitativa de empresas fracasadas y no fracasadas, aunque no coincida con la proporción de empresas de ambos tipos en la población (Wilson y Sharda, 1994). También se les ha criticado por su ausencia de fundamento teórico (Altman y Saunders, 1998).

¹⁷ Piramuthu *et al.* (1998) consiguen reducciones en el tiempo de aprendizaje y mejoras en la precisión de la clasificación usando metodología de construcción de rasgos.

¹⁸ En esta cuestión discrepan otros autores, que afirman que las redes neuronales generan estimaciones de probabilidades posteriores (Zhang *et al.*, 1999).

Dentro de la categoría de las redes neuronales, los mapas autoorganizativos son un modelo de red neuronal no supervisado. Esta técnica, utilizada por Serrano-Cinca en 1996, facilita una visión gráfica intuitiva sobre la información relativa al riesgo de quiebra. De acuerdo con Kaski *et al.* (2001), sus ventajas son: que proporciona al analista una idea más precisa de la situación de la empresa; que permite identificar diferentes tipos de comportamientos o trayectorias de las empresas; y que permite determinar la dependencia entre los indicadores contables y la situación de quiebra¹⁹. Además de los inconvenientes propios de las redes neuronales, como el tratamiento de “caja negra” que se utiliza para el manejo de datos, los mapas de rasgos autoorganizativos presentan otras desventajas (Lee *et al.*, 2005): menor precisión que las redes neuronales supervisadas (cuando existe vector objetivo); es necesario decidir el número de grupos que se va a incluir; e identificar las características del grupo para poder realizar la clasificación.

1.3.4.2. Algoritmo de particiones recursivas

En el grupo de aprendizaje máquina se encuentran los algoritmos de particiones recursivas, como los árboles de decisión (Marais *et al.*, 1984; Frydman *et al.*, 1985). El algoritmo de particiones recursivas o iterativas está basado en el desarrollo de un árbol de clasificaciones binarias (empresas sanas y fracasadas), que en cada nudo clasifica en dos grupos mediante una función univariante, con un ratio como variable independiente (un mismo ratio puede usarse más de una vez en el proceso). Esta técnica no paramétrica genera unas reglas del tipo “si, entonces” binarias por producto, que se pueden utilizar como un sistema experto de aviso temprano.

No está sujeta a algunas restricciones del análisis discriminante, como la normalidad de la distribución de probabilidad y la homoscedasticidad. Frydman *et al.* (1985) llegan a un nivel de clasificación correcta de empresas fracasadas y sanas superior al análisis discriminante. Aunque en comparación con las redes neuronales, que funcione mejor uno u otro modelo depende del conjunto de datos

¹⁹ Según Kaski *et al.* (2001), las dos últimas ventajas suponen un avance en el contraste empírico de algunas de las aportaciones teóricas de Argenti (1976).

sobre el que se aplique. Uno de los inconvenientes que presenta esta técnica es que no permite establecer comparaciones entre empresas que estén en el mismo grupo porque todas tienen la misma clasificación. También parece ser más sensible a los costes y a la información previa que los modelos de análisis discriminante (Altman, 1993).

1.3.4.3. Técnicas de escalamiento multidimensional

Esta metodología fue utilizada por Mar Molinero y Ezzamel (1991) para la predicción del fracaso empresarial por presentar una interpretación intuitiva. Aunque los resultados de acierto en la clasificación de las empresas no difieren de los métodos estadísticos tradicionales, su uso está justificado dado que no es necesaria la comprobación del modelo ni la transformación de los datos originales con el fin de que cumplan determinadas propiedades. La representación gráfica que muestra debe tomarse como una herramienta donde se plasman las relaciones financieras de las empresas sanas y fracasadas y la forma en que evolucionan en el tiempo.

1.3.4.4. Otros métodos de inteligencia artificial

Entre el amplio grupo de métodos que abarca la inteligencia artificial, también se han aplicado a los estudios sobre fracaso empresarial los conjuntos aproximados o los algoritmos genéticos (Ravi Kumar y Ravi, 2007). Los conjuntos aproximados (Slowinski y Zopounidis, 1995; McKee, 2000) generan reglas del tipo “si, entonces” aplicadas a valores ordinales para realizar tareas de clasificación. Parecen funcionar mejor que la regresión logística, el análisis discriminante y los árboles de decisión, pero en general son poco precisos y resultan sensibles a cambios en los datos. Los algoritmos genéticos (Shin y Lee, 2002) se basan en el principio evolutivo, generando soluciones hacia valores óptimos, los cuales dependerán de la adecuada codificación de dichas soluciones. Las dificultades que presenta esta metodología son las generales a todos los modelos de inteligencia artificial, no pudiendo mejorar la precisión de clasificación que consiguen otras técnicas.

1.3.5. Análisis envolvente de datos (DEA)

Una línea de investigación reciente utiliza la metodología de fronteras DEA (análisis envolvente de datos o *Data Envelopment Analysis* en terminología inglesa) como una nueva herramienta para la evaluación de quiebra. El DEA es una técnica que compara la eficiencia de las unidades productivas a partir de los datos de sus *inputs* y sus *outputs*. Troutt *et al.* (1996) plantean la utilización de DEA en análisis de riesgo de quiebra, pero sólo desde el punto de vista teórico. Según estos autores, la técnica puede utilizarse para desarrollar una frontera o límite de aceptación en la toma de decisiones basada en casos. Si se respetan las escasas premisas del modelo: monotonía condicional, convexidad del grupo aceptable, muestra representativa y selectividad no restringida, los casos que se sitúen sobre o por encima de la frontera serían aceptables. Las ventajas del DEA son (Premachandra *et al.*, 2009; 2011): que no establece limitaciones a priori sobre la relación entre los *inputs* y los *outputs*; no necesita que se especifique la distribución de las variables²⁰; no requiere una muestra de gran tamaño, a diferencia de otras metodologías estadísticas y econométricas; y genera puntuaciones de desempeño individual respecto a la muestra de estudio.

Paradi *et al.* (2004) usan un modelo DEA no radial mientras que Cielen *et al.* (2004) utilizan un modelo DEA radial (el modelo CCR) en comparación con análisis discriminante. Algunos problemas en la aplicación del DEA en estos trabajos iniciales sobre fracaso empresarial fueron: la existencia de datos negativos en las variables financieras (el modelo CCR no admite variables negativas); el desequilibrio de las poblaciones de empresas fracasadas y no fracasadas; y el hecho de que el fracaso sea un proceso dinámico a lo largo del tiempo. La etapa inicial de esta metodología se centró en la resolución de estos inconvenientes (Sueyoshi, 2005; Chang y Kuo, 2008; Shetty *et al.*, 2012). En general, se han utilizado muestras y cantidades de variables pequeñas cuando se han aplicado innovaciones en los métodos; y variables similares a las empleadas en trabajos previos cuando los autores comparan el desempeño del DEA con

²⁰ Los ratios contables tienden a seguir distribuciones no normales que requieren transformaciones de datos para cumplir con la premisa del DA sobre distribución multivariante normal. Además, las pymes muestran valores con dispersión elevada en cada ratio, lo que contribuye a empeorar el problema. Esta es la razón de que DEA sea propuesto por los autores que lo utilizan como un buen método para cualquier tipo de muestra.

metodologías de uso frecuente como el análisis discriminante o las técnicas binomiales.

Además, la tarea de separar las variables entre *inputs* y *outputs* puede verse como una tarea subjetiva y difícil (Sueyoshi y Goto, 2009c). Premachandra *et al.* (2009) usan un modelo no radial (el modelo aditivo) para comparar la efectividad de los modelos DEA y de las regresiones logit en la detección y predicción de quiebras. El modelo aditivo tiene la propiedad de invarianza de la traslación, por lo que permite el uso de datos negativos. Tratando de resolver las otras limitaciones del DEA en este ámbito, otros trabajos (Sueyoshi y Goto, 2009a; 2009b; y 2009c) combinan el modelo aditivo de DEA con análisis discriminante (DEA-DA), consiguiendo una mayor capacidad analítica. El DEA-DA supera el problema de desequilibrio de las poblaciones de los dos tipos de empresas, fracasadas y no fracasadas, porque permite controlar la importancia de los dos grupos. Kuo (2007) propone un modelo DEA de estratificación asimétrica, que elabora dos fronteras, una con las peores entre las empresas no quebradas y otra con las quebradas, y, a partir de ahí, elimina los diferentes estratos de solapamiento entre ambas fronteras para mejorar la precisión en la clasificación.

Unos pocos trabajos se han centrado en la mejor manera de reducir el número de variables para aplicar DEA, aunque solo en estudios aplicados al desempeño (*performance*) de las organizaciones y no al análisis de fracaso empresarial. Algunos autores (Seifor y Thrall, 1990; Jenkins y Anderson, 2003) señalan que a medida que el número de factores (*inputs* más *outputs*) aumenta, la capacidad de discriminar entre empresas (DMUs) disminuye y no es suficiente que el número de empresas sea mayor que el número de factores. Boussofiane *et al.* (1991) y Friedman y Sinuany-Stern (1998) son más precisos y limitan el número de variables *input* y *output* a no más de un tercio de las empresas en el análisis. En cambio, un gran número de variables en el análisis es capaz de explicar una mayor porción de las diferencias entre las empresas (Golany y Roll, 1989) obteniendo puntuaciones de eficiencia incluso más elevadas y DMUs más eficientes (Nunamaker, 1985).

De ahí que el objetivo sea usar el mayor número de variables *input* y *output* potencialmente discriminantes, pero un número total tan pequeño como sea posible. En cuanto a la mejor manera de reducir variables, la primera medida, sería omitir las variables altamente correlacionadas con el resto de variables, pero Nunamaker (1985) encuentra que una variable redundante en un modelo de regresión puede no serlo en un modelo DEA. Norman y Stoker (1991) destacan la necesidad de una relación lógica causal que explique por qué las variables influyen en el desempeño. Jenkins y Anderson (2003) usan análisis de regresión y correlación para identificar qué variables podrían ser eliminadas del modelo DEA con la menor pérdida de información²¹. Pero, al igual que en los estudios anteriores, los autores encuentran que los resultados del modelo DEA varían ampliamente dependiendo de qué variables correlacionadas se omiten o se incluyen. Wagner y Shimshak (2007) desarrollan un procedimiento paso a paso que consiste en maximizar (o minimizar) la media del cambio de la eficiencia cuando se agregan u omiten variables en el análisis. Pero estos autores concluyen diciendo que la determinación del “mejor” modelo para representar cualquier situación dada debe apoyarse en los criterios y el conocimiento que los gestores tienen sobre la situación real que se representa.

1.3.6. Otros modelos

Finalmente, algunos autores apuntan la conveniencia de combinar varias medidas para obtener una mayor precisión en el análisis del fracaso empresarial. Park y Han (2002) proponen una estructura de razonamiento analógico denominado análisis multicriterio o proceso de análisis jerárquico (AHP). Una debilidad importante de esta metodología es la sensibilidad de los modelos a la presencia de características irrelevantes en las variables y, además, no mejora los resultados de otros métodos. Canbas *et al.* (2005) componen un sistema de aviso temprano integrado (IEW) aplicado a bancos, donde incluyen análisis discriminante, regresión logística, probit y análisis de componentes principales. Por su parte, Xu y Zhang (2009) incluyen el modelo *Z-Score* de Altman, el *O-Score* de Ohlson y la distancia al fallido basada en la teoría de opciones. Dentro del

²¹ Su método de estadística multivariante mide la pérdida de información por la proporción de la varianza total en todas las variables perdidas, omitiendo una o más variables.

grupo de métodos de inteligencia artificial, la integración parece ser la forma de contrarrestar las deficiencias de los métodos individuales. Por ejemplo, Serrano-Cinca (1996) integra análisis discriminante lineal con redes neuronales y mapas de rasgos autoorganizativos; Ahn *et al.* (2000) proponen un modelo híbrido que combina conjuntos aproximados y redes neuronales; y Min *et al.* (2006) construyen otro modelo híbrido con algoritmos genéticos y máquina de vector de apoyo. En todos los casos, el poder predictivo es mejor que el individual de cada uno de estos modelos.

En el Anexo II se recogen por orden cronológico los autores de los trabajos, tanto extranjeros como españoles, que hemos revisado y la clase de metodología empleada en cada trabajo. Se puede apreciar que en España los primeros trabajos sobre fracaso empresarial aparecen a mediados de los años 80, con casi veinte años de retraso respecto al inicio de esta línea de investigación en el ámbito internacional. Los trabajos de Laffarga *et al.* (1985; 1987) pueden considerarse pioneros en nuestro país. El primero aplica ANOVA y análisis discriminante al sector bancario, mientras que el segundo incorpora la regresión logística. La utilización en Gabás (1990) del análisis de particiones recursivas, técnica que se había empezado a aplicar al fracaso empresarial en 1984, indica ya un notable recorte en el retraso. Y a partir de ese momento se puede decir que el retraso en la aplicación de los nuevos modelos desaparece, de manera que autores españoles como Mar Molinero y Ezzamel (1991) o Serrano Cinca (1996) publican modelos innovadores en el ámbito internacional, y otros autores aplican modelos de inteligencia artificial de forma simultánea a su desarrollo en los países más avanzados en la investigación sobre la cuestión.

En conjunto, la evolución de las metodologías muestra una tendencia hacia la utilización de modelos con menos restricciones y hacia la mejora en las formas de medir la significatividad de las variables. Sin embargo, cuando la evolución en los modelos conlleva una relajación en el rigor metodológico, se condiciona la validez del estudio porque pueden aparecer diversos problemas. Los más frecuentes son: la correlación entre variables, cuando algunos ratios empleados comparten

numerador o denominador; y la heteroscedasticidad en los residuos de las regresiones, tal como señalan Correa *et al.* (2003) y Acosta y Fernández (2007).

En el apartado dedicado a los modelos, queremos señalar que la aplicación a muestras específicas, como sectores concretos o empresas en condiciones determinadas, homogéneas para toda la muestra (tamaño, ciclo de vida, etc.); junto con la todavía escasa incorporación de otras variables de naturaleza no financiera, que se ha mencionado en el apartado anterior, han permitido el desarrollo de modelos de detección y predicción del fracaso empresarial más precisos.

Así, por ejemplo, se desarrollan modelos en varios sectores (Lincoln, 1984; Mensah, 1984; McGahan y Porter, 1997; Román *et al.*, 2001) o en sectores concretos, como el de seguros (Rodríguez Acebes, 1990; López *et al.* 1994; Mora, 1994b; Lee y Urrutia, 1996), la banca (Laffarga *et al.*, 1985; Rodríguez Fernández, 1986; Pina Martínez 1989; Bell *et al.*, 1990; Whalen, 1991; Serrano y Martín, 1993; Wheelock y Wilson, 2000; Canbas *et al.*, 2005), el comercio minorista (Bhargava *et al.*, 1998; McGurr y DeVaney, 1998), las empresas constructoras (Sueyoshi y Goto, 2009b) u otros (McDonald y Morris, 1984; Platt y Platt, 2002; Acosta y Fernández, 2007; Sueyoshi y Goto, 2009a).

Algunos trabajos se centran en grupos de empresas específicos, con algún tipo de característica homogénea. Por ejemplo, hay estudios para pymes (García *et al.* 1995; Ramírez, 1996; Lizarraga, 1997; Gallego *et al.* 1997a; López *et al.*, 1998; Correa *et al.*, 2003; De la Torre *et al.*, 2005; Rodríguez y Díaz, 2005; Madrid y García, 2006; Gómez *et al.* 2008; Altman y Sabato 2005 y 2007; Altman *et al.*, 2008) o sólo para empresas medianas (Lizarraga, 1997), o para zonas geográficas menores (Gandía *et al.*, 1995; Ferrando y Blanco, 1998; Crespo, 2000; Correa *et al.*, 2003; Rubio Misas, 2008). Dewaelheyns y Van Hulle (2004; 2006) muestran que el poder predictivo de las variables explicativas es diferente para empresas que forman parte de un grupo frente a las individuales.

1.4. Conclusiones a la primera parte

Tal como señalaba De Andrés (2000) constatamos que no existe una teoría aceptada académicamente del fracaso empresarial. Sobre el concepto de empresa fracasada, resulta más objetivo adoptar como tales a las empresas que han sido oficialmente declaradas en situación de quiebra, suspensión, o similar, pero eso plantea el problema de que la declaración está sujeta a la normativa nacional correspondiente, y puede no implicar una distinción clara, en cuanto a características económico-financieras o de otro tipo, respecto a empresas en dificultades no declaradas oficialmente. La consideración de situaciones diferentes de fracaso y de fases distintas del deterioro de las empresas se ha demostrado empíricamente que influye sobre los modelos, ya sea con ponderaciones distintas para las variables explicativas o incluso con la incorporación de algunas variables explicativas distintas en los modelos.

Respecto a la teoría sobre el fracaso empresarial que subyace a los modelos, esta parece quedar asentada en los planteamientos y desarrollos iniciales de autores clásicos como Beaver o Argenti, que se basan en el razonamiento económico sobre el funcionamiento de las empresas y en el análisis de casos, más o menos formalizados, de empresas fracasadas. Pero esa teoría básica no es suficiente para concretar modelos y variables válidos con carácter general y universal. Las condiciones específicas de la normativa de cada país, como se ha mencionado antes, y las condiciones ajenas al mero funcionamiento económico-financiero, también determinantes para el fracaso empresarial, impiden que la teoría general vaya más allá de esas grandes líneas maestras. Queda pendiente de resolver si los modelos son de detección o de predicción de empresas fracasadas, puesto que el concepto utilizado de fracaso y el modelo desarrollado no encajan con las intenciones de predecir que declaran muchos autores.

Siguiendo el razonamiento inductivo de Altman, es precisamente la evidencia empírica la que ha ratificado si las variables derivadas de los razonamientos de los

autores mencionados siguen teniendo vigencia y son aplicables al análisis del fracaso de la población concreta de que se trate, situada en un periodo y en una zona geográfica determinados. En las dos últimas décadas, el avance de la línea de investigación se ha producido en las técnicas empíricas. Resulta patente que hay una evolución en las metodologías, han ido apareciendo nuevas formas y planteamientos para soslayar limitaciones que planteaban las metodologías anteriores, pero eso no ha implicado la desaparición de éstas. Las razones son: que no se ha demostrado un poder predictivo mejor que fuese atribuible en exclusiva a la metodología; y que cualquiera de las metodologías que han ido apareciendo presenta sus propias limitaciones. El resultado es la coexistencia de las distintas metodologías y un desarrollo paralelo, debido a la mejora de herramientas de cálculo (nuevas pruebas y contrastes estadísticos, matemáticos o econométricos) y a la mejora de la información disponible (aumento del número de datos o de variables, información internacional más homogénea, etc.).

Como causa o consecuencia de esas grandes líneas maestras de la teoría sobre el fracaso empresarial, unas pocas variables parecen estar con gran frecuencia en los modelos: las más generales serían rentabilidad económica y equilibrio económico-financiero. Pero, para evitar el problema de las innumerables especificaciones posibles de los ratios que aproximan un mismo rasgo de la empresa, nuestro trabajo va más allá de la frecuencia de las variables, de manera que hemos identificado los rasgos de la empresa detrás de las variables explicativas. Así, para la muestra de trabajos empíricos analizados, hemos identificado los rasgos siguientes por orden de importancia (frecuencia): rentabilidad, endeudamiento y equilibrio económico-financiero, que aparecerían en todos los trabajos, incluso con más de un ratio por rasgo en algunos casos; seguidos de estructura económica, márgenes y rotaciones, que aparecen con una gran frecuencia, quedando cualquier otro grupo (identificativo de rasgo) a bastante distancia de los anteriores. Nótese que la mayor dispersión en la formulación del ratio de endeudamiento hace que no saliera en segundo lugar en el estudio de frecuencia de ratios individuales.

Las aportaciones más recientes muestran que, cuando se añaden variables empresariales ajenas a los estados financieros, o variables del entorno de la empresa (por ejemplo, sectoriales), es posible conseguir mejoras en la capacidad explicativa de los modelos, si bien los resultados varían notablemente de unos países a otros. En todo caso, el número de trabajos es aún escaso para que la evidencia resulte concluyente. Las variables macroeconómicas, también consideradas escasamente hasta ahora, han aportado evidencias contradictorias.

En cuanto a la tendencia a ajustar los modelos a grupos específicos de empresas con características homogéneas, por tamaño, situación geográfica, actividad económica, fase del ciclo de vida de la empresa, etc., esto sólo puede hacerse cuando la información es suficiente para que la muestra tenga el tamaño crítico mínimo para aplicar la metodología correspondiente. Por esta razón, salvo en países muy grandes y con información ampliamente disponible, como los Estados Unidos, es difícil hacer estudios por país, o zonas geográficas inferiores. La alternativa es realizar estudios internacionales para aumentar el tamaño muestral, pero en este caso la normativa y las condiciones macroeconómicas diversas plantean otros problemas de comparabilidad de la información, que los modelos deben resolver. El problema de la normativa se vería notablemente mermado con la convergencia internacional de la normativa contable, como se destacaba en la introducción de la tesis.

Ante la pregunta de si cada uno de los modelos, metodologías, conceptos y ratios utilizados en la literatura del fracaso empresarial es válido en el espacio y en el tiempo para otros datos, o tenemos que diseñarlos *ad-hoc* para cada país, región, ciudad, actividad, etc., nuestro estudio sugiere que una parte es común mientras otra depende de las condiciones del grupo específico analizado y va evolucionando a medida que lo hacen las condiciones del entorno. Cuanto más rápido cambien las condiciones, menos útil resultará un modelo estático, que no permita incorporar esas variaciones, o que siendo dinámico se alimente de información alejada en el tiempo de la fecha de predicción.

En general, detectamos dos problemas en las aproximaciones previas que limitan de forma notable la aplicación de los modelos de predicción de fracaso empresarial a la gestión del riesgo de crédito. El primero es el resultado que se obtiene de esos modelos: una variable dicotómica que dice si la empresa fracasa o no fracasa pero que no permite grados en la distancia al éxito o el fracaso. El segundo atañe a la ausencia, en la mayoría de los estudios, de factores ajenos a la propia empresa analizada, de manera que, en épocas de cambio, hasta que la empresa no sufre en sus variables los efectos de esos valores externos, los modelos no son capaces de detectar el deterioro. En el caso de la información contable, las variables obtenidas de esta importante fuente se refieren al estado patrimonial y a las rentas de varios meses atrás, por lo que un modelo que utilice entre sus variables ratios contables puede tardar en detectar problemas lo suficiente para que sea demasiado tarde.

Añadimos una reflexión sobre la escasa presencia de variables macroeconómicas en los modelos de predicción del fracaso empresarial, hasta ahora. El tipo de interés de mercado, la tasa impositiva, el índice de inflación, la evolución del PIB u otras variables similares recogen en parte algunas de las causas que pueden contribuir al éxito o fracaso de las empresas, pero las variables contables sólo varios meses más tarde se verán afectadas por el efecto de la variación de esas variables. Así, en épocas de cambio, las variables del entorno pueden contribuir a la predicción antes que las variables de la propia empresa. Sin embargo, nuestro estudio muestra un muy escaso número de trabajos que incluyan este tipo de variables y en algunos de ellos parecen aportar poco poder predictivo. Sí parecen variables relevantes cuando el número de años permite incluir en la muestra fases expansivas y recesivas de la economía. Por lo que quizá sería necesario estimar modelos que recojan variables macroeconómicas a partir de un número de años que incluya las oscilaciones de ciclo económico.

2. SELECCIÓN DE VARIABLES EN EL ANÁLISIS DEL FRACASO EMPRESARIAL: SELECCIÓN DE MEDIAS VS. SELECCIÓN DE MEDIANAS

2.1. Introducción a la segunda parte

Como ha quedado de manifiesto en la primera parte de esta tesis, la ausencia de una teoría sólida sobre fracaso empresarial hace que, a pesar del gran número de trabajos empíricos y teóricos sobre la cuestión, esté todavía en desarrollo el mapa conceptual de relaciones entre los factores explicativos. No hay un marco conceptual que vincule: el comportamiento económico de las empresas, aproximado por variables internas, mayormente obtenidas de la información contable; los factores externos, tanto sectoriales como macroeconómicos, aproximados por algunas variables externas disponibles; y el fracaso efectivo o la posibilidad de fracaso de esas empresas. Como consecuencia, el contraste empírico de hipótesis en este ámbito tropieza con problemas en uno o varios de estos tres elementos clave en esa teoría incompleta del fracaso empresarial: el concepto de fracaso, el modelo de identificación o predicción propuesto y la forma de seleccionar variables que puedan actuar como factores discriminantes.

En relación con la tercera cuestión, varios autores (Keasey y Watson, 1991; Platt y Platt, 1990; Sueyosi y Goto 2009c) han señalado que la selección de un conjunto adecuado de variables para aplicar un modelo en particular puede parecer trivial pero es una parte importante de la evaluación de la quiebra. No obstante, los avances en esta línea han llegado a un nivel de desarrollo mucho menor que las metodologías para evaluar el fracaso empresarial.

Si en la primera parte de la tesis nos ocupamos de la selección de variables a partir de los resultados obtenidos por la literatura previa, en esta segunda parte

nos centramos en la selección de variables basada en análisis estadísticos. En el apartado 1.2, aparte de identificar las variables que la muestra de trabajos empíricos analizados encuentran significativas con mayor frecuencia, hemos contribuido con una identificación de rasgos económicos subyacentes a esas variables, que permite una mejor ordenación de los factores más relevantes para la discriminación entre empresas fracasadas y sanas. En esta segunda parte realizamos el mismo estudio pero aplicado a trabajos empíricos sobre pymes, confirmado que los rasgos económicos discriminantes son los mismos, si bien, cambia un poco el orden. El endeudamiento, la estructura económica, la rentabilidad, las rotaciones, el equilibrio y el margen aparecen como los rasgos económicos con más poder discriminante, lo cual orienta nuestra selección de las variables individuales que se van a utilizar en los análisis empíricos realizados en la segunda y en la tercera partes de la tesis.

Para llevar a cabo nuestro estudio sobre la selección estadística de variables, escogemos el grupo de estudio más problemático para el tratamiento de los datos: las pequeñas empresas. En primer lugar, aplicamos el contraste de diferencias en medias, que es el más comúnmente utilizado, y después un contraste de diferencias en medianas (*Rank Sum Test*) introducido recientemente en trabajos sobre fracaso empresarial (Sueyoshi y Goto, 2009a; 2009b; 2009c). Nuestros resultados muestran que las variables significativas difieren radicalmente usando uno u otro contraste. En este sentido, la principal contribución de esta segunda parte de la tesis es la identificación de las implicaciones de usar diferencias en medias o diferencias en medianas para seleccionar variables, cuando se utiliza un variado grupo de modelos incluidos en el análisis discriminante, probit, logit y análisis envolvente de datos (DEA). Nuestros resultados revelan que la naturaleza de la muestra determina no sólo el contraste estadístico más apropiado para la selección de variables, sino también los métodos más apropiados para evaluar el fracaso de las empresas. Además, nuestro trabajo incorpora una variable cualitativa, sobre incidencias de pago, que contribuye a la literatura previa con nueva evidencia sobre el papel relevante que este tipo de información desempeña en el análisis de fracaso empresarial de pequeñas empresas.

2.2. Variables en la literatura previa sobre pymes

En la primera parte de la tesis hemos hecho una revisión de la literatura en cuanto a variables y metodologías que resultan pertinentes tener en cuenta para la interpretación de los resultados de esta segunda parte. En el caso de las variables, hemos revisado un amplio número de trabajos empíricos previos para identificar cuáles se han hallado significativas con mayor frecuencia y los rasgos económicos que esas variables aproximan. En este momento, dado que nuestros análisis empíricos se van a orientar a empresas pequeñas, hemos repetido el estudio de frecuencias de variables individuales y de rasgos económicos subyacentes a esas variables para el subgrupo de trabajos empíricos donde se analizan pymes. La información completa sobre los trabajos empíricos revisados y las variables significativas se incluye en una tabla en el Anexo III. En la interpretación de los resultados hay que considerar que el número de trabajos analizados es mucho menor (23 trabajos); y que la mayoría de ellos se refieren a pymes españolas²².

²² A nivel internacional, la proporción de trabajos sobre fracaso en empresas grandes y cotizadas es mucho mayor. Pero en España el mercado bursátil es pequeño, quizá por eso hay una buena proporción de estudios sobre pymes.

Tabla 5. Ratios más frecuentes en pymes

VARIABLES EXPLICATIVAS	Nº de trabajos
Activo Circulante/Pasivo Circulante	8
Cash Flow Recursos Generados/Activo Total	8
BAIT/Activo Total	7
Deuda Total/Activo Total	7
Activo Circulante/Activo Total	6
Ventas/Activo Total	6
BAIT/Gastos Financieros	5
Beneficio Neto/Activo Total	5
Beneficio Neto/Ventas	4
Gastos Financieros/Pasivo exigible	4
Disponible+Realizable/Activo Total	4
Capital Circulante/Activo Total	4
Patrimonio Neto/Activo Total	4
Beneficios No distribuidos/Activo Total	3

Como se aprecia en las Tablas 5 y 6, hay pocas diferencias en cuanto a los factores financieros más relevantes. Ocho de los ratios que salían con mayor frecuencia en general, aparecen entre los diez más frecuentes en las pymes (en la Tabla 5 aparecen en azul nueve de los diez ratios más frecuentes del estudio general realizado para cualquier tamaño de empresa). El segundo ratio en frecuencia en esta muestra de estudios sobre pymes (*cash flow* sobre activo total) podemos considerarlo prácticamente equivalente al situado en el puesto octavo en general (*cash flow* sobre pasivo total). Con menor importancia, en el puesto diez, aparecen los gastos financieros sobre el pasivo exigible, si bien en el puesto ocho aparece un ratio que no teníamos en el estudio general y que relaciona los gastos financieros con el BAIT. Más alejado, en el puesto catorce, encontramos el ratio de beneficios no distribuidos sobre activo total. Esta menor importancia resulta coherente con la menor representatividad de los fondos propios de las pymes en lo que se refiere a la evolución de este tipo de empresas. El menor tamaño en términos absolutos hace que la empresa pueda evolucionar positiva o negativamente sin que capital y reservas varíen proporcionalmente, o que aumentos o reducciones de capital y reservas supongan variaciones porcentuales

elevadas de estas partidas. Esto haría que su capacidad como indicador de proximidad a la quiebra sea menos fiable que en las empresas grandes.

Tabla 6. Factores económicos subyacentes a las variables en pymes

Categorías de ratios	Items	% s/ Total	Ratios distintos
Endeudamiento	44	16,36%	26
Estructura Económica	43	15,99%	21
Rentabilidad	41	15,24%	22
Rotaciones	34	12,64%	21
Equilibrio Económico - Financiero	33	12,27%	18
Margen	28	10,41%	16
Subtotal	223	82,90%	124
Otros ratios	38	14,13%	23
Variables	8	2,97%	8
Totales	269	100,00%	155

En cuanto a los factores económicos subyacentes, esto es, los aspectos de la empresa que han resultado relevantes con mayor frecuencia en la muestra de trabajos sobre pymes para la identificación o predicción del fracaso empresarial, los resultados coinciden aunque cambia el orden de importancia. Tal como se podía esperar, las variables contables, y en concreto las que aproximan rentabilidad, endeudamiento, equilibrio económico-financiero, estructura económica, margen y rotaciones, ganan peso frente a otras variables. En estos trabajos sobre pymes representan el 83% de las variables significativas frente a un 72% en los estudios para cualquier tamaño de empresa. Los indicadores financieros relativos a endeudamiento y estructura económica parecen resultar más significativos que los indicadores de rentabilidad. También es un resultado esperable que pierdan peso las variables que no son ratios contables (3% en pymes, frente al 7% en estudios para cualquier tamaño de empresa).

De este modo, confirmamos que el primer nivel de selección de variables (a partir de la literatura previa) es un buen punto de partida para la selección estadística de

variables que se va a acometer en esta segunda parte, aunque la muestra de trabajos incluyera empresas de cualquier tamaño, en vez de solo pequeñas.

2.3. Metodologías que se van a aplicar

Este epígrafe contiene una breve descripción de los métodos utilizados, tal como se aplican en *Stata*²³, el software elegido para ejecutar los modelos. La relación de instrucciones para *Stata* se recoge en el Anexo IV y los resultados en el Anexo V. Las metodologías son cuatro modelos de análisis discriminante, logit, probit y DEA, que aplicamos sobre los mismos datos, a efectos de comparación. Además, se describen estadísticamente las diez variables seleccionadas previamente y se explica su papel en la evaluación del fracaso empresarial. En la última parte se presentan los resultados obtenidos en la aplicación de modelos de fracaso empresarial.

Los modelos utilizados de análisis discriminante son: el análisis discriminante lineal (LDA), el análisis discriminante cuadrático (QDA), el análisis discriminante logístico (LogDA) y el análisis discriminante del K-ésimo vecino más próximo (KnnDA). Con todos ellos aplicamos probabilidades previas iguales²⁴ y, por tanto, cálculos geométricos (la otra alternativa, que aplicamos también como análisis de robustez, es considerar probabilidades proporcionales, en cuyo caso la probabilidad de clasificación dentro de un grupo es proporcional al área

²³ *Stata Corporation. Stata statistical software: release 11.0.*

²⁴ Uno de los problemas habituales en los modelos de análisis discriminante y regresión logística es la incorporación de probabilidades a priori para los grupos respectivos: empresas fracasadas y empresas no fracasadas (Premachandra *et al.*, 2009). Eisenbeis (1977) sugiere que se asuma una ocurrencia igualmente probable en ambos grupos de una muestra concreta. Si bien esto resulta adecuado para muestras formadas por parejas de empresas, cuando se analiza una población o una muestra aleatoria, es poco probable que el porcentaje potencial de empresas fracasadas sea del 50%.

Debido al elevado coste de clasificar como empresas no fracasadas si luego fracasan, la clasificación de este grupo es más meticulosa. Así, para el estudio de empresas fracasadas, con el establecimiento de probabilidades a priori por encima de la proporción de estas empresas en la población, se puede tener en cuenta ese coste de clasificación errónea de las empresas. Es por ello que, en nuestro trabajo, hemos decidido usar proporciones previas iguales (*equal priors*) en las tablas que se presentan. En todo caso, los resultados obtenidos usando proporciones previas proporcionales (*proportional priors*) son similares.

geométrica del mismo, y ciertamente no son iguales). Con KnnDA queremos hallar la línea recta entre los puntos, para lo que usamos la distancia euclídea.

2.3.1. Análisis discriminante lineal (LDA)

A partir de la aproximación de Fisher (1936) podemos encontrar variables canónicas que tengan máximo poder discriminante para clasificar nuevos elementos entre G poblaciones, teniendo en cuenta que la derivación de las funciones discriminantes canónicas supone estimar los valores de los coeficientes de ponderación u_0, u_1, \dots, u_k de las variables X_j de la expresión matemática:

$$f_{gi} = u_0 + u_1x_{1gi} + u_2x_{2gi} + \dots + u_kx_{kgi} \quad , i = 1, 2, \dots, n \quad , g = 1, 2, \dots, G$$

Donde X_1, X_2, \dots, X_k son las variables discriminantes observadas en n individuos que clasificamos a priori en G grupos; f_{gi} es el valor (puntuación) en la función discriminante para el caso i en el grupo g ; x_{jgi} es el valor de la variable discriminante X_j para el caso i en el grupo g .

Los valores u_0, u_1, \dots, u_k permiten obtener combinaciones lineales de las variables discriminantes que hacen máxima la separación entre los grupos establecidos, con el objeto de maximizar las diferencias entre los centroides o valores medios de cada grupo, siguiendo el criterio de máxima variabilidad entre grupos respecto a la variabilidad intra grupos. La matriz de sumas de cuadrados y productos cruzados (T) tiene como expresión:

$$t_{jh} = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{jgi} - \bar{x}_j)(x_{hgi} - \bar{x}_h)$$

Donde x_{jgi} y x_{hgi} son los valores de las variables j y h respectivamente, para el individuo i en el grupo g mientras que \bar{x}_j y \bar{x}_h son los valores medios respectivos de las variables j y h para todos los individuos. A partir de la matriz T se obtiene la matriz total de varianzas-covarianzas. La matriz de sumas de cuadrados y

productos cruzados intragrupos (W) es similar a la anterior, con la diferencia de que las desviaciones de cada variable se calculan respecto a la media del grupo al que pertenece el individuo:

$$w_{jh} = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{jgi} - \bar{x}_{jg})(x_{hgi} - \bar{x}_{hg})$$

Cuando no hay diferencias entre los centroides de grupo, todos los elementos de la matriz W son iguales a los elementos de T , en caso contrario, los elementos de W serán más pequeños que los respectivos términos de T , siendo la diferencia entre ambos la denominada matriz de sumas de cuadrados y productos cruzados entre grupos (B):

$$b_{jh} = t_{jh} - w_{jh} \sum_{g=1}^G n_g (\bar{x}_{jg} - \bar{x}_j)(\bar{x}_{hg} - \bar{x}_h)$$

Tomando como base las matrices T y W podemos expresar las sumas de cuadrados de la combinación lineal de las variables discriminantes como:

$$SCE = v' B v \text{ y } SCI = v' W v$$

Donde v es el vector columna formado por los pesos v_1, v_2, \dots, v_k que lleva a la expresión del criterio de discriminación a maximizar:

$$\frac{v' B v}{v' W v} = \lambda$$

Siendo λ un escalar denominado autovalor o valor propio (*eigenvalue*) que suele tomarse como criterio para medir la diferencia entre grupos. De lo que se trata es de determinar el valor de los coeficientes v_1, v_2, \dots, v_k que maximicen el criterio de discriminación de λ . Esto supone calcular la derivada parcial de λ respecto a cada componente del vector v e igualarla a cero:

$$\frac{\delta\lambda}{\delta v} = \frac{\delta(v'Bv/v'Wv)}{\delta v} = \frac{2[Bv - \lambda Wv]}{v'Wv} = 0$$

Que resulta equivalente a $Bv - \lambda Wv = (B - \lambda W)v = 0$

Suponiendo que la matriz W es no singular, es posible calcular su matriz inversa W^{-1} y aplicándola a la expresión anterior tenemos:

$$W^{-1}(B - \lambda W)v = (W^{-1}B - \lambda I)v = 0$$

La resolución de la ecuación da como resultado los autovalores de la matriz $W^{-1}B$ y sus vectores asociados, que se corresponden, respectivamente con las soluciones de λ y v .

2.3.2. Análisis discriminante cuadrático (QDA)

Introducido por Smith (1947), parte del supuesto de que si admitiendo la normalidad de las observaciones, la hipótesis de igualdad de varianzas no fuera admisible, el procedimiento para resolver el problema es clasificar la observación en el grupo con máxima probabilidad. A posteriori será necesario minimizar la función:

$$\min_{j \in \{1, \dots, G\}} \left[\frac{1}{2} \log |V_j| + \frac{1}{2} (x_0 - \mu_j)' V_j^{-1} (x_0 - \mu_j) - \ln(C_{j\pi_j}) \right]$$

Ahora el término $x_0' V_j^{-1} x_0$ no puede anularse, al depender del grupo y las funciones discriminantes no son lineales y tendrán un término de segundo grado.

Como vemos, tanto el análisis discriminante lineal como el análisis discriminante cuadrático asumen que las observaciones siguen una distribución normal multivariante. A diferencia del análisis lineal, el cuadrático permite que los grupos tengan matrices de covarianzas diferentes. Dada la naturaleza de la población que

estudiamos, debemos tener en cuenta una restricción en cuanto a los tamaños de la muestra: si un grupo es relativamente pequeño, la estimación de la matriz de covarianzas para ese grupo puede no estar representando adecuadamente la covarianza de ese grupo y generar imprecisiones en la clasificación.

2.3.3. Análisis discriminante Logístico (LogDA)

El análisis discriminante logístico es un método parcialmente paramétrico, que podemos situar en una posición intermedia entre los métodos paramétricos, tales como el LDA y el QDA, y los métodos no paramétricos, como el análisis discriminante del K-ésimo vecino más próximo (Knn) (Albert y Lesaffre, 1986). En vez de plantear premisas sobre la distribución de los datos dentro de cada grupo, el análisis discriminante logístico se apoya en la premisa de que los ratios de probabilidad de los grupos tienen forma exponencial. Si bien este método no requiere asumir una forma específica funcional del tipo $f_i(x)$ $i \in (1, 2, \dots, G)$, debe cumplir algunos supuestos para su correcta aplicación. Se utiliza en situaciones en las que las variables predictoras no están distribuidas normalmente y algunas o todas las variables son discretas o categóricas. Una ventaja adicional de este modelo es que, cuando las variables son normales, también verifican el modelo.

De acuerdo con la regla de asignación que surge de minimizar la probabilidad total de clasificar mal para una clasificación binaria, una observación se asigna a G_1 , si se cumple que:

$$p_1 f_1(x) \geq p_2 f_2(x)$$

aplicando el modelo logit y suponiendo que las probabilidades a priori son las mismas:

$$P(y = 1/x_i) = \frac{P(x_i/y = 1)P(y = 1)}{P(x_i)} = C \prod_{j=1}^P P_{1j}^{x_{ij}} (1 - P_{1j})^{1-x_{ij}}$$

la transformación logística será:

$$g_i = \log \frac{P(y=1/x_i)}{1-P(y=1/x_i)} = \sum x_{ij} \log \left(\frac{p_{1j}}{p_{2j}} \right) + \sum (1 - x_{ij}) \log \left(\frac{1-p_{1j}}{1-p_{2j}} \right)$$

que es una función lineal en las variables que podemos escribir como

$$g_i = b_0 + b_1'x_i \text{ siendo } b_0 = \sum \log \left(\frac{1-p_{1j}}{1-p_{2j}} \right) \text{ y } b_1' = \sum \log \left[\frac{p_{1j}(1-p_{2j})}{p_{2j}(1-p_{1j})} \right]$$

Dado que la regresión logística multinomial proporciona los fundamentos para el LogDA, este tipo de análisis discriminante es apropiado tanto para variables discriminantes de tipo binario como para las de tipo continuo.

2.3.4. Análisis discriminante del K-ésimo vecino más próximo

Si bien esta metodología fue utilizada inicialmente por Fix y Hodges en 1951, no fue publicada en esa época (Silverman y Jones, 1989) y los tratamientos avanzados son bastante recientes (McLachlan, 2004). Es un método de discriminación no paramétrico, que compara cada observación con un número determinado de las observaciones más próximas. A diferencia de otros métodos de análisis discriminante, éste puede utilizarse en poblaciones no normales y es capaz de distinguir grupos con forma irregular.

El K-ésimo vecino más próximo puede considerarse uno de los algoritmos de aprendizaje máquina más simples: un objeto se clasifica según el voto mayoritario de sus vecinos. Un problema clave de este método es la elección del valor de k , que es un número entero positivo pequeño²⁵. La práctica habitual es tomar $k = \sqrt{n_g}$, donde n_g es un tamaño del grupo promedio. En nuestro caso, hemos probado distintas opciones para k , aunque solo se muestra el modelo que ha obtenido un mejor desempeño, que es el modelo con $k = 2$.

Este método permite ponderar las aportaciones de los vecinos, para que los vecinos más próximos contribuyan más a la media que los más distantes. Una

²⁵ Cuando varios vecinos están a la misma distancia, *Stata* toma un valor de k más amplio que los incluya. Si seleccionamos $k = 2$, y dentro de la distancia de los dos vecinos más próximos, hay cuatro vecinos, los cálculos incluyen la información de los cuatro.

forma de medir estas ponderaciones es por la distancia a los vecinos (generalización de la interpolación lineal):

1. Se define una medida de distancia entre dos puntos, habitualmente la distancia euclídea o la de Mahalanobis²⁶.
2. Se calcula la distancia al punto que se debe clasificar x_0 de todos los puntos de la muestra.
3. Se seleccionan los k puntos muestrales más próximos al que pretendemos clasificar. Se calcula la proporción de estos k puntos que pertenecen a cada una de las poblaciones. Se clasifica el punto x_0 en la población con mayor frecuencia de los puntos entre los k . Para el caso particular de $k = 1$ el método consiste en asignar el objeto al grupo al que pertenece el elemento más próximo.

2.3.5. Logit

La palabra "logit" fue acuñada por Berkson (1944) y considera el problema de la discriminación entre dos poblaciones (P_1, P_2) para una muestra de n elementos del tipo (y_i, x_i) , donde y_i es el valor del elemento de la variable binaria y x_i es un vector de variables explicativas. El modelo de regresión $y = \beta_0 + \beta'_1 x + u$ trata de prever el valor de la variable binaria ficticia en un nuevo elemento, cuando se conocen las variables x . Si llamamos p_i a la probabilidad de que y tome el valor 1 (pertenezca a la población P_2 cuando $x = x_i$; $p_i = P(y = 1|x_i)$, la variable y es binomial puntual y toma los valores uno y cero con probabilidades p_i y $1 - p_i$. Su esperanza será $E[y/x_i] = p_i \times 1 + (1 - p_i) \times 0 = p_i$, concluimos que $p_i = \beta_0 + \beta'_1 x_i$. Este modelo estimado por mínimos cuadrados conduce a una nueva regla de clasificación ya que maximiza la separación entre los grupos, sea cual sea la distribución de datos. Para que la variable respuesta nos proporcione la probabilidad de pertenecer a cada población y que la respuesta sea cero o uno, debemos transformar la variable respuesta en $p_i = F(\beta_0 + \beta'_1 x_i)$. La clase de funciones no decrecientes acotadas entre cero y uno es la clase de funciones de

²⁶ Si bien la distancia de Mahalanobis puede resultar interesante en análisis multivariantes porque también tiene en cuenta las correlaciones, no es aplicable en nuestro caso por la utilización de una variable discreta, la variable incidencias (junto con el resto de variables continuas, que son los ratios contables).

distribución, por lo que el problema se resuelve tomando como F cualquier función de distribución. Entre las posibilidades podemos truncar el modelo de regresión del siguiente modo:

$$\begin{aligned} p_i &= 1 && \text{si } \beta_0 + \beta'_1 x_i \geq 1 \\ p_i &= \beta_0 + \beta'_1 x_i && \text{si } 0 < \beta_0 + \beta'_1 x_i < 1 \\ p_i &= 0 && \text{si } \beta_0 + \beta'_1 x_i \leq 0 \end{aligned}$$

Esta solución no es satisfactoria ya que, en la teoría, un pequeño incremento de x produce en los extremos un salto muy grande, cuando sería más lógica una evolución gradual. En la práctica, la estimación del modelo es difícil debido a la discontinuidad. Otra posibilidad es tomar como F la función de distribución logística dada por:

$$p_i = \frac{1}{1 + e^{-\beta_0 - \beta'_1 x_i}}$$

Esta función tiene la ventaja de la continuidad, y como $1 - p_i = \frac{e^{-\beta_0 - \beta'_1 x_i}}{1 + e^{-\beta_0 - \beta'_1 x_i}} = \frac{1}{1 + e^{\beta_0 + \beta'_1 x_i}}$ resulta que $g_i = \log \frac{p_i}{1 - p_i} = \beta_0 + \beta'_1 x_i$ que es un modelo lineal. La variable logit denominada g_i representa en una escala logarítmica la diferencia entre las probabilidades de pertenecer a ambas poblaciones, y al ser una función lineal de las variables explicativas nos facilita la explicación y la interpretación del modelo.

La estimación del modelo logit se realiza mediante la aplicación del modelo de máxima verosimilitud. Si todas las variables son binarias independientes y llamamos $P_1 = (p_{11}, \dots, p_{1p})$ a los parámetros de la primera población y $P_2 = (p_{21}, \dots, p_{2p})$ a los de la segunda y observamos un elemento $x_1 = (x_{i1}, \dots, x_{ip})$ tendremos, suponiendo que las probabilidades a priori sean las mismas:

$$P(y = 1|x_i) = \frac{P(x_i|y = 1)P(y = 1)}{P(x_i)} = c \prod_{j=1}^p p_{ij}^{x_{ij}} (1 - p_{1j})^{1-x_{ij}}$$

siendo $c = P(y = 1)/P(x_i)$. La transformación logística será:

$$g_i = \log \frac{P(y = 1|x_i)}{1 - P(y = 1|x_i)} = \sum x_{ij} \log(p_{1j}/p_{2j}) + \sum (1 - x_{ij}) \log[(1 - p_{1j})/(1 - p_{2j})]$$

que es una función lineal de las variables, que podemos escribir $g_i = \beta_0 + \beta'_1 x_i$ donde $\beta_0 = \sum \log[(1 - p_{1j})/(1 - p_{2j})]$ y $\beta'_1 = \sum \log[p_{1j}(1 - p_{2j})/p_{2j}(1 - p_{1j})]$

Una ventaja adicional de este modelo es que cuando las variables son normales, también verifican el modelo logístico. *Stata* lleva a cabo la maximización siguiendo el estimador de la varianza Huber/White/sandwich y su versión agrupada. En cuanto a la colinearidad de las variables, por defecto, *Stata* comprueba los datos y omite las variables que presenten este problema.

2.3.6. Probit

El análisis Probit se originó en el campo de los bioensayos. Fue Bliss (1934) quien sugirió el término probit por la contracción de “*probability unit*”. Probit ajusta modelos de máxima verosimilitud con una variable dependiente dicotómica que toma valores 0/1. Siguiendo el modelo de logit, en este caso la función de densidad es:

$$p_i = \int_{-\infty}^{\beta_0 + \beta'_1 x_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv$$

donde v es una variable normal estandarizada, es decir $v \sim N(0,1)$.

Los modelos logit y probit obtienen resultados similares, aunque las estimaciones de los parámetros de los dos modelos no son directamente comparables. La razón es que la varianza de la variable normal estándar (probit) es uno, mientras que la varianza de la distribución logística (logit) es $\pi^2/\sqrt{3}$. Al igual que en logit, *Stata* comprueba los datos que se analizan con probit y, de forma automática, elimina las variables con problemas de colinearidad.

2.3.7. Análisis Envolvente de Datos (DEA)

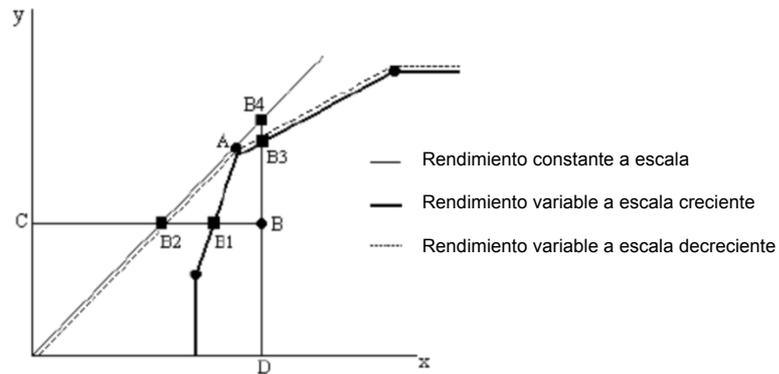
La metodología del análisis envolvente de datos (*Data Envelopment Analysis*, DEA) surge a raíz de la tesis doctoral de Rhodes (1978) y puede considerarse una extensión de los trabajos de Farrell (1957) que proporcionaba una medida satisfactoria de eficiencia productiva. El primer trabajo empírico basado en DEA lo realizaron Charnes, Cooper y Rhodes (1978).

El DEA es una técnica no paramétrica que utiliza la programación matemática permitiendo la construcción de una superficie envolvente, frontera eficiente o función de producción empírica a partir de los datos disponibles del conjunto de elementos objeto de estudio, denominados unidades (DMUs). DEA califica los elementos que determinan la envolvente como eficientes y el resto como ineficientes, proporcionando una evaluación de la eficiencia relativa de cada una de las unidades. La medida de la eficiencia de una unidad mediante DEA implica dos pasos:

1. La construcción de un conjunto de posibilidades de producción.
2. La estimación de la máxima valoración del *output* o de la máxima contracción del *input* dentro del conjunto de posibilidades de producción.

Los modelos DEA pueden ser clasificados en función del tipo de medida de eficiencia que proporcionan en: radiales y no radiales; según la orientación del modelo en: orientados al *input*, orientados al *output* o con orientación *input-output*, o de acuerdo con la tipología de los rendimientos a escala en: constantes o variables (crecientes y decrecientes).

Gráfico 1. Fronteras DEA



Fuente: Coll y Blasco (2006)

En el gráfico 1 puede observarse que la unidad A está en la frontera, por tanto, es eficiente; en cambio, la unidad B está por debajo de la frontera, lo que significa que no es eficiente. Para alcanzar la unidad B la eficiencia, si seguimos una orientación al *input*, debería situarse en la posición B2 con rendimientos constantes a escala (CRS) y rendimientos variables a escala decrecientes (DRS), mientras que debería situarse en B1 para alcanzar la eficiencia con rendimientos variables a escala (VRS). Si la orientación fuera al *output* los puntos para conseguir la eficiencia serían el B4 para CRS y B3 para DRS y VRS.

El modelo DEA-CCR desarrollado por Charnes *et al.* (1978) proporciona medidas eficientes radiales y rendimientos constantes a escala. La eficiencia técnica relativa de cada una de las DMUs se define como el cociente entre la suma ponderada de los *output* ($\sum_{r=1}^s u_r y_{r0}$) y la suma ponderada de los *input* ($\sum_{i=1}^m v_i x_{i0}$). La programación del modelo DEA-CCR orientado al *input* se define como:

$$\text{Max}_{u,v} h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}}$$

$$\text{Sujeto a: } \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \leq 1 \quad j = 1, 2, \dots, n$$

$$u_r, v_i \geq 0$$

donde se consideran n DMUs ($j = 1, 2, \dots, n$) cada una de las cuales utiliza los mismos *input* para obtener los mismos *output*, pero en diferentes cantidades. x_{ij} representa las cantidades de *input* ($i = 1, 2, \dots, m$) consumidos por la j -ésima DMU. y_{rj} representa las cantidades observadas de *output* ($j = 1, 2, \dots, s$) producidos por la j -ésima DMU. u_r, v_i representan los pesos o multiplicadores de los *outputs* e *inputs*, respectivamente. Si la solución óptima es $h_0^* = 1$, indica que la DMU es eficiente en relación con las otras DMUs; en caso contrario, será ineficiente.

En 1979 Charnes, Cooper y Rhodes sustituyen la condición de no negatividad ($u_r, v_i \geq 0$) por una condición de positividad estricta ($u_r, v_i \geq \varepsilon$), donde ε es un infinitésimo no-arquimediano²⁷. En la mayoría de las aplicaciones de DEA, el modelo empleado en la evaluación de la eficiencia es la forma envolvente

$$\text{Min}_{\theta, \lambda} z_0 = \theta$$

$$\text{Sujeto a: } Y\lambda \geq y_0$$

$$\theta x_0 \geq X\lambda$$

$$\lambda \geq 0$$

donde λ es el vector ($n \times 1$) de pesos o intensidades, siendo λ_j la intensidad de la unidad j . Y θ denota la puntuación de eficiencia de la DMU_0

El modelo DEA-BCC de Banker, Charnes y Cooper (1984) es una extensión del modelo DEA-CCR diferenciándose en que DEA-BCC introduce el supuesto de rendimientos variables a escala, crecientes o decrecientes. La forma fraccional de este modelo con orientación al *input* viene expresada por:

$$\text{Max}_{(u,v,k)} h_0 = \frac{u^T y_0 + k_0}{v^T x_0}$$

$$\text{Sujeto a: } \frac{u^T Y_j + k_0}{v^T X_j} \leq 1 \quad j = 1, 2, \dots, n$$

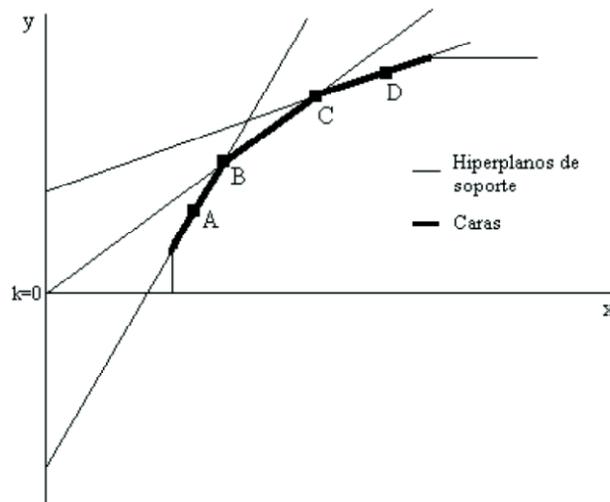
²⁷ En la práctica, un número muy pequeño.

$$u_r, v_i \geq I\varepsilon$$

k_0 no restringida

El término k es el valor del intercepto en el eje *output* de la proyección de cada segmento que define la frontera (Norman y Stoker, 1991).

Gráfico 2. Segmentos de frontera



Fuente: Coll y Blasco (2006)

La forma envolvente de este modelo, expresado como un modelo único a resolver en un proceso de dos etapas, será:

$$\text{Max}_{\theta, \lambda, s^+, s^-} z_0 = \theta - \varepsilon (Is^+ + Is^-)$$

$$\text{Sujeto a: } Y\lambda \geq y_0 + s^+$$

$$X\lambda = \theta x_0 - s^-$$

$$\vec{1}\lambda = 1$$

$$\lambda, s^+, s^- \geq 0$$

donde $\vec{1}\lambda = 1$ es la restricción de convexidad, Is^+ es el vector de holguras *output* y Is^- es el vector de holguras *input*.

A diferencia del resto de metodologías que aplicamos en este trabajo, el DEA aplicado a un número elevado de empresas requiere recursos informáticos potentes con gran capacidad de memoria y lleva bastante tiempo. Para que el lector se haga una idea, daremos un ejemplo. Mientras el resto de metodologías, aplicadas a una muestra de 3.193 empresas no fracasadas y 37 fracasadas, tardaba solo unos segundos en llegar a los resultados en un ordenador normal, para cualquiera de los números de variables empleados (entre 3 y 10), DEA tuvo que ser aplicado en un centro de supercomputación y cada proceso de cálculo ocupó desde varios días (con el menor número de variables) hasta algo más de un mes (con 10 variables). Es comprensible que DEA se aplique a cantidades pequeñas de DMUs; de hecho, su capacidad para ser aplicado a muestras pequeñas es una de sus ventajas reconocidas.

Hemos usado la metodología DEA en varias versiones: rendimientos constantes a escala (CRS), rendimientos variables a escala (VRS) y rendimientos no crecientes a escala (DRS), cada uno de ellos con orientación tanto al *input* como al *output*. Mostramos solo los resultados obtenidos con CRS, método con el que hemos conseguido los mejores resultados. Que CRS ofrezca los mejores resultados es consistente con *inputs* y *outputs* a escala lineal sin que eso signifique aumentos o reducciones de la salud financiera de la empresa²⁸. Como consecuencia, esperamos resultados similares si aplicamos medidas orientadas al *input* o medidas orientadas al *output*²⁹.

²⁸ En cambio, DRS y VRS implican un efecto distorsión del tamaño sobre los rendimientos a escala. DRS obtiene variaciones de los *outputs* menos que proporcionales ante variaciones de los *inputs*. VRS permite incorporar rendimientos a escala crecientes o decrecientes.

²⁹ En la literatura sobre la metodología DEA es bien conocido que cuando la tecnología presenta globalmente rendimientos a escala constantes, las medidas de eficiencia técnica radiales orientadas al *input* y al *output* son idénticas (Banker *et al.*, 1984).

2.4. Los datos y la selección estadística de variables

Todos los datos se han obtenido de la base de datos de *Iberinform*. Esta base de datos recoge la información contable oficial aportada por las empresas al Registro Mercantil; la información sobre la situación de fracaso (en situación concursal, disuelta o extinguida) que las empresas publican en el Boletín Oficial del Registro Mercantil (BORM); y los datos sobre incidencias de pago, recopilados de diferentes fuentes: Boletín Oficial del Estado, Boletines Oficiales de las Provincias, Boletines Oficiales de las Comunidades Autónomas, sitio web de la Seguridad Social y prensa (periódicos).

El análisis empírico se aplica a empresas españolas de la construcción, clasificadas como pequeñas según los criterios contables oficiales. Para la formación de la muestra se establecieron las siguientes restricciones:

- CNAE: Dado que las cuentas anuales utilizan una codificación establecida en el Real Decreto 475/2007, de 13 de abril, por el que se aprueba la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009), se eligió la sección F- Construcción con epígrafes superiores o iguales a 4100 e inferiores a 4399.
- Modo Balance: Se seleccionaron las empresas que, en el modo de balance, presentaban “PymeNPGC”, que son aquellas que cumplen los requisitos para presentar el modelo pyme correspondiente a la Orden JUS/206/2009, de 28 de enero, por el que se aprueban nuevos modelos para la presentación en el Registro Mercantil de las cuentas anuales de los sujetos obligados a su publicación. Las condiciones que establece la Orden ministerial citada para utilizar el modelo pyme en el balance, la cuenta de pérdidas y ganancias, el estado de cambios en el patrimonio neto y la memoria, son que durante dos ejercicios consecutivos se deben reunir, a la fecha de cierre de cada uno de ellos, al menos dos de las circunstancias siguientes:
 - Total activo: menos de 2.850.000 €.

- Importe neto de la cifra de negocio: menos de 5.700.000 €.
- Número medio de trabajadores: menos de 50.

En ningún caso pueden utilizar este modelo de presentación las empresas que hayan emitido valores admitidos a negociación en mercados regulados o sistemas multilaterales de negociación, de cualquier Estado miembro de la Unión Europea, o que formen parte de un grupo de sociedades que formulen o debieran haber formulado cuentas anuales consolidadas o que su moneda funcional sea distinta del euro.

- Resultado: Se han seleccionado aquellas empresas que tienen el balance “cuadrado”.
- Número de Empleados Fijos: Se han seleccionado aquellas empresas que declaren un número de empleados menor o igual a 50.
- Número de Empleados Temporales: Se han seleccionado aquellas empresas que declaren un número de empleados menor o igual a 50.

Con estas premisas, quedaron 3.426 empresas del ámbito geográfico de la Comunidad Autónoma de Castilla y León. Al calcular los ratios que constituyen las variables explicativas del trabajo, fueron eliminadas aquellas empresas con valores indefinidos. La muestra final está formada por las 3.230 empresas con valores reales para todos los ratios, de las cuales 3.193 son empresas no fallidas y 37 son empresas fallidas en 2008. La tasa de insolvencia es aproximadamente del 1,15%.

En primer lugar, presentamos las estadísticas y los resultados de los modelos usando todos los datos contables y, a continuación, presentamos la misma información sobre estadísticas y resultados de los modelos, una vez que los ratios contables han sido winsorizados³⁰.

³⁰ La winsorización consiste en la sustitución de los valores extremos por el valor del percentil que se toma como extremo deseable. En nuestro caso se winsoriza al 1%, o sea, que los valores por debajo

La selección inicial de variables realizada a partir de la literatura previa está formada por los siguientes ratios:

- Deuda Total / Activo Total
- Activo Circulante / Pasivo Circulante
- BAIT / Activo Total
- Beneficio Neto / Activo Total
- Activo Circulante / Activo Total
- Beneficios No Distribuidos / Activo Total
- Gastos Financieros / Pasivo Exigible
- Recursos Generados / Pasivo Exigible
- Beneficio Neto / Ventas
- Ventas / Activo Total

Los primeros ocho ratios son las variables individuales más frecuentes encontradas significativas en la literatura previa analizada en la primera parte de la tesis (Tabla 2, apartado 1.2). Los ratios noveno y décimo han sido seleccionados como los representantes más frecuentes en nuestro estudio para los rasgos no representados en las ocho variables anteriores, margen sobre beneficios y rotación (Tabla 3, apartado 1.2). Dado que estos diez ratios financieros han resultado ser los ratios significativos con mayor frecuencia en la literatura previa sobre fracaso empresarial, los tomamos como buenos representantes de la fortaleza o debilidad económica y financiera de las empresas. A continuación, se explica el comportamiento esperado de cada uno de estos ratios en los modelos de fracaso empresarial. También se explica la construcción y razonamiento de la variable cualitativa incidencias.

TD / TA, Deuda Total / Activo Total, es una medida de endeudamiento que indica una obligación financiera a largo plazo. Un aumento en el apalancamiento

del percentil 1 se sustituyen por el valor del punto situado en el percentil; y los valores por encima del percentil 99 se sustituyen por el valor del punto situado en el percentil.

aumentaría la probabilidad de dificultades financieras, ya que una reducción en los flujos de efectivo podría significar no disponer de fondos suficientes para pagar las deudas, lo que daría como resultado la quiebra. En general, se puede esperar una relación positiva entre endeudamiento y fracaso empresarial.

CA / CL, Activo Circulante / Pasivo Circulante, es una medida de equilibrio económico-financiero a corto plazo. Dado que las deudas comerciales son una parte importante en los pasivos de las pequeñas empresas (Altman *et al.* 2008), y el crédito a clientes se extiende en períodos de crisis financiera, el fracaso de las pequeñas empresas debería estar fuertemente influenciado por este tipo de deudas.

BAIT / TA, Beneficios Antes de Intereses e Impuestos / Activo Total, es una medida de la rentabilidad económica de las empresas.

NI / TA, Beneficio Neto / Activo Total, es una medida más general de la rentabilidad económica.

CA / TA, Activo Circulante / Activo Total, es una medida de la estructura económica de la empresa, que nos dice la proporción de activos a corto plazo dentro del total.

FE / TD, Gastos Financieros / Deuda Total, es una medida del coste financiero indicativo del endeudamiento y del riesgo. Dado que las empresas en dificultades tienden a acumular más deuda y que la deuda adicional suele resultar más cara, se puede esperar una relación positiva con el fracaso empresarial.

RP / TA, Beneficios No Distribuidos / Activo Total, es una medida de la rentabilidad acumulada y puede ser una medida de la edad de la empresa. Indica la capacidad de la empresa para protegerse contra los posibles riesgos futuros, por lo tanto, se puede esperar una correlación negativa con el fracaso de empresas.

CF / TD, Recursos Generados (*Cash Flow*) / Deuda Total, es una medida de la liquidez que la empresa es capaz de crear. Esperamos una relación negativa con el fracaso empresarial.

NI / SL, Beneficio Neto / Ventas. El margen es el principal inductor de la rentabilidad. Cuanta mayor proporción de ventas se convierta en beneficio, más

eficiente estará siendo la empresa. Esto sugiere una relación negativa con el fracaso empresarial.

SL / TA, Ventas / Activo Total. La rotación actúa como un multiplicador del margen para aumentar la rentabilidad cuando es positiva. Por tanto, esperamos una relación negativa con el fracaso.

Incidencias, indica la ocurrencia de determinados incumplimientos en los créditos de la empresa, ya sea en contratos de crédito con entidades financieras, en pagos de operaciones comerciales o en deudas con algún organismo de la Administración Pública. Distinguimos dos grupos de incidencias: las judiciales y las reclamaciones de los organismos de la Administración Pública. En las incidencias judiciales se incluyen todas las reclamaciones presentadas contra la empresa ante un tribunal, bien sea por una persona física o bien por una persona jurídica. Las reclamaciones de los organismos de la Administración Pública incluyen deudas con la Seguridad Social, la Hacienda Pública y las Administraciones Locales, por conceptos tales como cuotas sociales, multas, impuestos, etc. impagados a su vencimiento. Teniendo en cuenta su contenido, construimos la variable incidencias del siguiente modo: toma el valor 0 si la empresa no tiene pendientes reclamaciones (ni judiciales ni con los organismos de las Administraciones Públicas); toma el valor 1 si tiene uno de los dos tipos de reclamaciones; y toma el valor 2 si tiene reclamaciones de los dos tipos (judiciales y de los organismos de la Administración Pública).

Como idea general, podemos partir de la hipótesis de que las empresas con indicadores de liquidez bajos, como los flujos de efectivo (*cash flow*), con indicadores de rentabilidad reducidos, como los resultados netos o los resultados de explotación, con indicadores de desempeño pobres, como el margen o la rotación, y, en cambio, con valores elevados en los indicadores de endeudamiento y riesgo, tales como los gastos financieros, se espera que sean más vulnerables a la insolvencia y, por tanto, con más probabilidades de fracasar.

A continuación se detalla cómo contribuyen estas variables en la evaluación del fracaso empresarial. La Tabla 7 muestra la matriz de correlación entre los ratios

financieros y dos variables no contables específicas por empresa: las incidencias de pago y el número de trabajadores, además de la variable fracaso. Los coeficientes de correlación son relativamente bajos, excepto entre los ratios de rentabilidad (R3, R4 y R7), y entre algunos ratios que tienen como divisor el total de activos.

Tabla 7. Panel A. Matriz de correlaciones

	Failure	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	incident	N.emp
Failure	1.0000												
R1= TD/TA	-0.0011	1.0000											
R2= CA/CL	-0.0029	-0.0023	1.0000										
R3= BAIT/TA	-0.0108	-0.9057*	0.0031	1.0000									
R4= NI/TA	-0.0052	-0.9773*	0.0023	0.9692*	1.0000								
R5= CA/TA	0.0147	-0.0024	0.0260	-0.0049	0.0354	1.0000							
R6= FE/TD	0.1593*	-0.0006	0.0375	0.0024	0.0018	-0.0126	1.0000						
R7= RP/TA	-0.0021	0.9995*	-0.0015	-0.8940*	-0.9714*	-0.0011	-0.0006	1.0000					
R8= CF/TD	0.0471*	-0.0036	0.4113*	0.0343	0.0189	0.0096	0.5289*	-0.0018	1.0000				
R9= NI/SL	0.0012	0.0002	0.0000	0.0084	0.0044	-0.0062	0.0027	0.0001	0.0240	1.0000			
R10= SL/TA	0.0038	0.8138*	-0.0032	-0.6477*	-0.7883*	-0.3350*	-0.0144	0.8173*	-0.0037	0.0015	1.0000		
incidents	0.1794*	0.1057*	-0.0081	-0.0914*	-0.0978*	-0.0199	0.0041	0.1051*	-0.0227	0.0014	0.0808*	1.0000	
N.emp	-0.0031	-0.0035	-0.0018	0.0051	0.0033	-0.0134	0.0054	-0.0014	0.0003	0.0015	-0.0002	0.0024	1.0000

NOTAS: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.
* Nivel de significación al 1%.

Panel B. Matriz de correlaciones con ratios winsorizados

	Failure	wR1	wR2	wR3	wR4	wR5	wR6	wR7	wR8	wR9	wR10	incident	N.emp
Failure	1.0000												
wR1= TD/TA	0.0546*	1.0000											
wR2= CA/CL	-0.0047	-0.2632*	1.0000										
wR3= BAIT/TA	-0.0906*	-0.5407*	0.0550*	1.0000									
wR4= NI/TA	-0.0999*	-0.5879*	0.0656*	0.9719*	1.0000								
wR5= CA/TA	0.0215	-0.1396*	0.1469*	0.1595*	0.1724*	1.0000							
wR6= FE/TD	0.0193	-0.0274	0.1200*	0.0830*	0.0011	-0.0811*	1.0000						
wR7= RP/TA	-0.0127	0.3276*	-0.0047	-0.1929*	-0.2100*	-0.0788*	-0.0615*	1.0000					
wR8= CF/TD	-0.0325	-0.4451*	0.2458*	0.5925*	0.6009*	0.0401	-0.0240	0.0182	1.0000				
wR9= NI/SL	-0.0384	-0.2497*	-0.1024*	0.3873*	0.3974*	0.0778*	-0.0056	-0.0880*	0.3025*	1.0000			
wR10= SL/TA	0.0189	0.1653*	-0.1781*	-0.0521*	-0.0957*	0.0208	0.1062*	0.1581*	0.0433	0.1265*	1.0000		
incidents	0.1794*	0.1497*	-0.0346	-0.1304*	-0.1421*	-0.0025	0.0542*	0.0081	-0.1134*	-0.0679*	-0.0026	1.0000	
N.emp	-0.0031	-0.0088	-0.0021	0.0181	0.0119	-0.0212	0.0696*	0.0003	0.0081	0.0097	-0.0109	0.0024	1.0000

NOTAS: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.
* Nivel de significación al 1%.

Después de winsorizar las variables, todos los coeficientes de correlación son todavía más bajos, excepto entre los ratios de rentabilidad R3 y R4 (Tabla 7, panel B). En la Tabla 8, se puede apreciar que usando los datos completos de contabilidad los valores de la media son muy diferentes de los valores de la mediana y tanto la desviación típica como la asimetría son muy altos. Una vez que los datos han sido winsorizados al nivel del 1% (Tabla 8, Panel B), los valores medios están más cerca de los valores medianos y tanto las desviaciones típicas como la asimetría han disminuido considerablemente. Las variables obtienen los

signos de correlación esperados. En el Anexo VI se incluyen los histogramas de frecuencias de las diez variables analizadas, lo que además permite ver la diferente distribución de las variables ratio, y lo alejadas que se encuentran de una distribución normal en la mayoría de los casos.

Tabla 8. Panel A. Estadísticos descriptivos

Failure	variable	n	Mean	median	St.dev.	skewness
not-failed	R1= TD/TA	3193	1.464248	0.7783218	35.13914	56.26852
not-failed	R2= CA/CL	3193	4.992888	1.274884	65.81954	30.69193
not-failed	R3= BAIT/TA	3193	-0.0345993	0.0444596	2.19314	-41.647
not-failed	R4= NI/TA	3193	-0.0876411	0.0159537	3.849023	-51.79087
not-failed	R5= CA/TA	3193	0.7268043	0.7864155	0.3425963	9.195928
not-failed	R6= FE/TD	3193	0.0319486	0.0214282	0.0583653	13.70667
not-failed	R7= RP/TA	3193	0.810277	0.1062388	32.54369	56.31467
not-failed	R8= CF/TD	3193	0.1890471	0.0608538	2.319081	23.02114
not-failed	R9= NI/SL	3193	-0.5259805	0.0124977	26.23841	-48.29314
not-failed	R10= SL/TA	3193	1.815489	1.381083	22.46602	18.59462
not-failed	incidents	3193	0.0720326	0	0.3133578	4.695205
not-failed	N.emp	1955	32.46496	6	588.1141	36.44888
Failed	R1= TD/TA	37	1.09844	0.9751387	1.010565	3.7524
Failed	R2= CA/CL	37	3.23058	1.057278	11.17866	5.727102
failed	R3= BAIT/TA	37	-0.2554667	-0.0164002	0.7124744	-2.656331
failed	R4= NI/TA	37	-0.2745696	-0.0354618	0.6439843	-2.230388
failed	R5= CA/TA	37	0.774168	0.8938961	0.2734852	-1.460193
failed	R6= FE/TD	37	0.4296611	0.032181	2.422027	5.832577
failed	R7= RP/TA	37	0.1619858	0.0299024	0.6085422	3.317699
failed	R8= CF/TD	37	1.418611	-0.0286602	14.66338	4.637255
failed	R9= NI/SL	37	-0.2360054	-0.0441338	0.5619837	-3.127511
failed	R10= SL/TA	37	2.612674	1.475444	5.620027	5.245937
failed	incidents	37	0.6216216	0	0.720777	0.694475
failed	N.emp	16	12.1875	9	11.55404	1.095393

NOTAS: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.

Panel B. Estadísticos descriptivos, ratios winsorizados

failure	variable	n	Mean	median	St.dev.	skewness
not-failed	wR1= TD/TA	3193	0.7876563	0.7783218	0.4490908	2.483903
not-failed	wR2= CA/CL	3193	2.783909	1.274884	5.769318	5.38778
not-failed	wR3= BAIT/TA	3193	0.022242	0.0444596	0.220357	-2.688866
not-failed	wR4= NI/TA	3193	-0.0056056	0.0159537	0.1911801	-3.040909
not-failed	wR5= CA/TA	3193	0.7263177	0.7864155	0.2373353	-0.9497068
not-failed	wR6= FE/TD	3193	0.0297533	0.0214282	0.032074	1.978609
not-failed	wR7= RP/TA	3193	0.1956719	0.1062388	0.3814027	2.493171
not-failed	wR8= CF/TD	3193	0.1211922	0.0608538	0.3389109	2.552918
not-failed	wR9= NI/SL	3193	-0.0638995	0.0124977	0.475318	-6.155788
not-failed	wR10= SL/TA	3193	1.661431	1.381083	1.448189	2.268183
not-failed	incidents	3193	0.0720326	0	0.3133578	4.695205
not-failed	N.emp	1955	32.46496	6	588.1141	36.44888
failed	wR1= TD/TA	37	1.019683	0.9751387	0.6442166	1.777834
failed	wR2= CA/CL	37	2.527121	1.057278	6.940696	5.563727
failed	wR3= BAIT/TA	37	-0.1692527	-0.0164002	0.4425078	-1.603176
failed	wR4= NI/TA	37	-0.1896237	-0.0354618	0.4113469	-1.710292
failed	wR5= CA/TA	37	0.7744578	0.8938961	0.2726751	-1.451836
failed	wR6= FE/TD	37	0.0355791	0.032181	0.0330995	2.367284
failed	wR7= RP/TA	37	0.1500051	0.0299024	0.4884795	2.808293
failed	wR8= CF/TD	37	0.0164868	-0.0286602	0.5752716	2.43246
failed	wR9= NI/SL	37	-0.2360054	-0.0441338	0.5619837	-3.127511
failed	wR10= SL/TA	37	1.919005	1.475444	1.894715	1.989572
failed	incidents	37	0.6216216	0	0.720777	0.694475
failed	N.emp	16	12.1875	9	11.55404	1.095393

NOTAS: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.

Un método bien conocido (Peel *et al.*, 1986; Grice e Ingram, 2001; Sueyoshi y Goto, 2009a; 2009b y 2009c; Jin *et al.*, 2011) para seleccionar variables discriminantes es el análisis de diferencias en medias (análisis univariante simple) entre las empresas fracasadas y las no fracasadas.

En la tabla 9 presentamos las medias y las desviaciones típicas de las variables para las empresas fracasadas y las no fracasadas, por separado. En la última columna se muestran las diferencias en medias entre ambos grupos junto con el estadístico t, estableciendo la hipótesis nula de que las diferencias de medias son iguales a cero. Al analizar los resultados del contraste, podemos ver que R6=FE/TA (Gastos Financieros/ Deuda Total) es 42,96% en el grupo de las empresas fracasadas y sólo 3,19% en las no fracasadas, R8=CF/TD (Recursos Generados/ Deuda Total) alcanza un 141,86% en las fracasadas. Este resultado

espurio proviene de una sola empresa que tiene un valor de 8.328%. Una vez eliminada esta empresa el ratio toma un valor de -0,8553 siendo significativa la diferencia de medias. Las empresas no fracasadas toman un valor medio de 18,9%. La variable incidencias parece ser un factor discriminante relevante, con valores medios de 0,62 para la empresas fracasadas pero solo de 0,07 para las no fracasadas. Por tanto, para estas tres variables se rechaza la hipótesis nula y las diferencias en medias se consideran significativas. El grupo de las empresas no fracasadas supera al grupo de las fracasadas en estos factores con un nivel de significación del 1%, para un valor del estadístico $t=\pm 2,5773$.

Teniendo en cuenta los resultados obtenidos, las empresas fracasadas generan menos recursos (*cash flow*), tienen mayores gastos financieros y sufren más incidencias de pago que las empresas no fallidas. El análisis de diferencias en medias produce los resultados esperados considerando el análisis de correlación de las variables explicativas con la variable fracaso.

Tabla 9. Estadísticos descriptivos, fracasadas/no fracasadas, diferencias en medias

Variable	Failed firms n = 37 Mean (Std. dev.)	Not-failed firms n = 3193 Mean (Std. dev.)	Difference in means (t-Statistic)
R1= TD/TA	1.09844 (1.010565)	1.464248 (35.13914)	0.3658081 (0.0633)
R2= CA/CL	3.23058 (11.17866)	4.992888 (65.81954)	1.762308 (0.1628)
R3= BAIT/TA	-0.2554667 (0.7124744)	-0.0345993 (2.19314)	0.2208674 (0.6121)
R4= NI/TA	-0.2745696 (0.6439843)	-0.0876411 (3.849023)	0.1869285 (0.2953)
R5= CA/TA	0.774168 (0.2734852)	0.7268043 (0.3425963)	-0.0473637 (-0.8378)
R6= FE/TD	0.4296611 (2.422027)	0.0319486 (0.0583653)	-0.3977125 (-9.1707)***
R7= RP/TA	0.1619858 (0.6085422)	0.810277 (32.54369)	0.6482912 (0.1212)
R8= CF/TD	1.418611 (14.66338)	0.1890471 (2.319081)	-1.229564 (-2.6770)***
R9= NI/SL	-0.2360054 (0.5619837)	-0.5259805 (26.23841)	-0.2899751 (-0.0672)
R10= SL/TA	2.612674 (5.620027)	1.815489 (22.46602)	-0.7971844 (-0.2157)
incidents	0.6216216 (0.720777)	0.0720326 (0.3133578)	-0.5495891 (-10.3621)***
Num.empl	12.1875 (11.55404)	32.46496 (588.1141)	20.27746 (0.1379)

NOTAS: La muestra contiene 3.230 observaciones durante 2008. Definición de las variables: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.
*** nivel de significación del 1% basado en el test de dos colas.

Pero este tipo de análisis (diferencias en medias) requiere que las poblaciones que se comparan sigan una distribución normal. Además, una buena parte de las variables contables presentan distribuciones con una elevada dispersión (ver

distribuciones de frecuencias en el Anexo VI), por lo que los valores de la mediana resultan ser más representativos que los valores de la media. Por lo tanto, realizamos una prueba no paramétrica para la comparación de los dos grupos: el *Rank Sum Test* de Wilcoxon (también llamado Mann-Whitney-Wilcoxon). Este contraste ha sido utilizado para seleccionar variables en procesos de evaluación de fracaso empresarial en trabajos recientes (Premachandra *et al.*, 2009; Sueyoshi y Goto, 2009a; 2009b; 2009c).

Las principales ventajas de este método son: que no se asume que la distribución de las poblaciones que se comparan sea normal; usa únicamente rangos; y no es sensible a los valores extremos (*outliers*). A cambio, hay que considerar algunas desventajas: los métodos no paramétricos son, a menudo, menos sensibles (potentes) para encontrar las verdaderas diferencias ya que no utilizan toda la información de la muestra, sino solo los rangos; necesitan el conjunto completo de datos, no solo el resumen de las estadísticas; y los resultados no incluyen intervalos de confianza que cuantifiquen los rangos de posibilidad de encontrar la verdadera diferencia entre las poblaciones.

Como se muestra en la Tabla 10, nuestras variables están sesgadas en ambas muestras (fracasadas y no fracasadas). Los resultados del *Rank Sum Test* de Wilcoxon indican que los valores de la mediana para las empresas fracasadas y no fracasadas son significativamente diferentes para las siguientes variables: R1, R3, R4, R8, R9 e incidencias, para un nivel de significación del 1%, R5 y R7 para un nivel del 5%; y R2 para un nivel de significación del 10%.

Estos resultados sugieren que, excepto para R10 y R6, los ratios contables usados en el análisis son apropiados para la clasificación de empresas en fracasadas y saludables. Usando diferencias en medianas, se puede apreciar que los valores de R8 no se ven afectados por los valores extremos (Tabla 9), a diferencia de lo que ocurre cuando se usan diferencias en medias (Tabla 8). En relación con el número de empleados, como medida del tamaño de la empresa, ninguno de los contrastes de diferencias encuentran este factor significativo para

distinguir entre empresas fracasadas y no fracasadas. Por lo tanto, no lo hemos utilizado en los análisis multivariantes posteriores.

Aun cuando una variable concreta sea o no significativa para un determinado contraste de diferencias, no conocemos la contribución exacta de esa variable en análisis no paramétricos. Por eso, vamos a mantener todos los ratios seleccionados de acuerdo con la literatura previa (discriminantes con más frecuencia) y justificados por el razonamiento económico como inductores relevantes de la solvencia de la empresa. Estamos de acuerdo con Premachandra *et al.* (2009) en que el propósito de los análisis de fracaso empresarial es investigar si, en general, las variables seleccionadas son capaces de diferenciar entre dos grupos, siendo la diferencia individual entre los dos grupos menos importante.

Tabla 10. Estadísticos descriptivos, fracasadas/no fracasadas, Rank sum test

Variable	Failed firms n = 37 Median (Std. dev.) [Skewness]	Not-failed firms n = 3193 Median (Std. dev.) [Skewness]	Two sample Wilcoxon rank-sum test between medians (Z) [p-value> (Z)]
R1= TD/TA	0.9751387 (1.010565) [3.7524]	0.7783218 (35.13914) [56.26852]	-3.551 [0.0004]***
R2= CA/CL	1.057278 (11.17866) [5.727102]	1.274884 (65.81954) [30.69193]	1.715 [0.0864]*
R3= BAIT/TA	-0.0164002 (0.7124744) [-2.656331]	0.0444596 (2.19314) [-41.647]	3.093 [0.0020]***
R4= NI/TA	-0.0354618 (0.6439843) [-2.230388]	0.0159537 (3.849023) [-51.79087]	4.112 [0.0000]***
R5= CA/TA	0.8938961 (0.2734852) [-1.460193]	0.7864155 (0.3425963) [9.195928]	-1.977 [0.0480]**
R6= FE/TD	0.322181 (2.422027) [5.832577]	0.0214282 (0.0583653) [13.70667]	-1.641 [0.1007]
R7= RP/TA	0.0299024 (0.6085422) [3.317699]	0.1062388 (32.54369) [56.31467]	2.272 [0.0231]**
R8= CF/TD	-0.0286602 (14.66338) [4.637255]	0.0608538 (2.319081) [23.02114]	4.476 [0.0000]***
R9= NI/SL	-0.0441338 (0.5619837) [-3.127511]	0.0124977 (26.23841) [-48.29314]	4.302 [0.0000]***
R10= SL/TA	1.475444 (5.620027) [5.245937]	1.381083 (22.46602) [18.59462]	-0.436 [0.6625]
incidents	0 (0.720777) [0.694475]	0 (0.3133578) [4.695205]	-10.838 [0.0000]***
Num.empl	9 (11.55404) [1.095393]	6 (588.1141) [36.44888]	-0.716 [0.4737]

NOTAS: La muestra contiene 3.230 observaciones durante 2008. Definición de las variables: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.

* nivel de significación del 10% basado en el test de dos colas.

** nivel de significación del 5% basado en el test de dos colas.

*** nivel de significación del 1% basado en el test de dos colas.

2.5. Evaluación del fracaso empresarial

En este apartado aplicamos los cuatro métodos de análisis discriminante: el lineal (LDA), el cuadrático (QDA), el logístico (LogDEA) y el del K-ésimo vecino más próximo (KnnDA), así como dos modelos de regresión binaria, logit y probit.

En cada modelo se utiliza el grupo total de variables (diez ratios) y posteriormente los grupos reducidos de variables que resultan significativas en la selección de medias y en la selección de medianas a los niveles de significación del 5% y del 1%. Todos los modelos se han recalculado incluyendo la variable cualitativa incidencias, que resulta significativa al 1% en todos los casos. No podemos mostrar datos del modelo DEA para datos no winsorizados porque no hemos obtenido resultados en todas los modelos (y particularmente en el modelo CSR) solo para los grupos de variables más grandes (10 y 7 ratios).

Tabla 11. Análisis Discriminante Lineal (LDA)

LDA	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios +inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	3,158 (98.90)	3,003 (94.05)	2,636 (82.56)	3,004 (94.08)	3,102 (97.15)	3,005 (94.11)	3,171 (99.31)	3,007 (94.17)
Failed n= 37	2 (5.41)	20 (54.05)	19 (51.35)	19 (51.35)	6 (16.22)	19 (51.35)	2 (5.41)	20 (54.05)
Total well classified n= 3,230	3,160 (98.44)	3,023 (94.17)	2,655 (82.71)	3,023 (94.17)	3,108 (96.82)	3,024 (94.21)	3,173 (98.85)	3,027 (94.30)

NOTAS: Esta tabla contiene los resultados obtenidos con el modelo LDA tomando distribuciones iguales. Diez ratios incluye: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: R1, R3, R4, R8, R9; al nivel del 5% se incluyen: R1, R3, R4, R5, R7, R8, R9. La selección de medias a nivel del 1% (igual que al nivel del 5%) incluye: R6, R8.

Con el análisis discriminante lineal se obtiene el mejor porcentaje de empresas bien clasificadas con la selección de medias. La inclusión de la variable incidencias mejora considerablemente la selección de empresas fracasadas, mientras que reduce un elevado número de empresas no fracasadas bien clasificadas, resultando un porcentaje global peor (excepto con 7 ratios).

Tabla 12. Análisis Discriminante Cuadrático (QDA)

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios +inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	378 (11.84)	390 (12.21)	134 (4.20)	142 (4.45)	98 (3.07)	103 (3.23)	3,161 (99.00)	3,111 (97.43)
Failed n= 37	36 (97.30)	37 (100.00)	35 (94.59)	36 (97.30)	35 (94.59)	35 (94.59)	2 (5.41)	7 (18.92)
Total well classified n= 3,230	414 (12.82)	427 (13.22)	169 (5.23)	178 (5.51)	133 (4.12)	138 (4.27)	3,163 (97.93)	3,118 (96.53)

NOTAS: Esta tabla contiene los resultados obtenidos con el modelo QDA tomando distribuciones iguales. Diez ratios incluye: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TA es Gastos Financieros/ Deuda Total; CF/TA es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: R1, R3, R4, R8, R9; al nivel del 5% se incluyen: R1, R3, R4, R5, R7, R8, R9. La selección de medias a nivel del 1% (igual que al nivel del 5%) incluye: R6, R8.

El análisis discriminante cuadrático muestra una gran capacidad para identificar a las empresas fracasadas con diez variables, que mejora ligeramente cuando se añade la variable incidencias. Esta capacidad se mantiene con menos variables cuando se utiliza la selección de medianas (7 y 5 variables), pero no cuando las variables son seleccionadas con las medias (2 variables). Por el contrario, las empresas no fracasadas no se identifican bien con 10, 7 o 5 ratios, sin la variable incidencias, y sólo un poco mejor con ella. El poder discriminante de las 2 variables seleccionadas con medias es excelente para las empresas no fracasadas y muy pobre para las empresas fracasadas.

Tabla 13. Análisis Discriminante Logístico (LogDA)

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios +inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	1,896 (59.38)	2,921 (91.48)	1,691 (52.96)	2,998 (93.89)	2,952 (92.45)	3,006 (94.14)	2,839 (88.91)	2,995 (93.80)
Failed n= 37	26 (70.27)	23 (62.16)	28 (75.68)	19 (51.35)	6 (16.22)	19 (51.35)	8 (21.62)	20 (54.05)
Total well classified n= 3,230	1,922 (59.50)	2,944 (91.15)	1,719 (53.22)	3,017 (93.41)	2,958 (91.58)	3,025 (93.65)	2,847 (88.14)	3,015 (93.34)

NOTAS: Esta tabla contiene los resultados obtenidos con el modelo LogDA tomando distribuciones iguales. Diez ratios incluye: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TA es Gastos Financieros/ Deuda Total; CF/TA es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: R1, R3, R4, R8, R9; al nivel del 5% se incluyen: R1, R3, R4, R5, R7, R8, R9. La selección de medias a nivel del 1% (igual que al nivel del 5%) incluye: R6, R8.

Los resultados obtenidos con el análisis discriminante logístico muestran que las empresas no fracasadas se identifican mejor con un menor número de variables, especialmente cuando se añade la variable de incidencias. Por el contrario, las empresas fracasadas se identifican mejor con un número mayor de ratios. Con 10 o 7 ratios, añadir la variable incidencias incluso reduce poder discriminante, pero cuando se utilizan menos ratios, añadir esta variable mejora considerablemente la capacidad de discriminación de los modelos. De hecho, para 7, 5, y 2 ratios con la variable incidencias, el poder discriminante respecto a las empresas fracasadas es estable.

Tabla 14. Análisis Discriminante del Knn vecino más próximo (KnnDA)

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	3,151 (98.68)	3,163 (99.06)	3,153 (98.75)	3,161 (99.00)	3,163 (99.06)	3,156 (98.84)	3,150 (98.65)	3,160 (98.97)
Failed n= 37	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)
Total well classified n= 3,230	3,188 (98.70)	3,200 (99.07)	3,190 (98.76)	3,198 (99.01)	3,200 (99.07)	3,193 (98.85)	3,187 (98.67)	3,197 (98.98)

NOTAS: Esta tabla contiene los resultados obtenidos con el modelo KnnDA tomando distribuciones iguales. Diez ratios incluye: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: R1, R3, R4, R8, R9; al nivel del 5% se incluyen: R1, R3, R4, R5, R7, R8, R9. La selección de medias a nivel del 1% (igual que al nivel del 5%) incluye: R6, R8.

Usando el análisis discriminante del K-ésimo vecino más próximo, se puede apreciar una excelente capacidad de discriminación, tanto para las empresas no fracasadas como para las empresas fracasadas. Todas las empresas que han resultado fracasadas quedan identificadas con cualquier selección de variables. En cuanto a las empresas no fracasadas, cuando se utilizan solo ratios, cualquier selección de ratios según las medianas (7 o 5 ratios) se comporta mejor que la selección de ratios según las medias (2 ratios). Además, 5 ratios obtienen mejores resultados que 7 ratios, y 7 ratios funcionan mejor que 10 ratios. Con la metodología no paramétrica la selección con diferencias en medianas se comporta mejor que la selección con diferencias en medias.

Tabla 15. Análisis logit y probit

	All ratios		Median selection (5% and 1%)				Mean selection (1%)	
	10 ratios	10 ratios +inc.	7 ratios	7 ratios + inc.	5 ratios	5 ratios +inc.	2 ratios	2 ratios +inc.
LOGIT								
Prob>chi2	0.0038***	0.0000***	0.4951	0.0000***	0.5832	0.0000***	0.0003***	0.0000***
McFadden's R2	0.0642	0.1674	0.0158	0.1216	0.0093	0.1157	0.0406	0.145
McFadden's Adj R2	0.01	0.108	-0.024	0.077	-0.02	0.081	0.026	0.125
BIC	-25631.91	-25665.59	-25636.61	-25671.31	-25650.15	-25685.06	-25687.02	-25721.17
Area under ROC	0.7101	0.8052	0.6737	0.7894	0.5288	0.7099	0.6921	0.8116
PROBIT								
Prob>chi2	0.0032***	0.0000***	0.5916	0.0000***	0.6121	0.0000***	0.0003***	0.0000***
McFadden's R2	0.0653	0.1712	0.0138	0.1252	0.0088	0.1215	0.0405	0.1512
McFadden's Adj R2	0.011	0.112	-0.026	0.081	-0.021	0.087	0.026	0.131
BIC	-25632.39	-25667.13	-25635.78	-25672.74	-25649.95	-25687.41	-25686.99	-25723.69
Area under ROC	0.7145	0.8077	0.6835	0.784	0.5612	0.7118	0.6909	0.8105

Panel B. Resultados del análisis logit

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3,193 (100.00)	3,193 (100.00)
Failed n= 37	2 (5.41)	2 (5.41)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (5.41)	2 (5.41)
Total well clasifie n= 3,230	3,195 (98.92)	3,195 (98.92)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,195 (98.92)	3,195 (98.92)

Panel C. Resultados del análisis probit

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	2 ratios number (percentage)	2 ratios +inc. number (percentage)
Not-failed n=3,193	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3,193 (100.00)	3,193 (100.00)
Failed n= 37	1 (2.70)	1 (2.70)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (5.41)	1 (2.70)
Total well clasifie n= 3,230	3,194 (98.89)	3,194 (98.89)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,195 (98.92)	3,194 (98.89)

NOTAS: Esta tabla contiene los resultados obtenidos con los modelos Logit y Probit. Diez ratios incluye: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: R1, R3, R4, R8, R9; al nivel del 5% se incluyen: R1, R3, R4, R5, R7, R8, R9. La selección de medias a nivel del 1% (igual que al nivel del 5%) incluye: R6, R8.

Los resultados logit y probit en nuestro estudio muestran que:

- Para la misma selección de variables, ambos modelos obtienen prácticamente los mismos resultados, clasificando correctamente las empresas no fracasadas.
- Se obtienen menores valores en el R^2 de McFadden a medida que disminuye el número de variables.
- Para un mismo nivel de significación (1%) la selección de medias obtiene mejores resultados que la selección de medianas. De hecho, utilizando la

selección de medianas para reducir el número de ratios (5 o 7 ratios) los modelos no son significativos.

- Añadir la variable incidencias mejora considerablemente la calidad de los modelos, haciendo significativa cualquier selección de ratios.
- Una vez que hemos calculado el R^2 ajustado de McFadden con el fin de evitar la influencia del número de variables, apreciamos que la capacidad discriminante de los modelos es bastante inferior³¹. Sin embargo, la selección de medias (2 variables) muestra un poder discriminante mejor que el modelo con 10 ratios. Y el mismo patrón se mantiene cuando se añade la variable incidencias.
- Otra posibilidad de comparar los modelos logit y probit es utilizar BIC (*Bayesian Information Criterion*), basado en la probabilidad de los modelos en cuestión y en sus grados de libertad, siendo mejor el ajuste del modelo cuanto más negativo es el BIC. Los resultados del contraste BIC son mejores para aquellos modelos con menor número de variables, es decir, para la selección de medias (2 ratios).
- Por último, hemos utilizado el área bajo la curva ROC para comparar la bondad del ajuste de los diferentes modelos³². En este caso, parece que los métodos con más variables tienden a comportarse mejor que los métodos con un menor número de variables, excepto para el modelo con 2 ratios e incidencias, que obtiene el mejor resultado.

³¹ El R^2 ajustado elimina los casos relacionados con el marginal del grupo con más casos. Indica el porcentaje de clasificaciones correctas por encima de los que se obtienen si se ponen todas las empresas en el grupo con el marginal más alto. Esta prueba muestra si el modelo tiene una capacidad real para generar una buena clasificación.

³² Las curvas ROC (*Receiver Operating Characteristics*) las utilizan Altman *et al.* (2008) para evaluar el desempeño de los modelos logit. En una curva ROC la tasa de verdaderos positivos (sensibilidad) se representa en función de la tasa de falsos positivos (especificidad 100) para diferentes puntos de corte de un parámetro. Cada punto de la curva ROC representa un par sensibilidad / especificidad correspondiente a un umbral de decisión en particular para discriminar empresas que han fracasado de las que no. El área bajo la curva ROC es una medida de la precisión en la discriminación, donde el valor 1 significa que un modelo es perfecto. Tanto el coeficiente de Gini como el estadístico Kolmogorov-Smirnov (KS), comúnmente utilizados por los analistas de *scoring*, derivan de las curvas ROC.

2.6. Análisis de datos winsorizados

Con el fin de evitar la influencia de los valores extremos en algunas estadísticas, pero conservando el mismo número de observaciones, hemos winsorizado las variables³³. Para este procedimiento los datos inferiores al 1% se igualan al 1^{er} percentil mientras que el 1% de los datos superiores se igualan al percentil 99^o. La técnica reduce el efecto de los valores extremos y acerca la media de los datos corregidos a la mediana con una reducción en la desviación típica. Además de los resultados con las metodologías utilizadas hasta ahora, se muestra una tabla con los resultados del DEA para cada combinación de las variables contables seleccionadas estadísticamente.

³³ A diferencia del procedimiento utilizado, por ejemplo, en Altman *et al.* (2008), en el que los valores extremos se eliminan, restringiendo los rangos a los datos entre los percentiles 1 y 99. Como procedimiento alternativo, se han realizado transformaciones Box Cox con el fin de estandarizar las variables, y luego usar el test de Shapiro Wilk. Solo se han podido normalizar las variables de R2 a R5 sin buenos resultados. Por lo tanto, la normalización no es una buena alternativa para reducir la dispersión de nuestras variables.

Tabla 16. Estadísticos descriptivos, fracasadas/no fracasadas diferencias en medias, ratios winsorizados

Variable	Failed firms n = 37 Mean (Std. dev.)	Not-failed firms n = 3193 Mean (Std. dev.)	Difference in means (t-Statistic)
wR1= TD/TA	1.019683 (0.6442166)	0.7876563 (0.4490908)	-0.2320266 (-3.10639)***
wR2= CA/CL	2.527121 (6.940696)	2.783909 (5.769318)	0.2567876 (0.2685147)
wR3= BAIT/TA	-0.1692527 (0.4425078)	0.022242 (0.220357)	0.1914947 (5.168998)***
wR4= NI/TA	-0.1896237 (0.4113469)	-0.0056056 (0.1911801)	0.1840181 (5.706902)***
wR5= CA/TA	0.7744578 (0.2726751)	0.7263177 (0.2373353)	-0.0481401 (-1.224531)
wR6= FE/TD	0.0355791 (0.0330995)	0.0297533 (0.032074)	-0.0058258 (-1.098113)
wR7= RP/TA	0.1500051 (0.4884795)	0.1956719 (0.3814027)	0.0456668 (0.7215575)
wR8= CF/TD	0.0164868 (0.5752716)	0.1211922 (0.3389109)	0.1047053 (1.849157)*
wR9= NI/SL	-0.2360054 (0.5619837)	-0.0638995 (0.475318)	0.1721059 (2.184988)**
wR10= SL/TA	1.919005 (1.894715)	1.661431 (1.448189)	-0.257574 (-1.071418)
incidents	0.6216216 (0.720777)	0.0720326 (0.3133578)	-0.5495891 (-10.36207)***
Num.empl	12.1875 (11.55404)	32.46496 (588.1141)	20.27746 (0.1378802)

NOTAS: La muestra contiene 3.230 observaciones durante 2008. Definición de las variables: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.

* nivel de significación del 10% basado en el test de dos colas.

** nivel de significación del 5% basado en el test de dos colas.

*** nivel de significación del 1% basado en el test de dos colas.

Tabla 17. Estadísticos descriptivos fracasadas/no fracasadas, Rank sum test, ratios winsorizados

Variable	Failed firms n = 37	Not-failed firms n = 3193	Two sample Wilcoxon rank-sum test between medians (Z)
	Median (Std. dev.) [Skewness]	Median (Std. dev.) [Skewness]	[p-value > (Z)]
wR1= TD/TA	0.9751387 (0.6442166) [1.777834]	0.7783218 (0.4490908) [2.483903]	-3.551 [0.0004]***
wR2= CA/CL	1.057278 (6.940696) [5.563727]	1.274884 (5.769318) [5.38778]	1.714 [0.0866]
wR3= BAIT/TA	-0.0164002 (0.4425078) [-1.603176]	0.0444596 (0.220357) [-2.688866]	3.094 [0.0020]***
wR4= NI/TA	-0.0354618 (0.4113469) [-1.710292]	0.0159537 (0.1911801) [-3.040909]	4.118 [0.0000]***
wR5= CA/TA	0.8938961 (0.2726751) [-1.451836]	0.7864155 (0.2373353) [-0.9497068]	-1.976 [0.0481]**
wR6= FE/TD	0.032181 (0.0330995) [2.367284]	0.0214282 (0.032074) [1.978609]	-1.638 [0.1015]
wR7= RP/TA	0.0299024 (0.4884795) [2.808293]	0.1062388 (0.3814027) [2.493171]	2.27 [0.0232]**
wR8= CF/TD	-0.0286602 (0.5752716) [2.43246]	0.0608538 (0.3389109) [2.552918]	4.474 [0.0000]***
wR9= NI/SL	-0.0441338 (0.5619837) [-3.127511]	0.0124977 (0.475318) [-6.155788]	4.302 [0.0000]***
wR10= SL/TA	1.475444 (1.894715) [1.989572]	1.381083 (1.448189) [2.268183]	-0.434 [0.6641]
incidents	0 (0.720777) [0.694475]	0 (0.3133578) [4.695205]	-10.838 [0.0000]***
Num.empl	9 (11.55404) [1.095393]	6 (588.1141) [36.44888]	-0.716 [0.4737]

NOTAS: La muestra contiene 3.230 observaciones durante 2008. Definición de las variables: TD/TA es Deuda Total/ Activo Total; CA/CL es Activo Circulante/ Pasivo Circulante; BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; NI/TA es Beneficio Neto / Activo Total; CA/TA es Activo Circulante/ Activo Total; RP/TA es Beneficios No Distribuidos/ Activo Total; FE/TD es Gastos Financieros/ Deuda Total; CF/TD es Recursos Generados/ Deuda Total; NI/SL es Beneficio Neto/ Ventas; y SL/TA es Ventas/ Activo Total.

** nivel de significación del 5% basado en el test de dos colas.
*** nivel de significación del 1% basado en el test de dos colas.

Con la nueva selección de variables obtenidas por diferencias en medias y por diferencias en medianas, y con los datos winsorizados, procedemos a realizar los cálculos de todos los modelos anteriores. En la selección de medianas con un nivel de significatividad del 1% se obtienen 5 ratios (wR1, wR3, wR4, wR8 y wR9), en la selección de medias, para un nivel de significatividad del 10% se obtienen los mismos ratios, por lo que no se repiten los resultados.

Tabla 18. Análisis Discriminante Lineal (LDA), ratios winsorizados

	All ratios			Median selection (5% and 1%)			Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	2,776 (86.94)	2,939 (92.05)	2,798 (87.63)	2,935 (91.92)	2,836 (88.82)	2,934 (91.89)	2,822 (88.38)	2,944 (92.20)
Failed n= 37	18 (48.65)	22 (59.46)	17 (45.95)	22 (59.46)	13 (35.14)	22 (59.46)	13 (35.14)	21 (56.76)
Total well classified n= 3,230	2,794 (87.04)	2,961 (92.24)	2,815 (87.69)	2,957 (92.12)	2,849 (88.75)	2,956 (92.09)	2,835 (88.32)	2,965 (92.37)

NOTAS: Esta Tabla contiene los resultados obtenidos con el modelo LDA tomando distribuciones iguales. Diez ratios incluyen: wR1=TD/TA es Deuda Total/ Activo Total; wR2=CA/CL es Activo Circulante/ Pasivo Circulante; wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; wR4=NI/TA es Beneficio Neto / Activo Total; wR5=CA/TA es Activo Circulante/ Activo Total; wR6=RP/TA es Beneficios No Distribuidos/ Activo Total; wR7=FE/TD es Gastos Financieros/ Deuda Total; wR8= CF/TD es Recursos Generados/ Deuda Total; wR9=NI/SL es Beneficio Neto/ Ventas; y wR10=SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye; wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1%.

Tras winsorizar los ratios, los resultados del análisis discriminante lineal muestran un patrón irregular. Ahora, el mejor porcentaje de empresas bien clasificadas se obtiene para la selección de medias al 1% después de añadir la variable incidencias, aunque es solo ligeramente mejor que el porcentaje del modelo con 10 ratios y la variable incidencias. En los modelos que solo toman ratios como variables independientes, un menor número de variables produce mejores clasificaciones en las empresas no fracasadas pero peores clasificaciones en las empresas fracasadas, excepto para el modelo con 3 ratios. La inclusión de la variable incidencias produce una mejora en la selección de las empresas fracasadas, haciendo que sea estable el número de empresas fracasadas bien clasificadas. A cambio, la disminución de empresas no fracasadas bien clasificadas es pequeña para la selección de medianas e insignificante para la selección de medias al 5%. En suma, para nuestra muestra, después de winsorizar los datos, LDA tiende a seleccionar mejor las empresas fracasadas, pero a cambio selecciona peor las empresas no fracasadas.

Tabla 19. Análisis Discriminante Cuadrático (QDA), ratios winsorizados

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	2,947 (92.30)	2,916 (91.32)	2,905 (90.98)	2,829 (88.60)	2,847 (89.16)	2,778 (87.00)	2,944 (92.20)	2,873 (89.98)
Failed n= 37	16 (43.24)	25 (67.57)	11 (29.73)	25 (67.57)	11 (29.73)	24 (64.86)	7 (18.92)	22 (59.46)
Total well classified n= 3,230	2,963 (92.31)	2,941 (91.62)	2,916 (90.84)	2,854 (88.91)	2,858 (89.03)	2,802 (87.29)	2,951 (91.93)	2,895 (90.19)

NOTAS: Esta Tabla contiene los resultados obtenidos con el modelo QDA tomando distribuciones iguales. Diez ratios incluyen: wR1=TD/TA es Deuda Total/ Activo Total; wR2=CA/CL es Activo Circulante/ Pasivo Circulante; wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; wR4=NI/TA es Beneficio Neto / Activo Total; wR5=CA/TA es Activo Circulante/ Activo Total; wR6=RP/TA es Beneficios No Distribuidos/ Activo Total; wR7=FE/TD es Gastos Financieros/ Deuda Total; wR8= CF/TD es Recursos Generados/ Deuda Total; wR9=NI/SL es Beneficio Neto/ Ventas; y wR10=SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye: wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1%.

Tanto en la selección de las medias como en la selección de las medianas se obtienen buenos resultados en la identificación de las empresas no fracasadas, aunque la selección de las medianas muestra que menos variables implican una reducción del poder discriminante, mientras que la selección de medias obtiene un poder discriminante mayor con un menor número de variables (ligeramente inferior a tomar los 10 ratios). Por otro lado, en la clasificación de las empresas fracasadas, más ratios obtienen una mejor clasificación, y la selección de medias funciona peor que la selección de medianas, al mismo nivel de significación. A diferencia del LDA, tras la winsorización de los ratios, QDA identifica mucho mejor las empresas no fracasadas, pero la identificación de empresas fracasadas es más pobre.

Tabla 20. Análisis Discriminante Logístico (LogDA), ratios winsorizados

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	2,285 (71.56)	2,716 (85.06)	2,277 (71.31)	2,735 (85.66)	2,715 (85.03)	2,826 (88.51)	2,747 (86.03)	2,861 (89.60)
Failed n= 37	26 (70.27)	26 (70.27)	24 (64.86)	26 (70.27)	15 (40.54)	23 (62.16)	15 (40.54)	22 (59.46)
Total well classified n= 3,230	2,311 (71.99)	2,742 (85.42)	2,301 (71.68)	2,761 (86.01)	2,730 (85.05)	2,849 (88.75)	2,762 (86.04)	2,883 (89.81)

NOTAS: Esta Tabla contiene los resultados obtenidos con el modelo LogDA tomando distribuciones iguales. Diez ratios incluyen: wR1=TD/TA es Deuda Total/ Activo Total; wR2=CA/CL es Activo Circulante/ Pasivo Circulante; wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; wR4=NI/TA es Beneficio Neto / Activo Total; wR5=CA/TA es Activo Circulante/ Activo Total; wR6=RP/TA es Beneficios No Distribuidos/ Activo Total; wR7=FE/TD es Gastos Financieros/ Deuda Total; wR8= CF/TD es Recursos Generados/ Deuda Total; wR9=NI/SL es Beneficio Neto/ Ventas; y wR10=SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye: wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1%.

Después de winsorizar los ratios, el LogDA muestra los siguientes resultados: de nuevo, las empresas no fracasadas se identifican mejor con un menor número de variables, especialmente cuando se añade la variable incidencias. Por el contrario, en las empresas fracasadas cuando el número de variables disminuye el número de empresas bien clasificadas es menor, aunque añadiendo la variable incidencias mejora el poder discriminante para cualquier selección de variables. En general, tras winsorizar los ratios, LogDA obtiene mejores resultados con las diferentes selecciones de variables. Al mejorar la identificación de las empresas no fracasadas para un número mayor de ratios, y la identificación de las empresas fracasadas para un número menor de ratios, los porcentajes totales de las empresas bien clasificadas mejoran para 10 y 7 ratios, sin la variable incidencias (cuando los resultados previos eran más pobres), pero empeoran en el resto de casos.

Tabla 21. Análisis Discriminante del K-ésimo vecino más próximo (KnnDA), ratios winsorizados

	All ratios			Median selection (5% and 1%)			Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	3,152 (98.72)	3,162 (99.03)	3,153 (98.75)	3,161 (99.00)	3,159 (98.94)	3,153 (98.75)	3,143 (98.43)	3,161 (99.00)
Failed n= 37	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)	37 (100.00)
Total well classified n= 3,230	3,189 (99.35)	3,199 (99.66)	3,190 (99.38)	3,198 (99.63)	3,196 (99.56)	3,190 (99.38)	3,180 (99.07)	3,198 (99.63)

NOTAS: Esta Tabla contiene los resultados obtenidos con el modelo KnnDA tomando distribuciones iguales. Diez ratios incluyen: wR1=TD/TA es Deuda Total/ Activo Total; wR2=CA/CL es Activo Circulante/ Pasivo Circulante; wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; wR4=NI/TA es Beneficio Neto / Activo Total; wR5=CA/TA es Activo Circulante/ Activo Total; wR6=RP/TA es Beneficios No Distribuidos/ Activo Total; wR7=FE/TD es Gastos Financieros/ Deuda Total; wR8= CF/TD es Recursos Generados/ Deuda Total; wR9=NI/SL es Beneficio Neto/ Ventas; y wR10=SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye: wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1%.

Los resultados de KnnDA con los ratios winsorizados mantienen el mismo patrón que con los datos sin winsorizar. En general, un excelente poder discriminante: todas las empresas fracasadas han sido correctamente identificadas y las empresas no fracasadas alcanzan un alto porcentaje de clasificación correcta, con cualquier selección de variables. En cuanto a las empresas no fracasadas, usando sólo los ratios, la selección de medianas obtiene mejores resultados para un nivel de significación del 1%. Y de nuevo, cuanto menor es el número de variables contables, mejor poder discriminante se obtiene. Cuando se añade la variable

incidencias, se puede apreciar un buen comportamiento, y bastante estable, para cualquier selección de variables.

Tabla 22. Análisis logit y probit, ratios winsorizados

	All ratios		Median selection (5% and 1%)				Mean selection (1%)	
	10 ratios	10 ratios +inc.	7 ratios	7 ratios + inc.	5 ratios	5 ratios +inc.	3 ratios	3 ratios +inc.
LOGIT								
Prob>chi2	0.0004*	0.0000*	0.0001*	0.0000*	0.0009*	0.0000*	0.0002*	0.0000*
McFadden's R2	0.0798	0.1545	0.0765	0.1529	0.0515	0.1329	0.0497	0.1298
McFadden's Adj R2	0.025	0.095	0.037	0.108	0.022	0.098	0.03	0.105
BIC	-25638.2	-25660.3	-25661.1	-25683.9	-25667.1	-25692	-25682.6	-25706.9
Area under ROC	0.7576	0.8326	0.7518	0.8288	0.6742	0.7733	0.6908	0.7686
PROBIT								
Prob>chi2	0.0003*	0.0000*	0.0000*	0.0000*	0.0010*	0.0000*	0.0002*	0.0000*
McFadden's R2	0.0814	0.1619	0.0785	0.1607	0.051	0.142	0.0485	0.1377
McFadden's Adj R2	0.027	0.102	0.039	0.116	0.021	0.107	0.029	0.113
BIC	-25638.8	-25663.3	-25661.9	-25687.1	-25667	-25695.7	-25682.1	-25710.1
Area under ROC	0.7679	0.8389	0.7608	0.835	0.7061	0.7875	0.7103	0.7849

Panel B. Resultados del análisis logit, ratios winsorizados.

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	3193 (100.00)	3192 (99.97)	3193 (100.00)	3192 (99.97)	3193 (100.00)	3193 (100.00)	3,193 (100.00)	3,193 (100.00)
Failed n= 37	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)
Total well classifie n= 3,230	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,193 (98.85)	3,194 (98.89)	3,193 (98.85)	3,194 (98.89)

Panel C. Resultados del análisis probit, ratios winsorizados.

	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios number (percentage)	10 ratios +inc. number (percentage)	7 ratios number (percentage)	7 ratios + inc. number (percentage)	5 ratios number (percentage)	5 ratios +inc. number (percentage)	3 ratios number (percentage)	3 ratios +inc. number (percentage)
Not-failed n=3,193	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3193 (100.00)	3,193 (100.00)	3,193 (100.00)
Failed n= 37	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)	0 (0.00)	1 (2.70)
Total well classifie n= 3,230	3,193 (98.85)	3,194 (98.89)	3,193 (98.85)	3,194 (98.89)	3,193 (98.85)	3,194 (98.89)	3,193 (98.85)	3,194 (98.89)

NOTAS: Esta Tabla contiene los resultados obtenidos con los modelos logit y probit. Diez ratios incluyen: wR1=TD/TA es Deuda Total/ Activo Total; wR2=CA/CL es Activo Circulante/ Pasivo Circulante; wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total; wR4=NI/TA es Beneficio Neto / Activo Total; wR5=CA/TA es Activo Circulante/ Activo Total; wR6=RP/TA es Beneficios No Distribuidos/ Activo Total; wR7=FE/TD es Gastos Financieros/ Deuda Total; wR8= CF/TD es Recursos Generados/ Deuda Total; wR9=NI/SL es Beneficio Neto/ Ventas; y wR10=SL/TA es Ventas/ Activo Total. La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye: wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1%.
*Nivel de significación del 1%.

Después de winsorizar los ratios, hay un primer efecto relevante: todos los modelos son significativos al nivel del 1%. Es decir, podemos obtener modelos con capacidad discriminante con cualquier selección de variables, tanto con la

selección de medias como con la selección de medianas. Si examinamos los resultados en detalle, encontramos lo siguiente:

- R^2 de McFadden muestra mejor poder discriminante para el modelo con 10 ratios, seguido por el modelo con 3 ratios, mejorando considerablemente cuando se añade la variable incidencias.
- En este caso, probit tiende a funcionar ligeramente mejor que logit, aunque en menos modelos que usando ratios sin winsorizar.
- En el mismo nivel de significación, la selección de medias es mejor al 1%, mientras que la selección de medianas es mejor al 5%. Y estos resultados se mantienen sin cambios, con el R^2 ajustado de McFadden.
- Podemos observar a través del estadístico BIC que usando un menor número de variables se obtienen mejores resultado. Por lo tanto, la selección de medias es mejor para cualquier nivel de significación, y, como siempre, un menor número de variables son más significativas para cualquier nivel con la selección de medias.
- Uso de las curvas ROC, una vez más las áreas bajo las curvas muestran que los modelos con un mayor número de variables dan un mejor resultado.

Tabla 23. DEA. Rendimiento constante a escala (CRS), ratios winsorizados.

Orientación	All ratios		Median selection (5% and 1%)				Mean selection(1%)	
	10 ratios		7 ratios		5 ratios		3 ratios	
	Input number (percentage)	Output number (percentage)						
Not-failed n=3,193	1,606 (50.30)	1,606 (50.30)	1,606 (50.30)	1,604 (50.23)	1,606 (50.30)	1,604 (50.23)	1,605 (50.27)	1,597 (50.02)
Failed n= 37	28 (75.68)	28 (75.68)	21 (56.76)	26 (70.27)	6 (16.22)	26 (70.27)	24 (64.86)	19 (51.35)
Total well clas n= 3,230	1,634 (50.90)	1,634 (50.90)	1,627 (50.69)	1,630 (50.78)	1,612 (50.22)	1,630 (50.78)	1,629 (50.75)	1,616 (50.34)

Diez ratios incluyen (con la clasificación *input/output* entre paréntesis): wR1=TD/TA es Deuda Total/ Activo Total (i); wR2=CA/CL es Activo Circulante/ Pasivo Circulante (o); wR3=BAIT/TA es Beneficio Antes de Intereses e Impuestos/ Activo Total (o); wR4=NI/TA es Beneficio Neto / Activo Total (o); wR5=CA/TA es Activo Circulante/ Activo Total (i); wR6=RP/TA es Beneficios No Distribuidos/ Activo Total (i); wR7=FE/TD es Gastos Financieros/ Deuda Total (o); wR8= CF/TD es Recursos Generados/ Deuda Total (o); wR9=NI/SL es Beneficio Neto/ Ventas (o); y wR10=SL/TA es Ventas/ Activo Total (i). La selección de medianas al nivel del 1% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, al nivel del 5% (7 ratios) incluye: wR1, wR3, wR4, wR5, wR7, wR8, wR9. La selección de medias a nivel del 1% (3 ratios) incluye: wR1, wR3, wR4, al nivel del 10% (5 ratios) incluye: wR1, wR3, wR4, wR8, wR9, los mismos ratios que al nivel del 1% de la selección de medianas.

Nuestros resultados con DEA muestran que los mejores modelos discriminantes son aquellos con un mayor número de variables, lo cual es consistente con Golany y Roll (1989) y Nunamaker (1985). Nuestra amplia muestra hace innecesaria la

precaución de utilizar un número de variables de *input* y de *output* no superior a un tercio del número de DMUs (Boussofiane *et al.*, 1991; Friedman y Sinuany-Stern, 1998). Por lo tanto, nuestro interés por reducir el número de variables se debe a las limitaciones de los equipos informáticos para ejecutar los modelos.

Como hemos elegido las variables en orden de importancia, el equilibrio entre el número de variables *input* y *output* no es una prioridad. Estudios anteriores (Jenkins y Anderson, 2003) sobre selección de variables para DEA sólo requieren/recomiendan que se mantengan al menos un *input* y un *output*, lo que se cumple en cualquier de los grupos de variables que hemos seleccionado. En nuestra selección inicial se puede apreciar que seis de las diez variables son *outputs* mientras que cuatro son *inputs*, aunque este hecho no origina una diferencia, cuando se toma el número total de variables, entre la orientación al *input* y la orientación al *output* de los CRS del DEA. Tal como se ha indicado previamente, cuando el modelo CRS es el que funciona mejor, se obtienen resultados iguales en la orientación al *input* y en la orientación al *output*.

No obstante, hay una diferencia notable cuando se utiliza un menor número de variables, especialmente con la selección de 5 variables. Considerando que, de 7 variables sólo 2 son *inputs*, y de 5, sólo 1 es *input*, nuestros resultados sugieren que un mayor número de variables *input* (*output*) realizan una mejor discriminación en orientaciones al *input* (*output*). En la última selección, con 1 *input* y 2 *outputs*, los mejores resultados del método de orientación al *input* pueden ser atribuidos al poder discriminante de las variables individuales en lugar de a los números relativos de *inputs* y *outputs*, que están, en este caso, mucho más cercanos.

En nuestro estudio, el desempeño del DEA es mejor en la clasificación de las empresas fracasadas que de las empresas no fracasadas, no sólo el método CRS (Tabla 23), sino también cuando utilizamos rendimientos variables a escala (VRS) y rendimientos decrecientes a escala (DRS) (no tabulados), tanto con orientación al *input* como con orientación al *output*. Las empresas no fracasadas que están bien clasificadas rondan el 50%, independientemente del método o la orientación.

Se pueden apreciar diferencias en las empresas fracasadas, para las que CRS se comporta mejor que VRS y DRS, y la orientación al *output* obtiene mejores resultados de clasificación cuando el número de *outputs* supera al número de *inputs*.

En comparación con el resto de las metodologías utilizadas en esta segunda parte de la tesis, DEA obtiene peores clasificaciones de empresas no fracasadas que el resto de los modelos de análisis discriminante. En cuanto a las empresas fracasadas, DEA se comporta mejor que el análisis discriminante lineal y que el cuadrático, obtiene resultados similares a los del análisis discriminante logístico, pero está bastante lejos de los excelentes resultados del modelo del K-ésimo vecino más próximo. Por lo que se refiere a logit y probit, si tomamos como medida global de la tasa de buena clasificación el área bajo la curva ROC, obtenemos en ambas metodologías resultados de clasificación mejores que en DEA para cualquier selección de variables (aunque DEA clasifica mucho mejor a las empresas fracasadas). Después de reducir el número de variables, DEA obtiene mejores resultados con la selección de medianas que con la de medias. Esto era un resultado esperado, dado que la selección de la mediana contiene un mayor número de variables.

2.7. Conclusiones a la segunda parte

Con esta segunda parte de la tesis, completamos un análisis exhaustivo de la selección de variables, que hemos identificado como uno de los tres elementos básicos en el estudio empírico del fracaso empresarial (junto con el concepto de fracaso, y el método utilizado para identificar o predecir el fracaso en la empresa). Así, realizamos un análisis complementario de variables en la literatura previa, tanto de aquellas que resultan significativas con más frecuencia como de aquellos rasgos que subyacen a esas variables, repitiendo el estudio realizado en la primera parte, pero esta vez sobre los trabajos empíricos aplicados a pymes.

Usando una muestra de 23 trabajos empíricos internacionales aplicados a pymes, hemos confirmado que los mismos rasgos subyacen a las variables explicativas en pymes que en empresas de cualquier tamaño, si bien endeudamiento y estructura económica parecen resultar más discriminantes que rentabilidad. Rotación, equilibrio económico-financiero y margen muestran un poder discriminante similar. Destacamos la mayor utilización y poder discriminante de los ratios contables frente a otras variables, que se justifica por la mayor dificultad para acceder a información no contable de empresas no cotizadas. En cuanto a las variables concretas halladas significativas con más frecuencia en los estudios sobre pymes, la coincidencia con el estudio para empresas de cualquier tamaño es muy elevada: ocho de los diez ratios aparecen en los primeros diez puestos y otro aparece sustituido por un ratio similar. Teniendo en cuenta que los estudios sobre pymes no son numerosos, y que la diferencia de variables significativas es muy pequeña, hemos considerado adecuada la selección realizada en la primera parte de la tesis para ser aplicada en el estudio empírico de esta segunda parte.

En cuanto a la selección de variables de segundo nivel, utilizando contrastes estadísticos, nuestro análisis empírico muestra que los diferentes procedimientos estadísticos generan diferentes selecciones de variables, excepto en condiciones muy específicas. La muestra de pequeñas empresas ofrece una amplia dispersión en la mayoría de los ratios contables calculados. Los valores extremos aumentan la desviación típica y la asimetría hace que la media se aleje de la mediana. Como consecuencia, la selección de variables realizada utilizando diferencias en medias y la selección con diferencias en medianas generan grupos de variables discriminantes significativas radicalmente diferentes.

Hemos aplicado las técnicas más utilizadas en el análisis del fracaso empresarial, el análisis discriminante (DA) y los análisis de regresión binomial logit y probit. Aparte del análisis discriminante lineal (LDA) y el análisis discriminante cuadrático (QDA), ambas técnicas paramétricas bien conocidas, hemos utilizado una técnica semi-paramétrica, el análisis discriminante logístico (LogDA), y una técnica no paramétrica, el análisis discriminante del K-ésimo vecino más próximo (KnnDA). En cuanto a la comparación de los dos tipos de regresiones binomiales, hemos

incorporado varios contrastes recientes para comparar la bondad del ajuste de logit y probit para los diferentes grupos de variables. Además, hemos utilizado una de las metodologías no paramétricas de más reciente aplicación al fracaso empresarial, el análisis envolvente de datos (DEA), en tres versiones diferentes, rendimientos constantes a escala (CRS), rendimientos variables a escala (VRS) y rendimientos decrecientes a escala (DRS), y con dos orientaciones para cada versión, al *input* y al *output*.

Un resultado interesante es que la selección de variables que utiliza diferencias en medias ofrece una mejor capacidad de discriminación en las metodologías paramétricas, mientras que las metodologías no paramétricas funcionan mejor con las selecciones de variables basadas en las medianas.

Después de refinar la muestra, winsorizando los ratios al nivel del 1%, con el fin de homogeneizar las estadísticas (pero manteniendo el tamaño de la población), encontramos que, con diferencias en medias, los ratios seleccionados como significativos son completamente diferentes de los seleccionados antes winsorizar. En cambio, la selección con diferencias en medianas mantiene los mismos ratios antes y después de winsorizar. Como se esperaba, tras la winsorización la selección de variables con medias está más cerca de la selección con medianas, aunque todavía son distintas para diferentes niveles de confianza.

Nuestro trabajo demuestra que el análisis de medias no es un método adecuado para seleccionar las variables discriminantes cuando éstas sufren una gran dispersión (sobre todo, si se trata de una dispersión asimétrica o hay valores extremos sesgando la media). Por otro lado, la selección con medias será una manera apropiada para seleccionar variables discriminantes cuando la población o la muestra analizada se distribuyen normalmente, o al menos los valores medios se aproximan a los valores medianos de esas variables. Aunque esta condición se atribuye al LDA, nuestro estudio muestra que un problema similar ocurre con otros métodos paramétricos, tales como el QDA, el logit y el probit.

Nuestros resultados empíricos indican que una población con variables muy dispersas requiere un procedimiento diferente de selección de variables en consonancia con la metodología que se va a utilizar para discriminar o clasificar a empresas fracasadas y no fracasadas: los métodos paramétricos funcionan mejor con la selección de medias; mientras que los métodos no paramétricos obtienen mejores resultados con la selección de medianas. Pero, además, como los análisis discriminantes no paramétricos consiguen clasificaciones más precisas que los paramétricos en la clasificación de empresas fracasadas y no fracasadas, nuestros resultados sugieren que la metodología paramétrica está sesgada cuando los datos de las variables presentan una gran dispersión. Apoyando la idea de que la mediana es más representativa que la media cuando una variable presenta dispersión, nuestros resultados desaconsejan el uso del contraste estadístico de diferencias en medias y de las metodologías paramétricas de evaluación de fracaso empresarial en empresas pequeñas.

Los modelos DEA no son tan buenos en la identificación de las empresas no fracasadas como las distintas versiones del análisis discriminante y los métodos binomiales. Este resultado es consistente con la evidencia empírica todavía escasa que compara DEA con DA, logit y probit. Así, Premachandra *et al.* (2009) encuentran que el DEA es mucho más potente en la evaluación correcta de empresas fracasadas que de empresas no fracasadas, y que a medida que disminuye la relación fracasadas/no fracasadas, el porcentaje global de evaluaciones correctas se reduce gradualmente. También encuentran que el modelo logístico se comporta mejor que el modelo DEA dentro de la muestra, aunque el DEA supera al logit en la identificación de empresas ajenas a la muestra.

Los mismos modelos empíricos se han aplicado después de incorporar una variable cualitativa: incidencias de pago. Nuestros resultados muestran una mejoría general y significativa en el poder explicativo de los modelos utilizados, confirmando, de esta manera, la contribución a la precisión de los modelos de fracaso que puede conseguirse añadiendo variables específicas por empresa

ajenas a los estados financieros, tal como se ha encontrado en un grupo de trabajos, aún escaso, de la literatura previa. Más específicamente, nuestra evidencia apoya los hallazgos de Altman *et al.* (2008) en una zona geográfica distinta, usando también una amplia muestra de pequeñas empresas, lo que confirmaría que las variables cualitativas son aún más importantes para los modelos de fracaso aplicados a pequeñas empresas, teniendo en cuenta que para una gran parte de las mismas la información financiera es bastante limitada.

En suma, este trabajo contribuye a la evidencia previa de selección de variables de tres formas:

- En primer lugar, hemos probado dos procedimientos estadísticos distintos de selección de variables, mostrando los resultados radicalmente diferentes que se pueden obtener y las implicaciones relevantes para la elección de la metodología teniendo en cuenta los estadísticos descriptivos de los datos. Después de una revisión bibliográfica amplia, parece que el nuestro es el primer estudio centrado en esta cuestión.
- En segundo lugar, mostramos que la selección de variables a partir de la literatura empírica previa aplicada a pymes coincide con la realizada para cualquier tamaño de empresas en la primera parte de la tesis. Estaría sesgada si sólo se consideran ratios individuales, debido a la variedad de especificaciones de los mismos, por lo que un análisis de los rasgos económicos subyacentes a esas variables ayuda a evitar ese sesgo.
- En tercer lugar, aportamos evidencia nueva sobre las ventajas de añadir información cualitativa a los modelos de fracaso, siendo aún escasa la evidencia previa sobre variables cualitativas en pymes. Nuestros resultados muestran la contribución relevante de la variable incidencias de pago, una variable cualitativa no contable, para mejorar el poder discriminante de todos los modelos aplicados en nuestro trabajo, para cualquier selección de variables.

Por último, contribuimos a la literatura previa, añadiendo nueva evidencia empírica sobre la aplicación de modelos de análisis discriminante, logit, probit y DEA, a

varias selecciones de variables, siendo el uso de modelos de análisis discriminante no paramétricos o semi-paramétricos y la comparación de modelos DEA (CRS, VRS y DRS con orientación al *input* y al *output* en cada caso) innovadores respecto a la evidencia previa.

3. DESCRIPCIÓN DEL FRACASO EMPRESARIAL MEDIANTE FRONTERAS DE RATIOS CONTABLES SELECCIONADAS CON DIFERENCIAS DE PERCENTILES

3.1. Introducción

Resulta indiscutible el papel que las pymes desempeñan en la economía. Según el informe *Retrato de las PYME 2012* publicado por el Ministerio de Industria, Energía y Turismo, las pymes españolas con menos de 49 empleados representaron en 2010 el 99,2% del total de empresas, mientras que en los 27 países de la Unión Europea, las empresas con menos de 50 empleados constituyeron el 98,7% del total, lo que representa, en ambos casos, una inmensa mayoría. Por su parte, el informe *EU SMEs in 2012: at the crossroads. Annual report on small and medium-sized enterprises in the EU, 2011/12* estima que en el ejercicio 2012, en los 27 países de la Unión Europea las empresas pequeñas y medianas representarán el 99,8% del total de empresas, el 67,4% del empleo total y el 58,1% del valor añadido bruto. Todas estas cifras, junto con la identificación de características particulares en las pymes, que aconsejan el desarrollo de herramientas de gestión de riesgo específicamente desarrolladas para ellas, justifican la selección de este grupo de empresas como objeto de nuestro estudio. En épocas de crisis, como la actual, es de suma importancia analizar los mecanismos que llevan a un alto porcentaje de las pymes a situaciones límite, cuando no a la quiebra.

En cuanto al sector de la construcción, podemos apreciar su importancia en el tejido productivo nacional, así como el reflejo que la crisis iniciada en 2007 ha tenido hasta ahora sobre los niveles de fracaso en este sector, observando algunos datos que el Instituto Nacional de Estadística (INE) ofrece en su página

web³⁴. A la vista de los datos podemos constatar que el estallido de la burbuja inmobiliaria ha sido muy importante en el sector de la construcción, en comparación con el resto de los sectores de la economía, por tanto es especialmente interesante como objeto de estudio.

Respecto a los sistemas de evaluación del fracaso empresarial que utilizan información financiera, en la primera parte de la tesis hemos detectado dos deficiencias relevantes, de las que no nos hemos ocupado hasta ahora:

1. No se calculan distancias a la quiebra de las empresas porque los sistemas de clasificación suelen ser dicotómicos: clasifican las empresas en el conjunto de las fracasadas o en el conjunto de las sanas.
2. La información financiera tomada en forma de ratios no capta con suficiente rapidez cambios en el entorno³⁵. Los modelos no suelen tener en cuenta los cambios de ese entorno en el tiempo. De hecho, puede haber una confusión entre la evolución de la empresa y la evolución del conjunto de empresas del sector o de la zona geográfica en estudio, originada por los cambios del entorno y no por el desempeño de cada empresa individual.

Para evitar estas dos deficiencias, el objetivo de este trabajo es desarrollar un modelo de evaluación de fracaso empresarial orientado a pymes que permita, por un lado, la construcción de *scores* para puntuar empresas o grupos de empresa según su distancia al fracaso; y, por otro, la detección de puntos débiles en las

³⁴ Fuente: www.ine.es.

	Dato base 2007 (a)	Dato base 2011 (b)	Diferencia 2011-2007	% Dif.
Personas ocupadas (miles):				
· Total CNAE-93	20.356,2	18.654,1	-1.702,1	-8,36%
· Construcción	2.680,6	1.497,9	-1.182,7	-44,12%
· % Construcción/Total	13,17%	8,03%	69,48%	
PIB a precios de mercado (millones):				
· Total CNAE-93	1.053.161	1.063.355	+10.194	+0,97%
· Construcción	131.074	98.546	-32.528	-24,82%
· % Construcción./Total	12,45%	9,27%		

(a) Datos de ocupados referidos al primer trimestre de 2007

(b) Datos de ocupados referidos al segundo trimestre de 2011

³⁵ Y en las pymes, con carácter general, es difícil disponer de indicadores financieros ajenos a la información contable, dada la escasa información externa que aportan.

empresas que ayude en su reestructuración, cuando los modelos de identificación y predicción de fallido las identifican como empresas en riesgo de fracaso.

El modelo enlaza con la selección estadística de variables para discriminar entre las empresas fracasadas, o en riesgo de fracaso, y las empresas sanas, cuestión ampliamente tratada en la segunda parte de la tesis. En concreto, dada la dispersión de las variables contables en las pymes, partimos de la diferencia en medianas entre empresas sanas y quebradas, que se ha mostrado más adecuado en términos generales que la diferencia en medias.

Si el grupo de referencia para una empresa está bien elegido, la metodología tiene en cuenta factores ambientales que afectan a todo el grupo por igual (factores sectoriales y macroeconómicos). Lo ideal sería estudiar cómo van variando las referencias a medida que lo hacen los factores ambientales de estos dos tipos, sectoriales y macroeconómicos, pero a corto plazo (un año) puede proporcionar referencias realistas de lo que en ese momento es una empresa saludable o fallida, y de la distancia que separa a una empresa de esas dos referencias.

Así, en esta tercera parte de la tesis, hemos desarrollado una metodología novedosa basada en el cálculo de percentiles estadísticos que permite:

- Seleccionar los indicadores financieros que resulten más discriminantes, a efectos del suceso de quiebra o fracaso empresarial, en función de la población objeto de estudio: zona geográfica, sector y periodo temporal.
- Determinar fronteras de fracaso con los ratios seleccionados.
- Calcular la distancia al fracaso para empresas o conjuntos de empresas, para cada indicador financiero, pero también, y esto es lo más interesante, teniendo en cuenta grupos de indicadores con rangos de valores no comparables entre sí (como es el caso de los ratios contables). En el primer caso, las distancias permiten detectar deficiencias en las parcelas específicas de la empresa que refleja o aproxima el indicador financiero, por lo que puede ser una herramienta de gestión útil. En el segundo caso,

nuestra metodología permite elaborar *rankings* de empresas según su distancia al fracaso para una población y periodo concretos.

Para el estudio empírico, analizamos un periodo temporal y un sector concretos para probar la metodología: empresas pequeñas del sector de la construcción en los años 2006-10. La primera parte del trabajo se realiza con 268 empresas fracasadas sobre un total de 12.197, desde 2006 a 2008, en una zona geográfica concreta, la Comunidad Autónoma de Castilla y León. En la segunda parte del trabajo, hacemos un seguimiento de 11.889 empresas constructoras activas de toda España, durante el trienio 2008-10, de las cuales 28 fracasaron en 2010, con el fin de evaluar la evolución de sus indicadores financieros con la metodología propuesta.

3.2. Revisión del fracaso empresarial en pymes

A continuación recogemos los trabajos empíricos relacionados con riesgo de quiebra en empresas pequeñas y medianas, cuyos planteamientos, metodologías y resultados pueden servir como base de comparación al presente estudio. Aunque encontramos trabajos pioneros como Edmister (1972) y Lincoln (1984), que muestran la utilidad de los ratios contables en la identificación y predicción de quiebra o insolvencia en empresas pequeñas y medianas, la proporción de estudios empíricos sobre este tipo de empresas es mucho menor que la de estudios aplicados a empresas grandes, especialmente cotizadas. Las diferencias en cuanto a la accesibilidad de la información y en cuanto a su calidad o fiabilidad son las razones mencionadas para esa falta de atención (Edmister, 1972; Labatut *et al.*, 2009), si bien en la literatura de los últimos años se puede apreciar un reconocimiento general de que las empresas pequeñas y medianas reúnen características particulares que requieren herramientas de gestión de riesgo específicamente desarrolladas para ellas (Dietsch y Petey, 2002; Altman y Sabato, 2007; Behr y Guttler, 2007; Altman *et al.*, 2008; Davydenko y Franks, 2008).

Frente a las dudas sobre la fiabilidad de los datos correspondientes a empresas pequeñas y medianas, porque proceden de cuentas anuales en formatos abreviados que no requieren el informe de auditoría (Labatut *et al.*, 2009), se ha encontrado evidencia de que las pymes presentan una información similar en calidad a la de las grandes empresas. López *et al.* (1997) en un estudio sobre las pymes asturianas concluyen que no es posible afirmar que el tamaño medio de las empresas con defectos en las cuentas anuales depositadas sea inferior al de las empresas sin defectos “de forma clara y rotunda”. Este estudio pone en evidencia que, en algunos casos, la conclusión es, precisamente, la contraria. Sobre la diferencia entre empresas auditadas (las empresas grandes están obligadas) y no auditadas (las empresas pequeñas y medianas no están obligadas), Alemany y Monllau (1997) analizan empresas que depositan sus cuentas en el Registro Mercantil de Barcelona y constatan, por un lado, que el porcentaje de error es mayor en las primeras (12,6%) que en las segundas (10,2%), y por otro que, salvo en las deficiencias por ausencia de algún documento, los errores aumentan a medida que aumenta el tamaño de la empresa. Adicionalmente, la utilidad de los ratios contables en el análisis de fracaso empresarial de pymes se puede apreciar en el hecho de que los modelos desarrollados para este grupo objetivo de empresas alcancen porcentajes de aciertos y errores similares a los conseguidos con el mismo tipo de modelos aplicados a empresas auditadas y cotizadas.

Como señalan Altman y Sabato (2005), el numeroso grupo de empresas pequeñas y medianas proporciona dinamismo a la economía gracias a unas estructuras más simples que les permiten responder con rapidez a los cambios en las condiciones económicas y atender a las necesidades de los consumidores locales. Eso puede inducir a crecimientos extraordinarios en épocas de auge, pero también a rápidos fracasos cuando las condiciones son adversas. En consecuencia, el rendimiento de las pymes resulta más volátil que el de las empresas grandes, dando lugar a dificultades financieras periódicas para un número significativo de pymes (Dannreuther y Kessler, 2010). Entre los trabajos empíricos de fracaso empresarial, Dietsch y Petey (2004) identifican ese mayor riesgo en una muestra de pymes alemanas y francesas, pero también encuentran una menor correlación con otras empresas (para los activos) que la que presentan las empresas grandes. Pompe y Bilderbeek (2005) confirman que la quiebra es

más difícil de predecir en las empresas jóvenes que en empresas establecidas, en un trabajo empírico sobre más de mil empresas quebradas en Bélgica. Estos resultados confirman la necesidad de analizar el fracaso empresarial en las empresas pequeñas y medianas con herramientas específicamente desarrolladas para ellas.

Tres de los elementos tratados en estudios empíricos previos sobre fracaso empresarial resultan determinantes en el diseño de nuestro experimento y su posterior interpretación: (1) el plazo de anticipación en la predicción de la quiebra y las variables que resultan en cada plazo más discriminantes; (2) la consideración de poblaciones heterogéneas que incluyan varios sectores frente al estudio de grupos homogéneos formados por empresas del mismo sector; y (3) el análisis de varias causas de fracaso frente a una única categoría de empresas fracasadas.

En cuanto al primer elemento, el periodo de observación previo a la quiebra, Edmister (1972) aprecia entre sus resultados diferenciales para pymes que es recomendable revisar al menos tres informes anuales consecutivos. Así lo hacen un grupo de autores que consideran en su estudio los datos contables publicados uno y dos años antes (López *et al.*, 1998; Correa *et al.*, 2003) o en los tres ejercicios anteriores a que se produzca el fracaso (Arquero *et al.*, 2008; Gómez *et al.*, 2008). En otros casos, se utilizan modelos de predicción a dos años, ya sea a causa de los datos disponibles (Ferrando y Blanco, 1998) o sin aducir razones para ello (Gallego *et al.*, 1997b; Psillaki *et al.*, 2010). Pero también hay modelos que analizan la información contable contenida en los últimos estados financieros antes del fracaso (Baixauli y Módica-Milo, 2010) o antes de la inscripción de la empresa en el proceso oficial de reorganización para empresas en dificultades (Laitinen, 2008). En el caso de Altman y Sabato (2007), el modelo se desarrolla para predecir la probabilidad de fallido (probability of default, PD) con un año de antelación para que se ajuste al modelo requerido a los bancos por Basilea II.

Debemos destacar el trabajo de Labatut *et al.* (2009), que incorpora el tiempo como variable fundamental en la detección del fracaso empresarial. Estos autores

proponen un modelo exponencial acotado superiormente que, aplicado sobre ratios contables, permite jerarquizarlos según su eficiencia temporal en alertar sobre una posible insolvencia empresarial. De acuerdo con sus resultados, la rentabilidad de la empresa resulta ser la primera variable en detectar problemas de insolvencia futuros, y se configura como el factor más importante de predicción de fracaso, especialmente la rentabilidad relacionada con la inversión financiada con fondos propios y la relacionada con las ventas obtenidas. También la deficiencia en la generación de flujos de tesorería resulta determinante ya que obliga a las empresas a endeudarse excesivamente.

Destacamos también el trabajo de Gómez *et al.* (2008), cuyo objetivo es precisamente comprobar la estabilidad y la validez temporal de los modelos de predicción de fracaso empresarial. Sus resultados indican que, “si bien los modelos resultantes para cada año son distintos, no existen diferencias significativas en los errores obtenidos por cada uno de ellos”. Esto sugiere que la distancia temporal a la quiebra no es un condicionante en la selección del modelo a aplicar, ya que aplicando modelos elaborados para uno, dos o tres años antes del fracaso empresarial, los resultados obtenidos serían similares.

No podemos deducir, pues, de la literatura previa, un plazo de anticipación idóneo en la predicción, pero seguimos la recomendación de Edmister (1972) y analizamos el fracaso de un grupo de empresas a partir de sus datos contables en sus últimas cuentas anuales publicadas, en el ejercicio anterior y en el ejercicio de dos años antes.

Respecto al segundo elemento, la utilización de muestras heterogéneas en cuanto a la actividad de las empresas, frente a la consideración de muestras por sector, Lincoln (1984) encuentra que en Australia hay diferencias sectoriales que hacen recomendable la estimación de modelos separados por sector para una mejor clasificación de las empresas. López *et al.* (1998), en un estudio aplicado a pymes en la Comunidad Valenciana, encuentran que el tipo de sector en que la empresa desarrolla su actividad es un factor determinante en la predicción del fracaso. Por

su parte, Psillaki *et al.* (2010) documentan efectos diferentes por sectores, tanto del apalancamiento como de las oportunidades de crecimiento (intangibles), sobre el fracaso empresarial de las pymes francesas.

En este caso, la evidencia empírica previa es unánime en la recomendación de llevar a cabo estudios por sector para llegar a modelos con mejor capacidad de identificación o predicción de fracaso. Nuestro estudio se centra en un sector concreto de una zona geográfica determinada, lo que asegura iguales condiciones macroeconómicas. Como, además, las empresas objeto de estudio se han seleccionado por tamaño, podemos considerar que se encuentran en similares condiciones de riesgo económico.

Finalmente, el tercer elemento, el análisis de varias causas de fracaso frente a una única categoría de empresas fracasadas, enlaza con la falta de una teoría asentada sobre el fracaso empresarial que defina de forma inequívoca el concepto de fracaso, tal como se trató en el primer apartado de la tesis. La consideración de empresas oficialmente declaradas quebradas o en suspensión de pagos frente a empresas que no lo están es un criterio de separación habitual, lo cual implica que empresas en dificultades o fallidas, de acuerdo con otros criterios menos estrictos, se clasifiquen como “no fracasadas” a pesar de su elevada probabilidad de fracaso, tal como acertadamente señalan Baixauli y Módica-Milo (2010). Cuanto más amplia sea la definición de fracaso, más elevada es la tasa de empresas que quedarán incluidas en esa categoría (Watson y Everett, 1996; Altman *et al.*, 2008). En cambio, empresas oficialmente disueltas pueden no haber fracasado. Por ejemplo, puede haber empresas que se venden por un buen precio o son absorbidas por razones estratégicas (Headd, 2003); también empresas que cambian de propietarios por retiro o enfermedad (Cochran, 1986).

En este sentido, resulta muy interesante el trabajo de Arquero *et al.* (2008), que siguen un procedimiento similar al diagnóstico médico para buscar los síntomas de la empresa con riesgo de fracaso y su evolución. Su análisis confirma que las empresas no se encuentran en situaciones similares ni siguen un proceso común.

Para un porcentaje considerable de su muestra de empresas (casi el 50%) el fracaso puede predecirse con medidas habituales, dado que presentan flujos de fondos negativos que reducen considerablemente los fondos propios hasta que quedan en cifras muy bajas o negativas. Sin embargo, un porcentaje superior al 25% de las empresas sufre una súbita degradación de la cobertura de deudas tras un periodo de inversiones significativas financiadas con pasivos. Finalmente, un grupo más reducido de empresas parecen estar en una situación relativamente estable, aunque la rentabilidad y las coberturas de pasivo sean muy bajas.

Estos resultados son consistentes con la falta de confirmación obtenida por Pompe y Bilderbeek (2005) para su hipótesis sobre el poder predictivo de diferentes categorías de ratios durante las sucesivas fases previas a la quiebra. Estos autores esperan que los ratios se vayan degradando en un orden determinado acorde con el razonamiento económico, esto es, primero los ratios de actividad y de rentabilidad, después los ratios de solvencia y, finalmente, los ratios de liquidez. Pero sus resultados encuentran que no hay un orden fijo en el que las diferentes categorías de ratios financieros empiecen a ser predictivos, ratios que evalúan distintas dimensiones de la posición financiera de la empresa muestran similares niveles de eficacia en la predicción. Entre las causas posibles, apuntan que la hipótesis podría ser válida solo para algunas de las empresas que acaban fracasando. Por ejemplo, una empresa saludable que emprende una inversión de coste elevado, utilizando demasiada deuda, puede enfrentarse a graves problemas cuando la rentabilidad de la inversión es menor de lo esperado, sobre todo si es menor que su coste de financiación. En este posible escenario, el deterioro de la rentabilidad y el de la solvencia se darían de forma simultánea, y la reducción de liquidez no resultaría una buena señal de la futura quiebra.

Algunos autores diferencian dos categorías de fracaso, según que las pymes sean inviables económicamente o que pasen por dificultades financieras transitorias (Franks y Torous, 1992). Se trata de una distinción relevante, porque las empresas que cuentan con capacidad suficiente para generar al menos un rendimiento normal, deberían ser reorganizadas cuando están en peligro de quiebra, o ser objeto de rehabilitación si ya hubieran quebrado (Cook *et al.*, 2012). Dos

interesantes trabajos sobre reorganización o recuperación de pymes son los de Laitinen (2008) y Cook *et al.* (2012). El primero muestra la mejora en la capacidad de los modelos para identificar las empresas con posibilidades de evitar el fracaso tras un proceso de reorganización, cuando se utiliza información no financiera tanto cualitativa como cuantitativa, además de la información contable financiera. El segundo muestra que la fortaleza en recursos ayuda en la tarea de identificar las empresas económicamente viables entre las pymes quebradas.

Nuestro trabajo tiene en cuenta los resultados previos sobre este tercer factor en varios sentidos. Por un lado, somos conscientes de que nuestra definición de fracaso (que toma empresas en situación concursal, disueltas y extinguidas) incluye en la categoría de no fracasadas empresas que pueden atravesar situaciones financieras desfavorables e incluso críticas; mientras que alguna de las empresas disueltas o extinguidas puede haber llegado a esa situación en buenas condiciones financieras, incluso de rentabilidad. Por otro lado, esa es la razón de que en nuestra metodología se establezcan tres medidas de referencia: valores de las empresas fracasadas, valores centrales (medianas) de la población y valores espejo (para representar a las empresas más saludables).

En cuanto a la aplicación de *scores* al análisis de fracaso empresarial en pymes, podemos mencionar como antecedente el trabajo de Baixauli y Mónica-Milo (2010). Para reducir la heterogeneidad de las empresas calificadas como saludables, estos autores construyen un indicador de fortaleza financiera³⁶, que permita identificar una muestra de empresas fuertes que se opongan a las empresas fallidas, con el fin de obtener modelos más precisos de predicción de quiebra. En palabras de los autores, se trata de un *score* para aproximar “la probabilidad de quiebra de las empresas saludables”. Está formado a partir de cuatro medidas de rentabilidad (rendimiento sobre activo, rotación, rendimiento sobre fondos propios y cambio del valor neto anual) y una de calidad de la información contable, basada en la opinión del auditor externo.

³⁶ El indicador de Baixauli y Mónica-Milo (2010, p. 63) se formula: $FS = \sum_{j=1}^4 I(FR_j \geq P_5) + d_5$, donde FR es el ratio financiero, P_5 es el quinto percentil, d_5 es una variable dicotómica que toma el valor 1 si la opinión del auditor es favorable y 0 en caso contrario, y $I(.)$ es la función indicador, que toma el valor 1 si FR es mayor o igual que P_5 y 0 en caso contrario.

A diferencia del *score* propuesto en Baixauli y Mónica-Milo (2010), nuestro trabajo plantea diferencias de percentiles respecto a la población³⁷ y toma los valores de los percentiles como variables continuas, lo cual nos permite obtener puntuaciones con mayor capacidad de discriminación entre grupos o entre empresas individuales. Otra diferencia es que Baixauli y Mónica-Milo (2010) solo toman un punto de referencia fijo en los percentiles (el percentil 5), mientras que en el presente trabajo nosotros adoptamos tres puntos de referencia: el percentil de los valores centrales de las empresas quebradas, el percentil central de la población y el percentil espejo frente al de las empresas quebradas. De este modo, nuestro modelo es flexible ante las condiciones del entorno en el periodo de estudio: en qué niveles están fracasando las empresas determina en qué niveles podemos considerar que las empresas no fracasadas se pueden considerar más saludables. Por ejemplo, si las empresas quiebran en percentiles de los indicadores financieros muy alejados de la mediana, la situación económica de la población (país, región, sector,..) será mejor y una empresa cualquiera obtendrá buenos resultados con poco esfuerzo competitivo. En este entorno, tiene sentido elegir a las empresas más saludables, alejadas de la mediana en sentido contrario, esto es, las que mejores puntuaciones consiguen en cada ratio.

3.3. Propuesta metodológica

Tal como señalan Labatut *et al.* (2009) la información contable de la pyme española permite elaborar una gran batería de ratios financieros que abarca las diversas áreas de la realidad empresarial. También señalan que algunos valores de estos ratios contables resultan incompatibles con una situación desahogada de la empresa, lo que les confiere cierto carácter de control sobre la solvencia de la empresa. De ahí que tenga sentido plantearse cuál de ellos es más eficiente en la labor de detección del fracaso empresarial.

³⁷ Como hemos visto, ellos toman percentiles dentro del grupo de las no quebradas.

Al igual que estos autores (Labatut *et al.*, 2009), pretendemos ver qué ratios explican mejor que otros la insolvencia, y la metodología que proponemos permite identificar cuáles nos alertan con mayor anticipación. Pero nuestro principal objetivo avanza un poco más en esta línea de investigación, para calcular *scores* que nos indiquen la cercanía o lejanía a valores de fracaso y, una vez determinada la distancia al fracaso (en esa población y momento) nos permite sugerir a las empresas en peligro qué ratios financieros concretos podrían mejorar para alcanzar los niveles medios de las empresas saludables.

Para describir nuestra metodología, empezamos por distinguir dos grupos dentro de una población (o muestra) G de empresas. G_s incluye las empresas no fracasadas, cuyo número total es n_s , y G_q las empresas fracasadas, cuyo número total es n_q . La población G está compuesta por n empresas ($i = 1, \dots, n$), siendo $n = n_s + n_q$. Definimos ahora un grupo f de indicadores financieros ($f = 1, \dots, h$), que se han seleccionado por su capacidad para discriminar entre empresas fracasadas y no fracasadas, a partir de la literatura empírica previa, en la primera parte de esta tesis. Cada indicador financiero puede ser un ratio financiero, como en nuestro análisis empírico, pero también podría ser otra modalidad de medida financiera, para la que los percentiles resulten ser un buen descriptor estadístico.

El primer paso de nuestro análisis empírico consiste en describir cómo se distribuyen los indicadores financieros seleccionados en la población de estudio (zona geográfica, sector y periodo concretos), calculando y representando las distribuciones de frecuencias de cada uno de los indicadores, tanto para el total de la población (una por año), G , como para las empresas fracasadas cada año, G_q . Para cada indicador financiero, se trata de identificar los puntos en la distribución que permitan situar las empresas fracasadas dentro del total, esto es, qué valores del indicador dan fracaso y qué percentiles de la distribución de la población de ese periodo representan.

El valor de una empresa cualquiera para un indicador financiero concreto se representa como x_{if} , y el valor de una empresa fracasada para cada indicador

financiero como z_{if} . Definimos los vectores columna que incluyen los valores para el f -simo indicador financiero: X_f , para el conjunto de empresas, Y_f , para empresas no fracasadas, y Z_f , para empresas fracasadas, de manera que $X_f = (Y_f \cap Z_f)$.

Calculamos el valor mediano \tilde{z}_f del vector Z_f y el percentil del valor \tilde{z}_f sobre el conjunto de valores de cada indicador financiero en el conjunto de la población (o muestra) considerada, $\pi(q)_f = p(\tilde{z}_f, X_f) \in [0,1]$. De manera similar, especificamos la posición de una empresa individual i para el indicador financiero f en la población $\pi_{if} = p(x_{if}, X_f) \in [0,1]$.

La distancia entre percentiles para una empresa i y para un indicador financiero f , d_{if} , se define por diferencia entre el percentil que el valor del indicador de la empresa ocupa dentro del conjunto de valores de toda la población para ese indicador y 0,5 que es el percentil de la mediana de la población para ese indicador. Hemos adoptado valores medianos en lugar de medios para representar la población G , y el grupo de empresas fracasadas G_q , en atención a la dispersión de los valores de los indicadores financieros en las empresas pequeñas que constituyen nuestro objeto de estudio. La mayor representatividad de los valores medianos sobre los medios ha quedado claramente demostrada en el estudio empírico llevado a cabo en la segunda parte de esta tesis.

$$\begin{cases} \text{Si } \pi(q)_f < 0,5 ; & d_{if} = 0,5 - \pi_{if} \\ \text{Si } \pi(q)_f > 0,5 ; & d_{if} = \pi_{if} - 0,5 \end{cases}$$

d_{if} es igual a 0 cuando la empresa tiene un valor para el indicador financiero f igual al valor mediano del indicador en la población global. La distancia es positiva (negativa) cuando el percentil para el grupo de fracaso es inferior a 0,5 y eso indica un mayor (menor) riesgo de quiebra; o bien, cuando el percentil es superior a 0,5, y eso sitúa a la empresa más (menos) próxima a la quiebra. La distancia para cada indicador financiero está incluida en el intervalo $[-0,5, 0,5]$.

De manera similar, podemos calcular la distancia entre percentiles para un grupo dentro de la población. En nuestro caso, el grupo de interés es G_q , el subconjunto de empresas fracasadas. De este modo, podemos medir de forma homogénea para cualquier indicador financiero la capacidad discriminante entre empresas fracasadas y no fracasadas de ese indicador. Conseguir una medida homogénea resulta especialmente importante en variables como los ratios financieros, que resultan ser variables con rangos y distribuciones muy diferentes³⁸.

$$\begin{cases} \text{Si } \pi(q)_f < 0,5 ; & d(q)_f = 0,5 - \pi(q)_f \\ \text{Si } \pi(q)_f > 0,5 ; & d(q)_f = \pi(q)_f - 0,5 \end{cases}$$

Cuanto mayor sea la distancia entre percentiles, $d(q)_f$, del indicador financiero f -simo en el grupo de empresas fracasadas, G_q , sobre la población considerada, G , mayor es su capacidad discriminante.

De acuerdo con Labatut *et al.* (2009), que un indicador financiero concreto tome un valor alejado de una situación empresarial desahogada no significa que la empresa esté próxima al fracaso. Para poder confirmar esa proximidad al fracaso, tiene que ocurrir que varios indicadores financieros se desvíen de la situación normal de salud financiera. En nuestro modelo, construimos un *score* sumando las diferencias en percentiles de todos aquellos indicadores financieros que se muestran discriminantes para el grupo de empresas fracasadas respecto a la población total. Esto es, el *score* va a estar compuesto por las mayores distancias, $d(q)_f$, que representan a los indicadores financieros con poder discriminante para analizar la proximidad al fracaso. En este caso, el valor del *score* se encuentra en el intervalo $[-0,5 \times h, 0,5 \times h]$, siendo h el número de indicadores financieros. De manera que para 10 indicadores financieros el *score* del grupo de empresas fracasadas (q) resultará un valor entre -5 y 5.

³⁸ En la segunda parte de esta tesis se ha mostrado como ninguno de los 10 ratios seleccionados presentan una distribución normal en la muestra de estudio; tampoco presentan normalidad multivariante conjunta; y tras un proceso de normalización, las variables siguen sin ser normales (ver también histogramas de las variables en el Anexo VI).

$$S(h, q) = \sum_{f=1}^h d(q)_f$$

Es posible seleccionar un número inferior h_1 de indicadores, cuando varios de los elegidos resulten poco discriminantes en la población y periodo de estudio. Por ejemplo, en nuestro análisis empírico posterior se aprecia como 5 de los 10 ratios financieros resultan más discriminantes en los tres años de estudio, por lo que se calcula un *score* con 10 ratios y otro con 5 ratios. El *score* calculado para un número de indicadores financieros inferior al total ($h_1 < h$) del grupo de empresas fracasadas (q) se formularía como:

$$S(h_1, q) = \sum_{f=1}^{h_1} d(q)_f$$

Una vez determinados los indicadores financieros que resultan más discriminantes en la población de estudio para distinguir las empresas fracasadas, es posible aplicar el/los *score/s* a grupos más pequeños de empresas dentro de la población (por tamaño, sector, subsector, etc.) y también a empresas individuales. Así, el *score* con el número total de indicadores financieros considerados (h), para una empresa individual (i), se calcularía del modo siguiente:

$$S(h)_i = \sum_{f=1}^h d_{if}$$

Para un número menor de indicadores financieros, h_1 , el *score* para una empresa individual (i) sería:

$$S(h_1)_i = \sum_{f=1}^{h_1} d_{if}$$

Para evitar el efecto tamaño derivado del número de indicadores financieros utilizados, es posible calcular una distancia media de percentiles en cualquiera de los casos anteriores, simplemente dividiendo por el número de indicadores. Para el grupo G_q de empresas fracasadas, el valor del *score* medio se encuentra en el intervalo $[0, 0,5]$. Para cualquier otro grupo, o para una empresa individual, el *score* medio para cualquier número de indicadores financieros estará en el intervalo $[-0,5, 0,5]$. Así, por ejemplo, se muestran a continuación la fórmula del *score* medio, calculado con el total de indicadores financieros considerados (h) para el grupo de empresas fracasadas (q), $\bar{S}(h, q)$; y el *score* medio, calculado con un número concreto de indicadores inferior al total, de una empresa individual (i), $\bar{S}(h_1)_i$.

$$\bar{S}(h, q) = \frac{\sum_{f=1}^h d(q)_f}{h};$$

$$\bar{S}(h_1)_i = \frac{\sum_{f=1}^{h_1} d_{if}}{h_1}$$

Es fácil deducir que los indicadores financieros con poco poder discriminante, esto es, con distancias próximas a 0, tendrán contribuciones pequeñas al *score* de las empresas fracasadas, $S(h, q)$, y a los *scores* de las empresas individuales, $S(h)_i$. Y también que, si se utilizan *scores* medios (\bar{S}) los indicadores financieros poco discriminantes reducirán el valor total, siendo aquí más interesante elegir un número de indicadores inferior, para poder comparar grupos y empresas.

Si se quiere dar mayor importancia a los indicadores más discriminantes y menos a los menos discriminantes, se puede utilizar un *score* medio ponderado por la distancia para ese indicador entre el grupo de fracasadas y el total, $d(q)_f$, que será siempre un valor positivo, por definición. Así, por ejemplo, se muestran a continuación la fórmula del *score* medio ponderado, calculado con el total de indicadores financieros (h) para el grupo de empresas fracasadas (q), $\bar{S}w(h, q)$; y el *score* medio ponderado calculado con un número concreto de indicadores inferior al total (h_1) para una empresa individual (i), $\bar{S}w(h_1)_i$.

$$\overline{Sw}(h, q) = \frac{\sum_{f=1}^h d(q)_f \cdot d(q)_f}{S(h, q)};$$

$$\overline{Sw}(h_1)_i = \frac{\sum_{f=1}^{h_1} d_{if} \cdot d(q)_f}{S(h_1, q)}$$

Finalmente, el tercer punto de referencia corresponde a los valores espejo de los percentiles, que nos permiten tomar como referencia valores mejores que la mediana de la población para cada indicador financiero, en atención a la naturaleza heterogénea de las empresas clasificadas como no fracasadas. Estamos hablando, para cada indicador financiero, de valores de los percentiles diametralmente opuestos a los valores medianos obtenidos por las empresas fracasadas, \tilde{z}_f . O sea, que los percentiles espejo serían los que se encuentran a igual distancia de la mediana en sentido contrario.

$$\pi(esp) = 1 - \pi(q)_f \in [0, 1]$$

Así, para cualquiera de las modalidades de *score* propuestas, obtenemos tres valores de referencia a efectos de comparación:

- El *score* calculado con la mediana para cada indicador financiero del grupo de empresas fracasadas, G_q .
- El *score* calculado con la mediana para cada indicador financiero de la población total, G , que incluye tanto empresas fracasadas como no fracasadas;
- El *score* calculado con los percentiles espejo de los valores medianos del grupo de empresas fracasadas. Esto nos da una referencia de empresas no fracasadas en mejor situación que la media. Si sus valores para cada ratio están en el percentil opuesto al de las empresas fracasadas, podemos atribuirles una buena situación financiera.

De este modo, cuando se trata de analizar la posición de un subgrupo de la población de empresas, o de una empresa individual, el *score* obtenido se

compara con el *score* del grupo de empresas fracasadas, $S(h, q)$, con el *score* de la población total, $S(h, g)$, y con el *score* de los valores espejo, $S(h, esp)$, que representa una referencia sobre empresas saludables más adecuada que G_s . Podemos situar a la empresa en varios niveles de proximidad al fracaso o a situaciones de salud financiera (recordemos que los *scores* están formados por distancias, cuanto mayor es el *score* en valores positivos mayor es el riesgo de fracaso de la empresa).

1. La empresa tiene un riesgo de fracaso muy elevado, mayor cuanto más alto sea el valor del *score*, $S(h)_i \geq S(h, q)$;
2. La empresa está expuesta a un riesgo de fracaso de moderado a elevado, mayor cuanto más alto sea el valor del *score*,

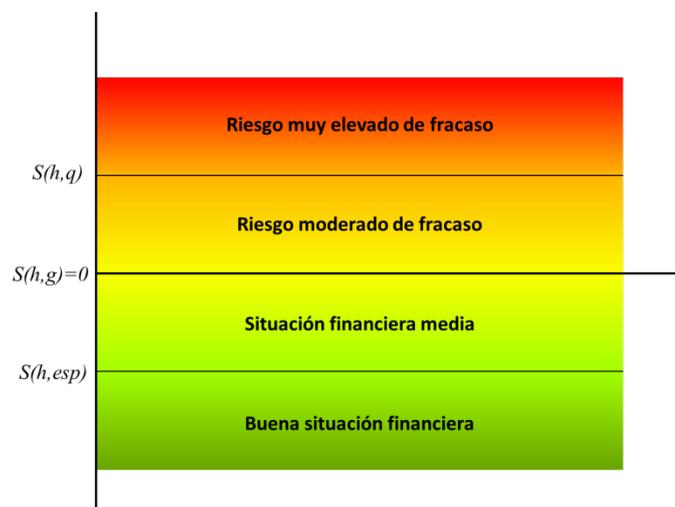
$$S(h, q) > S(h)_i \geq S(h, g) = 0;$$

3. La empresa se encuentra en una situación financiera media, mejor cuando el valor de los *scores* alcanza valores negativos más elevados,

$$S(h, g) = 0 > S(h)_i \geq S(h, esp);$$

4. La empresa se encuentra en una buena situación financiera, mejor cuanto más altos son los valores negativos de los *scores*, $S(h)_i < S(h, esp)$.

Gráfico 3. Niveles de scores



Para cada modalidad de *score* propuesto se pueden calcular distancias al fracaso, principal objetivo de esta parte de la tesis, aunque también es posible hallar distancias a los valores centrales (medianas) de la población y a los valores espejo, que indican valores representativos de empresas con mejor salud financiera dentro del grupo de no fracasadas. Así, la distancia al fracaso de un grupo, G_a , (que puede incluir empresas fracasadas y/o no fracasadas) perteneciente a la población en estudio, G , se define por la diferencia entre los *scores* de la misma categoría, esto es, por la diferencia entre el *score* del grupo G_a menos el *score* del grupo de empresas fracasadas G_q . Para dar coherencia al resultado de la distancia con la interpretación económica, aplicamos un cambio de signo, de manera que la distancia al fracaso es negativa cuando el grupo que se analiza está en peor situación que el grupo de referencia (grupo de fracasadas); y la distancia al fracaso es positiva cuando el grupo analizado está en mejor situación que el grupo de empresas fracasadas.

- La distancia al fracaso, D , con el número total de indicadores financieros (h) para el grupo G_a , será:

$$D(h, a) = -(S(h, a) - S(h, q)) = \sum_{f=1}^h d(q)_f - \sum_{f=1}^h d(a)_f = \sum_{f=1}^h (d(q)_f - d(a)_f)$$

- La distancia media al fracaso, \bar{D} , con el número total de indicadores financieros (h) para el grupo G_a , será:

$$\bar{D}(h, a) = -(\bar{S}(h, a) - \bar{S}(h, q)) = \frac{\sum_{f=1}^h d(q)_f}{h} - \frac{\sum_{f=1}^h d(a)_f}{h} = \frac{\sum_{f=1}^h (d(q)_f - d(a)_f)}{h}$$

- La distancia media ponderada al fracaso, \overline{Dw} , con el número total de indicadores financieros (h) para el grupo G_a , será:

$$\overline{Dw}(h, a) = -(\overline{Sw}(h, a) - \overline{Sw}(h, q)) = \frac{\sum_{f=1}^h d(q)_f \cdot d(q)_f}{S(h, q)} - \frac{\sum_{f=1}^h d(a)_f \cdot d(a)_f}{S(h, q)} = \frac{\sum_{f=1}^h (d(q)_f - d(a)_f) \cdot d(q)_f}{S(h, q)}$$

Para un número inferior de indicadores financieros, $h_1 < h$, las fórmulas serían las mismas, pero sustituyendo h por h_1 .

Si se trata de hallar distancias al fracaso para empresas individuales pertenecientes a la población G , a continuación se muestra la formulación para los

tres tipos de *scores* propuestos (S , \bar{S} y \overline{Sw}) y el número total de indicadores considerados (h).

- La distancia al fracaso, D , con el número total de indicadores financieros (h) para la empresa i , será:

$$D(h)_i = -(S(h)_i - S(h, q)) = \sum_{f=1}^h d(q)_f - \sum_{f=1}^h d_{if} = \sum_{f=1}^h (d(q)_f - d_{if})$$

- La distancia media al fracaso, \bar{D} , con el número total de indicadores financieros (h) para la empresa i , será:

$$\bar{D}(h)_i = -(\bar{S}(h)_i - \bar{S}(h, q)) = \frac{\sum_{f=1}^h d(q)_f}{h} - \frac{\sum_{f=1}^h d_{if}}{h} = \frac{\sum_{f=1}^h (d(q)_f - d_{if})}{h}$$

- La distancia media ponderada al fracaso, \overline{Dw} , con el número total de indicadores financieros (h) para la empresa i , será:

$$\overline{Dw}(h)_i = -(\overline{Sw}(h)_i - \overline{Sw}(h, q)) = \frac{\sum_{f=1}^h d(q)_f \cdot d(q)_f}{S(h, q)} - \frac{\sum_{f=1}^h d_{if} \cdot d(q)_f}{S(h, q)} = \frac{\sum_{f=1}^h (d(q)_f - d_{if}) \cdot d(q)_f}{S(h, q)}$$

De nuevo, la formulación para un número inferior de indicadores financieros, $h_1 < h$, es la misma, pero sustituyendo h por h_1 .

Una vez formuladas las distancias al fracaso, podemos reformular la posición de una empresa individual (o bien de un subgrupo cualquiera dentro de la población).

1. La empresa tiene un riesgo de fracaso muy elevado cuando la distancia es inferior a la del grupo de fracasadas (0). El riesgo es mayor cuanto más negativa sea la distancia al fracaso, $D(h)_i < D(h, q) = 0$;
2. La empresa está expuesta a un riesgo de fracaso de moderado a elevado cuando la distancia al fracaso se encuentra entre cero y la distancia de la población en su conjunto. El riesgo de fracaso es mayor cuanto menor sea la distancia al fracaso, $D(h, q) = 0 \leq D(h)_i < D(h, g)$;
3. La empresa se encuentra en una situación financiera media cuando la distancia al fracaso es mayor que la distancia de la población global aunque menor que la distancia espejo (doble distancia). El riesgo de

fracaso es menor cuando la distancia al fracaso alcanza valores positivos más altos, $D(h, g) \leq D(h)_i < D(h, esp)$;

4. La empresa se encuentra en una buena situación financiera cuando la distancia al fracaso es mayor que la distancia espejo. La situación es mejor, con menor riesgo de fracaso, cuanto más altos son los valores positivos de la distancia $D(h)_i > D(h, esp)$.

Gráfico 4. Niveles de distancias al fracaso



3.4. Estudio empírico: una aplicación al sector de la construcción

Como en la segunda parte, en esta tercera parte de la tesis, hemos seleccionado un sector concreto para probar la metodología: empresas pequeñas del sector de la construcción en la Comunidad Autónoma de Castilla y León. En cuanto al periodo temporal, se utiliza información de los años 2006 a 2010. La fuente utilizada es la base de datos proporcionada por *Iberinform* sobre información contable y sobre situación de fracaso empresarial (empresas en situación concursal, disueltas y extinguidas). Aplicamos la metodología propuesta en dos fases, un análisis del total de empresas por periodo y un análisis de evolución de un grupo determinado de empresas durante varios periodos.

En la primera fase, se hacen análisis anuales en 2006, 2007 y 2008 de todas las empresas fracasadas, dentro del conjunto de todas las empresas que contiene la base de datos en el sector analizado, en la Comunidad Autónoma de Castilla y León. El análisis anual muestra la capacidad de la metodología propuesta para seleccionar los ratios más discriminantes; caracterizar la población y el grupo de empresas fracasadas por sus distancias al fracaso, a las posiciones centrales de la población en estudio, y a un grupo de referencia de empresas con mejor salud financiera dentro de la población; trazar las tres fronteras de ratios mencionadas (fracaso, valores centrales de la población, empresas más saludables); caracterizar empresas individuales por su distancia a las fronteras anteriores; y apreciar la evolución de las fronteras año a año.

Para la selección de las empresas se establecieron las mismas restricciones que para el estudio empírico del apartado 2.4. de esta tesis: La población total está formada por 4.263 empresas en 2006, 5.037 empresas en 2007 y 3.427 empresas en 2008. Dado que utilizamos una serie de ratios en los cálculos, hemos eliminado aquellas empresas con valores indefinidos en esos ratios. De este modo, la población de empresas para la que se podían calcular las variables ha quedado formada por 4.143 empresas en 2006, 4.826 en 2007 y 3.230 en 2008.

En la segunda fase, se hace un análisis de seguimiento, durante los años 2008, 2009 y 2010, del grupo de empresas constructoras españolas que fracasan en 2010. Con este segundo tipo de análisis, se muestra que la metodología es aplicable a la predicción de fracaso, a la identificación de los ratios que dan las primeras señales de alarma y también como herramienta para orientar la reorganización de empresas en riesgo de fracaso. La población de empresas que existen en la base de datos durante los tres años 2008, 2009 y 2010, sin aparecer como fracasadas en ninguno de los tres periodos, es de 11.861; y las empresas fracasadas en 2010, que existían en los dos periodos anteriores, son 28.

Por un lado, utilizamos estas 28 empresas fracasadas en 2010 como muestra de control del análisis realizado en 2006-08. Por otro lado, estudiamos cómo evolucionan los *scores* para un mismo grupo de empresas en los periodos previos al fracaso. Finalmente, vemos si nuestros resultados sobre los ratios que se muestran más discriminantes en 2008, 2009 y 2010 son consistentes con la literatura previa sobre señales tempranas de alarma y sobre orden en el deterioro de los ratios.

En cuanto a las variables que nos permitirán analizar la situación de cada empresa, utilizamos los 10 ratios financieros seleccionados en la primera parte de la tesis a partir de una amplia muestra de trabajos empíricos nacionales e internacionales y del razonamiento económico. En la segunda parte de la tesis se ha comprobado, en un análisis de literatura previa sobre pymes, que la selección de variables no cambia significativamente según el tamaño de las empresas de la muestra.

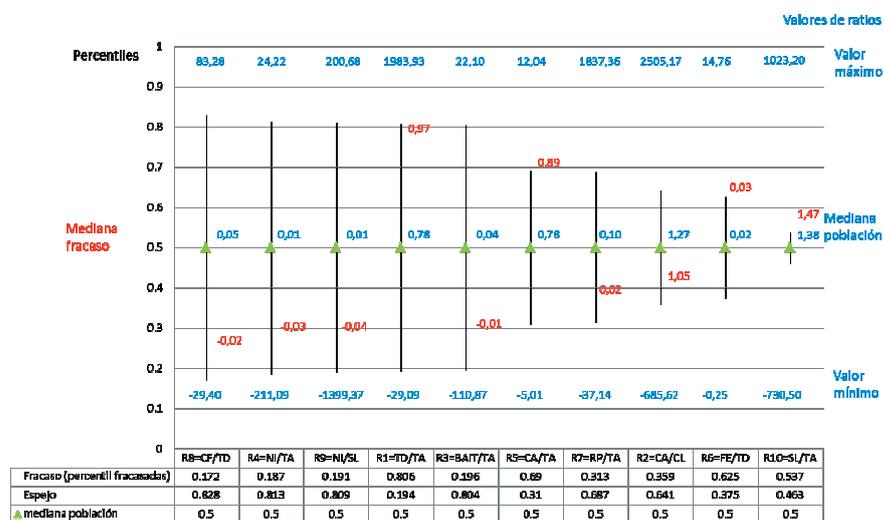
3.4.1. La elección de los ratios más discriminantes y la frontera gráfica de quiebra

Como se ha explicado previamente, los indicadores con mayor poder de discriminación entre empresas fracasadas y no fracasadas son los que reflejan mayor diferencia, d_f , entre los percentiles del valor mediano de cada ratio en empresas fracasadas, $\pi(q)_f$, y del total³⁹ de empresas, $\pi(g)_f$. Así, podemos trazar la frontera de quiebra de la población, con los ratios ordenados por percentil de quebradas más alejado de la mediana, y también la frontera espejo, $\pi(esp)_f$, de las mejores empresas para esos ratios⁴⁰. Los Gráficos 5, 6 y 7 muestran, en orden de mayor a menor poder discriminante, los 10 ratios utilizados, a efectos de identificación de las empresas quebradas dentro de la población de estudio. En la

³⁹ Nótese que $\pi(g)_f = 0,5$.

⁴⁰ Los valores espejo tienen utilidad en primer lugar para representar el gráfico, de manera que sea más fácil visualizarlo, ya que unos ratios tienen valores de fracaso altos (por ejemplo el endeudamiento) y otros ratios tienen valores de fracaso bajos (por ejemplo la rentabilidad). Además, construir una referencia con empresas que obtengan un *score* igual o superior a los valores espejo resulta consistente con los resultados de Baixauli y Mónica-Milo (2010), que obtienen mejores predicciones con modelos calibrados a partir de empresas quebradas o en suspensión de pagos frente a empresas fuertes identificadas por un *score* basado en varias medidas de rentabilidad.

Gráfico 7. Fracaso empresarial. Empresas constructoras de Castilla y León. 2008



Como se apuntaba en la introducción de esta tercera parte de la tesis, la aplicación empírica muestra cada año los ratios más discriminantes, que luego pueden utilizarse en distintas metodologías de identificación y predicción de fracaso (como por ejemplo las utilizadas en la segunda parte de esta tesis: análisis discriminante lineal, análisis discriminante cuadrático, análisis discriminante Knn, logit, probit y DEA). Nuestros resultados muestran que esos ratios más discriminantes, utilizando diferencias de percentiles son: en 2006, R8, R1, R7, R4, seguidos de R5, R6, R9; en 2007, R1, R8, R9, R4, seguidos de R5, R7, R6; y en 2008, R8, R4, R9, R1, R3 seguidos de R5, R7. La selección de variables coincide con la realizada en la segunda parte de la tesis utilizando la metodología *Rank Sum Test*. La única diferencia es que el *Rank Sum Test* se aplica para comparar el grupo de fracasadas con el de no fracasadas, mientras que la metodología propuesta en esta tercera parte de la tesis compara las fracasadas con el total de la población donde éstas están contenidas. Incluimos también en el Anexo VII la selección de ratios hecha con diferencias de percentiles, lo que ha permitido distinguir qué variables debían considerarse como *inputs* y cuáles como *outputs*, a efectos de la aplicación de la metodología DEA.

En segundo lugar, se aprecia la evolución en el tiempo de las fronteras de ratios. Las líneas verticales indican las distancias desde la mediana de la población hasta el percentil de fracaso y en sentido contrario hasta el percentil espejo. En rojo, se indica en las figuras el valor mediano para cada ratio en el grupo de empresas fracasadas. En azul, se indican los valores mínimo, mediano y máximo para cada ratio. En la tabla inferior de cada año, se muestran los valores de los percentiles para las empresas fracasadas, de los percentiles espejo y el percentil de la población (0,5).

En la población estudiada, los valores son similares en 2006 y 2007, pero hay un notable incremento de las distancias entre percentiles en 2008 (excepto para el ratio 10) y un empeoramiento de los valores medianos de los ratios de la población, entendiendo por empeoramiento: reducción cuando menor valor implica más cercanía al fracaso; e incremento cuando mayor valor significa más cercanía al fracaso. Este resultado es consistente con la evolución de este sector en nuestro país en el periodo de estudio (Gill de Albornoz y Giner, 2010).

3.4.2. Distancia a una frontera de quiebra calculada con scores de ratios financieros

El *score*, S , o distancia a la frontera de quiebra, se calcula sumando las diferencias de percentiles en todos los ratios considerados (h), o en aquellos ratios cuyo percentil para el valor mediano en empresas fracasadas es significativamente diferente de 0,5. Es decir, que hallamos los *scores* por acumulación de distancias en los ratios seleccionados. Para que sean *scores* comparables, entre unas empresas y otras o entre unos grupos y otros, el número de ratios seleccionados debe ser el mismo.

Un segundo *score*, \bar{S} , más fácil de interpretar, es la distancia media al fracaso, puesto que nos da un número que es una distancia entre percentiles de los valores medios de los ratios seleccionados en las dos poblaciones que se comparan, en nuestro caso, fracasadas frente al total.

Pero teniendo en cuenta que hablamos de poder discriminante diferente en cada ratio, la empresa estará más lejos o más cerca de la quiebra si los ratios más discriminantes así lo indican. Que una empresa tenga una gran distancia a los valores de fracaso en un ratio poco discriminante no debe tener la misma importancia en la calificación de la empresa como si la distancia en percentiles se da para un ratio con mayor poder discriminante. Por eso, proponemos el cálculo de una distancia media ponderada al fracaso, usando como ponderación la importancia que el ratio tiene en la población y periodo estudiados⁴¹.

En nuestro ejemplo, para 2008, los cinco ratios más discriminantes han sido: R8= (CF/TD), R4= (NI/TA), R9= (NI/SL), R1= (TD/TA) y R3= (BAIT/TA). De modo que calculada la frontera con 5 ratios, se puede dar un valor a esa frontera con *scores*:

- El *score* de distancia de la población al fracaso sería $S(5, q)$: $(0,5-0,172) + (0,5-0,187) + (0,5-0,191) + (0,5-0,194) + (0,5-0,196) = 1,5600$;
- El *score* de distancia media de la población al fracaso sería $\bar{S}(5, q)$: $1,56/5 = 0,3120$; y
- El *score* de distancia media ponderada de la población al fracaso sería $\overline{Sw}(5, q)$: $(0,328^2 + 0,313^2 + 0,309^2 + 0,306^2 + 0,304^2) / 1,56 = 0,3122$.

Calculada la frontera con los 10 ratios, la distancia en términos absolutos es mayor, pero dado el criterio de inclusión de los ratios, si se halla el valor medio, la distancia resulta menor⁴². La mecánica de selección de ratios, por la que los ratios añadidos son cada vez menos discriminantes hace que tomar más factores de la empresa, representados por ratios, pueda estar introduciendo más ruido en la

⁴¹ La ponderación es positiva por definición, porque así están definidas las distancias para todos los ratios entre el grupo de fracasadas y el total. Pero cuando los *scores* se calculan para empresas individuales u otros grupos dentro de G , es posible que se den distancias negativas y *scores* negativos. En estos casos, al ser la ponderación siempre positiva, no anula el efecto de los signos negativos.

⁴² Nos referimos al caso de comparar el grupo de empresas fracasadas con el total. Si estuviéramos comparando dos grupos de empresas dentro de G , o dos empresas concretas, podrían darse todo tipo de casos.

clasificación que capacidad para discriminar. Por eso resulta útil calcular la media ponderada.

En nuestro ejemplo, para 2008, con los 10 ratios por orden de capacidad discriminante, R8, R4, R9, R1, R3, R5, R7, R2, R6 y R10,

- El *score* de distancia de la población al fracaso es $S(10, q)$: $(0,5-0,172) + (0,5-0,187) + (0,5-0,191) + (0,5-0,194) + (0,5-0,196) + (0,5-0,31) + (0,5-0,313) + (0,5-0,359) + (0,5-0,375) + (0,5-0,463) = 2,2400$
- El *score* de distancia media de la población al fracaso queda $\bar{S}(10, q)$: $2,24/10 = 0,2240$
- Y el *score* de distancia media ponderada de la población al fracaso, $\overline{Sw}(10, q)$: $(0,328^2 + 0,313^2 + 0,309^2 + 0,306^2 + 0,304^2 + 0,19^2 + 0,187^2 + 0,141^2 + 0,125^2 + 0,037^2) / 2,24 = 0,2656$.

Comparando los valores de frontera, obtenidos con los *scores* para el grupo de empresas fracasadas, en cada uno de los tres años, se constata lo que habíamos visto gráficamente, que las distancias al fracaso son similares entre 2006 y 2007, mientras que en 2008 las distancias se amplían considerablemente. Se verifica, además, que los resultados son consistentes utilizando cualquiera de los *scores* propuestos. Calculando distancias de la población al fracaso mediante *scores* es posible ser más precisos en la comparación. Apreciamos como la distancia al fracaso se reduce ligeramente de 2006 a 2007, pero se incrementa aproximadamente en un 100% de 2007 a 2008.

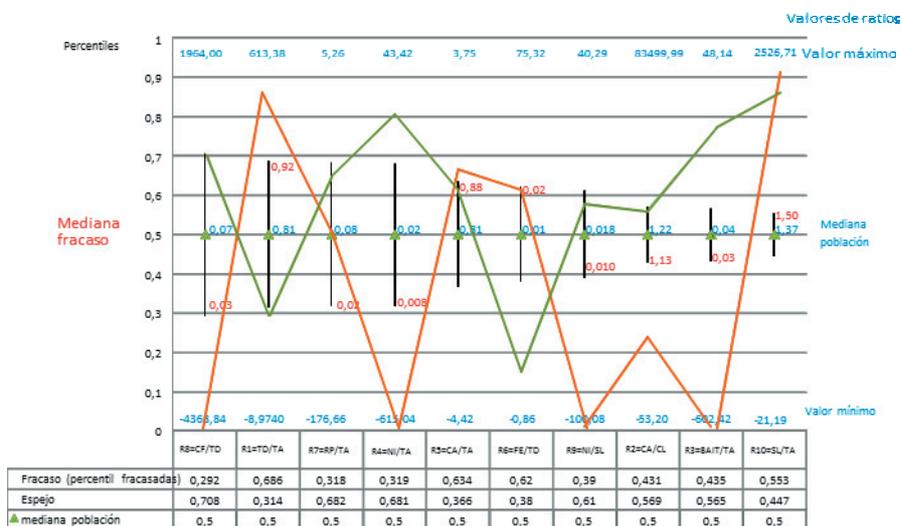
Tabla 24. Scores del grupo de empresas fracasadas con 5 y 10 ratios en 2006, 2007 y 2008.

Scores	2006	2007	2008
$S(5, q)$	0,8910	0,7690	1,5600
$S(10, q)$	1,3080	1,2260	2,2400
$S(5, q)$	0,1782	0,1538	0,3120
$S(10, q)$	0,1308	0,1226	0,2240
$\overline{Sw}(5, q)$	0,1814	0,1549	0,3122
$\overline{Sw}(10, q)$	0,1529	0,1333	0,2656

3.4.3. Situación relativa de empresas concretas

Una vez trazadas para una población la frontera de fracaso y la frontera espejo, por encima y por debajo de los valores medianos de los ratios en esa población, es posible situar cualquier empresa (o grupo de empresas) que se quieran analizar. Se pueden calcular y representar gráficamente las distancias para cada ratio a las fronteras. Además, se puede calcular numéricamente la distancia de la empresa a cada una de esas fronteras, mediante los *scores*, para caracterizar la empresa (o grupo de empresas) dentro de su población. En este apartado hemos seleccionado una empresa fracasada y una empresa no fracasada en 2006 para mostrar el potencial de caracterización dentro de la población.

Gráfico 8. Fracaso empresarial. Empresas constructoras de Castilla y León. 2006. Caracterización de empresas concretas



En el Gráfico 8, la línea naranja une los valores de los percentiles de cada ratio calculado para una empresa en situación de fracaso, en tanto que la línea verde une los valores de los percentiles de los ratios para una empresa no fracasada. Se puede ver gráficamente la distancia en percentiles al valor mediano de empresas fracasadas en ese periodo y población, lo cual resulta especialmente interesante para empresas fracasadas. También se aprecia la distancia a la mediana del conjunto de la población o al percentil espejo, ambas referencias de especial interés en el caso de las empresas sanas.

Calculando los distintos tipos de scores, situamos las empresas individuales en relación con la población, pudiendo ordenarlas según su distancia al fracaso. Para las dos empresas seleccionadas como ejemplo, representadas en el Gráfico 8, la Tabla 25 recoge los tres tipos de scores propuestos, para 10 y 5 ratios, así como las distancias al fracaso medidas por comparación entre los scores (empresas de interés respecto a empresas fracasadas). A primera vista, el gráfico ya permite apreciar:

- Que la empresa fracasada está claramente en situación de fracaso, porque todos los ratios están en peor situación o muy próximos (R7 y R6) a los valores de referencia para el grupo de fracasadas, G_q .

- Que en la empresa sana, los *scores* son negativos, porque casi todos los ratios están en mejor situación que la mediana de la población (excepto R5 y R10).

Tabla 25. *Scores* G_q y ejemplos fracasadas y no fracasadas, 2006.

Scores y Distancias	2006	Ej. Fracasada	Distancia al fracaso	Ej. no fracasada	Distancia al fracaso
$S(5, q), S(5)_i, D(5)_i$	0,8910	1,4250	-0,5340	-0,5650	1,4560
$S(10, q), S(10)_i, D(10)_i$	1,3080	3,0190	-1,7110	-0,9690	2,2770
$\bar{S}(5, q), \bar{S}(5)_i, \bar{D}(5)_i$	0,1782	0,2850	-0,1068	-0,1130	0,2912
$\bar{S}(10, q), \bar{S}(10)_i, \bar{D}(10)_i$	0,1308	0,3019	-0,1711	-0,0969	0,2277
$\overline{Sw}(5, q), \overline{Sw}(5)_i, \overline{Dw}(5)_i$	0,1814	0,2965	-0,1151	-0,1387	0,3201
$\overline{Sw}(10, q), \overline{Sw}(10)_i, \overline{Dw}(10)_i$	0,1529	0,2972	-0,1443	-0,1360	0,2889

Según esto caracterizamos las dos empresas seleccionadas de acuerdo con los niveles propuestos en el apartado de metodología, tanto con *scores* (S , \bar{S} , y \overline{Sw}) como con distancias al fracaso (D , \bar{D} , y \overline{Dw}). La empresa fracasada sitúa todos sus *scores* por encima de los valores medios del grupo de empresas fracasadas, por lo que el modelo le atribuye un riesgo de fracaso muy elevado.

$$[S(5)_i = 1,43] \geq [S(5, q) = 0,89]$$

$$[S(10)_i = 3,02] \geq [S(10, q) = 1,31]$$

$$[\bar{S}(5)_i = 0,29] \geq [\bar{S}(5, q) = 0,18]$$

$$[\bar{S}(10)_i = 0,30] \geq [\bar{S}(10, q) = 0,13]$$

$$[\overline{Sw}(5)_i = 0,30] \geq [\overline{Sw}(5, q) = 0,18]$$

$$[\overline{Sw}(10)_i = 0,30] \geq [\overline{Sw}(10, q) = 0,15]$$

En distancias al fracaso, la empresa fracasada tomada como ejemplo, muestra en todos los casos valores de distancia al fracaso negativos, lo que indica peores percentiles que las medianas de las empresas fracasadas. Es la posición equivalente de riesgo de fracaso muy elevado.

$$[D(5)_i = -0,53] < 0$$

$$[D(10)_i = -1,71] < 0$$

$$[\bar{D}(5)_i = -0,11] < 0$$

$$[\bar{D}(10)_i = -0,17] < 0$$

$$[\overline{Dw}(5)_i = -0,12] < 0$$

$$[\overline{Dw}(10)_i = -0,14] < 0$$

En cuanto a la empresa no fracasada tomada como ejemplo, vemos que los *scores* son negativos, y las distancias a la quiebra son positivas. Los *scores* de la empresa son todos menores que el *score* de las medianas de la población, que por definición es igual a 0, pero mayores que los *scores* formados para los percentiles espejo, $S(h, g) > S(h)_i \geq S(h, esp)$, de manera que esta empresa se encuentra en una situación financiera media, aunque buena, porque está más cerca de los valores espejo que de los valores medianos de la población.

$$[S(5, g) = 0] > [S(5)_i = -0,57] \geq [S(5, esp) = -0,89]$$

$$[S(10, g) = 0] > [S(5)_i = -0,97] \geq [S(5, esp) = -1,30]$$

$$[\bar{S}(5, g) = 0] > [\bar{S}(5)_i = -0,11] \geq [\bar{S}(5, esp) = -0,18]$$

$$[\bar{S}(10, g) = 0] > [\bar{S}(5)_i = -0,10] \geq [\bar{S}(5, esp) = -0,13]$$

$$[\overline{Sw}(5, g) = 0] > [\overline{Sw}(5)_i = -0,14] \geq [\overline{Sw}(5, esp) = -0,18]$$

$$[\overline{Sw}(10, g) = 0] > [\overline{Sw}(5)_i = -0,14] \geq [\overline{Sw}(5, esp) = -0,15]$$

Aplicando ahora las distancias a la quiebra, vemos que todas se encuentran por encima de la distancia a la quiebra de la población global pero por debajo de la distancia espejo, que sería justo el doble del *score* de referencia en 2006: $D(h, g) \leq D(h)_i < D(h, esp)$. De forma consistente, todos los indicadores de

distancia a la quiebra sitúan esta empresa en una situación financiera media, próxima a la posición financiera de las mejores empresas de la población en estudio.

$$[D(5, g) = 0,89] \leq [D(5)_i = 1,46] < [D(5, esp) = 1,78]$$

$$[D(10, g) = 1,31] \leq [D(10)_i = 2,28] < [D(10, esp) = 2,61]$$

$$[\bar{D}(5, g) = 0,18] \leq [\bar{D}(5)_i = 0,29] < [\bar{D}(5, esp) = 0,35]$$

$$[\bar{D}(10, g) = 0,13] \leq [\bar{D}(10)_i = 0,23] < [\bar{D}(10, esp) = 0,26]$$

$$[\overline{Dw}(5, g) = 0,18] \leq [\overline{Dw}(5)_i = 0,32] < [\overline{Dw}(5, esp) = 0,36]$$

$$[\overline{Dw}(10, g) = 0,15] \leq [\overline{Dw}(10)_i = 0,28] < [\overline{Dw}(10, esp) = 0,30]$$

Veamos un nuevo ejemplo con una empresa fracasada y otra sana, tomadas al azar, dentro de la población de 2007. El Gráfico 9 muestra la posición de los diez ratios en los percentiles de la población, y se puede apreciar su posición relativa respecto a la mediana de la población, respecto a la frontera de quiebra, y respecto a los percentiles espejo. Con el cálculo de las distancias por empresa (D , \bar{D} , y \overline{Dw}) se cuantifica la distancia global de la empresa al fracaso, pero además, al disponer de la información de las distancias individuales por ratios, se identifican los puntos fuertes y débiles reflejados en los indicadores financieros. Esas diferencias por ratio, que quedan explícitas en la metodología, pueden usarse a efectos de gestión de las empresas individuales.

Gráfico 9. Fracaso empresarial. Empresas constructoras de Castilla y León. 2007. Caracterización de empresas concretas



En la Tabla 26 se aprecia que las distancias al fracaso de la empresa fracasada son muy similares a las de la mediana de la población, cuando tomamos los 10 ratios. Aunque para los 5 ratios más discriminantes la situación de la empresa es mejor, esto es, más alejada del fracaso. En este caso, vemos como los *scores* ponderados corrigen el efecto de los ratios menos discriminantes y presentan valores medios más cercanos para 5 y 10 ratios.

Tabla 26. Scores y ejemplos fracasadas y no fracasadas, 2007.

Scores y Distancias	2007	Ej. Fracasada	Distancia al fracaso	Ej. no fracasada	Distancia al fracaso
$S(5, q), S(5)_i, D(5)_i$	0,7690	-0,2560	1,0250	-1,4900	2,2590
$S(10, q), S(10)_i, D(10)_i$	1,2260	0,1080	1,1180	-1,8860	3,1120
$\bar{S}(5, q), \bar{S}(5)_i, \bar{D}(5)_i$	0,1538	-0,0512	0,2050	-0,2980	0,4518
$\bar{S}(10, q), \bar{S}(10)_i, \bar{D}(10)_i$	0,1226	0,0108	0,1118	-0,1886	0,3112
$\overline{Sw}(5, q), \overline{Sw}(5)_i, \overline{Dw}(5)_i$	0,1550	-0,0371	0,1921	-0,3002	0,4552
$\overline{Sw}(10, q), \overline{Sw}(10)_i, \overline{Dw}(10)_i$	0,1333	0,0023	0,1310	-0,2052	0,3385

Pero caractericemos las dos empresas según los niveles propuestos en el apartado de metodología, tanto con *scores* (S , \bar{S} , y $\overline{S\bar{w}}$) como con distancias al fracaso (D , \bar{D} , y $\overline{D\bar{w}}$). La empresa fracasada ofrece puntuaciones de distinto signo en los *scores* calculados con 5 indicadores y los calculados con 10 indicadores.

$[S(5, g) = 0] > [S(5)_i = -0,26] \geq [S(5, esp) = -0,77]$, situación financiera media;

$[S(10, q) = 1,23] > [S(10)_i = 0,10] \geq [S(10, g) = 0]$, expuesta a riesgo de fracaso moderado;

$[\bar{S}(5, g) = 0] > [\bar{S}(5)_i = -0,05] \geq [\bar{S}(5, esp) = -0,15]$, situación financiera media;

$[\bar{S}(10, q) = 0,12] > [\bar{S}(10)_i = 0,01] \geq [\bar{S}(10, g) = 0]$, expuesta a riesgo de fracaso moderado;

$[\overline{S\bar{w}}(5, g) = 0] > [\overline{S\bar{w}}(5)_i = -0,04] \geq [\overline{S\bar{w}}(5, esp) = -0,16]$, situación financiera media;

$[\overline{S\bar{w}}(10, q) = 0,13] > [\overline{S\bar{w}}(10)_i = 0,002] \geq [\overline{S\bar{w}}(10, g) = 0]$, riesgo de fracaso moderado.

En distancias al fracaso, vemos como los indicadores muestran lo mismo: una situación financiera media, para los 5 ratios más discriminantes, y riesgo de fracaso muy moderado, al presentar valores similares a los centrales de la población, cuando se usan los 10 ratios:

$[D(5, g) = 0,77] \leq [D(5)_i = 1,02] < [D(5, esp) = 1,53]$;

$0 \leq [D(10)_i = 1,12] < [D(10, g) = 1,23]$;

$[\bar{D}(5, g) = 0,15] \leq [\bar{D}(5)_i = 0,21] < [\bar{D}(5, esp) = 0,30]$;

$0 \leq [\bar{D}(10)_i = 0,11] < [\bar{D}(10, g) = 0,12]$;

$$[\overline{Dw}(5, g) = 0,15] \leq [\overline{Dw}(5)_i = 0,19] < [\overline{Dw}(5, esp) = 0,31];$$

$$0 \leq [\overline{Dw}(10)_i = 0,131] < [\overline{Dw}(10, g) = 0,133].$$

Los ratios para los que esta empresa muestra mayor cercanía a la quiebra son: R1, R8, R6 y R2. Estos serían los indicadores financieros que la empresa tendría que mejorar si su situación de fracaso fuera reversible y se decidiera una reorganización.

La empresa no fracasada da valores claramente mejores que los obtenidos para la mediana de la población. Sin embargo, algunos ratios están situados en percentiles de fracaso, pudiendo interpretarse como indicadores que deben mejorarse para llegar a una situación financiera saludable. Concretamente, R6 indica gastos financieros elevados, y R2, (CA/CL), resulta muy bajo, si bien este último ratio resulta poco discriminante para esta población y periodo concretos.

3.4.4. Evolución de los indicadores financieros de empresas que fracasan

En este apartado vamos a hacer un seguimiento de empresas españolas que fracasaron en 2010, analizando con la metodología propuesta éste y los dos periodos previos al fracaso⁴³, 2009 y 2008. Nuestro objetivo es comprobar que los resultados de la metodología son consistentes con el razonamiento económico, y las distancias de percentiles se van ampliando a medida que se acerca el fracaso; pero también queremos comprobar que los resultados de identificación de fracaso obtenidos en el periodo 2006-08, para una comunidad autónoma concreta, pueden aplicarse a la predicción de quiebra en periodos posteriores y en zonas geográficas más amplias. Finalmente, podemos identificar los indicadores financieros más discriminantes en esos dos años previos, si bien los resultados obtenidos deben tomarse con cautela dado el número de empresas fracasadas

⁴³ No se ha podido extender el análisis de evolución a un número de años mayor debido al cambio de normativa contable que afecta al grupo objetivo de 2007 a 2008. Esto hubiera supuesto también una reducción significativa del grupo de empresas fracasadas.

(28) que cumplen la condición de estar activas durante los dos años previos⁴⁴ en la base de datos.

Como primer paso, se calcula el valor mediano, \tilde{z}_f , de las empresas fracasadas en 2010, para cada ratio, en ese año y en los dos anteriores. En el Anexo VIII se recogen las representaciones gráficas de esa evolución en los valores de la mediana para cada ratio.

Tabla 27. Medianas de ratios para empresas fracasadas y para empresas no fracasadas.

EMPRESAS FRACASADAS				EMPRESAS NO FRACASADAS			
Ratios	Mediana 2008	Mediana 2009	Mediana 2010	Ratios	Mediana 2008	Mediana 2009	Mediana 2010
R1	0.88367665	0.9736689	1.4772255	R1	0.7790409	0.764425	0.7640737
R2	1.061705	1.038557	0.62227095	R2	1.311502	1.360972	1.370086
R3	0.0517189	-0.05895485	-0.2849137	R3	0.0447217	0.0256129	0.0137286
R4	0.0065682	-0.09802225	-0.2837529	R4	0.0160181	0.0067646	0.0023021
R5	0.8032744	0.79556605	0.7041586	R5	0.799213	0.7885137	0.782233
R6	0.05028705	0.03888345	0.01401735	R6	0.0221526	0.0191608	0.0148811
R7	0.09856595	0.08420625	0.03167015	R7	0.1262436	0.1565207	0.156261
R8	0.02692695	-0.0720142	-0.2011695	R8	0.0644299	0.0395611	0.0235695
R9	0.00452075	-0.15621105	-0.50519875	R9	0.0116093	0.0062645	0.0027162
R10	1.40569	0.97212835	0.5705226	R10	1.536743	1.160619	0.9977077

Los ratios son: R1=TD/TA, deuda total/ activo total; R2=CA/CL, activo corriente/ pasivo corriente; R3=BAIT/TA, resultados antes de intereses e impuestos/ activo total; R4=NI/TA, beneficio neto/ activo total; R5=CA/TA, activo corriente/ activo total; R6= FE/TD, gastos financieros/deuda total; R7= RP/TA, resultados retenidos/ activo total; R8=CF/TD, *cash flow*/ deuda total; R9=NI/SL, beneficio neto/ ventas; y R10=SL/TA, ventas/ total activo.

A continuación, se calcula el lugar que las medianas anteriores ocupan dentro del conjunto de la población en estudio (11.889 empresas), o sea, el rango del percentil, $\pi(q)_f$, de esos valores medianos por ratio.

⁴⁴ En el Anexo IX se representan gráficamente, por ratio, los valores de las 28 empresas en los tres años de estudio.

Tabla 28. Percentiles de ratios para empresas fracasadas.

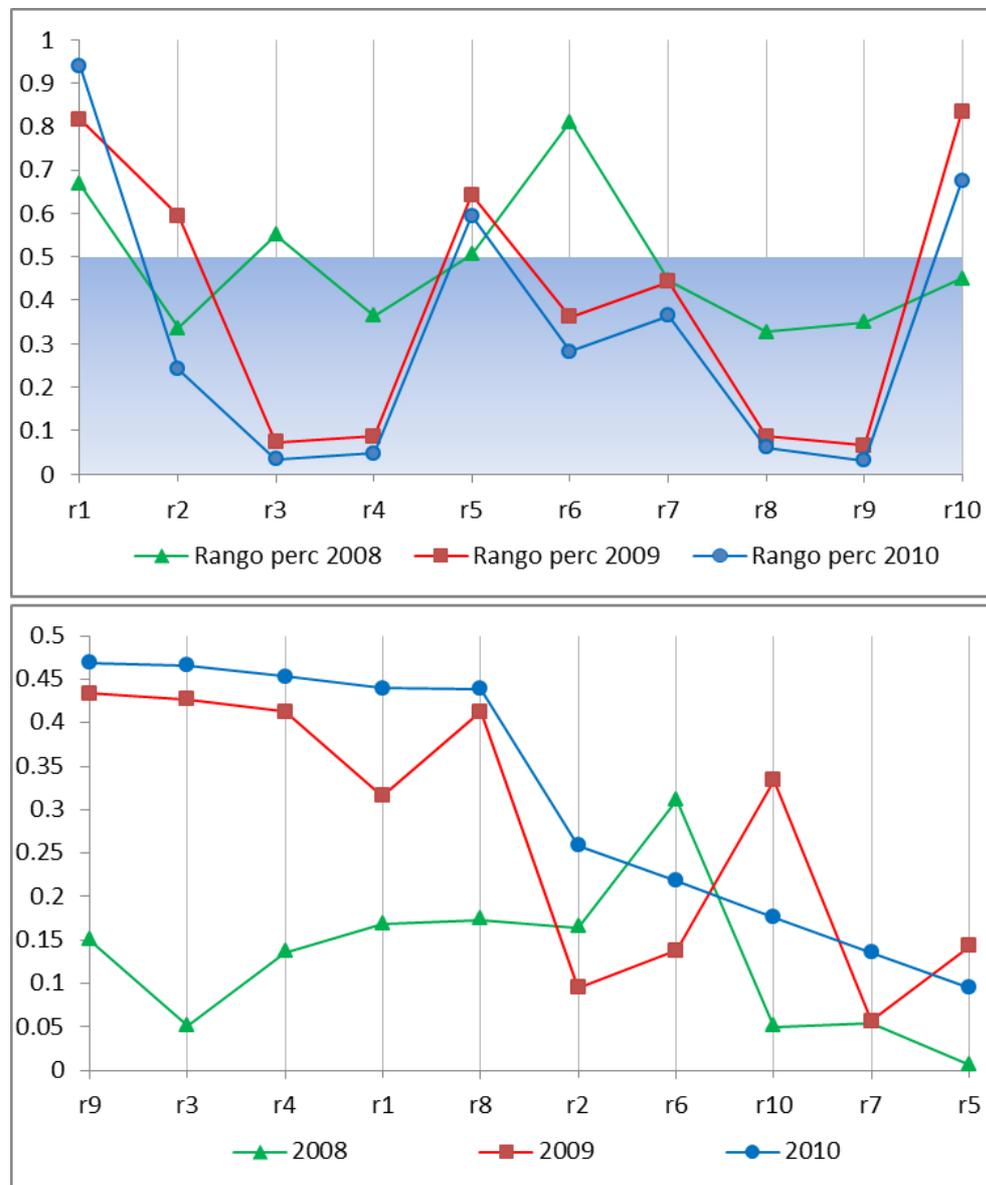
Ratios	Rango perc 2008	Rango perc 2009	Rango perc 2010
R1	0.668	0.816	0.94
R2	0.336	0.595	0.241
R3	0.551	0.073	0.034
R4	0.364	0.087	0.047
R5	0.506	0.643	0.595
R6	0.811	0.362	0.282
R7	0.445	0.443	0.364
R8	0.327	0.087	0.061
R9	0.35	0.066	0.031
R10	0.45	0.834	0.676

Los ratios son: R1=TD/TA, deuda total/ activo total; R2=CA/CL, activo corriente/ pasivo corriente; R3=BAIT/TA, resultados antes de intereses e impuestos/ activo total; R4=NI/TA, beneficio neto/ activo total; R5=CA/TA, activo corriente/ activo total; R6= FE/TD, gastos financieros/deuda total; R7= RP/TA, resultados retenidos/ activo total; R8=CF/TD, *cash flow* deuda total; R9=NI/SL, beneficio neto/ ventas; y R10=SL/TA, ventas/ total activo.

En esta tabla ya podemos ver que unos valores están más cerca y otros más lejos de la mediana de la población (percentil 0,5); cómo los percentiles de los ratios se van alejando de la mediana cuando está más próximo el fracaso; y qué ratios se alejan antes. Representamos la evolución de los percentiles de fracaso por ratio en 2008, 2009 y 2010 en el Gráfico 10.

Si colocamos los ratios por orden de distancia a la mediana en 2010 y hallamos la distancia euclídea para evitar el signo, vemos, en la segunda parte del Gráfico 10, que los ratios más discriminantes como indicadores de fracaso para esta población son R9, R3, R4, R1 y R8. También vemos que precisamente para esos cinco ratios entre la posición dos años antes y la posición en el año anterior hay una gran diferencia. El gráfico sugiere que podríamos haber hecho una buena predicción de la situación de fracaso de estos ratios con los valores del año anterior.

Gráfico 10. Percentiles de ratios. Evolución 2008-10



Los ratios son: R1=TD/TA, deuda total/ activo total; R2=CA/CL, activo corriente/ pasivo corriente; R3=BAIT/TA, resultados antes de intereses e impuestos/ activo total; R4=NI/TA, beneficio neto/ activo total; R5=CA/TA, activo corriente/ activo total; R6= FE/TD, gastos financieros/deuda total; R7= RP/TA, resultados retenidos/ activo total; R8=CF/TD, cash flow/ deuda total; R9=NI/SL, beneficio neto/ ventas; y R10=SL/TA, ventas/ total activo.

En la tabla siguiente se recogen los ratios con mayor poder discriminante cada año, medido en diferencia de percentiles entre las empresas fracasadas y las no fracasadas, $d(q)_f$. Así, comprobamos que dos años antes solo el ratio R6 (gastos financieros / deuda total) cuenta con buen poder discriminante; en 2009 los ratios

con buen poder discriminante coinciden con los de 2010, excepto en el caso de R10 (ventas/ total activo), que pierde importancia, siendo sustituido en la lista por R2 (activo corriente / pasivo corriente). Estos resultados son consistentes con los obtenidos por Pompe y Bilderbeek (2005) ya que el deterioro inicial no se muestra en ratios de actividad y rentabilidad, sino de coste financiero, y un año antes del fracaso no hay un orden en el que las diferentes categorías de ratios empiecen a tener capacidad discriminante: margen, rentabilidad económica, generación de *cash flow* y endeudamiento se muestran discriminantes en ambos periodos (2009 y 2010) y en niveles similares. En este sentido, nuestros resultados también son similares a los de Labatut *et al.* (2009) que señalan como variables más importantes para la predicción la reducción de rentabilidad y la falta de generación de flujos de tesorería que obligan a incrementar el endeudamiento y como contrastan Gill de Albornoz y Giner (2010), este incremento genera una probabilidad mayor de fracaso al entrar en una ciclo de recesión. Consistente con el estudio de ratios más significativos en la literatura sobre fracaso en pymes (desarrollado en el apartado 2.2 de esta tesis), R7 (resultados retenidos / activo total), tiene poco poder discriminante en cualquiera de los tres periodos.

Tabla 29. Ratios con mayor poder predictivo (diferencias de percentiles >0,25).

2010			2009			2008		
Orden	Ratio	Ponderación	Orden	Ratio	Ponderación	Orden	Ratio	Ponderación
1º	R9	0,469	1º	R9	0,434	1º	R6	0,311
2º	R3	0,466	2º	R3	0,427			
3º	R4	0,453	3º	R4	0,413			
4º	R1	0,440	4º	R8	0,413			
5º	R8	0,439	5º	R10	0,334			
6º	R2	0,259	6º	R1	0,316			

Los ratios son: R1=TD/TA, deuda total/ activo total; R2=CA/CL, activo corriente/ pasivo corriente; R3=BAIT/TA, resultados antes de intereses e impuestos/ activo total; R4=NI/TA, beneficio neto/ activo total; R5=CA/TA, activo corriente/ activo total; R6= FE/TD, gastos financieros/deuda total; R7= RP/TA, resultados retenidos/ activo total; R8=CF/TD, *cash flow* deuda total; R9=NI/SL, beneficio neto/ ventas; y R10=SL/TA, ventas/ total activo.

Aunque no hemos podido repetir el análisis para el periodo 2007-09, a causa del cambio contable, que podía distorsionar los valores de los ratios de 2007 a 2008, sí hemos realizado el estudio paralelo para empresas fracasadas en 2009, que la

base de datos muestra como no fracasadas en los dos años anteriores, dentro de la población de empresas no fracasadas en los tres periodos consecutivos. Comparando los percentiles de los ratios en 2009 con los de 2008, obtenemos resultados similares a los mostrados aquí para 2010 y 2009. Esto es, un año antes los ratios con mayor capacidad predictiva son: R9, R8, R4, R3 y R1.

Al situar los valores de las empresas fracasadas en cada periodo en el conjunto de valores de la población a la que pertenecen, la metodología de percentiles evita el efecto de la evolución de los ratios en la población de estudio. Si se da un empeoramiento de la situación financiera del sector, el deterioro en los ratios de las empresas puede no deberse a un peor desempeño que conduzca al fracaso, sino a esa tendencia general originada en factores macroeconómicos o sectoriales. Así, vemos que en la población estudiada los ratios R3, R4, R6, R8, R9 y R10 muestran un claro deterioro en las empresas no fracasadas. De ahí que la evolución de los valores de los ratios que se muestran en el Anexo IX no coincida en todos los casos con la evolución de los percentiles (ej. R10). El primer efecto interesante es que, a diferencia de los valores de los ratios, los percentiles del año anterior se sitúan muy próximos a los que se dan en el periodo de fracaso, sugiriendo que los percentiles de los ratios sobre los valores de la población tienen una mayor capacidad para la predicción a ese plazo. A continuación, cuantificamos esta lejanía o proximidad en términos de distancia al fracaso.

Tabla 30. Scores del grupo de empresas fracasadas en 2010 respecto a un grupo de no fracasadas estable, con 5 y 10 ratios en 2008, 2009 y 2010.

Scores	2008	2009	2010
$S(5, q)$	0,6780	2,0030	2,2670
$S(10, q)$	1,2640	2,7700	3,1510
$\bar{S}(5, q)$	0,1356	0,4006	0,4534
$\bar{S}(10, q)$	0,1264	0,2770	0,3151
$\bar{S}w(5, q)$	0,1501	0,4052	0,4537
$\bar{S}w(10, q)$	0,1827	0,3520	0,3814

Vemos como en el periodo del fracaso, 2010, la distancia en percentiles de los 10 ratios seleccionados y de los 5 más discriminantes son elevadas (3,15 sobre 5 y 2,27 sobre 2,5 en *scores* acumulados; 0,32 y 0,45 respectivamente sobre un máximo de 0,5 en los *scores* medios; y 0,38 y 0,45 respectivamente sobre un máximo de 0,5 en *scores* medios ponderados). En el periodo anterior al fracaso, las mismas empresas muestran valores muy próximos al fracaso en comparación con los datos de 2009 de las mismas empresas no fracasadas. La diferencia en percentiles para la media de los ratios es inferior a 0,04 con 10 ratios, y de 0,05 con 5 ratios. En cambio, dos años antes del fracaso, los *scores* se sitúan más próximos a las medianas de la población en estudio que a los valores medianos de fracaso de 2010. Por ejemplo, el *score* medio para 10 (5) ratios es 0,13 (0,14), que se encuentra más próximo a 0 (si todos los ratios estuvieran exactamente en el valor de la mediana de la población) que a 0,5 (si todos los ratios alcanzaran la distancia máxima en percentiles).

A continuación, hemos calculado los seis *scores* para las 11.861 empresas no fracasadas, para ver cuántas de ellas muestran la misma puntuación o superior que las empresas fracasadas en 2010. El porcentaje representaría el error tipo II, esto es, las empresas no fracasadas identificadas como fracasadas por la metodología.

Tabla 31. Tabla empresas con igual o superior puntuación cada año que las fracasadas en 2010.

<i>Scores</i>	2008		2009		2010	
	Nº empr.	%	Nº empr.	%	Nº empr.	%
$S(5, q)$	9.661	81,38%	1.526	12,85%	357	3,01%
$S(10, q)$	11.269	94,92%	4.380	36,90%	2.675	22,53%
$\bar{S}(5, q)$	9.661	81,38%	1.526	12,85%	357	3,01%
$\bar{S}(10, q)$	11.269	94,92%	4.380	36,90%	2.675	22,53%
$\bar{S}w(5, q)$	10684	90,00%	1.908	16,07%	480	4,04%
$\bar{S}w(10, q)$	11.342	95,54%	3.764	31,71%	2.212	18,63%

En 2010 vemos la capacidad del modelo para identificar empresas fracasadas. De las 11.861 sólo el 3% presentan un *score* que indique fracaso para los 5 ratios más discriminantes. Vemos cómo incorporar ratios poco discriminantes introduce ruido y hace que más empresas no fracasadas se clasifiquen como fracasadas. Un año antes, casi un 13% de las empresas no fracasadas alcanzan puntuaciones de fracaso, lo que significa que estaríamos clasificando correctamente más del 87% de las empresas no fracasadas. En cambio, dos años antes del fracaso, en 2008, las empresas que van a fracasar en dos años y las que no van a fracasar presentan muy poca diferencia, más del 80% de las empresas que no van a fracasar están en una posición similar (o peor) a las que fracasarán en 2010.

A continuación, la tabla 32 muestra cuántas empresas obtienen, en los tres años, y para cada *score*, puntuaciones iguales o superiores a la puntuación de fracaso de 2010. Nótese que las cifras coinciden en 2010, pero mejoran considerablemente en los dos años previos. Esto significa que las empresas que van a fracasar, cuanto más lejos de la fecha de fracaso, más difícil es predecir su quiebra. En nuestra población de estudio, a un año se obtiene una buena tasa de acierto y a dos años el modelo no serviría para predecir la empresas que van a fracasar. Sin embargo, las empresas que no van a fracasar pueden identificarse con tasas de error relativamente pequeñas aplicando *scores* de fracaso de otros años. En la tabla, vemos como el error un año antes para las empresas no fracasadas es muy pequeño y, dos años antes, el porcentaje de error se incrementa muy poco, de manera que se siguen manteniendo tasas de acierto por encima del 95% para los 5 ratios más discriminantes.

Tabla 32. Tabla empresas con igual o superior puntuación que $S(q)$ en 2010.

<i>Scores</i>	2008		2009		2010	
	Nº empr.	%	Nº empr.	%	Nº empr.	%
$S(5, q)$	517	4,36%	439	3,70%	357	3,01%
$S(10, q)$	2.607	21,96%	2.676	22,54%	2.675	22,53%
$\bar{S}(5, q)$	517	4,36%	439	3,70%	357	3,01%
$\bar{S}(10, q)$	2.607	21,96%	2.676	22,54%	2.675	22,53%
$\bar{S}w(5, q)$	622	5,24%	549	4,62%	480	4,04%
$\bar{S}w(10, q)$	2.204	18,57%	2.204	18,57%	2.212	18,63%

Los porcentajes de empresas no fracasadas con puntuaciones de fracaso en 2010 son coherentes con la literatura previa, en el sentido de que parte de las empresas no fracasadas pueden estar en dificultades financieras serias (Watson y Everett, 1996; Altman *et al.*, 2008; Baixauli y Mónica-Milo, 2010).

3.4.5. Evolución de empresas en riesgo de fracaso que se recuperan

Con objeto de mostrar el potencial de la metodología propuesta para la reorganización de empresas en riesgo de fracaso, hemos seleccionado otro grupo de empresas a efectos de seguimiento en el periodo 2008-2010. Una vez tomadas en el apartado anterior las empresas españolas pertenecientes al sector de la construcción, que en la base de datos existen en los tres años de este periodo, y aplicada la metodología hemos obtenido en 2010 los ratios más discriminantes entre el grupo de fracasadas y el grupo de no fracasadas, así como los *scores* que determinan la distancia a la quiebra en ese grupo. En este caso, aunque la tabla refleja todos los *scores*, vamos a fijarnos en los *scores* calculados con los 5 ratios más discriminantes, que son los que han mostrado en los apartados anteriores el mayor poder de identificación y predicción de fracaso.

Dentro de esta población de empresas que existen en los tres periodos, seleccionamos ahora un grupo de empresas que, partiendo de una situación de peligro de fracaso, consiguieron remontar, mejorando su situación financiera en los dos periodos siguientes. El proceso de selección del grupo de empresas que se va a analizar es el siguiente:

1. En 2008, seleccionamos todas las empresas existentes y no fracasadas en el periodo 2008-10 cuya puntuación en el *score* con cinco ratios, $S(5)_i$, es igual o superior al *score* de fracaso en 2010, $S(5, q)$. En esta situación de riesgo muy elevado de fracaso se encuentran 517 empresas de las 11.861 no fracasadas durante el periodo 2008-10.
2. En 2009, seleccionamos entre las anteriores todas las empresas cuya puntuación se ha reducido, lo cual implica una mejora en su situación de

solvencia de acuerdo con nuestro modelo. En esta situación están 465 empresas de las 517 anteriores.

3. En 2010, seleccionamos entre las anteriores todas las empresas cuya puntuación se haya reducido de nuevo, indicando la mejora progresiva de la situación de solvencia en el periodo considerado. Como resultado, obtenemos un grupo de estudio formado por 279 empresas.

Nuestra intención es identificar los indicadores financieros en los que estas empresas han mejorado para conseguir salir de la situación de peligro. Si rentabilidad, margen y generación de flujos son los principales indicadores de fracaso en pymes españolas (Labatut *et al.*, 2009) una empresa debería mejorar primero en estos factores para alejarse de la amenaza de insolvencia⁴⁵. Nuestra hipótesis es que las empresas se alejan de la situación de peligro si mejoran los ratios que resultan más discriminantes, que en este estudio concreto son: R9, R3, R4, R1 y R8, relativos al margen, la rentabilidad, el endeudamiento y la generación de fondos.

Si calculamos los *scores* y distancias al fracaso para el grupo de empresas en estudio, tal como se proponía en nuestra propuesta metodológica, esto es, hallando la mediana de cada ratio en el grupo y luego asignando a esa mediana el percentil que nos permita hallar *scores* y distancias, obtenemos los resultados que se exponen en la Tabla 33. El *score* para los cinco ratios más discriminantes $S(5, a)$ es de 2,33 en 2008, un valor próximo al máximo para 5 ratios (2,5) y superior a 2,26, que representa el *score* del grupo de fracaso tomado como referencia. En términos de distancia al fracaso, la cantidad es negativa. Ambos indicadores sitúan este grupo en riesgo muy elevado de fracaso, de acuerdo con las clasificaciones propuestas en el apartado 3.3. En 2009 y 2008 se reduce progresivamente el *score* en más de un tercio, aumentando la distancia al fracaso.

⁴⁵ No se han realizado trabajos sobre reorganización en empresas españolas, pero Laitinen (2008) recoge los indicadores del éxito de la reorganización en la escasa evidencia empírica internacional sobre la cuestión. La rentabilidad es un elemento presente en todos los casos, la liquidez aparece varias veces, y también se alude al apalancamiento. En cuanto a los resultados de Laitinen (2008), este autor encuentra que el éxito en la reorganización depende del ajuste de todos los factores, de la gestión orientada a la eficiencia y de una tasa de deuda suficiente que garantice la viabilidad de la empresa.

En ambos periodos la puntuación está entre 0 y 2,26 (lo que implica distancias al fracaso positivas y crecientes), indicando una exposición al riesgo de fracaso cada vez más moderado. En estas condiciones, el grupo de estudio se ajusta a nuestros requerimientos y parece adecuado para el análisis de la evolución de los factores financieros individuales.

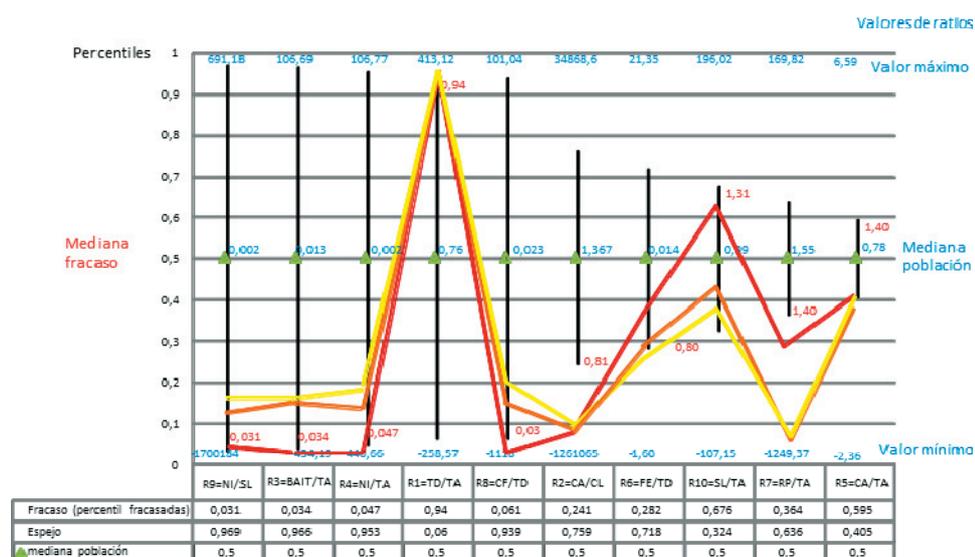
Tabla 33. Scores de fracaso en 2010 y scores y distancias del grupo a en 2008, 2009 y 2010.

Scores y Distancias	Referencia fracaso 2010	Grupo a					
		2008		2009		2010	
		Scores	Distancia al fracaso	Scores	Distancia al fracaso	Scores	Distancia al fracaso
$S(5, q), S(5, a), D(5, a)$	2,2670	2,3320	-0,0650	1,9210	0,3460	1,7360	0,5310
$S(10, q), S(10, a), D(10, a)$	3,1510	3,1570	-0,0060	2,8260	0,3250	2,6280	0,5230
$S(5, q), S(5, a), \bar{D}(5, a)$	0,4534	0,4664	-0,0130	0,3842	0,0692	0,3472	0,1062
$S(10, q), S(10, a), \bar{S}(10, a)$	0,3151	0,3157	-0,0006	0,2826	0,0325	0,2628	0,0523
$\bar{S}w(5, q), \bar{S}w(5, a), \bar{S}w(5, a)$	0,4537	0,4664	-0,0126	0,3837	0,0700	0,3466	0,1072
$\bar{S}w(10, q), \bar{S}w(10, a), \bar{S}w(10, a)$	0,3814	0,3941	-0,0128	0,3371	0,0443	0,3100	0,0714

En el gráfico 11, se recoge la caracterización del grupo en estudio (grupo *a*), las 279 empresas que partían en 2008 con puntuaciones de riesgo de fracaso elevado pero han ido mejorando progresivamente su situación financiera en 2009 y 2010. Las líneas verticales sitúan gráficamente, para cada ratio, la frontera de fracaso, los valores medianos de la población y la frontera espejo, medidos en percentiles, que aparecen numéricamente en la tabla situada al pie del gráfico. Estos son los valores de referencia donde se sitúan los percentiles alcanzados por el grupo en estudio. En azul se han añadido los valores mínimo, mediano y máximo de cada ratio en la población estudiada; y en rojo, se puede ver el valor de cada ratio que corresponde a la mediana de esos ratios para las empresas fracasadas. Las líneas de color sitúan el grupo de estudio en relación con las fronteras de ratios medidas en percentiles: la línea roja corresponde a los percentiles de los valores medianos del grupo en 2008; la línea naranja, a los percentiles en 2009; y la línea amarilla, a los percentiles en 2010.

Nuestros resultados indican claramente que, para el grupo en estudio, mejora la posición de los ratios R9, R3, R4 y R8, cuatro de los cinco ratios más discriminantes, relativos a margen, rentabilidad y generación de fondos, en consonancia con los resultados obtenidos por Labatut *et al.* (2009) sobre una muestra de pymes españolas, si bien este estudio no se orientaba a la reorganización. En cambio, no hay ninguna mejora en el ratio R1, representativo del endeudamiento, y esto resulta consistente con los resultados de Laitinen (2008) sobre la necesidad de obtener fondos ajenos suficientes para acometer la reorganización con ciertas garantías de éxito.

Gráfico 11. Fracaso empresarial. Grupo a. Empresas constructoras. España. 2008



En cuanto a los otros cinco ratios, dotados de menor poder discriminante, no parecen determinantes en la mejora de la situación financiera de la empresa, en el contexto de comparación entre fracasadas y no fracasadas. En el caso de R2 y R5, relativos a equilibrio económico-financiero y estructura económica, respectivamente, la mejora es muy pequeña; los gastos financieros se reducen, pero las empresas no fracasadas en esta población tenían gastos financieros más elevados que las fracasadas; la rotación se reduce, siendo elevada la de las

empresas fracasadas; y los beneficios no distribuidos se reducen considerablemente, aunque ya hemos señalado previamente que este es un ratio poco significativo para distinguir situaciones de fracaso en pymes. De esta forma, se confirma nuestra hipótesis de partida: la metodología de diferencias de percentiles permite identificar los indicadores financieros cuya mejora resulta más efectiva en la reorganización de las empresas en riesgo de fracaso. Hemos comprobado que es posible caracterizar grupos concretos de empresas dentro de la población, tanto de forma gráfica como en forma cuantitativa calculando los *scores* y las distancias al fracaso. También hemos probado que la posición relativa frente a la frontera de quiebra, a la posición central de la población y a la frontera espejo proporciona una considerable cantidad de información adicional, respecto a las metodologías tradicionales, susceptible de ser aplicada en múltiples procesos de evaluación de la empresa.

3.5. Conclusiones a la tercera parte

En esta tercera parte de la tesis se propone un modelo de medida de distancia al fracaso para empresas pequeñas, que constituyen el grupo menos tratado en la literatura empírica previa sobre fracaso empresarial. Las variables utilizadas en el modelo son ratios contables ampliamente contrastados para este fin, lo cual nos permite comparar los resultados de identificación de empresas fallidas obtenidos con la metodología propuesta con los resultados de aplicar modelos alternativos tanto de uso consolidado como otros de aparición reciente en esta línea de investigación. Los ratios se calculan a partir de información disponible en las cuentas anuales obligatorias que cualquier empresa debe presentar, lo cual es especialmente importante en empresas pequeñas, donde la información disponible es limitada.

El modelo consiste en el cálculo de una puntuación o *score* por empresa, donde en lugar de utilizar distancias entre ratios, que son variables heterogéneas (especialmente en empresas pequeñas y medianas), se utilizan distancias entre percentiles, lo que convierte todas las variables en homogéneas, de manera que pueden agregarse, para formar *scores*, con independencia de sus rangos y

distribuciones estadísticas. Además, los percentiles se calculan a partir de las medianas, lo cual resulta especialmente adecuado considerando la elevada dispersión de las variables contables en pymes.

Esta forma de comparar las variables es la principal contribución, porque permite establecer clasificaciones (o *rankings*) de las variables más relevantes en la distinción entre dos grupos; el *score* proporciona una puntuación por empresa o grupo de empresas que relacionamos con tres niveles de referencia: las medianas de la población, las medianas de las empresas fracasadas, y las medianas de las empresas saludables (medidas por los percentiles espejo); también se pueden establecer ponderaciones para un *score* determinado en referencia a un subgrupo (por ejemplo, un sector), donde unas variables concretas pueden ser más relevantes que otras para la distinción fracasadas / no fracasadas; a partir de los *scores*, se calculan las distancias a la quiebra (aunque también se pueden calcular las distancias a la situación normal de la población), o a la situación de empresa saludable, en función de cuál sea la situación de la empresa o grupo de empresas en estudio.

La metodología se puede aplicar a cualquier grupo donde el número de fallidas sea suficiente para establecer las referencias sobre posición de los ratios, que sirvan de comparación para el resto de empresas. En nuestro caso, evaluamos la distancia al fracaso de las empresas constructoras españolas en el periodo 2006-2010, con la realización de dos trabajos empíricos diferentes. En el primero, analizamos las empresas de la construcción fracasadas en Castilla y León en el periodo 2006-08, mediante el cálculo de *scores* y distancias al fracaso por año. En el segundo, hacemos un estudio de seguimiento desde 2008, de las empresas españolas de la construcción fracasadas en 2010. A la vista de los resultados obtenidos, podemos concluir que los *scores* y distancias a la quiebra, calculados con diferencias de percentiles sobre medianas, son eficientes en la identificación de empresas pequeñas fracasadas, o en riesgo de fracaso, obteniendo resultados comparables a los de metodologías de uso común en la literatura previa.

Los ratios de generación de fondos, rentabilidad económica y endeudamiento resultan ser los más discriminantes, tanto en nuestro trabajo por años (antes del cambio de normativa contable) para una Comunidad Autónoma, como en el estudio de seguimiento en todo el país para un periodo temporal posterior (tras el cambio de normas contables), en consonancia con resultados previos en nuestro país para el sector de la construcción y, en general, para pymes.

De las medidas propuestas, el *score* medio calculado con aquellos ratios con mayor poder discriminante (diferencias de percentiles superiores a 0,25) es el que ofrece mejores resultados en la predicción del fracaso. Un año antes del fracaso, se obtienen resultados de predicción muy buenos tanto para empresas que van a fracasar (por encima del 85%), como para empresas que no van a fracasar el próximo año (por encima del 95%). Con dos años de antelación el modelo sigue siendo muy efectivo para empresas que no van a fracasar, pero no es capaz de predecir las que fracasarán. Hay que tener en cuenta que este trabajo se aplica a empresas pequeñas, donde hay unas tasas muy elevadas de creación y desaparición de empresas en plazos muy cortos, especialmente en épocas de crisis como la que incluye nuestro periodo de estudio (Dietsch y Petey, 2004; Altman y Sabato, 2005; Dannreuther y Kessler, 2010).

En consonancia con el trabajo de Pompe y Bilderbeek (2005) no apreciamos que unas categorías de ratios se deterioren antes que otras, sino que varias dimensiones de la empresa parecen deteriorarse simultáneamente en los dos últimos periodos, aunque el deterioro sea creciente a medida que la empresa se acerca a su periodo de fracaso.

Finalmente, y a diferencia de otras metodologías que sirven para identificar empresas fallidas, esta tiene una utilidad potencial en la reorganización de empresas en riesgo de fracaso. Al determinar distancias individuales para cada variable ratio a valores medianos de empresas fallidas, a valores medianos de la población y a valores medianos de empresas saludables (valores espejo), esta metodología permite identificar los puntos fuertes y los puntos débiles en los

indicadores financieros de una empresa en referencia a un grupo, en un periodo y en una zona geográfica determinadas (las que sirven de base al estudio). Así, se conocen para cada empresa los ratios que están contribuyendo a aumentar la distancia al fracaso o, por el contrario, reduciendo esa distancia y poniendo a la empresa en situación de mayor peligro.

En el último estudio empírico, llevado a cabo en esta tercera parte de la tesis, se ha identificado un grupo de empresas en recuperación que partían de una situación de riesgo de fracaso muy elevada, y eso nos ha permitido confirmar que las variables identificadas como más discriminantes por nuestra metodología, relativas a margen, rentabilidad y generación de recursos son las primeras que mejoran las empresas en recuperación. Por el contrario, el endeudamiento, que es una variable también situada entre las más discriminantes por nuestra metodología (en consonancia con la literatura previa sobre identificación y predicción de fracaso), no se reduce en la fase inicial de la reorganización. Este resultado es consistente con la escasa literatura previa internacional sobre reorganización de empresas fracasadas o en riesgo de fracaso. Por tanto, la variable endeudamiento muestra un comportamiento diferencial respecto al resto de indicadores financieros con gran poder discriminante: su aumento es determinante para identificar la aproximación de las empresas al riesgo de fracaso, pero no es un buen indicador de que la empresa se aleja de esa situación de riesgo de fracaso, al menos en las fases iniciales de la recuperación.

Otra ventaja del cálculo de *scores* y distancias al fracaso mediante percentiles es que se tiene en cuenta la situación de la población completa, con lo cual se eliminan los posibles aumentos o disminuciones de puntuaciones y distancias a la quiebra debidos a movimientos generalizados de la población en estudio (sector, país, región), como ha quedado demostrado en el estudio de seguimiento. La evolución de una población explica en parte la evolución de los indicadores financieros de las empresas fracasadas o próximas al fracaso. De ahí la importancia de analizar estas últimas en el contexto sectorial del conjunto en el mismo periodo.

Los análisis empíricos realizados en esta tercera parte tienen algunas limitaciones. Aunque el número de empresas total estudiado es elevado, el número de empresas fallidas es pequeño en alguno de los años estudiados, y especialmente en el análisis de seguimiento. Nuestra muestra está formada por toda la población de empresas disponible en la base de datos para el sector y las zonas geográficas seleccionados, pero sería conveniente que la metodología propuesta se contrastara en otros sectores, zonas geográficas y periodos muestrales de seguimiento más largos, para obtener resultados más generales sobre la utilidad y aplicaciones del modelo. También debemos reconocer como limitación que en esta tercera parte de la tesis solo se consideran indicadores financieros, tomados de la información contable de las empresas, a pesar de que estamos de acuerdo con Cook *et al.* (2012), en que los factores importantes para la recuperación de empresas en dificultades no son solo la gestión financiera sino también la gestión comercial y la competitividad. Pero es difícil conseguir datos sobre estas dos últimas cuestiones para incluirlos en los modelos y debemos confiar en que la buena gestión en ambos aspectos desemboque en mejores datos financieros.

4. CONCLUSIONES FINALES

Al final de cada una de las tres partes en que se ha dividido esta tesis, hemos dedicado un apartado de conclusiones a situar los resultados en el contexto de la literatura previa y a destacar la contribución de nuestros análisis teóricos y empíricos a las respectivas corrientes de investigación dentro de la literatura sobre fracaso empresarial. No vamos a repetir en este momento cada una de esas explicaciones, sino solo a recoger el conjunto de resultados de forma esquemática, relacionando las contribuciones con las posibilidades de extensión futura de este trabajo de investigación.

Parte I

- La teoría que subyace a los modelos de fracaso empresarial es un marco conceptual en construcción, asentado sobre unos pocos modelos y teorías generales desarrollados en los inicios de esta línea de investigación por autores como Beaver y Argenti.
- Esta teoría ha sido incapaz de identificar indicadores que sirvan como variables en cualquier modelo sobre fracaso, lo que ha llevado a los autores a emplear regularidad estadística, donde la teoría no llega, para plantear modelos eficaces en periodos, zonas geográficas o poblaciones concretas, tal como reconoce Altman.
- El concepto de empresa fracasada más objetivo es el que toma las empresas declaradas en situación de quiebra, suspensión de pagos o similar, aunque esto plantea serios problemas a los modelos para distinguir entre las empresas declaradas fracasadas y las que están en serio riesgo de fracaso sin haber sido declaradas como tales.
- Nuevas metodologías de estudio o nuevas versiones de las anteriores se van proponiendo para eliminar limitaciones.
- Nuestro análisis de frecuencias de variables más significativas en la literatura previa ofrece resultados consistentes con los escasos trabajos similares. Las variables más frecuentes se refieren a endeudamiento,

equilibrio económico financiero, rentabilidad, estructura económica retención de beneficios gastos financieros y generación de recursos.

- Ante la limitación detectada en este tipo de análisis, a causa de la dispersión en las especificaciones de los ratios, contribuimos con un análisis de frecuencias por rasgos económicos subyacentes que corrige el sesgo. Los rasgos económicos más significativos, de acuerdo con nuestros resultados, son: rentabilidad, endeudamiento, equilibrio económico-financiero, estructura económica, márgenes y rotaciones.
- Identificamos en general dos problemas en los modelos presentes en la literatura previa: variables dicotómicas (fracaso/no fracaso) y escasa representación de indicadores ajenos a la empresa que se evalúa.

Parte II

- Nuestro estudio confirma que los mismos rasgos identificados como significativos en empresas de cualquier tamaño subyacen a las variables explicativas en pymes, aunque en un orden de importancia distinto, siendo los indicadores más importantes: endeudamiento, estructura económica y rentabilidad.
- Debido a la dispersión y a la distribución de las variables, es preferible utilizar diferencias en medianas que diferencias en medias para la selección estadística de variables. Este es un estudio que no hemos encontrado en la literatura previa, y constituye una contribución de esta tesis.
- La winsorización aproxima la selección de variables con diferencias en medias a la selección con diferencias en medianas.
- Las variables seleccionadas con diferencias en medias obtienen mejores resultados en metodologías paramétricas (análisis discriminante lineal, análisis discriminante cuadrático, logit y probit); mientras que las variables seleccionadas con diferencias en medianas obtienen mejores resultados en metodologías no paramétricas (análisis discriminante logístico, análisis discriminante del K-ésimo vecino más próximo y DEA). Las metodologías no paramétricas obtienen mejores resultados de evaluación.

- La distribución de los datos sesga en mayor medida las metodologías paramétricas que las no paramétricas.
- Los resultados obtenidos con DEA son consistentes con la escasa literatura previa sobre este modelo aplicado a fracaso empresarial, esto es, similares a los obtenidos con las metodologías más comunes (análisis discriminante paramétricos y análisis binomiales).
- La incorporación de la variable cualitativa sobre incidencias de pago mejora el poder discriminante de todos los modelos de análisis discriminante y binomiales utilizados.

Parte III

- Ante las limitaciones encontradas en los métodos de evaluación de quiebra analizados en la literatura previa (primera parte), y comprobadas en los métodos aplicados a nuestra muestra de estudio en la segunda parte, desarrollamos la metodología que se propone en la tercera parte, y que constituye la principal contribución de esta tesis. Por un lado, las clasificaciones dicotómicas (fracasadas / no fracasadas) no permiten medir la distancia de las empresas al fracaso. Las holguras del DEA sí dan una medida de esa distancia, aunque este modelo plantea limitaciones para su implementación en muestras con un número elevado de empresas. Por otro lado, la falta de consideración de variables del entorno, sesga los resultados de los modelos en función de la situación macroeconómica de la población en estudio.
- La metodología propuesta para el cálculo de *scores*, fronteras de fracaso y distancias al fracaso, a partir de diferencias de percentiles, resulta efectiva para seleccionar variables discriminantes y para clasificar a las empresas como fracasadas o no fracasadas, con resultados equivalentes o mejores que las metodologías tradicionales de análisis discriminante, las binomiales y el DEA.
- Además, la metodología propuesta calcula puntuaciones por empresa y por grupos de empresa, que permiten hallar su distancia a la frontera de fracaso, pero también a otros dos niveles de referencia: los valores centrales de la población y la frontera espejo que determina los valores de las mejores empresas dentro de la población.

- Al hacer los cálculos por diferencia a los valores de la población, esta metodología elimina el efecto que sobre *scores* y distancias tienen los cambios de la población debidos a factores que la afectan en su conjunto (tales como los factores macroeconómicos o los factores sectoriales si, como en nuestro caso, la población se encuadra en un sector concreto).
- Los ratios más discriminantes en pymes del sector de la construcción, aplicando nuestra metodología, son: generación de fondos, rentabilidad económica y endeudamiento.
- Con diferencias de percentiles iguales o superiores a 0,25, con un año de antelación, se obtienen resultados de predicción del fracaso superiores al 85% y de predicción de no fracaso superiores al 95%, para la muestra analizada de empresas españolas de la construcción de pequeño tamaño.
- Las variables más discriminantes, de acuerdo con nuestra metodología, son también los indicadores financieros que una empresa en riesgo de quiebra debe mejorar si entra en un proceso de reorganización, excepto el endeudamiento.
- El endeudamiento es un indicador financiero con elevado poder discriminante del riesgo de fracaso, pero una vez que la empresa se encuentra en situación de fracaso o de riesgo muy elevado, no es un buen indicador de la mejora de la empresa que se reorganiza.

En cuanto a las extensiones futuras del trabajo, por un lado, se puede señalar que nuestros resultados serían más consistentes si somos capaces de mejorar las muestras analizadas. Esto significa que:

- incrementar el número de trabajos de literatura empírica previa analizados nos puede ayudar a mejorar la percepción sobre variables y rasgos relevantes y cómo la importancia de algunos de ellos va cambiando en el tiempo;
- aplicar los contrastes estadísticos de selección de variables y las metodologías de evaluación del fracaso a sectores diferentes, y a poblaciones más amplias con un número de empresas fracasadas más elevado, contribuiría a la generalización de nuestros resultados;

Pero las futuras extensiones del trabajo que más pueden contribuir a la línea de investigación sobre fracaso empresarial son las aplicaciones de la metodología de fronteras y *scores* basados en diferencias de percentiles a numerosos tipos de estudios. Algunas posibilidades de la comparación de la situación de fracaso / no fracaso serían:

- poblaciones de empresas en distintas zonas geográficas (CCAA, países,...);
- poblaciones de empresas de distintos tamaños (ej. cotizadas, no cotizadas);
- poblaciones con características homogéneas en cuanto a su naturaleza o actividad, por ejemplo, según la distribución de la propiedad (familiares, con inversores externos);
- poblaciones con diferentes niveles de variables macroeconómicas, o bien la misma población según la evolución en el tiempo de determinados factores macroeconómicos.

Finalmente, las puntuaciones de frontera, medidas con los *scores*, y las distancias a la frontera de quiebra, se pueden utilizar en numerosos contrastes empíricos, con variables contables y no contables de las propias empresas, sectores o del entorno. Por ejemplo, el efecto del cambio de la normativa sobre fracaso en nuestro país puede haber cambiado el comportamiento de declaración de fracaso de las empresas y la metodología propuesta permite medir el posible cambio que se haya producido en ese comportamiento.

BIBLIOGRAFÍA

- Abad, C., Arquero, J.L. y Jiménez, SM. (2004). Procesos de fracaso empresarial. Identificación y contrastación empírica. *XI Encuentro de Profesores Universitarios de Contabilidad*.
- Acosta, E. y Fernández, F. (2007). Predicción del fracaso empresarial mediante el uso de algoritmos genéticos. *X Encuentro de Economía Aplicada*, Logroño, 14-15-16 de junio <http://www.unirioja.es/dptos/dee> DEPARTAMENTO DE ECONOMÍA Y EMPRESA.
- Ahn, B.S., Cho, S.S. y Kim, C.Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, 18(2), pp. 65-74.
- Albert, A., y Lesaffre, E. (1986). Multiple group logistic discrimination, *Computers and Mathematics with Applications*, 12(2), Part A, pp. 209-224.
- Aleman, J. y Monllau, J. (1997). Disponibilidad y defectos en la presentación de las cuentas anuales en el registro mercantil, *Técnica Contable*, 581, pp. 381-393.
- Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance*, 23(4), pp. 568-609.
- Altman, E.I. (1981). *Financial Handbook*. New York. John Wiley & Sons.
- Altman, E.I. (1983). *Corporate Financial Distress*. Chichester. John Wiley & Sons.
- Altman, E.I. (1993). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy*. New York. John Wiley & Sons.
- Altman, E.I., Haldeman, R. y Narayanan, P. (1977). Zeta analysis: a new model to identify bankruptcy risk of corporations, *Journal of Banking and Finance*, 1(1), June, pp. 29-54.
- Altman, E.I., Marco, G. y Varetto, F. (1994). Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience), *Journal of Banking and Finance*, 18(3), pp. 505-529.
- Altman, E.I. y Saunders, A. (1998). Credit risk measurement: developments over the last 20 years, *Journal of Banking and Finance*, 21(11-12), December, pp. 1721-1742.
- Altman, E.I. y Sabato, G. (2005). Effects of the new Basel capital accord on bank capital requirements for SMEs, *Journal of Financial Services Research*, 28(1-3), pp. 15-42.

- Altman, E.I. y Sabato, G. (2007). Modeling credit risk for SMEs: evidence from the U.S. market, *Abacus*, 43(3), pp. 332-357.
- Altman, E.I., Sabato, G. y Wilson, N. (2008). The value of qualitative information in SME risk management, *Working Paper*. Leonard N. Stern School of Business, New York University.
- Argenti, J. (1976). *Corporate collapse: the causes and symptoms*. New York. John Wiley & Sons.
- Arquero, J.L., Abad, M.C. y Jiménez, S.M. (2008). Procesos de fracaso empresarial en pymes, Identificación y contrastación empírica, *Revista Internacional de la Pequeña y Mediana Empresa*, 1(2), pp. 64-77.
- Atiya, A.F. (2001). Bankruptcy prediction for credit risk using neural networks: a survey and new results, *IEEE Transactions on Neural Networks*, 12(4), pp. 929-935.
- Aziz, M.A. y Dar, H.A., (2006). Predicting corporate bankruptcy: Where we stand? *Corporate Governance*, 6(1), pp.18-33.
- Baixaui, J.S. y Módica-Milo, A. (2010). The bias of unhealthy SMEs in bankruptcy prediction models, *Journal of Small Business and Enterprise Development*, 17(1), pp. 60-77.
- Balcaen, S. y Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems, *British Accounting Review*, 38(1), pp. 63-93.
- Banker, R.D., Charnes, A. y Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30 (9), pp. 1078-1092.
- Baum, C.F. (2006). *An introduction to modern econometrics using Stata*. Texas(USA). Stata Press.
- Barniv, R., Agarwal, A. y Leach, R. (1997). Predicting the outcome following bankruptcy filing: a three state classification using neural networks, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 6(3), pp. 177-194.
- Beaver, W.H. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research*, 4, January, pp. 71-111.
- Beaver, W.H. (1968). Alternative accounting measures and predictors of failure, *The Accounting Review*, 43(1), January, pp. 113-122.
- Beaver, W.H., Correia, M. y McNichols, M. (2009). Have changes in financial reporting attributes impaired informativeness? Evidence from the ability of financial ratios to predict bankruptcy, *Rock Center for Corporate Governance Working Paper No. 13*, Stanford University, December. Available at SSRN: <http://ssrn.com/abstract=1341305>.

- Beaver, W.H., McNichols, M. y Rhie, J. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy, *Review of Accounting Studies*, 10(1), pp. 93-122.
- Bechetti, L. y Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms, *Journal of Banking and Finance*, 27(11), pp. 2099-2120.
- Behr, P. y Güttler, A. (2007). Credit risk assessment and relationship lending: an empirical analysis of German small and medium-sized enterprises, *Journal of Small Business Management*, 45(2), April, pp. 194-213.
- Bell, T.B. (1997). Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 6(3), September, pp. 249-264.
- Bell, T.B., Ribar, G.S. y Verchio, J. (1990). Neural nets versus logistic regression: a comparison of each model's ability to predict commercial bank failures. En Srivastava, R.P. (ed.) *Auditing Symposium on Auditing Problems*, pp. 29-53.
- Bellovary, J.L., Giacomino, D.E. y Akers, M.D. (2007). A review of bankruptcy prediction studies: 1930 to present, *Journal of Financial Education*, 33(Winter), pp. 1-43.
- Berkson, J. (1944). Application of the logistic function to bio-assay, *Journal of the American Statistical Association*, 39(227), September, pp. 357-365.
- Bhargava, M., Dubelaar, C. y Scott, T. (1998). Predicting bankruptcy in the retail sector: an examination of the validity of key measures of performance, *Journal of Retailing and Consumer Services*, 5(2), pp. 105-117.
- Blanco, A., Irimia, A. y Oliver, M.D. (2012). The prediction of bankruptcy of small firms in the UK using logistic regression, *Análisis Financiero*, 118, pp. 32-40.
- Bliss, C.I. (1934). The method of probits, *Science*, 79, January, pp. 38-39.
- Blum, M. (1974). Failing company discriminant analysis, *Journal of Accounting Research*, 12(1), pp. 1-25.
- Bonsón Ponte, E., Escobar Rodríguez, T. y Martín Zamora, M.P. (1997a). Decision tree induction systems. Applications in accounting and finance. En E. Bonsón Ponte y G. Sierra Molina (ed): *Intelligent Technologies in Accounting and Business*. Proceedings of the III International Meeting on Artificial Intelligence in Accounting, Finance and Tax. Huelva.
- Bonsón Ponte, E., Escobar Rodríguez, T. y Martín Zamora, M.P. (1997b). Sistemas de inducción de árboles de decisión: utilidad en el análisis de crisis bancarias. *Ciberconta. Revista electrónica de Contabilidad*. Universidad de Zaragoza, Departamento de Contabilidad y Finanzas. (Disponible en <http://ciberconta.unizar.es/Biblioteca/0007/árboles.html>).

- Boussofiane, A., Dyson, R.G. y Thanassoulis, E. (1991). Applied data envelopment analysis, *European Journal of Operational Research*, 52(1), pp. 1-15.
- Calvo-Flores, A., García, D. y Madrid, A. (2006). Tamaño, antigüedad y fracaso empresarial. *Working Paper*. Universidad Politécnica de Cartagena.
- Canbas, S., Cabuk, A. y Kilic, S.B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structure: the Turkish case, *European Journal of Operational Research*, 166(2), October, pp. 528-546.
- Casey, C. y Bartczak, N. (1985). Using operating cash flow data to predict financial distress- some extensions, *Journal of Accounting Research*, 23(1), pp. 384-401.
- Chang, D.S. y Kuo, Y.C. (2008). An approach for the two-group discriminant analysis: an application of DEA, *Mathematical and Computer Modelling*, 47(9-10), May, pp. 970-981.
- Charalambous, C., Charitou, A. y Kaorou, F. (2000). Comparative analysis of artificial neural network models: application in bankruptcy prediction, *Annals of Operation Research*, 99(4), pp. 403-425.
- Charnes, A., Cooper, W.W. y Rhodes, E. (1978). Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2(6), pp 429-444.
- Charnes, A., Cooper, W.W. y Rhodes, E. (1979). Short communication: measuring the efficiency of decision making units, *European Journal of Operational Research*, 3, p. 339.
- Cielen, A., Peeters, L. y Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis, *European Journal of Operational Research*, 154(2), April, pp. 526-532.
- Cochran, B. (1986). Small business mortality rates: a review of the literature, *Journal of Small Business Management*, 19(4), pp. 50-9.
- Coll V. y Blasco, O.M. (2006). *Evaluación de la eficiencia mediante el análisis envolvente de datos. Introducción a los modelos básicos*. Universidad de Valencia.
- Collins, R.A. y Green, R.D. (1982). Statistical methods for bankruptcy forecasting, *Journal of Economics and Business*, 34(4), pp. 349-354.
- Cook, G.A.S., Pandit, N.R. y Milman, D. (2012). A resource-based analysis of bankruptcy law, SMEs and corporate recovery, *International Small Business Journal*, 30(3), pp. 275-293.
- Correa, A., Acosta, M. y González, A.L. (2003). La insolvencia empresarial: un análisis empírico para la pequeña y mediana empresa, *Revista de Contabilidad*, 6(12), pp. 47-79.

- Crespo Domínguez, M.A. (2000). Una aproximación a la predicción del fracaso empresarial mediante redes neuronales, *IX Encuentro de Profesores Universitarios de Contabilidad*, Las Palmas de Gran Canaria, pp. 591-607.
- Dambolena, I.G. y Khoury, S.J. (1980). Ratio stability and corporate failure, *Journal of Finance*, 35(4), September, pp. 1017-1026.
- Dannreuther, C. y Kessler, O. (2010). *Small firm finance and the political economy of risk*. London. Routledge.
- Daubie, M. y Meskens, N. (2002). Business failure prediction: a review and analysis of the literature, en Zopounidis, C. (Ed.) *New trends in banking management*, Physica-Verlag, pp. 71-86.
- Davydenko, S.A. (2007). When do firms default? A study of the default boundary. *AFA 2009 San Francisco Meetings Paper; EFA 2005 Moscow Meetings Paper; WFA 2006 Keystone Meetings Paper*, August. Available at SSRN: <http://ssrn.com/abstract=672343>.
- Davydenko, S.A. y Franks, J.R. (2008). Do bankruptcy codes matter? A study of defaults in France, Germany, and the U.K., *The Journal of Finance*, 63(2), April, pp. 565-608.
- De Andrés Suárez, J. (2000). Técnicas de inteligencia artificial aplicadas al análisis de la solvencia empresarial. *Documento de Trabajo núm. 206*, Universidad de Oviedo, Facultad de Ciencias Económicas.
- De Andrés Suárez, J. (2001). Statistical techniques vs. SEE5 algorithm. An application to a small business environment, *The International Journal of Digital Accounting Research*, 1(2), July, pp. 153-179.
- De Andrés Sánchez, J. (2005). Comparativa de métodos de predicción de la quiebra: redes neuronales artificiales vs. métodos estadísticos multivariantes, *Partida Doble*, 168, julio-agosto, pp. 105-113.
- De la Torre, J.M., Gómez, M.E. y Román, I. (2005). Análisis de sensibilidad temporal de los modelos de predicción de solvencia: una aplicación a las pymes industriales, *XIII Congreso AECA, Armonización y gobierno de la diversidad*, 22 a 24 de septiembre, Oviedo (recurso electrónico).
- De Miguel, L.J., Revilla, E., Rodríguez, J.M. y Cano, J.M. (1993). A comparison between statistical and neural network based methods for predicting bank failures, *Proceedings of the IIIth International Workshop on Artificial Intelligence in Economics and Management*, Portland (USA).
- Deakin, E.B. (1972). A discriminant analysis of predictors of business failure, *Journal of Accounting Research*, 10(1), pp. 167-179.
- Deakin, E.B. (1976). Distributions of financial accounting ratios: some empirical evidence, *The Accounting Review*, 51(1), January, pp. 90-96.

- Del Rey Martínez, E. (1996). Bankruptcy prediction in non-finance companies: An application based on artificial neural network models, En Sierra Molina, G. y Bonsón Ponte, E. (Eds.): *Intelligent Systems in Accounting and Finance*, Huelva, pp. 253-272.
- Demyanyk, Y. y Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods, *Omega*, 38(5), October, pp. 315-324.
- Dewaelheyns, N. y Van Hulle, C. (2004). The impact of business groups on bankruptcy prediction modeling, *Tijdschrift voor Economie en Management*, 49(4), pp. 623-645.
- Dewaelheyns, N. y Van Hulle, C. (2006). Corporate failure prediction modeling: Distorted by business groups' internal capital markets?, *Journal of Business Finance and Accounting*, 33(5-6), pp. 909-931.
- Dietsch, M. y Petey, J. (2002). The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements, *Journal of Banking and Finance*, 26(2-3), pp. 303-322.
- Dietsch, M. y Petey, J. (2004). Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs, *Journal of Banking and Finance*, 28(4), pp. 773-788.
- Dimitras, A., Zanakis S. y Zopounidis C. (1996). A survey of business failures with an emphasis on failure prediction methods and industrial applications, *European Journal of Operational Research*, 90(3), pp. 487-513.
- Dutta, S. y Shekhar, S. (1992). Bond rating: A non conservative application of neural networks. En *Neural networks in finance and investing*. Chicago: Probus Publishing, pp. 443-450.
- Edmister, R.O. (1972). An empirical test of financial ratio analysis for small business failure prediction, *Journal of Financial and Quantitative Analysis*, 7(2), March, pp. 1477-1493.
- Edmister, R.O. (1988). Combining human credit analysis and numerical credit scoring for business failure prediction, *Akron Business and Economic Review*, 19(3), pp. 6-14.
- Eisenbeis, R. (1977). Pitfalls in the application of discriminant analysis in business, finance and economics, *The Journal of Finance*, 32(3), June, pp. 875-900.
- Elam, R. (1975). The effect of lease data on the predictive ability of financial ratios, *The Accounting Review*, 50(1), January, pp. 25-43.
- Farrell, M.J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society, Series A*, 120, Part III, pp. 253-290.
- Fernández, E. y Olmeda, I. (1995). Bankruptcy prediction with artificial neural networks, *Lecture Notes in Computer Science*, 930, pp. 1142-1146.

- Ferrando, M. y Blanco, F. (1998). La previsión del fracaso empresarial en la comunidad valenciana: aplicación de los modelos discriminante y logit, *Revista Española de Financiación y Contabilidad*, 27(95), abril-junio, pp. 499-540.
- Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), September, pp. 179-188.
- Fletcher, D. y Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data, *Information and Management*, 24(3), March, pp. 159-167.
- Franks, J.R. y Torous, W.N. (1992). Lessons from a comparison of US and UK insolvency codes, *Oxford Review of Economic Policy*, 8(3), pp. 70-82.
- Friedman, L., y Sinuany-Stern, Z. (1998). Combining ranking scales and selecting variables in the DEA context: The case of industrial branches, *Computers and Operations Research*, 25(9), pp. 781-791.
- Frydman, H., Altman, E.I. y Kao, D. (1985). Introducing recursive partitioning for financial classification: The case of financial distress, *The Journal of Finance*, 40(1), March, pp. 269-291.
- Gabás Trigo, F. (1990). *Técnicas actuales de análisis contable: evaluación de la solvencia empresarial*. Madrid. Instituto de Contabilidad y Auditoría de Cuentas. Ministerio de Economía y Hacienda.
- Gallego, A.M., Gómez, J.C. y Yáñez, L. (1997a). Modelos de predicción de quiebras en empresas no financieras, *Actualidad Financiera*, 2(5), mayo, pp. 3-14.
- Gallego, A.M., Gómez, J.C. y Yáñez L. (1997b). Predicción de quiebras en empresas no financieras: una aplicación del modelo logit, *Revista Europea de Dirección y Economía de la Empresa*, 6(3), pp. 129-138.
- Gandía, J.L., López Gracia, J. y Molina, R. (1995). *Estudio empírico de la solvencia empresarial en la Comunidad Valenciana*. Valencia. Instituto Valenciano de Investigaciones Económicas, Junio.
- García, D., Arqués, A. y Calvo-Flores, A. (1995). Un modelo discriminante para evaluar el riesgo bancario en los créditos a empresas, *Revista Española de Financiación y Contabilidad*, 24(82), enero-marzo, pp. 175-200.
- García-Ayuso, M. (1995). La necesidad de llevar a cabo un replanteamiento de la investigación en materia de análisis de la información financiera, *Análisis Financiero*, 66, pp. 36-61.
- Gazengel, A. y Thomas, P. (1992). Les défaillances d'entreprises. *Les Cahiers de Recherche*, 105, 47 p., École Supérieure de Commerce de Paris.
- Gentry, J., Newbold, P. y Whitford, D. (1985). Classifying bankrupt firms with funds flow components, *Journal of Accounting Research*, 23(1), pp. 146-159.

- Gilbert, L.R., Menon, K. y Schwartx, K.B. (1990). Predicting bankruptcy for firms in financial distress, *Journal of Business, Finance and Accounting*, 17(1), pp. 161-171.
- Gill de Albornoz, B. y Giner, B. (2010). El fracaso empresarial en los sectores inmobiliario y de la construcción. 2008-2009. En Gill de Albornoz, B (Dir.), Fernández de Guevara, J, Giner, B. y Martínez, L. *Las empresas del sector de la construcción e inmobiliario en España: del boom a la recesión económica*, Madrid, Funcas, pp. 173-231.
- Golany, B. y Roll, Y. (1989). An application procedure for DEA, *Omega*, 17 (3), pp. 237-250.
- Gombola, M.J. y Ketz, J.E. (1983). A note on cash flow and classification patterns of financial ratios, *Accounting Research*, 58(1), January, pp. 105-114.
- Gómez, M.A., Torre, J.M. y Román, I. (2008). Análisis de sensibilidad temporal en los modelos de predicción de insolvencia: una aplicación a las pymes industriales, *Revista Española de Financiación y Contabilidad*, 37(137), enero-marzo, pp. 85-111.
- Graveline, J. y Kokalari, M. (2008). Credit risk. *Working Paper*, The Research Foundation of CFA Institute, November.
- Greenstein, M.M. y Welsh, M.J. (1996). Bankruptcy prediction using ex-ante neural networks and realistically proportioned testing sets. En Sierra Molina, G. y Bonsón Ponte, E. (Eds.): *Intelligent systems in accounting and finance*, Huelva, pp. 187-212.
- Grice, J.S. e Ingram, R.W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model, *Journal of Business Research*, 54(1), pp. 53-61.
- Grunert, J., Norden, L. y Weber, M. (2005). The role of non-financial factors in internal credit ratings, *Journal of Banking and Finance*, 29(2), pp. 509-531.
- Hair, J.F., Anderson, R.E., Tatham, R.L. y Black, W.C. (1999). *Análisis multivariante*. Madrid. Prentice-Hall.
- Hayden, E. (2003). Are credit scoring models sensitive with respect to default definitions? Evidence from the Australian market, *Dissertation Paper*, Department of Business Administration, Univesity of Vienna, Austria, pp.1-43.
- Headd, B. (2003). Redefining business success: Distinguishing between closure and failure, *Small Business Economics*, 21(1), pp. 51-61.
- Hill, N.T., Perry, S.E. y Andes, S. (1996). Evaluating firms in financial distress: An event history analysis, *Journal of Applied Business Research*, 13(13), pp. 60-71.
- Hillegeist, S.A., Keating, E.K., Cram, D.P. y Lundstedt, K.G. (2004). Assessing the Probability of Bankruptcy. *Review of Accounting Studies*, 9(1), pp. 5-34.

- Holder, M. (1984). *Le score de l'entreprise*. París. Nouvelles Editions Fiduciaires.
- Jacobson, T., Kindell, R., Lindé, J. y Roszbach, K. (2008). Firm default and aggregate fluctuations. *Working Paper*, Sveriges Riskbank, nº. 226, September. Available at SSRN: <http://ssrn.com/abstract=1471254>.
- Jenkins, L. y Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis, *European Journal of Operational Research*, 147(1), pp. 51-61.
- Jin, J.Y., Kanagaretnam, K. y Lobo, G.J. (2011). Ability of accounting and audit quality variables to predict bank failure during the financial crisis, *Journal of Banking and Finance*, 35(11), pp. 2811-2819.
- Jones, F.L. (1987). Current techniques in bankruptcy prediction, *Journal of Accounting Literature*, 6, pp. 131-164.
- Jones, S. y Hensher, D.A. (2004). Predicting firm financial distress: A mixed logit model, *The Accounting Review*, 79(4), pp. 1011-1038.
- Jones, S. y Hensher, D.A. (2008). *Advances in credit risk modelling and corporate bankruptcy prediction*. Cambridge University Press.
- Kaski, S., Sinkkonen, J. y Peltonen, J. (2001). Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks*, 12 (4), pp. 936-947.
- Keasey, K. y Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: a test of Argenti's hypothesis, *Journal of Business, Finance and Accounting*, 14(3), Autumn, pp. 335-354.
- Keasey, K. y Watson, R. (1988). The non-submission of accounts and small company financial failure prediction, *Accounting and Business Research*, 19(73), Winter, pp. 47-54.
- Keasey, K. y Watson, R. (1991). Financial distress prediction models: A review of their usefulness, *British Journal of Management*, 2(2), July, pp. 89-102.
- Ketz, J.E. (1978). The effect of general price-level adjustments on the predictability of financial ratios, *Journal of Accounting Research*, 16 supplement, pp. 273-284.
- Kiviluoto, K. (1998). Predicting bankruptcies with self organizing map, *Neurocomputing*, 21(1-3), pp. 191-201.
- Koh, H.C. (1991). Model predictions and auditor assessment of going concern status, *Accounting and Business Research*, 21(84), pp. 331-338.
- Koh, H.C. y Tan, S.S. (1999). A neural network approach to the prediction of going concern status, *Accounting and Business Research*, 29(3), pp. 211-216.

- Kuo, Y.C. (2007). *The data envelopment models for the application of two-group discriminant analysis*. Tesis Doctoral.
- Labatut, G., Pozuelo, J. y Veres, E.J. (2009). Modelización temporal de los ratios contables en la detección del fracaso empresarial de la pyme española, *Revista Española de Financiación y Contabilidad*, 38(143), julio-septiembre, pp. 423-448.
- Lachenbruch, P.A. (1975). *Discriminant Analysis*. New York. Hafner Press.
- Lacher, R.C., Coats, P.K., Sharma, S.C. y Fant, L.F. (1995). A neural network for classifying the financial health of a firm, *European Journal of Operational Research*, 85(1), pp. 53-65.
- Laffarga, J., Martín, J.L. y Vázquez, M.J. (1985). El análisis de la solvencia de las instituciones bancarias: propuesta de una metodología y aplicaciones a la banca española, *Esic-Market*, 48 (2º trim.), pp. 51-73.
- Laffarga J., Martín, J.L. y Vázquez, M.J. (1986a). El pronóstico a corto plazo del fracaso en las instituciones bancarias: metodología y aplicaciones a la banca española, *Esic-Market*, 53, (3º trim.), pp. 59-116.
- Laffarga J., Martín, J.L. y Vázquez, M.J. (1986b). El pronóstico a largo plazo del fracaso en las instituciones bancarias: metodología y aplicaciones al caso español, *Esic-Market*, 54, (4º trim.), pp. 113-167.
- Laffarga J., Martín, J.L. y Vázquez, M.J. (1987). Predicción de la crisis bancaria española: la comparación entre el análisis logit y el análisis discriminante, *Cuadernos de Investigación Contable*, 1(1), otoño, pp. 103-110.
- Laffarga J., Martín, J.L. y Vázquez, M.J. (1991). La predicción de la quiebra bancaria: el caso español, *Revista Española de Financiación y Contabilidad*, 20(66), enero-marzo, pp. 151-163.
- Laitinen, E.K. (1993). Financial predictors for different phases of the failure process, *Omega International Journal of Management Science*, 21(2), pp. 215-228.
- Laitinen, E.K. (2008). Data system for assessing probability of failure in SME reorganization, *Industrial Management and Data Systems*, 108(7), pp. 849-866.
- Laitinen, T. y Kankaanpää, M. (1999). Comparative analysis of failure prediction methods: The Finnish case, *The European Accounting Review*, 8(1), pp.67-92.
- Lee, K., Booth, D. y Alam, P. (2005). A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms, *Expert Systems with Applications*, 29(1), July, pp. 1-16.
- Lee, S.H. y Urrutia, J.L. (1996). Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models, *The Journal of Risk and Insurance*, 63(1), pp. 121-130.

- Lennox, C. (1999). Identifying failing companies: A re-evaluation of the logit, probit and DA approaches, *Journal of Economics and Business*, 51(4), July, pp. 347-364.
- Leshno, M. y Spector, Y. (1996). Neural network prediction analysis: The bankruptcy case, *Neurocomputing*, 10(2), March, pp. 125-147.
- Lev, B. (1971). Financial failure and informational decomposition measures. In *Accounting in perspective contributions to accounting thoughts by other disciplines*, edited by R. R. Sterling and W. F. Bentz. Cincinnati: Southwestern Publishing Co., pp. 102-111.
- Libby, R. (1975). Accounting ratios and the prediction of failure: Some behavioural evidence, *Journal of Accounting Research*, 13(1), Spring, pp. 150-161.
- Lincoln, M. (1984). An empirical study of the usefulness of accounting ratios to describe levels of insolvency risk, *Journal of Banking and Finance*, 8(2), pp. 321-340.
- Liou, D.K. y Smith, M. (2006). Macroeconomic variables in the identification of financial distress, *Working Paper*, May, Available at SSRN: <http://ssrn.com/abstract=900284>.
- Lizarraga Dallo, F. (1997). Utilidad de la información contable en el proceso de fracaso: análisis del sector industrial de la mediana empresa, *Revista Española de Financiación y Contabilidad*, 26(93), octubre-diciembre, pp. 871-915.
- Lizarraga Dallo, F. (1998). Modelos de predicción del fracaso empresarial: ¿funciona entre nuestras empresas el modelo de Altman de 1968?, *Revista de Contabilidad*, 1(1), enero-junio, pp. 137-164.
- Lo, A.W. (1986). Logic versus discriminant analysis, *Journal of Econometrics*, 31(2), pp. 151-178.
- López, A., De Andrés, J. y Rodríguez, E. (1997). Omisiones e incoherencias en las cuentas anuales: su influencia en las bases de datos con información contable, *IX Congreso AECA*, Salamanca, 25-27 septiembre, pp. 373-396.
- López, E. y Flórez, R. (1999). El análisis de solvencia empresarial utilizando redes neuronales autoasociativas: el modelo Koh-León, *Proceedings of the VI International Meeting on Advances in Computational Management*, Reus.
- López, E. y Flórez, R. (2000). Aplicación de dos modelos de redes neuronales artificiales para el análisis económico-financiero empresarial, *Revista Europea de Dirección y Economía de la Empresa*, 9(2), pp. 139-164.
- López, J., Gandía, J.L. y Molina, R. (1998). La suspensión de pagos en las pymes: una aproximación empírica, *Revista Española de Financiación y Contabilidad*, 27(94), enero-marzo, pp. 71-97.

- López, D., Moreno, J. y Rodríguez, P. (1994). Modelos de predicción del fracaso empresarial: aplicación a entidades de seguros en España, *Esic-Market*, 84, pp. 83-125.
- Madrid, A. y García, D. (2006). Factores que explican el fracaso empresarial en la pyme, *Gestión: Revista de Economía*, 36, marzo-junio, pp. 5-9.
- Manzanegue, M., Benegas, R. y García D. (2010). Diferentes procesos de fracaso empresarial: Un análisis dinámico a través de la aplicación de técnicas estadísticas clúster, *Revista Europea de Dirección y Economía de la Empresa*, 19(3), pp. 67-88.
- Mar Molinero, C. y Ezzamel, M. (1991). Multidimensional scaling applied to corporate failure, *Omega*, 19(4), pp. 259-274.
- Mar, C. y Serrano, C. (2001). Bank failure: A multidimensional scaling approach, *The European Journal of Finance*, 7(2), pp. 165-183.
- Marais, M., Patell, J. y Wolfson, M. (1984). The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications, *Journal of Accounting Research*, 22(1), pp. 87-118.
- Marose, R.A. (1992). A financial neural network application. En *Neural Networks in Finance and Investing*. Chicago: Probus Publishing, pp. 50-53.
- Martin, D. (1977). Early warning of bank failure, *Journal of Banking and Finance*, 1(3), pp. 249-276.
- Martínez, I. (1996). Forecasting company failure: Neural approach versus discriminant analysis: An application to Spanish insurance companies. En Sierra Molina, G. y Bonsón Ponte, E. (Eds.): *Intelligent systems in accounting and finance*, Huelva, pp. 169-185.
- Martínez, C., Navarro, M.V. y Sanz, F. (1989). Selección y explotación de los sistemas de alarma y prevención de quiebra, *Investigaciones Económicas*, (supl.) 13(3), pp. 135-141.
- Mateev, M., Poutziouris, P. y Ivanov, K. (2013). On the determinants of SME capital structure in Central and Eastern Europe: A dynamic panel analysis, *Research in International Business and Finance*, 27(1), January, pp. 28-51.
- McDonald, B.D. y Morris, M.H. (1984). The statistical validity of the ratio method in financial analysis: An empirical examination, *Journal of Business, Finance and Accounting*, 11(1), Spring, pp. 89-97.
- McGahan, A.M. y Porter, M.E. (1997). How much does industry matter, really?, *Strategic Management Journal*, 18, Summer, pp. 15-30.
- McGurr, P.T. y DeVaney, S.A. (1998). Predicting business failure of retail firms: An analysis using mixed industry models, *Journal of Business Research*, 43(3), pp. 169-176.

- McKee, T.E. (1990). Evaluation of enterprise continuity status via neural networks, Abstracts of the *Thirteenth Annual Congress of the European Accounting Association*, 72.
- McKee, T.E. (2000). Developing a bankruptcy prediction model via rough sets theory, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 9(3), September, pp. 159-173.
- McLachlan, G.J. (2004). *Discriminant analysis and statistical pattern recognition*. New York. John Wiley.
- Mensah, Y.M. (1984). An examination of the stationary of multivariate bankruptcy prediction models: A methodological study, *Journal of Accounting Research*, 22(1), pp. 380-395.
- Messier, W.F.Jr. y Hansen, J.V. (1988). Inducing rules for expert system development: An example using default and bankruptcy data, *Management Science*, 34(12), pp. 1403-1415.
- Meyer, P.A. y Pifer, H.W. (1970). Predictions of bank failures, *The Journal of Finance*, 25(4), September, pp. 853-868.
- Min, S.H., Lee, J. y Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, 31(3), October, pp. 652-660.
- Mora Enguádanos, A. (1994a). Limitaciones metodológicas de los trabajos empíricos sobre la predicción del fracaso empresarial, *Revista Española de Financiación y Contabilidad*, 24(80), pp. 709-732.
- Mora Enguádanos, A. (1994b). Los modelos de predicción del fracaso empresarial: una aplicación empírica del logit. *Revista Española de Financiación y Contabilidad*, 24(78), pp. 203-233.
- Norman, M. y Stoker, B. (1991). *Data envelopment analysis. The assessment of performance*. Chichester. England. John Wiley & Sons.
- Norton, C.L. (1976). *A Comparison of the abilities of general price level and conventional financial ratios to predict bankruptcy*, Arizona State University.
- Norton, C. y Smith, R. (1979). A comparison of general price level and historical cost financial statements in the prediction of bankruptcy, *The Accounting Review*, 54(1), January, pp. 72-87.
- Nunamaker, T.R. (1985). Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation, *Managerial and Decision Economics*, 6(1), pp. 50-58.
- Odom, M.D. y Sharda, R. (1992). A neural network model for bankruptcy prediction. En R.R. Trippi and E. Turban Eds. *Neural networks in finance and investing*. Chicago: Probus Publishing, pp. 163-168.

- Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, 18(1), pp. 109-131.
- Palepu, K.G. (1986). Predicting takeover targets: A methodological and empirical analysis, *Journal of Accounting and Economics*, 8(1), March, pp. 3-35.
- Paradi, J.C., Asmild, M. y Simak, P.C. (2004). Using DEA and worst practice DEA in credit risk evaluation, *Journal of Productivity Analysis*, 21(2), March, pp. 153-165.
- Park, C.S. y Han, I. (2002). A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction, *Expert Systems with Applications*, 23(3), pp. 255-264.
- Peel, M.J. y Peel, D.A. (1987). Some further empirical evidence on predicting private company failure, *Accounting and Business Research*, 18(69), pp. 57-66.
- Peel, M.J., Peel, D.A. y Pope, P.F. (1986). Predicting corporate failure. Some results for the UK corporate sector, *Omega*, 14(1), pp. 5-12.
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Madrid. McGraw Hill
- Pina Martínez, V. (1989). Estudio empírico de la crisis bancaria, *Revista Española de Financiación y contabilidad*, 28(58), enero-marzo, pp. 309-338.
- Piramuthu, S., Ragavan, H. y Shaw, M.J. (1998). Using feature construction to improve the performance of neural networks, *Management Science*, 44 (3), pp. 416-430.
- Platt, H.D. y Platt, M.B. (1990). Development of a class of stable predictive variables: The case of bankruptcy prediction, *Journal of Business, Finance and Accounting*, 17(1), pp. 31-51.
- Platt, H.D. y Platt, M.B. (1991). A note on the use of industry relative ratios in bankruptcy prediction, *Journal of Banking and Finance*, 15(6), pp. 1183-1194.
- Platt, H.D. y Platt, M.B. (2002). Predicting corporate financial distress: Reflections on choice-based sample bias, *Journal of Economics and Finance*, 26(2), pp. 184-199.
- Platt, H.D., Platt, M.B. y Pedersen, J.G. (1994). Bankruptcy discrimination with real variables, *Journal of Business, Finance and Accounting*, 21(4), June, pp. 491-510.
- Pompe, P.P.M. y Bilderbeek, J. (2005). The prediction of bankruptcy of small and medium sized industrial firms, *Journal of Business Venturing*, 20(6), November, pp. 847-868.
- Pozuelo, J. Labatut, G. y Veres, E.J. (2013). Validez de la información financiera en los procesos de insolvencia. Un estudio de la pequeña empresa española, *Cuadernos de Economía y Dirección de la Empresa*, en prensa, 12 pp.

- Premachandra, I.M., Bhabra, G.S. y Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique, *European Journal of Operational Research*, 193(2), pp. 412-424.
- Premachandra, I.M., Chen, Y. y Watson, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment, *Omega*, 39, pp. 620-626.
- Psillaki, M., Tsolas, I.E. y Margaritis, D. (2010). Evaluation of credit risk based on firm performance, *European Journal of Operational Research*, 201(3), pp. 873-881.
- Ramírez Comeig, I. (1996). La utilidad del análisis multivariante para evaluar la solvencia de las pequeñas empresas. *X Congreso Nacional de AEDEM*, Granada, junio, pp. 463-473.
- Ravi Kumar, P. y Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review, *European Journal of Operational Research*, 180(1), pp. 1-28.
- Rhodes, E. (1978). *Data envelopment analysis and approaches form measuring the efficiency of decision-making units with an application to program follow-through in U.S. education*, PhD Dissertation, School of Ruban and Public Affairs, Carnegie-Mellon University.
- Rodríguez Acebes, M.C. (1990). *La predicción de las crisis empresariales. Modelos para el sector de seguros*, Valladolid. Secretariado de Publicaciones, Universidad de Valladolid.
- Rodríguez Fernández, J.M. (1986). Crisis en los bancos privados españoles: un modelo logit, *Investigaciones Económicas*, (supl.), pp. 59-64.
- Rodríguez Fernández, J.M. (1987). Crisis en los bancos privados españoles: un modelo logit. *II Jornadas de Economía Industrial*, Madrid.
- Rodríguez Fernández, J.M. (1989a). Análisis de las insolvencias bancarias en España: un modelo empírico, *Moneda y Crédito*, 189, pp. 187-227.
- Rodríguez Fernández, J.M. (1989b). The crisis in Spanish private banks: A logit analysis, *Finance*, 10(1), junio, pp. 69-88.
- Rodríguez, M. y Díaz, F. (2005). La teoría de los *rough sets* y la predicción del fracaso empresarial: diseño de un modelo para las pymes, *XIII Congreso AECA, Armonización y gobierno de la diversidad*, 22 a 24 de septiembre, Oviedo (recurso electrónico).
- Rodríguez López, M. (2001). Predicción del fracaso empresarial en compañías no financieras: Consideración de técnicas de análisis multivariante de corte paramétrico, *Actualidad Financiera*, 6(6), pp. 27-42.

- Román, I., De La Torre, J.M., Castillo, P.A. y Merelo, J.J. (2002). Sectorial bankruptcy prediction analysis using artificial neural networks: The case of Spanish companies, *European Accounting Congress*, Copenhagen.
- Román, I., De la Torre, J.M. y Zafra, J.L. (2001). Análisis sectorial de la predicción del riesgo de insolvencia: un estudio empírico, *XI Congreso AECA: Empresa, Euro y Nueva Economía*, Madrid, 26-28 septiembre (recurso electrónico).
- Rose, P.S., Andrews, W.T. y Giroux, G.A. (1982). Predicting business failure: A macroeconomic perspective, *Journal of Accounting, Auditing and Finance*, 6(1), Fall, pp. 20-31.
- Rubio Misas, M. (2008). Análisis del fracaso empresarial en Andalucía: especial referencia a la edad de la empresa, *Cuadernos de CC.EE. Y EE.*, 54, pp. 35-56.
- Rughupathi, W., Schkade, L. y Raju, B.S. (1993). A neural network to bankruptcy prediction. En Trippi, R. y Turban, E. (editors) *Neural network in finance and investing*. Cambridge: Probus Publishing Company, pp. 159-176.
- Rumelt, R.P. (1997). How much does industry matter?, *Strategic Management Journal*, 12(3), pp. 167-185.
- Sanchís Arellano, A., Gil, J.A. y Heras Martínez, A. (2003). El análisis discriminante en la previsión de la insolvencia en las empresas de seguros de no vida, *Revista Española de Financiación y Contabilidad*, 32(16), enero-marzo, pp. 183-233.
- Santomero, A.M. y Vinso J.D. (1977). Estimating the probability of failure for commercial banks and the banking system, *Journal of Banking and Finance*, 1(2), October, pp. 185-205.
- Sarle, W.S. (1994). Neural networks and statistical models, *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, April.
- Scott, J. (1981). The probability of bankruptcy: A comparison of empirical predictions and theoretical models, *Journal of Banking and Finance*, 5(3), September, pp. 317-344.
- Seifor, L.M., y Thrall, R.M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis, *Journal of Econometrics*, 46(1-2), pp. 7-38.
- Serrano Cinca, C. (1994). *Las redes neuronales artificiales en el análisis de la información contable*. Tesis doctoral. Zaragoza. Universidad de Zaragoza.
- Serrano Cinca, C. (1996). Self organizing neural networks for financial diagnosis, *Decision Support Systems*, 17(3), pp. 227-238.
- Serrano Cinca, C. (1997). Feedforward neural networks in the classification of financial information, *European Journal of Finance*, 3(3), September, pp. 183-202.

- Serrano, C. y Martín, B. (1993). Predicción de la crisis bancaria mediante el empleo de redes neuronales artificiales, *Revista Española de Financiación y Contabilidad*, 22(74), pp. 153-176.
- Shetty, U., Pakkala, T.P.M., y Mallikarjunappa, T. (2012). A modified directional distance formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India, *Expert Systems with Applications*, 39(2), pp. 1988-1997.
- Shin, K.S. y Lee, Y.J. (2002). A genetic algorithm application in bankruptcy prediction modeling, *Expert Systems with Applications*, 23(3), pp. 321-328.
- Shin, K.S., Shin, T.S. y Han, I. (1998). Intelligent corporate credit rating system using bankruptcy probability matrix. Proceedings of the *IV International Conference on Artificial Intelligence and Emerging Technologies in Accounting, Finance and Tax*, Huelva.
- Shrieves, R.E. y Stevens, D.L. (1979). Bankruptcy avoidance as a motive for merger, *Journal of Financial and Quantitative Analysis*, 14(3), September, pp. 501-515.
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model, *Journal of Business*, 74(1), January, pp. 101-124.
- Silverman, B.W. y Jones M.C. (1989). E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951), *International Statistical Review / Revue Internationale de Statistique*. 57(3), December, pp. 233-238.
- Sinkey, J.F. (1975). A multivariate statistical analysis of the characteristics of problem banks, *The Journal of Finance*, 30(1), March, pp. 21-36.
- Slowinski, R. y Zopounidis, C. (1995). Application of the rough set approach to evaluation of bankruptcy risk, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 4(1), pp. 27-41.
- Smith, C. A. B. (1947). Some examples of discrimination. *Annals of Eugenics*, 13, May, pp. 272-282.
- Somoza López, A. (2001). La consideración de factores cualitativos, macroeconómicos y sectoriales en los modelos de predicción de la solvencia empresarial, *Papeles de Economía Española*, 89/90, pp. 402-426.
- Somoza López, A. (2002). Modelos de predicción de la insolvencia: la incorporación de otro tipo de variables. En Doldán, F. y Rodríguez, M. (Coord.) *La gestión del riesgo de crédito*, Madrid, AECA, pp. 139-173.
- Stein, J.H. y Ziegler, W. (1984). The prognosis and surveillance of risks from commercial credit borrowers, *Journal of Banking and Finance*, 8(2), June, pp. 249-268.
- Sueyoshi, T. (2005). Financial ratio analysis of the electric power industry, *Asia-Pacific Journal of Operational Research*, 22(3), pp. 349-376.

- Sueyoshi, T. y Goto, M. (2009a). Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA–discriminant analysis, *European Journal of Operational Research*, 196(1), pp. 289-311.
- Sueyoshi, T. y Goto, M. (2009b). DEA–DA for bankruptcy-based performance assessment: Misclassification analysis of Japanese construction industry, *European Journal of Operational Research*, 199(2), pp. 576-594.
- Sueyoshi, T. y Goto, M. (2009c). Methodological comparison between DEA (data envelopment analysis) and DEA–DA (discriminant analysis) from the perspective of bankruptcy assessment, *European Journal of Operational Research*, 199(2), pp. 561-575.
- Surkan, A.J. y Singleton, J.C. (1992). Neural networks for bond rating improved by multiple hidden layers. En R.R. Trippi y E. Turban, *Neural networks in finance and investing*, Chicago: Probus Publishing.
- Swicegood, P. y Clark, J.A. (2001). Off-site monitoring for predicting bank under performance: A comparison of neural networks, discriminant analysis and professional human judgment, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10(3), September, pp. 169-186.
- Taffler, R.J. (1982). Forecasting company failure in the UK using discriminant analysis and finance ratio data, *Journal of the Royal Statistical Society, Series A*, 145(3), pp. 342-358.
- Taffler, R.J. (1983). The assessment of company solvency and performance using a statistical model, *Accounting and Business Research*, 15(52), Autumn, pp. 295-307.
- Tam, K.Y. (1991). Neural network models and the prediction of bank bankruptcy, *Omega*, 19(5), pp. 429–445.
- Tam, K.Y. y Kiang, M.Y. (1992). Managerial applications of neural networks: The case of bank failure predictions, *Management Science*, 38(7), July, pp. 926-947.
- Troutt, M.D., Rai, A. y Zhang, A. (1996). The potential use of DEA for credit applicant acceptance systems, *Computers and Operations Research*, 23(4), April, pp. 405-408.
- Tsukuda, J. y Baba, S.I. (1994). Predicting Japanese corporate bankruptcy in terms of finance data using neural network, *Computers and Industrial Engineering*, 27(1-4), pp. 445-448.
- Viscione, J.A. (1985). Assessing financial distress, *The Journal of Commercial Bank Lending*, July, pp. 39-55.
- Wagner, J.M. y Shimshak, D.G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives, *European Journal of Operational Research*, 180(1), pp. 57-67.

- Watson, J. y Everett, J. (1996). Do small business have high failure rates?, *Journal of Small Business Management*, 34(4), pp. 45-52.
- Westgaard, S. y Van Der Wijst, N. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach, *European Journal of Operational Research*, 135(2), December, pp. 338-349.
- Whalen, G. (1991). A proportional hazard model of bank failure: An examination of its usefulness as an early warning model tool, *Federal Reserve Bank of Cleveland Economic Review*, 27(1), pp. 21-31.
- Wheelock, D.C. y Wilson, P.W. (2000). Why do banks disappear? The determinants of U.S. bank failures and acquisitions, *The Review of Economics and Statistics*, 82(1), February, pp. 127-138.
- Whittred, G.P. y Zimmer, I. (1984). Timeliness of financial reporting and financial distress, *The Accounting Review*, 59(2), April, pp. 297-295.
- Wilson, R.L. y Sharda, R. (1994). Bankruptcy prediction using neural networks, *Decision Support Systems*, 11(5), June, pp. 545-557.
- Wilcox, J.W. (1971). A gambler's ruin prediction of business failure using accounting data, *Sloan Management Review*, 12(3), September, pp. 1-10.
- Wilcox, J.W. (1973). A prediction of business failure using accounting data. *Empirical Research in Accounting: Selected Studies*, Supplement to *Journal of Accounting Research*, 11, pp. 163-179.
- Wilcox, J.W. (1976). The gambler's ruin approach to business risk, *Sloan Management Review*, 18(autumn), pp. 33-46.
- Wilson, R.L. y Sharda, R. (1994). Bankruptcy prediction using neural networks, *Decision Support Systems*, 11(5), pp. 545-557.
- Wu, Y., Gaunt, C. y Gray, S. (2010). A comparison of alternative bankruptcy prediction models, *Journal of Contemporary Accounting and Economics*, 6(1), June, pp. 34-45.
- Xu, M. y Zhang, C. (2009). Bankruptcy prediction: The case of Japanese listed companies, *Review of Accounting Studies*, 14(4), December, pp. 534-558.
- Zavgren, C.V. (1983). The prediction of corporate failure: The state of the art, *Journal of Accounting Literature*, 2(1), pp. 1-38.
- Zavgren, C.V. (1985). Assessing the vulnerability of failure of American industrial firms: A logistic analysis, *Journal of Banking and Finance*, 12(1), Spring, pp. 19-45.
- Zavgren, C.V. (1988). The association between probabilities of bankruptcy and market responses- a test of market anticipation, *Journal of Business, Finance and Accounting*, 15(1), pp. 27-45.

- Zhang, G.P., Hu, M.Y., Patuwo, B.E. e Indro, D.C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, *European Journal of Operational Research*, 116(1), July, pp. 16-32.
- Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22 (supplement), pp. 59-82.

ANEXOS

Anexo II Trabajos y metodologías

AÑO	AUTORES	METODOLOGÍA UTILIZADA
1966	Beaver	Análisis Univariante
1968	Altman	Análisis Discriminante Múltiple
	Beaver	Análisis Univariante
1970	Meyer y Pifer	Análisis Discriminante Múltiple
1972	Deakin	Análisis Discriminante Múltiple
	Edmister	Análisis Discriminante Múltiple
1974	Blum	Análisis Discriminante Múltiple
1975	Elam	Consideración de Métodos Contables Alternativos al Principio de Devengo
	Libby	Análisis Discriminante Múltiple
	Sinkey	Análisis Discriminante Múltiple
1976	Norton	Introducción de Índices Correctores de Inflación
1977	Altman, Haldeman y Narayanan	Análisis Discriminante Múltiple
	Martín	Análisis de Regresión Logística
	Santomero y Vinso	Análisis de Regresión Logística
1978	Ketz	Datos Ajustados al Nivel de Precios – Inflación
1979	Norton y Smith	Consideración de Métodos Contables Alternativos al Principio de Devengo
1980	Dambolena y Khoury	Análisis Discriminante Múltiple
	Ohlson	Análisis de Regresión Logística
1982	Rose, Andrews y Giroux	Influencia de Variables Macroeconómicas
	Collins y Green	Análisis Discriminante Múltiple
		Modelo de Probabilidad Lineal
Análisis de Regresión Logística		
1983	Gombola y Ketz	Análisis Discriminante Múltiple
	Taffler	Análisis Discriminante Múltiple
1984	Holder	Análisis Discriminante Múltiple
	Marais, Patell y Wolfson	Modelos de Partición Recursiva
	Mensah	Análisis de Regresión Logística Influencia de Variables Macroeconómicas
	Zmijewski	Análisis de Regresión Logística
	Lincoln	Análisis Discriminante Múltiple
1985	Casey y Bartczak	Análisis Discriminante Múltiple
		Análisis de Probabilidad Condicional
	Frydman, Altman y Kao	Modelos de Partición Recursiva
	Gentry, Newbold y Whitford	Análisis Discriminante Múltiple
		Análisis de Probabilidad Condicional
Zavgren	Análisis de Probabilidad Condicional	

AÑO	AUTORES	METODOLOGÍA UTILIZADA
		Análisis de Regresión Logística
	Laffarga, Martín y Vázquez	Análisis Univariante
		Análisis de Regresión Logística
1986	Peel, Peel y Pope	Introducción de Variables No Financieras
1987	Keasey y Watson	Análisis de Regresión Logística Introducción de Variables No Financieras
	Rodríguez Fernández	Análisis de Regresión Logística
	Laffarga, Martín y Vázquez	Análisis Discriminante Múltiple
		Análisis de Regresión Logística
1988	Edmister	Análisis Discriminante Múltiple
	Messier y Hansen	Inteligencia Artificial: Sistemas Expertos
1989	Martínez, Navarro y Sanz	Análisis de Regresión Logística
	Pina	Análisis de Regresión Logística
	Rodríguez Fernández	Análisis Univariante
Análisis de Regresión Logística		
1990	Bell, Ribar y Verchio	Inteligencia Artificial: Redes Neuronales
		Análisis de Regresión Logística
	McKee	Inteligencia Artificial: Sistemas Expertos
	Rodríguez Acebes	Análisis de Regresión Logística
	Gabás	Análisis Univariante
		Análisis Discriminante Múltiple
		Análisis de Regresión Logística
Modelos de Partición Recursiva		
1991	Koh	Análisis de Regresión Logística
	Platt y Platt	Análisis de Regresión Logística
	Mar Molinero y Ezzamel	Escalamiento Multidimensional
	Tam	Inteligencia Artificial: Redes Neuronales, BPNN
1992	Tam y Kiang	Inteligencia Artificial: Redes Neuronales, BPNN
	Surkan y Singleton	Inteligencia Artificial
	Dutta y Shekhar	Inteligencia Artificial
	Odom y Sharda	Inteligencia Artificial: Redes neuronales
	Marose	Inteligencia Artificial
1993	Fletcher y Goss	Inteligencia Artificial: Redes Neuronales, BPNN
	Rughupathi, Schkade y Raju	Inteligencia Artificial
	Serrano y Martín del Brio	Inteligencia Artificial: Redes Neuronales
	De Miguel, Revilla, Rodríguez y Cano	Inteligencia Artificial
1994	Altman, Marco y Varetto	Análisis Discriminante Lineal Inteligencia Artificial: Redes Neuronales, BPNN
	Platt, Platt y Pederson	Consideración de Métodos Contables Alternativos al

AÑO	AUTORES	METODOLOGÍA UTILIZADA
		Principio de Devengo
	Wilson y Sharda	Inteligencia Artificial: Redes Neuronales, BPNN
	Tsukuda y Baba	Inteligencia Artificial: Redes Neuronales, BPNN
	López, Moreno y Rodríguez	Análisis Univariante
		Análisis Discriminante Múltiple
	Mora	Análisis de Regresión Logística
Serrano Cinca	Inteligencia Artificial: Redes Neuronales	
1995	Slowinski y Zopounidis	Teoría de los Conjuntos Aproximados
	Lacher, Coats, Sharma y Faut	Inteligencia Artificial
	García, Arqués y Calvo-Flores	Análisis Discriminante Múltiple
	Fernandez y Olmeda	Análisis de Regresión Logística
Inteligencia Artificial: Redes Neuronales		
1996	Greenstein y Welsh	Inteligencia Artificial
	Serrano Cinca	Mapas Autoorganizativos
		Inteligencia Artificial: Redes Neuronales
		Análisis Discriminante Lineal
	Lesho y Spector	Inteligencia Artificial: Redes Neuronales, BPNN
	Del Rey	Inteligencia Artificial: Redes Neuronales
Martínez	Inteligencia Artificial	
1997	Barniv, Anurag y Leach	Inteligencia Artificial: Redes Neuronales, BPNN
	Bell	Inteligencia Artificial: Redes Neuronales, BPNN
	Lizarraga	Análisis Discriminante Múltiple
		Análisis de Regresión Logística
	Gallego, Gómez y Yáñez	Análisis de Regresión Logística
	Bonsón, Escobar y Martín	Inteligencia Artificial: Árboles de Decisión
Serrano Cinca	Análisis Univariante	
	Análisis de Regresión Logística	
	Inteligencia Artificial: Redes Neuronales	
1998	Shin, Shin y Han	Inteligencia Artificial
	Piramuthu, Ragavan y Shaw	Inteligencia Artificial: Redes Neuronales, BPNN
	Kiviluoto	Mapas Autoorganizativos
	López, Gandía y Molina	Análisis de Componentes Principales
		Análisis de Regresión Logística
	Ferrando y Blanco	Análisis Discriminante Múltiple
		Análisis de Regresión Logística
	Lizarraga	Análisis Discriminante Múltiple
Análisis de Regresión Logística		
1999	Koh y Tan	Inteligencia Artificial
	Lennox	Análisis Discriminante Múltiple
		Análisis de Regresión Logística

AÑO	AUTORES	METODOLOGÍA UTILIZADA
	Laitinen y Kankaanpää	Análisis Discriminante Lineal
		Análisis de Regresión Logística
		Modelos de Partición Recursiva
		Análisis de Supervivencia
		Redes Neuronales
		Juicio Humano de Expertos
	Zhang , Hu, Patuwo y Indro	Inteligencia Artificial
López y Flórez	Inteligencia Artificial: Redes Neuronales	
2000	McKee	Teoría de los Conjuntos Aproximados
	Ahn, Cho y Kim	Análisis Discriminante
		Inteligencia Artificial: Redes Neuronales
		Híbrido: Conjuntos Aproximados-Redes Neuronales
	Crespo	Análisis Discriminante Múltiple
Inteligencia Artificial: Redes Neuronales		
2001	Grice e Ingram	Análisis Discriminante Múltiple
	Westgaard y Wijst	Análisis de Regresión Logística
	Atiya	Inteligencia Artificial: Redes Neuronales, BPNN
	Swicegood y Clark	Inteligencia Artificial: Redes Neuronales, BPNN
		Análisis Discriminante Múltiple
		Juicio Humano de Profesionales
	Kaski, Sinkkonen y Peltonen	Mapas Autoorganizativos
	Rodríguez López	Análisis Discriminante Múltiple
		Análisis de Regresión Logística
	Somoza	Análisis de Regresión Logística
	De Andrés	Análisis Discriminante Múltiple
Análisis de Regresión Logística		
Inteligencia Artificial: Algoritmo SEE5		
2002	Shin y Lee	Inteligencia Artificial: Algoritmos Genéticos
	Park y Han	Análisis Multicriterio
	Platt y Platt	Análisis de Regresión Logística
	Román, De la Torre, Castillo y Merele	Análisis de Regresión Logística
		Inteligencia Artificial: Redes Neuronales
2003	Correa, Acosta y González	Análisis de Regresión Logística
		Inteligencia Artificial: Algoritmo SEE5
2004	Cielen, Peeters y Vanhoof	Análisis Envoltante de Datos, Radial (CCR)
	Paradi, Asmild y Simak	Análisis Envoltante de Datos, No Radial
2005	Beaver, McNichols y Rhie	Análisis de Regresión Logística Multiperiodo
	Lee, Booth y Alam	Inteligencia Artificial: Redes Neuronales, BPNN
		Mapas Autoorganizativos
Canbas, Cabuk y Kilic	Análisis Discriminante	

AÑO	AUTORES	METODOLOGÍA UTILIZADA
		Análisis de Regresión Logística
		Probit
		Análisis de Componentes Principales
	De la Torre, Gómez y Román	Análisis de Regresión Logística
	Rodríguez y Díaz	Teoría de los Conjuntos Aproximados
	De Andrés	Análisis de Regresión Logística Inteligencia Artificial: Redes Neuronales
2006	Min, Lee y Han	Inteligencia Artificial: Algoritmos Genéticos
		Máquina de Vector de Apoyo
	Calvo-Flores, García y Madrid	Análisis de Regresión Logística
		Análisis Cluster
2007	Kuo	Análisis Envoltente de Datos, Aditivo
2008	Jones y Hensher	Análisis de Regresión Logística
		Mixed Logit
	Arquero, Abad y Jiménez	Regresiones Nominales (multilogit)
2009	Beaver, Correia y McNichols	Análisis de Regresión Logística Multiperiodo
	Premachandra, Bhabra y Sueyoshi	Análisis Envoltente de Datos, Aditivo
		Análisis de Regresión Logística
	Sueyoshi y Goto (a, b y c)	Análisis Envoltente de Datos
		Análisis Envoltente de Datos-Análisis Discriminante
	Xu y Zhang	Análisis Discriminante Múltiple
Análisis de Regresión Logística		
Distancia al Fallido (Teoría Opciones)		
2010	Baixauli y Módica-Milo	Análisis de Regresión Logística
	Wu, Gaunt y Gray	Análisis Discriminante Múltiple
		Análisis de Regresión Logística
	Manzaneque, Benegas y García	Análisis Cluster
	Psillaki, Tsolas y Margaritis	Análisis Envoltente de Datos
Análisis de Regresión Logística		
2011	Jim, Kanagaretnam y Lobo	Análisis de Regresión Heckman's
2012	Blanco, Irimia y Oliver	Análisis de Regresión Logística
	Cook, Pandit y Milman	Análisis de Regresión Logística
	Shetty, Pakkala y Mallikarjunappa	Análisis Envoltente de Datos
2013	Mateev, Poutziouris y Ivanov	Modelo de Regresión Dinámica
	Pozuelo, Labatut y Veres	Análisis de Regresión Logística

VARIABLES PREDICTIVAS		Autores que han utilizado el ratio	
Ingresos de explotación/Activo fijo			2
Ingresos de explotación/Activo Total		X	2
Ingresos de explotación/Capital circulante			2
Ingresos de explotación/Ventas			1
Ingresos netos/Activo Total		X	3
Inmovilizado/Patrimonio neto			1
Innovación/Activo Total		X	1
Intereses/Activo Total		X	1
Intereses/Ventas			1
Intervalo sin crédito		X	1
Inversiones=Cash-Gastos Financieros/Activo circulante			1
Logaritmo neperiano del Activo Total		X	1
Márgen de explotación /Activo e explotación medio		X	1
Pasivo circulante/Activo Total		X	2
Pasivo circulante/Fondos Propios			1
Pasivo exigible/Ingresos de explotación		X	2
Pasivo exigible/Pasivo Total		X	1
Pasivo Fijo/Ventas			2
Pasivo Largo Plazo/Activo Total			1
Pasivo Largo Plazo/Fondos Propios		X	1
Pasivo Total/Activo Circulante			3
Pasivo Total/Activo Total		X	1
Pasivo Total/Fondos Propios		X	4
Patrimonio Neto/Activo Total			1
Plazo de pago de deudas		X	1
Provisiones/Ventas		X	1
Quick ratio (Activo circulante/Pasivo Circulante)	X		1
Realizable/Activo Circulante			2
Realizable/Activo Total			2
Recursos generados/Activo Total		X	2
Recursos generados/Fondos propios			1
Recursos generados/Ingresos de explotación		X	1

VARIABLES PREDICTIVAS		6	5	22	12	5	12	9	5	8	13	10	16	3	6	28	19	15	22	11	3	16	9	14	269
Recursos generados/ Pasivo exigible																									
Recursos generados/ Pasivo total																									
Rentabilidad Económica - Coste Medio Pasivo																									
Reservas/Activo Total																									
Resultado antes de impuestos/Activo Total																									
Resultado antes de impuestos/Fondos propios																									
Resultado antes de impuestos/Ventas																									
Resultado de actividades ordinarias/Activo Total																									
Resultado de actividades ordinarias/Fondos Propios																									
Resultado de actividades ordinarias/Ventas																									
Resultado del ejercicio/Valor añadido																									
Resultado neto/Activo Total																									
Resultado neto/Fondos Propios																									
Resultado neto-Realizable-Existencias/Activo Total																									
Retraso en la publicación																									
Tesorería+Deudores/Activo Total																									
Tesorería+Deudores/Pasivo circulante																									
Valor añadido/Activo Total																									
Valor añadido/Ingresos de explotación																									
Valor añadido/Valor de libros																									
Valor añadido/Ventas																									
Ventas/Activo Total																									
Ventas/Capital circulante-Existencias																									
Ventas/Clientes																									
Ventas/Deuda Total																									
Ventas/Deuda Total																									
Ventas/Número empleados medio																									
Ventas/Tesorería																									
Número de ratios utilizados por cada autor		6	5	22	12	5	12	9	5	8	13	10	16	3	6	28	19	15	22	11	3	16	9	14	269

Autores que han utilizado el ratio

- 1 Blanco, Irlanda y Oliver (2012)
- 2 Llano, Piñeiro y Rodríguez (2011)
- 1 Pederzoli y Torricelli (2010)
- 1 Manzanque, Benegas y García (2010)
- 3 Mora y González (2009)
- 3 Lugovskaya (2009)
- 3 Fantazini y Figini (2009)
- 3 Labatut, Pozuelo y Veres (2009)
- 3 Labatut, Pozuelo y Veres (2009)
- 3 Arquero, Abad y Jiménez (2008)
- 3 Gómez, Torre y Román (2008)
- 3 Somoza y Valverde (2007)
- 3 Altman y Sabato (2007)
- 3 Pompe y Bildereek (2005)
- 3 Altman y Sabato (2005)
- 3 Lizaraga (1998)
- 3 Ferrando y Blanco (1998)
- 3 López, Gandía y Molina (1998)
- 3 Ferrando y Blanco (1998)
- 3 Gallego, Gomez y Yañez (1997)
- 3 Lizaraga (1997)
- 3 Gandía, López Gracia y Molina (1995)
- 3 Keasey y Watson (1987)

Anexo IV Instrucciones programa STATA

```

**el delimitador de linea será ;**
#delimit;
**indica que los resultados serán mostrados por pantalla hasta el final. Es
decir, si los resultados ocupan más de una pantalla, la instrucción set more
off impide que los resultados vayan apareciendo poco a poco**
set more off;
** evita errores en el caso de que no haya podido ejecutarse un fichero .do
hasta el final**
capture log close;
log using "C:\Users\Javier Castaño\Desktop\DOCUMENTOS
TESIS\TESIS\salidatesis.log", replace;
use "C:\Users\Javier Castaño\Desktop\DOCUMENTOS
TESIS\TESIS\Mayte\datos tesis.dta", clear;
** winsorizar variables al 1%**
winsor R1_dt_at, gen(wR1) p(0.01);
winsor R2_ac_pc, gen(wR2) p(0.01);
winsor R3_bait_at, gen(wR3) p(0.01);
winsor R4_bn_at, gen(wR4) p(0.01);
winsor R5_ac_at, gen(wR5) p(0.01);
winsor R6_gf_dt, gen(wR6) p(0.01);
winsor R7_bnd_at, gen(wR7) p(0.01);
winsor R8__bn_amor__dt, gen(wR8) p(0.01);
winsor R9_margen_sobre_ventas__bn_incn, gen(wR9) p(0.01);
winsor R10_rotacion_activos_incn_at, gen(wR10) p(0.01);
** matriz de correlación con datos normales **
pwwcorr fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at
incidencias num_empl_total, star(.01);
** matriz de correlación con datos winsorizados**
pwwcorr fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias num_empl_total, star(1);
** estadística descriptiva separadas por grupos de fracasada=1 y no
fracasadas=0 sin winsorizar**
tabstat R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias num_empl_total if fracaso==1,
statistics( count mean median sd skewness ) columns(statistics);
tabstat R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias num_empl_total if fracaso==0,
statistics( count mean median sd skewness ) columns(statistics);
** estadística descriptiva separadas por grupos de fracasada=1 y no
fracasadas=0 datos winsorizar**

```

```

tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total if fracaso==0, statistics( count mean median sd skewness)
columns(statistics);
tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total if fracaso==0, statistics( count mean median sd skewness)
columns(statistics);
** Diferencia de medias por variable separado por fracaso =1 y fracaso=0,
nivel de significación del 99%**;
ttest R1_dt_at, by(fracaso) level(99);
ttest R2_ac_pc, by(fracaso) level(99);
ttest R3_bait_at, by(fracaso) level(99);
ttest R4_bn_at, by(fracaso) level(99);
ttest R5_ac_at, by(fracaso) level(99);
ttest R6_gf_dt, by(fracaso) level(99);
ttest R7_bnd_at, by(fracaso) level(99);
ttest R8__bn_amor__dt, by(fracaso) level(99);
ttest R9_margen_sobre_ventas__bn_incn, by(fracaso) level(99);
ttest R10_rotacion_activos_incn_at, by(fracaso) level(99);
ttest incidencias, by(fracaso) level(99);
ttest num_empl_total, by(fracaso) level(99);
** para el cálculo de la mediana, la desviación típica y la asimetría **;
tabstat R1_dt_at R2_ac_pc R5_ac_at R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R10_rotacion_activos_incn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at
incidencias num_empl_total, statistics( median sd skewness ) by(fracaso)
casewise nototal labelwidth(0) varwidth(0) columns(statistics) longstub;
** para el cálculo del test de Wilcoxon**;
ranksum R1_dt_at, by(fracaso);
ranksum R2_ac_pc, by(fracaso);
ranksum R3_bait_at, by(fracaso);
ranksum R4_bn_at, by(fracaso);
ranksum R5_ac_at, by(fracaso);
ranksum R6_gf_dt, by(fracaso);
ranksum R7_bnd_at, by(fracaso);
ranksum R8__bn_amor__dt, by(fracaso);
ranksum R9_margen_sobre_ventas__bn_incn, by(fracaso);
ranksum R10_rotacion_activos_incn_at, by(fracaso);
ranksum incidencias, by(fracaso);
ranksum num_empl_total, by(fracaso);
** cálculo LDA con 10 variables**;
discrim lda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at, group(fracaso);
** cálculo LDA con 10 variables+incidencias**;
discrim lda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias, group(fracaso);
** cálculo LDA con 7 variables**;

```

```
discrim lda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo LDA con 7 variables+incidencias**;
```

```
discrim lda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn incidencias,
group(fracaso);
** cálculo LDA con 5 variables**;
```

```
discrim lda R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo LDA con 5 variables+incidencias**;
```

```
discrim lda R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, group(fracaso);
** cálculo LDA con 2 variables**;
```

```
discrim lda R6_gf_dt R8__bn_amor__dt, group(fracaso);
** cálculo LDA con 2 variables+incidencias**;
```

```
discrim lda R6_gf_dt R8__bn_amor__dt incidencias, group(fracaso);
** cálculo QDA con 10 variables**;
```

```
discrim qda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at, group(fracaso);
** cálculo QDA con 10 variables+incidencias**;
```

```
discrim qda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias, group(fracaso);
** cálculo QDA con 7 variables**;
```

```
discrim qda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo QDA con 7 variables+incidencias**;
```

```
discrim qda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn incidencias,
group(fracaso);
** cálculo QDA con 5 variables**;
```

```
discrim qda R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo QDA con 5 variables+incidencias**;
```

```
discrim qda R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, group(fracaso);
** cálculo QDA con 2 variables**;
```

```
discrim qda R6_gf_dt R8__bn_amor__dt, group(fracaso);
** cálculo QDA con 2 variables+incidencias**;
```

```
discrim qda R6_gf_dt R8__bn_amor__dt incidencias, group(fracaso);
** cálculo LogDA con 10 variables**;
```

```
discrim logistic R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at,
group(fracaso);
** cálculo LogDA con 10 variables+incidencias**;
```

```

discrim logistic R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at
incidencias, group(fracaso);
** cálculo LogDA con 7 variables**;
```

```

discrim logistic R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo LogDA con 7 variables+incidencias**;
```

```

discrim logistic R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn incidencias,
group(fracaso);
** cálculo LogDA con 5 variables**;
```

```

discrim logistic R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);
** cálculo LogDA con 5 variables+incidencias**;
```

```

discrim logistic R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn incidencias, group(fracaso);
** cálculo LogDA con 2 variables**;
```

```

discrim logistic R6_gf_dt R8__bn_amor_dt, group(fracaso);
** cálculo LogDA con 2 variables+incidencias**;
```

```

discrim logistic R6_gf_dt R8__bn_amor_dt incidencias, group(fracaso);
** cálculo KnnDA con 10 variables**;
```

```

discrim knn R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at, k(2) group(fracaso) measure(L2);
** cálculo knnDA con 10 variables+incidencias**;
```

```

discrim knn R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias, k(2) group(fracaso)
measure(L2);
** cálculo knnDA con 7 variables**;
```

```

discrim knn R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn, k(2)
group(fracaso) measure(L2);
** cálculo KnnDA con 7 variables+incidencias**;
```

```

discrim knn R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor_dt R9_margen_sobre_ventas__bn_incn incidencias, k(2)
group(fracaso) measure(L2);
** cálculo KnnDA con 5 variables**;
```

```

discrim knn R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn, k(2) group(fracaso) measure(L2);
** cálculo KnnDA con 5 variables+incidencias**;
```

```

discrim knn R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn incidencias, k(2) group(fracaso)
measure(L2);
** cálculo KnnDA con 2 variables**;
```

```

discrim knn R6_gf_dt R8__bn_amor_dt, k(2) group(fracaso) measure(L2);
** cálculo KnnDA con 2 variables+incidencias**;
```

```
discrim knn R6_gf_dt R8__bn_amor__dt incidencias, k(2) group(fracaso)
measure(L2);
** cálculo de logit 10 variables **:
logit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 10 variables + incidencias **:
logit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn
R10_rotacion_activos_incn_at incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 7 variables **:
logit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn ;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 7 variables + incidencias **:
logit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 5 variables **:
logit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 5 variables + incidencias **:
logit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 2 variables **:
logit fracaso R6_gf_dt R8__bn_amor__dt ;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de logit 2 variables + incidencias **:
logit fracaso R6_gf_dt R8__bn_amor__dt incidencias;
fitstat;
```

```

Iroc, nograph;
estat classification;
** cálculo de probit 10 variables **;
probit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 10 variables + incidencias **;
probit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at
R6_gf_dt R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at
incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 7 variables **;
probit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn ;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 7 variables + incidencias **;
probit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8__bn_amor__dt R9_margen_sobre_ventas__bn_incn incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 5 variables **;
probit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 5 variables + incidencias **;
probit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias;
fitstat;
Iroc, nograph;
estat classification;
** cálculo de probit 2 variables **;
probit fracaso R6_gf_dt R8__bn_amor__dt ;
fitstat;
estat classification;
Iroc, nograph;
** cálculo de probit 2 variables + incidencias **;
probit fracaso R6_gf_dt R8__bn_amor__dt incidencias;

```

```
fitstat;
estat classification;
lroc, nograph;

** DATOS WINSORIZADOS **;

** Diferencia de medias por variable separado por fracaso =1 y fracaso=0,
nivel de significación del 99%**;
ttest wR1, by(fracaso) level(99);
ttest wR2, by(fracaso) level(99);
ttest wR3, by(fracaso) level(99);
ttest wR4, by(fracaso) level(99);
ttest wR5, by(fracaso) level(99);
ttest wR6, by(fracaso) level(99);
ttest wR7, by(fracaso) level(99);
ttest wR8, by(fracaso) level(99);
ttest wR9, by(fracaso) level(99);
ttest wR10, by(fracaso) level(99);
ttest incidencias, by(fracaso) level(99);
ttest num_empl_total, by(fracaso) level(99);
** para el cálculo de la mediana, la desviación típica y la asimetría **;
tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total, statistics( median sd skewness ) by(fracaso) nototal
columns(statistics);
** para el cálculo del test de Wilcoxon**;
ranksum wR1, by(fracaso);
ranksum wR2, by(fracaso);
ranksum wR3, by(fracaso);
ranksum wR4, by(fracaso);
ranksum wR5, by(fracaso);
ranksum wR6, by(fracaso);
ranksum wR7, by(fracaso);
ranksum wR8, by(fracaso);
ranksum wR9, by(fracaso);
ranksum wR10, by(fracaso);
ranksum incidencias, by(fracaso);
ranksum num_empl_total, by(fracaso);
** cálculo LDA 10 variables **;
discrim lda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10,
group(fracaso);
** cálculo LDA 10 variables + incidencias**;
discrim lda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias, group(fracaso);
** cálculo LDA 7 variables **;
discrim lda wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);
** cálculo LDA 7variables + incidencias**;
discrim lda wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, group(fracaso);
** cálculo LDA 5 variables **;
```

```

discrim lda wR1 wR3 wR4 wR8 wR9, group(fracaso);
** cálculo LDA 5variables + incidencias**;
```

```

discrim lda wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);
** cálculo LDA 3 variables **;
```

```

discrim lda wR1 wR3 wR4, group(fracaso);
** cálculo LDA 3 variables + incidencias**;
```

```

discrim lda wR1 wR3 wR4 incidencias, group(fracaso);
** cálculo QDA 10 variables **;
```

```

discrim qda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10,
group(fracaso);
** cálculo QDA 10 variables + incidencias**;
```

```

discrim qda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias, group(fracaso);
** cálculo QDA 7 variables **;
```

```

discrim qda wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);
** cálculo QDA 7variables + incidencias**;
```

```

discrim qda wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias,
group(fracaso);
** cálculo QDA 5 variables **;
```

```

discrim qda wR1 wR3 wR4 wR8 wR9, group(fracaso);
** cálculo QDA 5variables + incidencias**;
```

```

discrim qda wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);
** cálculo QDA 3 variables **;
```

```

discrim qda wR1 wR3 wR4, group(fracaso);
** cálculo QDA 3 variables + incidencias**;
```

```

discrim qda wR1 wR3 wR4 incidencias, group(fracaso);
** cálculo LogDA 10 variables **;
```

```

discrim logistic wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10,
group(fracaso);
** cálculo LogDA 10 variables + incidencias**;
```

```

discrim logistic wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias, group(fracaso);
** cálculo LogDA 7 variables **;
```

```

discrim logistic wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);
** cálculo LogDA 7variables + incidencias**;
```

```

discrim logistic wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias,
group(fracaso);
** cálculo LogDA 5 variables **;
```

```

discrim logistic wR1 wR3 wR4 wR8 wR9, group(fracaso);
** cálculo LogDA 5variables + incidencias**;
```

```

discrim logistic wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);
** cálculo LogDA 3 variables **;
```

```

discrim logistic wR1 wR3 wR4, group(fracaso);
** cálculo LogDA 3 variables + incidencias**;
```

```

discrim logistic wR1 wR3 wR4 incidencias, group(fracaso);
** cálculo KnnDA con 10 variables**;
```

```

discrim knn wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10, k(2)
group(fracaso) measure(L2);

```

```
** cálculo knnDA con 10 variables+incidencias**;  
discrim knn wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10  
incidencias, k(2) group(fracaso) measure(L2);  
** cálculo knnDA con 7 variables**;  
discrim knn wR1 wR3 wR4 wR5 wR7 wR8 wR9, k(2) group(fracaso)  
measure(L2);  
** cálculo KnnDA con 7 variables+incidencias**;  
discrim knn wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, k(2)  
group(fracaso) measure(L2);  
** cálculo KnnDA con 5 variables**;  
discrim knn wR1 wR3 wR4 wR8 wR9, k(2) group(fracaso) measure(L2);  
** cálculo KnnDA con 5 variables+incidencias**;  
discrim knn wR1 wR3 wR4 wR8 wR9 incidencias, k(2) group(fracaso)  
measure(L2);  
** cálculo KnnDA con 3 variables**;  
discrim knn wR1 wR3 wR4 , k(2) group(fracaso) measure(L2);  
** cálculo KnnDA con 3 variables+incidencias**;  
discrim knn wR1 wR3 wR4 incidencias, k(2) group(fracaso) measure(L2);  
** cálculo de logit 10 variables **;  
logit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10;  
fitstat;  
estat classification;  
lroc, nograph;  
** cálculo de logit 10 variables + incidencias **;  
logit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10  
incidencias;  
fitstat;  
estat classification;  
lroc, nograph;  
** cálculo de logit 7 variables **;  
logit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9 ;  
fitstat;  
estat classification;  
lroc, nograph;  
** cálculo de logit 7 variables + incidencias **;  
logit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias;  
fitstat;  
estat classification;  
lroc, nograph;  
** cálculo de logit 5 variables **;  
logit fracaso wR1 wR3 wR4 wR8 wR9;  
fitstat;  
estat classification;  
lroc, nograph;  
** cálculo de logit 5 variables + incidencias **;  
logit fracaso wR1 wR3 wR4 wR8 wR9 incidencias;  
fitstat;  
estat classification;
```

```
lroc, nograph;
** cálculo de logit 3 variables **;
logit fracaso wR1 wR3 wR4 ;
fitstat;
estat classification;
lroc, nograph;
** cálculo de logit 3 variables + incidencias **;
logit fracaso wR1 wR3 wR4 incidencias;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 10 variables **;
probit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 10 variables + incidencias **;
probit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 7 variables **;
probit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 7 variables + incidencias **;
probit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 5 variables **;
probit fracaso wR1 wR3 wR4 wR8 wR9;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 5 variables + incidencias **;
probit fracaso wR1 wR3 wR4 wR8 wR9 incidencias;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 3 variables **;
probit fracaso wR1 wR3 wR4 ;
fitstat;
estat classification;
lroc, nograph;
** cálculo de probit 3 variables + incidencias **;
```

```
probit fracaso wR1 wR3 wR4 incidencias;  
fitstat;  
estat classification;  
lroc, nograph;
```

```
log close;
```


Anexo V Salida de resultados del programa STATA

```

-----
name: <unnamed>
log: C:\Users\Javier Castaño\Desktop\DOCUMENTOS
TESIS\TESIS\salidatesis.log
log type: text
opened on: 3 Sep 2012, 20:09:42

. use "C:\Users\Javier Castaño\Desktop\DOCUMENTOS TESIS\TESIS\Mayte\datos
tesis.dta", clear;

. ** winsorizar variables al 1%**;
. winsor R1_dt_at, gen(wR1) p(0.01);

. winsor R2_ac_pc, gen(wR2) p(0.01);

. winsor R3_bait_at, gen(wR3) p(0.01);

. winsor R4_bn_at, gen(wR4) p(0.01);

. winsor R5_ac_at, gen(wR5) p(0.01);

. winsor R6_gf_dt, gen(wR6) p(0.01);

. winsor R7_bnd_at, gen(wR7) p(0.01);

. winsor R8__bn_amor__dt, gen(wR8) p(0.01);

. winsor R9_margen_sobre_ventas__bn_incn, gen(wR9) p(0.01);

. winsor R10_rotacion_activos_incn_at, gen(wR10) p(0.01);

. ** matriz de correlación con datos normales **;
. pwcorr fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor__dt R9_margen_sobre_ventas__bn
> _incn R10_rotacion_activos_incn_at incidencias num_empl_total, star(.01);

+-----+
| fracaso R1_dt_at R2_ac_pc R3_bai~t R4_bn_at R5_ac_at R6_gf_dt
-----+-----+
fracaso | 1.0000
R1_dt_at | -0.0011 1.0000
R2_ac_pc | -0.0029 -0.0023 1.0000
R3_bait_at | -0.0108 -0.9057* 0.0031 1.0000
R4_bn_at | -0.0052 -0.9773* 0.0023 0.9692* 1.0000
R5_ac_at | 0.0147 -0.0024 0.0260 -0.0049 0.0354 1.0000
R6_gf_dt | 0.1593* -0.0006 0.0375 0.0024 0.0018 -0.0126 1.0000
R7_bnd_at | -0.0021 0.9995* -0.0015 -0.8940* -0.9714* -0.0011 -0.0006

```

```

R8__bn_amo~t | 0.0471* -0.0036 0.4113* 0.0343 0.0189 0.0096 0.5289*
R9_margen_~n | 0.0012 0.0002 -0.0000 0.0084 0.0044 -0.0062 0.0027
R10_rotaci~t | 0.0038 0.8138* -0.0032 -0.6477* -0.7883* -0.3350* -0.0144
incidencias | 0.1794* 0.1057* -0.0081 -0.0914* -0.0978* -0.0199 0.0041
num_empl_t~l | -0.0031 -0.0035 -0.0018 0.0051 0.0033 -0.0134 0.0054

      | R7_bnd~t R8__bn~t R9_mar~n R10_ro~t incide~s num_em~l
-----+-----
R7_bnd_at | 1.0000
R8__bn_amo~t | -0.0018 1.0000
R9_margen_~n | 0.0001 0.0240 1.0000
R10_rotaci~t | 0.8173* -0.0037 0.0015 1.0000
incidencias | 0.1051* -0.0227 0.0014 0.0808* 1.0000
num_empl_t~l | -0.0014 0.0003 0.0015 -0.0002 0.0024 1.0000

. ** matriz de correlación con datos winsorizados**;
. pwcorr fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total, star(1);

      | fracaso wR1 wR2 wR3 wR4 wR5 wR6
-----+-----
fracaso | 1.0000
wR1 | 0.0546* 1.0000
wR2 | -0.0047 -0.2632* 1.0000
wR3 | -0.0906* -0.5407* 0.0550* 1.0000
wR4 | -0.0999* -0.5879* 0.0656* 0.9719* 1.0000
wR5 | 0.0215 -0.1396* 0.1469* 0.1595* 0.1724* 1.0000
wR6 | 0.0193 -0.0274 0.1200* 0.0830* 0.0011 -0.0811* 1.0000
wR7 | -0.0127 0.3276* -0.0047 -0.1929* -0.2100* -0.0788* -0.0615*
wR8 | -0.0325 -0.4451* 0.2458* 0.5925* 0.6009* 0.0401 -0.0240
wR9 | -0.0384 -0.2497* -0.1024* 0.3873* 0.3974* 0.0778* -0.0056
wR10 | 0.0189 0.1653* -0.1781* -0.0521* -0.0957* 0.0208 0.1062*
incidencias | 0.1794* 0.1497* -0.0346 -0.1304* -0.1421* -0.0025 0.0542*
num_empl_t~l | -0.0031 -0.0088 -0.0021 0.0181 0.0119 -0.0212 0.0696*

      | wR7 wR8 wR9 wR10 incide~s num_em~l
-----+-----
wR7 | 1.0000
wR8 | 0.0182 1.0000
wR9 | -0.0880* 0.3025* 1.0000
wR10 | 0.1581* 0.0433 0.1265* 1.0000
incidencias | 0.0081 -0.1134* -0.0679* -0.0026 1.0000
num_empl_t~l | 0.0003 0.0081 0.0097 -0.0109 0.0024 1.0000

. ** estadística descriptiva separadas por grupos de fracasada=1 y no fracasadas=0 sin
winsorizar**;
. tabstat R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8__bn_amo~t R9_margen_sobre_ventas__bn_incn R1
> 0_rotacion_activos_incn_at incidencias num_empl_total if fracaso==1, statistics(
count mean median sd skewness ) columns(sta

```

```
> tistics);
```

variable	N	mean	p50	sd	skewness
R1_dt_at	37	1.09844	.9751387	1.010565	3.7524
R2_ac_pc	37	3.23058	1.057278	11.17866	5.727102
R3_bait_at	37	-.2554667	-.0164002	.7124744	-2.656331
R4_bn_at	37	-.2745696	-.0354618	.6439843	-2.230388
R5_ac_at	37	.774168	.8938961	.2734852	-1.460193
R6_gf_dt	37	.4296611	.032181	2.422027	5.832577
R7_bnd_at	37	.1619858	.0299024	.6085422	3.317699
R8__bn_amo~t	37	1.418611	-.0286602	14.66338	4.637255
R9_margen_~n	37	-.2360054	-.0441338	.5619837	-3.127511
R10_rotaci~t	37	2.612674	1.475444	5.620027	5.245937
incidencias	37	.6216216	0	.720777	.694475
num_empl_t~1	16	12.1875	9	11.55404	1.095393

```
. tabstat R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8__bn_amo~t R9_margen_sobre_ventas__bn_incn R1
> 0_rotacion_activos_incn_at incidencias num_empl_total if fracaso==0, statistics(
count mean median sd skewness ) columns(sta
> tistics);
```

variable	N	mean	p50	sd	skewness
R1_dt_at	3193	1.464248	.7783218	35.13914	56.26852
R2_ac_pc	3193	4.992888	1.274884	65.81954	30.69193
R3_bait_at	3193	-.0345993	.0444596	2.19314	-41.647
R4_bn_at	3193	-.0876411	.0159537	3.849023	-51.79087
R5_ac_at	3193	.7268043	.7864155	.3425963	9.195928
R6_gf_dt	3193	.0319486	.0214282	.0583653	13.70667
R7_bnd_at	3193	.810277	.1062388	32.54369	56.31467
R8__bn_amo~t	3193	.1890471	.0608538	2.319081	23.02114
R9_margen_~n	3193	-.5259805	.0124977	26.23841	-48.29314
R10_rotaci~t	3193	1.815489	1.381083	22.46602	18.59462
incidencias	3193	.0720326	0	.3133578	4.695205
num_empl_t~1	1955	32.46496	6	588.1141	36.44888

```
. ** estadística descriptiva separadas por grupos de fracasada=1 y no fracasadas=0
datos winsorizar**;
```

```
. tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total if fracaso==0, statistics( count mean median sd
> skewness) columns(statistics);
```

variable	N	mean	p50	sd	skewness
wR1	3193	.7876563	.7783218	.4490908	2.483903
wR2	3193	2.783909	1.274884	5.769318	5.38778

```

wR3 | 3193 .022242 .0444596 .220357 -2.688866
wR4 | 3193 -.0056056 .0159537 .1911801 -3.040909
wR5 | 3193 .7263177 .7864155 .2373353 -.9497068
wR6 | 3193 .0297533 .0214282 .032074 1.978609
wR7 | 3193 .1956719 .1062388 .3814027 2.493171
wR8 | 3193 .1211922 .0608538 .3389109 2.552918
wR9 | 3193 -.0638995 .0124977 .475318 -6.155788
wR10 | 3193 1.661431 1.381083 1.448189 2.268183
incidencias | 3193 .0720326 0 .3133578 4.695205
num_empl_t~1 | 1955 32.46496 6 588.1141 36.44888

```

```

. tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total if fracaso==0, statistics( count mean median sd
> skewness) columns(statistics);

```

```

variable | N mean p50 sd skewness
-----+-----
wR1 | 3193 .7876563 .7783218 .4490908 2.483903
wR2 | 3193 2.783909 1.274884 5.769318 5.38778
wR3 | 3193 .022242 .0444596 .220357 -2.688866
wR4 | 3193 -.0056056 .0159537 .1911801 -3.040909
wR5 | 3193 .7263177 .7864155 .2373353 -.9497068
wR6 | 3193 .0297533 .0214282 .032074 1.978609
wR7 | 3193 .1956719 .1062388 .3814027 2.493171
wR8 | 3193 .1211922 .0608538 .3389109 2.552918
wR9 | 3193 -.0638995 .0124977 .475318 -6.155788
wR10 | 3193 1.661431 1.381083 1.448189 2.268183
incidencias | 3193 .0720326 0 .3133578 4.695205
num_empl_t~1 | 1955 32.46496 6 588.1141 36.44888

```

```

. ** Diferencia de medias por variable separado por fracaso =1 y fracaso=0, nivel de
significación del 99%**;
. ttest R1_dt_at, by(fracaso) level(99);

```

Two-sample t test with equal variances

```

-----+-----
Group | Obs Mean Std. Err. Std. Dev. [99% Conf. Interval]
-----+-----
0 | 3193 1.464248 .6218586 35.13914 -.1385122 3.067008
1 | 37 1.09844 .1661359 1.010565 .6466357 1.550243
-----+-----
combined | 3230 1.460057 .6147373 34.93742 -.1243375 3.044452
-----+-----
diff | .3658081 5.77775 -14.52549 15.25711
-----+-----
diff = mean(0) - mean(1) t = 0.0633
Ho: diff = 0 degrees of freedom = 3228

```

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.5252 Pr(|T| > |t|) = 0.9495 Pr(T > t) = 0.4748

. ttest R2_ac_pc, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  4.992888  1.164811  65.81954  1.990739  7.995037
1 | 37    3.23058  1.83776  11.17866  -1.767181  8.228341
-----+-----
combined | 3230  4.972701  1.151658  65.45227  2.004473  7.940929
-----+-----
diff |      1.762308  10.82408          -26.13518  29.65979
-----+-----
diff = mean(0) - mean(1)          t = 0.1628
Ho: diff = 0                      degrees of freedom = 3228
```

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.5647 Pr(|T| > |t|) = 0.8707 Pr(T > t) = 0.4353

. ttest R3_bait_at, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  -.0345993  .0388121  2.19314  -.1346324  .0654338
1 | 37    -.2554667  .1171301  .7124744  -.5740001  .0630667
-----+-----
combined | 3230  -.0371293  .0383925  2.181963  -.1360803  .0618216
-----+-----
diff |      .2208674  .3608198          -.7090926  1.150827
-----+-----
diff = mean(0) - mean(1)          t = 0.6121
Ho: diff = 0                      degrees of freedom = 3228
```

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.7298 Pr(|T| > |t|) = 0.5405 Pr(T > t) = 0.2702

. ttest R4_bn_at, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  -.0876411  .0681163  3.849023  -.263202  .0879199
1 | 37    -.2745696  .1058704  .6439843  -.5624824  .0133433
-----+-----
```

```

combined | 3230 -.0897824 .0673474 3.827563 -.2633605 .0837957
-----+-----
diff | .1869285 .6329722 -1.444465 1.818322
-----+-----
diff = mean(0) - mean(1) t = 0.2953
Ho: diff = 0 degrees of freedom = 3228

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.6161 Pr(|T| > |t|) = 0.7678 Pr(T > t) = 0.3839

```

. ttest R5_ac_at, by(fracaso) level(99);

Two-sample t test with equal variances

```

-----+-----
Group | Obs Mean Std. Err. Std. Dev. [99% Conf. Interval]
-----+-----
0 | 3193 .7268043 .0060629 .3425963 .7111779 .7424308
1 | 37 .774168 .0449607 .2734852 .6518981 .8964379
-----+-----
combined | 3230 .7273469 .0060156 .3418868 .7118425 .7428513
-----+-----
diff | -.0473637 .0565332 -.1930696 .0983422
-----+-----
diff = mean(0) - mean(1) t = -0.8378
Ho: diff = 0 degrees of freedom = 3228

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.2011 Pr(|T| > |t|) = 0.4022 Pr(T > t) = 0.7989

```

. ttest R6_gf_dt, by(fracaso) level(99);

Two-sample t test with equal variances

```

-----+-----
Group | Obs Mean Std. Err. Std. Dev. [99% Conf. Interval]
-----+-----
0 | 3193 .0319486 .0010329 .0583653 .0292865 .0346108
1 | 37 .4296611 .3981788 2.422027 -.6531801 1.512502
-----+-----
combined | 3230 .0365045 .0046739 .265634 .0244581 .0485508
-----+-----
diff | -.3977125 .0433677 -.5094864 -.2859385
-----+-----
diff = mean(0) - mean(1) t = -9.1707
Ho: diff = 0 degrees of freedom = 3228

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

```

. ttest R7_bnd_at, by(fracaso) level(99);

Two-sample t test with equal variances

```

-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193   .810277   .5759268   32.54369   -.6740998   2.294654
1 | 37     .1619858  .1000437   .6085422   -.1100816   .4340532
-----+-----
combined | 3230   .8028507  .5693309   32.35684   -.6645159   2.270217
-----+-----
diff |           .6482912   5.350979           -13.14307   14.43965
-----+-----
diff = mean(0) - mean(1)                                t = 0.1212
Ho: diff = 0                                           degrees of freedom = 3228

Ha: diff < 0      Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 0.5482  Pr(|T| > |t|) = 0.9036  Pr(T > t) = 0.4518

```

```
. ttest R8__bn__amor__dt, by(fracaso) level(99);
```

Two-sample t test with equal variances

```

-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193   .1890471  .0410409   2.319081   .0832696   .2948246
1 | 37     1.418611  2.410645   14.66338   -5.137102  7.974324
-----+-----
combined | 3230   .2031319  .0489228   2.780436   .0770405   .3292233
-----+-----
diff |           -1.229564   .4593034           -2.413351   -.0457769
-----+-----
diff = mean(0) - mean(1)                                t = -2.6770
Ho: diff = 0                                           degrees of freedom = 3228

Ha: diff < 0      Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 0.0037  Pr(|T| > |t|) = 0.0075  Pr(T > t) = 0.9963

```

```
. ttest R9__margen__sobre__ventas__bn__incn, by(fracaso) level(99);
```

Two-sample t test with equal variances

```

-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  -.5259805  .464342   26.23841  -1.722762   .670801
1 | 37    -2.360054  .0923896   .5619837  -4.872573   .0152466
-----+-----
combined | 3230  -.5226588  .4590236   26.08773  -1.705725   .660407
-----+-----
diff |           -2.899751   4.314239           -11.40929   10.82934
-----+-----
diff = mean(0) - mean(1)                                t = -0.0672

```



```

1 | 16 12.1875 2.88851 11.55404 3.675891 20.69911
-----+-----
combined | 1971 32.30036 13.1932 585.7246 -1.716035 66.31675
-----+-----
diff | 20.27746 147.0658 -358.9064 399.4613
-----+-----

```

```

diff = mean(0) - mean(1)          t = 0.1379
Ho: diff = 0                      degrees of freedom = 1969

```

```

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.5548    Pr(|T| > |t|) = 0.8903    Pr(T > t) = 0.4452

```

```

. ** para el cálculo de la mediana, la desviación típica y la asimetría **:
. tabstat R1_dt_at R2_ac_pc R5_ac_at R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R10_rotacion_activos_incn_at R8__bn_amor_
> _dt R9_margen_sobre_ventas__bn_incn R10_rotacion_activos_incn_at incidencias
num_empl_total, statistics( median sd skewness
> ) by(fracaso) casewise nototal labelwidth(0) varwidth(0) columns(statistics)
longstubs;
(option varwidth() outside valid range 8..16; 8 assumed)
(option labelwidth() outside valid range 8..32; 8 assumed)

```

```

fracaso variable | p50 sd skewness
-----+-----
0  R1_dt_at | .7725224 1.978093 10.4451
  R2_ac_pc | 1.253113 13.34974 15.50559
  R5_ac_at | .7687752 .361062 15.17528
  R3_bai~t | .0498902 .9310103 .6161073
  R4_bn_at | .0174166 .8417764 10.04655
  R5_ac_at | .7687752 .361062 15.17528
  R6_gf_dt | .0232892 .0513251 10.38918
  R7_bnd~t | .1135342 1.740371 9.410856
  R10_ro~t | 1.511289 16.98466 -40.97037
  R8__bn~t | .0693856 .6098472 -.794612
  R9_mar~n | .0123443 32.87636 -40.26592
  R10_ro~t | 1.511289 16.98466 -40.97037
  incide~s | 0 .319205 4.604918
  num_em~l | 6 588.1141 36.44888
-----+-----
1  R1_dt_at | .970955 1.378005 3.264092
  R2_ac_pc | 1.007679 16.96185 3.592969
  R5_ac_at | .8323201 .2881027 -1.415927
  R3_bai~t | -.0281351 .5194673 -2.867472
  R4_bn_at | -.0621735 .5376328 -3.134699
  R5_ac_at | .8323201 .2881027 -1.415927
  R6_gf_dt | .0335711 3.682029 3.614571
  R7_bnd~t | .0242397 .8190054 3.032377
  R10_ro~t | 1.553762 1.133234 .2333553
  R8__bn~t | -.0594517 20.83215 3.614197
  R9_mar~n | -.0697925 .4799963 -2.513418

```

```
R10_ro~t | 1.553762 1.133234 .2333553
incide~s | 1 .7719024 .1017525
num_em~l | 9 11.55404 1.095393
```

```
-----
. ** para el cálculo del test de Wilcoxon**;
```

```
. ranksum R1_dt_at, by(fracaso);
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5138266	5158291.5
1	37	79799	59773.5
combined	3230	5218065	5218065

```
unadjusted variance 31809464
adjustment for ties -.0056637
```

```
-----
adjusted variance 31809464
```

Ho: $R1_dt_at(\text{fracaso}=0) = R1_dt_at(\text{fracaso}=1)$

$z = -3.551$

Prob > |z| = 0.0004

```
. ranksum R2_ac_pc, by(fracaso);
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5167963	5158291.5
1	37	50102	59773.5
combined	3230	5218065	5218065

```
unadjusted variance 31809464
adjustment for ties -.0056637
```

```
-----
adjusted variance 31809464
```

Ho: $R2_ac_pc(\text{fracaso}=0) = R2_ac_pc(\text{fracaso}=1)$

$z = 1.715$

Prob > |z| = 0.0864

```
. ranksum R3_bait_at, by(fracaso);
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5175738	5158291.5
1	37	42327	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties 0

adjusted variance 31809464

Ho: $R3_bai \sim t(\text{fracaso}==0) = R3_bai \sim t(\text{fracaso}==1)$
 $z = 3.093$
 Prob > |z| = 0.0020

. ranksum R4_bn_at, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5181484	5158291.5
1	37	36581	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -.19822942

adjusted variance 31809464

Ho: $R4_bn_at(\text{fracaso}==0) = R4_bn_at(\text{fracaso}==1)$
 $z = 4.112$
 Prob > |z| = 0.0000

. ranksum R5_ac_at, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5147141.5	5158291.5
1	37	70923.5	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -888.35101

adjusted variance 31808576

Ho: R5_ac_at(fracaso==0) = R5_ac_at(fracaso==1)

z = -1.977

Prob > |z| = 0.0480

. ranksum R6_gf_dt, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso		obs	rank sum	expected
-----+-----				
0		3193	5149038	5158291.5
1		37	69027	59773.5
-----+-----				
combined		3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -23501.322

adjusted variance 31785963

Ho: R6_gf_dt(fracaso==0) = R6_gf_dt(fracaso==1)

z = -1.641

Prob > |z| = 0.1007

. ranksum R7_bnd_at, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso		obs	rank sum	expected
-----+-----				
0		3193	5171105.5	5158291.5
1		37	46959.5	59773.5
-----+-----				
combined		3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -44.012596

adjusted variance 31809420

Ho: R7_bnd~t(fracaso==0) = R7_bnd~t(fracaso==1)

z = 2.272

Prob > |z| = 0.0231

. ranksum R8__bn_amor__dt, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5183537	5158291.5
1	37	34528	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -.0056637

adjusted variance 31809464

Ho: $R8_bn \sim t(\text{fracaso}==0) = R8_bn \sim t(\text{fracaso}==1)$
 $z = 4.476$
 Prob > |z| = 0.0000

. ranksum R9_margen_sobre_ventas__bn_inc, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5182556	5158291.5
1	37	35509	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -.19822942

adjusted variance 31809464

Ho: $R9_mar \sim n(\text{fracaso}==0) = R9_mar \sim n(\text{fracaso}==1)$
 $z = 4.302$
 Prob > |z| = 0.0000

. ranksum R10_rotacion_activos_inc_at, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5155830	5158291.5
1	37	62235	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties 0

adjusted variance 31809464

Ho: $R10_ro \sim t(\text{fracaso}==0) = R10_ro \sim t(\text{fracaso}==1)$

z = -0.436

Prob > |z| = 0.6625

. ranksum incidencias, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5132884.5	5158291.5
1	37	85180.5	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -26313858

adjusted variance 5495606.6

Ho: $incide \sim s(\text{fracaso}==0) = incide \sim s(\text{fracaso}==1)$

z = -10.838

Prob > |z| = 0.0000

. ranksum num_empl_total, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	1955	1926009.5	1927630
1	16	17396.5	15776
combined	1971	1943406	1943406

unadjusted variance 5140346.67

adjustment for ties -24183.74

adjusted variance 5116162.93

Ho: $num_empl \sim l(\text{fracaso}==0) = num_empl \sim l(\text{fracaso}==1)$

z = -0.716

Prob > |z| = 0.4737

. ** cálculo LDA con 10 variables**;

. discrim lda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8_bn amor_dt R9_margen_sobre_ventas_bn_inc
> n R10_rotacion_activos_incn_at, group(fracaso);

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,158  35 | 3,193
          | 98.90  1.10 | 100.00
          |         |
          1 | 35     2 | 37
          | 94.59  5.41 | 100.00
-----+-----+-----
          Total | 3,193  37 | 3,230
          | 98.85  1.15 | 100.00
          |         |
          Priors | 0.5000 0.5000 |

```

```

. ** cálculo LDA con 10 variables+incidencias**;
. discrim lda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_inc
> n R10_rotacion_activos_incn_at incidencias, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,003  190 | 3,193
          | 94.05  5.95 | 100.00
          |         |
          1 | 17     20 | 37
          | 45.95  54.05 | 100.00
-----+-----+-----
          Total | 3,020  210 | 3,230
          | 93.50  6.50 | 100.00
          |         |

```

Priors | 0.5000 0.5000 |

```
. ** cálculo LDA con 7 variables**;  
. discrim lda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor_dt  
R9_margen_sobre_ventas__bn_incn, group(fracaso);
```

Linear discriminant analysis
Resubstitution classification summary

```
+-----+  
| Key |  
|-----|  
| Number |  
| Percent |  
+-----+  
| Classified  
True fracaso | 0 1 | Total  
-----+-----+-----  
0 | 2,636 557 | 3,193  
| 82.56 17.44 | 100.00  
| |  
1 | 18 19 | 37  
| 48.65 51.35 | 100.00  
-----+-----+-----  
Total | 2,654 576 | 3,230  
| 82.17 17.83 | 100.00  
| |  
Priors | 0.5000 0.5000 |
```

```
. ** cálculo LDA con 7 variables+incidencias**;  
. discrim lda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor_dt  
R9_margen_sobre_ventas__bn_incn incidencias, gro  
> up(fracaso);
```

Linear discriminant analysis
Resubstitution classification summary

```
+-----+  
| Key |  
|-----|  
| Number |  
| Percent |  
+-----+  
| Classified  
True fracaso | 0 1 | Total  
-----+-----+-----  
0 | 3,004 189 | 3,193  
| 94.08 5.92 | 100.00  
| |  
1 | 18 19 | 37
```

```

          | 48.65  51.35 | 100.00
-----+-----+-----
      Total | 3,022  208 | 3,230
          | 93.56  6.44 | 100.00
          |           |
      Priors | 0.5000  0.5000 |

```

```

.** cálculo LDA con 5 variables**;
.   discrim   lda   R1_dt_at   R3_bait_at   R4_bn_at   R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,102  91 | 3,193
          | 97.15  2.85 | 100.00
          |           |
          1 | 31    6 | 37
          | 83.78 16.22 | 100.00
-----+-----+-----
      Total | 3,133  97 | 3,230
          | 97.00  3.00 | 100.00
          |           |
      Priors | 0.5000  0.5000 |

```

```

.** cálculo LDA con 5 variables+incidencias**;
.   discrim   lda   R1_dt_at   R3_bait_at   R4_bn_at   R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----

```

```

      0 | 3,005   188 | 3,193
        | 94.11   5.89 | 100.00
        |         |
      1 |   18    19 |   37
        | 48.65  51.35 | 100.00
-----+-----+-----
Total | 3,023   207 | 3,230
        | 93.59   6.41 | 100.00
        |         |
Priors | 0.5000  0.5000 |

```

```

. ** cálculo LDA con 2 variables**;
. discrim lda R6_gf_dt R8__bn_amor__dt, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 3,171   22 | 3,193
        | 99.31   0.69 | 100.00
        |         |
      1 |   35    2 |   37
        | 94.59   5.41 | 100.00
-----+-----+-----
Total | 3,206   24 | 3,230
        | 99.26   0.74 | 100.00
        |         |
Priors | 0.5000  0.5000 |

```

```

. ** cálculo LDA con 2 variables+incidencias**;
. discrim lda R6_gf_dt R8__bn_amor__dt incidencias, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified

```

True fracaso	0	1	Total
0	3,007	186	3,193
	94.17	5.83	100.00
1	17	20	37
	45.95	54.05	100.00
Total	3,024	206	3,230
	93.62	6.38	100.00
Priors	0.5000	0.5000	

```

. ** cálculo QDA con 10 variables**;
. discrim qda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_inc
> n R10_rotacion_activos_incn_at, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

+-----+				
Key				

Number				
Percent				
+-----+				
Classified				
True fracaso	0	1	Unclassified	Total
0	378	2,809	6	3,193
	11.84	87.97	0.19	100.00
1	1	36	0	37
	2.70	97.30	0.00	100.00
Total	379	2,845	6	3,230
	11.73	88.08	0.19	100.00
Priors	0.5000	0.5000		

```

. ** cálculo QDA con 10 variables+incidencias**;
. discrim qda R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_inc
> n R10_rotacion_activos_incn_at incidencias, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

+-----+

```

| Key |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 Unclassified | Total
+-----+-----+-----+
0 | 390 2,797      6 | 3,193
  | 12.21 87.60    0.19 | 100.00
  |
1 | 0 37          0 | 37
  | 0.00 100.00    0.00 | 100.00
+-----+-----+-----+
Total | 390 2,834      6 | 3,230
      | 12.07 87.74    0.19 | 100.00
      |
Priors | 0.5000 0.5000      |

```

```

.** cálculo QDA con 7 variables**;
. discrim qda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 Unclassified | Total
+-----+-----+-----+
0 | 134 3,055      4 | 3,193
  | 4.20 95.68    0.13 | 100.00
  |
1 | 2 35          0 | 37
  | 5.41 94.59    0.00 | 100.00
+-----+-----+-----+
Total | 136 3,090      4 | 3,230
      | 4.21 95.67    0.12 | 100.00
      |
Priors | 0.5000 0.5000      |

```

```

.** cálculo QDA con 7 variables+incidencias**;
. discrim qda R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, gro
> up(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 Unclassified | Total
-----+-----+-----+-----+
      0 | 142 3,047      4 | 3,193
      | 4.45 95.43    0.13 | 100.00
      |
      1 | 1    36      0 | 37
      | 2.70 97.30    0.00 | 100.00
-----+-----+-----+-----+
      Total | 143 3,083      4 | 3,230
      | 4.43 95.45    0.12 | 100.00
      |
      Priors | 0.5000 0.5000      |

```

```

.** cálculo QDA con 5 variables**;
. discrim qda R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 Unclassified | Total
-----+-----+-----+-----+
      0 | 98 3,092      3 | 3,193
      | 3.07 96.84    0.09 | 100.00
      |
      1 | 2    35      0 | 37
      | 5.41 94.59    0.00 | 100.00
-----+-----+-----+-----+
      Total | 100 3,127      3 | 3,230
      | 3.10 96.81    0.09 | 100.00
      |
      Priors | 0.5000 0.5000      |

```

```
. ** cálculo QDA con 5 variables+incidencias**;  
.   discrim   qda   R1_dt_at   R3_bait_at   R4_bn_at   R8__bn_amor__dt  
R9_margen_sobre_ventas__bn__incn incidencias, group(fracaso);
```

Quadratic discriminant analysis
Resubstitution classification summary

```
+-----+  
| Key   |  
|-----|  
| Number |  
| Percent |  
+-----+  
      | Classified  
True fracaso | 0    1 Unclassified | Total  
-----+-----+-----+  
      0 | 103 3,087      3 | 3,193  
      | 3.23 96.68      0.09 | 100.00  
      |  
      1 | 2    35          0 | 37  
      | 5.41 94.59      0.00 | 100.00  
-----+-----+-----+  
      Total | 105 3,122      3 | 3,230  
      | 3.25 96.66      0.09 | 100.00  
      |  
      Priors | 0.5000 0.5000      |
```

```
. ** cálculo QDA con 2 variables**;  
.   discrim qda R6_gf_dt R8__bn_amor__dt, group(fracaso);
```

Quadratic discriminant analysis
Resubstitution classification summary

```
+-----+  
| Key   |  
|-----|  
| Number |  
| Percent |  
+-----+  
      | Classified  
True fracaso | 0    1 | Total  
-----+-----+-----+  
      0 | 3,161 32 | 3,193  
      | 99.00 1.00 | 100.00  
      |  
      1 | 35 2 | 37  
      | 94.59 5.41 | 100.00  
-----+-----+-----+  
      Total | 3,196 34 | 3,230
```

```

          | 98.95  1.05 | 100.00
          |          |
Priors | 0.5000  0.5000 |

```

```

. ** cálculo QDA con 2 variables+incidencias**;
. discrim qda R6_gf_dt R8__bn__amor__dt incidencias, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,111  82 | 3,193
          | 97.43  2.57 | 100.00
          |          |
          1 | 30    7 | 37
          | 81.08  18.92 | 100.00
-----+-----+-----
          Total | 3,141  89 | 3,230
          | 97.24  2.76 | 100.00
          |          |
Priors | 0.5000  0.5000 |

```

```

. ** cálculo LogDA con 10 variables**;
. discrim logistic R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn__amor__dt R9_margen_sobre_ventas__b
> n_incn R10_rotacion_activos_incn_at, group(fracaso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -196.55623
Iteration 2: log likelihood = -191.37988
Iteration 3: log likelihood = -190.07137
Iteration 4: log likelihood = -189.40395
Iteration 5: log likelihood = -189.21018
Iteration 6: log likelihood = -189.1853
Iteration 7: log likelihood = -189.18336
Iteration 8: log likelihood = -189.18335

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |

```

```

|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 1,896  1,297 | 3,193
      | 59.38  40.62 | 100.00
      |
      1 | 11    26 | 37
      | 29.73  70.27 | 100.00
-----+-----+-----
      Total | 1,907  1,323 | 3,230
      | 59.04  40.96 | 100.00
      |
      Priors | 0.5000  0.5000 |

```

```

. ** cálculo LogDA con 10 variables+incidencias**;
. discrim logistic R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8__bn_amor_dt R9_margen_sobre_ventas__b
> n_incn R10_rotacion_activos_incn_at incidencias, group(fracaso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -175.40438
Iteration 2: log likelihood = -173.48133
Iteration 3: log likelihood = -169.37731
Iteration 4: log likelihood = -168.57415
Iteration 5: log likelihood = -168.35417
Iteration 6: log likelihood = -168.31448
Iteration 7: log likelihood = -168.31229
Iteration 8: log likelihood = -168.30652
Iteration 9: log likelihood = -168.30527
Iteration 10: log likelihood = -168.30527

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,921  272 | 3,193
      | 91.48  8.52 | 100.00
      |

```

```

      1 | 14 23 | 37
      | 37.84 62.16 | 100.00
-----+-----+-----
Total | 2,935 295 | 3,230
      | 90.87 9.13 | 100.00
      |           |
Priors | 0.5000 0.5000 |

```

. ** cálculo LogDA con 7 variables**;

```

. discrim logistic R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_incn, group(frac
> aso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -199.8168
Iteration 2: log likelihood = -199.00873
Iteration 3: log likelihood = -198.97242
Iteration 4: log likelihood = -198.97109
Iteration 5: log likelihood = -198.97005
Iteration 6: log likelihood = -198.96667
Iteration 7: log likelihood = -198.95752
Iteration 8: log likelihood = -198.95737
Iteration 9: log likelihood = -198.95737

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0 1 | Total
-----+-----+-----
      0 | 1,691 1,502 | 3,193
      | 52.96 47.04 | 100.00
      |           |
      1 | 9 28 | 37
      | 24.32 75.68 | 100.00
-----+-----+-----
Total | 1,700 1,530 | 3,230
      | 52.63 47.37 | 100.00
      |           |
Priors | 0.5000 0.5000 |

```

. ** cálculo LogDA con 7 variables+incidencias**;

```

. discrim logistic R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_incn incidencias

```

```
> , group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -180.15585
Iteration 2: log likelihood = -177.8329
Iteration 3: log likelihood = -177.6849
Iteration 4: log likelihood = -177.67959
Iteration 5: log likelihood = -177.66603
Iteration 6: log likelihood = -177.58002
Iteration 7: log likelihood = -177.56506
Iteration 8: log likelihood = -177.56496
Iteration 9: log likelihood = -177.56496
```

Logistic discriminant analysis
Resubstitution classification summary

```
+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,998  195 | 3,193
        | 93.89  6.11 | 100.00
        |         |
      1 | 18    19 | 37
        | 48.65  51.35 | 100.00
-----+-----+-----
      Total | 3,016  214 | 3,230
            | 93.37  6.63 | 100.00
            |         |
      Priors | 0.5000  0.5000 |
```

```
. ** cálculo LogDA con 5 variables**;
.   discrim    logistic    R1_dt_at    R3_bait_at    R4_bn_at    R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -200.32511
Iteration 2: log likelihood = -200.28314
Iteration 3: log likelihood = -200.27961
Iteration 4: log likelihood = -200.27784
Iteration 5: log likelihood = -200.27536
Iteration 6: log likelihood = -200.26819
Iteration 7: log likelihood = -200.26766
Iteration 8: log likelihood = -200.26766
Iteration 9: log likelihood = -200.26766
```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 2,952  241 | 3,193
          | 92.45  7.55 | 100.00
          |         |
          1 | 31     6 | 37
          | 83.78 16.22 | 100.00
-----+-----+-----
          Total | 2,983  247 | 3,230
          | 92.35  7.65 | 100.00
          |         |
          Priors | 0.5000 0.5000 |

```

```

. ** cálculo LogDA con 5 variables+incidencias**;
. discrim logistic R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, group(fracaso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -180.80534
Iteration 2: log likelihood = -178.94009
Iteration 3: log likelihood = -178.87852
Iteration 4: log likelihood = -178.87236
Iteration 5: log likelihood = -178.83884
Iteration 6: log likelihood = -178.78317
Iteration 7: log likelihood = -178.77294
Iteration 8: log likelihood = -178.77273
Iteration 9: log likelihood = -178.77273

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total

```

```

-----+-----+-----
      0 | 3,006  187 | 3,193
        | 94.14  5.86 | 100.00
        |         |
      1 |   18   19 |   37
        | 48.65  51.35 | 100.00
-----+-----+-----
    Total | 3,024  206 | 3,230
          | 93.62  6.38 | 100.00
          |         |
    Priors | 0.5000  0.5000 |

```

```

. ** cálculo LogDA con 2 variables**;
. discrim logistic R6_gf_dt R8__bn_amor__dt, group(fracaso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -197.55943
Iteration 2: log likelihood = -194.57693
Iteration 3: log likelihood = -194.03114
Iteration 4: log likelihood = -193.96646
Iteration 5: log likelihood = -193.95343
Iteration 6: log likelihood = -193.95302
Iteration 7: log likelihood = -193.95302

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,839  354 | 3,193
        | 88.91  11.09 | 100.00
        |         |
      1 |   29   8 |   37
        | 78.38  21.62 | 100.00
-----+-----+-----
    Total | 2,868  362 | 3,230
          | 88.79  11.21 | 100.00
          |         |
    Priors | 0.5000  0.5000 |

```

```

. ** cálculo LogDA con 2 variables+incidencias**;
. discrim logistic R6_gf_dt R8__bn_amor__dt incidencias, group(fracaso);

```

Iteration 0: log likelihood = -202.15209
 Iteration 1: log likelihood = -177.72083
 Iteration 2: log likelihood = -174.50683
 Iteration 3: log likelihood = -172.86361
 Iteration 4: log likelihood = -172.83497
 Iteration 5: log likelihood = -172.83496

Logistic discriminant analysis
 Resubstitution classification summary

```
+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 2,995  198 | 3,193
          | 93.80  6.20 | 100.00
          |         |
          1 | 17    20 | 37
          | 45.95 54.05 | 100.00
-----+-----+-----
          Total | 3,012  218 | 3,230
          | 93.25  6.75 | 100.00
          |         |
          Priors | 0.5000 0.5000 |
```

```
. ** cálculo KnnDA con 10 variables**;  

. discrim knn R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at  

R8_bn_amor_dt R9_margen_sobre_ventas_bn_inc  

> n R10_rotacion_activos_incn_at, k(2) group(fracaso) measure(L2);
```

Kth-nearest-neighbor discriminant analysis
 Resubstitution classification summary

```
+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,151  42 | 3,193
          | 98.68  1.32 | 100.00
          |         |
```

```

      1 |  0  37 |  37
        | 0.00 100.00 | 100.00
-----+-----+-----
Total | 3,151  79 | 3,230
        | 97.55  2.45 | 100.00
        |           |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo knnDA con 10 variables+incidencias**;
. discrim knn R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas__bn_inc
> n R10_rotacion_activos_incn_at incidencias, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso |  0  1 | Total
-----+-----+-----
      0 | 3,163  30 | 3,193
        | 99.06  0.94 | 100.00
        |           |
      1 |  0  37 |  37
        | 0.00 100.00 | 100.00
-----+-----+-----
Total | 3,163  67 | 3,230
        | 97.93  2.07 | 100.00
        |           |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo knnDA con 7 variables**;
. discrim knn R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, k(2) group(frac
> aso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+

```

		Classified		
True fracaso	0	1	Total	
0	3,153	40	3,193	
	98.75	1.25	100.00	
1	0	37	37	
	0.00	100.00	100.00	
Total	3,153	77	3,230	
	97.62	2.38	100.00	
Priors	0.5000	0.5000		

```

. ** cálculo KnnDA con 7 variables+incidencias**;
. discrim knn R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, k(2)
> ) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

		Classified		
True fracaso	0	1	Total	
0	3,161	32	3,193	
	99.00	1.00	100.00	
1	0	37	37	
	0.00	100.00	100.00	
Total	3,161	69	3,230	
	97.86	2.14	100.00	
Priors	0.5000	0.5000		

```

. ** cálculo KnnDA con 5 variables**;
. discrim knn R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+

```

```

| Key |
|-----|
| Number |
| Percent |
+-----+
| Classified
True fracaso | 0 1 | Total
-----+-----+-----
0 | 3,163 30 | 3,193
| 99.06 0.94 | 100.00
| | |
1 | 0 37 | 37
| 0.00 100.00 | 100.00
-----+-----+-----
Total | 3,163 67 | 3,230
| 97.93 2.07 | 100.00
| |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 5 variables+incidencias**;
. discrim knn R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias, k(2) group(fracaso) me
> asure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key |
|-----|
| Number |
| Percent |
+-----+
| Classified
True fracaso | 0 1 | Total
-----+-----+-----
0 | 3,156 37 | 3,193
| 98.84 1.16 | 100.00
| | |
1 | 0 37 | 37
| 0.00 100.00 | 100.00
-----+-----+-----
Total | 3,156 74 | 3,230
| 97.71 2.29 | 100.00
| |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 2 variables**;
. discrim knn R6_gf_dt R8__bn_amor__dt, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number|
| Percent|
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 3,150  43 | 3,193
        | 98.65  1.35 | 100.00
        |         |
      1 |    0    37 |    37
        | 0.00 100.00 | 100.00
-----+-----+-----
      Total | 3,150  80 | 3,230
            | 97.52  2.48 | 100.00
            |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 2 variables+incidencias**;
. discrim knn R6_gf_dt R8__bn_amor__dt incidencias, k(2) group(fracaso)
measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number|
| Percent|
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 3,160  33 | 3,193
        | 98.97  1.03 | 100.00
        |         |
      1 |    0    37 |    37
        | 0.00 100.00 | 100.00
-----+-----+-----
      Total | 3,160  70 | 3,230
            | 97.83  2.17 | 100.00
            |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo de logit 10 variables **;
. logit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8_bn_amo_dt R9_margen_sobre_ventas_bn_i
> ncn R10_rotacion_activos_incn_at;

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -196.55623
Iteration 2: log likelihood = -191.37988
Iteration 3: log likelihood = -190.07137
Iteration 4: log likelihood = -189.38897
Iteration 5: log likelihood = -189.18564
Iteration 6: log likelihood = -189.18336
Iteration 7: log likelihood = -189.18335

```

```

Logistic regression                Number of obs =    3230
                                LR chi2(10)  =    25.94
                                Prob > chi2  =    0.0038
Log likelihood = -189.18335        Pseudo R2   =    0.0642

```

```

-----+-----
fracaso |   Coef.   Std. Err.      z    P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | .5941187   .2998368     1.98  0.048   .0064494   1.181788
R2_ac_pc | -.0042892   .0047589    -0.90  0.367  -.0136164   .0050381
R3_bait_at | .8904739   .8681268     1.03  0.305  -.8110233   2.591971
R4_bn_at | -.4111532   .6851729    -0.60  0.548  -1.754067   .9317611
R5_ac_at | 1.597136   .8345098     1.91  0.056  -.038473   3.232745
R6_gf_dt | 1.888189   .4591328     4.11  0.000   .9883052   2.788073
R7_bnd_at | -.650794   .3246748    -2.00  0.045  -1.287145  -.0144431
R8_bn_amo~t | -.2362523   .0702227    -3.36  0.001  -.3738862  -.0986183
R9_margen_~n | .0006848   .0094895     0.07  0.942  -.0179142   .0192838
R10_rotaci~t | .0250092   .0338547     0.74  0.460  -.0413449   .0913632
   _cons | -6.242101   .7345036    -8.50  0.000  -7.681702  -4.802501
-----+-----

```

Note: 3 failures and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for logit of fracaso

```

Log-Lik Intercept Only: -202.152   Log-Lik Full Model:    -189.183
D(3219):                378.367   LR(10):                25.937
                        Prob > LR:           0.004
McFadden's R2:          0.064   McFadden's Adj R2:      0.010
Maximum Likelihood R2:  0.008   Cragg & Uhler's R2:    0.068
McKelvey and Zavoina's R2: 0.212   Efron's R2:            0.041
Variance of y*:         4.176   Variance of error:      3.290
Count R2:               0.989   Adj Count R2:          0.054
AIC:                   0.124   AIC*n:                 400.367
BIC:                   -25631.918   BIC':                  54.865

```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.7101
```

```
. estat classification;
```

```
Logistic model for fracaso
```

		----- True -----		
Classified	D	~D	Total	
+	2	0	2	
-	35	3193	3228	
Total	37	3193	3230	

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

Sensitivity	Pr(+ D)	5.41%
Specificity	Pr(- ~D)	100.00%
Positive predictive value	Pr(D +)	100.00%
Negative predictive value	Pr(~D -)	98.92%
False + rate for true ~D	Pr(+ ~D)	0.00%
False - rate for true D	Pr(- D)	94.59%
False + rate for classified +	Pr(~D +)	0.00%
False - rate for classified -	Pr(D -)	1.08%
Correctly classified		98.92%

```
. ** cálculo de logit 10 variables + incidencias **;
```

```
. logit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8_bn_amor_dt R9_margen_sobre_ventas__bn_i
> ncn R10_rotacion_activos_incn_at incidencias;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -175.40438
Iteration 2: log likelihood = -173.48133
Iteration 3: log likelihood = -169.37731
Iteration 4: log likelihood = -168.57415
Iteration 5: log likelihood = -168.35417
Iteration 6: log likelihood = -168.31342
Iteration 7: log likelihood = -168.31142
Iteration 8: log likelihood = -168.30531
```


Logistic model for fracaso

Classified	True		Total
	D	~D	
+	2	0	2
-	35	3193	3228
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso != 0

```
-----
Sensitivity          Pr(+|D)  5.41%
Specificity         Pr(-|~D) 100.00%
Positive predictive value Pr(D|+) 100.00%
Negative predictive value Pr(~D|-) 98.92%
-----
```

```
-----
False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D  Pr(-|D)  94.59%
False + rate for classified + Pr(~D|+)  0.00%
False - rate for classified - Pr(D|-)  1.08%
-----
```

```
-----
Correctly classified          98.92%
-----
```

. ** cálculo de logit 7 variables **;

```
. logit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn ;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -199.8168
Iteration 2: log likelihood = -198.98766
Iteration 3: log likelihood = -198.97215
Iteration 4: log likelihood = -198.9702
Iteration 5: log likelihood = -198.95971
Iteration 6: log likelihood = -198.95737
Iteration 7: log likelihood = -198.95737
```

```
Logistic regression              Number of obs =   3230
                                LR chi2(7)    =    6.39
                                Prob > chi2    =   0.4951
Log likelihood = -198.95737      Pseudo R2    =   0.0158
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | .3015789   .3106628   0.97  0.332   -1.3073091   .7046669
R3_bait_at | .8197321   .6994687   1.17  0.241   -1.5512014   2.190665
R4_bn_at | -.6466425   .7127533  -0.91  0.364   -2.043613   .7503282
-----
```

```

R5_ac_at | .8398984 .6917457 1.21 0.225 -.5158983 2.195695
R7_bnd_at | -.3950922 .3825173 -1.03 0.302 -1.144812 .3546279
R8_bn_amo~t | .0439332 .0191662 2.29 0.022 .0063682 .0814983
R9_margen_~n | .0003751 .0095228 0.04 0.969 -.0182893 .0190395
_cons | -5.316297 .6821919 -7.79 0.000 -6.653369 -3.979226

```

Note: 1 failure and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for logit of fracaso

```

Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -198.957
D(3222): 397.915  LR(7): 6.389
Prob > LR: 0.495
McFadden's R2: 0.016  McFadden's Adj R2: -0.024
Maximum Likelihood R2: 0.002  Cragg & Uhler's R2: 0.017
McKelvey and Zavoina's R2: 0.402  Efron's R2: 0.010
Variance of y*: 5.503  Variance of error: 3.290
Count R2: 0.989  Adj Count R2: 0.000
AIC: 0.128  AIC*n: 413.915
BIC: -25636.610  BIC': 50.172

```

```
. lroc, nograph;
```

Logistic model for fracaso

```

number of observations = 3230
area under ROC curve = 0.6737

```

```
. estat classification;
```

Logistic model for fracaso

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |    0    0 |    0
- |   37  3193 |  3230
-----+-----+-----
Total |   37  3193 |  3230

```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

```

-----
Sensitivity      Pr( +| D)  0.00%
Specificity      Pr( -|~D) 100.00%
Positive predictive value  Pr(D| +)  .%
Negative predictive value  Pr(~D| -) 98.85%
-----

```

```
False + rate for true ~D   Pr( +|~D)  0.00%
False - rate for true D    Pr( -| D) 100.00%
False + rate for classified + Pr(~D| +)  .%
False - rate for classified - Pr(D| -)  1.15%
```

```
-----
Correctly classified          98.85%
-----
```

```
. ** cálculo de logit 7 variables + incidencias **;
. logit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn incidencias;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -180.15585
Iteration 2: log likelihood = -177.78821
Iteration 3: log likelihood = -177.68216
Iteration 4: log likelihood = -177.66447
Iteration 5: log likelihood = -177.56819
Iteration 6: log likelihood = -177.56496
Iteration 7: log likelihood = -177.56496
```

```
Logistic regression                Number of obs =   3230
                                LR chi2(8)   =   49.17
                                Prob > chi2   =   0.0000
Log likelihood = -177.56496        Pseudo R2   =   0.1216
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | -.0155526   .4453491   -0.03   0.972   -.8884209   .8573157
R3_bait_at | .9027537   .822943    1.10   0.273   -.710185    2.515692
R4_bn_at | -1.070332   1.037286   -1.03   0.302   -3.103375   .9627106
R5_ac_at | .8153708   .7123975    1.14   0.252   -.5809026   2.211644
R7_bnd_at | -.2025565   .4371679   -0.46   0.643   -1.05939    .6542767
R8__bn_amo~t | .0494846   .0190085    2.60   0.009   .0122285    .0867406
R9_margen~n | .000794    .0128298    0.06   0.951   -.0243519    .0259399
incidencias | 1.675266   .2144062    7.81   0.000   1.255038    2.095495
   _cons | -5.522948   .6870765   -8.04   0.000   -6.869593   -4.176303
-----
```

Note: 1 failure and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for logit of fracaso

```
Log-Lik Intercept Only:  -202.152   Log-Lik Full Model:   -177.565
D(3221):                 355.130   LR(8):                 49.174
                          Prob > LR:                 0.000
McFadden's R2:           0.122   McFadden's Adj R2:    0.077
Maximum Likelihood R2:   0.015   Cragg & Uhler's R2:   0.128
```

```

McKelvey and Zavoina's R2: 0.878   Efron's R2:          0.034
Variance of y*:          26.918   Variance of error:    3.290
Count R2:                 0.989   Adj Count R2:        0.000
AIC:                      0.116   AIC*n:               373.130
BIC:                      -25671.315 BIC':                 15.468

```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```

number of observations = 3230
area under ROC curve = 0.7894

```

```
. estat classification;
```

```
Logistic model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

```

Classified + if predicted Pr(D) >= .5
True D defined as fracaso != 0

```

```

-----
Sensitivity      Pr( +| D)  0.00%
Specificity     Pr( -|~D) 100.00%
Positive predictive value Pr( D| +)  .%
Negative predictive value Pr(~D| -) 98.85%

```

```

-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D Pr( -| D) 100.00%
False + rate for classified + Pr(~D| +)  .%
False - rate for classified - Pr( D| -)  1.15%

```

```

-----
Correctly classified          98.85%
-----

```

```

. ** cálculo de logit 5 variables **;
. logit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn__incn;

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -200.32511
Iteration 2: log likelihood = -200.28314
Iteration 3: log likelihood = -200.27875
Iteration 4: log likelihood = -200.27527

```


+	0	0	0
-	37	3193	3230
-----+-----+-----			
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

Sensitivity	Pr(+ D)	0.00%	
Specificity	Pr(- ~D)	100.00%	
Positive predictive value	Pr(D +)	.%	
Negative predictive value	Pr(~D -)	98.85%	

False + rate for true ~D	Pr(+ ~D)	0.00%	
False - rate for true D	Pr(- D)	100.00%	
False + rate for classified +	Pr(~D +)	.%	
False - rate for classified -	Pr(D -)	1.15%	

Correctly classified		98.85%	

```

. ** cálculo de logit 5 variables + incidencias **;
. logit fracaso R1_dt_at R3_bait_at R4_bn_at R8_bn_amor_dt
R9_margen_sobre_ventas__bn_incn incidencias;

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -180.80534
Iteration 2: log likelihood = -178.91896
Iteration 3: log likelihood = -178.87653
Iteration 4: log likelihood = -178.86209
Iteration 5: log likelihood = -178.78253
Iteration 6: log likelihood = -178.77273
Iteration 7: log likelihood = -178.77273

```

```

Logistic regression           Number of obs =   3230
                             LR chi2(6)   =   46.76
                             Prob > chi2   =   0.0000
Log likelihood = -178.77273   Pseudo R2   =   0.1157

```

fracaso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
R1_dt_at	-.1589971	.3301795	-0.48	0.630	-1.8061371	.4881429
R3_bait_at	.3771998	1.096403	0.34	0.731	-1.771711	2.526111
R4_bn_at	-.6569544	1.289362	-0.51	0.610	-3.184058	1.87015
R8_bn_amo~t	.0466673	.0187865	2.48	0.013	.0098466	.0834881
R9_margen_~n	.0010369	.0138981	0.07	0.941	-.0262029	.0282767
incidencias	1.665661	.2120986	7.85	0.000	1.249955	2.081366
_cons	-4.811805	.3176264	-15.15	0.000	-5.434342	-4.189269

Note: 1 failure and 0 successes completely determined.

. fitstat;

Measures of Fit for logit of fracaso

```

Log-Lik Intercept Only:  -202.152  Log-Lik Full Model:  -178.773
D(3223):                 357.545  LR(6):           46.759
                          Prob > LR:             0.000
McFadden's R2:          0.116  McFadden's Adj R2:  0.081
Maximum Likelihood R2:  0.014  Cragg & Uhler's R2:  0.122
McKelvey and Zavoina's R2: 0.817  Efron's R2:         0.028
Variance of y*:         17.972  Variance of error:   3.290
Count R2:                0.989  Adj Count R2:       0.000
AIC:                     0.115  AIC*n:              371.545
BIC:                     -25685.060  BIC':                1.723

```

. lroc, nograph;

Logistic model for fracaso

```

number of observations = 3230
area under ROC curve = 0.7099

```

. estat classification;

Logistic model for fracaso

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

```

-----
Sensitivity          Pr(+|D)  0.00%
Specificity          Pr(-|~D) 100.00%
Positive predictive value Pr(D|+)  .%
Negative predictive value Pr(~D|-) 98.85%
-----

```

```

False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D   Pr(-|D) 100.00%
False + rate for classified + Pr(~D|+)  .%
False - rate for classified - Pr(D|-)  1.15%
-----

```

```

Correctly classified          98.85%

```

```

-----
. ** cálculo de logit 2 variables **;
. logit fracaso R6_gf_dt R8__bn_amor__dt ;

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -197.55943
Iteration 2: log likelihood = -194.57693
Iteration 3: log likelihood = -194.03114
Iteration 4: log likelihood = -193.96646
Iteration 5: log likelihood = -193.95322
Iteration 6: log likelihood = -193.95302
Iteration 7: log likelihood = -193.95302

Logistic regression              Number of obs =   3230
                                LR chi2(2)   =    16.40
                                Prob > chi2   =    0.0003
Log likelihood = -193.95302      Pseudo R2   =    0.0406

-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
R6_gf_dt | 1.439997  .4274183   3.37  0.001   .6022728  2.277722
R8__bn_amo~t | -.160844  .0471816  -3.41  0.001  -.2533182  -.0683699
  _cons | -4.551977  .1723896 -26.41  0.000  -4.889854  -4.2141

-----

. fitstat;

Measures of Fit for logit of fracaso

Log-Lik Intercept Only: -202.152  Log-Lik Full Model:   -193.953
D(3227):                387.906  LR(2):                16.398
                        Prob > LR:                0.000
McFadden's R2:          0.041  McFadden's Adj R2:    0.026
Maximum Likelihood R2:  0.005  Cragg & Uhler's R2:   0.043
McKelvey and Zavoina's R2: 0.048  Efron's R2:          0.042
Variance of y*:         3.455  Variance of error:    3.290
Count R2:               0.989  Adj Count R2:        0.054
AIC:                   0.122  AIC*n:               393.906
BIC:                   -25687.020  BIC':                -0.238

. lroc, nograph;

Logistic model for fracaso

number of observations =   3230
area under ROC curve   =  0.6921

. estat classification;

```

Logistic model for fracaso

Classified	True		Total
	D	~D	
+	2	0	2
-	35	3193	3228
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$
 True D defined as fracaso != 0

Sensitivity	$\Pr(+ D)$	5.41%
Specificity	$\Pr(- \sim D)$	100.00%
Positive predictive value	$\Pr(D +)$	100.00%
Negative predictive value	$\Pr(\sim D -)$	98.92%
False + rate for true ~D	$\Pr(+ \sim D)$	0.00%
False - rate for true D	$\Pr(- D)$	94.59%
False + rate for classified +	$\Pr(\sim D +)$	0.00%
False - rate for classified -	$\Pr(D -)$	1.08%
Correctly classified		98.92%

. ** cálculo de logit 2 variables + incidencias **;
 . logit fracaso R6_gf_dt R8__bn_amor__dt incidencias;

Iteration 0: log likelihood = -202.15209
 Iteration 1: log likelihood = -177.72083
 Iteration 2: log likelihood = -174.50683
 Iteration 3: log likelihood = -172.85844
 Iteration 4: log likelihood = -172.83496
 Iteration 5: log likelihood = -172.83496

Logistic regression Number of obs = 3230
 LR chi2(3) = 58.63
 Prob > chi2 = 0.0000
 Log likelihood = -172.83496 Pseudo R2 = 0.1450

fracaso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
R6_gf_dt	1.516211	.4139142	3.66	0.000	.704954 2.327468
R8__bn_amo~t	-.1709743	.0481516	-3.55	0.000	-.2653497 -.0765988
incidencias	1.641301	.2086892	7.86	0.000	1.232277 2.050324
_cons	-5.004783	.2165071	-23.12	0.000	-5.429129 -4.580436

```
. fitstat;
```

Measures of Fit for logit of fracaso

```
Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -172.835
D(3226): 345.670  LR(3): 58.634
Prob > LR: 0.000
McFadden's R2: 0.145  McFadden's Adj R2: 0.125
Maximum Likelihood R2: 0.018  Cragg & Uhler's R2: 0.153
McKelvey and Zavoina's R2: 0.129  Efron's R2: 0.059
Variance of y*: 3.775  Variance of error: 3.290
Count R2: 0.989  Adj Count R2: 0.054
AIC: 0.109  AIC*n: 353.670
BIC: -25721.176  BIC': -34.394
```

```
. lroc, nograph;
```

Logistic model for fracaso

```
number of observations = 3230
area under ROC curve = 0.8116
```

```
. estat classification;
```

Logistic model for fracaso

```
----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |    2    0 |    2
- |   35  3193 |  3228
-----+-----+-----
Total |   37  3193 |  3230
```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

```
-----
Sensitivity      Pr( +| D)  5.41%
Specificity      Pr( -|~D) 100.00%
Positive predictive value  Pr(D| +) 100.00%
Negative predictive value  Pr(~D| -) 98.92%
```

```
-----
False + rate for true ~D  Pr( +|~D)  0.00%
False - rate for true D   Pr( -| D)  94.59%
False + rate for classified +  Pr(~D| +)  0.00%
False - rate for classified -  Pr( D| -)  1.08%
```

```
-----
Correctly classified      98.92%
-----
```

```

. ** cálculo de probit 10 variables **;
. probit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8_bn_amo~t R9_margen~n R10_rotaci~t
> incn R10_rotacion_activos_incn_at;

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -194.1176
Iteration 2: log likelihood = -189.31382
Iteration 3: log likelihood = -189.07135
Iteration 4: log likelihood = -188.95215
Iteration 5: log likelihood = -188.9449
Iteration 6: log likelihood = -188.94487
Iteration 7: log likelihood = -188.94487

```

```

Probit regression                Number of obs =   3230
                                LR chi2(10) =   26.41
                                Prob > chi2 =   0.0032
Log likelihood = -188.94487      Pseudo R2   =   0.0653

```

```

-----+-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | .2488802   .1305756   1.91  0.057   -.0070433   .5048038
R2_ac_pc | -.0019931   .0023146  -0.86  0.389   -.0065297   .0025435
R3_bait_at | .3923341   .4061439   0.97  0.334   -.4036934   1.188362
R4_bn_at | -.2010334   .3472464  -0.58  0.563   -.8816239   .4795571
R5_ac_at | .6046023   .3062966   1.97  0.048   .004272    1.204933
R6_gf_dt | .9112581   .2424345   3.76  0.000   .4360953   1.386421
R7_bnd_at | -.2739757   .1409536  -1.94  0.052   -.5502397   .0022883
R8_bn_amo~t | -.1100053   .036175   -3.04  0.002   -.1809069   -.0391037
R9_margen~n | .0003077   .003644   0.08  0.933   -.0068345   .0074499
R10_rotaci~t | .0093681   .0185304   0.51  0.613   -.0269508   .045687
   _cons | -2.968421   .2719723  -10.91  0.000   -3.501477   -2.435365
-----+-----

```

Note: 5 failures and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for probit of fracaso

```

Log-Lik Intercept Only:  -202.152   Log-Lik Full Model:  -188.945
D(3219):                 377.890   LR(10):                26.414
                          Prob > LR:                0.003
McFadden's R2:           0.065     McFadden's Adj R2:     0.011
Maximum Likelihood R2:   0.008     Cragg & Uhler's R2:    0.069
McKelvey and Zavoina's R2: 0.148   Efron's R2:            0.040
Variance of y*:          1.174     Variance of error:     1.000
Count R2:                 0.989     Adj Count R2:          0.027
AIC:                      0.124     AIC*n:                 399.890

```

BIC: -25632.394 BIC': 54.388

. lroc, nograph;

Probit model for fracaso

number of observations = 3230
area under ROC curve = 0.7145

. estat classification;

Probit model for fracaso

		----- True -----		
Classified		D	~D	Total
+		1	0	1
-		36	3193	3229
Total		37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

Sensitivity	Pr(+ D)	2.70%
Specificity	Pr(- ~D)	100.00%
Positive predictive value	Pr(D +)	100.00%
Negative predictive value	Pr(~D -)	98.89%

False + rate for true ~D	Pr(+ ~D)	0.00%
False - rate for true D	Pr(- D)	97.30%
False + rate for classified +	Pr(~D +)	0.00%
False - rate for classified -	Pr(D -)	1.11%

Correctly classified	98.89%
----------------------	--------

. ** cálculo de probit 10 variables + incidencias **;

. probit fracaso R1_dt_at R2_ac_pc R3_bait_at R4_bn_at R5_ac_at R6_gf_dt
R7_bnd_at R8_bn_amor_dt R9_margen_sobre_ventas__bn_
> incn R10_rotacion_activos_incn_at incidencias;

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -180.80142
Iteration 2: log likelihood = -168.00235
Iteration 3: log likelihood = -167.67454
Iteration 4: log likelihood = -167.56409
Iteration 5: log likelihood = -167.53999
Iteration 6: log likelihood = -167.53523
Iteration 7: log likelihood = -167.53522

```

Probit regression                Number of obs =   3230
                                LR chi2(11)  =   69.23
                                Prob > chi2   =   0.0000
Log likelihood = -167.53522      Pseudo R2   =   0.1712

```

```

-----+-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | .1821236 .1607357   1.13  0.257  -.1329125 .4971598
R2_ac_pc | -.002142 .0023794  -0.90  0.368  -.0068055 .0025214
R3_bait_at | .2890436 .4710583   0.61  0.539  -.6342137  1.212301
R4_bn_at | -.1680575 .4228125  -0.40  0.691  -.9967547 .6606396
R5_ac_at | .543004 .3154745   1.72  0.085  -.0753147  1.161323
R6_gf_dt | .8742418 .2612928   3.35  0.001   .3621174  1.386366
R7_bnd_at | -.2202832 .1656295  -1.33  0.184  -.5449111 .1043446
R8_bn_amo~t | -.100933 .0395232  -2.55  0.011  -.1783969 -.023469
R9_margen~n | .000382 .0047723   0.08  0.936  -.0089715 .0097356
R10_rotaci~t | .0131032 .0195727   0.67  0.503  -.0252586 .0514651
incidencias | .7509252 .1051971   7.14  0.000   .5447428 .9571077
   _cons | -3.060335 .2910584 -10.51  0.000  -3.630799 -2.489871
-----+-----

```

Note: 6 failures and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for probit of fracaso

```

Log-Lik Intercept Only:  -202.152   Log-Lik Full Model:   -167.535
D(3218):                 335.070   LR(11):                 69.234
                          Prob > LR:                 0.000
McFadden's R2:           0.171   McFadden's Adj R2:       0.112
Maximum Likelihood R2:   0.021   Cragg & Uhler's R2:       0.180
McKelvey and Zavoina's R2: 0.294   Efron's R2:               0.070
Variance of y*:          1.416   Variance of error:         1.000
Count R2:                 0.989   Adj Count R2:             0.027
AIC:                      0.111   AIC*n:                    359.070
BIC:                      -25667.134   BIC':                      19.649

```

```
. lroc, nograph;
```

Probit model for fracaso

```

number of observations =  3230
area under ROC curve   =  0.8077

```

```
. estat classification;
```

Probit model for fracaso

		----- True -----		
Classified	D	~D	Total	
+	1	0	1	
-	36	3193	3229	
Total	37	3193	3230	

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso != 0

```
-----
Sensitivity          Pr( +| D)  2.70%
Specificity         Pr( -|~D) 100.00%
Positive predictive Pr( D| +) 100.00%
Negative predictive Pr(~D| -) 98.89%
```

```
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D  Pr( -| D)  97.30%
False + rate for classified + Pr(~D| +)  0.00%
False - rate for classified - Pr( D| -)  1.11%
```

```
-----
Correctly classified          98.89%
```

. ** cálculo de probit 7 variables **;

```
. probit fracaso R1_dt_at R3_bait_at R4_bn_at R5_ac_at R7_bnd_at
R8_bn_amor_dt R9_margen_sobre_ventas_bn_incn ;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -199.92068
Iteration 2: log likelihood = -199.39057
Iteration 3: log likelihood = -199.38549
Iteration 4: log likelihood = -199.3832
Iteration 5: log likelihood = -199.37079
Iteration 6: log likelihood = -199.37072
Iteration 7: log likelihood = -199.37072
```

```
Probit regression          Number of obs =   3230
                          LR chi2(7)   =    5.56
                          Prob > chi2   =   0.5916
Log likelihood = -199.37072      Pseudo R2   =   0.0138
```

```
-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
R1_dt_at | .0639282  .1166143   0.55  0.584  -.1646317  .292488
R3_bait_at | .2363798  .2708834   0.87  0.383  -.294542  .7673017
R4_bn_at | -.242313  .3061378  -0.79  0.429  -.842332  .357706
R5_ac_at | .2674108  .2461921   1.09  0.277  -.2151169  .7499386
R7_bnd_at | -.0992933  .13903   -0.71  0.475  -.3717872  .1732005
```

```

R8__bn_amo~t | .0135516 .0101612 1.33 0.182 -.006364 .0334671
R9__margen_~n | .0002358 .0036823 0.06 0.949 -.0069813 .0074529
      _cons | -2.524606 .2257609 -11.18 0.000 -2.967089 -2.082122

```

Note: 1 failure and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for probit of fracaso

```

Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -199.371
D(3222):                398.741  LR(7):           5.563
                        Prob > LR:           0.592
McFadden's R2:          0.014  McFadden's Adj R2: -0.026
Maximum Likelihood R2:  0.002  Cragg & Uhler's R2:  0.015
McKelvey and Zavoina's R2: 0.233  Efron's R2:          0.001
Variance of y*:         1.304  Variance of error:    1.000
Count R2:               0.989  Adj Count R2:        0.000
AIC:                   0.128  AIC*n:              414.741
BIC:                   -25635.784  BIC':              50.999

```

```
. lroc, nograph;
```

Probit model for fracaso

```

number of observations = 3230
area under ROC curve = 0.6835

```

```
. estat classification;
```

Probit model for fracaso

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

```

Classified + if predicted Pr(D) >= .5
True D defined as fracaso != 0

```

```

-----
Sensitivity          Pr(+|D)  0.00%
Specificity          Pr(~|~D) 100.00%
Positive predictive value Pr(D|+)  .%
Negative predictive value Pr(~D|~) 98.85%
-----
False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D  Pr(~|D) 100.00%

```



```

Count R2:          0.989   Adj Count R2:          0.000
AIC:              0.115   AIC*n:         371.698
BIC:             -25672.746   BIC':         14.036

```

```
. lroc, nograph;
```

```
Probit model for fracaso
```

```

number of observations = 3230
area under ROC curve = 0.7840

```

```
. estat classification;
```

```
Probit model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```

-----
Sensitivity          Pr(+|D)  0.00%
Specificity          Pr(-|~D) 100.00%
Positive predictive value Pr(D|+)  .%
Negative predictive value Pr(~D|-) 98.85%
-----

```

```

False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D Pr(-|D) 100.00%
False + rate for classified + Pr(~D|+)  .%
False - rate for classified - Pr(D|-)  1.15%
-----

```

```
Correctly classified          98.85%
```

```
. ** cálculo de probit 5 variables **;
```

```
. probit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor__dt
R9_margen_sobre_ventas__bn_incn;
```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -200.58802
Iteration 2: log likelihood = -200.38068
Iteration 3: log likelihood = -200.37737
Iteration 4: log likelihood = -200.3721
Iteration 5: log likelihood = -200.36466
Iteration 6: log likelihood = -200.36463

```



```
-----+-----+-----
Total |    37    3193 |    3230
```

Classified + if predicted Pr(D) >= .5
True D defined as fracaso != 0

```
-----
Sensitivity          Pr( +| D)  0.00%
Specificity         Pr( -|~D) 100.00%
Positive predictive value Pr( D|+)  .%
Negative predictive value Pr(~D|-) 98.85%
```

```
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D  Pr( -| D) 100.00%
False + rate for classified + Pr(~D|+)  .%
False - rate for classified - Pr(D|-)  1.15%
```

```
-----
Correctly classified          98.85%
-----
```

```
. ** cálculo de probit 5 variables + incidencias **;
. probit fracaso R1_dt_at R3_bait_at R4_bn_at R8__bn_amor_dt
R9_margen_sobre_ventas__bn_incn incidencias;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -191.00577
Iteration 2: log likelihood = -177.85491
Iteration 3: log likelihood = -177.81946
Iteration 4: log likelihood = -177.7408
Iteration 5: log likelihood = -177.59522
Iteration 6: log likelihood = -177.59458
Iteration 7: log likelihood = -177.59458
```

```
Probit regression          Number of obs =    3230
                          LR chi2(6)   =    49.12
                          Prob > chi2   =    0.0000
Log likelihood = -177.59458      Pseudo R2    =    0.1215
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R1_dt_at | -0.0799617 .1279047   -0.63  0.532   -0.3306503   .1707268
R3_bait_at | .1189248 .2881619    0.41  0.680   -0.4458621   .6837117
R4_bn_at | -0.264715 .407261   -0.65  0.516   -1.062932   .5335019
R8__bn_amo~t | .0166146 .0101171    1.64  0.101   -0.0032145   .0364436
R9_margen~n | .000416 .005103    0.08  0.935   -0.0095857   .0104177
incidencias | .7631368 .1037015    7.36  0.000   .5598856   .966388
   _cons | -2.40722 .1175494  -20.48  0.000   -2.637613   -2.176828
-----
```

Note: 1 failure and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for probit of fracaso

```
Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -177.595
D(3223): 355.189  LR(6): 49.115
Prob > LR: 0.000
McFadden's R2: 0.121  McFadden's Adj R2: 0.087
Maximum Likelihood R2: 0.015  Cragg & Uhler's R2: 0.128
McKelvey and Zavoina's R2: 0.805  Efron's R2: 0.023
Variance of y*: 5.123  Variance of error: 1.000
Count R2: 0.989  Adj Count R2: 0.000
AIC: 0.114  AIC*n: 369.189
BIC: -25687.416  BIC': -0.634
```

```
. lroc, nograph;
```

Probit model for fracaso

```
number of observations = 3230
area under ROC curve = 0.7118
```

```
. estat classification;
```

Probit model for fracaso

```
----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230
```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

```
-----
Sensitivity      Pr( +| D)  0.00%
Specificity      Pr( -|~D) 100.00%
Positive predictive value  Pr( D| +)  .%
Negative predictive value  Pr(~D| -) 98.85%
```

```
-----
False + rate for true ~D  Pr( +|~D)  0.00%
False - rate for true D   Pr( -| D) 100.00%
False + rate for classified +  Pr(~D| +)  .%
False - rate for classified -  Pr( D| -)  1.15%
```

```
-----
Correctly classified      98.85%
-----
```

```
. ** cálculo de probit 2 variables **;
. probit fracaso R6_gf_dt R8_bn_amo~t ;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -194.05564
Iteration 2: log likelihood = -193.96661
Iteration 3: log likelihood = -193.9666
```

```
Probit regression                Number of obs =   3230
                                LR chi2(2)   =   16.37
                                Prob > chi2   =   0.0003
Log likelihood = -193.9666      Pseudo R2    =   0.0405
```

```
-----+-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R6_gf_dt | .7282704   .2230663   3.26  0.001   .2910686   1.165472
R8_bn_amo~t | -.0796621   .0283585  -2.81  0.005  -.1352438  -.0240805
   _cons | -2.314808   .0647997  -35.72  0.000  -2.441813  -2.187803
-----+-----
```

Note: 3 failures and 0 successes completely determined.

```
. fitstat;
```

Measures of Fit for probit of fracaso

```
Log-Lik Intercept Only:  -202.152   Log-Lik Full Model:  -193.967
D(3227):                 387.933   LR(2):                 16.371
                          Prob > LR:                 0.000
McFadden's R2:           0.040   McFadden's Adj R2:     0.026
Maximum Likelihood R2:   0.005   Cragg & Uhler's R2:    0.043
McKelvey and Zavoina's R2: 0.040   Efron's R2:           0.040
Variance of y*:          1.041   Variance of error:     1.000
Count R2:                 0.989   Adj Count R2:          0.054
AIC:                       0.122   AIC*n:                 393.933
BIC:                       -25686.993   BIC':                  -0.211
```

```
. estat classification;
```

Probit model for fracaso

```
----- True -----
Classified |   D    ~D | Total
-----+-----
+ |     2     0 |     2
- |    35   3193 |   3228
-----+-----
Total |    37   3193 |   3230
```

Classified + if predicted Pr(D) >= .5

True D defined as fracaso != 0

```
-----
Sensitivity          Pr( +| D)  5.41%
Specificity         Pr( -|~D) 100.00%
Positive predictive value Pr( D| +) 100.00%
Negative predictive value Pr(~D| -) 98.92%
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D Pr( -| D)  94.59%
False + rate for classified + Pr(~D| +)  0.00%
False - rate for classified - Pr( D| -)  1.08%
-----
Correctly classified          98.92%
-----
```

. lroc, nograph;

Probit model for fracaso

number of observations = 3230
area under ROC curve = 0.6909

. ** cálculo de probit 2 variables + incidencias **;
. probit fracaso R6_gf_dt R8__bn_ amor__dt incidencias;

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -183.48055
Iteration 2: log likelihood = -171.62175
Iteration 3: log likelihood = -171.57774
Iteration 4: log likelihood = -171.57759
Iteration 5: log likelihood = -171.57759

```
Probit regression          Number of obs = 3230
                          LR chi2(3) = 61.15
                          Prob > chi2 = 0.0000
Log likelihood = -171.57759      Pseudo R2 = 0.1512
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
R6_gf_dt | .7402816   .2210979   3.35  0.001   .3069377   1.173626
R8__bn_amo~t | -.0802974   .0287885  -2.79  0.005  -.1367219  -.0238729
incidencias | .7498254   .1022142   7.34  0.000   .5494892   .9501615
   _cons | -2.500666   .0800373  -31.24  0.000  -2.657536  -2.343795
-----
```

Note: 3 failures and 0 successes completely determined.

. fitstat;

Measures of Fit for probit of fracaso

```

Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -171.578
D(3226): 343.155  LR(3): 61.149
          Prob > LR: 0.000
McFadden's R2: 0.151  McFadden's Adj R2: 0.131
Maximum Likelihood R2: 0.019  Cragg & Uhler's R2: 0.159
McKelvey and Zavoina's R2: 0.095  Efron's R2: 0.058
Variance of y*: 1.105  Variance of error: 1.000
Count R2: 0.989  Adj Count R2: 0.027
AIC: 0.109  AIC*n: 351.155
BIC: -25723.691  BIC': -36.908

```

```
. estat classification;
```

```
Probit model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   1   0 |   1
- |  36 3193 | 3229
-----+-----+-----
Total |  37 3193 | 3230

```

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```

-----
Sensitivity          Pr(+|D)  2.70%
Specificity          Pr(-|~D) 100.00%
Positive predictive value Pr(D|+) 100.00%
Negative predictive value Pr(~D|-) 98.89%
-----
False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D  Pr(-|D)  97.30%
False + rate for classified + Pr(~D|+)  0.00%
False - rate for classified - Pr(D|-)  1.11%
-----
Correctly classified          98.89%
-----

```

```
. lroc, nograph;
```

```
Probit model for fracaso
```

```
number of observations = 3230
```

```
area under ROC curve = 0.8105
```

```
. ** DATOS WINSORIZADOS **;
```

```
. ** Diferencia de medias por variable separado por fracaso =1 y fracaso=0, nivel de significación del 99%**;
```

. ttest wR1, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
0 | 3193  .7876563 .0079476   .4490908   .7671724   .8081401
1 | 37    1.019683 .1059086   .6442166   .7316662   1.3077
-----+-----
combined | 3230  .7903142 .007959    .4523366   .7698009   .8108274
-----+-----
diff |      -.2320266 .0746933      -.4245377  -.0395155
-----+-----
diff = mean(0) - mean(1)                t = -3.1064
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.0010    Pr(|T| > |t|) = 0.0019    Pr(T > t) = 0.9990
```

. ttest wR2, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
0 | 3193  2.783909 .1020998   5.769318   2.52076    3.047058
1 | 37    2.527121 1.141043   6.940696  -.5759289  5.630171
-----+-----
combined | 3230  2.780967 .1017516   5.782859   2.518717   3.043217
-----+-----
diff |      .2567876 .956326      -2.208002  2.721578
-----+-----
diff = mean(0) - mean(1)                t = 0.2685
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.6058    Pr(|T| > |t|) = 0.7883    Pr(T > t) = 0.3942
```

. ttest wR3, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
0 | 3193  .022242 .0038997   .220357    .0121912   .0322929
1 | 37   -.1692527 .0727478   .4425078   -.3670893   .028584
-----+-----
combined | 3230  .0200485 .003958    .2249428   .0098474   .0302495
-----+-----
diff |      .1914947 .0370468      .0960121   .2869773
```

```
-----
diff = mean(0) - mean(1)                t = 5.1690
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 1.0000    Pr(|T| > |t|) = 0.0000    Pr(T > t) = 0.0000
```

```
. ttest wR4, by(fracaso) level(99);
```

Two-sample t test with equal variances

```
-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  -.0056056  .0033833   .1911801   -.0143257   .0031144
1 | 37    -.1896237  .067625   .4113469   -.3735289   -.0057185
-----+-----
combined | 3230  -.0077136  .003448   .1959619   -.0166004   .0011732
-----+-----
diff |      .1840181  .0322448           .1009118   .2671244
```

```
-----
diff = mean(0) - mean(1)                t = 5.7069
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 1.0000    Pr(|T| > |t|) = 0.0000    Pr(T > t) = 0.0000
```

```
. ttest wR5, by(fracaso) level(99);
```

Two-sample t test with equal variances

```
-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  .7263177  .0042001   .2373353   .7154924   .737143
1 | 37    .7744578  .0448275   .2726751   .6525501   .8963655
-----+-----
combined | 3230  .7268692  .0041838   .2377767   .7160861   .7376522
-----+-----
diff |      -.0481401  .039313           -.1494637   .0531836
```

```
-----
diff = mean(0) - mean(1)                t = -1.2245
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0      Ha: diff > 0
Pr(T < t) = 0.1104    Pr(|T| > |t|) = 0.2208    Pr(T > t) = 0.8896
```

```
. ttest wR6, by(fracaso) level(99);
```

Two-sample t test with equal variances

```
-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
```

```

-----+-----
      0 | 3193 .0297533 .0005676 .032074 .0282903 .0312162
      1 | 37 .0355791 .0054415 .0330995 .020781 .0503773
-----+-----
combined | 3230 .02982 .0005646 .0320866 .0283649 .0312751
-----+-----
diff | -.0058258 .0053053 -.0194995 .0078478
-----+-----
diff = mean(0) - mean(1)          t = -1.0981
Ho: diff = 0                      degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.1361    Pr(|T| > |t|) = 0.2722    Pr(T > t) = 0.8639

```

. ttest wR7, by(fracaso) level(99);

Two-sample t test with equal variances

```

-----+-----
Group | Obs   Mean  Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
      0 | 3193 .1956719 .0067497 .3814027 .1782754 .2130684
      1 | 37 .1500051 .0803055 .4884795 -.0683846 .3683947
-----+-----
combined | 3230 .1951488 .0067343 .3827337 .177792 .2125056
-----+-----
diff | .0456668 .0632892 -.1174519 .2087855
-----+-----
diff = mean(0) - mean(1)          t = 0.7216
Ho: diff = 0                      degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.7647    Pr(|T| > |t|) = 0.4706    Pr(T > t) = 0.2353

```

. ttest wR8, by(fracaso) level(99);

Two-sample t test with equal variances

```

-----+-----
Group | Obs   Mean  Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
      0 | 3193 .1211922 .0059977 .3389109 .1057338 .1366505
      1 | 37 .0164868 .0945741 .5752716 -.2407059 .2736796
-----+-----
combined | 3230 .1199928 .0060278 .3425759 .1044571 .1355284
-----+-----
diff | .1047053 .0566233 -.0412329 .2506436
-----+-----
diff = mean(0) - mean(1)          t = 1.8492
Ho: diff = 0                      degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0

```

Pr(T < t) = 0.9677 Pr(|T| > |t|) = 0.0645 Pr(T > t) = 0.0323

. ttest wR9, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  -.0638995  .0084117   .475318   -.0855796   -.0422194
1 | 37    -.2360054  .0923896   .5619837  -.4872573   .0152466
-----+-----
combined | 3230  -.065871   .0083868   .4766498  -.0874868  -.0442552
-----+-----
diff |      .1721059  .0787674           -.0309056   .3751174
-----+-----
diff = mean(0) - mean(1)                t = 2.1850
Ho: diff = 0                            degrees of freedom = 3228
```

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.9855 Pr(|T| > |t|) = 0.0290 Pr(T > t) = 0.0145

. ttest wR10, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  1.661431  .0256287   1.448189   1.595376   1.727485
1 | 37    1.919005  .3114892   1.894715   1.071915   2.766095
-----+-----
combined | 3230  1.664381  .025583    1.453959   1.598445   1.730318
-----+-----
diff |      -.257574  .2404047           -.8771819   .3620339
-----+-----
diff = mean(0) - mean(1)                t = -1.0714
Ho: diff = 0                            degrees of freedom = 3228
```

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.1420 Pr(|T| > |t|) = 0.2841 Pr(T > t) = 0.8580

. ttest incidencias, by(fracaso) level(99);

Two-sample t test with equal variances

```
-----+-----
Group | Obs   Mean   Std. Err.   Std. Dev.   [99% Conf. Interval]
-----+-----
0 | 3193  .0720326  .0055455   .3133578   .0577397   .0863254
1 | 37    .6216216  .118495    .720777    .2993763   .943867
-----+-----
combined | 3230  .0783282  .0057362   .3260084   .0635439   .0931125
```

```

-----+-----
diff |      -.5495891  .0530385      -.6862881  -.41289
-----+-----
diff = mean(0) - mean(1)                t = -10.3621
Ho: diff = 0                            degrees of freedom = 3228

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.0000    Pr(|T| > |t|) = 0.0000    Pr(T > t) = 1.0000

```

```
. ttest num_empl_total, by(fracaso) level(99);
```

Two-sample t test with equal variances

```

-----+-----
Group |  Obs   Mean  Std. Err.  Std. Dev.  [99% Conf. Interval]
-----+-----
0 |  1955  32.46496  13.30112  588.1141  -1.829951  66.75987
1 |   16  12.1875  2.88851  11.55404  3.675891  20.69911
-----+-----
combined |  1971  32.30036  13.1932  585.7246  -1.716035  66.31675
-----+-----
diff |      20.27746  147.0658      -358.9064  399.4613
-----+-----
diff = mean(0) - mean(1)                t = 0.1379
Ho: diff = 0                            degrees of freedom = 1969

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
Pr(T < t) = 0.5548    Pr(|T| > |t|) = 0.8903    Pr(T > t) = 0.4452

```

```
. ** para el cálculo de la mediana, la desviación típica y la asimetría **;
. tabstat wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias
num_empl_total, statistics( median sd skewness ) by(fracaso)
> nototal columns(statistics);
```

Summary for variables: wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10
incidencias num_empl_total
by categories of: fracaso

```

fracaso |  p50    sd skewness
-----+-----
0 | .7783218 .4490908 2.483903
  | 1.274884 5.769318 5.38778
  | .0444596 .220357 -2.688866
  | .0159537 .1911801 -3.040909
  | .7864155 .2373353 -.9497068
  | .0214282 .032074 1.978609
  | .1062388 .3814027 2.493171
  | .0608538 .3389109 2.552918
  | .0124977 .475318 -6.155788
  | 1.381083 1.448189 2.268183
  | 0 .3133578 4.695205

```

```

|      6 588.1141 36.44888
-----+-----
1 | .9751387 .6442166 1.777834
| 1.057278 6.940696 5.563727
| -.0164002 .4425078 -1.603176
| -.0354618 .4113469 -1.710292
| .8938961 .2726751 -1.451836
| .032181 .0330995 2.367284
| .0299024 .4884795 2.808293
| -.0286602 .5752716 2.43246
| -.0441338 .5619837 -3.127511
| 1.475444 1.894715 1.989572
|      0 .720777 .694475
|      9 11.55404 1.095393
-----

```

```

. ** para el cálculo del test de Wilcoxon**;
. ranksum wR1, by(fracaso);

```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

```

fracaso | obs rank sum expected
-----+-----
0 | 3193 5138264 5158291.5
1 | 37 79801 59773.5
-----+-----
combined | 3230 5218065 5218065

```

```

unadjusted variance 31809464
adjustment for ties -67.7888

```

```

-----
adjusted variance 31809396

```

```

Ho: wR1(fracaso==0) = wR1(fracaso==1)
z = -3.551
Prob > |z| = 0.0004

```

```

. ranksum wR2, by(fracaso);

```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

```

fracaso | obs rank sum expected
-----+-----
0 | 3193 5167957 5158291.5
1 | 37 50108 59773.5
-----+-----
combined | 3230 5218065 5218065

```

```

unadjusted variance 31809464
adjustment for ties -67.783136

```

adjusted variance 31809396

Ho: wR2(fracaso==0) = wR2(fracaso==1)

z = 1.714

Prob > |z| = 0.0866

. ranksum wR3, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
-----+-----			
0	3193	5175742	5158291.5
1	37	42323	59773.5
-----+-----			
combined	3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -67.783136

adjusted variance 31809396

Ho: wR3(fracaso==0) = wR3(fracaso==1)

z = 3.094

Prob > |z| = 0.0020

. ranksum wR4, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
-----+-----			
0	3193	5181516	5158291.5
1	37	36549	59773.5
-----+-----			
combined	3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -67.981366

adjusted variance 31809396

Ho: wR4(fracaso==0) = wR4(fracaso==1)

z = 4.118

Prob > |z| = 0.0000

. ranksum wR5, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5147145	5158291.5
1	37	70920	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -1256.2308

adjusted variance 31808208

Ho: $wR5(\text{fracaso}=0) = wR5(\text{fracaso}=1)$
 $z = -1.976$
 Prob > |z| = 0.0481

. ranksum wR6, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5149058	5158291.5
1	37	69007	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -24514.32

adjusted variance 31784950

Ho: $wR6(\text{fracaso}=0) = wR6(\text{fracaso}=1)$
 $z = -1.638$
 Prob > |z| = 0.1015

. ranksum wR7, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5171094.5	5158291.5
1	37	46970.5	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -111.79573

adjusted variance 31809352

Ho: wR7(fracaso==0) = wR7(fracaso==1)
 $z = 2.270$
 Prob > |z| = 0.0232

. ranksum wR8, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
-----+-----			
0	3193	5183526	5158291.5
1	37	34539	59773.5
-----+-----			
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -67.7888

adjusted variance 31809396

Ho: wR8(fracaso==0) = wR8(fracaso==1)
 $z = 4.474$
 Prob > |z| = 0.0000

. ranksum wR9, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
-----+-----			
0	3193	5182556	5158291.5
1	37	35509	59773.5
-----+-----			
combined	3230	5218065	5218065

unadjusted variance 31809464
 adjustment for ties -67.981366

adjusted variance 31809396

Ho: wR9(fracaso==0) = wR9(fracaso==1)
 $z = 4.302$
 Prob > |z| = 0.0000

. ranksum wR10, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5155842	5158291.5
1	37	62223	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -67.783136

adjusted variance 31809396

Ho: $wR10(\text{fracaso}=0) = wR10(\text{fracaso}=1)$

$z = -0.434$

Prob > |z| = 0.6641

. ranksum incidencias, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	3193	5132884.5	5158291.5
1	37	85180.5	59773.5
combined	3230	5218065	5218065

unadjusted variance 31809464

adjustment for ties -26313858

adjusted variance 5495606.6

Ho: $\text{incide}\sim\text{s}(\text{fracaso}=0) = \text{incide}\sim\text{s}(\text{fracaso}=1)$

$z = -10.838$

Prob > |z| = 0.0000

. ranksum num_empl_total, by(fracaso);

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

fracaso	obs	rank sum	expected
0	1955	1926009.5	1927630
1	16	17396.5	15776
combined	1971	1943406	1943406

unadjusted variance 5140346.67

adjustment for ties -24183.74

adjusted variance 5116162.93

Ho: num_em~l(fracaso==0) = num_em~l(fracaso==1)
 z = -0.716
 Prob > |z| = 0.4737

. ** cálculo LDA 10 variables **;
 . discrim lda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10, group(fracaso);

Linear discriminant analysis
 Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,776  417 | 3,193
      | 86.94  13.06 | 100.00
      |           |
      1 | 19    18 | 37
      | 51.35  48.65 | 100.00
-----+-----+-----
      Total | 2,795  435 | 3,230
      | 86.53  13.47 | 100.00
      |           |
      Priors | 0.5000  0.5000 |
  
```

. ** cálculo LDA 10 variables + incidencias**;
 . discrim lda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias,
 group(fracaso);

Linear discriminant analysis
 Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,939  254 | 3,193
  
```

```

      | 92.05  7.95 | 100.00
      |         |
      1 | 15    22 | 37
      | 40.54 59.46 | 100.00
-----+-----+-----
      Total | 2,954  276 | 3,230
      | 91.46  8.54 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo LDA 7 variables **;
. discrim lda wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,798  395 | 3,193
      | 87.63 12.37 | 100.00
      |         |
      1 | 20    17 | 37
      | 54.05 45.95 | 100.00
-----+-----+-----
      Total | 2,818  412 | 3,230
      | 87.24 12.76 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo LDA 7variables + incidencias**;
. discrim lda wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total

```

```

-----+-----+-----
      0 | 2,935  258 | 3,193
        | 91.92  8.08 | 100.00
        |         |
      1 |   15   22 |   37
        | 40.54 59.46 | 100.00
-----+-----+-----
    Total | 2,950  280 | 3,230
          | 91.33  8.67 | 100.00
          |         |
    Priors | 0.5000 0.5000 |

```

```

. ** cálculo LDA 5 variables **;
. discrim lda wR1 wR3 wR4 wR8 wR9, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0  1 | Total
-----+-----+-----
      0 | 2,836  357 | 3,193
        | 88.82 11.18 | 100.00
        |         |
      1 |   24   13 |   37
        | 64.86 35.14 | 100.00
-----+-----+-----
    Total | 2,860  370 | 3,230
          | 88.54 11.46 | 100.00
          |         |
    Priors | 0.5000 0.5000 |

```

```

. ** cálculo LDA 5variables + incidencias**;
. discrim lda wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);

```

Linear discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+

```

True fracaso	Classified		Total
	0	1	
0	2,934	259	3,193
	91.89	8.11	100.00
1	15	22	37
	40.54	59.46	100.00
Total	2,949	281	3,230
	91.30	8.70	100.00
Priors	0.5000	0.5000	

```
. ** cálculo LDA 3 variables **;
. discrim lda wR1 wR3 wR4, group(fracaso);
```

Linear discriminant analysis
Resubstitution classification summary

True fracaso	Classified		Total
	0	1	
0	2,822	371	3,193
	88.38	11.62	100.00
1	24	13	37
	64.86	35.14	100.00
Total	2,846	384	3,230
	88.11	11.89	100.00
Priors	0.5000	0.5000	

```
. ** cálculo LDA 3 variables + incidencias**;
. discrim lda wR1 wR3 wR4 incidencias, group(fracaso);
```

Linear discriminant analysis
Resubstitution classification summary

True fracaso	Classified		Total
	0	1	
0	2,822	371	3,193
	88.38	11.62	100.00
1	24	13	37
	64.86	35.14	100.00
Total	2,846	384	3,230
	88.11	11.89	100.00
Priors	0.5000	0.5000	

```

| Percent |
+-----+
| Classified
True fracaso | 0    1 | Total
-----+-----+-----+
0 | 2,944  249 | 3,193
  | 92.20  7.80 | 100.00
  |         |
1 | 16    21 | 37
  | 43.24 56.76 | 100.00
-----+-----+-----+
Total | 2,960  270 | 3,230
      | 91.64  8.36 | 100.00
      |         |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo QDA 10 variables **;
. discrim qda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10, group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
| Classified
True fracaso | 0    1 Unclassified | Total
-----+-----+-----+
0 | 2,947  245      1 | 3,193
  | 92.30  7.67    0.03 | 100.00
  |         |
1 | 21    16      0 | 37
  | 56.76 43.24    0.00 | 100.00
-----+-----+-----+
Total | 2,968  261      1 | 3,230
      | 91.89  8.08    0.03 | 100.00
      |         |
Priors | 0.5000 0.5000      |

```

```

. ** cálculo QDA 10 variables + incidencias**;
. discrim qda wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias,
group(fracaso);

```

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+

```

Key		Number		Percent	
+-----+					
Classified					
True fracaso	0	1	Unclassified	Total	
+-----+					
0	2,916	276	1	3,193	
	91.32	8.64	0.03	100.00	
1	12	25	0	37	
	32.43	67.57	0.00	100.00	
+-----+					
Total	2,928	301	1	3,230	
	90.65	9.32	0.03	100.00	
	Priors 0.5000 0.5000				

. ** cálculo QDA 7 variables **;
. discrim qda wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);

Quadratic discriminant analysis
Resubstitution classification summary

Key		Number		Percent	
+-----+					
Classified					
True fracaso	0	1	Total		
+-----+					
0	2,905	288	3,193		
	90.98	9.02	100.00		
1	26	11	37		
	70.27	29.73	100.00		
+-----+					
Total	2,931	299	3,230		
	90.74	9.26	100.00		
	Priors 0.5000 0.5000				

. ** cálculo QDA 7variables + incidencias**;
. discrim qda wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, group(fracaso);

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,829  364 | 3,193
        | 88.60  11.40 | 100.00
        |         |
      1 | 12    25 | 37
        | 32.43  67.57 | 100.00
-----+-----+-----
      Total | 2,841  389 | 3,230
            | 87.96  12.04 | 100.00
            |         |
      Priors | 0.5000  0.5000 |

```

. ** cálculo QDA 5 variables **;
. discrim qda wR1 wR3 wR4 wR8 wR9, group(fracaso);

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,847  346 | 3,193
        | 89.16  10.84 | 100.00
        |         |
      1 | 26    11 | 37
        | 70.27  29.73 | 100.00
-----+-----+-----
      Total | 2,873  357 | 3,230
            | 88.95  11.05 | 100.00
            |         |
      Priors | 0.5000  0.5000 |

```

. ** cálculo QDA 5variables + incidencias**;
. discrim qda wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 2,778  415 | 3,193
          | 87.00  13.00 | 100.00
          |         |
          1 | 13    24 | 37
          | 35.14  64.86 | 100.00
-----+-----+-----
          Total | 2,791  439 | 3,230
          | 86.41  13.59 | 100.00
          |         |
          Priors | 0.5000  0.5000 |

```

. ** cálculo QDA 3 variables **;
. discrim qda wR1 wR3 wR4, group(fracaso);

Quadratic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 2,944  249 | 3,193
          | 92.20  7.80 | 100.00
          |         |
          1 | 30    7 | 37
          | 81.08  18.92 | 100.00
-----+-----+-----
          Total | 2,974  256 | 3,230
          | 92.07  7.93 | 100.00
          |         |
          Priors | 0.5000  0.5000 |

```

. ** cálculo QDA 3 variables + incidencias**;

```
. discrim qda wR1 wR3 wR4 incidencias, group(fracaso);
```

Quadratic discriminant analysis
Resubstitution classification summary

```
+-----+
| Key   |
|-----|
| Number|
| Percent|
|-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,873  320 | 3,193
      | 89.98  10.02 | 100.00
      |         |
      1 | 15     22 | 37
      | 40.54  59.46 | 100.00
-----+-----+-----
      Total | 2,888  342 | 3,230
      | 89.41  10.59 | 100.00
      |         |
      Priors | 0.5000 0.5000 |
```

```
. ** cálculo LogDA 10 variables **;
. discrim logistic wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10,
group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -190.09894
Iteration 2: log likelihood = -186.40216
Iteration 3: log likelihood = -186.03038
Iteration 4: log likelihood = -186.02236
Iteration 5: log likelihood = -186.02235
```

Logistic discriminant analysis
Resubstitution classification summary

```
+-----+
| Key   |
|-----|
| Number|
| Percent|
|-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,285  908 | 3,193
      | 71.56  28.44 | 100.00
```

	11	26	37
1	29.73	70.27	100.00
-----+-----+-----			
Total	2,296	934	3,230
	71.08	28.92	100.00
-----+-----+-----			
Priors	0.5000	0.5000	

```
. ** cálculo LogDA 10 variables + incidencias**;  
. discrim logistic wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias,  
group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209  
Iteration 1: log likelihood = -176.50195  
Iteration 2: log likelihood = -171.32898  
Iteration 3: log likelihood = -170.92844  
Iteration 4: log likelihood = -170.92298  
Iteration 5: log likelihood = -170.92298
```

Logistic discriminant analysis
Resubstitution classification summary

+-----+			
Key			

Number			
Percent			
+-----+			
	Classified		
True fracaso	0	1	Total
-----+-----+-----			
0	2,716	477	3,193
	85.06	14.94	100.00
-----+-----+-----			
1	11	26	37
	29.73	70.27	100.00
-----+-----+-----			
Total	2,727	503	3,230
	84.43	15.57	100.00
-----+-----+-----			
Priors	0.5000	0.5000	

```
. ** cálculo LogDA 7 variables **;  
. discrim logistic wR1 wR3 wR4 wR5 wR7 wR8 wR9, group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209  
Iteration 1: log likelihood = -190.39495  
Iteration 2: log likelihood = -186.99974  
Iteration 3: log likelihood = -186.6895
```

Iteration 4: log likelihood = -186.68406
 Iteration 5: log likelihood = -186.68405

Logistic discriminant analysis
 Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,277  916 | 3,193
        | 71.31  28.69 | 100.00
        |         |
      1 | 13    24 | 37
        | 35.14  64.86 | 100.00
-----+-----+-----
      Total | 2,290  940 | 3,230
            | 70.90  29.10 | 100.00
            |         |
      Priors | 0.5000  0.5000 |
  
```

. ** cálculo LogDA 7variables + incidencias**;
 . discrim logistic wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, group(fracaso);

Iteration 0: log likelihood = -202.15209
 Iteration 1: log likelihood = -176.61768
 Iteration 2: log likelihood = -171.59725
 Iteration 3: log likelihood = -171.24047
 Iteration 4: log likelihood = -171.23608
 Iteration 5: log likelihood = -171.23608

Logistic discriminant analysis
 Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,735  458 | 3,193
        | 85.66  14.34 | 100.00
  
```

	11	26	37
	29.73	70.27	100.00
-----+-----+-----			
Total	2,746	484	3,230
	85.02	14.98	100.00
Priors	0.5000	0.5000	

```
. ** cálculo LogDA 5 variables **;
. discrim logistic wR1 wR3 wR4 wR8 wR9, group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -193.84815
Iteration 2: log likelihood = -191.82893
Iteration 3: log likelihood = -191.74681
Iteration 4: log likelihood = -191.74648
```

Logistic discriminant analysis
Resubstitution classification summary

+-----+			
Key			

Number			
Percent			
+-----+			
	Classified		
True fracaso	0	1	Total
-----+-----+-----			
0	2,715	478	3,193
	85.03	14.97	100.00
1	22	15	37
	59.46	40.54	100.00
-----+-----+-----			
Total	2,737	493	3,230
	84.74	15.26	100.00
Priors	0.5000	0.5000	

```
. ** cálculo LogDA 5variables + incidencias**;
. discrim logistic wR1 wR3 wR4 wR8 wR9 incidencias, group(fracaso);
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -178.43254
Iteration 2: log likelihood = -175.37832
Iteration 3: log likelihood = -175.28922
Iteration 4: log likelihood = -175.28906
```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,826  367 | 3,193
        | 88.51  11.49 | 100.00
        |         |
      1 | 14    23 | 37
        | 37.84  62.16 | 100.00
-----+-----+-----
      Total | 2,840  390 | 3,230
            | 87.93  12.07 | 100.00
            |         |
      Priors | 0.5000  0.5000 |

```

. ** cálculo LogDA 3 variables **;
. discrim logistic wR1 wR3 wR4, group(fracaso);

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -193.18965
Iteration 2: log likelihood = -192.13322
Iteration 3: log likelihood = -192.11166
Iteration 4: log likelihood = -192.11164

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,747  446 | 3,193
        | 86.03  13.97 | 100.00
        |         |
      1 | 22    15 | 37
        | 59.46  40.54 | 100.00
-----+-----+-----

```

```

Total | 2,769  461 | 3,230
      | 85.73 14.27 | 100.00
      |         |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo LogDA 3 variables + incidencias**;
. discrim logistic wR1 wR3 wR4 incidencias, group(fracaso);

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -179.02287
Iteration 2: log likelihood = -175.99543
Iteration 3: log likelihood = -175.91167
Iteration 4: log likelihood = -175.91156

```

Logistic discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----
      0 | 2,861  332 | 3,193
      | 89.60 10.40 | 100.00
      |         |
      1 | 15    22 | 37
      | 40.54 59.46 | 100.00
-----+-----+-----
Total | 2,876  354 | 3,230
      | 89.04 10.96 | 100.00
      |         |
Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 10 variables**;
. discrim knn wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10, k(2)
group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+

```

		Classified		
True fracaso	0	1	Total	
0	3,152	41	3,193	
	98.72	1.28	100.00	
1	0	37	37	
	0.00	100.00	100.00	
Total	3,152	78	3,230	
	97.59	2.41	100.00	
Priors	0.5000	0.5000		

```
. ** cálculo knnDA con 10 variables+incidencias**;  
. discrim knn wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias, k(2)  
group(fracaso) measure(L2);
```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

		Classified		
True fracaso	0	1	Total	
0	3,162	31	3,193	
	99.03	0.97	100.00	
1	0	37	37	
	0.00	100.00	100.00	
Total	3,162	68	3,230	
	97.89	2.11	100.00	
Priors	0.5000	0.5000		

```
. ** cálculo knnDA con 7 variables**;  
. discrim knn wR1 wR3 wR4 wR5 wR7 wR8 wR9, k(2) group(fracaso) measure(L2);
```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

		Classified		
True fracaso	0	1	Total	
0	3,162	31	3,193	
	99.03	0.97	100.00	
1	0	37	37	
	0.00	100.00	100.00	
Total	3,162	68	3,230	
	97.89	2.11	100.00	
Priors	0.5000	0.5000		

```

| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----+
      0 | 3,153  40 | 3,193
      | 98.75  1.25 | 100.00
      |         |
      1 | 0    37 | 37
      | 0.00 100.00 | 100.00
-----+-----+-----+
      Total | 3,153  77 | 3,230
      | 97.62  2.38 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 7 variables+incidencias**;
. discrim knn wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias, k(2) group(fracaso)
measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
-----+-----+-----+
      0 | 3,161  32 | 3,193
      | 99.00  1.00 | 100.00
      |         |
      1 | 0    37 | 37
      | 0.00 100.00 | 100.00
-----+-----+-----+
      Total | 3,161  69 | 3,230
      | 97.86  2.14 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 5 variables**;
. discrim knn wR1 wR3 wR4 wR8 wR9, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
+-----+-----+-----+
      0 | 3,159  34 | 3,193
      | 98.94  1.06 | 100.00
      |         |
      1 | 0      37 | 37
      | 0.00 100.00 | 100.00
+-----+-----+-----+
      Total | 3,159  71 | 3,230
      | 97.80  2.20 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 5 variables+incidencias**
. discrim knn wR1 wR3 wR4 wR8 wR9 incidencias, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis
Resubstitution classification summary

```

+-----+
| Key   |
+-----+
| Number |
| Percent |
+-----+
      | Classified
True fracaso | 0    1 | Total
+-----+-----+-----+
      0 | 3,153  40 | 3,193
      | 98.75  1.25 | 100.00
      |         |
      1 | 0      37 | 37
      | 0.00 100.00 | 100.00
+-----+-----+-----+
      Total | 3,153  77 | 3,230
      | 97.62  2.38 | 100.00
      |         |
      Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 3 variables**
. discrim knn wR1 wR3 wR4 , k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis

Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,143  50 | 3,193
          | 98.43  1.57 | 100.00
          |         |
          1 | 0    37 | 37
          | 0.00 100.00 | 100.00
-----+-----+-----
          Total | 3,143  87 | 3,230
          | 97.31  2.69 | 100.00
          |         |
          Priors | 0.5000 0.5000 |

```

```

. ** cálculo KnnDA con 3 variables+incidencias**;
. discrim knn wR1 wR3 wR4 incidencias, k(2) group(fracaso) measure(L2);

```

Kth-nearest-neighbor discriminant analysis

Resubstitution classification summary

```

+-----+
| Key   |
|-----|
| Number |
| Percent |
+-----+
          | Classified
True fracaso | 0    1 | Total
-----+-----+-----
          0 | 3,161  32 | 3,193
          | 99.00  1.00 | 100.00
          |         |
          1 | 0    37 | 37
          | 0.00 100.00 | 100.00
-----+-----+-----
          Total | 3,161  69 | 3,230
          | 97.86  2.14 | 100.00
          |         |
          Priors | 0.5000 0.5000 |

```

```

. ** cálculo de logit 10 variables **;
. logit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10;

```


	0	1	Total
+	0	0	0
-	37	3193	3230
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$
 True D defined as fracaso != 0

Sensitivity	$\Pr(+ D)$	0.00%
Specificity	$\Pr(- \sim D)$	100.00%
Positive predictive value	$\Pr(D +)$.%
Negative predictive value	$\Pr(\sim D -)$	98.85%
False + rate for true $\sim D$	$\Pr(+ \sim D)$	0.00%
False - rate for true D	$\Pr(- D)$	100.00%
False + rate for classified +	$\Pr(\sim D +)$.%
False - rate for classified -	$\Pr(D -)$	1.15%
Correctly classified		98.85%

. lroc, nograph;

Logistic model for fracaso

number of observations = 3230
 area under ROC curve = 0.7576

. ** cálculo de logit 10 variables + incidencias **;
 . logit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10 incidencias;

Iteration 0: log likelihood = -202.15209
 Iteration 1: log likelihood = -176.50195
 Iteration 2: log likelihood = -171.20084
 Iteration 3: log likelihood = -170.92324
 Iteration 4: log likelihood = -170.92298
 Iteration 5: log likelihood = -170.92298

Logistic regression Number of obs = 3230
 LR chi2(11) = 62.46
 Prob > chi2 = 0.0000
 Log likelihood = -170.92298 Pseudo R2 = 0.1545

fracaso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
wR1	-.4393767	.5243667	-0.84	0.402	-1.467117 .5883633
wR2	-.0257876	.0402642	-0.64	0.522	-.1047041 .0531288
wR3	2.566873	3.311311	0.78	0.438	-3.923177 9.056923

```

wR4 | -6.367191  3.816585  -1.67  0.095  -13.84756  1.113178
wR5 |  1.802707  .8474429  2.13  0.033  .1417495  3.463665
wR6 |  2.915193  5.386322  0.54  0.588  -7.641805  13.47219
wR7 | -0.9051272  .4735433  -1.91  0.056  -1.833255  .0230006
wR8 |  1.054801  .5743624  1.84  0.066  -.0709284  2.180531
wR9 | -0.1469095  .2936907  -0.50  0.617  -.7225327  .4287136
wR10 | -0.0240154  .1231102  -0.20  0.845  -.2653069  .2172761
incidencias | 1.438678  .2311489  6.22  0.000  .9856349  1.891722
_cons | -6.07439  .8297887  -7.32  0.000  -7.700746  -4.448034
-----

```

```
. fitstat;
```

Measures of Fit for logit of fracaso

```

Log-Lik Intercept Only:  -202.152  Log-Lik Full Model:  -170.923
D(3218):                 341.846  LR(11):           62.458
                          Prob > LR:              0.000
McFadden's R2:          0.154  McFadden's Adj R2:    0.095
Maximum Likelihood R2:  0.019  Cragg & Uhler's R2:    0.163
McKelvey and Zavoina's R2: 0.177  Efron's R2:          0.036
Variance of y*:         4.000  Variance of error:    3.290
Count R2:               0.989  Adj Count R2:        0.000
AIC:                   0.113  AIC*n:               365.846
BIC:                   -25660.358  BIC':                26.424

```

```
. estat classification;
```

Logistic model for fracaso

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   1   1 |   2
- |  36 3192 | 3228
-----+-----+-----
Total |   37 3193 | 3230

```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

```

-----
Sensitivity          Pr( +| D)  2.70%
Specificity          Pr( -|~D) 99.97%
Positive predictive value  Pr(D| +) 50.00%
Negative predictive value  Pr(~D| -) 98.88%
-----
False + rate for true ~D  Pr( +|~D) 0.03%
False - rate for true D   Pr( -| D) 97.30%
False + rate for classified +  Pr(~D| +) 50.00%
False - rate for classified -  Pr( D| -) 1.12%

```

```
-----
Correctly classified          98.85%
-----
```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.8326
```

```
. ** cálculo de logit 7 variables **;
. logit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9 ;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -190.39495
Iteration 2: log likelihood = -186.73793
Iteration 3: log likelihood = -186.68413
Iteration 4: log likelihood = -186.68405
Iteration 5: log likelihood = -186.68405
```

```
Logistic regression                Number of obs = 3230
                                LR chi2(7) = 30.94
                                Prob > chi2 = 0.0001
Log likelihood = -186.68405        Pseudo R2 = 0.0765
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
wR1 | -0.0426857   .4376003   -0.10   0.922   -9.003666   .8149952
wR3 |  3.6293   2.804265   1.29   0.196   -1.866959   9.125559
wR4 | -7.452953   3.176552   -2.35   0.019   -13.67888   -1.227026
wR5 |  1.724322   .8119493   2.12   0.034   .1329307   3.315714
wR7 | -0.9418714   .4400797   -2.14   0.032   -1.804412   -0.079331
wR8 |  .7691471   .4869327   1.58   0.114   -1.1852236   1.723518
wR9 | -0.0679917   .2662158   -0.26   0.798   -.589765   .4537817
_cons | -6.000777   .7598539   -7.90   0.000   -7.490063   -4.51149
-----
```

```
. fitstat;
```

```
Measures of Fit for logit of fracaso
```

```
Log-Lik Intercept Only: -202.152   Log-Lik Full Model: -186.684
D(3222): 373.368   LR(7): 30.936
Prob > LR: 0.000
McFadden's R2: 0.077   McFadden's Adj R2: 0.037
Maximum Likelihood R2: 0.010   Cragg & Uhler's R2: 0.081
McKelvey and Zavoina's R2: 0.124   Efron's R2: 0.025
Variance of y*: 3.757   Variance of error: 3.290
```

```
Count R2:          0.989   Adj Count R2:          0.000
AIC:              0.121   AIC*n:          389.368
BIC:             -25661.157   BIC':          25.626
```

```
. estat classification;
```

```
Logistic model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230
```

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```
-----
Sensitivity          Pr( +| D)  0.00%
Specificity          Pr( -|~D) 100.00%
Positive predictive value Pr( D| +)  .%
Negative predictive value Pr(~D| -) 98.85%
```

```
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D Pr( -| D) 100.00%
False + rate for classified + Pr(~D| +)  .%
False - rate for classified - Pr( D| -)  1.15%
```

```
-----
Correctly classified          98.85%
-----
```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.7518
```

```
. ** cálculo de logit 7 variables + incidencias **;
. logit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9 incidencias;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -176.61768
Iteration 2: log likelihood = -171.48584
Iteration 3: log likelihood = -171.23639
Iteration 4: log likelihood = -171.23608
Iteration 5: log likelihood = -171.23608
```

```
Logistic regression          Number of obs = 3230
```

```

LR chi2(8) = 61.83
Prob > chi2 = 0.0000
Log likelihood = -171.23608 Pseudo R2 = 0.1529

```

fracaso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
wR1	-.345037	.495486	-0.70	0.486	-1.316172	.6260977
wR3	3.328869	2.937757	1.13	0.257	-2.42903	9.086767
wR4	-6.925118	3.391167	-2.04	0.041	-13.57168	-.2785522
wR5	1.657813	.8213473	2.02	0.044	.0480018	3.267624
wR7	-.9073765	.4723506	-1.92	0.055	-1.833167	.0184137
wR8	.8501034	.4789527	1.77	0.076	-.0886266	1.788833
wR9	-.1320701	.2690662	-0.49	0.624	-.6594302	.3952901
incidencias	1.449306	.2296002	6.31	0.000	.9992978	1.899314
_cons	-6.051433	.77926	-7.77	0.000	-7.578754	-4.524111

```
. fitstat;
```

Measures of Fit for logit of fracaso

```

Log-Lik Intercept Only: -202.152 Log-Lik Full Model: -171.236
D(3221): 342.472 LR(8): 61.832
Prob > LR: 0.000
McFadden's R2: 0.153 McFadden's Adj R2: 0.108
Maximum Likelihood R2: 0.019 Cragg & Uhler's R2: 0.161
McKelvey and Zavoina's R2: 0.171 Efron's R2: 0.035
Variance of y*: 3.967 Variance of error: 3.290
Count R2: 0.989 Adj Count R2: 0.000
AIC: 0.112 AIC*n: 360.472
BIC: -25683.973 BIC': 2.810

```

```
. estat classification;
```

Logistic model for fracaso

Classified	True		Total
	D	~D	
+	1	1	2
-	36	3192	3228
Total	37	3193	3230

```

Classified + if predicted Pr(D) >= .5
True D defined as fracaso != 0

```

Sensitivity	Pr(+ D)	2.70%
Specificity	Pr(~ ~D)	99.97%

```

Positive predictive value   Pr( D| +)  50.00%
Negative predictive value   Pr(~D| -) 98.88%
-----
False + rate for true ~D   Pr( +|~D)  0.03%
False - rate for true D    Pr( -| D)  97.30%
False + rate for classified + Pr(~D| +)  50.00%
False - rate for classified - Pr( D| -)   1.12%
-----
Correctly classified       98.85%
-----

```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```

number of observations = 3230
area under ROC curve = 0.8288

```

```

. ** cálculo de logit 5 variables **;
. logit fracaso wR1 wR3 wR4 wR8 wR9;

```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -193.84815
Iteration 2: log likelihood = -191.77406
Iteration 3: log likelihood = -191.74652
Iteration 4: log likelihood = -191.74648
Iteration 5: log likelihood = -191.74648

```

```

Logistic regression               Number of obs = 3230
                                LR chi2(5) = 20.81
                                Prob > chi2 = 0.0009
Log likelihood = -191.74648      Pseudo R2 = 0.0515

```

```

-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
wR1 |  -.282446  .353163  -0.80  0.424  -1.0146327  .4097407
wR3 |  3.708092  2.729268   1.36  0.174  -1.641175  9.05736
wR4 | -6.927969  3.095361  -2.24  0.025  -12.99476  -8.611733
wR8 |  .471166  .5174283   0.91  0.363  -0.5429748  1.485307
wR9 | -0.0611807 .2642753  -0.23  0.817  -0.5791508  .4567895
_cons | -4.588111  .3310461 -13.86  0.000  -5.23695  -3.939273
-----

```

```
. fitstat;
```

```
Measures of Fit for logit of fracaso
```

```

Log-Lik Intercept Only: -202.152  Log-Lik Full Model: -191.746
D(3224):                 383.493  LR(5):                20.811

```

```

                                Prob > LR:          0.001
McFadden's R2:                 0.051   McFadden's Adj R2:   0.022
Maximum Likelihood R2:         0.006   Cragg & Uhler's R2:   0.055
McKelvey and Zavoina's R2:    0.061   Efron's R2:          0.023
Variance of y*:                3.505   Variance of error:   3.290
Count R2:                      0.989   Adj Count R2:       0.000
AIC:                           0.122   AIC*n:              395.493
BIC:                           -25667.192   BIC':               19.590

```

```
. estat classification;
```

```
Logistic model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```

-----
Sensitivity          Pr(+|D)  0.00%
Specificity          Pr(-|~D) 100.00%
Positive predictive value Pr(D|+)  .%
Negative predictive value Pr(~D|-) 98.85%
-----

```

```

False + rate for true ~D Pr(+|~D)  0.00%
False - rate for true D Pr(-|D)    100.00%
False + rate for classified + Pr(~D|+)  .%
False - rate for classified - Pr(D|-)  1.15%
-----

```

```
Correctly classified          98.85%
-----
```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```

number of observations = 3230
area under ROC curve = 0.6742

```

```
. ** cálculo de logit 5 variables + incidencias **;
. logit fracaso wR1 wR3 wR4 wR8 wR9 incidencias;
```

```

Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -178.43254
Iteration 2: log likelihood = -175.35512

```



```
-----
Sensitivity          Pr( +| D)  2.70%
Specificity          Pr( -|~D) 100.00%
Positive predictive value Pr( D| +) 100.00%
Negative predictive value Pr(~D| -) 98.89%
-----
```

```
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D  Pr( -| D)  97.30%
False + rate for classified + Pr(~D| +)  0.00%
False - rate for classified - Pr( D| -)  1.11%
-----
```

```
-----
Correctly classified          98.89%
-----
```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.7733
```

```
. ** cálculo de logit 3 variables **;
. logit fracaso wR1 wR3 wR4 ;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -193.18965
Iteration 2: log likelihood = -192.11542
Iteration 3: log likelihood = -192.11164
Iteration 4: log likelihood = -192.11164
```

```
Logistic regression          Number of obs = 3230
                             LR chi2(3) = 20.08
                             Prob > chi2 = 0.0002
Log likelihood = -192.11164   Pseudo R2 = 0.0497
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
wR1 | -0.2587969   .3513135   -0.74   0.461   -0.9473587   .4297649
wR3 |  3.722627    2.7117    1.37   0.170   -1.592207    9.037461
wR4 | -6.628814    3.0642   -2.16   0.031  -12.63454   -6.230927
_cons | -4.541243    .3252336  -13.96   0.000   -5.178689   -3.903797
-----
```

```
. fitstat;
```

```
Measures of Fit for logit of fracaso
```

```
Log-Lik Intercept Only: -202.152   Log-Lik Full Model: -192.112
D(3226):                 384.223   LR(3):                 20.081
```

```

                Prob > LR:                0.000
McFadden's R2:    0.050  McFadden's Adj R2:    0.030
Maximum Likelihood R2:    0.006  Cragg & Uhler's R2:    0.053
McKelvey and Zavoina's R2:    0.061  Efron's R2:    0.024
Variance of y*:    3.505  Variance of error:    3.290
Count R2:    0.989  Adj Count R2:    0.000
AIC:    0.121  AIC*n:    392.223
BIC:    -25682.623  BIC':    4.160

```

```
. estat classification;
```

```
Logistic model for fracaso
```

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |   0   0 |   0
- |  37 3193 | 3230
-----+-----+-----
Total |  37 3193 | 3230

```

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```

-----
Sensitivity      Pr( +| D)  0.00%
Specificity      Pr( -|~D) 100.00%
Positive predictive value  Pr( D| +)  .%
Negative predictive value  Pr(~D| -) 98.85%
-----

```

```

False + rate for true ~D  Pr( +|~D)  0.00%
False - rate for true D   Pr( -| D) 100.00%
False + rate for classified +  Pr(~D| +)  .%
False - rate for classified -  Pr( D| -)  1.15%
-----

```

```
Correctly classified      98.85%
```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
```

```
area under ROC curve = 0.6908
```

```
. ** cálculo de logit 3 variables + incidencias **;
```

```
. logit fracaso wR1 wR3 wR4 incidencias;
```

```
Iteration 0: log likelihood = -202.15209
```

```
Iteration 1: log likelihood = -179.02287
```

```
Iteration 2: log likelihood = -175.99543
```



```

Specificity          Pr( ~|~D) 100.00%
Positive predictive value Pr( D|+) 100.00%
Negative predictive value Pr(~D|-) 98.89%
-----
False + rate for true ~D Pr(+|~D) 0.00%
False - rate for true D Pr(-|D) 97.30%
False + rate for classified + Pr(~D|+) 0.00%
False - rate for classified - Pr(D|-) 1.11%
-----
Correctly classified          98.89%
-----

```

```
. lroc, nograph;
```

```
Logistic model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.7686
```

```
. ** cálculo de probit 10 variables **;
```

```
. probit fracaso wR1 wR2 wR3 wR4 wR5 wR6 wR7 wR8 wR9 wR10;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -188.99831
Iteration 2: log likelihood = -185.72157
Iteration 3: log likelihood = -185.69976
Iteration 4: log likelihood = -185.69975
```

```

Probit regression              Number of obs = 3230
                             LR chi2(10) = 32.90
                             Prob > chi2 = 0.0003
Log likelihood = -185.69975    Pseudo R2 = 0.0814

```

```

-----+-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
wR1 |   .0746465   .1923355   0.39  0.698   -3.023242   .4516173
wR2 |  -0.0124421   .0157599  -0.79  0.430   -0.0433309   .0184466
wR3 |   .9192189   1.489191   0.62  0.537   -1.999542   3.83798
wR4 |  -2.524215   1.693014  -1.49  0.136   -5.842461   .7940317
wR5 |   .7001611   .3066338   2.28  0.022   .0991699   1.301152
wR6 |   1.862694   2.203954   0.85  0.398   -2.456976   6.182364
wR7 |  -0.4432024   .1908193  -2.32  0.020   -0.8172013  -0.0692035
wR8 |   .4563071   .2389049   1.91  0.056   -0.0119379   .9245522
wR9 |  -0.0476218   .1241728  -0.38  0.701   -0.2909959   .1957524
wR10 | -0.0190171   .047182   -0.40  0.687   -0.1114922   .073458
_cons | -2.956476   .3101747  -9.53  0.000   -3.564407  -2.348545
-----+-----

```

```
. fitstat;
```


Classified	----- True -----		Total
	D	~D	
+	1	0	1
-	36	3193	3229
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$
 True D defined as fracaso != 0

Sensitivity	$\Pr(+ D)$	2.70%
Specificity	$\Pr(- \sim D)$	100.00%
Positive predictive value	$\Pr(D +)$	100.00%
Negative predictive value	$\Pr(\sim D -)$	98.89%
False + rate for true ~D	$\Pr(+ \sim D)$	0.00%
False - rate for true D	$\Pr(- D)$	97.30%
False + rate for classified +	$\Pr(\sim D +)$	0.00%
False - rate for classified -	$\Pr(D -)$	1.11%
Correctly classified		98.89%

. lroc, nograph;

Probit model for fracaso

number of observations = 3230
 area under ROC curve = 0.8389

. ** cálculo de probit 7 variables **;
 . probit fracaso wR1 wR3 wR4 wR5 wR7 wR8 wR9;

Iteration 0: log likelihood = -202.15209
 Iteration 1: log likelihood = -189.41439
 Iteration 2: log likelihood = -186.29596
 Iteration 3: log likelihood = -186.27934
 Iteration 4: log likelihood = -186.27934

Probit regression	Number of obs =	3230
	LR chi2(7) =	31.75
	Prob > chi2 =	0.0000
Log likelihood = -186.27934	Pseudo R2 =	0.0785

fracaso	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
wR1	.1145792	.1841005	0.62	0.534	- .2462512 .4754096
wR3	1.457822	1.278564	1.14	0.254	-1.048117 3.963761

```

wR4 | -2.965688  1.456961  -2.04  0.042  -5.82128  -.110097
wR5 |  .6449897  .300833   2.14  0.032  .0553679  1.234612
wR7 | -4.511928  .1886467  -2.39  0.017  -.8209336  -.0814519
wR8 |  .3549366  .1998174   1.78  0.076  -.0366984  .7465716
wR9 | -.046265   .1139308  -0.41  0.685  -.2695653  .1770354
_cons | -2.953055  .2887933  -10.23  0.000  -3.51908  -2.387031
-----

```

```
. fitstat;
```

Measures of Fit for probit of fracaso

```

Log-Lik Intercept Only:  -202.152  Log-Lik Full Model:  -186.279
D(3222):                 372.559  LR(7):             31.746
                          Prob > LR:                 0.000
McFadden's R2:           0.079  McFadden's Adj R2:   0.039
Maximum Likelihood R2:   0.010  Cragg & Uhler's R2:  0.083
McKelvey and Zavoina's R2: 0.078  Efron's R2:         0.018
Variance of y*:         1.084  Variance of error:   1.000
Count R2:                0.989  Adj Count R2:       0.000
AIC:                     0.120  AIC*n:              388.559
BIC:                     -25661.966  BIC':               24.816

```

```
. estat classification;
```

Probit model for fracaso

```

----- True -----
Classified |   D   ~D | Total
-----+-----+-----
+ |    0    0 |    0
- |   37  3193 |  3230
-----+-----+-----
Total |   37  3193 |  3230

```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso != 0

```

-----
Sensitivity           Pr( +| D)  0.00%
Specificity           Pr( -|~D) 100.00%
Positive predictive value  Pr(D| +)  .%
Negative predictive value  Pr(~D| -) 98.85%
-----

```

```

False + rate for true ~D  Pr( +|~D)  0.00%
False - rate for true D   Pr( -| D) 100.00%
False + rate for classified +  Pr(~D| +)  .%
False - rate for classified -  Pr( D| -)  1.15%
-----

```

```

Correctly classified           98.85%
-----

```


BIC: -25687.114 BIC': -0.331

. estat classification;

Probit model for fracaso

Classified	True		Total
	D	~D	
+	1	0	1
-	36	3193	3229
Total	37	3193	3230

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso $\neq 0$

Sensitivity	$\Pr(+ D)$	2.70%
Specificity	$\Pr(- \sim D)$	100.00%
Positive predictive value	$\Pr(D +)$	100.00%
Negative predictive value	$\Pr(\sim D -)$	98.89%
False + rate for true ~D	$\Pr(+ \sim D)$	0.00%
False - rate for true D	$\Pr(- D)$	97.30%
False + rate for classified +	$\Pr(\sim D +)$	0.00%
False - rate for classified -	$\Pr(D -)$	1.11%
Correctly classified		98.89%

. lroc, nograph;

Probit model for fracaso

number of observations = 3230

area under ROC curve = 0.8350

. ** cálculo de probit 5 variables **;
. probit fracaso wR1 wR3 wR4 wR8 wR9;

Iteration 0: log likelihood = -202.15209

Iteration 1: log likelihood = -194.69551

Iteration 2: log likelihood = -191.84403

Iteration 3: log likelihood = -191.83812

Iteration 4: log likelihood = -191.83812

Probit regression

Number of obs	=	3230
LR chi2(5)	=	20.63
Prob > chi2	=	0.0010
Log likelihood	=	-191.83812
Pseudo R2	=	0.0510

```
-----+-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
wR1 | -.0440674  .1492984  -0.30  0.768  -.336687  .2485521
wR3 |  1.529136  1.222252   1.25  0.211  -.8664343  3.924707
wR4 | -2.781741  1.39303   -2.00  0.046  -5.512029  -.0514534
wR8 |  .19833  .1981614   1.00  0.317  -.1900592  .5867192
wR9 | -.0422817  .1120632  -0.38  0.706  -.2619215  .1773581
_cons | -2.37722  .1389823  -17.10  0.000  -2.649621  -2.10482
-----+-----
```

```
. fitstat;
```

Measures of Fit for probit of fracaso

```
Log-Lik Intercept Only:  -202.152  Log-Lik Full Model:  -191.838
D(3224):                 383.676  LR(5):             20.628
                          Prob > LR:                0.001
McFadden's R2:          0.051  McFadden's Adj R2:  0.021
Maximum Likelihood R2:  0.006  Cragg & Uhler's R2:  0.054
McKelvey and Zavoina's R2: 0.036  Efron's R2:         0.017
Variance of y*:         1.038  Variance of error:  1.000
Count R2:                0.989  Adj Count R2:       0.000
AIC:                     0.123  AIC*n:             395.676
BIC:                    -25667.009  BIC':              19.773
```

```
. estat classification;
```

Probit model for fracaso

```
----- True -----
Classified |   D    ~D | Total
-----+-----+-----
+ |   0    0 |   0
- |   37  3193 |  3230
-----+-----+-----
Total |   37  3193 |  3230
```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso != 0

```
-----+-----
Sensitivity           Pr(+|D)  0.00%
Specificity           Pr(-|~D) 100.00%
Positive predictive value  Pr(D|+)  .%
Negative predictive value  Pr(~D|-) 98.85%
-----+-----
False + rate for true ~D  Pr(+|~D)  0.00%
False - rate for true D   Pr(-|D)  100.00%
False + rate for classified +  Pr(~D|+)  .%
```


Log likelihood = -192.3415 Pseudo R2 = 0.0485

```
-----+-----
fracaso |   Coef.  Std. Err.   z  P>|z|  [95% Conf. Interval]
-----+-----
wR1 | -0.0473528  .1492055  -0.32  0.751  -0.3397901  .2450845
wR3 |  1.598959  1.21286   1.32  0.187  -0.7782035  3.976122
wR4 | -2.74236  1.384715  -1.98  0.048  -5.456352  -0.283679
_cons | -2.347156  .1354602 -17.33  0.000  -2.612653  -2.081659
-----+-----
```

. fitstat;

Measures of Fit for probit of fracaso

```
Log-Lik Intercept Only:  -202.152  Log-Lik Full Model:  -192.341
D(3226):                 384.683  LR(3):              19.621
                          Prob > LR:                0.000
McFadden's R2:          0.049  McFadden's Adj R2:    0.029
Maximum Likelihood R2:  0.006  Cragg & Uhler's R2:    0.051
McKelvey and Zavoina's R2: 0.036  Efron's R2:           0.018
Variance of y*:         1.038  Variance of error:    1.000
Count R2:                0.989  Adj Count R2:         0.000
AIC:                     0.122  AIC*n:                392.683
BIC:                     -25682.163  BIC':                 4.620
```

. estat classification;

Probit model for fracaso

```
----- True -----
Classified |   D      ~D | Total
-----+-----+-----
+ |   0      0 |   0
- |   37    3193 |  3230
-----+-----+-----
Total |   37    3193 |  3230
```

Classified + if predicted $\Pr(D) \geq .5$

True D defined as fracaso != 0

```
-----+-----
Sensitivity           Pr( +| D)  0.00%
Specificity           Pr( -|~D) 100.00%
Positive predictive value  Pr(D| +)  .%
Negative predictive value  Pr(~D| -) 98.85%
-----+-----
False + rate for true ~D  Pr( +|~D)  0.00%
False - rate for true D   Pr( -| D) 100.00%
False + rate for classified +  Pr(~D| +)  .%
False - rate for classified -  Pr( D| -)  1.15%
```

```
-----
Correctly classified          98.85%
-----
```

```
. lroc, nograph;
```

```
Probit model for fracaso
```

```
number of observations = 3230
area under ROC curve = 0.7103
```

```
. ** cálculo de probit 3 variables + incidencias **;
. probit fracaso wR1 wR3 wR4 incidencias;
```

```
Iteration 0: log likelihood = -202.15209
Iteration 1: log likelihood = -184.745
Iteration 2: log likelihood = -174.33308
Iteration 3: log likelihood = -174.31815
Iteration 4: log likelihood = -174.31815
```

```
Probit regression                Number of obs = 3230
                                LR chi2(4) = 55.67
                                Prob > chi2 = 0.0000
Log likelihood = -174.31815      Pseudo R2 = 0.1377
```

```
-----
fracaso |   Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
wR1 | -1.1546698   .1650297   -0.94   0.349   -1.478122   -.1687825
wR3 |  1.734921   1.27252    1.36   0.173   -.7591717   4.229014
wR4 | -2.808646   1.456475   -1.93   0.054   -5.663284   .0459923
incidencias | .6898876   .1062441    6.49   0.000   .4816531   .8981222
_cons | -2.417059   .1474311  -16.39   0.000   -2.706019   -2.1281
-----
```

```
. fitstat;
```

```
Measures of Fit for probit of fracaso
```

```
Log-Lik Intercept Only: -202.152   Log-Lik Full Model: -174.318
D(3225):                348.636   LR(4):                55.668
                        Prob > LR:                0.000
McFadden's R2:          0.138   McFadden's Adj R2:      0.113
Maximum Likelihood R2:  0.017   Cragg & Uhler's R2:     0.145
McKelvey and Zavoina's R2: 0.078   Efron's R2:             0.030
Variance of y*:         1.085   Variance of error:      1.000
Count R2:                0.989   Adj Count R2:           0.027
AIC:                     0.111   AIC*n:                  358.636
BIC:                     -25710.129   BIC':                   -23.347
```

```
. estat classification;
```

```
Probit model for fracaso
```

Classified	True		Total
	D	~D	
+	1	0	1
-	36	3193	3229
Total	37	3193	3230

```
Classified + if predicted Pr(D) >= .5
```

```
True D defined as fracaso != 0
```

```
-----
Sensitivity          Pr( +| D)  2.70%
Specificity          Pr( -|~D) 100.00%
Positive predictive value Pr( D| +) 100.00%
Negative predictive value Pr(~D| -) 98.89%
-----
```

```
-----
False + rate for true ~D Pr( +|~D)  0.00%
False - rate for true D  Pr( -| D)  97.30%
False + rate for classified + Pr(~D| +)  0.00%
False - rate for classified - Pr( D| -)  1.11%
-----
```

```
-----
Correctly classified          98.89%
-----
```

```
. lroc, nograph;
```

```
Probit model for fracaso
```

```
number of observations = 3230
```

```
area under ROC curve = 0.7849
```

```
. log close;
```

```
name: <unnamed>
```

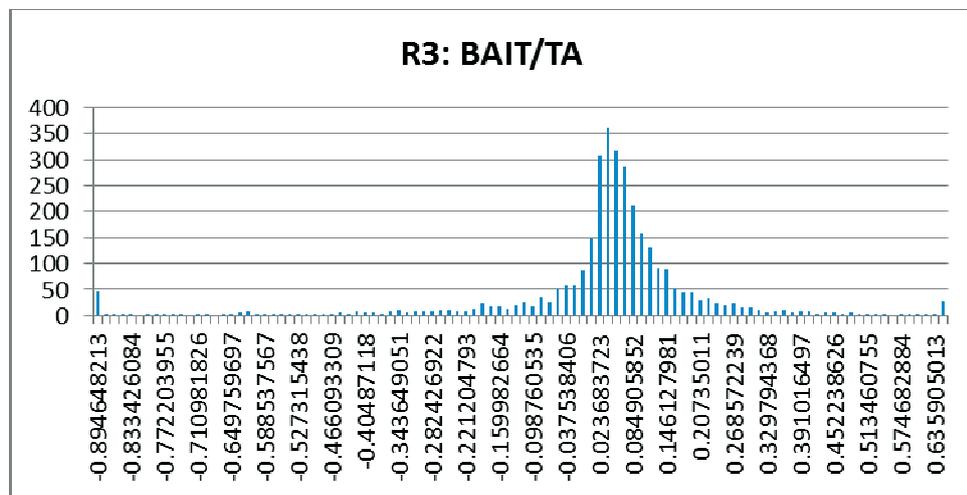
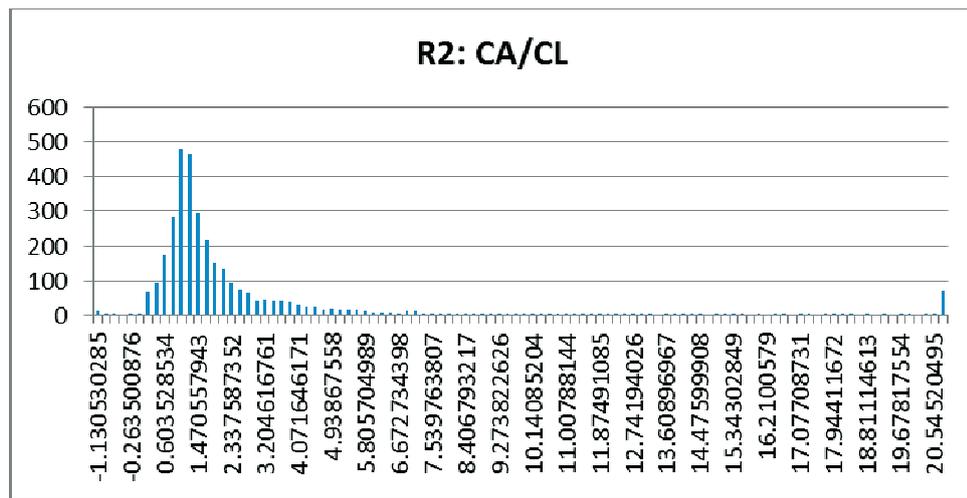
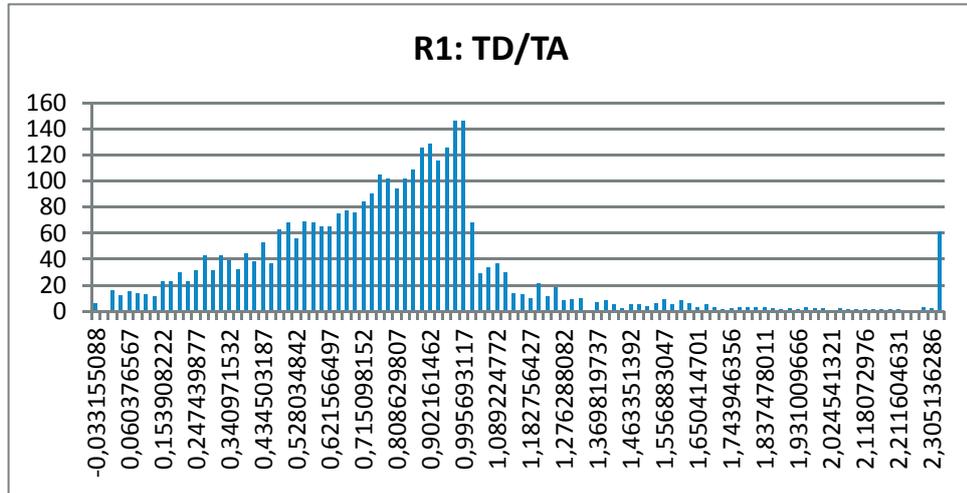
```
log: C:\Users\Javier Castaño\Desktop\DOCUMENTOS
```

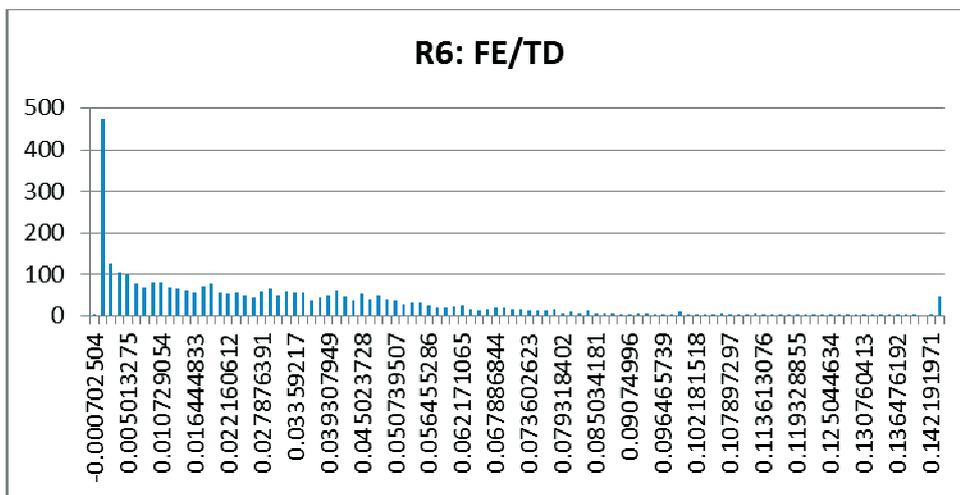
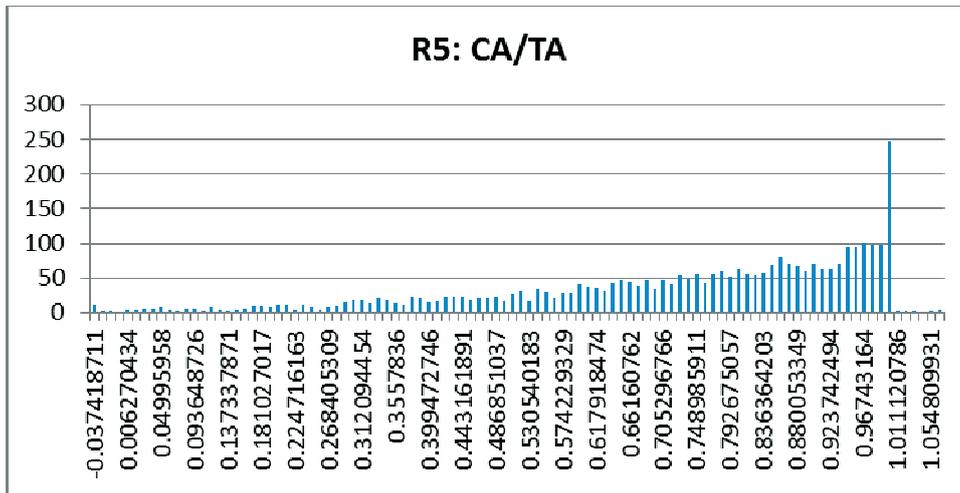
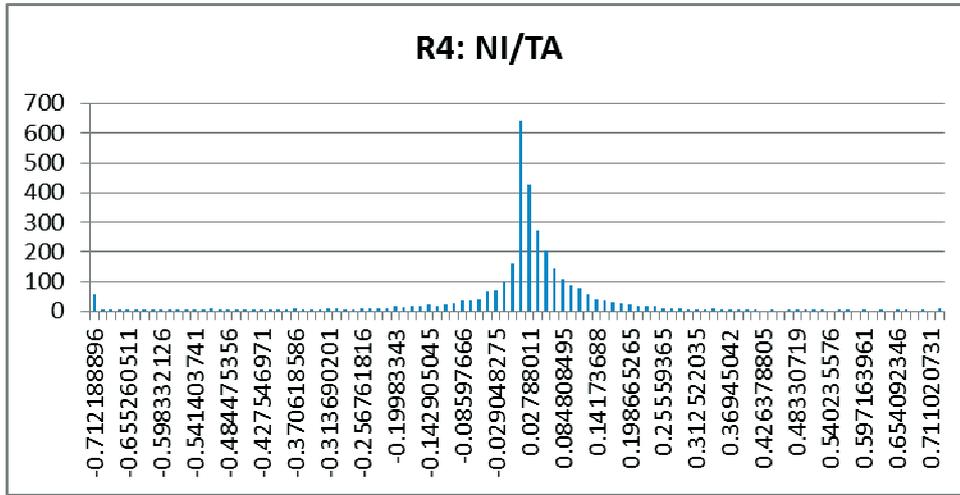
```
TESIS\TESIS\salidatesis.log
```

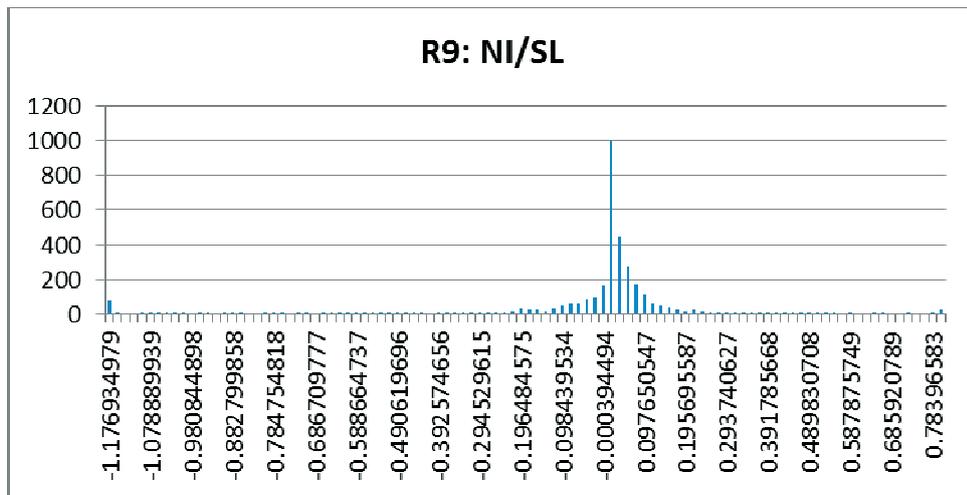
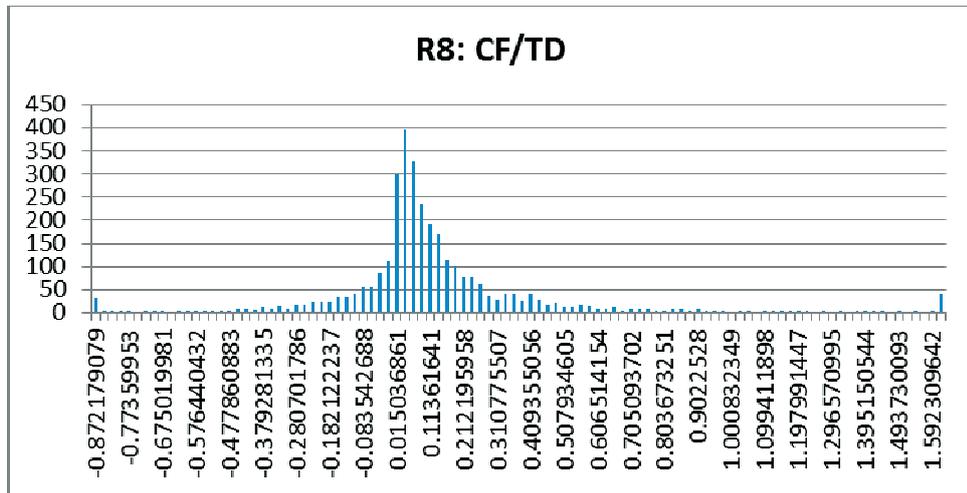
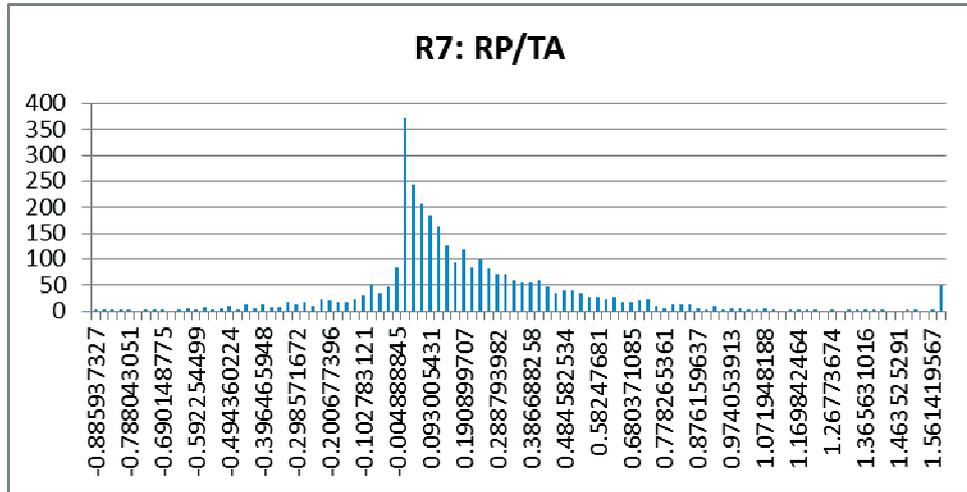
```
log type: text
```

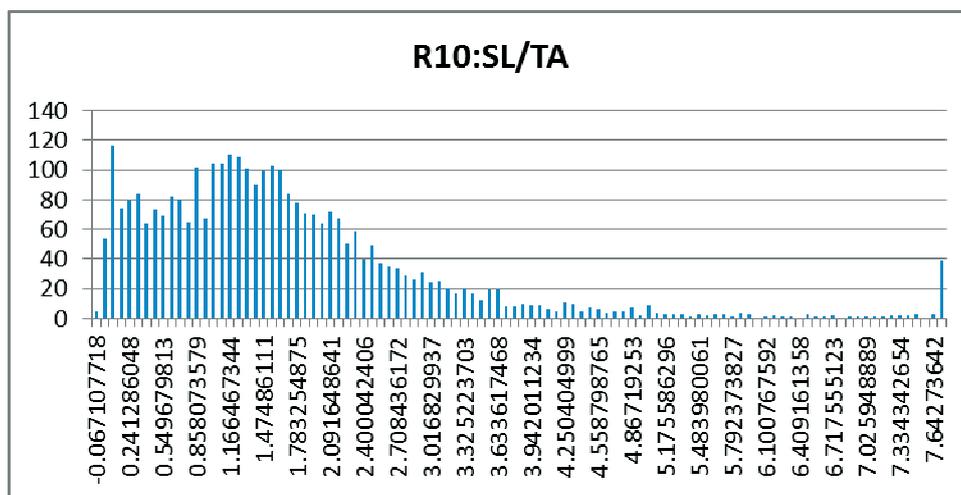
```
closed on: 3 Sep 2012, 20:11:19
```

Anexo VI Histogramas de frecuencias









Anexo VII Selección de ratios para DEA e identificación de inputs-outputs

Como hemos visto en esta tercera parte de la tesis, las diferencias de percentiles muestran cuándo las empresas fracasadas presentan valores de un ratio superiores y cuándo inferiores a los valores representativos de las empresas sanas. Para los tres años analizados en el estudio empírico anual, ese valor es consistente con el razonamiento económico. A efectos de la metodología DEA, aplicada en la segunda parte de esta tesis, las diferencias de percentiles permiten seleccionar qué ratios deben tratarse como *inputs* y cuáles como *outputs*.

Ratio	Puesto	BMP/SMP	DEA role
R1=TD/TA	2,1,4=2,3	SMP	input
R2=CA/CL	8,9,8=8,3	BMP	output
R3=BAIT/TA	9,10,5=8	BMP	output
R4=NI/TA	4,4,2=3,3	BMP	output
R5=CA/TA	5,5,6=5,3	SMP	input
R6=FE/TD	6,7,9=7,3	SMP	input
R7=RP/TA	3,6,7=5,3	BMP	output
R8=CF/TD	1,2,1=1,3	BMP	output
R9=NI/SL	7,3,3=4,3	BMP	output
R10=SL/TA	10,8,10=9,3	SMP	input

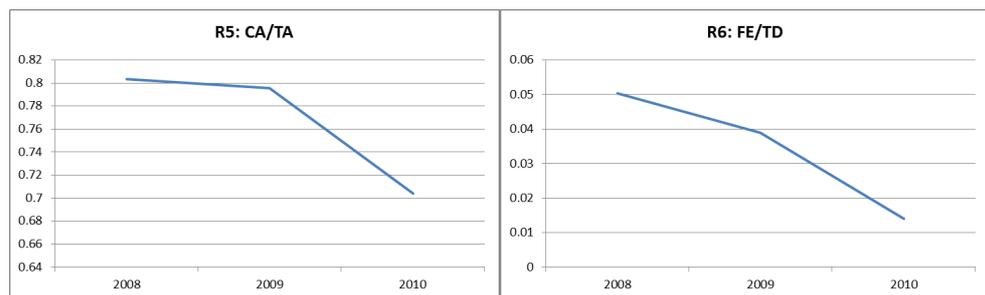
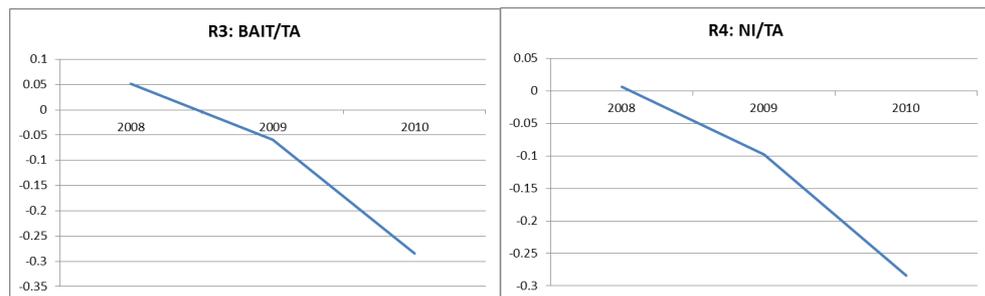
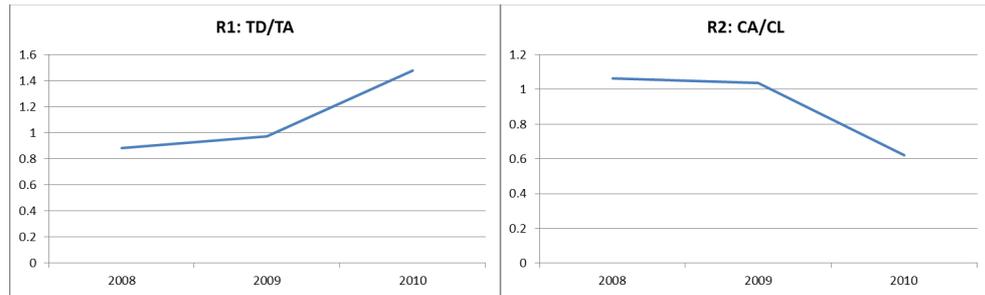
BMP: mediana fracaso bajo mediana población; SMP: mediana fracaso sobre mediana población.

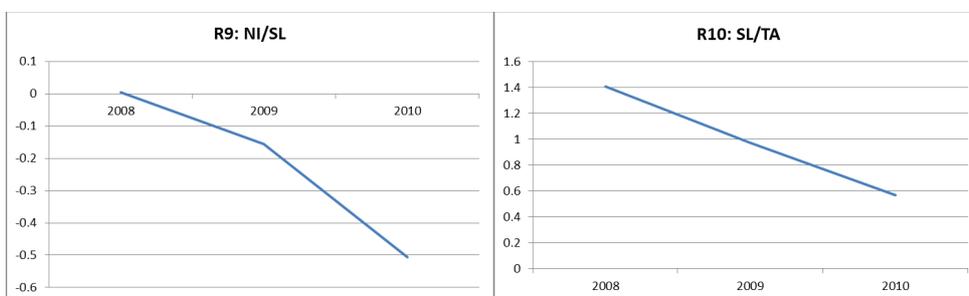
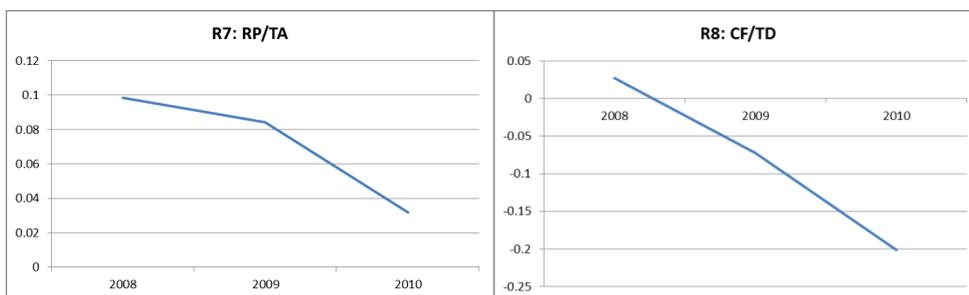
Si se quieren seleccionar un número de ratios para varios periodos, podríamos utilizar un criterio simple de ordenación a partir de la ordenación anterior. En la tabla anterior se ha calculado el puesto medio del ratio en los tres años del estudio. Así, obtenemos en amarillo los cuatro ratios con más poder discriminante, en amarillo claro, dos ratios con un poder discriminante intermedio, y en blanco, otros cuatro ratios con menor poder discriminante. Esto nos permite seleccionar los ratios en mayor o menor número, según la capacidad para asumir variables (o la necesidad de variables adicionales) de la metodología de análisis de fracaso que estemos utilizando. Pero resulta más útil clasificar los ratios para un periodo de varios años usando la distancia media al fracaso (DMF) de cada ratio en esos años. Como se puede ver en la tabla que viene a continuación, se consigue mayor

precisión en la clasificación. Para la población utilizada, vemos que R7 tiene mayor poder discriminante que R5, aunque ocupaban un puesto medio equivalente para los tres periodos en la tabla anterior.

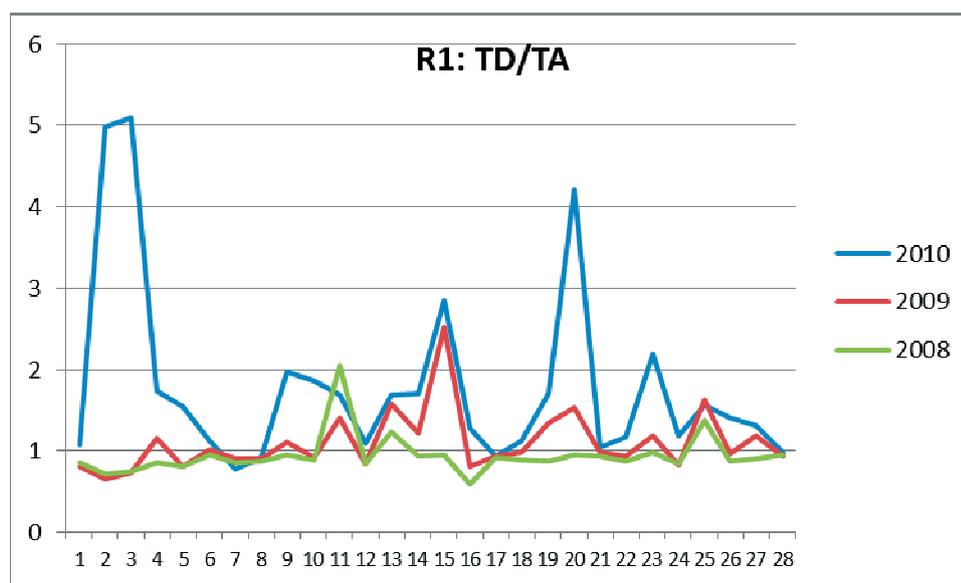
Ratio	Puesto DMF	BMP/SMP	DEA role
R1=TD/TA	0,218	SMP	input
R2=CA/CL	0,093	BMP	output
R3=BAIT/TA	0,14466667	BMP	output
R4=NI/TA	0,217	BMP	output
R5=CA/TA	0,15033333	SMP	input
R6=FE/TD	0,11833333	SMP	input
R7=RP/TA	0,16366667	BMP	output
R8=CF/TD	0,23266667	BMP	output
R9=NI/SL	0,19333333	BMP	output
R10=SL/TA	0,06033333	SMP	input

Anexo VIII Representación gráfica de medianas de ratios de empresas españolas fracasadas en 2010. Sector de la construcción. Periodo 2008-2010

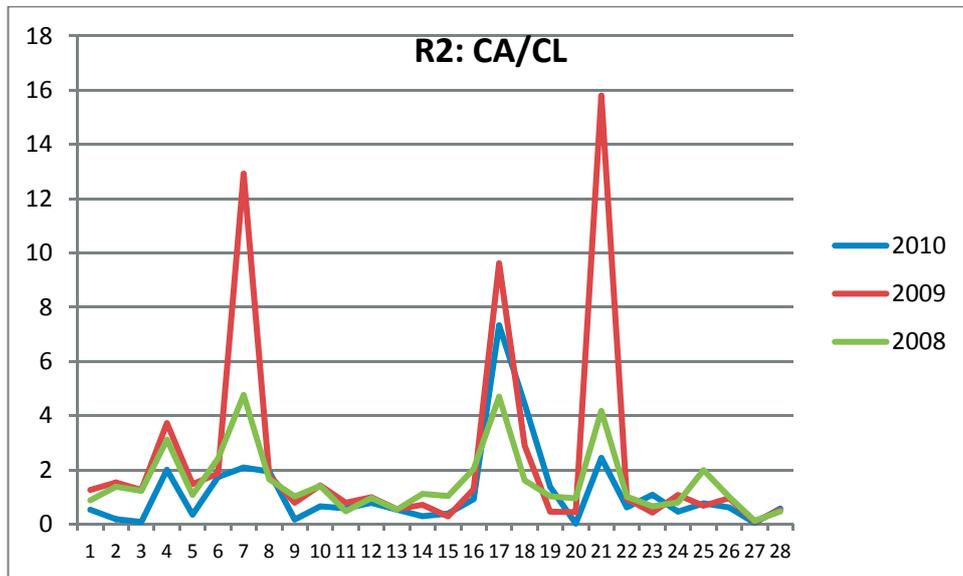




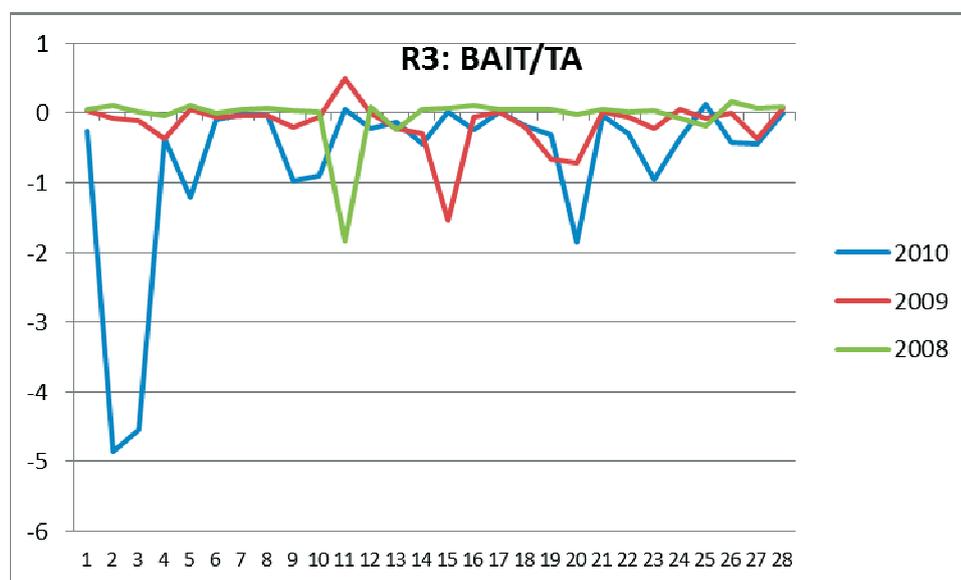
Anexo IX Representación de los valores de los ratios, de 2008 a 2010, para las 28 empresas fracasadas en 2010



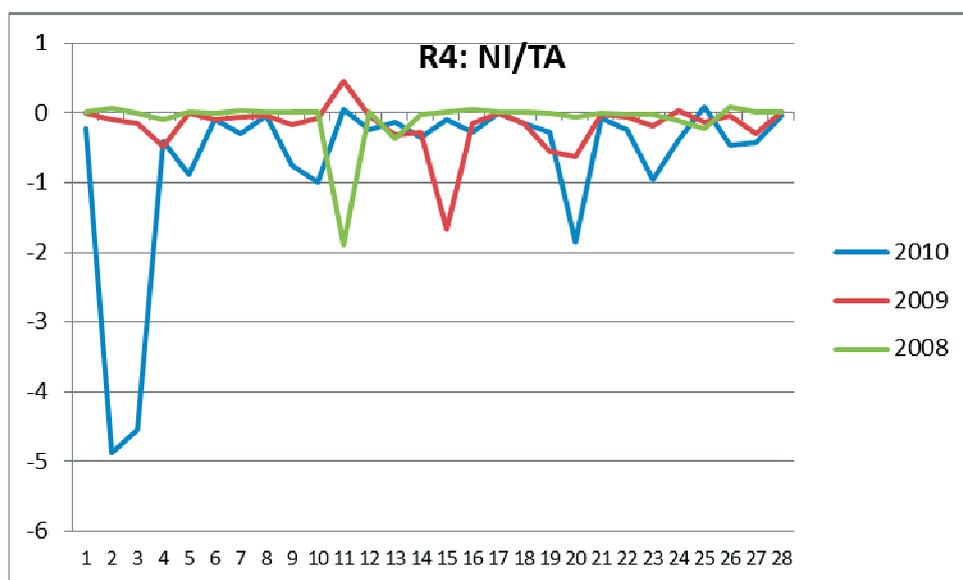
R1: TD/TA, deuda total/ activo total, es una medida de endeudamiento que indica una obligación financiera a largo plazo. Un aumento en el apalancamiento aumentaría la probabilidad de dificultades financieras en tanto que una reducción en los flujos de efectivo podría significar no disponer de fondos suficientes para pagar las deudas, lo que daría como resultado la quiebra. Los datos representados en el gráfico confirman el incremento de endeudamiento en las empresas que fracasan durante los últimos tres periodos.



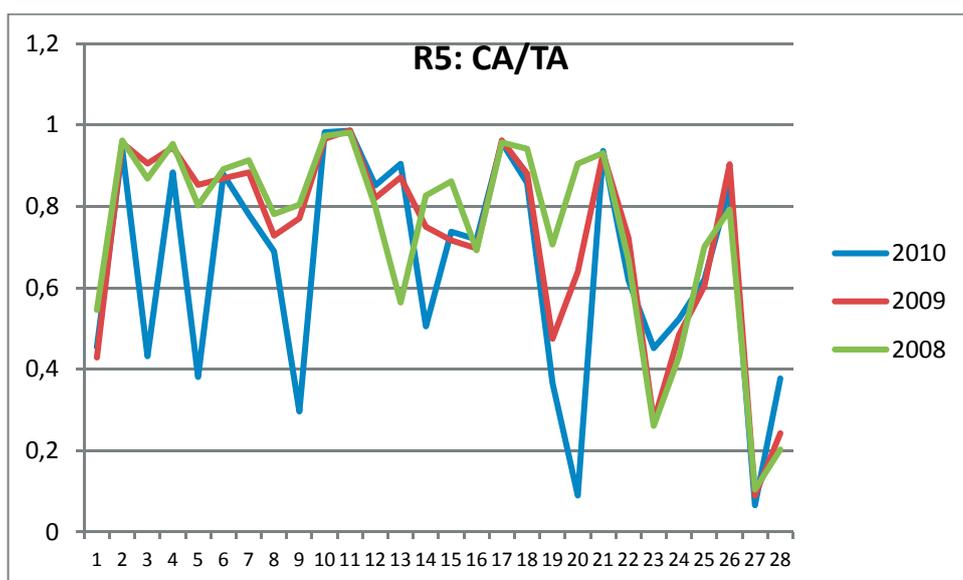
R2: CA/CL, activo corriente/ pasivo corriente, es una medida de equilibrio económico-financiero a corto plazo. Dado que las deudas comerciales son una parte importante en los pasivos de las pequeñas empresas (Altman *et al.*, 2008), y el crédito a clientes se extiende en períodos de crisis financiera, las quiebras de pequeñas empresas están fuertemente influenciadas por este tipo de deudas. Por tanto, a medida que nos acercamos al momento de la quiebra el denominador es mayor y el numerador más pequeño lo que nos da como consecuencia que los valores van reduciéndose. Se puede ver esta tendencia en los valores representados en el gráfico.



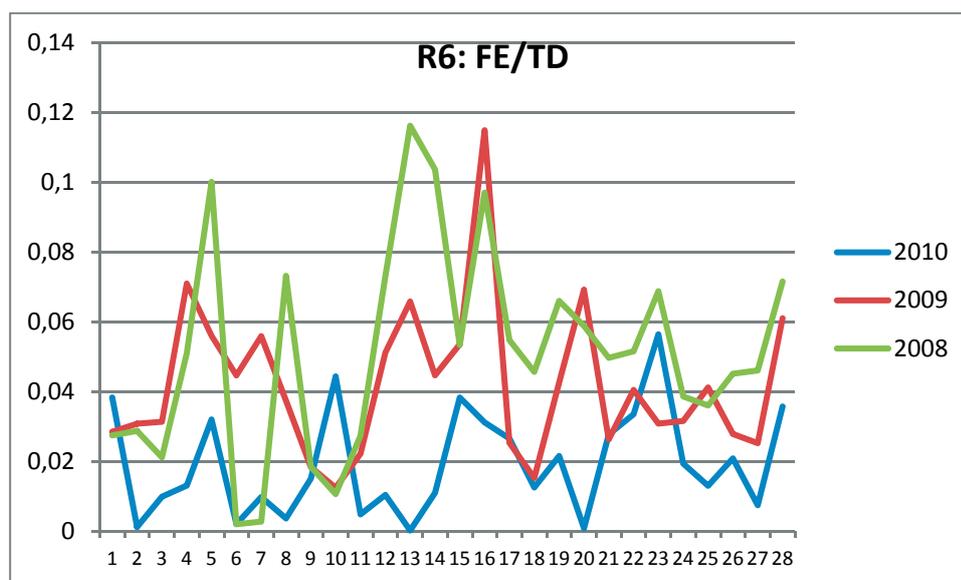
R3: BAIT/TA, beneficios antes de intereses e impuestos/ activo total, es una medida de rentabilidad económica de la empresa. El empeoramiento del negocio, reflejado en su reducción de beneficio económico, es una causa frecuente de quiebra o proximidad a la quiebra. El gráfico confirma que en los dos años previos a fracaso este ratio ya estaba muy deteriorado y, además que en esos tres años la tendencia general es la reducción en el grupo estudiado.



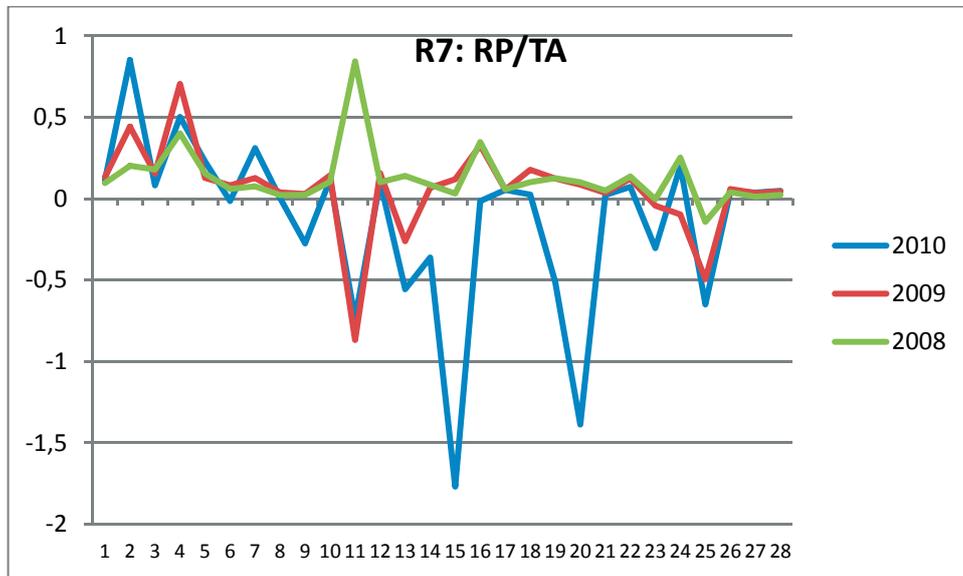
R4: NI/TA, beneficio neto / activo total, es una medida más general de rentabilidad económica. Al igual que el ratio anterior, el gráfico muestra un deterioro temprano, en general, y que en los tres últimos años el deterioro aún progresa hacia valores negativos.



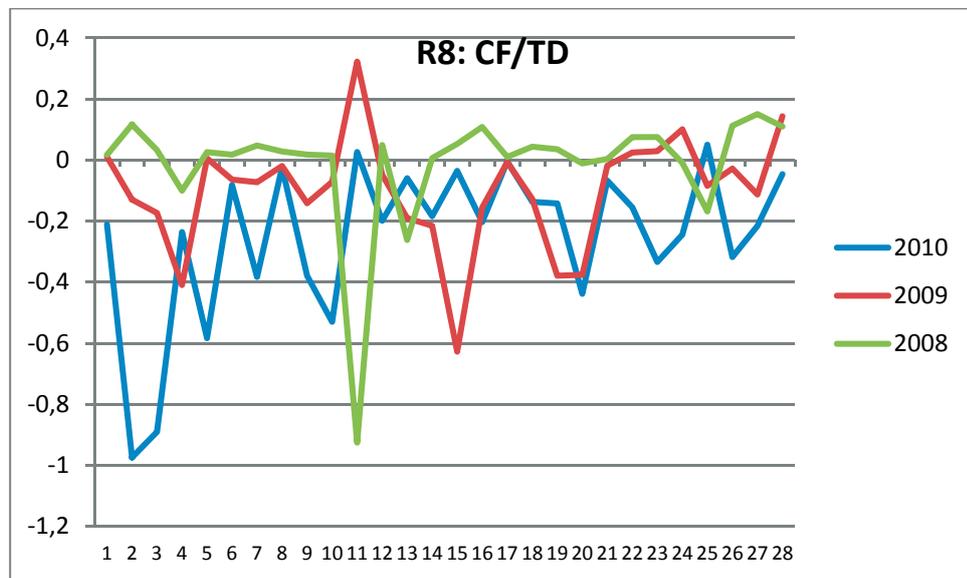
R5: CA/TA, activo corriente/ activo total, es una medida de estructura económica. La teoría nos dice que a medida que la empresa se acerca a la quiebra el valor de los activos líquidos y realizables se va reduciendo, por lo que el ratio se reduce, como confirma la gráfica en la mayoría de los casos estudiados.



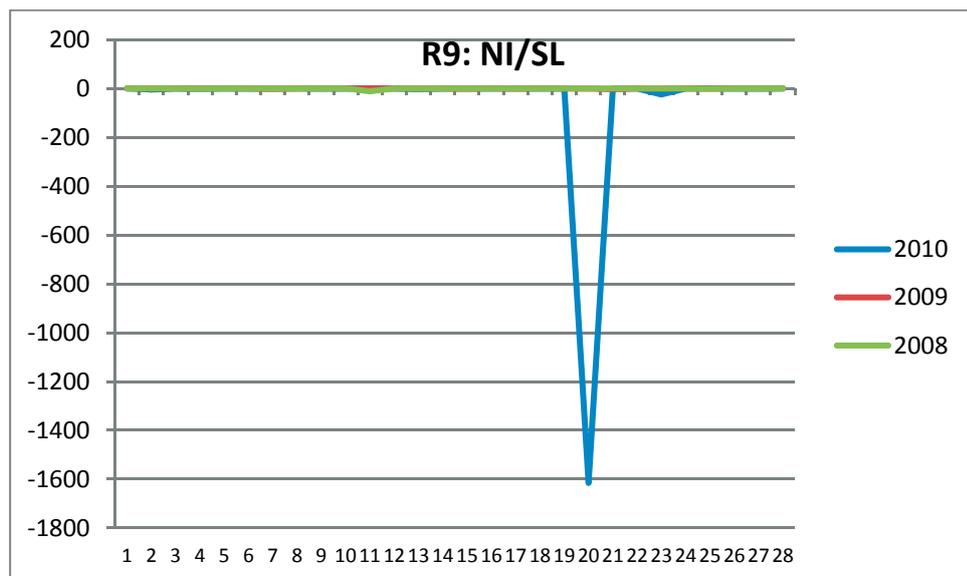
R6: FE/TD, gastos financieros/ deuda total, es una medida del coste financiero por unidad de deuda, indicativo del riesgo. El valor del numerador depende del apalancamiento de la empresa, aunque también está sujeto a la evolución de los tipos de interés del mercado financiero. El denominador es previsible que vaya aumentando ya que a medida que la empresa genera menos resultados, tiene que compensar con deudas la reducción de sus fondos de rotación. La proporción de aumento de los gastos financieros (numerador) es inferior a la del incremento de la financiación (denominador).



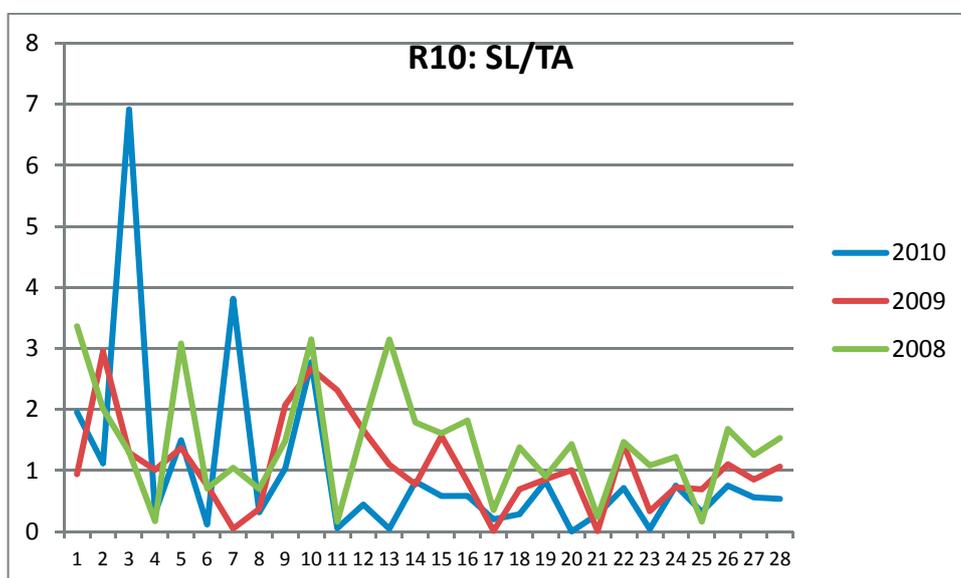
R7: RP/TA, beneficio no distribuido/ activo total, es una medida de la rentabilidad acumulada y puede ser una medida de la edad de la empresa, en aquellas con crecimiento estable de los fondos acumulados. Indica la capacidad de la empresa para protegerse contra los posibles riesgos futuros, por lo tanto, se puede esperar una evolución negativa a medida que la empresa se aproxima a la quiebra, tal como el gráfico confirma.



R8: CF/TD, recursos generados (*Cash Flow*) deuda total, es una medida de liquidez. Las empresas, a medida que se acercan a una situación de quiebra, encuentran más dificultades para obtener recursos líquidos, tal y como se aprecia en el gráfico.



R9: NI/SL, beneficio neto/ ventas, o margen sobre ventas, es el inductor principal de la rentabilidad económica. Un margen mayor indica más eficiencia. Al acercarse al momento de la quiebra las empresas reducen su beneficios pasando a tener pérdidas, de ahí los valores próximos a cero o negativos del ratio. En esta gráfica no se aprecia bien debido a que la empresa 20 presenta un fuerte valor negativo que distorsiona la imagen. Sin embargo, el descenso se aprecia claramente en la segunda parte del gráfico 10 que contiene la evolución de las distancias euclídeas por ratio en el periodo 2008-10.



R10: SL/TA, ventas/ activo total, o rotación, actúa como un multiplicador del margen para aumentar la rentabilidad económica, cuando el margen es positivo. En los casos de proximidad a la quiebra, la reducción de ventas hace que la rotación se aproxime a cero, como confirma la gráfica para la gran mayoría de casos de la muestra.



GOBIERNO
DE ESPAÑA

MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD