

UNIVERSITAT DE BARCELONA

FACULTAT DE CIÈNCIES ECONÒMIQUES I EMPRESARIALS

**LOS MODELOS CONTABLE-FINANCIEROS DE  
PREDICCIÓN DE LA INSOLVENCIA EMPRESARIAL.  
UNA APORTACIÓN Y SU APLICACIÓN A UNA MUESTRA  
DE EMPRESAS DE LOS SECTORES TEXTIL Y  
CONFECCIÓN DE LA PROVINCIA DE BARCELONA  
(1994-1997)**

DEPARTAMENTO DE CONTABILIDAD  
PROGRAMA DE DOCTORADO EN CONTABILIDAD  
Y AUDITORÍA DE CUENTAS, BIENIO 1992-1994

TESIS DOCTORAL PRESENTADA POR  
**ANTONIO SOMOZA LÓPEZ**  
PARA EL ACCESO AL TÍTULO DE DOCTOR EN  
CIENCIAS ECONÓMICAS Y EMPRESARIALES

DIRECTOR: DR. JOSEP VALLVERDÚ CALAFELL  
TUTOR: DR. JAVIER CASTILLO NAVARRO

Barcelona, febrero de 2000

estructura económica con rotación<sup>6</sup>, por último, también encontramos alta correlación entre un ratio de liquidez con estructura económica<sup>7</sup>.

Parece necesario para acabar esta sección referirse a lo que Thomas señala en cuanto a limitaciones del análisis de correlación “La conclusión de lo expuesto es que no debe darse mucha importancia a un coeficiente de correlación muestral significativo cuando viene dado *aisladamente* y que, para deducir vínculos causales hace falta disponer de una cierta cantidad de información adicional” [1986:126].

#### 6.4.2. ANÁLISIS DE COMPONENTES PRINCIPALES

##### 6.4.2.1. Introducción a la técnica

En este apartado vamos a tratar de reducir el número de variables a considerar para aplicar el análisis dicotómico de Beaver a las seleccionadas.

El análisis de componentes principales es una técnica factorial de reducción de la dimensionalidad, partiendo del conjunto de ratios correlacionados que hemos escogido en esta investigación, se trata de encontrar una serie de factores tales que no estén correlacionados entre sí, y que una selección de éstos últimos retenga una gran parte de la información original. Lógicamente, si partiésemos de un conjunto de variables incorrelacionadas desde el principio, dicha técnica no sería necesaria.

Como se puede intuir, hay que elegir un criterio para la selección de dichos factores. En este caso los primeros serán aquellos que tengan una mayor varianza, o dicho de otra manera, tal que la dispersión en ellos se conserve máxima, y se irán seleccionando conforme ésta vaya decreciendo.

---

<sup>5</sup> Ratio 22 y 30 coincidencia en el fondo de maniobra.

<sup>6</sup> Ratio 24 con 39: coincidencia en el activo fijo.

<sup>7</sup> El 9 y el 27.

La formulación de la técnica aparece en el anexo 6.9, aquí solamente planteamos unas breves referencias cuya explicación más detallada puede encontrarse allí y en los manuales citados en la bibliografía.

Partimos de un conjunto de ratios:

$$R_1, R_2, R_3, \dots, R_p$$

y mediante una combinación de éstos con su vector propio, asociado a la vez a su valor propio proyectamos dichos ratios sobre cada componente:

$$Z_{1i} = R_{1i}U_{11} + R_{2i}U_{12} + R_{3i}U_{13} + \dots + R_{1p}U_{1p} \text{ (primera componente)}$$

$$Z_{2i} = R_{1i}U_{21} + R_{2i}U_{22} + R_{3i}U_{23} + \dots + R_{2p}U_{2p} \text{ (segunda componente)}$$

y así sucesivamente hasta la componente j:

$$Z_{jp} = R_{1p}U_{j1} + R_{2p}U_{j2} + R_{3p}U_{j3} + \dots + R_{jp}U_{jp} \text{ (componente j)}$$

El valor propio asociado ( $\lambda$ ) a cada componente representa su varianza o la dispersión retenida por dicho factor. Si las variables estuvieran tipificadas entonces la proporción de variabilidad retenida por una componente h-ésima es:

$$\frac{\lambda_h}{p}$$

dado que en este caso la varianza total del espacio muestral viene dado por el número de componentes.

De la misma manera es interesante saber la correlación que guarda cada variable original con cada componente, que en el caso de que sean tipificadas será:

$$r_{xjv1} = u_{1j} \cdot \sqrt{\lambda_1}$$

En este caso tenemos la correlación de la variable  $x_j$  con el primer componente encontrado, esto es, aquel que posee la mayor varianza.

Para saber el número de componentes a retener no existe una única solución. El programa estadístico utilizado para realizar esta tesis retiene por defecto el número de componentes asociadas a un valor propio sobre las variables estandarizadas superior a 1, o lo que es lo mismo, que el valor propio sea superior a la media aritmética de esos valores. Hay otros criterios, tales como los que suponen un comportamiento normal en los valores propios o, también, basados en el gráfico de sedimentación, sin embargo, se ha optado por seguir el criterio proporcionado por el SPSS.

#### 6.4.2.2. Resultados obtenidos

Tal y como vimos en el apartado anterior, muchas de las variables presentadas tenían elementos comunes, ya sea en el numerador o en el denominador. Es por ello que no resulta extraño el primer resultado del análisis, el programa ha retenido 16 componentes principales de 59 variables que habían quedado después de la primera selección efectuada en el capítulo 5.

Con éstas se retiene el 84,17% de la dispersión total de la muestra y, a partir del componente 37, los factores subsiguientes no añaden mayor información (el 39 tiene un valor propio de 0,122 y los siguientes presentan decimales de dos dígitos, anexo 6.10).

Con 16 componentes la representación de algunas variables originales no es demasiado buena, por esta razón se ha procedido a incrementar el número de componentes a 20 y 25. El motivo por el cual se han elegido dichos componentes es porque con 20 el porcentaje de dispersión retenida alcanza el 90%, y con 25 el 95 %. Si comparamos la representación de esas variables en las tres decisiones (medidas a través de las comunalidades de cada ratio en los tres casos) encontramos lo siguiente (figura 6.8):

	16	20	25	<u>RATIOS</u>
R17	0,581	0,755	0,974	Gastos financieros a fondos ajenos
R35	0,165	0,990	0,999	Ventas a fondo de maniobra
R37	0,535	0,896	0,980	Ventas a fondos propios
R44	0,549	0,760	0,984	Ventas a derechos medios de cobro
R48	0,357	0,942	0,987	Ventas a activo circulante menos exist.
R70	0,566	0,689	0,975	Dotación de reservas a reservas.

Figura 6.8. Representación de ratios en función de componentes.

Como se puede observar al incrementar el número de componentes escogidas, el ratio peor representado en 16 factores, incrementa su representatividad en mucho más del 50% (el 35, del 0,165 a 0,990 en 20) y lo mismo sucede con los restantes ratios.

FACTOR	CON 16 FACTORES	CON 20 FACTORES	CON 25 FACTORES
FACTOR 1	R56, R50, R55, R57, R21, R64, R61, R12*	R56, R50, R55, R57, R21, R64, R61, R12*	R56, R50, R55, R57, R21, R61, R64, R12*.
FACTOR 2	R03, R06, R01, R05, R04, R29	R03, R06, R01, R05, R04, R29	R03, R06, R01, R05, R04, R29.
FACTOR 3	R25, R54, R14, R51, R10.	R25, R54, R14, R10, R51.	R25, R54, R14, R10, R51.
FACTOR 4	R24, R39, R26, R71.	R24, R39, R26, R71.	R24, R39, R26, R71.
FACTOR 5	R72, R45, R27, R02, R12*.	R45, R72, R27, R02, R12*	R62, R53, R52.
FACTOR 6	R62, R53, R52	R62, R53, R52	R72, R45, R27, R12*, R02, R09
FACTOR 7	R28, R23, R31.	R23, R28, R31.	R23, R28, R31.
FACTOR 8	R19, R20, R09, R13.	R19, R20, R09	R20, R19, R09.
FACTOR 9	R34, R32, R36	R34, R36, R32	R36, R34, R32.
FACTOR 10	R63, R66, R67	R63, R66	R63, R66
FACTOR 11	R18, R15	R18, R15	R18, R15.
FACTOR 12	R11, R69	R69, R11	R69, R11.
FACTOR 13	R22, R30	R22, R30	R22, R30.
FACTOR 14	R49, R07	R49, R07	R49, R07.
FACTOR 15	R48.	R65, R67	R67, R64
FACTOR 16	R13, R37*	R13	R13
FACTOR 17		R37, R38	R29, R04.
FACTOR 18		R17, R44	R48
FACTOR 19		R48	R37
FACTOR 20		R35	R44
FACTOR 21			R70
FACTOR 22			R38
FACTOR 23			R17
FACTOR 24			R35
FACTOR 25			R65

Figura 6.9. Ratios agrupados por factores.

El siguiente paso consiste en investigar qué hay en cada uno de los factores, para ello se ha confeccionado la figura 6.9 que representa los ratios agrupados por componentes en los tres casos estudiados, y cuyo coeficiente de correlación con la componente en cuestión sea superior a 0,5.

La metodología empleada para realizar el análisis es partir de la primera columna e ir comparando con las dos restantes. Éstos son los resultados:

- ❑ FACTOR 1: sea cual sea el número de componentes retenido, todos los ratios comparten uno de estos dos elementos, el beneficio antes de impuestos más amortizaciones y provisiones o bien el beneficio antes/después de impuestos. Ello indica que se trata de una componente relacionada con la rentabilidad de la empresa o con los recursos generados por ésta. A destacar que en el caso de utilizar 20 y 25 componentes, aparece un ratio con correlación negativa, se trata del R12 (*deuda total a pasivo total*).
- ❑ FACTOR 2: los ratios tienen una característica común; partidas cuya liquidez es inmediata o casi inmediata. En todos ellos aparecen en el numerador partidas tales como las cuentas financieras o la tesorería.
- ❑ FACTOR 3: en este caso, y comparando los tres grupos, encontramos una coincidencia total en los elementos de cada conjunto y con una característica compartida por todos ellos, el patrimonio neto.
- ❑ FACTOR 4: el común denominador a todas las variables es el inmovilizado ya sea en el numerador o en el denominador.
- ❑ FACTOR 5: en el caso de 25 componentes los ratios que forman parte de esta componente se sitúa en el sexto factor, en vez del quinto. Nos encontramos aquí con un conjunto heterogéneo de ratios: R72 (*reservas a activo total*), R27 (*capital circulante a activo total*), R45 (*fondo de maniobra a ventas*), R02 (*cuentas financieras más valores negociables a pasivo circulante*), R12 (*deuda total a pasivo total*), y R09 (*activo circulante a pasivo circulante*), éste último sólo presente en el grupo de 25 factores. El que tiene mayor correlación con la componente es el ratio 72 en dos de los tres casos analizados (16 y 25 factores), que representa la autofinanciación empresarial; los restantes guardan una asociación estrecha con éste y, además, el ratio 12 tiene signo negativo en su correlación con la componente (endeudamiento), lo cual confirma que se trata de un factor representativo del enriquecimiento de la empresa.

- ❑ FACTOR 6: para 25, estos ratios están en el factor 5. Los ratios que forman parte de esta componente relacionan el beneficio y el beneficio antes de impuestos más amortizaciones y provisiones con el capital. Se trata de la rentabilidad sobre los recursos propios.
- ❑ FACTOR 7: los tres ratios relacionan mercaderías con activo circulante (ratio 23), cuentas a cobrar (ratio 28) y ventas (ratio 31). Podría decirse que es un componente que refleja la gestión de existencias.
- ❑ FACTOR 8: para 20 componentes están en el factor 6. Dos de los tres ratios relacionan deuda a largo plazo con activo (ratio 19) y con capital (ratio 20), sólo el ratio 09 relaciona activo circulante a pasivo circulante. Esta componente podría interpretarse como solvencia a largo plazo de la empresa, ya sea a través del endeudamiento (respecto al activo o al capital) o la capacidad de la empresa para devolver sus obligaciones a corto plazo.
- ❑ FACTOR 9: forman parte tres ratios, el 34 (*activo a ventas*), 32 (*deudores a ventas*) y 36 (*acreedores a ventas*). La característica común que los une es la relación de ventas con otras partidas del balance, por lo cual podemos atribuir a este factor la rotación.
- ❑ FACTOR 10: los ratios comunes en los tres casos relacionan ventas con número de empleados (R63) y gastos de personal (R66), y en el caso de 16 componentes se añade el 67 que relaciona patrimonio neto más exigible a largo plazo a número de empleados. Podemos ver la relación de este factor con la política de personal de la empresa.
- ❑ FACTOR 11: los dos ratios que forman parte de esta componente relacionan una magnitud de beneficio - o de recursos generados - con los intereses pagados por la empresa. Por lo tanto, podríamos decir que se trata de un factor de retribución de los fondos ajenos, o dicho de otra forma, qué parte de los recursos generados se dirigen a retribuir dichos fondos.
- ❑ FACTOR 12: aparecen dos ratios, el 11, que relaciona deudas totales con capitales propios y el 69, reservas con fondos propios. El nexo de unión son aquí los capitales propios ya sea, relacionándolos con las deudas (endeudamiento), o con la autofinanciación. Podríamos designarlo como representativo de esta partida.
- ❑ FACTOR 13: Los dos ratios que forman parte de éste se refieren a, por una parte, la deuda total con el fondo de maniobra (R22) y las existencias con el fondo de maniobra (R30). Por lo tanto, aquí lo común es el fondo de maniobra.
- ❑ FACTOR 14: los ratios 7 (*intereses a efectivo más inversiones financieras temporales*) y 49 (*ventas a efectivo*) tienen como característica común el efectivo de la empresa.

- ❑ FACTOR 15: el ratio 48 relaciona las ventas con el activo circulante menos las existencias, por consiguiente esta agrupación representa la rotación del circulante.
- ❑ FACTOR 16: aparecen disgregados en dos grupos para 20 y 25 factores (componentes 16 y 17 en 20 y 16 y 19 en 25). En los tres casos el ratio predominante es el 13 (*deuda a largo plazo a patrimonio neto*) con lo cual representaría el endeudamiento a largo plazo de la empresa respecto a sus fondos propios. El ratio 37, *ventas a fondos propios*, relaciona qué parte de las ventas se financia con dichos fondos, si bien tiene correlación negativa respecto al componente y con un coeficiente menor que el 13 (0,823 frente a -0,556). La razón estriba en que este ratio proporciona una información que ningún otro lo suministra, como es el endeudamiento a largo plazo.

Para el resto de factores no se ha realizado este análisis por dos razones: la primera, porque las restantes componentes están formadas por un ratio considerado en el resto de factores de una manera implícita o explícita (tal y como puede verificarse en la figura 6.9) y la segunda, está en la propia esencia de la técnica; si lo que queremos es reducir la dimensionalidad del espacio inicial de variables, carece de sentido incrementar el número de componentes a añadir.

Resumiendo, podemos subrayar:

1. Como puede observarse a medida que avanzamos en el número de componentes a considerar, el número de ratios que entran a formar parte en éstas se reducen. Así, a partir del factor 10 para 16 componentes (salvo el factor 12) sólo hay dos ratios en cada nuevo factor. Ello es significativo de la reducción en la aportación de información de cada nueva componente. Los últimos factores no son más que particularizaciones de los primeros: el caso del factor 9, rotaciones, con el factor 15, o el factor 1, rentabilidad de la empresa, con el 6, rentabilidad de los recursos propios, etc.
2. Cada componente se ha interpretado en función de los resultados. Se ha de admitir, no obstante, que el hecho de asignar una peculiaridad a cada una de estas 16 variables no es concluyente, y ello porque dentro de cada una de ellas hemos encontrado algún ratio cuya relación con el resto no era fácilmente interpretable o consistente con el resto. En consecuencia, lejos de ser algo definitivo está sujeto a otras posibles interpretaciones no coincidentes con ésta.

Dado que alguno de estos factores han sido difíciles de interpretar y ante el dilema de utilizar éstos o los ratios más correlacionados, se ha optado por la segunda alternativa. Los motivos que nos han llevado a tal elección son los siguientes:

1. La simplicidad de utilizar un ratio - que mejor o peor- es claramente definido y observable.
2. La interpretación de los resultados que de ello se pueda extraer, dado que en el caso de los factores no es fácilmente atribuible a una característica.

He aquí los ratios más correlacionados con cada uno de los factores (figura 6.10):

FACTORES	RATIOS	FORMULACIÓN
FACTOR 1	R56	(B.A.T.+ amortz.+ provis.) / ventas
FACTOR 2	R03	(Tesorería + val. neg.) / pasivo circulante
FACTOR 3	R25	Inmovilizado / patrimonio neto
FACTOR 4	R24	Activo circulante / activo fijo
FACTOR 5	R72	Reservas / activo total
FACTOR 6	R62	(B.A.T+ amortz.+ provis.) / capital social
FACTOR 7	R28	Cuentas a cobrar / existencias
FACTOR 8	R19	Deuda a L.P. / activo total
FACTOR 9	R34	Activo medio / ventas
FACTOR 10	R63	Ventas / número de empleados medios
FACTOR 11	R18	(B.A.T+ amortz.+ provis.)/ gastos financieros
FACTOR 12	R11	Deudas totales / capitales propios
FACTOR 13	R22	Deudas totales / fondo de maniobra
FACTOR 14	R49	Ventas / tesorería media
FACTOR 15	R48	Ventas / (act. circulante menos existencias) medio
FACTOR 16	R13	Deuda a largo plazo / patrimonio neto

*Figura 6. 10. Ratios seleccionados a través del análisis de componentes principales.*

Es inevitable reconocer que el hecho de optar por utilizar las variables más correlacionadas con estos factores tiene limitaciones importantes:

- a) Estos ratios sean específicos de esta muestra y no generalizables a otras [Chen & Shimerda, 1981].
- b) No hemos conseguido anular la correlación entre variables, lo cual sí hubiera sido posible al utilizar directamente los factores.

Pese a ello, pensamos que las ventajas operativas y de interpretación que brinda contrarrestan las desventajas que ocasiona.

Si comparamos estos resultados con los obtenidos por otras investigaciones que emplean la misma metodología, resulta obligado consultar el artículo de Chen y Shimerda [1981] quienes, al comparar varios estudios<sup>8</sup> que emplearon el análisis factorial de componentes principales, encuentran siete factores comunes a todos ellos, cuales son los que aparecen en la figura 6.11, a los cuales habría que añadir un octavo encontrado por Gombola y Ketz [1983a y b], el *cash-flow*. Posteriormente, Zavgren [1985] confirma la validez de esos factores en un estudio empírico.

COMPONENTE	RATIOS MÁS REPRESENTATIVOS
Rentabilidad sobre la inversión	-B.A.I.T. a Activo total (*) -B.A.I.T. a ventas -B.A.T. a ventas -Inversión neta a patrimonio neto -Inversión neta a activo total
Rotación del capital	-Ventas a activo total (*)
Apalancamiento financiero	-Deuda a largo plazo a patrimonio neto (*) -Deuda a largo plazo a activo total (*) -Pasivo total a activo total -Activo circulante a pasivo circulante (*)
Solvencia a corto plazo	-Activo circulante a activo total -Fondo de maniobra a activo total
Posición de efectivo o liquidez	-Efectivo a activo total (*)
Rotación de existencias	-Activo circulante a ventas
Rotación de deudores	-Ventas a activos líquidos (*)

**Figura 6.11.** Componentes comunes según Chen y Shimerda [1981].

En nuestro caso los 16 factores incluyen estas agrupaciones, lo cual viene confirmado por los ratios dado que, salvo el de rotación de existencias, figuran entre los seleccionados (señalados con un asterisco).

En cuanto al nexo de estas variables con el objetivo del estudio podemos decir que, en principio, y hasta no obtener más evidencia, todas ellas guardan una estrecha relación. Efectivamente, la rentabilidad es el factor clave para entender por qué una empresa fracasa o no, primero por una ralentización en las rotaciones (así como en el margen) que se traduce a corto plazo en un problema de circulante (de existencias) y de liquidez, y a largo plazo en el debilitamiento del patrimonio empresarial. Por otra parte, la falta de

<sup>8</sup> Entre ellos destacan: Pinches *et al.* [1973]; Stevens [1973]; Libby [1975a y b].

generación de recursos tiene una incidencia directa en la política de inversiones, lo cual acaba redundando en el debilitamiento de la posición competitiva de la empresa. Lo acabado de mencionar está sujeto al contraste de hipótesis, pero, ahora por el momento, nos permite enlazar lo ya obtenido con el problema planteado.

## 6. 5. ANÁLISIS DICOTÓMICO DE BEAVER

El artículo publicado por Beaver [1966] marca un hito en lo que actualmente se consideran los estudios empíricos de predicción de la insolvencia, en éste se aplican dos técnicas univariantes a una muestra de empresas. Estas son:

1. El *análisis de perfiles*: inicialmente utilizado en psicología, tal y como indica su nombre, consiste en la representación de cada uno de los ratios para cada estado (solventia e insolvencia) y en cada año previo. Como el propio autor señala, permite analizar la diferencia entre ambos tipos de empresas, pero no proporciona su cuantificación. No tiene en cuenta la dispersión de los ratios, sino únicamente su media.
2. El *test de clasificación dicotómica*: a diferencia del anterior es un test predictivo, basado en la prueba y error, cuya finalidad última es seleccionar qué ratio permite una mejor discriminación entre ambos grupos de empresas. La metodología consiste en su ordenación de forma ascendente, de tal forma que a través de diferentes puntos de corte, se halle aquel que minimice el número de errores, o lo que es lo mismo maximice el porcentaje de predicciones correctas. Esta prueba se aplica para cada año anterior a la insolvencia.

Sólo añadir que la investigación de Beaver, además de ser una de las pioneras en este campo de investigación, consiguió los mejores resultados, no sólo porque para el año previo el porcentaje de error de su mejor ratio *cash-flow a deuda total* fue sólo de un 13%; sino porque lo logró sin los refinamientos estadísticos que más tarde se aplicarían.

Ahora, vamos a seleccionar los años con los que vamos a trabajar, de tal forma que medien nueve meses, al menos, entre la fecha en que se presenta la suspensión de pagos y el último estado financiero analizado [Ohlson, 1980]. Con este fin se adjunta el anexo 6.11 en que, por una parte, se calcula la diferencia temporal mencionada y, por otra, los ejercicios que se han recogido para las empresas suspendidas. Lógicamente, los mismos años han sido seleccionados para realizar el emparejamiento con las empresas no suspendidas.

Como se puede vislumbrar, la cantidad de información disminuye en el momento que incorporamos esta restricción ya que consideramos únicamente los tres ejercicios anteriores (con un promedio de meses desde el último incorporado al estudio y la fecha de presentación de la suspensión de 15 meses). El objetivo último es lograr que el modelo que obtengamos no incorpore información ya sesgada por las dificultades financieras manifiestas.

#### 6. 5. 1. ANÁLISIS DE PERFILES

Los resultados del análisis de perfil que vamos a exponer se han elaborado a partir de los ratios resultantes de ANOVA y los de las componentes principales, se han agrupado atendiendo a los factores expuestos en el capítulo 4<sup>9</sup>:

- LIQUIDEZ Y SOLVENCIA A CORTO PLAZO: en primer lugar, el ratio 03 (análisis de componentes principales) de liquidez o solvencia a corto plazo muestra como las empresas suspendidas, partiendo de una situación mejor, presentan un deterioro que las lleva a una posición muy por debajo de la situación inicial; en contraste con las empresas no suspendidas, que exhiben una tendencia más o menos constante.
  
- ENDEUDAMIENTO O SOLVENCIA A LARGO PLAZO, queda caracterizado por los siguientes ratios:

---

<sup>9</sup> En los ratios 24, 72, 28, 18, 22, 48 y 49 se ha realizado una detección de valores extremos debido a que la forma en que se comportaban las empresas suspendidas no parecía tener una explicación lógica. En todos ellos sólo se ha extraído un valor de los considerados.

1. El ratio 13 o *deuda a largo plazo a patrimonio neto* (análisis de componentes principales), presenta un mayor porcentaje de deuda a largo plazo para las empresas no suspendidas que suspendidas, aunque para el año previo a la insolvencia ambos grupos se acercan mucho. En este caso ha sido necesario borrar un valor extremo.<sup>10</sup>
2. Ratio 18 (análisis de componentes principales) *beneficio antes de impuestos más amortización y provisiones a intereses*, muestra una tendencia alcista para las empresas no suspendidas, lo cual pone de manifiesto una mejora en los recursos generados por este grupo; mientras que para las empresas suspendidas, podemos constatar como la caída de beneficios provoca una caída del ratio. La influencia del valor extremo es fácilmente identificable<sup>11</sup>.
3. Ratio 21 (ANOVA): recursos generados por el beneficio antes de impuestos más la amortización y la provisión con la deuda total, manifiesta un deterioro constante en las empresas suspendidas para el período considerado, en tanto que en el otro grupo crece del tercero al segundo año, y después se mantiene más o menos constante.
4. Ratio 11 (análisis de componentes principales) relaciona las deudas totales con los capitales propios: en los periodos considerados se observa que las empresas no suspendidas tienen un porcentaje superior de éstos sobre aquellos, salvo en el último ejercicio en que se llega a una posición coincidente. Ello puede ser debido, por una parte, a que las empresas no suspendidas realicen un esfuerzo de reducción del endeudamiento y en las fracasadas, en cambio, éste sea menor al partir de una posición menos endeudada. En todo caso, aquí aparece un diferente comportamiento de las instituciones financieras al confiar claramente en aquellas empresas con una buena salud financiera.
5. Ratio 22 (análisis de componentes principales) relaciona la deuda total con el fondo de maniobra en este caso y para las empresas no suspendidas se constata un incremento sostenido. No ocurre lo mismo en las empresas suspendidas, las cuales, y sin sacar ningún valor extremo, muestran una caída y luego una subida no fácilmente justificable, encontramos también dicho comportamiento pese a extraer un valor extremo, si bien el vaivén no es tan pronunciado<sup>12</sup>. A destacar que para el año previo los valores de ambos grupos se aproximen tanto.

---

<sup>10</sup> La empresa CINTEX para el año 1993 dos años previos al fracaso.

<sup>11</sup> La empresa RIUSEC ha sido extraída para el año 1993 que es el último año antes de la suspensión de pagos.

<sup>12</sup> La empresa Hilaturas de Exportación para el año 1994 es el valor extremo, siendo éste el último año previo a la petición de suspensión de pagos.

6. Ratio 19 (análisis de componentes principales) que relaciona la deuda a largo plazo con el activo total, siempre se muestra con unos valores superiores para las empresas suspendidas, lo cual es síntoma inequívoco de un apalancamiento mayor en éstas.
- ESTRUCTURA ECONÓMICA: cuatro ratios son representativos de esta característica:
1. El ratio 28 (análisis de componentes principales): cuentas a cobrar a existencias, muestra una tendencia ligeramente alcista en los tres años anteriores para el grupo de las empresas no suspendidas. El comportamiento del grupo de las suspendidas exhibe un mayor porcentaje de deudores sobre ventas hasta el año anterior a la suspensión en que éste cae a una posición por debajo del otro grupo, también ha sido necesario también eliminar una observación extrema<sup>13</sup>.
  2. El ratio 25 (análisis de componentes principales) que relaciona el inmovilizado con el patrimonio neto (estructura económica), exhibe un importante deterioro en los dos primeros ejercicios y remonta en el último año, aunque dicho deterioro es muy superior en el grupo de las suspendidas.
  3. El ratio 29 (ANOVA) que relaciona activos líquidos con existencias, presenta un deterioro en la liquidez de la empresa insolvente (a través de sus activos líquidos) del tercero al segundo año antes de la fecha de suspensión, y un mejoramiento en el último anterior aunque de manera muy débil, en todo caso, quedando en una posición muy alejada del otro grupo.
  4. El ratio 24 (análisis de componentes principales) relaciona activo circulante a fijo y constatamos una mejor posición solvente por parte de las empresas no suspendidas que suspendidas. Sin eliminar el valor extremo, el año previo a la suspensión ambos grupos coinciden, mas, al extraerlo<sup>14</sup>, el grupo de empresas suspendidas se mantiene finalmente en una posición por debajo del otro grupo.
- ROTACIONES: también aquí sucede algo parecido (ratios 34 y 36 para ANOVA y 34, 48 y 49 para componentes principales), las empresas con dificultades financieras tienen menor rotación en sus activos respecto a las empresas sanas, así como un porcentaje mayor de acreedores sobre sus ventas (ratio 36) y una menor liquidez sobre éstas (ratio 49). A destacar que en los ratios 48 y 49 ha sido necesario borrar valores

---

<sup>13</sup> Es la empresa CINCTEX para el año 1993, dos años previos al fracaso.

<sup>14</sup> Es la empresa SOLFEL, para el año 1993 el año previo a la suspensión.

extremos en cada uno de ellos<sup>15</sup> para poder llegar a tal conclusión, pues tal y como se puede observar en el anexo 6. 12, ambas figuras son muy difíciles de interpretar con todas las observaciones.

- RENTABILIDAD: los ratios utilizados han sido en el análisis ANOVA los ratios 50, 52, 53, 55, 56, 57 y 62 y para las componentes principales el 56 y 62. Los ratios son consistentes con lo que se esperaba en cuanto a que las empresas que posteriormente suspenden pagos presentan una menor rentabilidad los tres años precedentes respecto al otro grupo, aun partiendo en alguno de estos casos de una situación similar. A destacar que en los ratios 50, 55, 57 se yuxtaponen los valores en la situación de partida.
- AUTOFINANCIACIÓN: los dos ratios representativos, el 71 para ANOVA y el 72 para componentes principales, se comportan de forma similar. En cuanto al ratio 71 se constata una menor autofinanciación de las empresas con dificultades financieras respecto al otro grupo; por el contrario, el ratio 72 de reservas a activo total manifiesta una mejor posición de las empresas suspendidas en el tercer año anterior, no obstante, al aproximarse el momento de la insolvencia se da una caída en picado de este grupo. A señalar que si no se borrara el valor extremo<sup>16</sup>, la interpretación sería la contraria.
- PRODUCTIVIDAD: lo mismo puede decirse en cuanto a los ratios de productividad, es decir, el grupo de empresas suspendidas muestra una menor productividad en el periodo considerado, respecto al grupo de no suspendidas (ratios 63 y 64 para ANOVA y 63 para componentes principales).

Resumiendo, podemos decir que las empresas que suspenden pagos se caracterizan en última instancia por una dificultad de solvencia o liquidez, lo cual provoca esa situación extrema. Si bien, este último desencadenante tiene unas causas que se pueden identificar en los años anteriores; su rentabilidad siempre se mantiene por debajo del otro tipo de empresas debido a una menor rotación del activo y a una dificultad de convertirlo en líquido. Este problema afecta progresivamente, por una parte, a la estructura económica de

---

<sup>15</sup> Para el ratio 48 se ha borrado el estado financiero de DIMODA, año 1992 (año previo al fracaso) y para el ratio 49 los estados financieros de DUOTEX, año 1992 (3 años antes de la suspensión) y YOTEX año 1991 (3 años previos).

<sup>16</sup> El valor extremo para el ratio 72 corresponde a la empresa Fabricante de artículos de mercería para el año 1994 o el último previo a la presentación de insolvencia.

la empresa, con una menor inversión en inmovilizado; a un mayor endeudamiento que, casi siempre, y para las empresas pequeñas, es a corto plazo, y a su autofinanciación, lo cual repercute en sus posibilidades de supervivencia futuras. Por último, parece arriesgado pronunciarse sobre la productividad, dado que en los dos ratios utilizados viene medida por las ventas y el beneficio, con lo cual no hay evidencia suficiente para poder atribuir a los empleados cualquier responsabilidad en el fracaso de la empresa.

### 6.5.2. ANÁLISIS DICOTÓMICO

Como anteriormente se ha comentado, consiste en encontrar un punto de corte para cada ratio, tal que minimice el número de errores. Aunque tiene limitaciones importantes es necesario utilizarlo para examinar las variables individualmente y, sobre todo, cuando una variable pueda ser de interés para el investigador [Jones,1987].

Es por ello que se va a aplicar este análisis a aquellas variables que mejor comportamiento han demostrado hasta el momento; es decir, las que superaban la comparación de medias de los ratios y estaban dentro de uno de los 16 factores antes comentados. Siguiendo este criterio tan restrictivo encontramos que dichas variables son:

- Ratio 56: (BAT + amortizaciones y provisiones) a Ventas → RENTABILIDAD
- Ratio 62: (BAT + amortizaciones y provisiones) a Capital social → RENTABILIDAD
- Ratio 34: Activo a ventas → ROTACIÓN
- Ratio 63: Ventas a número de empleados → PRODUCTIVIDAD

Además se va a aplicar el análisis al mejor ratio de Beaver, cash-flow a deuda total en nuestro caso es:

- Ratio 21: (BAT+ amortizaciones y provisiones) a deuda total → ENDEUDAMIENTO C  
SOLVENCIA A LARGO PLAZO.

Los resultados se muestran en la siguiente figura 6.12:

RATIO	SDO.	AÑO -3					AÑO -2					AÑO -1				
		CORTE	✓	X	X I	X II	CORTE	✓	X	X I	X II	CORTE	✓	X	X I	X II
R21	Endeudamiento y solvencia largo plazo	0,117746	61%	39%	20%	80%	0,0383151	68%	32%	65%	35%	0,0690853	73%	27%	38%	62%
R34	Rotación	0,272051	63%	37%	48%	52%	0,5219332	71%	29%	33%	67%	0,889820319	69%	31%	87%	13%
R56	Rentabilidad	0,3224735	59%	41%	86%	74%	0,08580114	69%	31%	12%	88%	0,0245531	71%	29%	50%	50%
R62	Rentabilidad	-0,1390129	59%	41%	86%	14%	0,422282	69%	31%	44%	56%	0,41741011	71%	29%	50%	50%
R63	Productividad	0,3565794	66%	34%	62%	52%	11.215.942	67%	33%	41%	59%	0,1561562	68%	32%	54%	28%
		14.857342	66%	34%	62%	38%						12.249.063				46%

✓ : porcentaje de aciertos.

X : porcentaje de errores totales.

X I: porcentaje de errores tipo I dentro del total de errores: empresas suspendidas consideradas como no suspendidas.

X II: porcentaje de errores tipo II dentro del total de errores: empresas no suspendidas consideradas como suspendidas.

Figura 6.12. Resultados del análisis dicotómico aplicado a los mejores ratios en ANOVA y componentes principales.

Tal y como se desprende de la figura anterior, podemos constatar que el ratio 21 sigue siendo el mejor para separar ambos grupos y en el año previo, con lo cual sería válida la referencia de Beaver a su teoría de la reserva, expuesta en el anexo del capítulo 2 [1966: 86] “*the most crucial factor seems to be the net liquid-asset flow supplied to the reservoir*”.

Otra cuestión importante es que para cada año hay un cambio en el mejor ratio, véase por ejemplo como para el tercer año previo es el 63 (*ventas a número de empleados*<sup>17</sup>) y para el penúltimo año el 34 (*activos a ventas*).

También es interesante investigar qué tipo de errores son los encontrados, en concreto:

- Error tipo I:** *clasificar una empresa suspendida como una empresa sana.*
- Error tipo II:** *clasificar una empresa sana como una empresa suspendida.*

Como queda patente en la extensa literatura sobre el tema, el coste de los errores es difícilmente cuantificable; según el usuario el coste de error tipo I es mucho mayor que el tipo II, puesto que el primero será siempre un coste de oportunidad asociado a la no elección de dicha empresa (ya sea para conceder un préstamo, para invertir en ella, o cualquier otro motivo); en cambio, el tipo I llevará ligado la pérdida de parte o la totalidad de capital de la inversión (entiéndase aquí en un sentido amplio: concesión de préstamos, inversión en capital, etc.) lo cual será un coste fácilmente cuantificable.

Siguiendo este criterio (de minimización del coste I) encontramos que los ratios 21 (20%) y 34 (26%) son los mejores para el tercer año previo; el ratio 56 lo es para el segundo (12%), seguido por el ratio 34 (33%) y, para el último previo, vuelve a serlo el 56 (35%) y el R21 (38%). Asimismo, no se constata un incremento progresivo del error tipo I conforme nos acercamos al momento de la suspensión de pagos, observado por Beaver [1966], sino que es en el segundo año previo, donde se produce el mínimo error I para el ratio 56 (12%), incrementándose para el último previo.

---

<sup>17</sup> Este ratio tiene en el año -2, 16 observaciones perdidas, es decir, para las que no hay un valor, teniendo en cuenta que partíamos de 52 observaciones, ello supone el 30%, lo que permite poner un interrogante acerca de su funcionalidad.

Es importante relacionar estos resultados con los de otras investigaciones anteriores, así el porcentaje de clasificaciones correctas es muy inferior al obtenido por Beaver (en su investigación dicho porcentaje era del 87% para el año previo, si bien aquí sólo llega al 73%), en concreto hay el doble de errores. Otros estudios han utilizado esta metodología: Deakin [1972] para el año previo y con el ratio de Beaver alcanza un porcentaje de errores superior, en concreto el 20%; Casey y Bartzack [1984] para los ratios de cash-flow operativo con la deuda total llegan al 72% de aciertos para el año previo y 75% al relacionar cash-flow con pasivo circulante en ese mismo periodo; por último, entre los trabajos españoles, cabe destacar el de Laffarga *et al.* [1991] con unos porcentajes del 90 al 95% para el año previo en los ratios de rentabilidad y liquidez del sector bancario y los de Lizarraga [1995a, 1997c] quien, al aplicar el mejor ratio de Beaver en el primer trabajo llega al 90% de aciertos para el año previo, y en el segundo selecciona como mejor ratio el capital circulante de las operaciones a deuda total con un porcentaje de aciertos del 87,50%.

De todo ello se desprende que nuestros resultados no son satisfactorios en cuanto a exactitud, si bien están en la misma línea de los demás autores.

En cuanto a las causas de este porcentaje superior de errores, puede tener dos posibles explicaciones. Por una parte, Beaver no consideró lo que más tarde Ohlson [1980] expondría con claridad, la necesidad de escoger un estado anterior a la crisis legal, pero que no la incluyera en sus cifras y, por otra parte, la calidad de la información con la que estamos trabajando (puesta de manifiesto en el anterior capítulo), así como el número de empresas de que disponemos, no permiten albergar grandes esperanzas acerca de la exactitud que de éstos se obtenga.

Nos hemos preguntado si el año del fracaso tenía alguna influencia sobre el porcentaje de aciertos. En concreto, y para el último previo, se ha observado el comportamiento del mejor ratio (ratio 21) separando la muestra según los tres años de suspensión considerados, esto es, empresas que presentaron la suspensión en el año 1994, 1995 y 1996. He aquí los resultados (figura 6.13):

RATIO 21	NÚM. OBS.	✓	✗	✗ I	✗ II
AÑO 1994	18	72%	28%	20%	80%
AÑO 1995	14	71%	29%	75%	25%
AÑO 1996	14	92%	7%	-	100%
TOTALES	46				

*Figura 6. 13. Porcentajes de aciertos y fracasos por año. Tipos de errores.*

Efectivamente, el año del fracaso juega un papel relevante en la discriminación entre ambos grupos, lo cual parece indicar que el ciclo económico subyacente ha de ser considerado como una variable a tener en cuenta para los siguientes capítulos.

Las limitaciones a esta metodología de investigación fueron expuestas en el anexo del capítulo 2 al examinar el artículo de Beaver, si bien vale la pena recordar que provienen de la sensibilidad de la prueba a la magnitud del ratio [Beaver, 1966]; de la no consideración explícita de los diferentes costes asociados a los dos tipos de errores [Netter, 1966]; de llegar a clasificaciones conflictivas entre diversos ratios [Zavgren, 1983] y, finalmente, de la incapacidad para considerar la multidimensionalidad de una firma [Altman, 1968].

La cuestión inmediata es si del análisis univariante se deriva alguna conclusión definitiva acerca del comportamiento de los ratios para el objetivo planteado. Resulta muy difícil poder llegar a manifestar una preferencia por un ratio, dado que ninguno de ellos logra ser el que menos porcentajes de errores consigue para cada año previo, y de forma consecutiva.

Tampoco queda claro poder entrever un comportamiento diferente entre ambos tipos de empresas, efectivamente, si en el tercer año es el ratio de productividad o, en su defecto, el de endeudamiento el que marca la diferencia; en el segundo, es la rotación, y en el año previo, el de solvencia a largo plazo o endeudamiento.

## 6. 6. CONCLUSIONES

En este capítulo se ha dado el primer paso hacia la consecución de un modelo que permita explicar, más que predecir, la crisis financieras de las empresas textiles. Sabemos que los ratios no siguen una distribución normal, lo cual dificulta la tarea a realizar, además, hemos reducido el número de variables a considerar mediante la comparación de medias y el análisis factorial de las componentes principales. Finalmente, éstas han sido examinadas desde dos puntos de vista: el gráfico y el clasificadorio.

Del gráfico, a través del análisis de perfiles, hemos llegado a la conclusión que preveíamos, las empresas suspendidas muestran el comportamiento esperado en las características examinadas: menor rotación, menor rentabilidad, mayor endeudamiento sobre todo a corto plazo.

Desde el punto de vista clasificadorio, los ratios escogidos no permiten alcanzar un punto de corte tan deseable como el de otros estudios, puesto que tenemos hechos probados que así lo corroboran.

Aunque en este momento las dudas que surgen sobrepasan las evidencias conseguidas, aplazamos la indagación de éstas con mayor profundidad a un capítulo posterior.

# CAPÍTULO 7: ANÁLISIS MULTIDIMENSIONAL DE LOS RATIOS

## 7.1. INTRODUCCIÓN

Hemos llegado a uno de los capítulos más esenciales de esta tesis en que se van a exponer los modelos calculados a partir de la consideración simultánea de varias variables, en nuestro caso, los ratios elaborados a partir de los estados contables.

Está dividido en ocho partes: en la primera, se hará una introducción a las dos técnicas estadísticas aplicadas, el análisis discriminante múltiple y el logit; la segunda, mostrará los resultados alcanzados a partir de la información considerada hasta este momento. Seguidamente, se intentará mejorarlos a través de la transformación logarítmica. No cabe duda que sin la aplicación de los modelos a otros datos, su utilidad y significación queda en entredicho; por esta razón, se ha procedido en el siguiente apartado a validar estos resultados con una muestra contemporánea a la utilizada. Finalmente, plantaremos, después de la interpretación de los resultados, una nueva selección de los ratios basada en el análisis de componentes principales.

## 7.2. LOS MODELOS MULTIVARIANTES : EL ANÁLISIS DISCRIMINANTE MÚLTIPLE Y EL LOGIT

### 7. 2. 1. EL ANÁLISIS DISCRIMINANTE MÚLTIPLE

El planteamiento del problema en ambos es cómo obtener una respuesta cualitativa, a partir de un conjunto de variables independientes. En este sentido se adapta perfectamente al objeto de estudio: la respuesta que pretendemos alcanzar es cualitativa y binaria (suspensión *versus* no suspensión de pagos) a través de una información cuantitativa basada en los estados financieros de una muestra de empresas.

El análisis discriminante múltiple fue el que se utilizó con anterioridad en las investigaciones sobre insolvencia empresarial y, por lo tanto, el que trataremos en primer lugar. Como en el caso de las componentes principales, es una técnica factorial de reducción de la dimensionalidad, pero a diferencia de ésta, aquí buscamos un plano (subespacio en general), tal que, proyectadas las empresas en éste, se maximice la distancia entre grupos y se minimice dentro de cada grupo [Sierra, 1986; Uriel, 1995]. Ello se consigue a través del cálculo de los vectores propios asociados a los valores propios de la matriz:

$$[D^{-1}E]$$

donde :  $D^{-1}$ : matriz de variación dentro de los grupos.  
 E: matriz de variación entre grupos.

y el criterio para la obtención de la función discriminante es:

$$\text{Maximizar } \lambda = \frac{\text{Variabilidad entre grupos}}{\text{Variabilidad intragrupos}}$$

A partir de esta expresión se obtienen las funciones discriminantes que son las que generarán las puntuaciones de cada empresa (lo que habitualmente conocemos como *scores*).

$$\begin{aligned} D_I &= u_1 R_{1,0} + u_2 R_{2,0} + u_3 R_{3,0} + \dots + u_p R_{p,0} \\ D_{II} &= u_1 R_{1,1} + u_2 R_{2,1} + u_3 R_{3,1} + \dots + u_p R_{p,1} \end{aligned}$$

De esta forma conseguimos:

- $D_i$  : función discriminante para cada grupo, i: I, II.
- $R_{ij}$ : ratio i para cada una de las empresas de los grupos j, en nuestro caso 2.
- $u_i$  : vectores propios asociados a los valores propios.

Cada función está formada por una combinación lineal de ratios (de 1 a p) y hay una función para cada uno de los dos grupos considerados (0: empresas no suspendidas y

1: empresas suspendidas), a estas funciones así obtenidas se les denomina *Funciones discriminantes de Fisher*. De esta forma se clasifica cada empresa en aquel grupo en que consiga una mayor puntuación. Una alternativa a este planteamiento consiste en la sustracción de las dos funciones, lográndose una única, y asignando cada empresa en función de un punto de corte, como más adelante comprobaremos.

El análisis discriminante puede entenderse como una extensión del análisis de regresión [Peña, 1994b] y el procedimiento de obtención de estas funciones como el cálculo de la *distancia de Mahalanobis* a los centros de ambas poblaciones, es decir:

Asignar la empresa al grupo en que la distancia con su centro sea mínima:

$$M^2_{I,I} = (R_I - \bar{R}_I)' V^{-1} (R_I - \bar{R}_I)$$

donde  $(R_I - \bar{R}_I)$  = distancia del ratio  $i$  al centro del primer grupo (igual para el segundo) .

$V$  = inversa de la matriz de covarianzas entre grupos.

Se ha de reconocer, sin embargo, que el análisis discriminante tiene una serie de supuestos básicos para que su utilización sea totalmente óptima [Uriel, 1995: 279]:

1. Las covarianzas de todos los grupos deben de ser las mismas (hipótesis de homoscedasticidad).
2. Cada uno de los grupos posea una distribución normal multivariante.
3. Independencia de cada una de las muestras multivariantes en cada uno de los grupos.

Si nos atenemos a lo dicho en el capítulo anterior, no se cumplía la segunda en casi ninguno de los ratios para su formulación original.

Esta técnica permite, además, plantear una serie de cuestiones referentes a los resultados:

1. *Hipótesis de igualdad de medias*, o dicho de otra forma, si el análisis es o no significativo. Dicha prueba se realiza a través de la lambda de Wilks ( $\Lambda$ ) que se define como:

$$H_0 : \mu_0 = \mu_1 = \dots = \mu_K$$

$$\Lambda = \frac{|D|}{|T|}$$

donde  $|D|$  y  $|T|$  son los determinantes de la variación dentro de cada grupo y total. Su distribución es compleja, pero se aproxima a una distribución F a través de la  $T^2$  de Hottelling (para dos grupos) o la  $V$  de Barlett que es la utilizada por el programa SPSS [Sierra, 1986]:

$$V = \frac{(J + K)}{2} [N - 1 - \frac{(J + K)}{2}] * \ln \Lambda \sim \chi^2_{J(K-1)}$$

Donde N : número de empresas.

J: número de variables.

K: número de grupos, en nuestro caso 2 (suspensión de pagos *versus* no suspensión).

2. *Importancia de cada una de las funciones discriminantes*, en este caso basta con ponderar el valor propio obtenido asociado a cada función, con la suma de los valores propios. Asimismo una medida de la bondad del ajuste de cada función se representa a través del *coeficiente de correlación canónica*, cuya expresión matemática es la siguiente:

$$\eta = \sqrt{\frac{\lambda}{1 + \lambda}}$$

3. *Valoración de las variables independientes*, o lo que es lo mismo, cuáles de ellas cooperan en mayor medida a la separación de los grupos, se realiza mediante un cálculo iterativo. Se parte de una función con una única variable: aquella que proporcione un mayor ajuste de la función o, tal que la distancia de Mahalanobis sea mínima, y se van incorporando variables independientes paso a paso siguiendo los mismos criterios que la primera. Este procedimiento se denomina *selección hacia delante (stepwise forward)*. También se puede partir en sentido inverso, de una función con todas las variables consideradas ir eliminando aquellas cuya reducción en el ajuste sea mínima *selección hacia detrás (stepwise backward)*. En cada paso puede entrar y salir un ratio, y se fija un nivel de *tolerancia* ligado a la correlación de cada

una de las variables con las existentes en el modelo. Además de los dos señalados, existen una variedad de métodos similares a éstos [Eisenbeis, 1977; Altman *et al.*, 1981]<sup>1</sup>.

4. *Las clasificaciones de las empresas:* si se parte de unas probabilidades de pertenencia a cada grupo del 50% (tal y como las muestras están confeccionadas), la obtención de un punto de corte es inmediata a partir de las puntuaciones de todas las empresas:

$$Z = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$$

$Z > \bar{Z}_1$  asignar la empresa al primer grupo.  
 $Z < \bar{Z}_2$  asignar la empresa al segundo grupo.

En el caso de que pudiéramos disponer de probabilidades previas de los dos grupos y de los costes de clasificación incorrecta, dicho punto de corte se modificaría con:

$$Z = \frac{\bar{Z}_1 + \bar{Z}_2}{2} - \ln \frac{\pi_{II} * \text{Coste (I/II)}}{\pi_I * \text{Coste (II/I)}}$$

donde:

$\pi_i$  = probabilidad previa de pertenencia al grupo  $i$ , en nuestro caso  $i = 0, 1$ .

Coste (I/II) : coste de clasificación incorrecta de pertenencia al grupo I cuando se pertenece al contrario.

$Z > \bar{Z}_1$  asignar la empresa al primer grupo.

$Z < \bar{Z}_2$  asignar la empresa al segundo grupo.

Además, es posible asignar la probabilidad de pertenencia a uno de los grupos, aplicando el teorema de Bayes:

$$P(g/D) = \frac{\pi_g \times \text{Prob (D / g)}}{\sum_{i:1}^G \pi_i \times \text{Prob (D / g)}}$$

<sup>1</sup> Eisenbeis [1977] enumera seis principales: el estadístico F univariante, los coeficientes de la función discriminante ponderados por la dispersión dentro de cada grupo, los métodos de escalonamiento hacia delante y hacia detrás, el borrado condicional y la ponderación de cada coeficiente por la diferencia de las medidas de cada grupo y dividido por las diferencias en las puntuaciones medias.

$P(g / D)$ : probabilidad condicionada de que una empresa pertenezca a un grupo, a partir de las puntuaciones discriminantes.

$\pi_i$ : probabilidades a priori de pertenencia a un grupo.

$P(D / g)$ : probabilidad de la puntuación observada suponiendo la pertenencia a un grupo  $G$ .

### 7.2.2. EL ANÁLISIS LOGIT

Se trata, como en el caso anterior, de construir un modelo de respuesta cualitativa en el cual, la variable dependiente toma dos valores discretos: 0 (ausencia del evento a estudiar) o 1 (existencia de éste). En nuestro caso, como es sabido, se trata de ver a través de los ratios extraídos de las cuentas anuales si una empresa puede ser clasificada como suspendida o no suspendida.

La formulación teórica del modelo implica [Peña, 1994b; Novales, 1993; Clar *et al.* 1997]:

$$P_i = \begin{cases} 0 : \text{la empresa no suspende pagos} \\ 1 : \text{la empresa suspende pagos.} \end{cases}$$

Donde  $P_i$  es la probabilidad estimada por cada empresa.

Este modelo tiene la siguiente formulación:

*Modelo Logit*

$$P_i = F(R_i' \beta) = \frac{e^{R_i' \beta}}{1 + e^{R_i' \beta}}$$

la probabilidad del evento (para nosotros, la insolvencia financiera) es una variable aleatoria que se distribuye según una función logística. Donde:

$$P_i = e^{R_i' \beta} \quad \text{o} \quad \log \frac{P_i}{1 - P_i} = R_i' \beta$$

$P_i$ : probabilidad de suspensión de pagos.

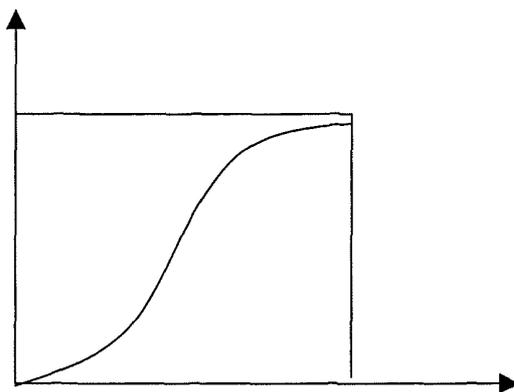
$R_i'$ : matriz del conjunto de características observables, ratios, de las empresas utilizadas en nuestro estudio.

$\beta$ : estimadores de los coeficientes de las variables independientes.

en forma extensa:

$$P_i = e^{Y_j - \beta_0 + R_{1j}\beta_{1j} + R_{2j}\beta_{2j} + R_{3j}\beta_{3j} + \dots + R_{kj}\beta_{kj}}$$

Gráficamente:



*Figura 7.1. Función de distribución logística.*

La principal ventaja de la especificación del modelo logit respecto al análisis discriminante es la ausencia de restricciones (o hipótesis de partida). También, con respecto al modelo de regresión lineal con variable dependiente cualitativa, presenta varias cualidades: la función de distribución está acotada entre 0 y 1, lo cual garantiza que los valores no sean superiores a 1 ni inferiores a cero; el incremento de probabilidad depende de la situación inicial de la empresa, lo cual no ocurre en el modelo de probabilidad lineal y, para acabar, solventa el problema de la heteroscedasticidad en las perturbaciones.

Aunque es una extensión del modelo de regresión, no todo lo asociado con éste es válido aquí porque:

1. La aplicación de los mínimos cuadrados ordinarios no conduce a una estimación eficiente e insesgada de los coeficientes  $\beta$ . En este caso se aplica el método de la función de máxima verosimilitud, siendo ésta la probabilidad conjunta de las N observaciones o:

$$L = \Pi_{y=1} F(R_i' \beta) \Pi_{y=0} (1 - F(R_i' \beta))$$

de tal forma que se procede a realizar una estimación del modelo a través de un algoritmo iterativo. En primer lugar, se obtienen los valores de los parámetros a través de un modelo de probabilidad lineal, ésto es:

$P_i = R_i' \beta$  donde la estimación de  $\beta$  se calcula a través de los mínimos cuadrados generalizados:

$\beta = (R'R)^{-1}(R'V Y)$  siendo  $V$  la matriz de varianzas de los residuos o perturbaciones:

$$V = \begin{bmatrix} (1-P_1)P_1 & 0 & 0 \\ 0 & (1-P_2)P_2 & 0 \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ 0 & 0 & (1-P_n)P_n \end{bmatrix}$$

y, se van repitiendo las estimaciones hasta que la diferencia entre dos consecutivas lleve a unos valores que sean menores a unos fijados a priori, de esta forma se obtienen unos estimadores insesgados y eficientes [Peña, 1994b].

2. Así como en el modelo de regresión normal, los coeficientes de cada ratio nos hubiesen proporcionado información acerca de cómo afecta el hecho de que haya una variación en un ratio a la probabilidad de insolvencia, en este caso, esta interpretación no es posible, dado que dicha probabilidad depende de la situación de partida de la empresa, no es [Clar *et al.*, 1997]:

$$\frac{\delta P_i}{\delta R_j} = \beta_j \text{ como en el modelo lineal.}$$

sino:

$$\frac{\delta P_i}{\delta R_j} = (P_i (1-P_i))\beta_j .$$

Donde  $P_i$  son las probabilidades estimadas de suspensión de pagos. Si es posible determinar entre dos variables explicativas, cuál de ellas tiene un mayor efecto en la probabilidad, ello se realiza de forma marginal:

$$\frac{\delta P_i}{\delta R_i} = (P_i (1-P_i))\beta_i \cdot \beta_i$$

$$\frac{\delta P_i}{\delta R_j} = (P_i (1-P_i))\beta_j \cdot \beta_j$$

de tal manera que si este cociente es superior a uno, es más importante el ratio  $R_i$ , y al revés el  $R_j$ . Para el contraste de hipótesis acerca de la significación individual de cada variable independiente:

$$H_0: \beta_j = 0$$

se puede utilizar el mismo contraste que en el modelo de la regresión.

3. Tampoco se debería utilizar el coeficiente  $R^2$  del modelo de regresión normal, sino una comparación de las funciones de verosimilitud con todos los parámetros o de forma restringida [Novalés, 1993; Clar *et al.* 1997]:

$$\rho = 1 - \frac{\ln L (\beta_{\text{no restringida}})}{\ln L (\beta_{\text{restringida}})} \quad R^2 \text{ de Mc Fadden.}$$

donde:

$\ln L (\beta_{\text{no restringida}}$ ): es el logaritmo neperiano de la función de verosimilitud con todos los parámetros.

$\ln L (\beta_{\text{restringida}}$ ): es la función de verosimilitud sólo con el término independiente.

de tal forma que cuanto más cercano esté a cero, menor ajuste se produce por el modelo.<sup>2</sup> Con esta misma formulación es posible obtener la razón de verosimilitud:

$$RV = \frac{L(\beta_{\text{no restringida}})}{L(\beta_{\text{restringida}})} \sim \text{chi-cuadrado}_{k-1}$$

que permite contrastar la hipótesis:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

Así como se ha determinado una función de distribución logística, es posible suponer que esta variable se distribuya a través de una normal, en cuyo caso sería:

$$P_i \Rightarrow F(\beta'X_i) = \int_{-\infty}^{\beta'X} \frac{1}{\sqrt{2\pi}} * e^{(-t^2/2)} dt.$$

con lo cual estaríamos delante de un modelo PROBIT, que no será objeto de esta tesis.

### 7. 2. 3. LIMITACIONES

En cuanto al análisis discriminante múltiple, las limitaciones provienen de los supuestos básicos en los cuales se fundamenta. Encontramos en Eisenbeis [1977] y, posteriormente en Altman *et al.* [1981], un exhaustivo estudio de cada uno de éstos así como la repercusión en los resultados Aquí, de forma esquemática seguiremos la misma pauta, completándolo con las aportaciones de otros autores:

1. Ausencia de normalidad en las variables independientes: inciden en los test de igualdad de medias basados en *Hotelling* así como en la clasificación obtenida. Una posible solución es la transformación de variables para aproximarse a la

<sup>2</sup> También se podría utilizar la suma de cuadrados de los errores (con o sin ponderación) o el número de

normalidad, aunque ello tenga efectos en el propio análisis, dado que ni las relaciones entre éstas, ni la posición relativa de cada empresa se mantienen constantes.

2. Matrices de covarianzas no iguales en los grupos: afecta a los test de diferencias en las medias de los grupos y a la utilidad de las transformaciones en el espacio reducido, así como al criterio de clasificación, conduciendo a resultados sesgados. Una posibilidad para eludirlo proviene de la utilización de la clasificación cuadrática<sup>3</sup> en vez de la lineal, pero manteniéndose el supuesto de la normalidad (de hecho el modelo ZETA™ de Altman *et al.* [1977a] consigue mejores resultados al aplicarlo). En el caso de que los grupos estén más alejados, las diferencias entre ambas clasificaciones serán menores, pese a que si hay grandes solapaciones funcionará mejor el cuadrático. Collins y Green [1982] consideran que ambos funcionan aproximadamente igual. Marks & Olive [1984] comparan el análisis cuadrático con el lineal y encuentran que el tamaño de la muestra –bajo el supuesto de normalidad- influye en el modelo, de tal forma que si es grande (con valores grandes de la lambda de Wilks) se muestra preferencia por el cuadrático y, por el contrario, si es pequeño (con una lambda de Wilks también pequeña), por el lineal.
3. Grupos no discretos ni diferentes, o dicho de otra forma, una empresa sólo puede formar parte de un grupo (en nuestro caso, suspendidas *versus* no suspendidas), en caso contrario se originan problemas debido a la falta de objetividad en esa segmentación.

Altman plantea en el mismo trabajo tres cuestiones más:

- La significación relativa de las variables independientes: como es sabido, a diferencia del análisis de regresión no se obtiene un único coeficiente en las funciones discriminantes, por lo cual no se puede contrastar su significación. Ello

---

observaciones exactas.

<sup>3</sup> La diferencia entre la clasificación cuadrática y la lineal radica en utilizar logaritmos neperianos tanto en las probabilidades previas como en las propias funciones.

ha originado una variedad de métodos condicionales (como los anteriormente mencionados) para evaluar la contribución de cada ratio<sup>4</sup>.

- La reducción de la dimensionalidad: a través, o bien de la eliminación de variables poco significativas o de funciones con esa misma característica. En el caso de las variables se utilizan métodos condicionales, de tal manera que se simula como quedaría la lambda de Wilks considerando únicamente un subconjunto de variables, y aplicando el correspondiente estadístico F. En cuanto a la eliminación de funciones, se compara el porcentaje que cada valor propio asociado a una función representa sobre el total de valores propios, y muestra la variabilidad capturada por cada dimensión. Tal y como Altman *et al.* [1981] advierten, ello entraña un riesgo: rechazar variables o funciones sin plantearse los efectos que se puedan derivar.
- Problemas en la clasificación, son los más severos, siendo:
  1. *La selección de las probabilidades a priori y sus costes*: en cuanto a las probabilidades a priori, a falta de conocimiento previo, se pueden utilizar las proporciones en la muestra, aunque ello genere el problema de saber si coincide con las de la población, y se plantea si se deben escoger las de un año o las del promedio del período investigado. Respecto a la incorporación de los costes de error, poco se ha realizado en este campo y se hace necesario investigar más sobre este punto, sobre todo porque depende del usuario del modelo.
  2. *La valoración de los errores de clasificación*: los métodos más utilizados consisten en aplicar la función a la propia muestra para validar los resultados; utilizar una muestra externa (*secundaria o de validación*); o el llamado procedimiento *jackknife* es decir, la estimación para todas las empresas menos una, y aplicarla a ésta, de tal forma que al repetirlo para el resto de empresas, obtendremos unos resultados validados. La mayoría de estos métodos presumen una distribución normal de la muestra.  
Joy y Tollefson [1975] proponen dos tipos de validación: una sobre una muestra contemporánea a la que ha servido de base para estimar el modelo y otra posterior para validar los resultados. Igualmente, para la valoración de la efectividad del modelo proponen tres tipos de comparaciones: asignar todas las empresas al grupo más grande; de forma proporcional, o sea, considerando que

---

<sup>4</sup> En Dietrich [1984] se afirma que si la finalidad del investigador es predictiva, entonces la importancia a la adherencia de estos supuestos es menor que si lo que se pretende es identificar los factores y explicar su relación con el fracaso empresarial.

la probabilidad de pertenencia a ambos grupos es igual y, finalmente, la eficiencia condicional, dado que una empresa es suspendida o no, cuál es la probabilidad de que sea clasificada en cada uno de los dos grupos.

La conclusión de todo lo comentado es - como afirma Altman *et al.* [1981, p. 161] - que “*One must simply temper the conclusions reached by recognizing that the empirical results represent approximations that may be significantly biased*”.

La cuestión inmediata es si el modelo de probabilidad condicional, el LOGIT, permite aliviar estas limitaciones y, por lo tanto, llegar a resultados más fiables. Para contestarla, lo primero que debemos recordar es que, al no estar sujeta a ninguno de los supuestos del análisis discriminante, es una aproximación mucho más llamativa desde el punto de vista teórico [Collins & Green, 1982].

No obstante, los resultados no son del todo concluyentes acerca de la superioridad de este tipo de modelos, en concreto, Collins & Green [1982] califican de “modesto” el incremento en la clasificación correcta y apuntan a una reducción del error tipo I al ser utilizados; por otra parte, Jones [1987] afirma que en los rangos medios de probabilidad es en donde se sitúan la mayoría de las firmas y son en éstos, precisamente, en los que las probabilidades son más sensibles a las variables independientes (como puede comprobarse en la figura 7.1, un pequeño cambio en éstas permite cambiar la clasificación de la firma). Además, no hay una interpretación directa del efecto de las variables independientes sobre dicha probabilidad, tampoco es fácil incorporar las probabilidades previas ni los costes de los errores, tal y como sucede en el análisis discriminante.

Jones [1987] y Collins & Green [1982] se manifiestan a favor de este tipo de modelos respecto al análisis discriminante múltiple, siempre y cuando estén debidamente ponderados, es decir -tal y como Zmijewski [1984] planteó su investigación- incorporando un ajuste que tenga en cuenta la proporción de empresas insolventes en la población.

Hamer [1983] realiza una comparación sobre una muestra de empresas de diferentes conjuntos de variables y modelos estadísticos, poniendo especial énfasis en el logit y el análisis discriminante. Los resultados (ya expuestos en el anexo del capítulo 2) no muestran una superioridad de uno de los modelos respecto al otro, especialmente entre el

análisis discriminante lineal y el logit, ni tampoco cuando se relacionan éstos con el análisis cuadrático. Lo [1986] también llega a la conclusión de que no se puede rechazar la hipótesis de que ambos procedimientos son equivalentes. Press & Willson [1978] consideran que si las poblaciones son normales y con idénticas matrices de varianza y covarianza, entonces los estimadores del análisis discriminante son preferibles a los del logit.

Hay otras limitaciones debidas a la propia metodología de la investigación y no a problemas atribuibles a la técnica estadística utilizada:

- a. La inclusión dentro de las empresas no suspendidas de firmas con dificultades empobrece innecesariamente el modelo, aunque su inclusión sobrestimaría el poder discriminante obtenido por éste [Jones, 1987; Mora, 1994b].
- b. La utilización de un modelo para el período analizado o la realización de un modelo para cada año previo, lo cual plantea cuál es el adecuado y, adicionalmente, en que año previo estamos situados [Jones, 1987; Zavgren 1983].
- c. La no consideración de variables potencialmente discriminantes, tales como el tamaño [Jones, 1987].
- d. La utilización de varios años en los cuales el fracaso empresarial ha tenido diferente incidencia, con una repercusión inmediata en las probabilidades previas así como el recurrente uso del 50% de cada tipo de empresas en la muestra, lejos de las probabilidades existentes en la población [Eisenbeis, 1977; Zavgren, 1983]. En la investigación de Zmijewski [1984], los resultados, al comparar un modelo ponderado probit con otro sin ponderación, ponen de manifiesto la existencia de sesgo en los resultados.<sup>5</sup>
- e. Los retrasos en el suministro de la información contable y el hecho de trabajar con datos incompletos son dos fuentes de sesgo en este tipo de investigaciones [Dietrich, 1984]. Para el caso de los datos completos, de nuevo la investigación de Zmijewski [1984] proporciona importantes evidencias. Las empresas con alta probabilidad de quiebra son las que tienen menor probabilidad de presentar datos completos, lo que puede conducir a una baja estimación de la probabilidad de insolvencia técnica, al ser desestimadas.

---

<sup>5</sup> En concreto, este autor llega a la conclusión de que, al incorporar menor proporción de empresas quebradas, resultan menores estimadores de probabilidad de quiebra. Asimismo, las tasas de clasificación correctas se

Todo lo comentado permite hacernos una idea de las limitaciones cuando se realizan investigaciones como la presentada.

### 7.3. APLICACIÓN DE LOS MODELOS A LA MUESTRA OBJETO DE INVESTIGACIÓN

Llegados a este punto debemos aplicar las dos técnicas estadísticas acabadas de mencionar a los datos objeto de nuestro estudio. La metodología que se va a emplear consiste en construir dos modelos: uno, con el análisis discriminante múltiple y otro, paralelamente, con el logit. Ambos serán confeccionados partiendo de la muestra primaria y:

1. Para todas las variables del capítulo 5 (o sea, 59 variables).
2. Para los mejores ratios extraídos del análisis factorial de componentes principales y el ANOVA (4 variables).

Asimismo, se hará una comparación:

- a) Período a período (resultando por lo tanto tres modelos).
- b) Un modelo global para los tres ejercicios previos.

Al igual que hicimos en el análisis dicotómico, también aquí se van a seleccionar los tres ejercicios últimos antes de la suspensión de pagos y dejando como mínimo 9 meses entre el último estado contable depositado en el Registro Mercantil y la fecha de presentación de la suspensión de pagos ante el juzgado. Ello repercute en la composición de la muestra, la cual queda ahora tal y como aparece en el anexo 7.1 (el cual no es más que el anexo 5.1, pero teniendo en cuenta lo comentado en el capítulo 6).

La primera cuestión a tratar es la aparición de valores ausentes (*valores perdidos*) ya que el programa no los considera al calcular la función. Ello, que a primera vista no tendría por qué ocasionar grandes problemas, en realidad, sí que ha sido importante puesto

---

incrementan cuando se utilizan los parámetros menos sesgados y el sesgo aumenta, conforme se incrementan las diferencias entre la probabilidad de la muestra y la población.

que de los 154 casos <sup>6</sup> (en nuestro procedimiento, los 59 ratios de cada empresa para cada año previo) el programa sólo se quedaba con 35, lo cual hacía desconfiar de la representatividad de la función.

Para solventarlo hemos experimentado con la media de cada ratio. Hemos supuesto que en aquellos casos en que no disponíamos del valor, éste no se desviaría de la media. Nos planteamos tres posibilidades en este proceso de sustitución:

- Una media global para todos los ratios.
- Una media para cada grupo o estado; suspensas - no suspensas.
- Una media para cada grupo y período previo: suspensas - no suspensas para el ejercicio -3, -2 y -1.

Ninguna de las tres opciones es realmente segura. Por un lado, si únicamente considerásemos las empresas y los ejercicios con ratios completos se hacía imposible conseguir el propósito inicial de la investigación, un modelo multivariante y representativo de las causas de la insolvencia empresarial. Por otra parte, suponer que el valor de un ratio determinado es igual a su media es mucho suponer, y más cuando la desviación estándar en algunos de ellos es realmente grande. Es por esta razón que hemos procedido a calcular medias cada vez más cercanas a lo que, desde un punto de vista apriorístico, sería lo deseable; el propio valor del que no disponemos<sup>7</sup>.

Seguidamente, hemos realizado una última comparación con una función construida a partir de los cuatro ratios seleccionados como mejores, tanto por el análisis de componentes principales como por el estudio del ANOVA.

Este proceso presenta una ventaja, analizar la robustez de la función para supuestos diferentes, esto es, la sensibilidad de la función al cambio de la composición de la muestra y al número de datos que en ella entran.

---

<sup>6</sup> Fabricante de artículos de Mercería para el año 1990 ha sido eliminado de la base de datos para el programa estadístico ya que, de los 59 ratios considerados, sólo disponíamos de 4.

<sup>7</sup> Ciertamente, una media por grupos será más cercana al valor perdido que una media global. De forma paralela, una media por grupos y año previo es más cercana que la calculada por grupos.

Tanto en el análisis discriminante como en el logit se ha optado por una metodología escalonada o "a pasos" (*stepwise*), ello quiere decir que se parte del ratio que mayor información aporta a la función, y se van incorporando variables en cada etapa. Además, en cada paso, se valora la posibilidad de sacar uno de los que ya están dentro de la función, si no supera un valor crítico. De esta forma nos aseguramos que las funciones resultantes contengan los ratios más significativos desde el punto de vista estadístico.

Por último señalar que en todo el proceso se ha supuesto unas probabilidades previas del 50% de cada grupo. En el punto 7.4. trataremos este punto con mayor detenimiento.

### 7. 3. 1. RESULTADOS DEL ANÁLISIS DISCRIMINANTE MÚLTIPLE

Los resultados más importantes son los siguientes:

F1. Función construida con 35 casos:

<b>F1 = 1,63484 R05 + 1,47142 R07 + 1,09949 R19 + 0,69791 R20 -2,75085 R22 +6,59293 R30 - 1,99027 R45 +5,41709 R48 + 2,24011 R67 +2,97174 R69 + 0,71901 R71</b>					
Resultados de la clasificación:					
	Grupo Actual	Casos	0	1	
Grupo	0	77	63	14	24 casos de empresas no suspendidas
			81,8%	18,2%	11 casos de empresas suspendidas
Grupo	1	77	46	32	Lambda Wilks : 0,063989
			58,4%	41,0%	Correl canónica: 0,9675
Porcentaje de casos correctamente clasificados: 61,69%					

F2. Función construida con 154 casos en los cuales la media total se ha aplicado a los valores perdidos:

<b>F2 = 0,39996 R18 + 0,478 R27 + 0,79705 R36 -0, 53610 R56</b>					
Resultados de la clasificación:					
	Grupo Actual	Casos	0	1	
Grupo	0	77	59	18	77 casos de empresas no suspendidas
			76,6%	23,4%	77 casos de empresas suspendidas
Grupo	1	77	26	51	Lambda Wilks : 0,811972
			33,8%	66,2%	Correl canónica: 0, 4336
Porcentaje de casos correctamente clasificados: 71,43%					

F3. Función construida con 154 casos en los cuales la media de cada grupo ha reemplazado los valores perdidos:

F3 = 0,352 R12 -0,919 R19 + 0,766 R20 -0,495 R36 + 0,416 R56					
Resultados de la clasificación:			Composición muestra:		
Grupo	Grupo Actual	Casos	0	1	77 casos de empresas no suspendidas
	0	77	65	12	77 casos de empresas suspendidas
			84,4%	15,6%	Lambda Wilks : 0,689
Grupo	1	77	26	51	Correl canónica: 0,558
			33,8%	66,2%	
Porcentaje de casos correctamente clasificados:			75,3%		

F4. Función construida con 154 casos en los cuales, la media de cada grupo y para cada ejercicio previo ha reemplazado los valores perdidos:

F4 = 0,36513 R12 -0,92517 R19 + 0,74799 R20 -0,50970 R36 + 0,41758 R56					
Resultados de la clasificación:			Composición muestra:		
Grupo	Grupo Actual	Casos	0	1	77 casos de empresas no suspendidas
	0	77	65	12	77 casos de empresas suspendidas
			84,4%	15,6%	Lambda Wilks : 0,686728
Grupo	1	77	25	52	Correl canónica: 0,5597
			32,5%	67,5%	
Porcentaje de casos correctamente clasificados:			75,97%		

F5. Función de cuatro variables escogidas como las mejores tanto por los componentes principales, como por el análisis de la varianza (aquí el procedimiento no ha sido por etapas, sino que se han incluido todas las variables y considerando 109 casos para los cuales se disponían de todos los ratios):

F5: 0,84377 R56 -0,26208 R34 - 0,5026 R62 + 0, 32348 R63					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	58 casos de empresas no suspendidas
	0	77	61	16	51 casos de empresas suspendidas
			79,2%	20,8%	Lambda Wilks : 0,820819
Grupo	1	77	36	41	Correl canónica: 0,4233
			46,8%	53,2%	
Porcentaje de casos correctamente clasificados:			66,23%		

En cada cuadro aparece la función resultante así como los resultados de la clasificación y la composición de la muestra, a partir de la cual se han obtenido esos parámetros. Cabe resaltar que los valores críticos de entrada para entrar a formar parte de cada una de estas funciones se basan en la minimización de la lambda de Wilks, siendo para la F de entrada 3,84 y para la salida 2,71.

Dos índices adicionales nos ayudan a completar el análisis: la correlación canónica y la lambda de Wilks. La primera nos informa de la parte de la dispersión atribuida a diferencias entre grupos y, por lo tanto, el poder discriminante de cada función (sería equivalente a  $R^2$  en la regresión); la segunda nos proporciona una indicación de cómo la función ha separado ambos grupos y, cuanto más baja sea, mejor, ya que de esta forma la dispersión se debe a las diferencias entre grupos, pero no dentro de cada uno.

Para evaluar lo conseguido partiremos de los resultados de la clasificación y de los dos índices acabados de mencionar, seguidamente, pasaremos a las variables independientes seleccionadas a través del proceso por etapas.

En cuanto a los resultados de la clasificación, la función F4 se muestra más eficaz desde un punto de vista global consiguiendo un porcentaje de clasificación correcta del 76% frente al 61,3% de empresas correctamente clasificadas por la F1. Por consiguiente, si nuestro objetivo es la predicción, nos decantaríamos por la F4.

Si prestamos atención a la lambda de Wilks, todas las funciones son significativas al 1% y, si nos centramos en el coeficiente de correlación canónica, vemos como la primera se muestra muy superior a las siguientes. Sin embargo, tales índices están calculados para los 35 valores únicamente y no sobre los 154 casos en que ha sido validada.

En cuanto a los ratios seleccionados, lo primero que sorprende es su variedad; por ejemplo, la primera función no comparte ninguno con el resto. No obstante, sí hay un ratio que aparece en cuatro de los cinco modelos, el R56 (*beneficio antes de impuestos más amortizaciones y provisiones / importe neto de la cifra de negocios*) que es de rentabilidad. Definitivamente, parece que la rentabilidad juega un papel determinante para la discriminación entre estos dos tipos de empresas ya que también estaba presente en el capítulo 6. Además, el ratio 36 (*acreedores totales / importe neto de la cifra de negocios*) de rotación (aunque una mejor interpretación sería como capacidad de devolución de los fondos ajenos con las ventas generadas, dado que el signo con que aparece no permite una comprensión unívoca) aparece en tres de las cinco funciones.

Por consiguiente, parece que el problema de la insolvencia se relaciona con la rentabilidad o la rotación.

Destacar, por último, que tanto la función F3 como F4 tienen los mismos ratios dado que la diferencia es más bien pequeña. Ambas se basan en una media diferente para cada estado, pero además, la F4 puntualiza el ejercicio previo en que se encuentra esa media.

Valorar la última función resulta difícil ya que no consigue el porcentaje de aciertos que había conseguido la F4 y el ajuste de esta función a la dispersión de la muestra está lejos de ser mejor que en la F1. Por ello, es posible concluir que los ratios, que en el ámbito individual habían conseguido destacar como fuertemente discriminantes o bien, como importantes portadores de contenido informativo; cuando se combinan para obtener un índice, pierden todo protagonismo.

### 7.3.1.1. Comparación para cada año previo

De forma paralela a cómo se ha actuado en el anterior epígrafe, vamos a comparar cuatro modelos: uno para cada año previo, considerando por una parte los valores perdidos y, por otra, sustituyendo éstos por una media para cada ratio. He aquí los resultados:

- Tercer año previo:

F31. Función con valores perdidos validada sobre la muestra:

<b>F31: -9,97255 R01 + 5,44370 R02 -1,25810 R03 +16,67237 R05+1,97996 R07 -0,78030 R09 -1,10693 R10 -6,60861 R18</b>				
Resultados de la clasificación:			Composición de la muestra:	
Grupo	Grupo Actual	Casos	0	1
	0	26	14	12
			53,8%	46,2%
Grupo	1	25	6	19
			24%	76%
Porcentaje de casos correctamente clasificados:			64,71%	
			6 casos de empresas no suspendidas	
			4 casos de empresas suspendidas	
			Lambda Wilks : 0,132192	
			Correl canónica: 0,9316	

Esta función no es estadísticamente significativa al 1%.

F32. Sustituyendo los valores perdidos por la media, el programa no selecciona ninguna función.

F33. Sustituyendo los valores perdidos por la media de los grupos, tenemos la siguiente función:

F33 = -0,849 R52 + 1,182 R70					
Resultados de la clasificación:			Composición muestra:		
Grupo	Actual	Casos	0	1	25 casos de empresas no suspendidas
0		26	23	3	26 casos de empresas suspendidas
			88,5%	11,5%	Lambda Wilks : 0,494
Grupo	1	25	6	19	Correl canónica: 0,756
			24%	76 %	
Porcentaje de casos correctamente clasificados:			82,4%		

F34. Sustituye los valores perdidos por la media de cada grupo y para cada año previo.

F34 = 0,49328 R17 + 0,93533 R19 - 1,15478 R20 - 0,71399 R25 + 0,77536 R70					
Resultados de la clasificación:			Composición muestra:		
Grupo	Actual	Casos	0	1	25 casos de empresas no suspendidas
0		26	22	4	26 casos de empresas suspendidas
			84,6%	15,4%	Lambda Wilks : 0,567365
Grupo	1	25	5	20	Correl canónica: 0,6578
			20%	80 %	
Porcentaje de casos correctamente clasificados:			82,35%		

Tanto F33 como F34 son significativas al 1%, según el estadístico asociado a la lambda de Wilks.

- Segundo año previo:

F21. Función con valores perdidos validada sobre la muestra:

F21: 12,41842 R01 + 1,33238 R02 - 0,23320 R03 + 7,31723 R04 - 13,34947 R05 - 0,30422 R07 - 6,55821 R15					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Actual	Casos	0	1	6 casos de empresas no suspendidas
0		26	21	5	3 casos de empresas suspendidas
			80,8%	19,2 %	Lambda Wilks : 0,440872
Grupo	1	26	16	10	Correl canónica: 0,7477
			61,5%	38,5%	
Porcentaje de casos correctamente clasificados:			59,62%		

La lambda de Wilks no es significativa al 1% y, por lo tanto, no podemos rechazar la hipótesis de igualdad de medias, o lo que es lo mismo, que la función no sea estadísticamente significativa.

F22. Sustituyendo los valores perdidos por la media, calculamos la siguiente función:

F22: R50					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	26 casos de empresas no suspendidas
	0	26	22	4	26 casos de empresas suspendidas
			84,6%	15,4%	Lambda Wilks : 0, 861214
Grupo	1	26	17	9	Correl canónica: 0,3725
			65,4%	34,6%	
Porcentaje de casos correctamente clasificados:			59,62%		

F23. Función resultante de sustituir los valores perdidos por la media de cada estado:

F23: -0,937 R19 +0,875 R20 +0,441 R37 +0,629 R50					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	26 casos de empresas no suspendidas
	0	26	25	1	26 casos de empresas suspendidas
			96,2%	3,8 %	Lambda Wilks : 0,589
Grupo	1	26	5	21	Correl canónica: 0,641
			19,2%	80,8%	
Porcentaje de casos correctamente clasificados:			88,5%		

F24. Función resultante de sustituir los valores perdidos por la media de cada grupo y para cada ejercicio previo:

F24: -0,96180 R19 +0,87246 R20 +0,44598 R37 + 0,62881 R50					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	26 casos de empresas no suspendidas
	0	26	25	1	26 casos de empresas suspendidas
			96,2%	3,8 %	Lambda Wilks : 0,572992
Grupo	1	26	5	21	Correl canónica: 0,6535
			19,2%	80,8%	
Porcentaje de casos correctamente clasificados:			88,46%		

Tanto F22, como F23 y F24 son significativas al 1%

-Último año previo:

F11. Función con valores perdidos validada sobre la muestra:

F11: R44					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	25 casos de empresas no suspendidas
	0	25	19	6	26 casos de empresas suspendidas
			76%	24%	Lambda Wilks : 0, 615788
Grupo	1	26	16	10	Correl canónica: 0,6198
			61,6%	38,5%	
Porcentaje de casos correctamente clasificados:			56,86%		

F12. Sustituyendo los valores perdidos por la media, calculamos la siguiente función:

F12: 0,58179 R38 + R57					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	25 casos de empresas no suspendidas
	0	25	19	6	26 casos de empresas suspendidas
			76%	24%	Lambda Wilks : 0, 707945
Grupo	1	26	7	19	Correl canónica: 0,5404
			27%	73%	
Porcentaje de casos correctamente clasificados:			74,51%		

F13. Sustituyendo los valores perdidos por las medias de cada estado:

F13: 0,579 R38 + 0,983 R57					
Resultados de la clasificación:			Composición de la muestra:		
Grupo	Grupo Actual	Casos	0	1	25 casos de empresas no suspendidas
	0	25	19	6	26 casos de empresas suspendidas
			76%	24%	Lambda Wilks : 0,707
Grupo	1	26	7	19	Correl canónica: 0,541
			26,9%	73,1%	
Porcentaje de casos correctamente clasificados:			74,5%		

F14. Sustituyendo los valores perdidos por las medias de cada estado y para cada año previo:

F14: 0,47225 R02 + 0,66410 R38 +0,95746 R56				
Resultados de la clasificación:			Composición de la muestra:	
Grupo	Grupo Actual	Casos	0	1
	0	25	21	4
			84%	16%
Grupo	1	26	6	20
			23,1%	76,9%
Porcentaje de casos correctamente clasificados:			80,39%	
			25 casos de empresas no suspendidas	
			26 casos de empresas suspendidas	
			Lambda Wilks :0,632882	
			Correl canónica: 0,6059	

Todas las funciones para el último año previo son estadísticamente significativas al 1%, la única que está por encima del valor crítico es F11, pero con 0,0105, por lo tanto, muy cercana al rechazo de la hipótesis nula.

Merece la pena hacer algunos comentarios a estos resultados:

- Aquellas funciones que han sido construidas a partir de considerar los valores perdidos como medias alcanzan porcentajes de éxitos muy superiores a aquellas otras en que se han estimado sin estos casos.
- Si tuviéramos que elegir una función por año previo, desde el punto de vista clasificatorio, seleccionaríamos para el tercero o bien la F33 o la F34; para el segundo, la F23 y, para el primero, la F14. Si consideramos el ajuste de la función a los puntos, es decir la lambda de Wilks y la correlación canónica - como anteriormente comentamos - las basadas en la muestra con valores perdidos, el ajuste siempre es mejor que en el resto por el reducido número de empresas que entran a formar parte.
- Tampoco es posible afirmar que conforme nos acerquemos al momento de la suspensión de pagos las funciones consigan unos mejores resultados. Aquí -y a diferencia de otros estudios en que ello se ha puesto de relieve- si bien en el segundo año alcanza la clasificación más exacta (88% para F23 y F24), cuando pasamos al último previo, desciende el porcentaje de exactitud conseguida (80% para F14). Aplazamos la búsqueda de una explicación hasta más adelante, en concreto, en el momento que los resultados, aplicando el logit, confirmen o desmientan esta afirmación.

- Hay que destacar que las funciones que tienen en cuenta la totalidad de las situaciones con medias totales (F22, F12) muestran unos resultados llamativos; así, en el tercer año el programa no llega a ninguna función; en el segundo, el modelo es un único ratio de rentabilidad (R50) y, en el año previo, selecciona dos ratios para entrar en el modelo (R38 y R50). Las causas provienen del hecho de trabajar con un número muy reducido de datos para cada ejercicio y, paralelamente, con un número muy elevado de valores perdidos sustituidos por la media (la misma para ambos tipos de empresa). Lógicamente, son razones más que suficientes para que el programa en un caso, se vea incapaz de discriminar entre los dos grupos y en otro, escoja aquellos ratios en que haya menos valores sustituidos por dicho subrogado.
- En cuanto a variables comunes, para cada ejercicio previo los ratios que aparecen con más frecuencia son: para el tercer ejercicio previo, no hay ningún ratio compartido por las funciones; para el segundo, el ratio 50 (*beneficio antes de impuestos a ventas*) aparece en tres de las cuatro funciones y el R19 (*deuda a largo plazo a activo*), R20 (*deuda a largo plazo a capital social*) y R37 (*ventas a fondos propios*) en dos de las cuatro funciones; para el anterior a la presentación de la suspensión de pagos, los ratios comunes son el 38 (*ventas a financiación básica*) y el 57 (*beneficios más amortizaciones y provisiones a activo total medio*).

### 7. 3. 2. RESULTADOS DEL MODELO LOGIT

Vamos a proceder usando la misma metodología anterior, es decir, la comparación de un modelo global con otros calculados período a período y, en ambos casos, probando las diferentes posibilidades ofrecidas en cuanto a sustitución de valores perdidos.

Es necesario resaltar que, también aquí, se ha utilizado un proceso por etapas en el cual a cada variable se le imponía un criterio de entrada (basado en el *coeficiente de Rao*<sup>8</sup>) y otro de salida (basado en *Wald*<sup>9</sup>), de tal forma que en el modelo final sólo quedasen

---

<sup>8</sup> Contrasta la hipótesis nula de que el coeficiente asociado a una determinada variable es cero  $H_0: \beta_j = 0$ . Es un criterio de entrada, tal que si la variable a la que hace referencia fuera seleccionada en el siguiente paso, la información que aportaría no sería significativa. En cada etapa será seleccionada aquella que presente el mínimo p-valor asociado y menor al valor crítico prefijado 0,05.

<sup>9</sup> Contrasta la hipótesis  $H_0: \beta_j = 0$  para las variables que ya están dentro del modelo, es un criterio de salida y se interpreta que en el caso de que la variable saliera del modelo, la información perdida no sería

aqueños ratios realmente significativos, en concreto, el valor crítico para seleccionar una variable es 0,05 y para eliminarla es 0,1.

En cada cuadro de los que aparecen a continuación, se muestran los porcentajes de clasificación correcta y la composición de la muestra sobre la cual se ha procedido a calcular la función (en el caso de Z1 no es coincidente ya que ésta se ha aplicado al resto de casos que el programa no aceptaba para construirla). Además, se incluyen tres índices del ajuste del modelo sobre la nube de puntos; el  $-2$  *logaritmo de la verosimilitud*<sup>10</sup>, la *bondad del ajuste*<sup>11</sup> y el *Chi-cuadrado del modelo*<sup>12</sup>. En general, cuanto menores sean sus valores, indicarán que las funciones se ajustan más a la nube de puntos.

Dado que el programa estadístico no estandariza las variables al introducirlas en la función logit (sí en cambio en el discriminante), se ha procedido a tipificarlas, con ello lo que se ha pretendido es evitar el problema de trabajar con valores muy diversos (desde decimales hasta millones); si nos conformamos con dejarlo tal cual, la interpretación puede conducirnos a conclusiones erróneas (anticipemos que, así y todo, nos hemos hallado con algunos ratios cuyo análisis no ha sido fácil), por todo ello, en adelante se muestran dos formulaciones diferentes con los mismos resultados: los ratios originales y estandarizados.

Estos han sido los principales modelos:

- Modelo basado en 35 casos:

Los ratios originales aparecen tal y como se formularon, los tipificados con una Z delante del ratio.

---

significativa. Es similar al contraste  $t$  de la regresión (véase punto 7.2.). En cada etapa será eliminada aquella variable cuyo máximo p-valor asociado al estadístico y mayor que el valor crítico 0,1.

<sup>10</sup> Como ya se comentó en el punto anterior, es la probabilidad conjunta de los resultados observados con los parámetros estimados. Dado que es un valor inferior a 1 se expresa en negativo y es una medida del ajuste del modelo, de tal forma que si el ajuste es perfecto entonces la probabilidad es 1 y el logaritmo es 0 [SPSS,1992].

<sup>11</sup> Es la suma de los cuadrados de los errores y compara las probabilidades observadas con aquellas que predice el modelo y, por lo tanto, divide los residuos entre esas probabilidades. Se distribuye como una Chi-cuadrado con N-k grados de libertad. Su formulación es  $Z^2 = \frac{\sum \text{residuos}^2}{P(1-P)}$  donde P= predicción de valores.

<sup>12</sup> Es una forma alternativa al  $R^2$  de Mc. Fadden o, lo que es lo mismo, la diferencia entre la función de verosimilitud sólo con el término independiente y la función con todos los ratios. Contrasta la hipótesis nula de que todos los coeficientes excepto el constante son cero y se distribuye como una función *Chi-cuadrado*. Es comparable a la F del modelo de regresión.

$Z1 = -5,9495 R25 + 29,8448 R36 - 0,00000017 R67 - 5,0012$ $(4,9578) \quad (19,7832) \quad (0,0000017) \quad (4,4759)$ $Z1 = -297,102 ZR25 + 8,0851 ZR36 - 72,9890 ZR67 + 28,2948$ $(247,5804) \quad (5,3594) \quad (41,3875) \quad (28,0550)$				
casos	0	1	% correcto	24 casos de empresas no suspendidas 11 casos de empresas suspendidas -2 log likelihood: 139,855 Bondad de ajuste: 177,099 Chi-cuadrado: 10,801
0	45	13	77,59	
1	27	24	47,06	
-----				
Total			63,30	

Como se puede apreciar, los ratios muestran unos parámetros demasiado elevados en la forma estandarizada. Los ratios 25, 36 y 67 no son significativos al 5%, pero el modelo globalmente (tanto en una circunstancia como en otra) sí lo es al 1%. Aunque aquí se muestran los resultados validados para el resto de la muestra, el porcentaje de aciertos sobre la submuestra que los generó alcanza el 93,15%.

- Modelo basado en 154 casos, con valores perdidos sustituidos por la media total:

$Z2 = -4,4723 R12 + 2,5558 R19 + 4,0667 R36 - 14,1765 R56 - 0,000000087R64 + 0,0179R66 + 1,2373$ $(1,3805) \quad (1,2682) \quad (1,0795) \quad (3,8180) \quad (0,000000048) \quad (0,0117) \quad (0,8845)$ $Z2 = -0,8415 ZR12 + 0,3984 ZR19 + 1,0945 ZR36 - 1,1193 ZR56 - 0,4847ZR64 + 0,2990ZR66 + 0,2053$ $(0,2597) \quad (0,1977) \quad (0,2905) \quad (0,3014) \quad (0,2670) \quad (0,1962) \quad (0,1980)$						
casos	0	1	% correcto	77 casos de empresas no suspendidas 77 casos de empresas suspendidas -2 log likelihood: 170,965 Bondad de ajuste: 142,421 Chi-cuadrado: 42,524		
0	58	19	75,32			
1	28	49	63,64			
-----						
Total			69,48			

Aquí el modelo es significativo al 1% y, de los ratios seleccionados, sólo el 64 y 66 no son significativos al 5%, aunque están cerca de ese valor.

-Modelo basado en 154 casos, con valores perdidos sustituidos por la media para cada estado:

$Z3 = -0,0536 R13 + 24,0879 R19 - 1,0476 R20 + 3,7150 R36 - 2,5768$ $(0,0259) \quad (4,2312) \quad (0,1949) \quad (1,0704) \quad (0,6362)$ $Z3 = -0,6503 ZR13 + 3,8064 ZR19 - 5,5441 ZR20 + 1,0004 ZR36 - 0,1924$ $(0,3138) \quad (0,6686) \quad (1,0314) \quad (0,2882) \quad (0,2432)$				
casos	0	1	% correcto	77 casos de empresas no suspendidas 77 casos de empresas suspendidas -2 log likelihood: 128,581 Bondad de ajuste: 551,576 Chi-cuadrado: 84,908
0	64	13	83,12	
1	13	50	83,12	
-----				
Total			83,12	

Este modelo es significativo al 1% y todas sus variables son también significativas al 5%.

- Modelo con valores perdidos sustituidos por la media de cada estado y para cada año previo.

$Z4 = 21,3035 R19 - 0,8250 R20 + 3,2094 R36 - 2,4922$ (3,8124) (1,666) (0,9418) (0,5834) $Z4 = 3,3848 ZR19 - 4,3953 ZR20 + 0,8648 ZR36 - 0,0943$ (0,6057) (0,8876) (0,2538) (0,2421)				
casos	0	1	% correcto	77 casos de empresas no suspendidas
				77 casos de empresas suspendidas
0	65	12	84,42	-2 log likelihood: 136,730
1	15	62	80,52	Bondad de ajuste: 417, 725
Total			82,47	Chi-cuadrado: 76,759

Tanto el modelo como los ratios son significativos al 1%.

- Modelo con los ratios resultantes del capítulo 6 (análisis de componentes principales y anova):

$Z5 = 1,4726 R34 - 11,9303 R56 + 0,0084 R62 - 0,0000000072 R63 - 0,7475$ (0,8616) (4,0804) (0,0384) (0,000000117) (0,7885) $Z5 = 0,8991 ZR34 - 0,9481 ZR56 + 0,0554 ZR62 - 0,1638 ZR63 - 0,0949$ (0,5260) (0,3243) (0,2537) (0,2656) (0,2216)				
casos	0	1	% correcto	77 casos de empresas no suspendidas
				77 casos de empresas suspendidas
0	46	12	79,31	-2 log likelihood: 125,999
1	22	29	56,86	Bondad de ajuste: 114,705
Total			68,71	Chi-cuadrado: 24,658

Modelo significativo al 1%, no son significativas ni el R34, ni R62 y R63 (5%).

Comparando los modelos, las variables que forman parte de éstos son diferentes, lo que induce a pensar que estamos delante de funciones muy inestables. Las únicas variables que aparecen con más frecuencia son la 19 (*beneficio antes de impuestos más amortizaciones y provisiones a gastos financieros, de solvencia a largo plazo*) y la 36 (*acreedores totales a cifra neta de negocios, de rotación*). Si comparamos con lo obtenido por los análisis discriminantes, el único ratio común es éste último, aunque es interesante apuntar que en las dos técnicas aparecen cuatro componentes en los ratios: importe neto de

la cifra de negocios, beneficio antes de impuestos más amortizaciones y provisiones (por lo tanto recursos generados) y el exigible total.

Si tuviéramos que escoger uno de los cinco modelos, desde el punto de vista clasificatorio elegiríamos el Z3 (con medias para cada estado) y atendiendo al ajuste, el Z1, si bien, tal y como se comentó para el discriminante, carece de mérito al estar calculado sobre los 35 casos que sirvieron para calcular el modelo. Excluido éste, la función que consigue índices más altos es la última, es decir, aquella que tiene en cuenta los ratios del capítulo 6.

### 7.3.2.1. Análisis logit para cada ejercicio previo

Como en el caso anterior, se ha procedido a analizar los resultados para cada año previo en las dos alternativas posibles: con valores perdidos y sustituyendo éstos por la media del ratio:

#### Tercer año previo

- Modelo con valores perdidos:

$Z_{31} = 1,1335 R_{24} - 8,1179$ $(0,8627) \quad (6,1576)$ $Z_{31} = 35,3627 ZR_{24} + 10,5767$ $(25,7923) \quad (7,6901)$				
casos	0	1	% correcto	6 casos de empresas no suspendidas
-----	-----	-----	-----	4 casos de empresas suspendidas
0	8	18	30,77	-2 log likelihood: 69,803
1	19	16	64	Bondad de ajuste: 50,596
-----	-----	-----	-----	Chi-cuadrado: 0,879
Total			47	

El modelo no es significativo al 1% y en este caso el ratio 24 no es significativo al 5%. Los porcentajes aquí representados se refieren a la validación para el resto de casos. Para la submuestra que la ha originado, el porcentaje asciende al 80% de aciertos totales y tampoco el ratio 24 es significativo al 5%.

- El programa no selecciona ninguna variable para construir un modelo sustituyendo los valores perdidos por la media de cada ratio sin distinción de estado (Z32: inexistente).

- Modelo con valores perdidos sustituidos por la media de cada ratio y distinguiendo cada estado:

$Z33 = 8,2119 R17 - 0,4239 R52 + 2,7254 R70 - 2,2450$ $(4,5712) \quad (0,1556) \quad (0,7442) \quad (0,8628)$ $Z33 = 0,7570 ZR17 - 1,7036 ZR52 + 28,5982 ZR70 - 1,1396$ $(0,4214) \quad (0,6255) \quad (7,8095) \quad (0,4617)$				
casos	0	1	% correcto	26 casos de empresas no suspendidas 25 casos de empresas suspendidas -2 log likelihood: 43,465 Bondad de ajuste: 188,352 Chi-cuadrado: 27,216
0	22	4	84,62	
1	4	21	84	
Total			84,31	

El modelo es significativo al 1% y todos los ratios son significativos al 5% (salvo el ratio 17)<sup>13</sup>.

- Modelo con valores perdidos sustituidos por la media de cada ratio, distinguiendo cada estado y para cada ejercicio previo:

$Z34 = 8,0268 R13 + 168,2912 R19 + 5,5479 R23 - 14,3825 R25 - 2,3524 R31 - 3,2627$ $(4,8514) \quad (88,0965) \quad (3,3787) \quad (8,1856) \quad (1,3036) \quad (3,6449)$ $Z34 = 98,5907 ZR13 + 26,7385 ZR19 + 56,4990 ZR23 - 718,223 ZR25 - 44,9072 ZR31 + 84,6932$ $(59,5883) \quad (13,9970) \quad (34,4086) \quad (408,7682) \quad (24,8857) \quad (47,6143)$				
casos	0	1	% correcto	26 casos de empresas no suspendidas 25 casos de empresas suspendidas -2 log likelihood: 9,825 Bondad de ajuste: 14,333 Chi-cuadrado: 60,857
0	26	0	100	
1	1	24	96	
Total			98,04	

Es un modelo significativo al 1%, no hay ninguna variable significativa al 5%.

*Segundo año previo*

- Modelo con valores perdidos para el segundo año previo:

$Z21 = 13,4855 R36 - 7,8772$ $(9,4637) \quad (5,2270)$ $Z21 = 3,6533 ZR36 - 1,0354$ $(2,5638) \quad (1,0438)$				
casos	0	1	% correcto	6 casos de empresas no suspendidas 3 casos de empresas suspendidas -2 log likelihood: 63,655 Bondad de ajuste: 50,366 Chi-cuadrado: 7,027
0	20	6	76,90	
1	10	15	60	
Total			68,63	

<sup>13</sup> El ratio 70 tiene un exponencial del parámetro asociado que resulta incomprensiblemente alto (supera los 10 dígitos) y mide el incremento que se produce en la probabilidad de la insolvencia al incrementarse en una unidad dicha variable.

El modelo es significativo al 5%, pero no al 1% en la muestra original, y el ratio 36 no es significativo al 5%, tanto para los 9 casos sobre los cuales se calculó la función, como sobre los 41 sobre los cuales se validó, y que son los aquí representados.

- Modelo sin valores perdidos: sustituidos por la media de cada ratio sin distinguir entre estado.

$Z22 = -0,0302 R37 - 16,4311 R50 + 0,4705$ (0,0196) (7,6894) (0,4581)				
$Z22 = -4,3486 ZR37 - 1,3421 ZR50 + 0,3828$ (2,8252) (0,6281) (0,3966)				
casos	0	1	% correcto	26 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 57,504 Bondad de ajuste: 47,022 Chi-cuadrado: 14,853
0	23	3	88,46	
1	10	16	61,54	
Total			75	

El modelo es significativo al 1%, pero el ratio 37 no es significativo al 5%.

- Modelo con valores perdidos sustituidos por la media de cada ratio y para cada estado:

$Z23 = -0,0356 R37 - 31,0010 R50 + 0,1384 R52 - 0,00000017 R67 - 0,4676$ (0,0214) (13,4648) (0,1080) (0,00000007534) (0,6175)				
$Z23 = -5,1357 ZR37 - 2,5336 ZR50 + 0,5564 ZR52 + 6,4497 ZR67 + 1,3198$ (3,0903) (1,1004) (0,4344) (2,7331) (0,6396)				
casos	0	1	% correcto	26 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 49,872 Bondad de ajuste: 41,906 Chi-cuadrado: 22,215
0	21	5	80,77	
1	10	16	61,54	
Total			71,15	

El modelo es significativo al 1% y ni la variables R52 ni R37 son significativas al 5%.

- Modelo con valores perdidos reemplazados por la media de cada ratio, para cada estado y para cada ejercicio previo:

$Z24 = -0,0592 R07 + 35,0228 R19 - 1,5483 R20 - 9,4776 R21 + 0,0398$ $(0,0846) \quad (10,2566) \quad (0,4981) \quad (3,9900) \quad (0,7531)$ $Z24 = -3,2269 ZR07 + 5,5645 ZR19 - 8,2487 ZR20 - 1,7282 ZR21 - 0,5997$ $(4,6147) \quad (1,6296) \quad (2,6539) \quad (0,7275) \quad (0,7738)$				
casos	0	1	% correcto	26 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 31,442 Bondad de ajuste: 31,785 Chi-cuadrado: 40,646
0	22	4	84,62	
1	4	22	84,62	
-----				
Total			84,62	

La función es significativa al 1%, los ratios son todos significativos al 5% excepto el R07.

- Último ejercicio previo

- Modelo con valores perdidos para el último año previo:

$Z11 = -1,2749 R44 + 3,5830$ $(0,7902) \quad (2,8022)$ $Z11 = -25,7617 ZR44 - 5,6228$ $(15,9675) \quad (3,1223)$				
casos	0	1	% correcto	12 casos de empresas no suspendidas 4 casos de empresas suspendidas -2 log likelihood: 67,612 Bondad de ajuste: 49,130 Chi-cuadrado: 1,602
0	17	8	68	
1	9	6	24	
-----				
total			46	

La función es significativa al 1%, pero el único ratio que entra a formar parte de ésta no lo es (al 5%). Al aplicarlo sobre los casos cuyos resultados están aquí expuestos, sucede lo mismo.

- Modelo con valores perdidos sustituidos por la media de los ratios sin distinguir entre estado:

$Z12 = -0,0828 R38 - 13,5245 R57 + 1,3448$ $(0,0453) \quad (4,2074) \quad (0,5667)$ $Z12 = -7,0529 ZR38 - 1,7092 ZR57 - 0,5464$ $(3,8601) \quad (0,5317) \quad (0,4852)$				
casos	0	1	% correcto	25 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 52,707 Bondad de ajuste: 56,384 Chi-cuadrado: 17,975
0	19	6	76	
1	6	20	76,92	
-----				
Total			76,47	

La función es significativa al 1% y de los dos ratios, mientras que el ratio 57 es significativo, el 38 no lo es al 5%.

- Modelo con valores sustituidos por la media de los ratios distinguiendo entre estados:

$Z13 = -0,0838 \text{ R}38 -13,3793 \text{ R}57 +1,3392$ $(0,0451) \quad (4,2099) \quad (0,5638)$ $Z13 = -7,1402 \text{ ZR}38 -1,6908 \text{ ZR}57 -0,5499$ $(3,8431) \quad (0,5320) \quad (0,4822)$				
casos	0	1	% correcto	25 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 52,602 Bondad de ajuste: 55,844 Chi-cuadrado: 18,079
0	19	6	76	
1	6	20	76,92	
Total			76,47	

La función es significativa al 1% y el ratio 38 es el único no significativo al 5%.

- Modelo obtenido al sustituir los valores perdidos por la media de cada ratio, teniendo en cuenta el estado y el año previo en que se encuentra:

$Z14 = -0,1233 \text{ R}38 -25,4812 \text{ R}56 + 1,73$ $(0,0601) \quad (7,9375) \quad (0,7000)$ $Z14 = -10,4983 \text{ ZR}38 -2,0142 \text{ ZR}56 -0,6247$ $(5,1151) \quad (0,6274) \quad (0,5596)$				
casos	0	1	% correcto	25 casos de empresas no suspendidas 26 casos de empresas suspendidas -2 log likelihood: 48,956 Bondad de ajuste: 41,852 Chi-cuadrado: 21,726
0	20	5	80	
1	7	19	73,08	
Total			76,47	

La función es significativa al 1% y los ratios son significativos al 5%.

En primer lugar, se constata como el mejor modelo para el tercero y el segundo año anterior a la insolvencia es aquel en que los valores perdidos se sustituyen por la media de cada estado y en cada ejercicio previo: Z34 (98%) y Z24 (84,62%). No, en cambio, para el último previo, ya que  $Z12 = 76,47\%$ ,  $Z13 = 76,47\%$  y  $Z14 = 76,47\%$ .

Asimismo, y como se había comentado, no hay una tendencia positiva en el porcentaje de éxitos conforme nos acercamos al momento del evento, es más, para el

último ejercicio previo hay una reducción en el número de empresas que se clasifican correctamente, lo cual parece sugerir que los estados contables muestran una estructura más diferenciada en el tercer año antes de la suspensión de pagos que en el último (el mejor modelo está situado en el tercer año previo).

Las razones que pueden inducir a este comportamiento son diversas, únicamente señalemos que, o bien realmente, se produce esta mejoría en la situación empresarial del grupo insolvente como consecuencia de una serie de reformas internas con ese objetivo o bien, las cuentas anuales están debidamente “maquilladas” para salir a la luz pública y su presentación en los juzgados.

Los modelos son muy inestables y así no encontramos variables compartidas entre ellos o que se repitan ejercicio a ejercicio. Incluso, dentro de cada período, al utilizar diferentes sustitutos de los valores perdidos, afecta a la selección de los ratios produciendo un conjunto diferente en cada caso (únicamente en el segundo año previo, en tres de los cuatro modelos se repite una sola variable R37 *importe neto de la cifra de negocios a fondos propios*, de rotación).

Es necesario interpretar estos resultados con cautela por los pocos casos con los que el programa ha trabajado para cada ejercicio previo y la poca significación que de ellos se derivan.

Como último punto a tratar, es necesario hacer una comparación entre el análisis discriminante y el logit.

Mientras que en el análisis discriminante hay una regularidad en las variables utilizadas a lo largo de las funciones, ello no ocurre en el logit, lo cual parece indicar una mayor volatilidad cuando este tipo de formulación elige los ratios.

Por otra parte, en lo que respecta a porcentajes de clasificación correctos, se observa lo siguiente:

- En las funciones globales, la mejor función discriminante (F4) consigue un porcentaje de aciertos del 75,97 %, mientras que su análoga (Z4) la sobrepasa

alcanzando un 82,47% y no es ni siquiera la mejor (la Z3 con 83,12%). Además, comparten tres de los cuatros ratios utilizados (Z4: R19, R20 R36 frente F4: R12, R19, R20, R36 y R56).

- Año tras año:
  - Tercero previo: la mejor función discriminante (F33) consigue un porcentaje de aciertos (82,4%) inferior a su análoga LOGIT (Z33, 84,31%), que no es tampoco la mejor (Z34: 98,06%). En este caso vuelven a compartir variables (Z33: R17, R52, R70 y F33: R52, R70).
  - Segundo año previo: la mejor función discriminante (F23: 88,5%) no es superada por su análoga logit (Z23: 71,15%), siendo la mejor la Z24 (84,62%). En este caso las funciones Z23 y F23 comparten dos variables (R37 y R50), pero no se produce una coincidencia como en los anteriores.
  - Último año previo: la función discriminante supera el logit. En este caso la mejor función es la F14 (80,39%) cuya eficacia supera a su correspondiente logit Z14 (76,47%). Los ratios comunes a ambas son el R38 y R56.

Por lo tanto, podemos concluir:

- El análisis logit se comporta mejor en las funciones globales, mientras que para cada ejercicio previo no se ha demostrado su superioridad, aun cuando hay que reconocer que la diferencia entre los dos procedimientos no es elevada: F23 consigue un 88,5% de aciertos frente a Z24 con un 84,62%, y F14 con un 80,39% frente a Z14 con un 76,47%.
- Ante la presencia de valores perdidos, la mejor opción siempre es sustituirlos por la media de cada grupo para la función global. En cada año previo, el segundo y el tercero muestran una superioridad cuando se sustituyen por una media para cada grupo y en cada ejercicio anterior, lo cual, aunque quizás no requería tan extensa justificación, puesto que se podía intuir, merece ser destacado.
- Ante la disyuntiva entre modelos globales o por cada ejercicio, basta comparar los resultados cambiantes y, a veces, sorprendentes aquí detallados para decantarse por los primeros; pero, si ello no fuera suficiente, es el momento de apuntar lo que sugiere

Zavgren [1983] acerca de la poca utilidad de este tipo de funciones cuando en la práctica se desconoce el año en que se encuentra la empresa a la cual aplicar el modelo.

En lo que a continuación se expone, trabajaremos con el logit sin dejar olvidado el análisis discriminante cuando sea perceptiva su utilización.

## 7.4. VALIDACIÓN EN LA MUESTRA SECUNDARIA

### 7.4.1. MUESTRA SECUNDARIA CONTEMPORÁNEA

Como ya se comentó en su momento, lejos de desaprovechar las empresas con datos incompletos para los tres ejercicios anteriores a la suspensión de pagos, se ha procedido a su incorporación como muestra de validación; aquella que nos permitirá evaluar la eficacia de los modelos antes conseguidos con datos diferentes a aquellos otros que los generaron.

En este apartado procederemos de una forma similar a lo que se hizo en el capítulo 5, aunque de forma más sintética. En primer lugar, en el anexo 7.2 hay el listado de empresas que forman parte de esta muestra secundaria que, tal y como se puede observar, no es más que el anexo 5.1 para aquellas firmas que allí no figuraban sombreadas. También aquí se ha tenido en cuenta los 9 meses que deben transcurrir entre las últimas cuentas anuales presentadas en el Registro Mercantil y la fecha de la suspensión de pagos.

Seguidamente se ha procedido a investigar en qué formato presentaban las cuentas anuales, ello figura en el anexo 7.3 y, como de él se desprende, tanto el número de empresas como sus cuentas anuales han menguado enormemente.

A continuación, presentamos algunas estadísticas que permitan comprender con mayor detalle el tipo de información que poseemos, partiendo del anexo 7.3 al cual se le ha sustraído la empresa Peñatex para el ejercicio 1993 puesto que - tal y como allí se indica- no comienza su actividad hasta enero de 1994.

valores absolutos	Balance			cuenta de pérdidas y ganancias			Memoria			Informe de auditoria		
	1994	1995	1996	1994	1995	1996	1994	1995	1996	1994	1995	1996
año suspensión												
formato normal	4	8	9	4	8	9	3	5	6			
formato abreviado	17	13	6	17	13	6	12	8	4	3	1	6
<b>TOTAL ESTADOS</b>	<b>21</b>	<b>21</b>	<b>15</b>	<b>21</b>	<b>21</b>	<b>15</b>	<b>15</b>	<b>13</b>	<b>10</b>	<b>3</b>	<b>1</b>	<b>6</b>
porcentajes sobre total	Balance			cuenta de pérdidas y ganancias			Memoria			Informe de auditoria		
	1994	1995	1996	1994	1995	1996	1994	1995	1996	1994	1995	1996
año suspensión												
formato normal	0,2	0,38	0,6	0,2	0,38	0,6	0,2	0,38	0,6			
formato abreviado	0,8	0,62	0,4	0,8	0,62	0,4	0,8	0,62	0,4	1	1	1
<b>TOTAL</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

*Figura 7.2. Estadísticas de la muestra secundaria contemporánea en números de estados contables disponibles.*

Una vez más, el formato abreviado vuelve a predominar en las empresas escogidas, lo cual repercute en el propio análisis y, como consecuencia, en la falta de fiabilidad necesaria para evaluar esas cuentas anuales al no estar verificadas por un experto independiente.

En este caso, además, cabe subrayar que una vez eliminadas las empresas sobre las que no poseemos ningún estado contable, el emparejamiento realizado anteriormente no ha sido posible (por ello han sido descartadas de la muestra original), quedando formada por las siguientes sociedades de cada tipo (figura 7.3):

	1994	1995	1996	TOTAL	PORCENTAJE
<b>INSOLVENTES</b>	5	5	0	10	42
<b>SOLVENTES</b>	4	5	5	14	58
<b>TOTAL</b>	<b>9</b>	<b>10</b>	<b>5</b>	<b>24</b>	<b>100</b>

*Figura 7.3. Número de empresas que de cada tipo en la muestra secundaria.*

Como en el capítulo anterior, hemos procedido a confeccionar un anexo con comentarios para estas empresas - tal y como se hizo en el anexo 5.4 - con el objetivo de desarrollar aquellos puntos que creemos puedan ser de interés para una comprensión más adecuada de los resultados. Aparecen en el anexo 7.4 el cual, además de dichos comentarios, incluye las opiniones de los auditores en aquellos casos donde estaban presentes (dado que aquí disponemos de un número demasiado reducido como para dedicarle uno en concreto).

#### 7.4.2. RESULTADOS DE LA APLICACIÓN DE LOS MODELOS SELECCIONADOS EN EL PUNTO 7.3

En este apartado validamos los modelos anteriormente obtenidos con esta muestra contemporánea.

Utilizamos para la validación las funciones Z3 globalmente, y para cada ejercicio previo: Z34, Z24 y Z14. Estas dos últimas, en realidad, no eran las mejores puesto que eran sobrepasadas por el discriminante, pero nos decantamos de forma explícita hacia el logit por las ventajas que éste ofrece.

Clasificamos ahora los datos de la muestra secundaria de la misma manera a cómo lo habíamos realizado en la original, es decir, sustituyendo los valores perdidos por la media para cada estado en el global; y la media en cada estado y año previo, para los correspondientes a cada ejercicio. Medias que, por otra parte, son recalculadas únicamente sobre la muestra secundaria.

En cada uno de los siguientes cuadros hemos incorporado estos resultados:

- *Función original y resultados de la clasificación en la muestra de validación contemporánea.* Al lado de los porcentajes figuran el número de casos que componen la muestra de validación y los dos índices del ajuste de la función en la muestra primaria (éstos son los mismos a los originales al no modificar la función).

- Función obtenida con las mismas variables que en la muestra original, pero sobre una muestra que incorpore tanto la original como la de validación y resultados de clasificación (las hemos representado con una comilla). Al lado de éstos figuran el número de casos que componen la nueva muestra y los índices de ajuste de la función en la nube de puntos.

**Modelo global para los tres ejercicios previos**

$Z3 = -0,0536 R13 + 24,0879 R19 - 1,0476 R20 + 3,7150 R36 - 2,5768$ $Z3 = -0,6503 ZR13 + 3,8064 ZR19 - 5,5441 ZR20 + 1,0004 ZR36 - 0,1924$				
casos	0	1	% correcto	34 casos de empresas no suspendidas 23 casos de empresas suspendidas -2 log likelihood: 128,581 Bondad de ajuste: 551, 576 Chi-cuadrado: 84,908
0	28	6	82,35	
1	20	3	13,04	
Total			54,39	
$Z3' = 0,0082 R13 + 7,9925 R19 - 0,2774 R20 + 0,0431 R36 - 0,2219$ (0,0161) (2,0781) (0,0811) (0,1631) (0,2609) $Z3' = -0,0870 ZR13 + 1,3092 ZR19 - 8,8012 ZR20 + 0,0519 ZR36 - 1,5375$ (0,1706) (0,3404) (2,5713) (0,1965) (0,5380)				
casos	0	1	% correcto	111 casos de empresas no suspendidas 110 casos de empresas suspendidas -2 log likelihood: 235, 098 Bondad de ajuste: 198, 206 Chi-cuadrado: 56,836
0	87	24	78,38	
1	35	65	65,00	
Total			72,04	

La función Z3' es significativa al 1% y las únicas variables no significativas al 5% son el ratio 13 y 36.

Para cada año previo hemos procedido a realizar los mismos pasos que en Z3 y éstos son los resultados:

**Tercer ejercicio previo**

$Z_{34} = 8,0268 R_{13} + 168,2912 R_{19} + 5,5479 R_{23} - 14,3825 R_{25} - 2,3524 R_{31} - 3,2627$ $Z_{34} = 98,5907 ZR_{13} + 26,7385 ZR_{19} + 56,4990 ZR_{23} - 718,223 ZR_{25} - 44,9072 ZR_{31} + 84,6932$				
casos	0	1	% correcto	11 casos de empresas no suspendidas 8 casos de empresas suspendidas -2 log likelihood: 9,825 Bondad de ajuste: 14,333 Chi-cuadrado: 60,857
0	2	9	18,18	
1	2	6	75	
Total			42, 11	
$Z_{34}' = 0,3409 R_{13} - 0,8842 R_{19} + 0,2631 R_{23} - 0,1560 R_{25} - 0,0978 R_{31} - 0,4502$ (0,1087) (2,0830) (0,1976) (0,0542) (0,0717) (0,5812) $Z_{34}' = 3,6724 ZR_{13} - 0,1504 ZR_{19} + 2,5135 ZR_{23} - 6,6972 ZR_{25} - 1,6647 ZR_{31} + 0,4338$ (1,1705) (0,3544) (1,8875) (2,3294) (1,2062) (0,3890)				
casos	0	1	% correcto	37 casos de empresas no suspendidas 33 casos de empresas suspendidas -2 log likelihood: 75,934 Bondad de ajuste: 95,259 Chi-cuadrado: 20,878
0	24	13	64,86	
1	9	24	72,73	
Total			68,57	

La función  $Z_{34}'$  es significativa al 1% pero las variables  $R_{19}$ ,  $R_{23}$  y  $R_{31}$  no son significativas al 5%.

**Segundo ejercicio previo**

$Z_{24} = -0,0592 R_{07} + 35,0228 R_{19} - 1,5483 R_{20} - 9,4776 R_{21} + 0,0398$ $Z_{24} = -3,2269 ZR_{07} + 5,5645 ZR_{19} - 8,2487 ZR_{20} - 1,7282 ZR_{21} - 0,5997$				
casos	0	1	% correcto	11 casos de empresas no suspendidas 8 casos de empresas suspendidas -2 log likelihood: 31,442 Bondad de ajuste: 31,785 Chi-cuadrado: 40,646
0	10	1	90,91	
1	8	0	0	
Total			52, 63	
$Z_{24}' = -0,0142 R_{07} + 12,9713 R_{19} - 0,6797 R_{20} - 5,8184 R_{21} + 0,7100$ (0,0271) (4,7618) (0,2433) (2,4437) (0,4949) $Z_{24}' = -0,6660 ZR_{07} + 2,2067 ZR_{19} - 23,3694 ZR_{20} - 1,0313 ZR_{21} - 4,6985$ (1,2653) (0,8101) (8,3653) (0,4332) (1,7799)				
casos	0	1	% correcto	37 casos de empresas no suspendidas 34 casos de empresas suspendidas -2 log likelihood: 65,171 Bondad de ajuste: 58,809 Chi-cuadrado: 33,129
0	31	6	83,78	
1	10	24	70,59	
Total			77,46	

El modelo Z24' es significativo al 1%, los ratios son todos significativos excepto el R07 al 5%.

### Último ejercicio previo

$Z14 = -0,1233 R38 - 25,481 R56 + 1,7300$ $Z14 = -10,4983 ZR38 - 2,0142 ZR56 - 0,6247$				
casos	0	1	% correcto	12 casos de empresas no suspendidas 7 casos de empresas suspendidas -2 log likelihood: 48,956 Bondad de ajuste: 41,852 Chi-cuadrado: 21,726
0	3	9	25,00	
1	2	5	71,43	
Total			42,11	
$Z14' = -0,0186 R38 - 0,1145 R56 + 0,0237$ (0,0289) (0,8103) (0,3235) $Z14' = -1,3586 ZR38 - 0,0215 ZR56 - 0,2157$ (2,1070) (0,1519) (0,2853)				
casos	0	1	% correcto	37 casos de empresas no suspendidas 33 casos de empresas suspendidas -2 log likelihood: 96,330 Bondad de ajuste: 70,052 Chi-cuadrado: 0,482
0	36	1	97,30	
1	29	4	12,12	
Total			57,14	

El modelo Z14' no es significativo al 5% y tampoco ninguna de las variables lo es.

Lo presentado no deja ningún margen de dudas: los modelos pierden gran parte de la efectividad demostrada en su cálculo al ser validados sobre una muestra contemporánea. Como se desprende de los resultados precedentes, no hay ninguna de las anteriores funciones que, aplicada sobre la muestra secundaria, exhiba un porcentaje de aciertos similar a la conseguida en la primaria. La cuestión es, sin embargo, un poco peor de lo que en principio se esperaba, puesto que los modelos, al ser utilizados para clasificar los casos en la muestra de validación, rondan unos porcentajes que plantean serias dudas acerca de su utilidad; sirva de ejemplo la función Z14, en la cual las clasificaciones conseguidas son del 42% y al ser comparado con el azar (50%), permite rechazar el modelo como útil.

No obstante, frente a esta imagen tan negativa de lo obtenido, recalculamos cada función con la muestra secundaria (funciones Z'). Efectivamente, si nos centramos en los porcentajes de clasificación correcta con la información proporcionada por la muestra de

validación los resultados mejoran, aun cuando no lleguen a los porcentajes de exactitud que habíamos conseguido anteriormente; así, por ejemplo, todos ellos rondan el 70%, excepto en el caso de la función Z14.

La cuestión más inmediata es si consiguen mayor exactitud identificando la insolvencia. No hay una respuesta contundente dado que, si bien parece que así sucede, cuando se utiliza la muestra original para clasificar la secundaria en el último ejercicio previo (véase la Z14), ello no ocurre ni en el modelo global Z3, ni en el de dos años anteriores (Z24). Por último, cuando se calcula de nuevo cada modelo, tampoco parece haber una tendencia seguida por todas las funciones en orden a discriminar mejor un estado que otro (mientras que la Z3, Z24 y Z14 consiguen mejores resultados para las empresas solventes; la Z34 se decanta por las insolventes).

En cuanto al ajuste conseguido, los dos índices ofrecidos ponen de manifiesto lo que de hecho el porcentaje de aciertos muestra: la pérdida de representatividad de las empresas en la función. Pensamos que no es necesario insistir, puesto que los resultados son irrefutables.

Cabe plantearse si los ratios siguen siendo estadísticamente significativos al incorporar la muestra secundaria contemporánea. Para ello reunimos en la figura 7.4 de la siguiente página los resultados de contrastar la hipótesis nula del parámetro asociado a una variable independiente a través del estadístico de *Wald* o, dicho de otra manera, la comparación con un determinado valor crítico prefijado (es decir, 0,05) en las funciones con los ratios recalculados para las dos submuestras (o sea, las funciones con una comilla).

Se observa un incremento en los ratios no significativos cuando las funciones se validan en una posterior. No obstante, encontramos una excepción, la Z34, puesto que en la muestra primaria ninguno de los ratios es significativo al 5%; pero, al incorporar la secundaria, el ratio 25 y el 13 aparecen como significativos. A destacar también que la función Z14' está compuesta de dos ratios y los dos no son significativos, lo cual provoca que ésta no sea globalmente significativa, a pesar de que al trabajar únicamente sobre la primaria sí lo era.

FUNCIÓN	VARIABLES				
	Ratio	Ratio	Ratio	ratio	ratio
Z3'	R13	R36			
Z3					
Z34'	R19	R23	R31		
Z34	R19	R23	R31	R25	R13
Z24'	R07				
Z24	R07				
Z14'	R38	R56			
Z14					

*Figura 7.4. Ratios no significativos por estadístico Wald.*

Ello ha sido motivo suficiente para plantearse cómo han influido los valores perdidos en la representatividad de las variables. Para ello hemos reunido los ratios que presentaban un mayor número de valores perdidos y hemos confeccionado la siguiente figura 7.5. Estos han sido los resultados:

RATIOS	VALORES PERDIDOS	TOTAL	PORCENTAJE
R13, R19, R20	29	57	50,8
R63, R64, R65, R67	24	57	42,1
R70	41	51	80,3

*Figura 7.5. Ratios con mayores valores perdidos en la muestra secundaria.*

Nos podríamos preguntar si ello no sucedía en la muestra primaria, pasemos a comprobarlo (figura 7.6):

RATIOS	VALORES PERDIDOS	TOTAL	PORCENTAJE
R13, R19, R20	62	154	40%
R63	45	154	29%
R64, R65, R67	43	154	28%
R70	81	154	52%

*Figura 7.6. Ratios con mayores valores perdidos en la muestra primaria.*

Como se desprende, el número de valores perdidos en la muestra primaria era inferior a la muestra secundaria, lo cual ha repercutido en los resultados antes indicados.

De todo lo comentado, podemos concluir que la incorporación de datos incompletos afecta de forma significativa y negativa a los modelos resultantes del punto anterior, lo cual pone de manifiesto que este tipo de información debe manejarse con precaución, así como los resultados que de ella se infieran.

Como siguiente asunto a tratar, hemos querido probar la sensibilidad de la técnica utilizada. Se ha procedido de la siguiente forma, el mejor modelo para el último año previo no era una función logit, sino discriminante; pues bien, comprobemos en qué medida se ha visto afectada dicha función ante la incorporación de la muestra secundaria y cómo la ha clasificado:

F14: 0,472 R02 + 0,664 R38 + 0,957 R56				
casos	0	1	% correcto	12 casos de empresas no suspendidas 7 casos de empresas suspendidas Lambda Wilks: 0,632882 Correlac. Canónica: 0,6059
0	7	5	58,32	
1	2	5	71,40	
Total			63,20	
F14': 0,998R02 + 0,435 R38 -0,117 R56				
casos	0	1	% correcto	37 casos de empresas no suspendidas 33 casos de empresas suspendidas Lambda Wilks: 0,902 Correlación canónica: 0,313
0	20	17	54,1	
1	14	19	57,6	
Total			55,7	

La función F14 es significativa por la lambda de Wilks al 5%, no en cambio la F14'.

Aquí no incorporamos la función tipificada, puesto que el propio programa, al realizar los cálculos, estandariza las variables.

Al igual que sucedía en el análisis logit, también el discriminante muestra unos peores resultados para la muestra de validación, tanto en F14 como en F14', por lo tanto, todo ello nos está confirmando las anteriores aseveraciones.

Sin perjuicio de lo acabado de comentar, el análisis discriminante permite llegar a una mejor clasificación cuando la función primaria se utiliza para la clasificación del segundo grupo, lo cual no sucedía en el logit.

Para acabar nos hemos planteado dos cuestiones adicionales:

1. Investigar las variables que escoge cada función con la muestra primaria y de validación (serán representadas mediante el nombre de la función y comillas).
2. Si cambiamos el 50% como criterio de clasificación e incorporamos la composición de la muestra, indagamos si con ello se consiguen mejores resultados.

En cuanto al primer punto, las funciones se muestran en la figura 7.7.

Si nos centramos en los ratios seleccionados y los cotejamos con aquellos obtenidos sobre la muestra primaria, llegamos a las siguientes conclusiones:

- a) Para la función global los ratios comunes son R19 y el R20 y, mientras que en la primera se utilizaban 4 variables independientes, aquí se utilizan 6. El valor de la constante tampoco es parecido, aunque negativa como en el anterior.

- b) Para las funciones por ejercicio previo sólo la Z24" comparte ratios con su homónima primaria (el ratio R19 y R20), asimismo el término independiente es inferior.

En cuanto a los porcentajes de clasificación y el ajuste conseguido:

- a) Para la función global los porcentajes conseguidos por la primera (Z3: 83,12%) son superiores a los conseguidos por ésta (Z3": 75,36%). Así, el logaritmo de verosimilitud se mantiene en unas cifras superiores a la función calculada sobre la muestra primaria, mientras que la bondad del ajuste era superior en la muestra primaria.
- b) Para las funciones por ejercicio previo, consiguen unos porcentajes de aciertos similares a los conseguidos por la calculada anteriormente, excepto para Z34 que es ampliamente superada por la calculada sobre la muestra primaria (Z34: 98,04%, frente a Z34": 75,71%). Hay una leve mejoría cuando se utilizan las funciones con ratios seleccionados para las dos muestras en: Z24:84,62% vs. Z24": 87,32% y Z14: 76,47% vs. Z14"; 77,14%.

Si realizamos la comparación con las funciones recalculadas y los ratios originales (aquellas que llevan una comilla al lado de la función) podemos verificar como en todos los casos aquellas en las que se han seleccionado los ratios "a medida" ganan en exactitud a aquellas otras en que ello no sucede.

En cuanto a las probabilidades previas, el programa utilizado nos permite incorporarlas en el caso del análisis discriminante múltiple, pero no en el caso del logit.

Para solucionarlo hemos cambiado el punto de corte. Por defecto, el logit lo sitúa en la probabilidad 0,5, por lo tanto, implícitamente, la composición de la muestra es coincidente con este valor. Lo que hemos hecho es analizar la clasificación que obtendríamos con los modelos primarios y con éstos últimos, si la probabilidad previa coincidiera con la composición de las dos muestras.

Los resultados son casi coincidentes con los anteriores, por lo que aquí únicamente se exponen los resultados de la función Z3 y la discriminante F14.

**FUNCIÓN LOGIT Z3'**

<u>punto corte: 0,52 (composición)</u>				<u>punto corte 0,5</u>			
	solv	insolv	porcentaje		solv	insolv	porcentaje
solvente	91	20	81,98%	solvente	87	24	78,38%
insolvente	38	62	62%	insolvente	35	65	65 %
TOTAL			72,04				72,51

**FUNCIÓN DISCRIMINANTE F14'**

<u>punto corte: 0,52 (composición)</u>				<u>punto corte 0,5</u>			
	solv	insolv	porcentaje		solv	insolv	porcentaje
solvente	20	17	54,1%	solvente	26	11	70,3%
insolvente	14	19	57,6%	insolvente	20	13	39,4%
TOTAL			55,7%				55,7%

Es obvio que cualquier intento de mejorar los resultados a través de incorporar las probabilidades previas basadas en la composición de la muestra es baladí, ya sea porque dicha composición está muy cerca del 50%, o porque al incrementar los aciertos en un estado, los reducimos en el otro; es un juego de suma cero.

Funciones	COMPOSICIÓN			PORCENTAJES DE ACIERTOS			MEDIDAS DE AJUSTE		
	Insolvente	solventes	solventes	solventes	Insolvente	totales	A	B	C
<b>Z3<sup>o</sup></b> : 0,0191 R18 + 8,9418 R19 -0,2931 R20 - 14,1871 R21 + (0,0083) (2,3238) (0,0893) (3,3998) 15,1997 R55 + 4,6920 R72 - 0,0449 (4,3917) (1,5652) (0,3376)	111	100		76,58	74	75,36	206,904	247,941	85,031
<b>Z3<sup>o</sup></b> : 0,8660 ZR18 + 1,4647 ZR19 - 9,2989 ZR20 - 2,5147 ZR21 + (0,3756) (0,3807) (2,8323) (0,6026) 1,8145 ZR55 + 0,7040 ZR72 - 1,5370 (0,5243) (0,2349) (0,5808)									
<b>Z34<sup>o</sup></b> : -0,1005 R62+ 0,00000078 R65 +1,3450 R70 -2,9147 (0,0460) (0,00000031) (0,4177) (0,9561)	37	33		81,08	69,70	75,71	77,556	71,479	19,256
<b>Z34<sup>o</sup></b> : -0,9177 ZR62 + 0,7343 ZR65 + 12,2806 ZR70 - 0,7730 (0,4198) (0,2920) (3,8141) (0,3439)									
<b>Z24<sup>o</sup></b> : 0,0597 R15 +14,6457 R19 -0,5555R20 -26,4377 R56 (0,0744) (6,3217) (0,3188) (10,4573) -1,1986 R70 +1,5283. (0,5060) (0,6169)									
<b>Z24<sup>o</sup></b> : 1,3698 ZR15 + 2,4916 ZR19 -19,0966 ZR20 - 4,9573 ZR56 - (1,7073) (1,0755) (10,9593) (1,9609) 10,9438 ZR70 - 2,0887 (4,6204) (2,3447)	37	34		83,78	91,18	87,32	42,422	42,104	55,879

**Figura 7.7.** Selección de los ratios a través de la consideración conjunta de las dos muestras para su cálculo.

Funciones	COMPOSICIÓN			PORCENTAJES DE ACIERTOS			MEDIDAS DE AJUSTE		
	Insolvente	solventes		Insolvente	solventes		A	B	C
		solvente	totales		solvente	totales			
Z14": $-1,4209 R02 - 0,00000069 R64 + 0,6862$ (0,7018) (0,000000227) (0,5064)	37	33	86,49	66,67	77,14	74,875	106,445	21,937	
Z14": $-0,7596 ZR02 - 3,2882 ZR64 - 0,4981$ (0,3784) (1,1057) (0,3018)	37	33	78,4	75,8	77,14	0,762751	0,4871	-----	
F14": $0,52962 R01 + 0,79478 R64$									

Z3": es significativo al 1% y todos los ratios son significativos al 5%.

Z34": es significativo al 1% y todos los ratios son significativos al 5%.

Z24": es significativa al 1% y los ratios 15 y 20 no son significativos al 5%.

Z14": es significativa al 1% y todos los ratios son significativos al 5%.

**A:** -2 logaritmo de verosimilitud para el análisis LOGIT y la lambda de Wilks para el discriminante. **B:** Bondad del ajuste para LOGIT y correlación canónica para el discriminante. **C:** Chi-cuadrado del modelo.

*Figura 7.7. Selección de los ratios a través de la consideración conjunta de las dos muestras para su cálculo.*

## 7. 5. LA TRANSFORMACIÓN DE VARIABLES: UN INTENTO DE MEJORAR LOS MODELOS

En este apartado vamos a intentar mejorar los modelos conseguidos en el anterior y con este fin seguiremos la siguiente metodología de trabajo:

- a. Identificar los valores atípicos a través de los residuos, es decir, aquellos que muestren un valor absoluto elevado [Novales, 1993] y, en todo caso, superior a dos veces la desviación típica.
- b. Eliminar esos casos.
- c. Reestimar los modelos.

Además, hemos probado otra formulación en los ratios: su logaritmo neperiano.

Por lo tanto, en lo que a continuación se expone, el lector podrá descubrir las diferencias entre los modelos con todos los casos y aquellos otros en los que los más influyentes han sido rechazados. Simultáneamente, se plantea la comparación para cada una de sus formulaciones.

### 7. 5. 1. LA CONSIDERACIÓN DE LOS VALORES ATÍPICOS EN LA FORMULACIÓN ORIGINAL

En este apartado recogemos la primera de las comparaciones. El programa identifica los valores influyentes como aquellos cuyo residuos *estudiantizados* sean superiores a dos. Dichos residuos miden el cambio en la desviación típica del modelo si el caso analizado es eliminado de la estimación. Siguen una distribución t-Student con  $n-k-2$  grados de libertad<sup>1</sup>.

---

<sup>1</sup> Se expresa como:  $Z = \text{residuo} / \text{desviación típica}$ , en el numerador aparece el residuo como diferencia entre la probabilidad observada del evento y la predecida y en el denominador, la desviación típica estimada. Con la peculiaridad de que ésta se ha calculado omitiendo el valor influyente [Novales, 1993]. Se distribuye como una t-student con  $n-k-2$  grados de libertad, siendo  $n$  el número de casos y  $k$  el número de restricciones.

En las siguientes figuras presentamos los resultados en cuanto a formulación y clasificación de las funciones Z3', Z34', Z24' y Z14' (figuras 7.8 y 7.9) así como las correspondientes a Z3'', Z34'', Z24'' y Z14'' (figuras 7.10 y 7.11).

Los valores atípicos detectados se muestran a continuación:

1. Para **Z3'** el programa ha detectado tres influyentes:
  - HISITEX (1992): solvente y perteneciente a la muestra primaria, tercer ejercicio previo.
  - ROB-TEX (1994): solvente y perteneciente a la muestra secundaria para el último anterior a la suspensión de pagos.
  - FITEXVA (1993): insolvente, pertenece a la muestra secundaria, para el último previo.
  
2. Para cada ejercicio previo:
  - **Z34'**:
    - Manufacturas John Peter S.A (1992): solvente, muestra secundaria, tercer ejercicio previo.
  - **Z24'**:
    - Duel, S.A (1991): solvente, muestra primaria, segundo ejercicio previo.
    - Turbofil: S.A.(1993), solvente, muestra secundaria, segundo año previo.
  - **Z14'**:
    - Rob-tex, S.A.(1994): solvente, muestra secundaria, último año previo.
  
3. Para **Z3''** se han detectado cinco casos influyentes:
  - HISITEX (1992): solvente, muestra primaria, tercer ejercicio previo.
  - CINCTEX (1991): insolvente, muestra primaria, tercer ejercicio previo.
  - Manufacturas de tricots artisans (1990): insolvente, muestra secundaria, tercer ejercicio previo.
  - FITEXVA (1993): insolvente, muestra secundaria para el último previo.
  
4. Para cada ejercicio previo:
  - **Z34''**:
    - CATEXT (1990): solvente, muestra primaria, tercer ejercicio previo.
    - EGARMODA (1992): solvente, muestra primaria, tercer ejercicio previo.
    - DYELMA (1990): insolvente, muestra primaria, tercer ejercicio previo.
  - **Z24''**:
    - TURBOFIL (1993) solvente, muestra secundaria, segundo ejercicio previo.
  - **Z14''**:
    - Absorba (1993) solvente, muestra primaria, último ejercicio previo.
    - DIFSA (1993) insolvente, muestra primaria, último ejercicio previo.
    - LOMTEX (1995), insolvente, muestra primaria, último ejercicio previo.

Función		FUNCIONES SIN VALORES INFLUYENTES			
Z3'	ORIGINALES	Z3' = -0,0126 R13 + 10,8615 R19 - 0,4665R20 + 1,4004 R36 - 0,7720 (0,0198) (2,3312) (0,995) (0,4035) (0,2955)			
	TIPIFICADAS	Z3' = -0,1344 ZR13 + 1,7775 ZR19 - 14,8986 ZR20 + 1,2627 ZR36 - 2,6720 (0,2111) (0,3815) (3,1765) (0,3638) (0,6402)			
Z34'	ORIGINALES	Z34' = 0,7688 R13 + 6,6064 R19 + 0,4868 R23 - 0,3379 R25 - 0,1869 R31 - 0,2475 (0,2755) (3,7030) (0,2990) (0,1403) (0,1003) (0,7170)			
	TIPIFICADAS	Z34' = 8,2736 ZR13 + 1,0723 ZR19 + 4,6611 ZR23 - 14,5437 ZR25 - 3,1527 ZR31 + 1,0254 (2,9645) (0,6010) (2,8632) (6,0372) (1,6915) (0,6174)			
Z24'	ORIGINALES	Z24' = 0,0701 R07 + 19,4931 R19 - 1,0024 R20 - 7,0102 R21 + 0,6454 (0,0561) (6,2088) (0,3172) (2,8275) (0,5352)			
	TIPIFICADAS	Z24' = 3,2751 ZR07 + 3,3162 ZR19 - 34,4620 ZR20 - 1,2429 ZR21 - 6,4437 (2,6213) (1,0563) (10,9046) (0,5012) (2,2002)			
Z14'	ORIGINALES	Z14' = -0,0786 R38 - 24,2565 R56 + 1,1366 (0,0461) (7,4981) (0,5262)			
	TIPIFICADAS	Z14' = -5,7333 ZR38 - 4,5483 ZR56 - 0,1799 (3,3637) (1,4060) (0,4003)			

**Figura 7.8.** Estadísticos de clasificación de las diferentes funciones teniendo en cuenta los valores atípicos y extrayéndolos.

FUNCIONES	COMPOSICIÓN DE LA MUESTRA				PORCENTAJES DE ACIERTOS				MEDIDAS DE AJUSTE							
	SOLVENTE		INSOLV.		SOLVENTES		INSOLVENTES		TOTAL		-2LOG		BONDAD		CHI-CUADRADO	
Z3'	Con atípicos	110	111	78,38	65	72,04	235,098	198,206	56,836							
	Sin atípicos	99	109	77,06	75,76	76,44	213,053	197,057	74,805							
Z34'	Con atípicos	33	37	64,86	72,73	68,57	75,934	95,259	20,878							
	Sin atípicos	33	36	69,44	81,82	75,36	60,706	57,954	34,818							
Z24'	Con atípicos	34	37	83,78	70,59	77,46	65,171	58,809	33,129							
	Sin atípicos	34	35	85,71	79,41	82,61	55,617	51,906	40,023							
Z14'	Con atípicos	33	37	97,30	12,12	57,14	96,330	70,052	0,482							
	Sin atípicos	36	33	77,78	69,70	73,91	71,166	59,127	24,358							

Todas las funciones son significativas al 1%.

**Figura 7. 9.** Estadísticos de clasificación de las diferentes funciones teniendo en cuenta los valores atípicos y extrayéndolos.

Función		FUNCIONES SIN VALORES INFLUYENTES					
Z3"	ORIGINALES	Z3" = 0,0338 R18 + 15,5966 R19 - 0,6529 R20 - 23,4213 R21 + 25,9505 R55 + 5,5723 R72 + 0,4418 (0,0151) (3,1466) (0,1353) (4,6408) (5,8812) (1,9096) (0,3799)					
	TIPIFICADAS	Z3" = 1,5366 ZR18 + 2,5548 ZR19 - 20,7134 ZR20 - 4,1515 ZR21 + 3,0979 ZR55 + 0,8361 ZR72 - 3,8390 (0,6859) (0,5154) (4,2925) (0,8226) (0,7021) (0,2865) (0,8668)					
Z34"	ORIGINALES	Z34" = 0,1011 R62 + 0,00000104 R65 + 2,4143 R70 - 4,5202 (0,0999) (0,00000036) (0,6287) (1,2157)					
	TIPIFICADAS	Z34" = 0,8559 ZR62 + 0,9818 ZR65 + 22,1343 ZR70 - 1,0784 (0,8453) (0,3403) (5,7639) (0,3999)					
Z24"	ORIGINALES	Z24" = 0,0568 R15 + 25,0690 R19 - 1,0167 R20 - 26,9580 R56 - 1,5270 R70 + 1,8083 (0,0754) (9,3781) (0,4401) (11,1499) (0,7451) (0,7092)					
	TIPIFICADAS	Z24" = 1,3063 ZR15 + 4,2681 ZR19 - 35,0327 ZR20 - 5,0646 ZR56 - 13,9421 ZR70 - 4,9551 (1,7355) (1,5967) (15,1651) (2,0948) (6,8036) (2,9598)					
Z14"	ORIGINALES	Z14" = -1,7328 R02 - 0,0000026 R64 + 0,8020 (0,9751) (0,000000838) (0,6440)					
	TIPIFICADAS	Z14" = - 0,9380 ZR02 - 12,7315 ZR64 - 1,1331 (0,5279) (4,0698) (0,4393)					

**Figura 7.10.** Estadísticos de clasificación de las diferentes funciones teniendo en cuenta los valores atípicos y extrayéndolos.

FUNCIONES	COMPOSICIÓN MUESTRA		PORCENTAJES DE ACIERTOS				MEDIDAS DE AJUSTE			
	SOLVENTE	INSOLVENTE	SOLVENTES	INSOLVENTES	TOTAL	-2LOG.	BONDAD	CHI-CUADRADO		
Z3"	Con atípicos	111	100	76,58	74	75,36	206,904	247,291	85,031	
	Sin atípicos	110	97	80	79,38	79,71	171,560	159,833	114,586	
Z34"	Con atípicos	37	33	81,08	69,70	75,71	77,556	71,479	19,256	
	Sin atípicos	35	32	88,57	78,13	83,58	61,503	65,249	31,244	
Z24"	Con atípicos	37	34	83,78	91,18	87,32	42,422	42,104	55,879	
	Sin atípicos	36	34	86,11	91,18	88,57	35,793	34,758	61,190	
Z14"	Con atípicos	37	33	86,49	66,67	77,14	74,875	106,445	21,937	
	Sin atípicos	36	31	88,89	77,42	83,58	47,585	59,465	44,923	

Todas las funciones son significativas al 1%.

Figura 7. 11. Estadísticos de clasificación de las diferentes funciones teniendo en cuenta los valores atípicos y extrayéndolos.

La extracción de estos valores ha modificado los parámetros asociados a los ratios en todos los casos. No es posible encontrar un patrón de comportamiento en los cambios, de tal forma que en unos casos un ratio puede modificarse al alza o a la baja. Lo mismo podría decirse con respecto a la significación de cada uno de ellos, que aquí no se incluye para no hacer excesivamente larga esta exposición.

Resulta también difícil valorar si el cambio es más importante en el primer grupo de funciones que en el segundo, siendo diferente según el tipo de función que tomemos como referencia (por ejemplo, el cambio en las variables de la función Z14' supera el triple de los valores con atípicos, sin embargo, en el caso de la Z14'' no llega al 1,5).

Por el contrario, sí podemos afirmar con generalidad algunos aspectos relevantes en cuanto al porcentaje de aciertos y a las medidas de ajuste:

1. En todas las funciones la extracción de los valores atípicos mejora los resultados de la clasificación, lo cual siendo previsible, resulta en algunas realmente significativo, en las funciones Z14' es del 30% (el mayor porcentaje de todos) en el primer caso, y en Z14'' es 8,3%.
2. Por grupos, tanto las solventes como las insolventes mejoran sus éxitos, ahora bien, en Z3' y Z14' se observa un empeoramiento del grupo solvente; en la primera función no es importante (de 78% a 77%), en la segunda sí lo es (del 97,31% a 77,78%). Aunque también es cierto que en esta última el valor influyente sesgaba de forma muy significativa la función hacia el primer grupo en detrimento del segundo (después de borrado éste, el porcentaje de aciertos pasa de 12,12% a 69,70%).
3. Las medidas de ajuste en todos los casos estudiados presentan valores inferiores, lo cual está en concordancia con lo acabado de mencionar.

#### 7. 5. 2. LA TRANSFORMACIÓN LOGARÍTMICA

Procedamos ahora a realizar la transformación logarítmica. Lo que se expondrá a continuación ha sido calculado sin los valores influyentes.

El cálculo del logaritmo neperiano para cada una de las variables independientes acarrea un problema adicional en su tratamiento estadístico, como es que al estar acotado en valores positivos o cero, no se calculan aquellos ratios con valor negativo. La repercusión inmediata es la aparición de más valores perdidos dentro de la muestra.

Aquí, no obstante, no hemos procedido a sustituirlos por las respectivas medias como en situaciones anteriores. La razón es sencilla, una cosa es plantear una suposición del tipo “a falta de información, al ratio de la empresa *se le supone* no muy diferente a la media correspondiente”, y otra muy diferente “a falta de posibilidad de calculo *se le supone* no muy diferente a la media correspondiente”, en este segundo planteamiento no es necesario suponer nada, puesto que el ratio subyacente existe en realidad.

Lógicamente, el hecho de trabajar con valores perdidos reduce tanto la muestra como la calidad de los resultados que de ella se obtiene.

Los resultados se presentan en las siguientes figuras 7.12, 7.13, 7.14 y 7.15, a destacar que hemos incluido únicamente la formulación tipificada para poder realizar la comparación en cada modelo. La repercusión de dicha transformación es diferente para cada una de las funciones.

- Para las funciones totales:
  - En Z3' y Z3'' la transformación logarítmica conduce a peores resultados que el cálculo a través de las variables originales. Apuntemos, que las medidas de ajuste son menores en esta formulación.
- Para las funciones anuales:
  - Z34' y Z34'' presentan mejores medidas de ajuste, pero, mientras Z34' conduce a una mayor exactitud que en la muestra original, Z34'' obtiene menor porcentaje de aciertos.
  - En Z24' y Z24'' tienen porcentajes de aciertos inferiores. El ajuste es mejor en la primera, no, en cambio, en la segunda.
  - En Z14' los porcentajes de clasificación empeoran respecto a la original y las medidas de ajuste se sitúan en valores inferiores. En Z14'' el porcentaje de aciertos totales es superior en la formulación logarítmica que en la original, siendo en este caso la mejor de todas ellas, y lo mismo puede decirse con respecto a las medidas de ajuste.

FUNCIONES		FUNCIONES	
Z3'	TIPIFICADA	$Z3' = -0,1344 ZR13 + 1,7775 ZR19 - 14,8986 ZR20 + 1,2627 ZR36 - 2,6720$ (0,2111) (0,3815) (3,1765) (0,3638) (0,6402)	
	LOGARITMICA	$Z3' = 0,8115 LR13 - 0,2519 LR19 - 0,5045 LR20 + 0,6379 LR36 + 0,6649$ (0,2349) (0,3182) (0,1665) (0,3307) (0,7971)	
Z34'	TIPIFICADA	$Z34' = 8,2736 ZR13 + 1,0723 ZR19 + 4,6611 ZR23 - 14,5437 ZR25 - 3,1527 ZR31 + 1,0254$ (2,9645) (0,6010) (2,8632) (6,0372) (1,6915) (0,6174)	
	LOGARITMICA	$Z34' = 1,0320 LR13 + 0,2380 LR19 + 2,1516 LR23 - 1,1931 LR25 - 2,2823 LR31 + 3,6328$ (0,7614) (0,8261) (1,2227) (0,7712) (1,1954) (2,8888)	
Z24'	TIPIFICADA	$Z24' = 3,2751 ZR07 + 3,3162 ZR19 - 34,4620 ZR20 - 1,2429 ZR21 - 6,4437$ (2,6213) (1,0563) (10,9046) (0,5012) (2,2002)	
	LOGARITMICA	$Z24' = 0,4148 LR07 + 1,8082 LR19 - 1,2641 LR20 - 0,8627 LR21 + 2,3612$ (0,2824) (0,6380) (0,4541) (0,4087) (1,5985)	
Z14'	TIPIFICADA	$Z14' = -5,7333 ZR38 - 4,5483 ZR56 - 0,1799$ (3,3637) (1,4060) (0,4003)	
	LOGARITMICA	$Z14' = -0,4422 LR38 - 0,7041 LR56 - 2,2174$ (0,3886) (0,3636) (1,1455)	

Figura 7.12. Estadísticos de clasificación de las diferentes funciones (transformación logarítmica).