



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

(ICADE)

Análisis de la sostenibilidad financiera de las Cajas Rurales a través de modelos logit y regresión de Cox. Propuesta de un indicador sintético de salud financiera.

TESIS DOCTORAL

Doctorando: Ldo. D. Antonio Madera del Pozo

Directores: Prof. Dr. D. Antonio Rúa Vieites

Prof. Dr. D^a Natalia Cassinello Plaza

Madrid, abril de 2017

A mi Capitán Tapón

AGRADECIMIENTOS

A través de estas líneas quiero mostrar mi más sincero agradecimiento a todas aquellas personas que directa o indirectamente han contribuido a la elaboración de esta Tesis.

En primer lugar quiero y debo dar las GRACIAS en “mayúsculas” a mis dos tutores, Antonio y Natalia, por el intenso trabajo, por las guías y consejos y, sobre todo, por vuestra confianza, creísteis en este Proyecto a pesar de las adversidades surgidas.

No puedo olvidar a Juan José Barbachano, por su lucha y tesón para que la Tesis pudiera ver la luz, con ánimos que fueron tan bienvenidos en momentos de flaqueza.

Igualmente quiero destacar la labor de María Coronado, ya que de su contribución inicial se sustenta la idea principal de esta Tesis. Gracias por todo.

A mi familia profesional, a Ignacio y Obdulia, así como a la Unión Nacional de Cooperativas de Crédito. Gracias por toda vuestra comprensión y ayuda.

A mis amigos, por sus muestras de afecto, destacando a nuestra Rosi “Fermina” y Alberto, por sus intensas revisiones que han permitido cazar esos pequeños detalles.

A mis padres. Gracias por todo aquello de lo que os habéis privado para que nunca nos haya faltado acceso a la formación, sin duda la base que sustenta y da sentido a este trabajo.

Ana, gracias de todo corazón. Gracias por esas cómplices miradas en momentos que tanto las necesitaba. Gracias por tu infinita paciencia, comprensión y ayuda. Perdona por el tiempo que te he robado de esos ratitos y escapadas que tanto nos gustan. A partir de ahora a disfrutar juntos de esta pequeña familia que en breve se convertirá en tres, ese sí que será nuestro mayor Proyecto.

ÍNDICE GENERAL

ÍNDICE DE TABLAS	11
ÍNDICE DE ILUSTRACIONES.....	13
GLOSARIO DE TÉRMINOS.....	15
<u>CAPÍTULO PRIMERO: INTRODUCCIÓN</u>	17
1.1 PLANTEAMIENTO Y JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	19
1.2 OBJETIVOS.....	22
1.3 METODOLOGÍA Y ENFOQUE EPISTEMOLÓGICO.....	23
1.4 ESTRUCTURA DE LA TESIS.....	25
1.5 SINOPSIS	28
<u>CAPITULO SEGUNDO:ESTADO DE LA CUESTIÓN</u>	31
2.1 SINÓPSIS DEL CAPÍTULO	33
2.2 METODOLOGÍA SEGUIDA EN LA ELABORACIÓN DE LA REVISIÓN BIBLIOGRÁFICA.....	34
2.3 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA EMPRESARIAL.....	37
2.4 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA BANCARIA.....	40
2.4.1. Definición de insolvencia bancaria	42
2.4.2. Revisión de las técnicas utilizadas para la construcción de modelos de previsión de insolvencia bancaria.....	43
2.4.3. La investigación sobre insolvencias bancarias en España.	49
2.5 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA DE COOPERATIVAS DE CRÉDITO.....	54
2.5.1 Definición de la insolvencia de cooperativas de crédito	54
2.5.2 Revisión de las técnicas utilizadas para la construcción de modelos de previsión de insolvencia de cooperativas de crédito.	56
2.6 CONCLUSIONES DEL CAPÍTULO.....	60
<u>CAPÍTULO TERCERO:EL ESTUDIO DE LA SITUACIÓN FINANCIERA DE LAS CAJAS RURALES....</u>	63
3.1 SINOPSIS DEL CAPÍTULO	65
3.2 HIPÓTESIS DEL INVESTIGADOR.....	66
3.3 MODELOS DE ALERTA TEMPRANA.....	68
3.3.1 Regresión Logística binaria o Logit.....	72
3.3.1.1 Definición formal del modelo logit.....	72
3.3.1.2 Críticas al uso del modelo logit.....	74
3.3.2 Regresión logística multiperiodo.....	76
3.3.3 Regresión de Riesgos Proporcionales o Regresión de Cox.....	78
3.4 POBLACIÓN Y MUESTRA	80
3.4.1 Las Cajas Rurales.....	81

3.4.2	Fuentes de información	86
3.4.3	Limitaciones para la obtención de la información.....	87
3.4.4	Muestra para la estimación y validación del modelo.....	88
3.4.5	Segmentación temporal	89
3.5	VARIABLE DEPENDIENTE	91
3.6	VARIABLES EXPLICATIVAS	98
3.6.1	Ratios Financieras	102
3.6.1.1	Capitalización	102
3.6.1.2	Activo.....	103
3.6.1.3	Equipo gestor.....	105
3.6.1.4	Rentabilidad.....	106
3.6.1.5	Endeudamiento.....	108
3.6.1.6	Liquidez.....	109
3.6.2	Ratios macroeconómicas.....	110
3.6.3	Otras ratios	111
3.7	CONCLUSIONES DEL CAPÍTULO.....	114
CAPÍTULO CUARTO: RESULTADOS DEL ESTUDIO EMPÍRICO		117
4.1	SINOPSIS DEL CAPÍTULO	119
4.2	ANÁLISIS UNIVARIANTE	120
4.3	SELECCIÓN DE VARIABLES EXPLICATIVAS.....	138
4.3.1	Métodos tradicionales	140
4.3.2	Análisis factorial.....	141
4.4	EXPOSICIÓN DE LOS RESULTADOS OBTENIDOS DE LA ESTIMACIÓN DE LOS MODELOS LOGIT Y REGRESIÓN DE COX.....	148
4.4.1	Estimación de la Regresión Logística Binaria (Logit estático)	149
4.4.1.1	Periodo 1: desde 1992 hasta 1994.....	151
4.4.1.2	Periodo 2: desde 2000 hasta 2002.....	151
4.4.1.3	Periodo 3: desde 2002 hasta 2007.....	158
4.4.1.4	Periodo 4: desde 2011 hasta 2014.....	158
4.4.1.5	Periodo completo: 1.991-2.015.....	161
4.4.2	Estimación de la regresión logística multiperiodo	167
4.4.3	Modelo de Riesgos Proporcionales o Modelo de Cox.....	174
4.4.3.1	Modelo de Cox con covariables dependientes del tiempo.....	176
4.4.3.2	Modelo de Cox estratificado	181
4.5	CONCLUSIONES DEL CAPÍTULO.....	185

CAPÍTULO QUINTO: PROPUESTA DE UN ÍNDICADOR SINTÉTICO DE SALUD FINANCIERA DE LAS CAJAS RURALES	189
5.1 SINOPSIS DEL CAPÍTULO	191
5.2 METODOLOGÍA APLICABLE.....	192
5.3 CONSTRUCCIÓN DE UN INDICADOR SINTÉTICO PARA MEDIR LA SALUD FINANCIERA DE LAS CAJAS RURALES	195
5.4 EXPOSICIÓN DE RESULTADOS.....	201
5.5 CONCLUSIONES DEL CAPÍTULO.....	205
CAPÍTULO SEXTO: PREVISIONES FUTURAS SOBRE EL SECTOR DE CAJAS RURALES A PARTIR DE LOS MODELOS ESTIMADOS.....	207
6.1 SINOPSIS DEL CAPÍTULO	209
6.2 PREVISIONES FUTURAS.....	209
CONCLUSIONES	215
BIBLIOGRAFÍA.....	223
ANEXOS.....	249

ÍNDICE DE TABLAS

Tabla 1: Principales publicaciones sobre quiebra empresarial clasificadas en función de la técnica empleada.	38
Tabla 2: Principales publicaciones sobre la quiebra de entidades financieras (excepto cooperativas de crédito). Clasificación por metodología.....	46
Tabla 3: Resumen de las investigaciones en materia de insolvencia bancaria en España.	53
Tabla 4: Revisión bibliográfica de la definición de insolvencia de una cooperativas de crédito.	55
Tabla 5: Revisión bibliográfica de las técnicas estadísticas utilizadas para el estudio de la insolvencia de cooperativas de crédito.	71
Tabla 6: Distribución territorial de las cajas rurales entre 1915 y 1933	84
Tabla 7: Casos eliminados de la muestra	88
Tabla 8: Criterios seguidos en la determinación de la variable dependiente en la investigación sobre insolvencia de cooperativas de crédito.....	93
Tabla 9: Clasificación de las Cajas Rurales que cumplen con los criterios de insolvencia establecidos.	95
Tabla 10: Análisis de los procesos de fusión de Cajas Rurales	96
Tabla 11: Principales indicadores utilizados en las investigaciones sobre insolvencias de Cooperativas de Crédito.	100
Tabla 12: Resumen de ratios financieras utilizadas para el análisis de la sosteibilidad financiera de las Cajas Rurales españolas.....	113
Tabla 13: Observaciones de Cajas Rurales incluidas en la muestra.....	120
Tabla 14: Principales características de la muestra.	122
Tabla 15: Matriz de correlaciones entre las variables explicativas	124
Tabla 16: Test de Kaiser-Meyer-Olkin y Test de la esfericidad de Barlett.....	143
Tabla 17: Total de la varianza explicada por los distintos factores	144
Tabla 18: Matriz de componentes	145
Tabla 19: Matriz de componentes rotados	146
Tabla 20: Codificación de la variable dependiente	151
Tabla 21: Codificación de las variables independientes categóricas	152
Tabla 22: Historial de Iteraciones	153
Tabla 23: Parámetros estimados en la construcción del modelo.....	153
Tabla 24: Prueba de la bondad del ajuste del modelo	155
Tabla 25: Tabla de contingencia para la prueba de Hosmes y Lemeshow	156
Tabla 26: Prueba de la bondad del ajuste. Prueba de Hosmer y Lemeshow.	156
Tabla 27: Clasificación del modelo logit para el periodo 2.000-2.002	157
Tabla 28: Parámetros estimados en el modelo logit del periodo 2011-2014.....	159
Tabla 29: Prueba de la bondad del ajuste del modelo	160
Tabla 30: Tabla de clasificación del modelo logit para el periodo 2011-2014	160
Tabla 31: Parámetros estimado del modelo logit para el periodo 1991-2014	162
Tabla 32: Bondad del ajuste del modelo logit.....	163
Tabla 33: Tabla de clasificación del modelo logit estático.....	164
Tabla 34: Clasificación de las entidades a partir del modelo logit con selección de variables por componentes principales	166
Tabla 35: Variables introducidas en el modelo logit multiperiodo	168
Tabla 36: Tabla de clasificación modelo logit multiperiodo.....	172

Tabla 37: Modelo logit multiperiodo seleccionado por componentes principales.....	172
Tabla 38: Tabla de clasificación del modelo logit multiperiodo seleccionado por componentes principales.....	173
Tabla 39: variables significativas en el modelo de Cox.....	176
Tabla 40: Contraste de la hipótesis general del modelo.	179
Tabla 41: Estratos considerados para la estimación del modelo de Cox.....	183
Tabla 42: Variables significativas en la estimación de los diferentes modelos	186
Tabla 43: Comparación del grado de clasificación correcta en la estimación de los modelos logit.....	188
Tabla 44: Coeficientes de correlación	197
Tabla 45: Prueba t de igualdad de medias	198
Tabla 46: Varianza total explicada	198
Tabla 47: Prueba KMO y Bartlett.....	199
Tabla 48: Matriz de coeficiente de puntuación de componente	199
Tabla 49: Estadísticos descriptivos asociados al indicador sintético	203
Tabla 50: Media del ISF y Prueba t en función de la ubicación geográfica.....	205
Tabla 51: Puntuaciones máximas y mínimas otorgadas por el indicador sintético para las cajas rurales operativas en el año 2.016.	211
Tabla 52: Clasificación correcta de los diferentes modelos logit	219

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Sinópsis de la Tesis	29
Ilustración 2: Ubicación del Capítulo segundo dentro de la Sinopsis de la Tesis	33
Ilustración 3: Tipos de documentos utilizados en la revisión bibliográfica	35
Ilustración 4: Número de revistas en función de su calificación JCR y Scopus	36
Ilustración 5: Distribución de las publicaciones utilizadas por años	36
Ilustración 6: Ubicación del Capítulo Tercero dentro de la Sinopsis de la Tesis	65
Ilustración 7: Distribución de las bajas de Cajas Rurales entre 1974-2015.	94
Ilustración 8: Cajas Rurales insolventes incorporadas a la muestra Comparada con la población del periodo.....	94
Ilustración 9: Ubicación del Capítulo Cuarto dentro de la sinopsis de la Tesis	119
Ilustración 10: Antigüedad media de las Cajas Rurales españolas entre 1991 y 2015.....	125
Ilustración 11: Evolución del número medio de empleados en las cajas rurales fusionadas	126
Ilustración 12: Evolución del número medio de empleados de las cajas rurales operativas.....	126
Ilustración 13: Evolución del número medio de oficinas de cajas rurales fusionadas	127
Ilustración 14: Evolución del número medio de oficinas de cajas rurales operativas.....	127
Ilustración 15: Evolución de la ratio de Capitalización en el periodo de estudio.	128
Ilustración 16: Evolución de la ratio de tamaño a lo largo del periodo de estudio.....	129
Ilustración 17: Evolución de la ratio negocio a lo largo del periodo de estudio.	130
Ilustración 18: Evolución de la ratio coste del riesgo a lo largo del periodo de estudio	131
Ilustración 19: Evolución de la ratio de esfuerzo en provisiones	132
Ilustración 20: Evolución de la ratio eqg1 a lo largo del periodo de estudio.	133
Ilustración 21: Evolución la ratio eficiencia a lo largo del periodo de estudio	133
Ilustración 22: Evolución de la ratio rentab a lo largo del periodo de estudio	134
Ilustración 23: Evolucion de la ratio ROA a lo largo del periodo de estudio	135
Ilustración 24: Evolución de la ratio roe a lo largo del periodo de estudio.....	135
Ilustración 25: Evolución de la ratio apalancamiento a lo largo del periodo de estudio	136
Ilustración 26: Evolución de la ratio deuda a lo largo del periodo de estudio	137
Ilustración 27: Evolución de la ratio funding gap a lo largo del periodo de estudio	137
Ilustración 28: Evolución de la ratio liquid a lo largo del periodo de estudio.....	138
Ilustración 29: Clasificación del modelo logit estático	164
Ilustración 30: Clasificación del modelo logit estático a partir de la selección por componentes principales.....	166
Ilustración 31: Clasificación del modelo logit multiperiodo.....	170
Ilustración 32: Clasificación de la muestra de validación del logit multiperiodo	171
Ilustración 33: Probabilidades medias estimadas a partir de datos económicos y financieros de hasta cuatro años anteriores a la fusión.....	173
Ilustración 34: Comprobación de la hipótesis de riesgos proporcionales mediante la gráfica de los residuos de Schoenfeld.....	180
Ilustración 35: Comprobación gráfica de la condición de riesgos proporcionales	181
Ilustración 36: Probabilidad de supervivencia estimada a partir del modelo de riesgos proporcionales de Cox estratificado.....	184
Ilustración 37: Función de supervivencia a partir del estimador de kaplan-meier.....	185
Ilustración 38: Ubicación del Capítulo Quinto dentro de la sinopsis de la Tesis	191
Ilustración 39: Representación gráfica de las puntuaciones obtenidas del indicador sintético	201
Ilustración 40: Distribución anual de la puntuación media del indicador sintético.....	202

Ilustración 41: Gráfico de cargas puntuaciones indicador sintético	203
Ilustración 42: Resultados del indicador sintético en función de la ubicación geográfica.....	204
Ilustración 43: Ubicación del Capítulo Sexto dentro de la Sinopsis de la Tesis	209
Ilustración 44: Gráfico de cajas de las puntuaciones otorgadas por el indicador sintético a las Cajas Rurales operativas en la actualidad.....	211
Ilustración 45: Puntuaciones otorgadas por el indicador sintético para cada Caja Rural.....	212
Ilustración 46: Distribución de la probabilidad de insolvencia estimada para las Cajas Rurales operativas en la actualidad.....	213
Ilustración 47: Probabilidad de fusión para la muestra de Cajas Rurales operativas en la actualidad	214

GLOSARIO DE TÉRMINOS

- BAKIS: *Bakred Information System* (Alemania)
- BASILEA: hace referencia a los Acuerdos de Basilea de 1988, en los que se fijaba el capital mínimo que debían presentar las entidades financieras en función de sus activos ponderados por riesgo.
- CAEL: siglas de *Capital, Assets, Equity & Leverage*.
- CAMEL: siglas de *Capital, Assets, Management, Equity & Leverage*.
- CAREL: *Indicadores Econômico-financeiros de Instituições sob Acompanhamento e Controle*
- FDIC: *Federal Deposit Insurance Corporation* de EE.UU.
- FIMS: *Financial Institution Monitoring System*
- LTD: *Loan to Deposits*
- OCC: *Currency Administration of National Banks*
- ORAP: *Organization and Reinforcement of Preventive Action* (Francia)
- PATROL: Sistema de rating establecido por el Banco de Italia para evaluar la banca de dicho país.
- PEARL: es una adaptación del CAMEL al entorno de las cooperativas de crédito.
- PIB: Producto Interior Bruto
- RAST: *Risk Analysis Support Tool* (Holanda)
- RATE: *Risk Assessment Tools of Supervision and Evaluation* (Reino Unido)
- SAABA: *Support System for Banking Analysis* (Francia)
- SIP: Sistema Institucional de Protección
- UBSS: *Uniform Bank Surveillance Screen*
- UNACC: Unión Nacional de Cooperativas de Crédito.
- WOCU: World Council of Credit Unions

CAPÍTULO PRIMERO

INTRODUCCIÓN

1.1 PLANTEAMIENTO Y JUSTIFICACIÓN DE LA INVESTIGACIÓN

Dentro de la literatura académica son numerosas las investigaciones que tratan de determinar las causas que subyacen a las crisis bancarias¹, básicamente porque estos acontecimientos están rodeados de factores que pueden influir negativamente en el crecimiento económico de un país², tanto por la restricción del crédito y sus efectos sobre empresas y familias, como por la elevada cantidad de recursos que se hacen necesarios para el rescate de una entidad financiera en una situación de insolvencia³.

En el ámbito internacional las primeras investigaciones se remontan a las publicaciones de Secrist (1938) y Altman (1968), si bien en España esta línea de investigación es más tardía, iniciada con la publicación de Laffarga *et al* (1987) y centrada desde entonces en el estudio de la insolvencia de bancos y cajas de ahorros, dotando de reducida importancia a las Cajas Rurales, cooperativas de crédito⁴ vinculadas al sector agrícola, por su baja representatividad, y ello a pesar de estar inmersas en procesos de concentración que agudizan su riesgo sistémico.

Aunque la existencia de procesos de insolvencias dentro del sector de Cajas Rurales es un acontecimiento que reviste de cierto carácter excepcional, tan solo habiéndose constado una intervención por parte del Banco de España en los últimos treinta años, de un análisis pormenorizado del sector se desprende la existencia de procesos de concentración que

¹ Para mayor información al respecto, en el Capítulo Segundo de esta Tesis se incluye una recapitulación de las principales publicaciones encontradas sobre el estudio y predicción de la insolvencia bancaria, catalogadas en función de la técnica estadística empleada por sus autores. Entre otros trabajos, caben destacar las publicaciones de Altman (1968), Meyer *et al.* (1970), Sinkey (1975), Martin (1977), Marino *et al* (1983), Pantalone *et al* (1987), Demirgüç *et al* (1999), Logan (2001), Andersen (2008) o Poghosyan *et al* (2011).

² Álvarez (2008) afirma que las crisis financieras tienen una traslación a la economía real a través de los canales de crédito, patrimonial, confianza y fiscal, detallados en el Capítulo Segundo de esta Tesis.

³ Serrano (2013) indica que la reestructuración del sistema financiero español iniciada en 2007, con la bancarización de las cajas de ahorro para ganar tamaño y mejorar solvencia, implicó ayudas públicas que alcanzaron los 60.000 MM €.

⁴ Esta línea de investigación es relativamente reciente, siendo pocas las investigaciones existentes en España, ya que las pocas localizadas se refieren en su mayoría a cooperativas de crédito europeas y latinoamericanas ubicadas en países en los que estas instituciones gozan de mayor relevancia, siendo destacables los trabajos de Simon (1980), Dabós (1998), Bressan *et al* (2004), Porath (2006), Cabo *et al* (2010), Bressan *et al* (2010), Lima (2012) y Fiordelisi *et al.* (2013)

podrían estar buscando la sostenibilidad económica y/o financiera de las entidades implicadas, como ponen de manifiesto Porath (2006), Cabo *et al.* (2010) y Lima (2012) en sus respectivos estudios sobre la banca cooperativa en Alemania, Portugal y Argentina.

El objetivo de esta Tesis no es otro que el de analizar los procesos de concentración de las Cajas Rurales a través de modelos logit y regresión de Cox, para determinar si esos procesos fueron respuesta a situaciones de debilidad económica y/o financiera que amenazaban la sostenibilidad de las entidades implicadas, con el fin de prever este tipo de situaciones en un futuro. Además se propone un indicador sintético que permita un acercamiento más profundo a la realidad económica y financiera de este sector.

El interés de este estudio subyace, tanto en el riesgo sistémico indicado como en la importancia que las Cajas Rurales tienen dentro del ámbito rural y agrario (Navarro; 2004 y Palomo *et al.*; 2010a), no solo porque financian a un sector sobre el que no siempre el resto del sistema financiero ha estado interesado en cubrir (Palacios, 2003), sino que también porque su alta proximidad al cliente⁵ se traduce en la prestación de servicios financieros en municipios que por su menor densidad poblacional pueden estar viéndose negativamente afectados por la concentración sectorial y la reducción de oficinas de otro tipo de entidades financieras.

En este sentido, a diciembre de 2015 las Cajas Rurales estaban presentes en algo más de 1.290 municipios españoles, el 40% de los cuales presentaba una población inferior a los 2.000 habitantes. De hecho, estas entidades no se caracterizan por su presencia en municipios de elevada densidad poblacional, ya que el 80% de las localidades en las que estaban presentes apenas superaban los 25.000 habitantes.

⁵ Palomo *et al.* (2010a) señala que una de las fortalezas de las cooperativas de crédito es su mayor proximidad al cliente en comparación al resto de entidades, porque estas entidades mantienen un trato preferencial con sus socios a lo que ofrecen productos a medida.

Por sucursales la situación es similar, concentrando el 21% de las oficinas en localidades inferiores a los 800 habitantes (incluyendo municipios inferiores a los 300 habitantes en los que prácticamente se posicionan como la única entidad financiera operativa), otro 59% en municipios de población inferior a los 11.500 habitantes, y el resto en capitales de provincias en las que llegan a mantener varias sucursales operativas.

Para alcanzar el objetivo planteado, en esta Tesis se recurre a dos modelos estadísticos ampliamente utilizados en las investigaciones sobre insolvencias de bancos y cooperativas de crédito en un ámbito nacional e internacional⁶, concretamente un modelo de respuesta binaria (logit) y un modelo de supervivencia (regresión de riesgos proporcionales de Cox).

El uso de modelos logit en el estudio de la insolvencia empresarial y bancaria está generalizado, aunque Shumway (2001), Nam *et al.*, (2008) y Nuñez *et al.* (2011) lo critican por el hecho de considerar la quiebra como un acontecimiento estático cuando en realidad se trata de un proceso continuado en el tiempo. Por ello proponen el uso de una versión extendida del modelo logit (Shumway, 2001) que considere el carácter dinámico de los datos, aunque es cierto que aún no existiría evidencia empírica suficiente para confirmar que una técnica funciona mejor que la otra.

En base a lo anterior, en esta Tesis se han utilizado ambos tipos de modelos logit (versión estática y dinámica), además del modelo de supervivencia de Cox, no solo para determinar si los procesos de fusiones de Cajas Rurales acaecidos entre 1991 y 2015 buscaban la sostenibilidad económica y/o financiera de las entidades implicadas, sino que también para conocer las causas y, sobre todo, prever este tipo de situaciones en un futuro.

Módica *et al* (2012) pone de manifiesto que el uso de modelos de este tipo no es suficiente para conocer la salud financiera de una empresa, dado que el acontecimiento de quiebra y

⁶ En el Capítulo Segundo se recoge una recapitulación de las investigaciones existentes.

no quiebra es un fenómeno que difícilmente puede ser medidos a través de variables dicotómicas por tratarse de un proceso de deterioro continuado en el tiempo, proponiendo por ello la utilización de indicadores sintéticos.

Es por ello por lo que en esta Tesis se construye un indicador sintético de salud financiera a partir de variables que miden aspectos económicos y financieros de cada una de ellas para observar si es posible tener un acercamiento más profundo de la realidad económica y/o financiera de este sector a partir del empleo del indicado estimado.

Como se ha puesto de manifiesto, dentro de la investigación sobre insolvencias bancarias es habitual el uso de modelos estadísticos, más o menos complejos, que anticipan situaciones de insolvencia de bancos, si bien no han sido localizadas aplicaciones de indicadores sintéticos que midan la salud financiera del sector bancario.

Por ello la propuesta planteada en esta Tesis es novedosa, no solo por la aplicación de modelos logit y regresión de Cox a un sector poco estudiado como son las Cajas Rurales, sino que también por suponer una primera aplicación de indicadores sintéticos para medir la salud financiera de un tipo de entidad financiera española.

1.2 OBJETIVOS

Con esta Tesis se contribuye a un conocimiento financiero más profundo de las Cajas Rurales españolas analizando la fusión como parte de un proceso de reestructuración del sector, buscando la determinación de los siguientes aspectos:

- a. Si los procesos de concentración de Cajas Rurales acaecidos entre 1991 y 2015 buscaban la sostenibilidad económica y/o financiera de las entidades involucradas, tal y como observan Porath (2006), Cabo *et al.* (2010) y Lima (2012) en su estudio de las

insolvencias de cooperativas de crédito en Alemania, Argentina y Portugal, respectivamente.

- b. Si los modelos de respuesta binaria (logit) y supervivencia (Cox) son eficaces para anticipar la pérdida de sostenibilidad económica y/o financiera que obligue a la Caja Rural a formar parte de un proceso de fusión, como ocurre en la mayor parte de investigaciones sobre insolvencias bancarias, en las que los modelos logit y supervivencia son los más extendidos por presentar los mayores porcentajes de aciertos en su aplicación sobre la muestra de validación.
- c. Si la información que se desprende de los estados financieros de las Cajas Rurales es suficiente para construir los modelos anteriores, determinando que factores económicos y financieros son determinantes para permitir actuaciones que eviten debilidades futuras.
- d. Si la propuesta de un indicador sintético que mida la salud financiera de las Cajas Rurales es eficiente en la discriminación de las entidades que formaron parte de los procesos de fusión y, por tanto, es factible su uso en el conocimiento de la situación económica y financiera de las mismas.

1.3 METODOLOGÍA Y ENFOQUE EPISTEMOLÓGICO.

Para conseguir el objetivo planteado, se propone seguir el enfoque epistemológico de la economía positiva o cuantitativa propuesto por Friedman (1953), consistente en fallar en desaprobar las hipótesis de partida a partir de técnicas estadísticas.

Para Friedman (1953) *“la economía positiva no abre juicios de valor”* porque explica las relaciones entre diferentes acontecimientos a través de teorías formuladas sobre la base de procedimientos empíricos, tal y como postula el positivismo en su defensa del conocimiento científico a partir de datos empíricos (Meza, 2003).

En este sentido, Friedman (1953) afirma que *“el último objetivo de una ciencia positiva es el desarrollo de una teoría o hipótesis que produzca predicciones válidas y significativas sobre fenómenos todavía no observados”*.

La teoría solo puede ser puesta en evidencia si su predicción no es suficiente para explicar los acontecimientos observados, si bien *“una teoría no puede probarse comparando sus supuestos directamente con la realidad. Un realismo completo es inalcanzable y la cuestión de si una teoría es bastante realista puede resolverse sólo comprobando si sus predicciones son lo bastante buenas para el propósito que se persigue, o si son mejores que las predicciones de las teorías alternativas”* (Friedman, 1953).

La aplicación del enfoque de la economía positiva a esta Tesis se realizará a partir de su división en dos partes:

- a. Marco teórico: que incluye una revisión bibliográfica de las principales investigaciones existentes en relación a la insolvencia bancaria y de cooperativas de crédito, a partir de las cuales se establecerán hipótesis que permitan explicar los procesos de fusiones de las Cajas Rurales en España.
- b. Marco empírico: en la que se llevará a cabo una recolección de los datos para su posterior tratamiento mediante diversas técnicas estadísticas con el fin de verificar el cumplimiento de las hipótesis establecidas en el marco teórico. Por tanto, se enfoca en conocer la realidad del fenómeno a través de la descripción exacta de los aspectos que la determinan para inferir su comportamiento

Por su objeto de estudio, esta Tesis se puede encuadrar dentro del área de conocimiento de las ciencias sociales, concretamente dentro de la rama de la economía financiera y contabilidad en su variante aplicada a las entidades financieras.

1.4 ESTRUCTURA DE LA TESIS.

Como se ha definido en el epígrafe anterior, esta Tesis sigue el enfoque epistemológico de la economía positiva para analizar si los procesos de concentración de Cajas Rurales están relacionados con la sostenibilidad económica y/o financiera de las entidades implicadas, habiéndose estructurado en dos partes, un marco teórico y un marco empírico.

El marco teórico se inicia con el capítulo segundo, en el que se llevará a cabo una revisión bibliográfica de las principales investigaciones existentes sobre las insolvencias dentro del sector de las cooperativas de crédito, tanto nacional como internacional.

Para ello se comenzará con una breve exposición de la investigación en materia de quiebras empresariales, línea de investigación que podría ser considerada precursora en el uso de modelos para la predicción de situaciones de insolvencias a partir de una serie de variables explicativas.

En este sentido, con esta Tesis no se pretende aportar nuevos conocimientos al estudio de la insolvencia empresarial, sino que se busca ampliar el ya existente sobre cooperativas de crédito, por ello en este segundo capítulo tan solo se recogerán las bases en las que se ha fundamentado esas investigaciones sobre empresas.

La investigación sobre insolvencias de cooperativas de crédito alcanza su máximo desarrollo en la primera década del siglo XXI, principalmente en países latinoamericanos, si bien es cierto que aún serían pocas las investigaciones realizadas sobre este sector. De hecho, en España no han sido encontradas investigaciones similares.

Por ese motivo dentro de este capítulo se incluye un análisis detallado de la situación actual de la investigación académica sobre insolvencias de entidades financieras con el fin de conocer las principales líneas de investigación, metodologías y conclusiones, ya que servirán de base para el posterior desarrollo del marco empírico.

Aunque el volumen de investigaciones sobre la quiebra de entidades financieras españolas es reducido, se ha creído conveniente dedicar un apartado a la recopilación de las encontradas, resaltando técnicas y principales conclusiones para determinar si en España esta área de investigación ha seguido en términos generales las líneas observadas en el resto de investigadores.

Esta primera parte teórica continúa en el capítulo tercero con el desarrollo del procedimiento a seguir para la estimación de los modelos que permitan alcanzar los objetivos fijados, detallándose la muestra a utilizar y los principales problemas encontrados en su construcción.

Este capítulo continúa con la definición de la variable dependiente a considerar en esta Tesis, que surge de una combinación de los criterios cualitativos considerados por el Registro de Banco de España en la baja de entidades financieras, con una serie de criterios cuantitativos contruidos a partir de la situación económica y financiera mostrada por cada una de las Cajas Rurales consideradas en la muestra a lo largo del periodo de estudio.

Para la selección de las variables explicativas que formarán parte de los modelos se recurrirá a lo observado en otras publicaciones sobre el sector de cooperativas de crédito, las empleadas por Banco de España en su Boletín Estadístico, o las consideradas en diferentes normativas aplicables, con las que se trata de medir aspectos similares a los establecidos por el sistema CAMEL.

Finalmente se detallan las técnicas estadísticas a emplear, una combinación de modelos de respuesta binaria con modelos basados en análisis de supervivencia, técnicas que han sido seleccionadas por observarse un uso mayoritario en la mayor parte de las publicaciones revisadas.

La segunda parte, el marco empírico, se inicia en el capítulo cuarto, en el que se detallan los resultados obtenidos de la aplicación de la metodología descrita en el capítulo anterior.

En este sentido se iniciará con un análisis univariante de la muestra, no solo por permitir un conocimiento inicial de la realidad del sector, sino que también por servir como herramienta para determinar la existencia de observaciones atípicas que pudieran estar influenciando negativamente en la calidad predictiva de los modelos buscados.

Dentro del análisis univariante se incluye un análisis factorial por dos motivos principales:

- a. Observar si las variables explicativas pueden ser reducidas a un conjunto de componentes principales que mantengan el poder explicativo y que podrían ser utilizados en la estimación de los modelos descritos.
- b. Seleccionar las variables explicativas que presentan una mayor correlación con los componentes principales como método alternativo para estimar los modelos predictivos.

En este sentido uno de los aspectos fundamentales que condicionan la calidad de las predicciones de un modelo estadístico lo constituye el propio método de selección de las variables que forman parte del mismo.

Es habitual que esta selección se realice en función del grado de significación aportada por la variable considerada, no obstante se pretende observar si utilizando un método de selección basado en los componentes principales se consigue mejorar la calidad predictiva de los modelos.

El capítulo cuarto finalizará con la exposición de los resultados obtenidos de la estimación de los modelos a partir de las técnicas logit y regresión de Cox, comparándolos en función de la bondad de su clasificación a partir del Error Tipo I, consistente en clasificar a una

entidad operativa dentro del grupo de entidades insolventes, y el Error Tipo II, contrario al anterior, clasificar a una entidad insolvente dentro del grupo de las saneadas.

Surge la cuestión de si se podría llegar a mejores resultados a partir de la construcción de un indicador sintético, ya que existen autores que critican el uso de modelos probabilísticos binomiales por obviar que la insolvencia no es un proceso estático, sino que es el resultado de un proceso continuo de deterioro.

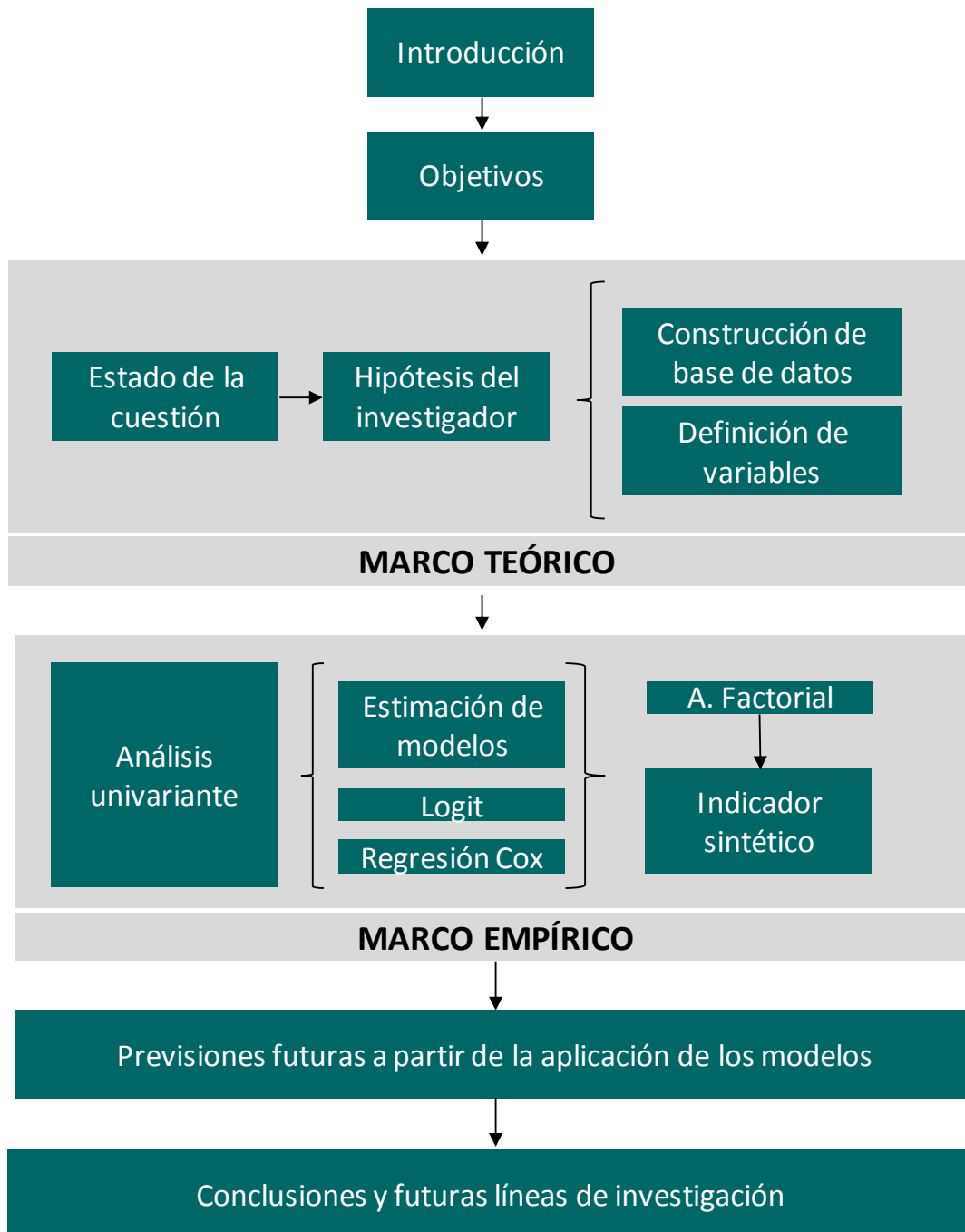
Por ello en el capítulo quinto se construirá un indicador sintético que tratará de medir la salud financiera de una Caja Rural a partir de un conjunto de variables explicativas, para la posterior comparación de los resultados con los obtenidos en el capítulo anterior tras la estimación de los modelos probabilísticos.

Se cerrará con un capítulo sexto que se analiza la situación actual del sector a partir de la aplicación de los modelos anteriores y un capítulo final en el que se recapitulan las conclusiones alcanzadas.

1.5 SINOPSIS

Para facilitar el seguimiento de la Tesis al lector, en la Ilustración 1 se ha recogido la estructura, diferenciando entre el marco teórico y el marco empírico.

ILUSTRACIÓN 1: SINÓPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

CAPÍTULO SEGUNDO:

ESTADO DE LA CUESTIÓN

2.1 SINÓPSIS DEL CAPÍTULO

A través de este capítulo se realiza una recopilación y descripción de las principales publicaciones existentes sobre el estudio de la quiebra de cooperativas de crédito.

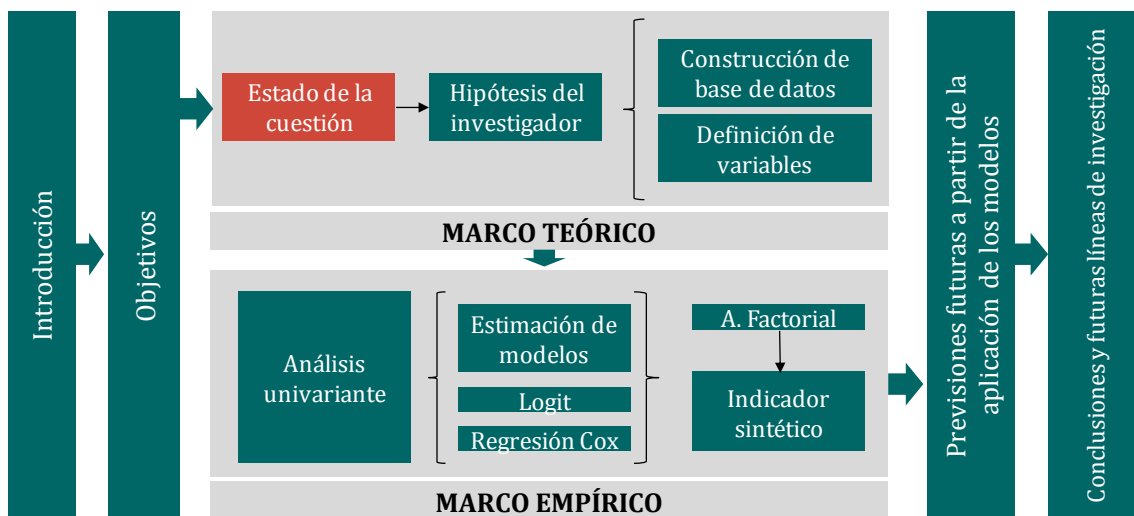
Para ello se comienza con una breve reseña sobre la investigación en materia de insolvencia empresarial, precursora de las investigaciones sobre insolvencia bancaria.

Se continúa profundizando en el estudio de la quiebra de entidades financieras, con especial mención a las técnicas empleadas y a los resultados obtenidos, principalmente en el estudio de la banca en España.

En la última parte de este capítulo se lleva a cabo una descripción detallada de la investigación académica en materia de insolvencia de cooperativas de crédito en un enfoque internacional, dado que no han sido localizadas investigaciones sobre insolvencias de cooperativas de crédito españolas.

Para facilitar el seguimiento de la Tesis, en la Ilustración 2 se refleja la sinopsis de la Tesis.

ILUSTRACIÓN 2: UBICACIÓN DEL CAPÍTULO SEGUNDO DENTRO DE LA SINOPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

2.2 METODOLOGÍA SEGUIDA EN LA ELABORACIÓN DE LA REVISIÓN BIBLIOGRÁFICA.

Para la localización de los documentos bibliográficos se ha recurrido a diversas fuentes documentales, en las que se han realizado búsquedas recurriendo a los siguientes descriptores: *Cajas Rurales, Cooperativas de Crédito, Insolvencia, Solvencia, Quiebra, Logit, Datos de Panel, Modelo de Cox, Análisis de Supervivencia, Indicadores sintéticos.*

En el caso de bases de datos anglosajonas, los descriptores a utilizar han sido equivalentes en su traducción inglesa: *Credit Unions, Solvency, Insolvency, Bankruptcy, Cox Proportional Hazard Model, Logit, Panel Data, Survival Analysis, Synthetic indicators.*

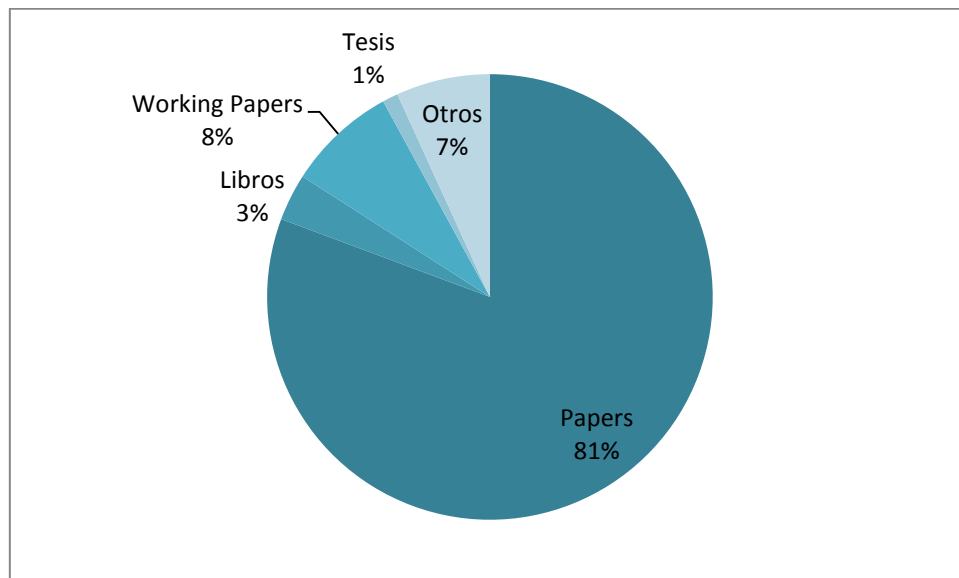
Para la elaboración de la revisión bibliográfica se han recurrido a las siguientes fuentes documentales, definidas por la Biblioteca de la Universidad Pontificia Comillas como:

- a) *Business Source Complete*: recoge artículos de las principales publicaciones académicas que versan sobre temas empresariales desde 1986.
- b) *EconLit*: el servicio de biblioteca de la Universidad Pontificia Comillas indica que esta base de datos es propiedad de la American Economic Association. Contiene algo más de 1.010.900 registros (desde 1969 hasta la actualidad) que cubren casi todas las áreas relacionadas con la economía.
- c) *Academic Search Complete*: base de datos multidisciplinar de EbscoHost. Contiene el texto completo de más de 4 600 publicaciones periódicas. Además recoge índices, resúmenes de más de 8.200 revistas y 7.200 sumarios y abstracts de "peer-reviewed". Se actualiza diariamente.
- d) *E-Journals*: Base de datos que, a través de EBSCO, da acceso a artículos de cientos de publicaciones electrónicas de carácter financiero.
- e) *Google Académico*: es el buscador de Google enfocado al mundo académico, especializado en literatura científica y académica.

- f) Biblioteca de Banco de España.
- g) Unión Nacional de Cooperativas de Crédito.
- h) Biblioteca de la Universidad Pontificia Comillas.
- i) Biblioteca de la Facultad de Estadística de la Universidad Complutense de Madrid.

Para la realización de esta Tesis se ha revisado un importante volumen de publicaciones profesionales y académicas, de las que se ha seleccionado un total de 108 documentos correspondientes a publicaciones en *journals*, *working papers*, tesis doctorales y documentos técnicos elaborados por bancos centrales, como se desglosa en la Ilustración 3.

ILUSTRACIÓN 3: TIPOS DE DOCUMENTOS UTILIZADOS EN LA REVISIÓN BIBLIOGRÁFICA

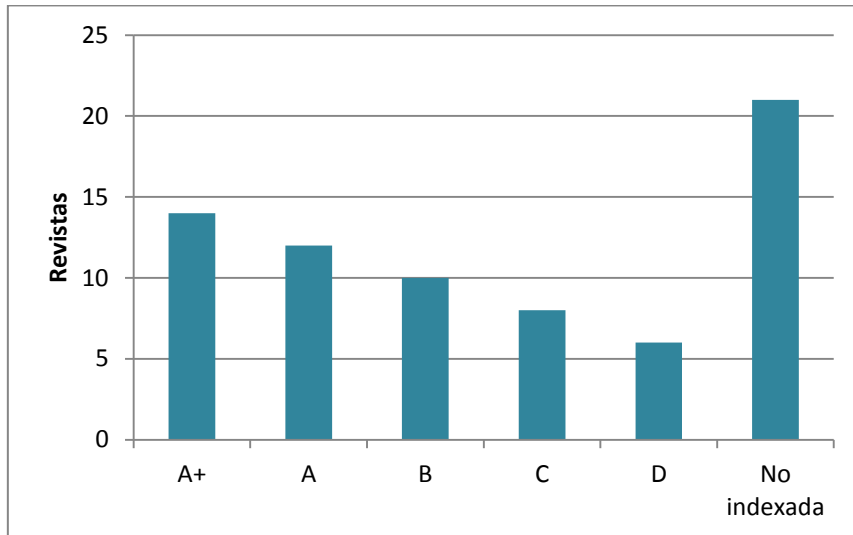


FUENTE: ELABORACIÓN PROPIA

En la Ilustración 4 se ha recogido la clasificación de las revistas utilizadas en función del cuartil de las categorías de JCR en el que se posicionan. En este sentido se puede observar que el 20% de las revistas se encuentran dentro del primer cuartil de las categorías de JCR (A+), el 17% dentro del primer cuartil de SCOPUS (A) y posiciones estables de JCR

(excluyendo el cuarto cuartil); y el 14% en el cuarto cuartil de JCR y en el segundo, tercer y cuarto cuartil de SCOPUS (B).

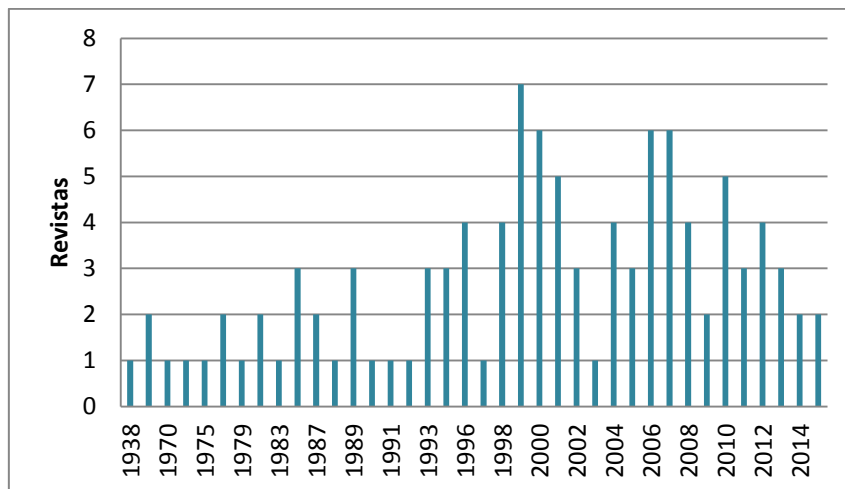
ILUSTRACIÓN 4: NÚMERO DE REVISTAS EN FUNCIÓN DE SU CALIFICACIÓN JCR Y SCOPUS



FUENTE: ELABORACIÓN PROPIA

En la Ilustración 5 se incluye una distribución temporal de las publicaciones utilizadas, observándose que la mayor parte corresponden a la primera década del siglo XXI, si bien se incluyen algunas publicaciones anteriores por su relevancia.

ILUSTRACIÓN 5: DISTRIBUCIÓN DE LAS PUBLICACIONES UTILIZADAS POR AÑOS



FUENTE: ELABORACIÓN PROPIA

2.3 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA EMPRESARIAL.

Con esta Tesis no se pretende llevar a cabo una exhaustiva revisión bibliográfica sobre las insolvencias empresariales, no solo porque ya existen excelentes trabajos que incluyen completas recopilaciones de las principales publicaciones existentes⁷, sino que también porque el objetivo buscado no es otro que el de determinar si las fusiones de Cajas Rurales están directamente relacionadas con la sostenibilidad financiera de las entidades implicadas.

Sin embargo se podría decir que el origen de la investigación sobre quiebras bancarias se encuentra muy vinculada a las publicaciones sobre insolvencia empresarial⁸ realizadas por Beaver (1967) y Altman (1968), en las que a partir del uso de ratios financieras que eran comparadas con los valores que los autores consideraban normales, pudieron confirmar que las quiebras de corporaciones empresariales en Estados Unidos estaban relacionadas con una baja rentabilidad, mayor deuda y problemas de liquidez.

Por ello se ha creído conveniente llevar a cabo en este epígrafe una breve descripción de las líneas de investigación seguidas sobre insolvencias empresariales, ya que servirán como base para el posterior desarrollo del estudio de la quiebra de entidades financieras.

Una de las principales dificultades a las que se enfrenta el estudio de la quiebra empresarial, y que también afecta a la insolvencia bancaria, es la ausencia de una teoría formal que permita explicar el fracaso empresarial (Tascón *et al*, 2011), por ello los diferentes investigadores han tratado de explicar estos acontecimientos a partir del

⁷ Para mayor información, en las publicaciones de Tascón *et al*. (2011), Caballo (2011) y Laguillo (2015).se encuentra una excelente recopilación de los trabajos existentes en esta materia.

⁸ Redondo *et al* (2014) se refiere en esta línea al nacimiento de la investigación sobre insolvencias bancarias.

empleo de ratios contables (que miden principalmente rentabilidad y endeudamiento) previamente utilizados en otros estudios de la misma índole, con la que estiman modelos estadísticos que permiten discriminar entre empresas sanas e insolventes.

En este sentido, en la Tabla 1 se han recopilado las principales investigaciones encontradas en el ámbito de la insolvencia empresarial, clasificadas en función de la técnica estadística empleada para su elaboración.

TABLA 1: PRINCIPALES PUBLICACIONES SOBRE QUIEBRA EMPRESARIAL CLASIFICADAS EN FUNCIÓN DE LA TÉCNICA EMPLEADA.

TÉCNICA	PUBLICACIONES
ANÁLISIS UNIVARIANTE Y DISCRIMINANTE	Beaver (1967), Altman (1968), Deakin (1972), Emister (1972), Deakin (1972), Blum (1974), Appetiti (1984), Gombola <i>et al.</i> (1987), Hing-ling Lau (1987), Emister (1988), Aziz <i>et al.</i> (1988), Koh y Killough (1990), Johnsen y Melicher (1994), Altman (1994), Yi-Chung <i>et al.</i> (2005), Dietrich <i>et al.</i> (2005), Lin (2009), Tzong-Huei (2009), Susicky (2011), Vavrina <i>et al.</i> (2013), Sangjae y Wu Sing (2013)
ANÁLISIS LOGIT / PROBIT	Martin (1977), Ohlson (1980), Harmer (1983), Zavgren (1985), Casey y Bartczak (1985), Aziz <i>et al.</i> (1988), Gilbert <i>et al.</i> (1990), Plat <i>et al.</i> (1991), Fletcher y Goss (1993), Pope (1995), Charalambous <i>et al.</i> (2000), Yi-Chung <i>et al.</i> (2005), Tzong-Huei (2009), Shuk-Wern <i>et al.</i> (2011), Chen <i>et al.</i> (2011), Vavrina <i>et al.</i> (2013)
MODELOS DE TRANSICIÓN Y SUPERVIVENCIA	Lane <i>et al.</i> (1986), Luoma and Laitinen (1991), Honjo (2000), Shumway (2001), Chava <i>et al.</i> (2004), Beaver <i>et al.</i> (2004), Buhler <i>et al.</i> (2005), Mínguez (2006), He y Kamath (2006), Treewichayapong <i>et al.</i> (2011)
REDES NEURONALES Y OTRAS TÉCNICAS	Casey (1980), Tam and Kiang (1990), Odom y Sharda (1990), Coats and Fant (1991), Tam (1991), Tam and Kiang (1992), Altman <i>et al.</i> (1994), Libby (1995), Yi-Chung <i>et al.</i> (2005), Tzong-Huei (2009), Chen <i>et al.</i> (2011), Kwak <i>et al.</i> (2012), Sangjae y Wu Sing (2013), Vavrina <i>et al.</i> (2013).

FUENTE: ELABORACIÓN PROPIA A PARTIR DE TASCÓN *ET AL.* (2011), CABALLO (2011) Y LAGUILLO (2015).

Las primeras técnicas se basaban en la aplicación de análisis univariante, es decir en el uso de pruebas como el análisis de la varianza (Tascón *et al.*, 2011) con la que trataron de encontrar una explicación al fracaso empresarial a partir del uso de una serie de ratios contables, si bien esta metodología no permitía realizar predicciones.

Posteriormente se dio paso al análisis discriminante lineal multivariante, técnica consistente en la estimación de una regresión que clasifica a los individuos a partir de la puntuación obtenida de un conjunto de variables y de unos niveles de corte. Un ejemplo de

aplicación de esta metodología es el modelo Z-Score de Altman *et al* (1977), que desarrolla una función en la que a partir de cinco ratios obtiene una puntuación Z que es comparada con el valor de corte establecida por el autor, lo que le permite diferenciar entre empresas solventes e insolventes.

Sin embargo, el análisis discriminante ha sido criticado por basarse en dos supuestos restrictivos como son la existencia de una relación lineal entre las variables y la hipótesis de normalidad de las ratios contables, de ahí que autores como Martin (1977) propusiesen el uso de regresiones logísticas (logit) para solucionar este problema.

La técnica logit es un tipo de regresión logística que permite calcular la probabilidad de que una empresa forma parte del grupo de las sanas o de las insolventes, ya que se construye a partir de una variable dependiente dicotómica que toma valores en el intervalo $[0, 1]$, con supuestos menos restrictivos que las técnicas anteriores (Tascón *et al*, 2011).

Los modelos de transición o supervivencia se basan en un mecanismo de funcionamiento muy similar al de las regresiones logísticas, si bien complementan a los anteriores por permitir conocer, además de la probabilidad de supervivencia, el tiempo que transcurre desde que una empresa solvente cambia al estado de insolvente.

Esta metodología ha sido ampliamente utilizada en la ciencia sanitaria por permitir conocer la influencia que un conjunto de factores puede ejercer sobre la supervivencia de un paciente, si bien su aplicación al estudio de la insolvencia empresarial se inicia con la publicación de Lane *et al* (1986).

Finalmente, y con el desarrollo de la informática y la mejora del software de tratamiento de datos, en la década de 1990 se comienzan a aplicar las técnicas de inteligencia artificial para la predicción de la insolvencia empresarial.

Tascón *et al* (2011) define a estas técnicas como “*programas de ordenador capaces de generar conocimiento a través del análisis de ratios, utilizando posteriormente ese conocimiento para realizar inferencias sobre nuevos datos*”, siendo pioneros los trabajos de Casey (1980), Odom *et al* (1990), Tam *et al* (1992) principalmente.

Dentro de la inteligencia artificial existen multitud de técnicas, si bien Tascón *et al.* (2011) destaca la “*teoría de los conjuntos borrosos, las redes neuronales, el aprendizaje de máquinas (árboles de decisión), la teoría de los conjuntos aproximados, las máquinas de vectores de apoyo y la inteligencia computacional*”. Para mayor información sobre las mismas, en Ravi *et al* (2008) se encuentra una excelente recapitulación de los principales trabajos existentes.

Para la evaluación de la capacidad predictiva de los modelos es habitual observar la consecución de los denominado Error Tipo I, falso positivo; y el Error Tipo II, que indicaría un falso negativo, siendo el mejor modelo predictivo aquel que sea capaz de minimizar ambos tipos de errores, aunque para el inversor el coste de incurrir en el Error Tipo II es notablemente superior al del Tipo II (Sinkey, 1975).

2.4 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA BANCARIA.

Según Thomson (1989) hasta la década de 1970 el sector bancario de EE.UU. se mantuvo prácticamente ajeno a las quiebras. Sin embargo con el agudizamiento de las crisis bancarias, sobre todo a partir de la década de 1980, el volumen de investigaciones académicas que trataban de determinar sus causas creció considerablemente por observar que estas crisis, antes de ser causa de las recesiones económicas, constituían uno de los mecanismos que las amplificaban, básicamente por estar acompañadas de crisis

cambiarías, de deuda interna y externa, y de crisis monetarias con episodios de inflación y deflación (Martin *et al*, 2012).

De hecho, como afirma Álvarez (2008) *“las crisis financieras tienen una traslación a la economía real a través de los canales de crédito, patrimonial, confianza y fiscal”*, y que el propio autor describe de la siguiente forma:

- a) El canal de crédito está relacionado con las restricciones de financiación tanto para empresas como hogares por la menor liquidez disponible en el interbancario, en cierta medida explicado por el deterioro de los niveles de solvencia de las entidades financieras y la pérdida de confianza entre ellas.
- b) El canal patrimonial se fundamenta en la reducción del consumo y de la inversión de empresas y hogares, lo que afecta negativamente al crecimiento económico.
- c) El canal de confianza se resume básicamente en la pérdida de confianza de los agentes económicos, lo que nuevamente afecta negativamente a sus decisiones de inversión y por tanto potencia la desaceleración económica.
- d) El canal fiscal muy relacionado con la necesidad de recursos públicos para el rescate de entidades financieras problemáticas, lo que en última instancia afecta al contribuyente por el hecho de que los gobiernos se ven obligados a destinar partidas presupuestarias financiadas con los recursos fiscales, partidas que por tanto dejarán de estar disponibles para otras finalidades públicas.

Gama *et al* (2011) afirma que conocer la situación financiera de una entidad no solo preocupa a sus accionistas, sino que también importa a los gobiernos, las empresas y las personas que mantienen sus recursos en esas entidades.

2.4.1. Definición de insolvencia bancaria

Al igual que ocurre en la quiebra empresarial, no existe un consenso mayoritario sobre el momento en el que un banco entra en situación de insolvencia ya que en muchas ocasiones depende de la legislación existente en cada uno de los países, línea sobre la que se incluiría Sinkey (1975) en su definición de entidad problemática como aquella que al incumplir con alguna ley o regulación, presenta incertidumbres sobre su solvencia futura.

Muy relacionada con la definición anterior estaría Laffarga *et al* (1991) quien en un análisis de la situación de las entidades en España considera que la quiebra bancaria llega con la intervención de la entidad por parte de Banco de España a través de los mecanismos dispuestos para ello. Matias *et al* (1996) y Logan (2001) defienden la misma postura en el estudio de las entidades brasileñas e inglesas, respectivamente.

Otro grupo de autores definen la quiebra bancaria acudiendo a la situación financiera de la propia entidad. Entre ellos Janot (2009), que considera que la insolvencia se alcanza cuando el patrimonio neto se vuelve negativo, o Gama *et al* (2011) que incide en la importancia del coeficiente de liquidez en la evolución futura de la entidad.

Una mezcla de los casos anteriores sería la expuesta por Andersen (2008), quien considera que la insolvencia de un banco llega cuando se produce su liquidación, su fusión o adquisición; o si el valor tomado por el TIER1⁹ se sitúa por debajo del 8%.

⁹ TIER1 es una medida del capital adoptada en los Acuerdos de Basilea (mayor información en nota 19).

2.4.2. Revisión de las técnicas utilizadas para la construcción de modelos de previsión de insolvencia bancaria.

Podría decirse que el estudio de la quiebra bancaria se inició en EE.UU. por la necesidad del regulador de encontrar un sistema que permitiese evaluar la solvencia para controlar que no se tomaban posiciones arriesgadas que pudieran amenazar la integridad del Sistema.

En este sentido Swicegood *et al.* (2001) señalan la existencia de dos tipos de evaluaciones, las denominadas *on-site*, que inicialmente eran realizadas cada 12-18 meses, y que ya Cole *et al.* (1995a) había descrito anteriormente como inspecciones *in situ* para evaluar, además de su situación financiera, el cumplimiento de la legislación vigente, la capacidad del equipo gestor y el sistema de control interno; y las evaluaciones *off-site*, realizadas mediante un sistema de alerta temprana enfocado a predecir la evolución de esa entidad a partir de información económica y financiera que periódicamente era suministrada al regulador, permitiendo anticipar las actuaciones que se hiciesen necesarias sobre aquellas entidades que acusasen los primeros síntomas de inestabilidad. Estos sistemas *off-site* fueron evolucionando a medida que la informática y la estadística se desarrollaron.

Para la ejecución de las evaluaciones *on-site*, las agencias reguladoras¹⁰ de los Estados Unidos crearon en 1979 el denominado Sistema CAMEL¹¹, que clasificaba a las entidades en función de puntuaciones obtenidas tras analizar la solvencia, calidad de activos, equipo directivo, rentabilidad y liquidez.

¹⁰ En EE.UU. existen tres agencias reguladoras: la Reserva Federal, la *Federal Deposit Insurance Corporation* (FDIC) y la *Currency Administrator of National Banks* (OCC).

¹¹ Las siglas CAMEL se refieren a *Capital protection, Assets quality, Management competence, Earnings strenght, Liquidity risk*.

El Sistema CAMEL presentaba el inconveniente de que únicamente reflejaba la situación financiera en un momento determinado y no permitía efectuar predicciones sobre la evolución futura (Gaytan *et al.*, 2002).

De hecho, Cole *et al.* (1995a) y Hirtle *et al.* (1999) en estudios anteriores observaron que las calificaciones asignadas por el Sistema CAMEL carecían de validez transcurridos seis o doce trimestres desde la fecha de cierre de los estados financieros, por lo que teniendo en cuenta que la información de las entidades se hacía pública trascurrido un importante lapso de tiempo, el cálculo del CAMEL por sí mismo carecía de validez para la toma de decisiones a largo plazo.

En el desarrollo de sistemas *off-site* en EE.UU., Cole *et al.* (1995a) señalan tres hitos:

- a. La creación en la década de 1980 del *Uniform Bank Surveillance Screen* (en adelante *UBSS*) por parte de la Reserva Federal estadounidense, y que estuvo vigente hasta 1.993.

Putnam (1983) realiza una descripción de su funcionamiento, si bien simplemente destacar que este sistema utilizaba la información financiera disponible para tratar de identificar aquellas entidades cuyas ratios se habían deteriorado en comparación con el grupo de comparación.

- b. La puesta en marcha del CAEL (*Capital, Equity, Assets y Leverage*) en la misma época que el *UBSS*, por la *Federal Deposit Insurance Corporation* (en adelante *FDIC*). Este sistema mide los mismos inputs que el CAMEL con la excepción de la labor desempeñada por el equipo gestor.
- c. En 1993 se pone en marcha el *Financial Insitution Monitoring System* (en adelante *FIMS*), que buscaba corregir los errores de los sistemas *off-site* anteriores, y que

según Cole *et al.* (1995a) no eran otros que la selección de las ratios y el peso de cada una dentro del modelo de predicción.

Si bien en Cole *et al.* (1995a) se realiza una buena descripción del funcionamiento de este sistema, tan solo destacar que *FIMS* consta de dos de dos partes, el *FIMS Rating*, que evalúa la situación actual de la entidad, y el *FIMS Risk Rank*, que realiza predicciones sobre la situación futura de la entidad a partir de información financiera trimestral de dos años de antigüedad. Para determinar los inputs de ambas partes y el peso que cada variable explicativa tiene en el modelo, el *FIMS* realiza diversos test financieros que substituyen a los criterios subjetivos de los procedimientos *off-site* anteriores.

En Europa la creación de los sistemas *off-site* es algo más tardía, básicamente en la década de 1990. Entre otros, Sahajwala *et al.* (2000) destaca la fundación en 1993 del PATROL en Italia; en 1997 del *Organization and Reinforcement of Preventive Action* (ORAP) y el *Support System for Banking Analysis* (SAABA) en Francia y el *Bakred Information System* (BAKIS) en Alemania; en 1998 el *Risk Assessment Tools of Supervision and Evaluation* (RATE) en Reino Unido; y en 1999 el *Risk Analysis Support Tool* (RAST) en Holanda. Para mayor información sobre los métodos estadísticos y los inputs utilizados en cada uno de los sistemas *off-site* europeos, en Sahajwala *et al.* (2000) encontramos una excelente descripción.

Desde un punto de vista académico, la bibliografía que versa sobre el estudio y predicción de la insolvencia bancaria es amplia, aunque el origen de todas ellas se encuentra en las publicaciones de Secrist (1938) y Meyer *et al.* (1970) relativa a la insolvencia de instituciones financieras en los EE.UU.

En la Tabla 2 se recogen las principales publicaciones encontradas en relación a la insolvencia de entidades financieras, clasificadas en función de la técnica estadística

utilizada, observándose un uso mayoritario de las técnicas basadas en modelos de respuesta binaria (logit/probit) entre el conjunto de investigadores.

TABLA 2: PRINCIPALES PUBLICACIONES SOBRE LA QUIEBRA DE ENTIDADES FINANCIERAS (EXCEPTO COOPERATIVAS DE CRÉDITO). CLASIFICACIÓN POR METODOLOGÍA.

TÉCNICA	PUBLICACIONES
Análisis discriminante ¹²	Secrist (1938); Altman (1968); Meyer <i>et al.</i> (1970); Stuhr <i>et al.</i> (1974); Sinkey (1975); Sinkey (1978); Ohlson (1980), Rose <i>et al.</i> (1985), Sriram <i>et al.</i> (1996), Ahumada <i>et al.</i> (1999), Swicegood <i>et al.</i> (2001), Sajter (2005), Aziz <i>et al.</i> (2006), Doganay <i>et al.</i> (2006), Ozkan <i>et al.</i> (2007), Pasiouras <i>et al.</i> (2007), Gaganis <i>et al.</i> (2008), Guimaraes <i>et al.</i> (2008), Pasiouras <i>et al.</i> (2009), Gaganis <i>et al.</i> (2010), Jordan <i>et al.</i> (2010), Crespo (2011),
Logit y Probit	Sinkey (1975), Martin (1977); Hanweck (1977), Ohlson (1980), Maddala (1983), Bovenzi <i>et al.</i> (1983); West (1985) Pantalone <i>et al.</i> (1987); Abrams <i>et al.</i> (1987), Kolari <i>et al.</i> (1987), Cox <i>et al.</i> (1989), Belongia <i>et al.</i> (1990), Espahbodi (1991), Whalem (1991), Thompson (1992); Cramer (1991), Greene (1993), Cole <i>et al.</i> (1995a y 1995b), Borges <i>et al.</i> (1996); Siriam <i>et al.</i> (1996), Frankel <i>et al.</i> (1996), Cole <i>et al.</i> (1998), Demirgüç <i>et al.</i> (1998); Hardy <i>et al.</i> (1999); Hutchinson <i>et al.</i> (1999); Glick <i>et al.</i> (1999); Demirgüç <i>et al.</i> (1999); Estrella <i>et al.</i> (2000), Logan (2001), Maghalanes (2001), Kolari <i>et al.</i> (2002), , Godlewski (2003), Sajter (2005), Montgomery <i>et al.</i> (2005), Distinguin <i>et al.</i> (2006), Pasiouras <i>et al.</i> (2007), Pasiouras <i>et al.</i> (2007), Distinguin <i>et al.</i> (2008), Gaganis <i>et al.</i> (2008), Gaganis <i>et al.</i> (2010), Tatom <i>et al.</i> (2011), Poghosyan <i>et al.</i> (2011); Karminsky <i>et al.</i> (2012), Shafer (2012), DeYoung (2013)
Modelos de Supervivencia	Lane <i>et al.</i> (1986), Whalen (1991), Henebry (1997), Anastasi (1998); Wheelock <i>et al.</i> (2000); Magalhaes (2001), Molina (2002), Meyers <i>et al.</i> (2012),
Redes Neuronales y otros	Odom <i>et al.</i> (1990), Espahbodi (1991), Tam <i>et al.</i> (1992); Salchenberger <i>et al.</i> (1992); Serrano (1993); kumar <i>et al.</i> (1995); Bell (1997), Swicegood <i>et al.</i> (2001), Ravi <i>et al.</i> (2008), Boyacioglu <i>et al.</i> (2009), Etemadi <i>et al.</i> (2009), Messai <i>et al.</i> (2015).

FUENTE: ELABORACIÓN PROPIA

Aunque los primeros estudios se basaban en el empleo de técnicas univariantes a partir del cálculo de ratios financieras para un conjunto de entidades que posteriormente eran comparados con la tendencia general, Meyer *et al.* (1970) es pionero en el cálculo de regresiones lineales para la estimación de la insolvencia de entidades financieras.

Para ello el autor estimó una regresión lineal a partir de nueve variables económicas y financieras de una treintena de entidades financieras insolventes que posteriormente fueron emparejadas con similar número de entidades solventes. Con todo ello el autor

¹² Dentro de esta denominación se incluyen técnicas de análisis univariante (como por ejemplo el análisis de la varianza de los dos grupos definidos por la variable dependiente, es decir, entidades solventes e insolventes; como técnicas basadas en el cálculo de regresiones lineales

consiguió estimar un modelo que clasificó correctamente a más del 80% de las entidades solventes e insolventes.

Esta línea de investigación se ha mantenido hasta fechas recientes, siendo destacables los trabajos de Swicegood *et al.* (2001) y Crespo (2011), en los que analizan la efectividad de la regresión lineal en la estimación de la insolvencia bancaria.

Swicegood *et al.* (2001) analizaron la capacidad de las regresiones lineales, redes neuronales y los sistemas empleados por los supervisores bancarios para prever las insolvencias bancarias, confirmando la mayor efectividad de las redes neuronales y la menor capacidad de las regresiones lineales.

A conclusiones similares llegó Crespo (2011) en su análisis de la banca venezolana mediante la comparación de la efectividad del CAMEL y la regresión lineal.

De hecho, Laffarga *et al.* (1991) ya puso de manifiesto que la aplicación de regresiones lineales para la previsión de insolvencias bancarias presenta el inconveniente de asumir que las variables siguen una distribución normal en las dos poblaciones que define la variable dependiente, hipótesis difícil de sustentar en este tipo de problemas en los que la variable dependiente es de tipo cualitativo.

Por ello se ha generalizado el empleo de técnicas logit y probit, ya que además de evitar los errores de las regresiones lineales, permiten conocer la probabilidad de que una entidad pueda verse inmersa en un proceso de liquidación, ya que estas técnicas constituyen un tipo de regresión probabilística.

Podría decirse que Martin (1977) fue el pionero en el empleo de la técnica logit en el estudio de la quiebra bancaria, observando que las ratios que medían el capital, liquidez y beneficios eran las más significativas para explicar las causas, línea de investigación seguida Bovenzi *et al.* (1983), Pantalone *et al.* (1987), Thomson (1992), Borges *et al.*

(1996), Demirgüç *et al.* (1998), Hardy *et al.* (1999), Hutchinson *et al.* (1999), Glick *et al.* (1999), Demirgüç *et al.* (1999), Logan (2001) y Poghosyan *et al.* (2011), entre otros

Si bien es cierto que en la ciencia sanitaria está extendido el uso de los modelos de supervivencia para estudiar la efectividad de un tratamiento sobre una población, Lane *et al.* (1986) fue pionero en el empleo de esta herramienta para el estudio de la quiebra bancaria, estimando un modelo en el que a partir de seis ratios consiguió estimar el tiempo que transcurría desde que se producían los primeros síntomas hasta que el banco se veía inmerso en un proceso de insolvencia. Autores como Cole *et al.* (1995b), Anastasi (1998), Wheelock *et al.* (2000) y Magalhaes (2001), entre otros, continuaron empleando esta línea de investigación.

Aunque el empleo de redes neuronales en la literatura económica se remonta a White (1988) para la predicción del comportamiento de las acciones de IBM en Bolsa, no es hasta el desarrollo de la informática cuando se empieza a producir un empleo masivo de esta técnica en el estudio de la insolvencia bancaria. En este sentido uno de los primeros estudios que encontramos es el realizado por Tam (1992) para un conjunto de entidades en EE.UU. Posteriormente, autores como Salchenberger *et al.* (1992), Kumar *et al.* (1995) y Bell (1997), Swicegood *et al.* (2001), Ravi *et al.* (2008), Boyacioglu *et al.* (2009), Etemadi *et al.* (2009) o Messai *et al.* (2015), entre otros.

Siguiendo a Lopez *et al.* (2015) en la actualidad existen nuevas herramientas basadas en la inteligencia artificial, cada vez más desarrolladas, como podría ser *Random Forest* (Selvas Aleatorias), *Support Vector Machines* (Máquinas de Vectores de Soporte), *Genetic Algorithm* o técnicas híbridas (que emplearían una combinación de las anteriores). Entre otros autores encontramos a Olmeda *et al.* (1997), Canbas *et al.* (2005), Min *et al.* (2006) y Boyacioglu *et al.* (2009).

2.4.3. La investigación sobre insolvencias bancarias en España.

Aunque en España la investigación en materia de insolvencia bancaria ha tenido un desarrollo reciente, las primeras investigaciones encontradas se remontan a la crisis económica del periodo 1977 y 1985, en la que la desaparición de un elevado número de entidades financieras permitió a los investigadores disponer de una base de datos suficiente para analizar las características subyacentes en dichos procesos de insolvencia.

En este sentido uno de los primeros trabajos encontrados es el de Laffarga *et al* (1987), que compara la eficacia de la regresión lineal y de la regresión logit en el estudio de la crisis financiera de 1977-1985. Para ello se basó en cuatro ratios financieras que medían composición y tamaño del activo, capitalización y rentabilidad, con las que pudieron confirmar la mayor capacidad predictiva de los modelos logit.

Posteriormente Pina (1989), recurriendo a la misma secuencia temporal, construyó un modelo logit con datos económicos y financieros anteriores al año en el que la entidad se declaraba en situación de insolvencia (en su caso, la intervención por parte del Fondo de Garantía de Depósitos), mediante el cual pudieron confirmar que las ratios que medían la liquidez y la rentabilidad resultaban ser significativas para la previsión de la insolvencia bancaria.

En esa línea continuó Fernández (1989), observando a través de un modelo logit que las ratios financieras que medían la composición del activo y la rentabilidad resultaban ser más significativas para la determinación de la probabilidad de insolvencia durante la crisis de 1977.

Martinez *et al* (1989) utilizó datos del *Anuario Estadístico de la Banca*¹³ para la estimación de dos modelos logit, uno predictivo de la probabilidad de insolvencia y otro de control para analizar las actuaciones a corto plazo que podrían tener efecto sobre la probabilidad de insolvencia, observando que ambos modelos presentaban altos niveles de clasificación de las entidades financieras que cesaron actividad entre 1980 y 1983.

Laffarga *et al* (1991) llevaron a cabo una comparación de la capacidad del modelo de regresión lineal y el modelo logit para prever los procesos de insolvencias acaecidos en la crisis financiera de 1978 a 1983, confirmando que la regresión lineal, además de violar el supuesto de normalidad, mostraba una menor capacidad de clasificación que el modelo logit.

Con Serrano *et al.* (1993) se evoluciona en la complejidad de las técnicas usadas, ya que recurrieron al uso de redes neuronales, principalmente mapas autoorganizados, logrando un mayor porcentaje de acierto que los modelos clásicos, si bien con la desventaja de desconocer el procedimiento seguido por la red neuronal para clasificar a las distintas entidades, así como los pesos de las variables utilizadas.

De Miguel *et al* (1993) actualizan el estudio de Fernández (1989) utilizando redes neuronales, observando que este nuevo modelo propuesto proporcionaba mejores resultados, si bien por las limitaciones propias de las redes neuronales, no conocían los ratios que ejercían mayor influencia

Olmeda *et al* (1997) recurren a una muestra de bancos españoles que entre 1977 y 1985 habrían quebrado, estimando diferentes modelos de previsión de insolvencia, básicamente análisis discriminante, análisis logit, redes neuronales y árboles de propagación, a partir

¹³ Publicado anualmente por la Asociación Española de Banca (AEB), fundada en 1987 y que asocia profesionalmente a todos los bancos operativos en España (nacionales y extranjeros).

de las mismas ratios financieras utilizadas por Pina (1989), confirmando nuevamente que la red neuronal era la que ofrecía los mejores resultados.

Mar-Molinero *et al* (2001) utilizan técnicas de escalados multidimensionales para analizar las características de los bancos que se declararon insolventes en España entre 1973 y 1990, todo ello a partir de las mismas ratios financieras ya utilizadas por Pina (1989), confirmando la mayor significatividad de aquellas ratios vinculadas con la rentabilidad.

Sanchez (2003) construye una red neuronal para estudiar la insolvencia bancaria en España a partir de la muestra previamente utilizada por Laffarga *et al* (1986) y Serrano *et al.* (1993), con nueve ratios contables previamente reducidas mediante análisis factorial. Observaron un mayor poder de clasificación en las redes neuronales frente a los modelos basados en análisis discriminante y respuesta binaria, si bien indican que estas redes neuronales deben ser entendidas como complementarias a otros métodos tradicionales de previsión por la dificultad de interpretación de los coeficientes asociados.

Sagarra *et al* (2011) centran su estudio en el caso de las cajas de ahorro españolas en el año 2006, todo ello a partir de una serie de ratios financieras y mediante la aplicación de análisis factorial, logit, análisis envolvente de datos y escalados multidimensionales, confirmando que las variables que miden capitalización y rentabilidad son las más significativas para explicar la quiebra de este tipo de entidades financieras.

Quesada *et al.* (2011) se basan en análisis univariante de una serie de ratios financieras para analizar la situación los bancos y cajas de ahorro durante la crisis de 2006 hasta 2010, si observando que las cajas de ahorro en problemas presentaban menores niveles de liquidez que los bancos. Además de ello, las cajas de ahorro se encontraban más endeudadas, y mantenían similares ratios de consistencia, garantía y solidez, siendo los bancos los que presentan los mayores niveles de rentabilidad.

Climent (2013) trata de determinar los factores explicativos de la solvencia de bancos y cajas de ahorro entre 2009 y 2009. Para ello utilizan una serie de ratios financieros que miden diferentes aspectos de la estructura económico-financiera de las entidades implicadas, para la posterior estimación de una regresión lineal. Con todo ello confirman que la calidad de los activos, la morosidad, la dependencia de los mercados mayoristas, el tamaño y la importancia de las participadas en relación al activo, se posicionan como los principales determinantes de la solvencia de las entidades financieras en este periodo, poniendo de manifiesto marcadas diferencias entre bancos y cajas de ahorro, ya que los bancos son los que obtienen los mayores niveles de solvencia.

Gutierrez *et al.* (2014) utilizaron el análisis logit para comprobar si los resultados de las pruebas de estrés llevadas a cabo por el Banco Central Europeo podrían haber sido previstos a partir de una regresión logit, concluyendo que el modelo propuesto hubiera presentado un alto porcentaje de acierto a partir de las ratios que miden la estructura y calidad del capital de las entidades analizadas.

Redondo *et al* (2014) recurren al análisis discriminante lineal multivariante para estudiar las características de las entidades financieras (bancos, cajas de ahorros y cooperativas de crédito) con problemas en España entre 2008 y 2010. Aunque los autores verifican el no cumplimiento de las hipótesis de los modelos lineales (normalidad e igualdad en la matriz de varianzas y covarianzas) ponen de manifiesto la existencia de investigaciones anteriores que indican que aún en esos casos el análisis discriminante lineal muestra resultados satisfactorios. Con todo ello confirman la mayor significatividad en las ratios que median la calidad del activo, exposición inmobiliaria y capitalización, con un porcentaje de clasificación correcta que ronda el 88%.

Fernández (2015) estudia la crisis bancaria de 2009 hasta 2013 a partir de una serie de ratios financieras incluidas en el análisis CAMEL con el fin de determinar si estas ratios

podrían haber sido utilizadas como indicadores precoces, todo ello a partir del uso de diversas técnicas estadísticas (escalado multidimensional, técnicas paramétricas y no paramétricas) observando alto poder de estimación en todas ellas, si bien los árboles de decisión se configurarían como las más eficientes.

A modo de resumen, en Tabla 3 siguiente se han detallado las publicaciones analizadas, clasificadas en función de la técnica y ratios financieras más significativas.

TABLA 3: RESUMEN DE LAS INVESTIGACIONES EN MATERIA DE INSOLVENCIA BANCARIA EN ESPAÑA.

AUTORES	AÑO	VARIABLES SIGNIFICATIVAS	TÉCNICA
Pina	1989	(Activo circulante - Tesorería)/ Activo total Activo circulante / Pasivo exigible Beneficio neto / Neto	Logit
Fernández	1989	Inmovilizado / Activo total Fondos Propios Libres / Inv. Crediticias Resultado Neto / Activo total medio	Logit
Martínez <i>et al</i>	1989	Pasivo a corto plazo / pasivo exigible Crédito en moneda extranjera / crédito total Acreedores en moneda ext./acreedores total Caja / pasivo a corto plazo Fondos propios / activos ponderados Fondos propios / activo total Inmovilizado / activo Beneficio neto / activo total	Logit
De Miguel <i>et al</i>	1993	No especificados	Redes neuronales
Serrano <i>et al</i>	1993	No especificados	Redes neuronales
Olmeda <i>et al</i>	1997	Rentabilidad	Varias
Sánchez	2003	Nueve ratios financieras reducidas mediante análisis factorial	Redes neuronales
Sagarra <i>et al.</i>	2011	Rentabilidad y capitalización	Varias
Quesada <i>et al</i>	2011	Liquidez, endeudamiento y rentabilidad	Análisis discriminante
Climent	2013	Morosidad y acceso a interbancario	Análisis discriminante
Gutiérrez <i>et al.</i>	2014	Capitalización	Logit
Redondo <i>et al.</i>	2014	Activo, exposición inmobiliaria y capitalización	Análisis multivariante
Fernández	2015	No especificadas	Varias

FUENTE: ELABORACIÓN PROPIA

2.5 PRINCIPALES PUBLICACIONES EN MATERIA DE INSOLVENCIA DE COOPERATIVAS DE CRÉDITO.

El volumen de investigaciones relacionadas con el estudio y la predicción de la insolvencia de cooperativas de crédito es reducido.

En España tan solo ha sido encontrada la publicación de Redondo *et al* (2014) en la que estudia, de forma agregada junto al resto de los integrantes del sector financiero, los factores asociados a los procesos de insolvencias de estas entidades entre 2008 y 2010, no habiendo sido localizada alguna otra investigación centrada en exclusiva en materia de cooperativas de crédito. Por el contrario fuera de España, aun siendo limitadas, se encuentran investigaciones en EE.UU. (Simon, 1980), Argentina (Dabos, 1996), Canadá (Pille, 1998), Brasil (Gama *et al.*, 2004; Braga *et al.*, 2006; Gama *et al.*, 2011; y Carvalho *et al.*, 2015); Alemania (Porath, 2006), Portugal (Cabo *et al.*, 2010; y Lima, 2012) e Italia (Maggiolini *et al.*, 2005; y Fiordelisi *et al.*, 2013).

2.5.1 Definición de la insolvencia de cooperativas de crédito

En el caso de las cooperativas de crédito la definición de quiebra es un poco más compleja que la ya expuesta para el caso de las entidades financieras, ya que en muchas ocasiones las cooperativas con problemas son asistidas por otros miembros, bien mediante fusiones y adquisiciones, o bien a través de ayudas que pueden ocultar la verdadera realidad financiera de la institución.

Maggiolini *et al.* (2005) considera que la quiebra de una cooperativa de crédito se produce cuando el regulador ordena su liquidación o bien se autoriza la fusión con otra entidad.

Para Porath (2006), y en línea con la definición anterior, la insolvencia de la cooperativa de crédito llega cuando se hace necesaria la intervención del regulador o del auditor; o bien, cuando la entidad presenta pérdidas superiores al 25% del capital, incluyendo reestructuraciones por fusiones.

Para Cabo *et al.* (2010) una cooperativa se encuentra en situación de insolvencia cuando se fusiona con otra cooperativa, posición que Lima (2012) defiende y complementa con la circunstancia de percibir asistencia financiera del regulador o de otros miembros.

En la Tabla 4 se ha recogido una recapitulación de las definiciones dadas sobre el concepto de insolvencia por el conjunto de investigadores, observándose que en la mayor parte de los casos se los distintos autores hacen referencia la intervención por parte del Regulador o de cualquier otro organismo superior.

TABLA 4: REVISIÓN BIBLIOGRÁFICA DE LA DEFINICIÓN DE INSOLVENCIA DE UNA COOPERATIVAS DE CRÉDITO.

AUTORES	AÑOS	DEFINICIÓN DE INSOLVENCIA
Porath	2006	Intervención por parte del regulador o auditor, en base a normativa legal pero incluyendo reestructuraciones por fusiones. Pérdidas representativas del 25% del capital.
Gama <i>et al.</i>	2004	Consideración de entidades insolventes por el CREDIMINAS, órgano responsable del control de las cooperativas de crédito rural del Estado de Minas Gerais (Brasil).
Dabós	1996	Cooperativas de crédito suspendidas, colocadas en el fondo fiduciario o absorbidas.
Gama <i>et al.</i>	2011	Toda cooperativa que dejó de enviar balances al Banco de Brasil, presentaba patrimonio neto negativo, o cumplía con alguna de las condiciones establecidas por el Banco Central de Brasil para su cese: 1) Suspendida 2) En liquidación extrajudicial. 3) En liquidación ordinaria
Lima	2012	Aquellas cooperativas de crédito que hayan recibido asistencia financiera por parte de los fondos de asistencia de Portugal.
Cabo <i>et al.</i>	2010	Considera a la entidad insolvente si: 1) El patrimonio neto es negativo 2) Intervención del regulador o de la patronal del sector. 3) Integrada en otra cooperativa de crédito. 4) Declaración de oficial de insolvencia.

(Continuación Tabla 4)

AUTORES	AÑOS	DEFINICIÓN DE INSOLVENCIA
Braga <i>et al.</i>	2006	Considera insolvente a aquellas cooperativas de crédito brasileñas que están obligadas a mantener un 35% de sus depósitos retenidos en cumplimiento de una resolución del Banco de Brasil.
Fiordelisi <i>et al.</i>	2013	Si se produce la intervención o si se produce su liquidación.
Carvalho <i>et al.</i>	2015	Considera los siguiente eventos: 1) Cierre 2) Suspensión 3) Liquidación (judicial y extrajudicial).

FUENTE: ELABORACIÓN PROPIA

2.5.2 Revisión de las técnicas utilizadas para la construcción de modelos de previsión de insolvencia de cooperativas de crédito.

Se podría considerar a Dabós (1996) como pionero en aplicar regresiones para la determinación de las causas asociada a la insolvencia de cooperativas de crédito en Argentina, si bien es cierto que este propio autor hace referencia a la existencia de una investigación anterior (Simon, 1980) que ya hizo uso de esta misma técnica.

Concretamente Dabós (1996) centró su investigación en Argentina entre 1994 y 1995, años en los que observó la quiebra de nueve cooperativas de crédito sobre las treinta y ocho existentes. Combinando un modelo probit con ratios definidos por el sistema *CAMEL*, y utilizando información económica y financiera del año inmediatamente anterior al de la declaración de insolvencia, consiguió estimar un modelo que clasificó correctamente al 89% de las entidades quebradas y no quebradas.

Por su parte, Gama *et al.* (2004) utilizaron un modelo logit para el estudio de la insolvencia de cooperativas de crédito en Brasil entre 1998 y 2001, disponiendo de una muestra de 103 entidades de las que 11 cumplieron con la condición de insolvencia considerada por los autores. A partir de una serie de variables previamente utilizadas en el ámbito

empresarial, observaron la significatividad de las variables que medían la capitalización, liquidez y crecimiento del negocio, construyendo un modelo que consiguió clasificar correctamente al 81,82% de las cooperativas insolventes y al 98,91% de las cooperativas solventes.

Posteriormente estos mismos autores afinaron el estudio anterior con la publicación de Gama *et al.* (2011), manteniendo la técnica, pero incrementando tanto la muestra (hasta las 112 cooperativas) como el periodo de estudio (desde 1995 hasta 2008). En esta ocasión los inputs del modelo provinieron de los indicadores financieros propuesto por el *World Council of Credit Unions*¹⁴, conocido como *Sistema Pearls*¹⁵. Sin embargo, los resultados obtenidos fueron ligeramente inferiores a los de Gama *et al.* (2004), ya que el nuevo modelo tan solo clasificó correctamente al 72,32% de las cooperativas insolventes y al 96,14% de las solventes.

Por su parte, y aun manteniendo el uso de la técnica logit, Lima (2012) enfocó su estudio en las cooperativas de crédito portuguesas entre 1999 y 2006. Para ello recurrió al *CAMEL* y utilizó información financiera de los dos años anteriores a la quiebra, además de variables macroeconómicas que medían PIB y tasa de paro, observando que tan solo resultaron ser significativas las ratios que medían calidad de los activos, eficiencia del equipo gestor (gastos operativos / ingresos brutos) y la dimensión (a mayor tamaño, menor probabilidad de quiebra), careciendo de importancia los inputs macroeconómicos. Con todo ello, su modelo clasificó correctamente la totalidad de las cooperativas de crédito analizadas.

¹⁴El World Council of Credit Unions (WOCCU), fundado en 1971, es una asociación internacional de cooperativas de crédito, algo más de 57.000 asociadas presente en 103 países, que busca el desarrollo de sus miembros.

¹⁵ El Sistema PEARL es una adaptación del CAMEL al entorno de las cooperativas de crédito. Consta de una serie de indicadores contables y financieros utilizados por el WOCCU desde 1990 que evalúan las siguientes áreas de las cooperativas de crédito: Protección, Estructura Financiera, Calidad de los Activos, Rentabilidad, Liquidez y Crecimiento.

Porath (2006) compara la situación de las cooperativas de crédito con la mostrada por las cajas de ahorro alemanas desde 1993 hasta 2002. Para ello estudió las causas que explicaban la quiebra de ambos tipos de entidades mediante el uso de información contable no pública de dos años antes de la declaración de insolvencia, recurriendo a un modelo binario para datos de panel, concretamente una versión mejorada de la regresión logit con variables dependientes que varían a lo largo del tiempo, todo ello a partir de información financiera de 15.456 cooperativas y complementado con inputs macroeconómicos. Con todo ello observó que el mismo tipo de ratios servía para explicar la quiebra tanto en las cooperativas como en las cajas de ahorro, aunque por la propia naturaleza de este último grupo de entidades, los efectos de los cambios financieros resultaron ser más pronunciados en las cajas de ahorros.

En lo que respecta al uso de técnicas basadas en el análisis de supervivencia, es Maggiolini *et al.* (2005) el pionero en utilizarlo sobre las cooperativas de crédito de Italia en la década de 1990, comparando los tiempos de supervivencia de las cooperativas que se crearon en el periodo de estudio en relación con las creadas en años anteriores. Observaron que la alta competitividad existente en el mercado junto con la presencia de otras cooperativas de crédito en los territorios de las nuevas cooperativas, constituían las dos principales amenazas sobre los tiempos de supervivencia de las nuevas entidades.

Por su parte, Braga *et al.* (2006) se centra en la insolvencia de las cooperativas de crédito en Brasil. Para ello utilizaron información contable semestral desde 2001 hasta 2003 con la que calcularon diecinueve ratios de las que tan solo resultaron ser significativas seis, construyendo un modelo que permitió clasificar correctamente al 83% de las entidades solventes y al 79% de las insolventes.

Cabo *et al.* (2010) aplicaron el análisis de supervivencia para el estudio de la quiebra de cooperativas de crédito en Portugal entre 1995 y 2009, periodo en el que observaron que

el número de cooperativas se redujo desde las 220 en la década de 1990, hasta las 91 entidades a finales de 2009. Utilizaron 12 variables financieras previamente empleadas en investigaciones anteriores, con las que al final resultaron ser significativas tan solo dos, la que denominan ratio de transformación, o también conocido como *loan to deposits* (inversión crediticia / depósitos de la clientela) y otros gastos estructurales (gastos administrativos / activo total), aunque no indican nada sobre la capacidad predictiva del modelo y la situación de los errores Tipo I y Tipo II.

Fiordelisi *et al.* (2013) centran su estudio en la insolvencia de cooperativas de crédito en Italia entre 1997 y 2002, país en el que observaron que la tasa de quiebra era cuatro veces superior a la presentada por el resto de entidades financieras del sistema, por ello se preguntaron si la eficiencia y el equipo gestor se configuraba como una de las principales causas. Para dar respuesta a este interrogante recurrieron a un modelo de supervivencia en el que a partir de una serie de inputs que medían eficiencia y rentabilidad, el entorno macroeconómico y aspectos macroeconómicos llegaron a la conclusión de que las cooperativas con mayor rentabilidad y capitalización eran las que al mismo tiempo denotaban las mayores tasas de supervivencia.

Finalmente Carvalho *et al.* (2015) estudian los factores determinantes de la insolvencia de cooperativas de crédito en Brasil entre 1995 y 2009 mediante diferentes tipos de análisis de supervivencia, concluyendo que la regresión de Cox es la que ofrece los mejores resultados. Se plantean si la rentabilidad podría ser considerada justificante de la quiebra de este tipo de entidades, observando que no existía evidencia empírica suficiente para afirmar dicha hipótesis, sobre todo porque este tipo de entidades fomentan eficiencia antes que el beneficio ya que sus clientes son también sus socios. De las 38 ratios financieras inicialmente consideradas, finalmente 15 resultan ser significativas en el estudio, principalmente las relacionadas con el fondeo, el tamaño y la política de

inversiones, observando diferencias entre ellas en función de los diferentes tipos de insolvencias considerados.

2.6 CONCLUSIONES DEL CAPÍTULO.

A lo largo de este capítulo se ha tratado de exponer las principales líneas de investigación existentes sobre el estudio de la insolvencia de cooperativas de crédito, partiendo de una exposición inicial de la investigación en materia de quiebra empresarial y bancaria.

Se observa que en términos generales el conjunto de investigadores trata de determinar la relación entre insolvencias bancarias, y la situación económica, financiera, bursátil, y/o macroeconómica de años anteriores, mediante el uso de técnicas estadísticas o econométricas como son la regresión lineal, regresión probabilística, técnicas de supervivencia o redes neuronales.

Algunos autores ponen de manifiesto que los modelos estimados a partir de la regresión lineal presentan el inconveniente de no cumplir con las hipótesis básicas de esta técnica, como son la normalidad y homocedasticidad, si bien otro conjunto de autores indica que aún en esos casos el análisis discriminante lineal muestra resultados satisfactorios, justificando su uso aún en nuestros días.

Se observa que el mayor volumen de investigaciones se centra en el uso de los modelos probabilísticos logit y probit, con los que es posible conocer la probabilidad de que una entidad financiera pueda presentar una situación de insolvencia en un plazo no superior a tres años, no habiendo encontrado estudios significativos que demuestren la validez de estos modelos a plazos de tiempo superiores.

El análisis de supervivencia se postula como una buena alternativa a los modelos de respuesta binaria anteriores porque además de acercar la probabilidad de insolvencia, mide la influencia del tiempo sobre esta situación, acentuándose el empleo entre el conjunto de investigadores de una técnica ampliamente utilizada en el contexto de la ciencia sanitaria.

Uno de los principales problemas a los que se enfrentan el conjunto de investigadores es el de la selección de los inputs o variables explicativas con poder suficiente para la construcción de modelos eficaces, ya que no existe una teoría cierta que explique el funcionamiento de la insolvencia bancaria y que pueda servir de base para la selección de los inputs.

No obstante se observa una tendencia generalizada de las ratios que se basan en el modelo CAMEL, formado por ratios que miden el capital, la calidad del activo, el equipo gestor, los fondos propios y el endeudamiento. De hecho, por regla general la capitalización, calidad de la cartera crediticia y el acceso a fuentes de financiación se postulan como los principales indicadores en la explicación de la insolvencia bancaria.

Se constata que el volumen de investigaciones sobre las insolvencias de cooperativas de crédito es reducido, mayoritariamente en un ámbito internacional y siendo prácticamente inexistentes en ámbito nacional.

A nivel internacional se observa que en los últimos años se está produciendo un agudizamiento en el volumen de investigaciones disponibles sobre los procesos de insolvencias de cooperativas de crédito, y aunque es cierto la mayor parte de los investigadores ponen de manifiesto que estos procesos de insolvencia revierten cierto carácter excepcional, una de las principales diferencias se encuentra en el hecho de ser habitual que las cooperativas de crédito con problemas perciban asistencia del resto de

cooperativas (sea mediante fusiones o ayudas) que distorsionan el cómputo de procesos de insolvencia dentro de este sector.

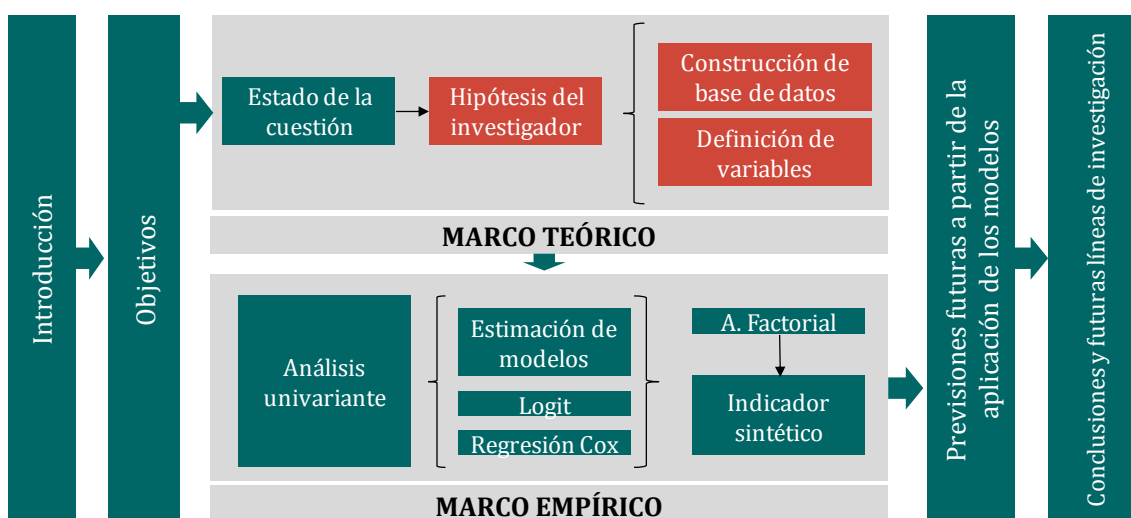
CAPÍTULO TERCERO:

EL ESTUDIO DE LA SITUACIÓN FINANCIERA DE LAS CAJAS RURALES.

3.1 SINOPSIS DEL CAPÍTULO

A lo largo de este capítulo se describe el procedimiento seguido para el estudio de la situación financiera de las Cajas Rurales, siendo su ubicación dentro de la sinopsis general de la Tesis la recogida en la Ilustración 6.

ILUSTRACIÓN 6: UBICACIÓN DEL CAPÍTULO TERCERO DENTRO DE LA SINOPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

Se comienza definiendo las hipótesis a contrastar para la consecución del objetivo fijado, que no es otro que la contribución al conocimiento financiero de las Cajas Rurales determinando si los procesos de fusiones acaecidos dentro del sector pueden ser explicados por motivos de sostenibilidad económica y/o financiera de las entidades implicadas, con el fin último de prever este tipo de situaciones en un futuro. Para ello se ha construido un panel de datos que combina la dimensión temporal (desde 1991 hasta 2015) con la dimensión transversal (características económicas, financieras, sectoriales y coyunturales de cada una de las Cajas Rurales que forman parte del estudio), que ha sido utilizada para la construcción de dos modelos estadísticos basados en regresión logística binaria (o logit) y supervivencia, ampliamente utilizadas por otros autores dentro de la literatura sobre insolvencias empresariales y bancarias.

3.2 HIPÓTESIS DEL INVESTIGADOR

Para la consecución de los objetivos expuestos en el Capítulo primero de esta Tesis, la contribución a un conocimiento económico y financiero más profundo de las Cajas Rurales españolas, se plantea contrastar las siguientes hipótesis:

PRIMERA HIPÓTESIS

Marco teórico: Porath (2006), Cabo *et al.* (2010) y Lima (2012) observan en sus respectivos estudios sobre las cooperativas de crédito en Alemania, Portugal y Brasil que dentro del sector es habitual la existencia de procesos de fusión o asistencia financiera entre las cooperativas de crédito explicados por la debilidad financiera de las entidades implicadas, a diferencia del sector bancario en el que el volumen de insolvencias manifiestas es superior.

Hipótesis planteada: en España los procesos de fusiones de Cajas Rurales acaecidos entre 1991 y 2015 estuvieron motivados por la propia situación económica y financiera de las Cajas Rurales implicadas.

Método de contraste: a través de un modelo logit y una regresión de riesgos proporcionales de Cox se tratará de determinar la existencia de significatividad de las variables explicativas sobre el comportamiento de la variable dependiente (variable dicotómica que toma el valor uno si la Caja Rural ha formado parte de un proceso de fusión, y cero en el resto de casos)

SEGUNDA HIPÓTESIS

Marco teórico: la revisión bibliográfica pone de manifiesto el uso mayoritario de ratios financieras que miden aspectos del sistema CAMEL para estimar modelos que permitan prever la insolvencia de entidades financieras, confirmando que aquellas entidades que mantienen los mayores niveles de capitalización, rentabilidad y liquidez muestran una menor probabilidad de insolvencia. Por el contrario, las que muestran menor aversión al riesgo muestran mayores probabilidades de insolvencia.

Hipótesis planteada: las variables recogidas en el sistema CAMEL permiten explicar los procesos de fusiones de Cajas Rurales, esperándose una influencia significativa de todas aquellas variables que midan la capitalización, rentabilidad y liquidez, no siendo esperable influencia significativa de la variable tamaño, ya que se trata de entidades de pequeño tamaño, muy dependientes del territorio en el que se ubican.

Método de contraste: a través de los modelos de respuesta binaria (logit) y riesgos proporcionales de Cox definidos en la hipótesis primera.

TERCERA HIPÓTESIS

Marco teórico: entre el conjunto de investigadores no existe consenso sobre la influencia de las variables macroeconómicas en la previsión de los procesos de insolvencia de entidades financieras, si bien este hecho puede estar explicado por el uso de modelos estadísticos estáticos que no tienen en cuenta la influencia del tiempo sobre las variables explicativas.

Hipótesis planteada: las variables macroeconómicas influyen sobre la probabilidad de que una Caja Rural forme parte de un proceso de fusión.

Método de contraste: a través de una variación del modelo logit que permite la incorporación de variables que varían a lo largo del tiempo, de forma similar a lo ya realizado por Shumway (2001) en su estudio sobre la insolvencia de corporaciones empresariales.

CUARTA HIPÓTESIS

Marco teórico: de la revisión bibliográfica efectuada no se observa diferencias en la probabilidad de insolvencia de entidades financieras en función del territorio en los que se ubican, si bien las causas que las motivan pueden diferir de unos territorios a otros.

Hipótesis planteada: se espera una influencia positiva de la variable que considera la ubicación geográfica de las Cajas Rurales, ya que existe concentración de las Cajas Rurales en la zona central y el levante de la península ibérica.

Método de contraste: a través de los modelos estadísticos especificados en las hipótesis anteriores.

QUINTA HIPÓTESIS

Marco teórico: no han sido halladas investigaciones académicas que recurran al uso de indicadores sintéticos para valorar la salud financiera de entidades financieras, si bien la bibliografía existente sobre la construcción de indicadores sintéticos para la estimación de otros aspectos macroeconómicos muestran resultados muy favorables.

Hipótesis planteada: el indicador sintético construido a partir de los ratios que miden la situación económica y financiera de las Cajas Rurales muestra la capacidad de discriminar las entidades más saneadas de las que muestran mayores incertidumbres.

Método de contraste: construcción de un indicador sintético a partir del método del análisis factorial por componentes principales.

3.3 MODELOS DE ALERTA TEMPRANA.

Laitinen *et al.* (1999) pone de manifiesto que en el estudio de la insolvencia de empresarial no existe consenso entre el conjunto de investigadores sobre qué técnicas funcionan mejor, si bien es cierto que en su amplia mayoría se han basado en la aplicación de análisis univariante, regresiones lineales y regresiones probabilísticas.

El análisis univariante constituyó el primer acercamiento al estudio de las causas de subyacen en la quiebra empresarial y bancaria. Este método se basa en la construcción de una serie de ratios financieras que, a partir de su comparación con el valor que el investigador consideraba como normal, permitían discriminar entre individuos solventes e insolventes, si bien fue criticado por el sesgo introducido por el investigador en la determinación del punto de corte considerado normal (Anastasi *et al.*, 1998).

Posteriormente esta línea de investigación se complementó con la estimación de regresiones lineales que asignaban una puntuación a cada individuo en base a una serie de variables explicativas (Altman *et al.*, 1977), si bien estos modelos se fundamentan en supuestos restrictivos, principalmente que cada grupo definido por la variable dependiente se ajusta a una distribución normal, además de igualdad de las matrices de las covarianzas.

El hecho de que la variable dependiente sea dicotómica implica que los grupos definidos por esta variable dependiente no se adaptan a una distribución normal, y aunque existen autores que indican que estos modelos son suficientemente robustos para proporcionar resultados eficientes, son numerosos los investigadores que proponen recurrir al uso de regresiones probabilísticas binarias, principalmente la regresión logit y la regresión probit, que permiten explicar en términos de probabilidad el comportamiento de la variable dependiente a partir de una serie de variables explicativas.

Estas regresiones probabilísticas no pueden ser estimadas por mínimos cuadrados ordinarios por presentar inconsistencias en su estimación (Ramajo *et al.*, 2002), siendo por ello necesaria su estimación a partir de máxima verosimilitud (parámetros que maximizan la probabilidad de acierto).

Aunque se explicará con detenimiento en el epígrafe siguiente de este capítulo, Anastasi *et al.* (1998) define el modelo de respuesta binaria como:

$$P(Y = 1) = F(\beta \cdot X) \tag{EQ1}$$

$$P(Y = 0) = 1 - F(\beta \cdot X)$$

Donde F es la función de distribución acumulada, P(Y=1) es la probabilidad de que la variable dependiente de tipo binario tome el valor 1, y β son los parámetros que miden la influencia de las variables explicativas sobre la probabilidad,

Aunque los modelos probabilísticos permiten conocer la probabilidad de ocurrencia del suceso estudiado, los modelos de supervivencia complementan esta información estimando el tiempo que tarda un individuo en mostrar los primeros síntomas de insolvencia.

Como indica Anastasi *et al.* (1998) los modelos de supervivencia consisten en un *análisis de corte transversal de los tiempos de duración observados para todos los individuos de la población*, considerando que la función de distribución del tiempo es similar para todos ellos.

La principal crítica de estos modelos es el hecho de trabajar con datos censurados a la derecha, es decir, se supone que todos los individuos han vivido hasta el momento de inicio del estudio, pero nada se indica sobre su supervivencia real por escaparse del periodo abarcado en el estudio.

Formalmente los modelos de supervivencia se integran por dos componentes principales: la función de supervivencia propiamente dicha, que se denota por $S(t)$; y la función de riesgo $h(t)$, ambas estimadas a partir de los tiempos de supervivencia observados para un conjunto de individuos, hallándose en Cordón *et al* (2014) una buena descripción de los fundamentos teóricos en los que se basan.

En este sentido Cordón *et al* (2014) indica que la función de supervivencia $S(t)$ se define como la probabilidad de sobrevivir al siguiente instante de tiempo:

$$S(t) = Prob(T > t) = \int_t^{\infty} f(u)du \quad \text{EQ2}$$

Siendo T una variable aleatoria continua no negativa con función de densidad $f(t)$ que cumple que la probabilidad acumulada es igual a 1.

Por su parte, la función de riesgo $h(t)$ mide la probabilidad de que un determinado individuo i presente el suceso en el siguiente instante de tiempo:

$$h(t) = \lim_{\Delta t \rightarrow \infty} \frac{P[t < T \leq t + \Delta t | T > t]}{\Delta t} \quad \text{EQ3}$$

Si T es continua, entonces se cumple que:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log[S(t)]}{dt} \quad \text{EQ4}$$

siendo el riesgo acumulado el definido como:

$$h(t) = \int_0^t h(u)du = -\log[S(t)] \quad \text{EQ5}$$

Por lo tanto, y en función de las definiciones anteriores, se puede considerar que $h(t)\Delta t$ representa, aproximadamente, la probabilidad de que un individuo en el momento t presente el suceso en el siguiente instante de tiempo.

Aunque en el análisis de supervivencia no es necesario conocer la distribución de probabilidad a la que se ajusta el tiempo de supervivencia, si se conoce, las inferencias serán más precisas al reducir los errores de los estimadores (modelos paramétricos).

Cordón *et al.* (2014) defiende que es preferible utilizar un modelo paramétrico cuando se esté seguro del modelo correcto, y en caso contrario, recurrir a modelos no paramétricos, como por ejemplo la regresión de Cox.

Finalmente existen técnicas más complejas basadas en el ámbito de las redes neuronales que no son objeto de estudio en esta Tesis porque estos modelos no permiten conocer la influencia que cada una de las variables explicativas (X) ejerce sobre la probabilidad de insolvencia de los individuos, de forma que su uso no permitiría cumplir con el objetivo establecido de determinar si los Estados Financieros proporcionan suficiente información para explicar los procesos de integración de Cajas Rurales y su vinculación con la sostenibilidad económica y/o financiera de las entidades implicadas.

Además de ello, en la Tabla 5 se observa que entre el conjunto de investigadores sobre insolvencias de cooperativas de crédito ha sido habitual el empleo de regresiones probabilísticas (logit/probit) y análisis de supervivencia, no habiendo sido encontradas investigaciones que recurran al uso de redes neuronales. Por todo ello en esta Tesis se utilizará finalmente la regresión logística binaria y el modelo de supervivencia de Cox.

TABLA 5: REVISIÓN BIBLIOGRÁFICA DE LAS TÉCNICAS ESTADÍSTICAS UTILIZADAS PARA EL ESTUDIO DE LA INSOLVENCIA DE COOPERATIVAS DE CRÉDITO.

TÉCNICA	INVESTIGACIONES
Análisis discriminante	Simon (1980)
Logit / Probit	Dabós (1996), Gama <i>et al.</i> (2004), Porath (2006), Gama <i>et al.</i> (2011) y Lima (2012),
Supervivencia	Maggiolini <i>et al.</i> (2005), Braga <i>et al.</i> (2006), Cabo <i>et al.</i> (2010), Fiordelisi <i>et al.</i> (2013) y Carvalho <i>et al.</i> (2015)

FUENTE: ELABORACIÓN PROPIA

3.3.1 Regresión Logística binaria o Logit.

3.3.1.1 Definición formal del modelo logit

La regresión logística binaria relaciona una variable dependiente de tipo dicotómico con una serie de variables explicativas (X), siendo utilizado en aquellos casos en los que no es conveniente utilizar la regresión simple o múltiple, lineal o no (Silva *et al.*, 2004) por no cumplirse los supuestos básicos, esto es, porque presentan la no normalidad de los errores, heterocedasticidad, valores de la probabilidad fuera del intervalo $[0,1]$ y la presencia de un coeficiente de determinación que no es adecuado como medida de ajuste del modelo (Ramajo *et al.*, 2002).

Aunque el modelo logit no se ajusta a una distribución lineal, para su definición formal se partirá de una regresión lineal que relaciona una variable dependiente (Y) con una única variable independiente (X):

$$Y = \alpha + \beta X \quad \text{EQ6}$$

Siendo Y una variable dicotómica que toma el valor 1 en caso de que se produzca el cese de actividad de la Caja Rural, y el valor 0 en caso contrario. Por su parte, α es la constante del modelo y β el parámetro estimado para la variable independiente.

Interesa conocer la probabilidad de ocurrencia de cada suceso de la variable dependiente, de ahí que la probabilidad se ajuste de la forma recogida en EQ7:

$$Y = \frac{P}{1 - P} \quad \text{EQ7}$$

Siendo P la probabilidad de cese de actividad de la Caja Rural, y $1-P$ la probabilidad de ocurrencia de la supervivencia de la Entidad. Igualando ambas expresiones se llega a EQ8:

$$\frac{P}{1-P} = \alpha + \beta X \quad \text{EQ8}$$

El modelo lineal clásico no permite calcular la probabilidad de que la variable dependiente tome los valores 0 y 1, puesto que de su estimación pudieran derivarse valores de la variable dependiente que fueran superiores a 1, por ello es necesario realizar una transformación logarítmica de la expresión anterior como se recoge en EQ9:

$$\text{Ln}\left(\frac{P}{1-P}\right) = \alpha + \beta X \quad \text{EQ9}$$

Operando en ambos términos llegamos a la expresión general de la Regresión Logística Binaria o Logit recogida en EQ10:

$$P = \frac{1}{1 + e^{-\alpha - \beta X}} \quad \text{EQ10}$$

La expresión anterior relacionaría una variable dependiente con una única variable explicativa. En esta Tesis se pretende relacionar una variable dependiente Y, con un conjunto de k variables explicativas, por lo que la expresión general del modelo logit quedar definida en EQ11:

$$P = \frac{1}{1 + e^{-\alpha - \beta_1 X_1 - \beta_2 X_2 - \dots - \beta_k X_k}} \quad \text{EQ11}$$

Para la estimación de los parámetros del modelo se recurre al método de máxima verosimilitud ya que permite obtener aquellos valores de los coeficientes que maximizan la suma de los cuadrados de los residuos.

Según Silva *et al.* (2004) el proceso de hallar los parámetros que maximicen la función de verosimilitud es complicado y exige la aplicación de un procedimiento numérico, como el algoritmo de *Newton-Ralphson* que demanda el uso de iteraciones, y aunque no es objeto de esta Tesis definir los pasos para hallar los parámetros, a modo explicativo el

procedimiento se inicia calculando el valor del término constante que maximiza ese valor de la verosimilitud, sin incluir ninguna de las variables explicativas (EQ12):

$$a = \ln \frac{\sum_{i=1}^n Y_i}{n - \sum_{i=1}^n Y_i} \text{ y } b_1 = b_2 = \dots = b_k = 0 \quad \text{EQ12}$$

Siendo a, b_1, b_2, \dots, b_k las estimaciones de los parámetros $\alpha, \beta_1, \beta_2, \dots, \beta_k$.

Posteriormente se calculan iteraciones del modelo incluyendo una a una cada variable explicativa, observando los valores tomados por la verosimilitud con el fin de encontrar la respuesta que maximice esa verosimilitud.

Para la interpretación de los parámetros se calcula la primera derivada, llegando a la expresión recogida en EQ13:

$$\frac{dP}{dX_1} = \frac{-e^{-\alpha - \beta_1 X_1} \times (-\beta_1)}{(-e^{-\alpha - \beta_1 X_1})^2} = \beta_1 P(1 - P) \quad \text{EQ13}$$

Si β_1 es positivo, dP/dx también lo será. Siguiendo a Ramajo *et al.* (2002) en el caso general del modelo con K variables puede hacerse una interpretación alternativa de los coeficientes considerando que de la expresión dP/dx se obtiene que $\frac{\beta_j}{\beta_k} = \frac{dP/dx_j}{dP/dx_k}$ y, por tanto, la razón de componentes estimados proporciona una medida del cambio relativo en las probabilidades.

3.3.1.2 Críticas al uso del modelo logit

El uso de modelos logit para el estudio de la quiebra de empresas y entidades financieras ha sido criticado por ignorar el carácter dinámico de las estructuras económico financieras de las empresas y entidades analizadas (Shumway, 2001; Nam *et al.*, 2008; y Nuñez *et al.*, 2011), ya que la mayor parte de las investigaciones recurren a información económica y

financiera previa a la situación de insolvencia (o solvencia) del individuo, sin considerar el hecho de que la insolvencia es un proceso de deterioro continuado en el tiempo (Nam *et al.*; 2008)

Shumway (2001) puso de manifiesto que las investigaciones sobre insolvencias empresariales y bancarias que consideran el evento de quiebra como un acontecimiento dicotómico, se enfrentan al problema de que habitualmente el número de insolvencias que se producen en un único año es inferior al de los individuos que continúan siendo solventes, por ello muchos autores suelen incluir en la muestra insolvencias de años anteriores, emparejadas con individuos solventes de los mismos ejercicios, para mejorar la calidad muestral, si bien continuaría la pérdida del carácter temporal de la información.

Por ello Shumway (2001) propuso una variación del modelo logit con la que mejoró las predicciones de los investigadores anteriores, observando que variables explicativas que habían resultado significativas bajo los modelos logit estáticos, dejaban de serlo mediante el uso de esta nueva técnica.

En línea similar, Nam *et al.* (2008) en su estudio sobre la quiebra empresarial en Korea entre 1999 y 2000, confirmaron que los modelos logit multiperiodo muestran mayor poder predictivo que los modelos logit estáticos.

Aunque esta versión multiperiodo del modelo logit propuesta por Shumway (2001) se explicará en el epígrafe siguiente, su definición formal se basa en el uso de modelos logit que incluyen covariables que varían con el tiempo.

En el ámbito de estudio de las cooperativas de crédito, Porath (2006) ha sido el único investigador que ha utilizado una regresión binaria multiperiodo. Concretamente a través de un probit multiperiodo estudia la insolvencia de cooperativas de crédito y cajas de ahorro en Alemania entre 1993 y 2002.

Como indican Schmidt *et al.* (2010) no todos los autores coinciden en que la estimación de los modelos logit multiperiodo sean mejores que las obtenidas por los modelos estáticos, básicamente por el hecho que estos nuevos modelos pueden presentar problemas de parsimonia que justifiquen la mayor capacidad predictiva observada en las pruebas sobre individuos que no formaron parte de la muestra inicial.

Fantazzini *et al.* (2009) en un estudio de la quiebra de pequeñas corporaciones observan que los modelos logit estáticos demuestran una mayor capacidad predictiva en el caso de individuos no incluidos inicialmente en la muestra, conclusión a la que también llegan Schmidt *et al.* (2010) en su comparación sobre el poder predictivo de ambos tipos de modelos en base a 102 empresas cotizadas de EE.UU., corroborando así hallazgos similares de Fuertes *et al.* (2006).

Como se ha puesto de manifiesto no existe consenso suficiente entre el conjunto de investigadores sobre si el modelo logit estático funciona mejor que el logit multiperiodo, sobre todo en lo que refiere a la aplicación del modelo para la posterior estimación de probabilidad sobre individuos que no formaron parte de la muestra inicial, por este motivo en esta Tesis se evaluarán ambos modelos sobre la muestra de datos referentes a las Cajas Rurales españolas.

3.3.2 Regresión logística multiperiodo

Como se ha introducido en el subepígrafe anterior, el concepto de regresión logística multiperiodo se fundamenta en la investigación de Shumway (2001) en la que critica el uso del logit estático por obviar el carácter dinámico de los datos introducidos.

En este sentido, Shumway (2001) indica que la mayor parte de las bases de datos que se utilizan para crear modelos de previsión de insolvencia empresarial incluyen información de los n individuos que existen entre $t=1$ hasta $t=T$, periodo durante el cual una compañía puede quebrar, sobrevivir o simplemente salir del estudio, por ello define una variable t_i para cada individuo que recoge el tiempo que transcurre hasta que se produce su insolvencia o hasta que abandona el estudio.

Si bien es cierto que Shumway (2001) se centró en el análisis de la probabilidad de insolvencia empresarial, Cole (2009) fue pionero en aplicar logit multiperiodo al estudio de la insolvencia bancaria en EE.UU., línea de investigación que posteriormente habría sido seguida por Porath (2006) o DeYoung (2013), entre otros.

Formalmente los modelos logit multiperiodo se basan en la definición del modelo logit estático, pero incluyendo covariables que dependen del tiempo (Duda *et al*, 2010), por ello su definición formal es muy similar a la expresión del modelo logit estático:

$$h(t, x) = \frac{1}{1 + e^{-(\beta'_1 x_t + \beta'_2 x_{it})}} \quad \text{EQ14}$$

Donde x_t es una variable dependiente del tiempo, también denominada función de riesgo basal, que para Shumway (2001) corresponde al logaritmo del tiempo de supervivencia del individuo, si bien existen diversas formas de especificación, incluyendo variables macroeconómicas (Nam *et al.*, 2008; y, Schmidt *et al.*; 2010) u de otro tipo de dependan del tiempo; por su parte, $\beta'_2 x_{it}$ corresponde a las propias características del individuo, covariables que también dependen del tiempo.

Como indica Shumway (2001) con este sistema se mejora el volumen de casos a introducir en los modelos, puesto que a diferencia de lo que ocurre con los modelos logit estático, en esta ocasión para cada uno de los individuos se consideran todas las observaciones disponibles.

En Shumway (2001) y Duda *et al* (2010) se encuentra una excelente justificación formal del funcionamiento del modelo logit estático, si bien simplemente señalar que estos modelos se basan en la función de supervivencia $S(t_i, x_i, \beta)$ que mide la probabilidad de que un individuo sobreviva al periodo t , siendo su expresión la siguiente:

$$S(t_i, x_i, \beta) = 1 - F(t_i, x_i, \beta) \quad \text{EQ15}$$

Donde $F(t_i, x_i, \beta)$ es la función de densidad acumulada asociada a la insolvencia.

Estos modelos se estiman por máxima verosimilitud. De hecho Shumway (2001) demostró que la función de verosimilitud de un modelo logit multiperiodo es similar a la de un modelo logit estático con función de riesgo $h(t, x) = F(t_i, x_i, \beta)$, por ello puede ser fácilmente estimado utilizando las técnicas de regresión logística (Duda *et al.*, 2010).

3.3.3 Regresión de Riesgos Proporcionales o Regresión de Cox

El Modelo de Cox es una variación de los modelos de supervivencia para aquellos casos en los que se desconoce la función de probabilidad a la que se ajusta el tiempo de supervivencia. En el manual de Cordon *et al.* (2014) se recoge una excelente descripción formal de este modelo, de ahí que se haya seguido para la elaboración de esta Tesis y al que referimos para mayor información sobre la justificación formal del mismo.

Brevemente indicar que Cox (1972) elaboró un método no paramétrico para expresar la función de riesgo $h(t; x)$ en función del tiempo (t) y de k variables independientes (X_k).

Cordon *et al.* (2014) recoge que esta regresión de Cox consta de dos términos, la función de riesgo base $h_0(t)$, y un segundo término, $\varphi(X)$, que está en función de las variables explicativas:

$$h(t; x) = h_0(t)\varphi(X) \quad \text{EQ16}$$

Concretamente la función de riesgo base registra el riesgo que presenta un individuo cuando las variables explicativas toman el valor cero, mientras que $\varphi(X)$ recoge el riesgo en función del valor tomado por las k variables explicativas (X_k) en el momento t para el individuo i , esta última ajustada mediante una función exponencial (Cordón *et al.*; 2014), siendo por ello su definición la siguiente:

$$h(t; x) = h_0(t)e^{\beta'x} \quad \text{EQ17}$$

El modelo de Cox tiene la consideración de semiparamétrico por el hecho de que tanto la función de riesgo base como $\varphi(X)$ pueden ser estimados independientemente. En este sentido, los k parámetros β_k asociados a las variables explicativas X_k se estiman mediante máxima verosimilitud, mientras que la función de riesgos se obtiene a partir de los datos analizados.

Según Cordón *et al.* (2014) la probabilidad de que un individuo i presente la situación estudiada en el momento t_j a partir de una series de variables explicativas X_j , se define de la siguiente forma:

$$\frac{h(t_j, x_j)}{\sum_{i \in R(t_j)} h(t_j, x_j)} \quad \text{EQ18}$$

Donde $t_1 \leq t_2 \leq \dots \leq t_m$ son los tiempos de fallo y $R(i)$ es el conjunto de observaciones en riesgo en cada instante del tiempo j .

En lo que respecta a la función de verosimilitud, Cordón *et al.* (2014) la define a partir de las probabilidades condicionales para los m tiempos de fallo, de la siguiente forma:

$$L(\beta) = \prod_{j=1}^m \frac{e^{\beta'x_j}}{\sum_{i \in R(t_j)} e^{\beta'x_j}} \quad \text{EQ19}$$

Como ya se ha comentado, los modelos de supervivencia presentan la peculiaridad de presentar observaciones censuradas (es decir, individuos sobre los que no se conoce la

supervivencia exacta por superar el periodo de estudio o por ser anterior al mismo). Esta censura se establece mediante el indicador δ_j , el cual toma el valor 0 si existen observaciones censuradas, y el valor 1 cuando el tiempo sea completo:

$$L(\beta) = \left[\prod_{j=1}^m \frac{e^{\beta x_j}}{\sum_{i \in R(t_j)} e^{\beta x_j}} \right]^{\delta_j} \quad \text{EQ20}$$

Así mismo, si existiesen observaciones con tiempos empatados la estimación se realiza por el método de máxima verosimilitud parcial con la corrección de Breslow¹⁶, ya que como indica Cordon *et al.* (2014) podrían generarse errores al ordenar los tiempos de supervivencia de los distintos individuos.

3.4 POBLACIÓN Y MUESTRA

Se dispone de una población de datos de panel que incluye información económica y financiera de cada una de las Cajas Rurales existentes en España entre 1980 y 2015.

Se ha extraído una muestra que incluye a todas las Cajas Rurales operativas a lo largo del periodo 1991 y 2015, habiendo sido descartados los años anteriores por la existencia de limitaciones de información que se detallarán a lo largo de este epígrafe.

¹⁶ Cordon *et al.* (2014) pone de manifiesto que la definición inicial del modelo de Cox (1972) se basa en la idea del tiempo continuo, por lo que dos individuos no pueden manifestar el acontecimiento de interés en el mismo instante. Sin embargo, desde un punto de vista discreto, sí es posible que se produzca la ocurrencia en el mismo instante de ahí la necesidad de recurrir a la modificación de Breslow indicada.

3.4.1 Las Cajas Rurales

Las Cajas Rurales son un tipo de cooperativa de crédito que combinan las características de una entidad de crédito con las de sociedad cooperativa, estando sujetas tanto a la normativa aplicable a las entidades de crédito¹⁷, como a la normativa que regula la actividad cooperativa¹⁸, a nivel estatal y autonómico.

En este sentido, la Ley 13/1989 define a las Cooperativas de Crédito como *“sociedades constituidas con arreglo a la presente Ley, cuyo objeto social es servir a las necesidades financieras de sus socios y de terceros mediante el ejercicio de las actividades propias de las entidades de crédito”*.

La principal diferencia existente entre una Caja Rural y una Cooperativa de Crédito radica en el carácter agrario de la primera, es decir, según establece la Ley 17/1999 de Cooperativas, la denominación de Caja Rural tan solo puede ostentarla aquellas cooperativas de crédito que *“incluyan una cooperativa agraria o al menos cincuenta socios personas físicas que sean titulares de explotaciones agrarias”*.

Al ser sociedades de crédito, las Cajas Rurales están capacitadas para el ejercicio de la actividad financiera, es decir, tomar recursos para la posterior financiación a sus clientes, si bien, su carácter de cooperativa le obliga a que esas operaciones de financiación deban realizarse en su mayor parte con sus propios socios.

De hecho, tienen la obligación de que las operaciones con los no socios no puedan superar el cincuenta por ciento de los recursos propios de la cooperativa, ya que, como afirma

¹⁷ Entre otras, dentro de la regulación estatal se incluye la Ley 10/2014 de ordenación, supervisión y solvencia de las entidades de crédito, y el RD 84/2015 que la desarrolla. A nivel comunitario, destaca el Reglamento 575/2013 del Parlamento Europeo sobre los requisitos prudenciales de las entidades de crédito y las empresas de inversión, y que modifica al Reglamento 648/2012.

¹⁸ Principalmente la Ley 13/1989 de Cooperativas de Crédito, así como el RD 84/1993 que lo desarrolla, además de la Ley 17/1999 de Cooperativas.

Rodero (2003) las Cajas Rurales buscan la *“concesión de créditos, con bajos tipos de interés y adaptados a la necesidad de la producción agraria”* de sus asociados.

Para ser socio de una Caja Rural es necesario participar en el capital social de la misma con acciones que tienen un nominal que varía en función de cada una de las Cajas Rurales, si bien la propia Ley 13/1989 establece que no podrá ser inferior a los 60,10 euros por acción, existiendo además limitaciones a la participación máxima que pueden ostentar individualmente cada uno de los socios, según se trate de personas físicas (máximo del 2,5% del capital social), o personas jurídicas (máximo del 20% del capital social), si bien en ningún caso las personas jurídicas que no sean cooperativas podrán poseer más del 50% del capital.

Atendiendo a las características anteriores, Palomo *et al* (2008a) define a las cooperativas como *sociedades de personas que participan como socios en su capital social, como empresarios en su dirección o gestión, y como clientes en su actividad financiera.*

Su carácter de cooperativa implica ciertas restricciones a la distribución de los beneficios, ya que su principal objetivo no es otro que maximizar la eficiencia de sus socios a través de los precios establecidos en las operaciones financieras, no obstante, la Ley 13/1989 establece que los beneficios deberán ser distribuidos al fondo de reserva obligatorio (20%), el fondo de educación y promoción (10%), siendo el resto disponible para remunerar a los socios, siempre y cuando la Caja Rural cumpla con los requisitos de solvencia establecidos en la regulación aplicable.

Existen diferentes tipos de Cajas Rurales, en función del ámbito territorial de actuación definidos en sus Estatutos fundacionales, diferenciados entre local, provincial (supone una asociación de las Cajas Rurales locales de la provincia, si bien su asociación es voluntaria) y nacional (que asocia voluntariamente a las Cajas Rurales provinciales). De hecho, el RD

84/1993 establece que las Cooperativas de Crédito no podrán realizar operaciones de activo, ni con sus socios, fuera de ese ámbito de actuación.

Palomo *et al* (2008a) señala que en los últimos años se ha producido una apertura en el negocio tradicional de las Cajas Rurales, dejando atrás esa especialización agraria, como parte de un proceso de bancarización de las cooperativas de crédito, permitiéndoles acceder a nuevos entornos a los que antes no tenían acceso.

A la par de ese proceso de bancarización de las cooperativas de crédito, Palomo *et al* (2008a) pone de manifiesto que se ha producido un proceso de expansión territorial de estas entidades mediante la apertura de nuevas sucursales, principalmente en las zonas de mayor desarrollo económico de España.

En lo que respecta a los orígenes del cooperativismo de crédito en España, en Soto (2003) se encuentra una minuciosa descripción tanto del inicio como de la evolución de este movimiento, poniendo de manifiesto que los orígenes se remontan a los años veinte del siglo XX para cubrir las necesidades del área rural ya que los bancos existentes en esa fecha mostraban poco interés en cubrir las necesidades financieras de este segmento. No obstante, al no contar con bancos centrales que proporcionaran recursos a las cooperativas de crédito creadas, como sí pasaba en otros países europeos, su desarrollo en el conjunto del sistema financiero español fue muy pequeño.

En la Tabla 6 se recoge la distribución de Cajas Rurales existentes en los primeros años del cooperativismo de crédito en España, denotándose en 1915 la existencia de un total de 525 Cajas Rurales que se encontraban muy concentradas en el entorno de Castilla y León, Cataluña y Navarra, y que evolucionó positivamente hasta las 616 entidades registradas en 1933.

TABLA 6: DISTRIBUCIÓN TERRITORIAL DE LAS CAJAS RURALES ENTRE 1.915 Y 1.933

AÑO	1915	1918	1924	1933
Andalucía	24	11	11	29
Aragón	53	48	46	42
Asturias	55	31	29	
Baleares	20	19	15	19
Canarias				
Cantabria	2	5	5	52
CLM	36	30	22	48
CyL	103	80	81	161
Cataluña	47	53	65	105
Extremadura	34	30	28	50
Galicia	9	6	6	12
Madrid	4	12	12	11
Murcia	10	10	10	8
Navarra	113	153	156	63
País Vasco	4	6	5	14
Rioja	1	2	2	5
Valencia	10	7	6	81
Total	525	503	501	616

FUENTE: SOTO (2003)

En la actualidad el segmento de las cooperativas de crédito españolas se integra por un total de 62 cooperativas de las 59 corresponden a Cajas Rurales, agrupadas a su vez en torno a la Unión Nacional de Cooperativas de Crédito (UNACC), la Asociación Española de Cajas Rurales y el Grupo Cooperativo Cajamar.

Bajo el prisma de la UNACC se asocian un total de 40 Cajas Rurales y 3 cooperativas de crédito diferentes, con un volumen de activos de 92.859 millones de euros y una cuota de mercado del 7,03%. Así mismo gestionan un total de 3.349 oficinas y 12.337 empleados, con más de 1,4 millones de socios cooperativistas.

Algunas de las Cajas Rurales integrantes en la UNACC forman parte de la Asociación Española de Cajas Rurales, una asociación creada en 1923 por 23 Cajas Rurales que abandonaron el Grupo Asociado de Crédito Agrícola, y que fue incrementándose posteriormente con la adhesión de nuevas Cajas Rurales. En la actualidad está conformado por un total de 29 Cajas Rurales que totalizan más de 59.000 millones de euros de activos y fondos propios por 4.800 millones de euros, con una red de oficinas de más de 2.350 sucursales y 8.400 empleados.

La Asociación de Cajas Rurales se basa en un sistema de integración federado que respeta la identidad individual de cada una de las entidades asociadas, pero que permite mejorar las estructuras de funcionamiento entre las Cajas Rurales asociadas.

Además de ello, existe un Fondo Interno de Solidaridad, de carácter mutuo y solidario, que permite asistir financieramente a los miembros con problemas, y que se nutre de las aportaciones de cada Caja Rural en función de su patrimonio neto.

El Grupo Caja Rural cuenta con tres pilares fundamentales: el Banco Cooperativo Español, la sociedad Rural Servicios Informáticos y el holding que conforma Seguros RGA:

- a. Banco Cooperativo Español, constituido en 1990 y participado por Cajas Rurales y la entidad alemana DZ Bank. Su misión principal es la de ser central bancaria de las Cajas Rurales asociadas.
- b. RSI, Rural Servicios Informáticos, constituida en 1986 para externalizar los servicios informáticos de las Cajas Rurales asociadas, si bien en la actualidad brindaría servicios a otras entidades del sector financiero.
- c. RGA, constituía en 1986 y respaldada por un total de 38 Cajas Rurales y la empresa alemana *R+V Allgemeine Versicherung AG*, y que desarrolla la actividad aseguradora de Grupo Caja Rural a través de las sucursales de las entidades asociadas.

Por su parte el Grupo Cooperativo Cajamar surge como resultado de un Sistema Institucional de Protección (SIP) conformado por un total de 19 Cajas Rurales, y constituyéndose actualmente como grupo consolidable de entidades de crédito autorizado por el Banco de España.

Bajo el Grupo Cajamar se extiende una red conformada por más de 1.191 oficinas y 6.036 empleados, gestionando un activo total de 39.166 millones de euros, y presente en la práctica totalidad de la geografía española.

Este Grupo se organiza en torno a la figura del Banco de Crédito Social Cooperativo, constituido en 2014 y que actúa como central financiera de las Cajas Rurales que conforman el Grupo Cooperativo Cajamar y otras Cajas Rurales asociadas, siendo su principal accionista Cajamar.

3.4.2 Fuentes de información

A continuación se exponen las fuentes utilizadas para la construcción de la población necesaria para la realización del análisis, la identificación de las Cajas Rurales que cesaron su actividad en el periodo de estudio y la información financiera necesaria de cada una de ellas.

Para la obtención de la información para la construcción de la población se ha recurrido a Banco de España, la Unión Nacional de Cooperativas de Crédito y el Grupo Cooperativo Cajamar.

Para la determinación del valor tomado por la variable dependiente se ha recurrido al Registro de Entidades de Banco de España, constatándose la disponibilidad de información sobre las Cajas Rurales que cesaron actividad desde el año 1965 hasta el año 2015.

Para la obtención de la información económica y financiera necesaria para la construcción de las variables independientes o explicativas se ha recurrido a las memorias sectoriales publicadas anualmente por la Unión Nacional de Cooperativas de Crédito, siendo las Memorias disponibles las correspondientes al periodo 1974 -2015.

3.4.3 Limitaciones para la obtención de la información.

Las Memorias anteriores a 1985 tan solo recogen unas breves notas sobre las principales magnitudes económicas y financieras presentadas por cada una de las Cajas Rurales, insuficientes para la construcción de las ratios financieras que serán descritas en epígrafe posterior, por ello se desecha incluir esta parte de la población en la muestra.

Las Memorias Anuales de los años 1986, 1987, 1988 y 1989 no recogen información económica y financiera de todas las Cajas Rurales existentes en dichos años, sobre todo de aquellas que cesaron actividad, de ahí que se descarte la utilización de dichos años en la muestra a utilizar en esta Tesis.

A partir de 1990, y considerando los datos disponibles sobre Cajas Rurales operativas e inactivas en el Registro de Banco de España, se constata la existencia de diecisiete Cajas Rurales sobre las que no se dispondría de información económica y financiera del ejercicio económico anterior al de la situación de insolvencia.

De las diecisiete Cajas Rurales, un total de seis habrían causado baja en los primeros días del año, mientras que las once restantes se habrían producido a lo largo del ejercicio.

Siguiendo el criterio de Bologna (2011) se opta por incluir a las Cajas Rurales cuya baja se produjo en los primeros días del ejercicio por asemejarse a bajas que se hubieran producido a finales del año anterior, siendo la información disponible para estas entidades la correspondiente a t-2, eliminándose el resto por falta de información.

En la Tabla 7 se desglosan las Cajas Rurales que finalmente han sido eliminadas de la muestra por disponer de información económica y financiera que no cumple con el criterio establecido en esta Tesis.

TABLA 7: CASOS ELIMINADOS DE LA MUESTRA

CASO	AÑO BAJA	MOTIVO EXCLUSIÓN
1	1992	Últimos EE.FF. 1990
2	1992	Últimos EE.FF. 1989
3	1993	Últimos EE.FF. 1990
4	1993	Últimos EE.FF. 1989
5	1994	Últimos EE.FF. 1991
6	1994	Últimos EE.FF. 1991
7	1999	Últimos EE.FF. 1997
8	2001	Últimos EE.FF. 1998
9	2001	Últimos EE.FF. 1998
10	2001	Últimos EE.FF. 1999
11	2008	Últimos EE.FF. 2006

FUENTE: ELABORACIÓN PROPIA

El cambio de moneda (pesetas a euros) obliga a convertir la divisa para los EE.FF. anteriores al año 2000, siendo el tipo de cambio oficial utilizado el publicado por Banco de España, es decir, un euro es equivalente a 166,386 pesetas.

3.4.4 Muestra para la estimación y validación del modelo.

Siguiendo el criterio de la mayor parte de los autores (Kumar *et al.*, 1995; Anastasi *et al.*, 1998; Lane, 1999; Magalhaes, 2001; Serra *et al.*, 2002; Ayala *et al.*, 2007a; Li *et al.*, 2011; Poghosyan *et al.*, 2011; Trussel *et al.*, 2012; y Apergis *et al.*, 2013, entre otros) se elaboraran dos muestras de forma aleatoria, una que será utilizada para la estimación de los modelos, y otra que servirá para comprobar la validez de los resultados.

La primera muestra estará compuesta por el 2/3 partes de las Cajas Rurales que hayan cumplido con la situación de insolvencia, reservándose 1/3 parte para la validación de los modelos. Ambas corresponderán a muestras de datos de panel en la que se dispondrá de observaciones para cada Caja Rural desde el año 1991 hasta 2015.

3.4.5 Segmentación temporal

Como se ha expuesto anteriormente, se trata de una muestra de datos de panel que se inician en 1991 y finalizan en 2015, un total de veinticinco años de actividad en los que se han registrado acontecimientos coyunturales que han podido afectar a la calidad de los datos financieros que se incorporarán al modelo.

Para realizar la segmentación temporal de la muestra se considerarán los intervalos temporales establecidos por las crisis bancarias que ha experimentado la economía española durante el periodo objeto de estudio.

Aunque Pons (2015) afirma que en el último siglo y medio la economía española habría experimentado un total de dieciocho crisis financieras, para Sudriá (2014) tan solo habría acaecido una crisis bancaria pura, la que se produjo a mediados de los años 1920 y que se derivó del ajuste económico y financiero una vez acabada la guerra europea por el ajuste del superávit comercial y por cuenta corriente del conflicto bélico.

Aunque el resto de crisis financieras habrían estado relacionadas con otras crisis (bursátiles y financieras), estas también afectaron sensiblemente al sector financiero español (Beltrán *et al.*, 2012).

En este sentido, y siguiendo a Beltrán *et al.* (2012), Sudriá (2014) y Pons (2015), podemos distinguir las siguientes crisis en la economía española:

- **Crisis de 1982 - 1985**, en la que el sistema bancario español quedó significativamente reducido y más saneado como consecuencia de la crisis que afectó al sector industrial en el contexto del incremento de los precios del petróleo, junto con su obsolescencia técnica, pérdida de competitividad y alta dependencia del crédito financiero. De hecho este último acontecimiento se configuró como el mecanismo de transmisión de la crisis industrial a la crisis bancaria, dado que la quiebra empresarial

afectó a la calidad de los activos bancarios, causando finalmente la liquidación de aquellas entidades menos preparadas para soportar esta situación, en torno a 56 entidades financieras sobre un total de 110, obligando a que finalmente fuesen intervenidas un total de 29 entidades, a las que se unirían 7 fusiones, 4 liquidaciones y la nacionalización de una veintena de entidades (Pons, 2015).

- **Crisis de 1991 -1995**, principalmente fue crisis monetaria y bursátil. Esta crisis tuvo sus orígenes en la burbuja inmobiliaria de Japón y la tensiones en torno a los precios del petróleo, parcialmente incentivadas por la Guerra del Golfo, y aunque es cierto que en España los efectos de la políticas keynesianas de incremento del gasto público (en el contexto de la Exposición Universal de Sevilla y las Olimpiadas de Barcelona) retrasaron las consecuencias de esta crisis mundial sobre nuestra economía, la finalización de esas grandes obras tuvieron consecuencias negativas sobre las Cuentas Públicas españolas, incrementando el déficit y el volumen de deuda pública sobre PIB, situación que se agravó por una gran sequía que se extendió hasta el año 1993 y que obligó al Gobierno a elaborar una Ley Financiera para tratar de amortiguar esta situación.

En el ámbito bancario Franganillo *et al.* (1995) indica que, tras la crisis bancaria de la década de los ochenta y la firma del Tratado de Maastrich, el sector bancario estuvo expuesto a cambios regulatorios que fomentaron su concentración. En este sentido entre 1990 y 1993 se produjo una mayor concentración del volumen de activos en torno a grandes grupos bancarios españoles, si bien la presencia de banca extranjera se incrementó en casi un 40% con respecto al número de entidades existentes a cierre de 1985.

- **Crisis de 2007 – 2013**, alentada principalmente por el crecimiento del crédito (Pons, 2015), todo ello como consecuencia de un exceso de liquidez en el sistema financiero.

De hecho, como indica Alvarez (2008) es frecuente encontrar en la historia económica episodios de crisis bancarias que han estado precedidas de una fase de abundante liquidez. Aunque en los primeros momentos de esta crisis la banca española, centrada principalmente en la banca minorista, se mantuvo estable, su larga duración y sus efectos sobre la valoración de los activos inmobiliarios terminó por afectar al sistema financiero español, obligando a la reestructuración y concentración sectorial, principalmente de las Cajas de Ahorro que desaparecieron en su totalidad. Situación similar fue la vivida por algunos bancos, que también terminarían siendo absorbidos por otros grupos bancarios de mayor tamaño, fomentando así la concentración sectorial.

Considerando las tres crisis financieras anunciadas anteriormente, en esta Tesis se dividirá la muestra en tres periodos:

- a. Periodo primero, que abarcaría los ejercicios económicos comprendidos entre 1991 y 1995.
- b. Periodo segundo, comprendido entre 1996 y 2007
- c. Periodo tercero, que incluirá los ejercicios económicos entre 2008 y 2015.

3.5 VARIABLE DEPENDIENTE

Para la estimación de los modelos estadísticos que permitan alcanzar el objetivo perseguidos con esta Tesis se define una variable dependiente de tipo dicotómico que tomará el valor 1 en el caso de que la Caja Rural haya dejado de existir a lo largo del periodo de estudio, y el valor 0 en el resto de situaciones.

En este punto surge una de las principales dificultades, que no es otra que la determinación de aquellos casos de Cajas Rurales que hayan dejado de existir por motivos de sostenibilidad financiera.

Como recapitula Fernández (2015), dentro del conjunto de investigaciones que versan sobre la predicción de insolvencias bancarias la determinación de la variable dependiente (la insolvencia de la entidad financiera) se ha realizado recurriendo a partir de los siguientes criterios:

- a. Sobre características cualitativas, no exentas de juicio de valor (Fernández, 2015), bien sea recurriendo a lo dispuesto en las distintas leyes aplicables, la normativa propia del regulador o supervisor, o bien por la realización de determinadas operaciones societarias (fusiones principalmente).
- b. En base a características cuantitativas, principalmente el nivel de capitalización, la existencia de pérdidas que pudieran derivar en una continua descapitalización de la entidad, la morosidad de la cartera de inversiones crediticias o la falta de liquidez, entre otros, si bien Fernández (2015) señala que este tipo de criterio también está expuesto a críticas porque determinadas operaciones contables pueden dilatar en el tiempo el verdadero valor que deberían tomar esas ratios financieras.
- c. Combinación de los dos procedimientos anteriores.

En el caso concreto de las cooperativas de crédito la determinación de esa situación de insolvencia se complica tanto por las propias características de este tipo de entidades como del sector en el que opera, en el que es habitual encontrar procesos de fusión o asistencia financiera entre ellas, como defienden Porath (2006), Cabo *et al.* (2010) y Lima (2012), que buscan la sostenibilidad financiera de las entidades implicadas.

En este sentido en la Tabla 8 se ha procedido a clasificar las investigaciones sobre insolvencia de cooperativas de crédito en función del criterio utilizado para la determinación de la variable dependiente, observándose que en la mayor parte de ellas se recurre a la consideración de aspectos cualitativos, no habiendo sido localizadas investigaciones que recurran en exclusiva a condiciones cuantitativas, siendo dos las publicaciones que combinan ambos tipos de criterios. Al no existir uniformidad entre el conjunto de investigadores, en esta Tesis se han combinado ambos criterios.

TABLA 8: CRITERIOS SEGUIDOS EN LA DETERMINACIÓN DE LA VARIABLE DEPENDIENTE EN LA INVESTIGACIÓN SOBRE INSOLVENCIA DE COOPERATIVAS DE CRÉDITO.

MÉTODO	INVESTIGACIONES
Cualitativo	Dabós (1.996), Gama <i>et al</i> (2.004), Braga <i>et al</i> (2.006), Gama <i>et al</i> (2.011), Lima (2.012), Fiordelisi (2.013) y Carvalho (2.015)
Cuantitativo	No encontradas
Mixto	Porath (2.006) y Cabo <i>et al</i> (2.010)

FUENTE: ELABORACIÓN PROPIA.

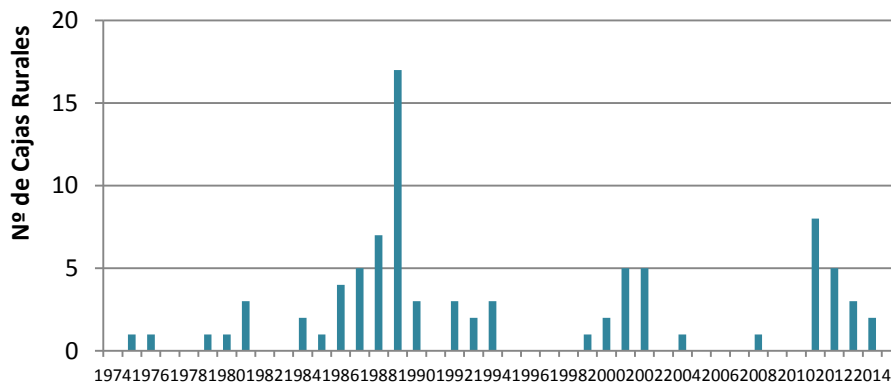
Concretamente, en la determinación de las características cualitativas se recurrirá al registro de entidades de Banco de España, que considera que una entidad causa baja por alguno de los siguientes motivos:

- a. Por acuerdo de la Comisión Ejecutiva de Banco de España.
- b. Por cesión global de activos y pasivos.
- c. Por disolución
- d. Por fusión por absorción
- e. Por fusión y creación de una nueva entidad
- f. Otros motivos

A partir de este criterio se observa que en el Registro de Entidades de Banco de España estaban censadas en torno a 150 Cajas Rurales a principios de 1974, cifra que fue reduciéndose hasta las 59 actuales principalmente por fusiones entre sus miembros.

Aunque de la Ilustración 7 se desprende que la mitad de las bajas de Cajas Rurales se produjeron entre 1974 y 1990, al carecer de información económica y financiera suficiente para la totalidad de las mismas, se opta por delimitar la dimensión temporal de la muestra al periodo 1991-2015, en el que desaparecieron un total de 42 Cajas Rurales, no habiéndose constatado bajas de Cajas Rurales en el Registro de Banco de España en el año 2015.

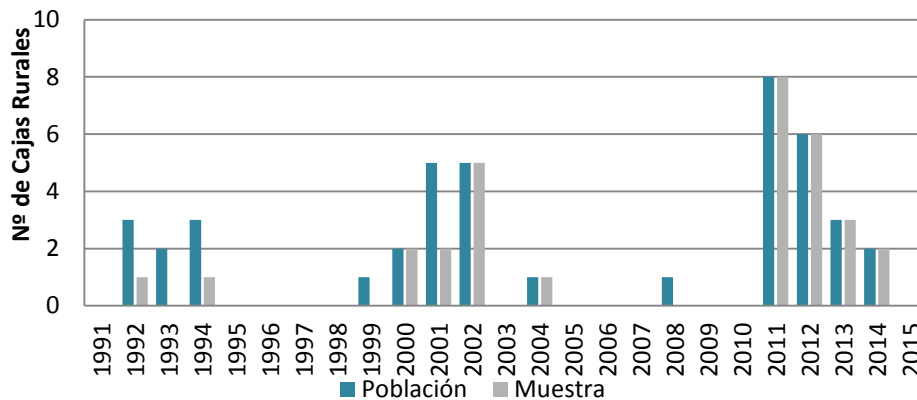
ILUSTRACIÓN 7: DISTRIBUCIÓN DE LAS BAJAS DE CAJAS RURALES ENTRE 1974-.015.



FUENTE: ELABORACIÓN PROPIA

Dadas las limitaciones de información indicadas en el epígrafe anterior, la muestra de Cajas Rurales que cumple con el criterio cualitativo queda reducida a un total de 31 Cajas Rurales (en su totalidad por fusiones con otras entidades), siendo su distribución temporal la detallada en la Ilustración 8.

ILUSTRACIÓN 8: CAJAS RURALES INSOLVENTES INCORPORADAS A LA MUESTRA COMPARADA CON LA POBLACIÓN DEL PERIODO.



FUENTE: ELABORACIÓN PROPIA

En lo que respecta al criterio cuantitativo, Porath (2006) considera que una cooperativa de crédito se encuentra en una situación problemática si, además de cumplir con alguno de los supuestos establecidos por el supervisor, presenta pérdidas superiores al 25% del capital.

Por su parte Cabo *et al* (2010) en su análisis de la banca cooperativa portuguesa, considera, además de características cualitativas similares a las utilizadas en esta Tesis, que una cooperativa de crédito manifiesta una situación delicada si presenta un patrimonio neto negativo.

Aplicando los criterios de ambos autores al caso de las Cajas Rurales españolas no se constata la presencia de ninguna Caja Rural con patrimonio neto negativo en ninguno de los ejercicios económicos objeto de estudio, si bien sí que se observa la existencia de Cajas Rurales que cumplirían con el criterio utilizado por Porath (2006).

En este sentido, en la Tabla 9 se puede observar que del total de treinta y una Cajas Rurales que causaron baja en el Registro de Entidades de Banco de España, tan solo cuatro mostraron pérdidas superiores al 25% del capital social en el año previo a la fusión.

Por el contrario se observa la existencia de veintiséis Cajas Rurales que, no habiendo cumplido con el criterio de haber sido dada de baja del Registro de Banco de España, habrían presentado pérdidas superiores al 25% del capital social en algún momento del periodo analizado.

TABLA 9: CLASIFICACIÓN DE LAS CAJAS RURALES QUE CUMPLEN CON LOS CRITERIOS DE INSOLVENCIA ESTABLECIDOS.

MÉTODO	CAJAS RURALES QUE CUMPLEN CON CADA CRITERIO	CAJAS RURALES QUE CUMPLEN CON AMBOS CRITERIOS
Cualitativo	31	4
Cuantitativo	26	4

FUENTE: ELABORACIÓN PROPIA

En esta Tesis se propone utilizar un criterio cuantitativo que combina el criterio de Porath (2006) con el comportamiento mostrado por las ratios que miden el apalancamiento y las provisiones:

- a. La ratio de apalancamiento mide la necesidad de financiación ponderando los Fondos Propios de la Caja Rural en función del Activo total, y para el que los Acuerdos de Basilea III establecen un valor mínimo del 3%.
- b. Esfuerzo en provisiones mide las dotaciones de provisiones por insolvencias sobre el margen neto.

En este sentido en la Tabla 10 se recoge un análisis individualizado de cada uno de los procesos de fusión de Cajas Rurales a partir de los criterios cuantitativos enunciados anteriormente.

TABLA 10: ANÁLISIS DE LOS PROCESOS DE FUSIÓN DE CAJAS RURALES

ID	Porath (2006)	Porath (2006) histórica	FGD	APALANCAMIENTO			ESFUERZO EN PROVISIONES		
				Cuartil inferior (año anterior)	Cuartil inferior (histórico)	Cuartil medio (año anterior)	Cuartil inferior (año anterior)	Cuartil inferior (histórico)	Cuartil medio (año anterior)
F1					X	X		X	X
F2								X	
F3					X	X		X	X
F4		X	X	X	X	X	X	X	X
F5					X	X	X	X	X
F6					X		X	X	X
F7				X	X	X	X	X	X
F8					X			X	
F9					X			X	X
F10									X
F11									X
F12			X		X	X		X	
F13					X	X	X	X	X
F14			X	X	X	X		X	X
F15				X	X	X	X	X	X
F16				X	X	X		X	
F17					X			X	

(Continuación Tabla 10)

ID	Porath (2006)	Porath (2006) histórico	FGD	APALANCAMIENTO			ESFUERZO EN PROVISIONES		
				Cuartil Inferior (año Anterior)	Cuartil Inferior (histórico)	Cuartil Medio (año anterior)	Cuartil Inferior (año Anterior)	Cuartil Inferior (histórico)	Cuartil Medio (año anterior)
F18		X			X			X	
F19				X	X	X		X	
F20	X	X		X	X	X	X	X	X
F21								X	X
F22				X	X	X			
F23	X	X			X		X	X	X
F24					X	X	X	X	X
F25				X	X	X			X
F26	X	X		X	X	X		X	
F27					X	X		X	
F28									
F29	X	X		X	X	X	X	X	X
F30							X	X	X
F31						X	X	X	X

FUENTE: ELABORACIÓN PROPIA

La columna denominada “Porath (2006)” recoge aquellas Cajas Rurales fusionadas que cumplirían con el criterio de presentar pérdidas superiores al 25% del capital social en el ejercicio económico inmediatamente anterior al proceso de fusión.

La segunda columna, denominada “Porath (2006) histórica”, refleja aquellas Cajas Rurales fusionadas que cumplirían con el criterio de presentar pérdidas superiores al 25% de capital social en alguno de los ejercicios anteriores a la ocurrencia del proceso de fusión.

Dentro de la columna denominada “FGD” se recogen aquellas Cajas Rurales fusionadas que habrían sido asistidas por el Fondo de Garantía de Depósitos en años anteriores al proceso de fusión.

En lo que respecta a la variable “apalancamiento”, se han clasificado a las Cajas Rurales en función de su posición dentro del peor cuartil calculado para el conjunto de Cajas Rurales (incluyendo fusionadas y no fusionadas), bien sea en el ejercicio económico anterior al

proceso de fusión (recogido en la columna cuartil inferior año anterior), bien en algún año anterior al proceso de fusión (columna cuartil inferior histórico). Adicionalmente, se ha incluido la clasificación de aquellas Cajas Rurales que se sitúan dentro del cuartil central en el año anterior al proceso de fusión (dentro de la columna cuartil medio año anterior).

De forma similar a lo descrito para la variable apalancamiento, el tratamiento de las Cajas Rurales fusionadas en función de la variable “Esfuerzo en provisiones” las clasifica en función de su posicionamiento en el peor cuartil, tanto con datos del año anterior como de toda la serie histórica, complementado con el cuartil medio del año anterior.

Según los resultados recogidos en la Tabla 10, ninguno de los procesos de fusión cumple con todos los criterios cuantitativos considerados en esta Tesis, si bien es cierto que la totalidad de las Cajas Rurales fusionadas, con la excepción de un único caso, cumpliría con la mayor parte de ellos.

Este hecho puede ser considerado un primer indicio de que los procesos de fusión de Cajas Rurales pudieron ser debido a motivos de solvencia de las entidades implicadas, justificando además su consideración para la estimación de modelos que permitan la previsión de estas situaciones en un futuro, si bien se descarta la inclusión de la Caja Rural fusionada que no cumple con ninguno de los criterios cuantitativos fijados.

3.6 VARIABLES EXPLICATIVAS

De la revisión bibliográfica llevada a cabo en el Capítulo Segundo se desprende que en la mayor parte de las investigaciones sobre insolvencia empresarial y bancaria juega un papel fundamental la información económica y financiera intrínseca de los propios individuos incluidos en los estudios.

Para la selección de dicha información, los diferentes autores analizados han seguido los siguientes criterios:

- a. Mediante el uso de las ratios integradas en el CAMEL, como se observa en Tam *et al.* (1992), Cole *et al.* (1995b), Hirtle *et al.* (1999), Wheelock *et al.* (2000); Sarkar *et al.* (2001), Nuxoll (2003); Andersen (2008), Poghosyan *et al.* (2011) y Crespo (2011) entre otros.
- b. Los que recurren a ratios utilizadas por reguladores u otras instituciones, como puede ser Borges *et al.* (1996) que recurre a las ratios recogidas en manuales no académicos para estimar un modelo de previsión de quiebras bancarias en Brasil; Bell (1997) que estima un modelo de previsión de insolvencia bancaria mediante redes neuronales y a partir de una serie de ratios financieras extraídas del informe elaborado por organismos gubernamentales de EE.UU.; Magalhaes (2001) que recurre a los *Indicadores Econômico-financeiros de Instituições sob Acompanhamento e Controle* (en adelante CAREL) del Banco Central de Brasil para construir un modelo de supervivencia que permita prever la insolvencia bancaria en Brasil; o Bernhardsen (2001) que utiliza el modelo SEBRA.
- c. El recurso a ratios previamente utilizadas por investigadores anteriores como ocurre en Sinkey (1975), Kumar *et al.* (1995), Anastasi *et al.* (1998), Laitinen *et al.* (1999), Shumway (2001), Logan (2001) o Tan *et al.* (2007), entre otros.

Además de la información económica y financiera, existen autores que dotan de importancia a variables externas, sobre todo el entorno macroeconómico o el equipo gestor.

En este sentido Meyer *et al.* (1970) observó que la situación económico financiera por sí sola no era suficiente para explicar la quiebra de un banco, afirmando que también

mostraban relevancia las variables que medían el entorno macroeconómico, la calidad del equipo gestor y, sobre todo, la profesionalidad de los empleados.

Martin *et al.* (2012) encontró evidencia suficiente para afirmar que una de las principales causas de la crisis bancaria que asoló España en 1866 fue la presencia de gestores inexpertos que de forma arriesgada tomaron recursos a corto plazo para financiar activos de larga maduración, principalmente acciones de empresas ferroviarias.

Nuxoll (2003) que no dudaba sobre el papel fundamental del equipo gestor, no encontró evidencia empírica suficiente para cuantificar la influencia que ejercían las variables macroeconómicas, por lo que no pudo aclarar la utilidad de incluir este tipo de información en los modelos predictivos.

En el ámbito de estudio de las Cajas Rurales, y a pesar de ser pocas las publicaciones relativas a la insolvencia de cooperativas de crédito, en la Tabla 11 se han recogido todos aquellos inputs que han resultado ser estadísticamente significativos en el análisis insolvencia de cooperativas de crédito.

TABLA 11: PRINCIPALES INDICADORES UTILIZADOS EN LAS INVESTIGACIONES SOBRE INSOLVENCIAS DE COOPERATIVAS DE CRÉDITO.

AUTORES	AÑO	VARIABLES SIGNIFICATIVAS	TÉCNICA
Dabós	1996	Patrimonio Neto / Activo Pasivo / Patrimonio Neto (Disponibilidades + Títulos públicos) / Depósitos (Patrimonio Neto – Inmovilizado) / Pasivo Gastos Explotación / Pasivo (Cartera irregular – provisiones) / Patrim. Neto ROE	Probit
Gama <i>et al.</i>	2004	Patrimonio Neto / Pasivo total Activos líquidos / Pasivo total Depósitos / Depósitos periodo anterior	Logit
Braga. <i>et al.</i>	2006	Liquidez Gastos salariales Crédito a la clientela / Fondos Propios	Análisis de Supervivencia
Porath	2006	Capital Reservas no distribuidas Pérdidas ocultas	Modelo Binario para Datos de Panel.

(Continuación Tabla 11)

AUTORES	AÑO	VARIABLES SIGNIFICATIVAS	TÉCNICA
Cabo <i>et al.</i>	2010	Cartera crediticia / Depósitos de la clientela Gastos administrativos / Activo total	Análisis de Supervivencia
Gama <i>et al.</i>	2011	Operaciones vencidas / Cartera total Capital / Activo total Activos ajenos / Activo total Ingresos / Gastos Administrativos	Logit
Lima	2012	Fallidos / Inversión crediticia Equipo gestor (gastos operativos / ingresos) Liquidez (depósitos / créditos)	Logit
Fiordelisi <i>et al.</i>	2013	Fondos Propios Inversiones crediticias Calidad de los activos Liquidez Tamaño Tasa de crecimiento del PIB	Análisis de Supervivencia
Carvalho <i>et al.</i>	2015	Crédito a la clientela / Activo total Tesorería / Activo total Gastos administrativos/ Activo total Depósitos / Pasivos financieros Pasivos sin coste / Pasivos financieros Ingresos financieros / Activo totales Activo fijo / (Patrimonio Neto + Pasivo a lp) Beneficio / Ingresos Gastos administrativos / Depósitos Otros ingresos operativos / Ingresos totales Comisiones / Ingresos operativos Ln (activo total)	Análisis de Supervivencia

FUENTE: ELABORACIÓN PROPIA

Al igual que sucede en el estudio de la insolvencia bancaria, para la determinación de las variables explicativas a utilizar en los modelos de previsión de insolvencias de cooperativas de crédito es habitual que los autores recurran:

- a. Variables que resultaron ser estadísticamente significativas en otras investigaciones, como ocurre en Braga *et al.* (2006), Porath (2006) y Cabo *et al.*, (2010);
- b. Otras técnicas como el Sistema PEARLS (Gama *et al.*, 2011), las ratios incluidas en el Sistema CAMEL (Lima, 2012; y Fiordelisi *et al.*; 2013) u otras técnicas del ámbito no académico (Gama *et al.*; 2004).

Para el objeto de estudio de esta Tesis se utilizarán ratios financieras ya utilizadas en otras investigaciones en las que además hayan resultado ser significativas, junto con ratios sectoriales utilizadas por Banco de España y los Acuerdos de Basilea.

3.6.1 Ratios Financieras

3.6.1.1 Capitalización

La capitalización es una medida de los recursos de los que dispone una entidad financiera para absorber pérdidas futuras que puedan afectar a su viabilidad, de ahí el amplio grado de utilización observado en las investigaciones sobre insolvencias bancarias analizadas.

De hecho, en todas las publicaciones sobre insolvencias de cooperativas de crédito, con la excepción de Braga *et al.* (2006), Cabo *et al.* (2010) y Lima (2012), esta medida resulta ser significativa a través de ratios financieras que adoptan diferentes formas.

Woon *et al.* (2011) afirma que una buena medida de la capitalización de una entidad financiera es la ratio TIER 1¹⁹, publicada por los Acuerdos de Basilea de 1988, que pone en valor el capital y los beneficios retenidos sobre los activos ponderados por riesgo, si bien la principal dificultad de este sistema subyace en cómo medir esos activos ponderados por riesgo, más si cabe cuando se trabaja con información contable pública con la que no es

¹⁹ En 1988 el Comité de Supervisión Bancaria de Basilea publicó el primer Acuerdo de Capital (Basilea I) que fijaba el capital mínimo que debían presentar las entidades financieras en función de sus activos ponderados por riesgo, también conocido como TIER 1, y que inicialmente se estableció en un mínimo del 8%. Se complementó mediante Basilea II, nuevos acuerdos que ampliaban la definición de activos ponderados por riesgo, al tiempo que permitió que las entidades financieras estableciesen modelos internos para establecer esas calificaciones de riesgo. Con la crisis financiera de 2007 se observó que estas medidas no fueron suficientes para mantener la integridad del sistema financiero, surgiendo los Acuerdos de Basilea III en 2008 mediante los que se buscan reforzar el colchón de capital en años de bonanza, al tiempo que se complementa con una nueva medida del apalancamiento de las entidades financieras.

posible tener un acercamiento a la calidad de los activos que conforman las masas patrimoniales de las entidades analizadas.

Por ello autores como Dabós (1996), Berger *et al.* (1997), Gama *et al.* (2011), Woon *et al.* (2011) y Li *et al.* (2010) proponen utilizar una ratio más simple que pone en relación el capital sobre el activo total de la entidad, ratio que encuentran significativa en sus respectivas investigaciones sobre insolvencias bancarias, siendo por ello finalmente utilizada en el desarrollo de esta Tesis:

$$\text{CAPITALIZACIÓN} = \frac{\text{Capital}}{\text{Activo total}}$$

Aunque en el epígrafe 3.4.1 de este Capítulo se ha realizado una detallada descripción de cómo se configura el capital en las cooperativas de crédito, simplemente señalar que en España se encuentra determinado por la Ley 13/1989, en la que se establece que las participaciones no podrán ser inferior a 60,10 euros por acción, existiendo además ciertas limitaciones a la participación máxima de los socios en función de si se tratan de personas físicas o jurídicas. Además de ello, cabe señalar que para ser cliente de una cooperativa de crédito es necesario ser socio de la misma.

3.6.1.2 Activo

Goddard *et al.* (2009) y Carvalho *et al.* (2015) observan que la probabilidad de insolvencia de las cooperativas de crédito se reduce cuanto mayor es el tamaño de la propia entidad.

Aunque existen diversas formas de medir el tamaño, en esta Tesis se considerará la ratio utilizada por Carvalho *et al.* (2015) construida a partir del activo total de las Cajas Rurales analizadas.

Al tratarse de una magnitud absoluta, se ha realizado una modificación de la ratio calculando el logaritmo neperiano de la misma:

$$\text{TAMAÑO} = \ln(\text{Activo total})$$

Muñoz (1999) indica que las propias condiciones del sector empresarial tienen un impacto inmediato en la solidez del sistema bancario por condicionar la calidad de la cartera de inversiones crediticias, ya que pueden afectar directamente a los niveles de capital y reservas de las entidades financieras en el caso de que se haga necesario el deterioro de las inversiones en determinadas empresas.

Por ello interesa conocer qué proporción de las inversiones de la Caja Rural están concentradas en el crédito a la clientela, ya que el conocimiento de esta magnitud pone de manifiesto la exposición de la Caja Rural a las insolvencias que pudieran manifestar sus clientes.

En este sentido, la variable NEGOCIO que se define a continuación pone en relación el crédito a la clientela sobre el activo total de la entidad:

$$\text{NEGOCIO} = \frac{\text{Crédito a clientes}}{\text{Activo total}}$$

Finalmente resulta necesario analizar la calidad de la cartera crediticia a través de dos ratios financieras que miden la importancia de las provisiones por insolvencias de clientes²⁰, más teniendo en cuenta que estas provisiones pueden ocasionar una caída de los resultados bancarios que puede afectar a la capacidad para absorber futuras pérdidas.

Para ello se utilizará la ratio denominada COSTE DEL RIESGO, empleada por Banco de España y que busca ponderar la dotación de provisiones por insolvencias sobre el crédito a la clientela; y el ESFUERZO EN PROVISIONES, también empleada por Banco de España, y

²⁰ Salas (2002) afirma que *“las entidades financieras dotan provisiones para insolvencias con la finalidad de cubrir el riesgo de crédito, el más importante al que se enfrenta un banco”*. Hasta el año 2.000 en España existían dos tipos de provisiones, una específica (que cubre los activos dudosos a medida que aparecen) y otra genérica, que se aplica a través de un porcentaje fijo sobre los activos sanos. A partir de ese año se incluye también la provisión estadística que cubre el riesgo latente que aparece en las carteras crediticias de las entidades desde que se concede la operación al cliente.

que pondera las dotaciones de provisiones en función del margen bruto (margen de intereses más las comisiones netas y resultados netos de operaciones financieras):

$$\text{COSTE DEL RIESGO} = \frac{\text{Dotación provisión insolvencias}}{\text{Crédito a clientes}}$$

$$\text{ESFUERZO EN PROVISIONES} = \frac{\text{Dotación provisión insolvencias}}{\text{Margen bruto}}$$

3.6.1.3 Equipo gestor

El equipo gestor es determinante en la propia evolución de la entidad, tal y como observan Meyer *et al.* (1970), Clarke (1988), Nuxoll (2003) y Martin *et al.* (2012), entre otros.

De hecho, Clarke (1988) afirma que las deficiencias del equipo gestor es una de las principales causas que explicó la crisis de entidades financieras en EE.UU. en la década de 1980.

Esta ratio es difícilmente cuantificable, si bien Clarke (1980) indica que una de las principales funciones del equipo gestor no es otra que la toma de decisiones sobre la política de inversiones de una entidad financiera, proponiéndose medir la capacidad del equipo gestor a través de la evolución interanual de la cartera de inversiones crediticias que se recoge en la siguiente ratio:

$$\text{EQG1: } \ln \left[\frac{\text{Credito a la clientela}_t}{\text{Credito a la clientela}_{t-1}} \right]$$

Existen investigadores que ponen de manifiesto los efectos perjudiciales que sobre la estabilidad financiera tiene un crecimiento rápido del crédito (Salas, 2002), hasta el punto de que una expansión acelerada del crédito impide a los gestores obtener información sobre la capacidad de pago de sus nuevos prestatarios (Muñoz, 1999), perjudicando a la calidad de la cartera de inversiones crediticias.

También se puede tener una aproximación de la capacidad del equipo gestor mediante la medida de la eficiencia, para ello, y siguiendo a Lima (2012), se medirá la proporción de gastos de explotación sobre los ingresos totales de la Caja Rural.

Para medir la eficiencia del negocio en esta Tesis se calculará la proporción de gastos de explotación sobre los ingresos netos, entendiendo estos últimos como la diferencia entre intereses y comisiones percibidas menos los intereses y comisiones satisfechas (también denominado margen bruto), tal y como se recoge a continuación:

$$\text{EFICIENCIA: } \frac{\text{Gastos de Explotación}}{\text{Margen bruto}}$$

3.6.1.4 Rentabilidad

Por su propia naturaleza, las cooperativas buscan maximizar el beneficio para sus asociados, no tanto por los mayores ingresos, sino que más bien por la vía de la eficiencia a partir de la reducción de los gastos (Carvalho *et al*; 2015), llegando incluso a pagar más por los depósitos de lo que finalmente es cobrado a sus clientes por los créditos concedidos (Taylor; 1979).

Como ya se ha especificado en epígrafes anteriores, son muchos los autores que incluyen variables que miden rentabilidad dentro del análisis de la insolvencia de cooperativas de crédito, en base a los cuales en esta Tesis se consideran los siguientes:

- a) Gastos de Explotación sobre Activo. Esta variable tan solo es utilizada en la publicación de Cabo *et al.* (2010), ya que ninguna otra investigación sobre la situación de las cooperativas de crédito la utiliza.

Con esta variable se busca conocer la rentabilidad de la actividad desde la óptica de los gastos de explotación, entendiendo que cuanto menor sea el valor de la ratio, menor

será la probabilidad de que la cooperativa incurra en pérdidas, básicamente porque las cooperativas con menor ratio estarán explotando un negocio más rentable.

$$\text{RENTAB} = \frac{\text{Gastos de Explotación}}{\text{Activo total}}$$

- b) Rentabilidad del Activo (*Return on Assets* o *ROA* en sus siglas inglesas): es tradicional el uso de esta ratio en la mayor parte de las investigaciones en materia de insolvencia de entidades financieras. Esta magnitud trata de cuantificar la rentabilidad del activo como el beneficio sobre el activo de la entidad.

Cabe tener en cuenta que la definición original de esta ratio utiliza el resultado antes de impuestos para eliminar la influencia fiscal. Sin embargo, las limitaciones halladas en la información económica y financiera disponible para cada una de las Cajas Rurales obliga a realizar una modificación en la ratio original consistente en el empleo del resultado neto (después de impuestos), tal y como se ha recogido a continuación, modificación que consideramos poco significativa puesto que se están comparando entidades que operan en un mismo entorno fiscal.

$$\text{ROA} = \frac{\text{Resultado neto}}{\text{Activo total}}$$

- c) Rentabilidad del capital (*Return on Equity* o *ROE* en sus siglas inglesas): al igual que ocurre con el *ROA* también resulta tradicional el empleo de esta ratio en el análisis financiero, principalmente en el ámbito profesional, si bien es cierto que Dabós (1.996) lo aplica en su análisis de la insolvencia de cooperativas de crédito.

Aunque lo habitual es partir del resultado antes de impuestos, para evitar los efectos fiscales sobre el resultado ya comentados, por las limitaciones en la información económica y financieras de cada una de las Cajas Rurales y considerando que todas operan en un mismo entorno fiscal, finalmente se opta por realizar una modificación

de la ratio financiera a partir del resultado neto, siendo la definición utilizada la siguiente:

$$\text{ROE} = \frac{\text{Resultado neto}}{\text{Fondos Propios}}$$

3.6.1.5 Endeudamiento

Recientemente Bologna (2011) ha analizado la relevancia de la estructura de financiación sobre la insolvencia de entidades financieras, observando que se trata de una de las principales causas en la última crisis económica.

De hecho, los Acuerdos de Basilea III introdujeron una nueva variable para medir la calidad del capital de una Entidad en base a su grado de endeudamiento, medida que además fue adoptada por la Comisión Europea en octubre de 2015 para limitar la excesiva deuda de las entidades financieras.

Esta ratio pondera el capital regulatorio (o fondos propios) sobre el activo total, habiéndose recogido en Basilea III una obligación mínima del 3% a partir del año 2018. Su definición formal es la reflejada a continuación:

$$\text{APALANCAMIENTO} = \frac{\text{Fondos Propios}}{\text{Activo total}}$$

Otra medida del endeudamiento es la que ofrece la ratio LTD, que pone en relación el crédito a la clientela sobre los depósitos de sus propios clientes, entendiendo que cuanto mayor sea la ratio, menor es la necesidad de recurrir a fuentes externas de financiación.

Formalmente esta ratio se puede definir de la siguiente forma:

$$\text{LTD} = \frac{\text{Crédito a la clientela}}{\text{Depósito a la clientela}}$$

Muy relacionado con el anterior es la medida del *Funding Gap* entendida como la necesidad de financiación mayorista y muy utilizada por los estudios que la Fundación BBVA²¹ realiza sobre el sector bancario.

Esta medida se calcula a partir de la diferencia entre los créditos a la clientela y los depósitos de clientes, tal y como se refleja en la siguiente expresión:

$$\text{FUNDING GAP} = \text{Crédito a cliente} - \text{Depósito de clientes}$$

3.6.1.6 Liquidez

La liquidez permite conocer la capacidad de una Entidad para atender sus compromisos financieros a corto plazo sin necesidad de buscar nuevas fuentes de financiación, siendo importante su control ya que de lo contrario puede ocurrir que entidades totalmente solventes puedan verse en una situación de insolvencia como consecuencia de su incapacidad para atender desfases de liquidez.

Autores como Gama *et al.* (2004), Braga *et al.* (2006), Lima (2012) y Fiordelisi *et al.* (2013) encuentran significativa esta ratio en el estudio de la insolvencia de cooperativas de crédito.

Para medir la liquidez se recurrirá a las partidas de caja y depósitos en bancos centrales, por entender que se pueden disponer de forma inmediata para atender posibles requerimientos de los depositantes de la entidad (Männasoo *et al.*, 2009), siendo su definición la siguiente:

$$\text{LIQUID} = \frac{\text{Caja y depósitos en Bancos Centrales}}{\text{Depósitos de clientes}}$$

²¹ La Fundación BBVA es una fundación española propiedad de Banco Bilbao Vizcaya Argentaria que busca promover el conocimiento fomentando la actividad investigadora en diferentes ramas, como son la economía, sociedad, medio ambiente, medicina, ciencias básicas, tecnología y cultura.

3.6.2 Ratios macroeconómicas

Como se ha puesto de manifiesto anteriormente, en el estudio de las crisis bancarias ha sido habitual que los investigadores centren el análisis en las características propias de las entidades que han experimentado la situación de insolvencia, si bien Muñoz (1999) indica que la inclusión de factores macroeconómicos puede ayudar a completar el mismo, ya que la probabilidad de que la insolvencia de una entidad financiera se convierta en sistémica es mayor cuando ésta se debe a condiciones macroeconómicas que pueden afectar a todas las entidades.

Para medir la influencia de macroeconómica se han introducido las siguientes variables:

a. A través de la variable TASA DE VARIACIÓN DEL PIB NACIONAL, calculada como la evolución del PIB a precios corrientes entre un ejercicio y otro, se pretende medir la evolución de la economía española de un año a otro, ya que es habitual que previo a las crisis bancarias se produzcan ciclos recesivos que acentúan la importancia de las crisis bancarias, como pone de manifiesto Muñoz (1999) en su análisis de las insolvencias de bancos latinoamericanos.

Más concretamente, Delgado *et al* (2004) indican que los errores en las políticas crediticias de las entidades suelen cometerse en la fase expansiva del ciclo, acentuando el riesgo de crédito por el crecimiento de la exposición y el paulatino deterioro de la calidad media de los nuevos acreditados. Por su parte Davis (1995) encuentra una estrecha relación entre la evolución del PIB y los tipos de interés, con las quiebras empresariales y bancarias.

b. La TASA DE PARO NACIONAL, calculada como la proporción de parados sobre la población activa a cierre de cada año, puede afectar a las crisis bancarias por su influencia sobre el riesgo de crédito, ya que se trata de un factor condicionante de la renta disponible

de las familias para hacer frente a sus operaciones crediticias, afectando a la necesidad de dotaciones por insolvencias y a la morosidad de las entidades financieras. Entre otros, Brookes *et al* (1994) observan una relación positiva entre tasas de paro y morosidad de los préstamos hipotecarios en Reino Unido.

c. La evolución de los tipos de interés también puede desempeñar un papel determinante en la situación crediticia de las entidades financieras, ya que autores como Delgado *et al.* (2004) ponen de manifiesto que la presencia de tipos de interés reducidos potencia el crecimiento del crédito a la clientela, acentuando a su vez el riesgo de crédito.

Para medir este comportamiento de los tipos de interés se emplearán dos medidas, los TIPOS DE INTERÉS DEL MERCADO INTERBANCARIO A 1 AÑO y los TIPOS DE INTERÉS DEL MERCADO HIPOTECARIO, ambos publicados por Banco de España.

3.6.3 Otras ratios

a. EMPLEADOS: esta variable recoge el dato referente al número de empleados presentado por la Caja Rural al cierre del ejercicio. La variable está expresada en logaritmos neperianos.

b. OFICINAS: incluye dos variables que recogen información sobre el número de oficinas operativas a cierre del ejercicio. Se definen de la siguiente forma:

OF1: calculada a partir del logaritmo neperiano del número de oficinas.

OF2: pondera el resultado neto en función del número de oficinas, tal y como se refleja en la siguiente expresión:

$$OF2 = \frac{\text{Resultado neto}}{\text{Número de oficinas}} \quad (EQ11)$$

c. UBICACIÓN: variable categórica introducida para considerar el ámbito territorial de actuación original. Esta variable toma los valores “norte”, “centro-sur” y “levante-Cataluña”.

d. SECCIÓN, variable que considera el ámbito territorial de la Caja Rural. Toma los valores “comarcal”, “local”, “provincial” y “otros”.

E. FGD: variable dicotómica que considera si la caja rural ha percibido algún tipo de ayudas por parte del fondo de garantía de depósitos a lo largo de su historia. al ser dicotómica tomaría el valor 1 en caso de haber percibido algún tipo de ayuda, y el valor 0 si nunca hubiera percibido ese tipo de asistencia.

F. PORATH: variable dicotómica que toma el valor 1 si se cumple con el criterio utilizado por Porath (2006) para definir si una cooperativa de crédito se encuentra en una situación de insolvencia, es decir, si las pérdidas del ejercicio inmediatamente anterior al de la situación de insolvencia representaron más del 25% del capital social.

En la Tabla 12 siguiente se han recogido las variables financieras, macroeconómicas y sectoriales definidas a lo largo de este epígrafe para el estudio de las Cajas Rurales:

TABLA 12: RESUMEN DE RATIOS FINANCIERAS UTILIZADAS PARA EL ANÁLISIS DE LA SOSTENIBILIDAD FINANCIERA DE LAS CAJAS RURALES.

TIPOLOGÍA	DENOMINACIÓN	RATIOS
CAPITALIZACIÓN	CAPITAL	$\frac{\text{Capital}}{\text{Activo total}}$
ACTIVO	TAMAÑO	log(activo total)
	NEGOCIO	$\frac{\text{Crédito a clientes}}{\text{Activo total}}$
	COSTERIESGO	$\frac{\text{Dotación provisión insolencias}}{\text{Crédito a la clientes}}$
RESULTADOS	ESFUERZOPROV	$\frac{\text{Dotación provisión insolencias}}{\text{Margen bruto}}$
GESTIÓN	EQG1	$\ln \left[\frac{\text{Credito a la clientela}_t}{\text{Credito a la clientela}_{t-1}} \right]$
	EFICIENCIA	$\frac{\text{Gastos de explotación}}{\text{Ingresos netos por intereses y comisiones}}$
RENTABILIDAD	RENTAB	$\frac{\text{Gastos de explotación}}{\text{Activo total}}$
	ROA	$\frac{\text{Resultado}}{\text{Activo total}}$
	ROE	$\frac{\text{Resultado}}{\text{Fondos Propios}}$
ENDAUDAMIENTO	APALANCAMIENTO	$\frac{\text{Fondos Propios}}{\text{Activo total}}$
	LTD	$\frac{\text{Créditos a clientes}}{\text{Depósitos de clientes}}$
	FUNDING GAP	Créditos a clientes - Depósitos de clientes
LIQUIDEZ	LIQUID	$\frac{\text{Caja y depósitos en bancos centrales}}{\text{Depósitos de clientes}}$
MACROECONÓMICAS	PIB	Tasa de crecimiento del PIB
	PARO	Tasa de paro a fin de año
	INTcp	Tipos de interés a corto plazo
	INTlp	Tipos de interés a largo plazo

(Continuación Tabla 12)

TIPOLOGÍA	DENOMINACIÓN	RATIOS
EMPLEADOS	EMPLEADOS	ln (nº empleados a cierre del ejercicio)
OFICINAS	OF1	ln (nº oficinas a cierre del ejercicio)
	OF2	$\frac{\text{Resultado neto}}{\text{Nº de oficinas}}$
TIPO DE CAJA RURAL	SECCIÓN	Tipo de Caja Rural: local/comarcal, provincial, regional o nacional.
TERRITORIO	UBICACIÓN	Territorio de actuación: norte, sur, este u oeste
FGD	FGD	Asistencia por parte del Fondo de Garantía de Depósitos. Variable categórica.
PORATH	PORATH	$\frac{\text{Resultado neto}}{\text{Capital}}$

FUENTE: ELABORACIÓN PROPIA

3.7 CONCLUSIONES DEL CAPÍTULO

A lo largo de este capítulo se ha realizado una descripción detallada de la metodología a emplear para la realización de la parte empírica. A modo de recapitulación se ha llegado a las siguientes conclusiones:

- a. Siguiendo la metodología epistemológica de la economía positiva, se han planteado una serie de hipótesis que son resultado del estudio de otras investigaciones realizadas sobre el sector financiero. Sobre las mismas se intentará demostrar su cumplimiento a lo largo de la parte empírica de esta Tesis.
- b. Muestra: Cajas Rurales existentes entre 1991 y 2015.

- c. Variable dependiente: Cajas Rurales que cumplan con el criterio de Banco de España y se encuentren en el peor cuartil de la ratio que mide el apalancamiento y el esfuerzo en provisiones. Bajo estos dos supuestos se considera un total de 30 Cajas Rurales.
- d. Variables explicativas, en base al sistema CAMEL y los indicadores utilizados por Banco de España en su Boletín Estadístico, además de ratios propuestos en diferente normativas y utilizados mayoritariamente por otros investigadores, se ha realizado una selección de veinticinco ratios clasificadas en cuatro grupos: ratios financieras, ratios coyunturales, ratios sectoriales y otras.
- e. En lo que respecta a la técnica estadística, se empleará un modelo de respuesta binaria en sus diferentes versiones y un modelo de supervivencia de Cox, por observarse en la revisión bibliográfica que la mayor parte de las investigaciones existentes demuestran que estas técnicas ofrecen los mejores resultados.

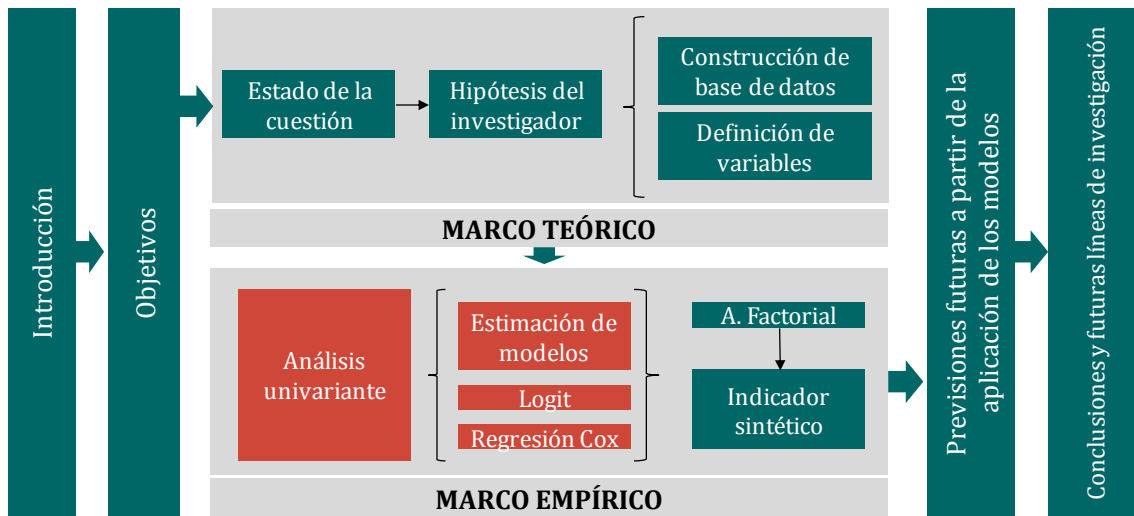
CAPÍTULO CUARTO

RESULTADOS DEL ESTUDIO EMPÍRICO

4.1 SINOPSIS DEL CAPÍTULO

A lo largo de este capítulo se realiza una descripción detallada de los resultados obtenidos de la estimación de los modelos de respuesta binaria y de supervivencia, siendo la ubicación del Capítulo dentro de la sinopsis de la Tesis la reflejada en la Ilustración 9.

ILUSTRACIÓN 9: UBICACIÓN DEL CAPÍTULO CUARTO DENTRO DE LA SINOPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

Se comienza describiendo las principales características de la muestra a partir de una serie de estadísticos descriptivos, con especial mención al caso de las Cajas Rurales que habrían cesado actividad a lo largo del periodo de estudio con el fin de corroborar si los procesos de fusión estuvieron relacionados con la sostenibilidad económica y/o financiera de las entidades implicadas, como ya se adelantó en el capítulo anterior en la descripción de la variable dependiente.

Se continúa detallando los resultados de los modelos de respuesta binaria definidos en el capítulo anterior, logit estático y logit multiperiodo, y el modelo de supervivencia o regresión de Cox, para la posterior comparación de la bondad del ajuste alcanzada entre los tres modelos y su aplicación a Cajas Rurales no incluidas en la muestra inicial.

4.2 ANÁLISIS UNIVARIANTE

Antes de proceder a estimar los modelos indicados en el Capítulo anterior se llevará a cabo un análisis univariante de la muestra utilizada, con el fin de conocer la composición de la misma.

Como ya se ha anticipado la muestra se ha extraído de un panel de datos que incorpora la totalidad de información económica y financiera disponible para cada una de las Cajas Rurales operativas y fusionadas a lo largo del periodo de estudio (1991 - 2015), es decir, un total de 1.764 observaciones, de las que 30 observaciones corresponden a Cajas Rurales fusionadas.

En la Tabla 13 se recoge la distribución de las observaciones en función de la ubicación geográfica de las Cajas Rurales consideradas, observándose que el mayor volumen corresponde a Cajas Rurales ubicadas en el levante español.

TABLA 13: OBSERVACIONES DE CAJAS RURALES INCLUIDAS EN LA MUESTRA

SITUACIÓN DE LA CAJA RURAL	UBICACIÓN GEOGRÁFICA			Total
	Norte	Centro-Sur	Levante-Cataluña	
Operativa	328	490	916	1.734
Fusionada	11	10	9	30

FUENTE: ELABORACIÓN PROPIA

Por su parte, en la Tabla 14 se desglosan los principales estadísticos de la muestra utilizada, confirmándose que las Cajas Rurales fusionadas mostraron políticas de inversión más agresivas, con unas ratios de NEGOCIO, ESFUERZOPROV y COSTERIESGO ligeramente superiores a las presentadas por conjunto de entidades no fusionadas, y alejadas de la media sectorial.

En términos de eficiencia ambos grupos se mantuvieron en sintonía, y aunque es cierto que en relación al activo la rentabilidad fue inferior en el conjunto de las entidades fusionadas, considerando la rentabilidad de los fondos propios la situación fue distinta, con un ROA superior en el grupo de entidades fusionadas.

En este sentido cabe señalar que las entidades fusionadas mostraron, en término medio, los mayores niveles de capitalización, con un grado de apalancamiento y liquidez por debajo de sus homólogas operativas, en parte justificado por esa política de inversión más agresiva ya comentada, y una ratio LTD que, a pesar de ser inferior a uno, se mostraba por encima de la media de las entidades operativas.

Para confirmar las diferencias observadas en los estadísticos descriptivos calculados se recurrirá al análisis de igualdad y homogeneidad de valores medios entre los dos grupos que determina la variable dependiente (operativa y fusionada) a través de una prueba *t* de igualdad de medias, resultados que han sido recogidos en la Tabla 14.

Con esta prueba *t* se contrasta la hipótesis de si las medias poblacionales son iguales (la media de la variable dependiente en cada nivel de la variable independiente), lo que significa que los grupos no difieren en la variable dependiente, y por tanto la variable explicativa es independiente de la variable dependiente.

H₀: las medias poblacionales son idénticas entre los dos grupos de Cajas Rurales

H₁: las medias poblacionales son diferentes entre los dos grupos de Cajas Rurales

En este sentido de los resultados recogidos en la **Tabla 14** se pone de manifiesto la existencia de diferencias en los valores medios únicamente en las variables CAPITALIZACIÓN, APALANCAMIENTO, TAMAÑO, NEGOCIO, ESFUERZOPROV, EFICIENCIA, ROA, ROE, EMP, OF1 y OF2, en las que se observa un valor de la significación inferior al

nivel del 0,05, en el resto de variables no se pondrían de manifiesto diferencias entre ambos grupos de Cajas Rurales operativas y fusionadas.

TABLA 14: PRINCIPALES CARACTERÍSTICAS DE LA MUESTRA.

VARIABLES	SITUACIÓN DE LA CAJA RURAL	DESCRIPTIVOS			PRUEBA t (IGUALDAD MEDIAS)	
		Media	Mínimo	Máximo	t	Sig.
Capital	Operativa	0,0168	0,000	0,1711	2,91	0,004*
	Fusionada	0,0249	0,0034	0,0622		
	Total	0,0169	0,000	0,1711		
Apalancamiento	Operativa	0,0873	-0,002	0,2116	-2,09	0,036*
	Fusionada	0,0758	-0,0091	0,1586		
	Total	0,0871	-0,0091	0,2116		
Tamaño	Operativa	5,1379	3,0637	7,6015	3,16	0,002*
	Fusionada	5,5573	3,7057	6,9525		
	Total	5,145	3,0637	7,6015		
Negocio	Operativa	0,5423	0,0489	4,4747	2,62	0,009*
	Fusionada	0,6618	0,0689	0,8863		
	Total	0,5443	0,0489	4,4747		
CosteRiesgo	Operativa	-0,007	-0,299	0,000	-0,65	0,514
	Fusionada	-0,0085	-0,0952	0,000		
	Total	-0,007	-0,299	0,000		
EsfuerzoProv	Operativa	-0,1212	-5,9878	0,1665	-3,54	0,000*
	Fusionada	-0,2943	-3,8094	0,000		
	Total	-0,1241	-5,9878	0,1665		
EQG1	Operativa	0,0474	-1,1668	1,325	-1,22	0,219
	Fusionada	0,0212	-0,0949	0,2299		
	Total	0,047	-1,1668	1,325		
Eficiencia	Operativa	0,6265	-0,8787	7,6401	3,87	0,000*
	Fusionada	0,7996	0,5366	1,5166		
	Total	0,6295	-0,8787	7,6401		
RENTAB	Operativa	0,0195	0,0024	0,2046	0,22	0,825
	Fusionada	0,0199	0,0102	0,0398		
	Total	0,0195	0,0024	0,2046		
ROA	Operativa	0,0088	-0,0796	0,1264	-6,14	0,000*
	Fusionada	-0,0003	-0,0863	0,0119		
	Total	0,0086	-0,0863	0,1264		
ROE	Operativa	0,1082	-1,1906	2,2444	-4,15	0,000*
	Fusionada	0,1203	-1,1284	0,2360		
	Total	0,112	-1,1906	2,2444		

(Continuación de la Tabla 14)

VARIABLES	SITUACIÓN DE LA CAJA RURAL	DESCRIPTIVOS			PRUEBA t (IGUALDAD MEDIAS)	
		Media	Mínimo	Máximo	t	Sig.
LTD	Operativa	0,7059	0,0619	33,7324	0,71	0,477
	Fusionada	0,8252	0,0957	1,2829		
	Total	0,708	0,0619	33,7324		
FundingGAP	Operativa	-28054	-6639663	6231011	0,27	0,783
	Fusionada	-8768	-433949	1123408		
	Total	-27726	-6639663	6231011		
LIQUID	Operativa	0,0229	0,0016	1,1825	-1,10	0,269
	Fusionada	0,0152	0,0042	0,0536		
	Total	0,0228	0,0016	1,1825		
EMP	Operativa	3,7674	0,6931	8,7808	2,69	0,007*
	Fusionada	4,5901	1,0986	7,5517		
	Total	3,7814	0,6931	8,7808		
OF1	Operativa	2,2144	0,000	7,1831	3,13	0,002*
	Fusionada	3,3272	0,000	6,1924		
	Total	2,2333	0,000	7,1831		
OF2	Operativa	171,8222	-1527,5	1089	-8,62	0,000*
	Fusionada	-233,707	-7126	264,5		
	Total	164,9255	-7126	1089		

* Significativa al 0,05

FUENTE: ELABORACIÓN PROPIA

Adicionalmente se ha llevado a cabo un cálculo de las correlaciones lineales entre las distintas variables explicativas, resultados que han sido recogidos en la Tabla 15, en la que se puede observar la existencia de coeficientes de correlación reducidos entre todas ellas, descartándose por tanto problemas de multicolinealidad.

Por variables, se constata la existencia de los mayores niveles de correlación entre las variables que miden la dimensión (TAMAÑO, EMP Y OF1) y el negocio, con las variables coyunturales y la calidad del negocio (COSTERIESGO y ESFUERZOPROV), principalmente, siendo la mayor parte de los coeficientes de correlación significativos a un nivel del 0,05.

TABLA 15: MATRIZ DE CORRELACIONES ENTRE LAS VARIABLES EXPLICATIVAS

	Capital	Apalancamiento	Tamaño	Negocio	CosteRiesgo	EsfuerzoProv	EQG1	Eficiencia	RENTAB1	ROA	ROE	LTD	FundingGAP	LIQUID	PIB	Paro	INT1	INT2	EMP	OF1	OF2
Capital	1	-,133**	,054*	,114**	-,053*	-,036	,048*	,076**	,277**	,127**	,029	-,010	,071**	,095**	,099**	,003	,092**	,090**	,129**	,202**	-,135**
Apalancamiento	-,133**	1	-,366**	,110**	,161**	,094**	-,047*	-,222**	-,201**	,161**	-,212**	,007	,034	-,137**	-,267**	-,158**	-,319**	-,330**	-,474**	-,517**	,361**
Tamaño	,054*	-,366**	1	,371**	-,008	-,166**	,017	,197**	-,073**	-,248**	-,071**	,150**	,087**	-,137**	-,278**	-,066**	-,283**	-,286**	,943**	,901**	-,176**
Negocio	,114**	,110**	,371**	1	,093**	-,156**	,114**	,129**	,122**	-,018	-,203**	,257**	,162**	,001	-,415**	-,304**	-,514**	-,534**	,329**	,280**	-,089**
CosteRiesgo	-,053*	,161**	-,008	,093**	1	,767**	,071**	-,041	-,128**	,157**	,027	,026	,013	-,009	-,060*	-,221**	-,170**	-,196**	-,058*	-,065**	,241**
EsfuerzoProv	-,036	,094**	-,166**	-,156**	,767**	1	,071**	-,131**	,043	,370**	,194**	-,058*	-,097**	,075**	,218**	-,176**	,124**	,101**	-,130**	-,119**	,405**
EQG1	,048*	-,047*	,017	,114**	,071**	,071**	1	-,022	,074**	,117**	,099**	,036	,165**	-,156**	,183**	-,101**	,094**	,080**	,042	,043	,049*
Eficiencia	,076**	-,222**	,197**	,129**	-,041	-,131**	-,022	1	,498**	-,264**	-,117**	,061	-,020	-,026	-,075**	,012	-,064**	-,070**	,251**	,267**	-,279**
RENTAB1	,277**	-,201**	-,073**	,122**	-,128**	,043	,074**	,498**	1	,245**	,118**	-,016	-,059*	,201**	,260**	,088**	,287**	,295**	,143**	,166**	-,170**
ROA	,127**	,161**	-,248**	-,018	,157**	,370**	,117**	-,264**	,245**	1	,659**	-,075**	-,043	,155**	,356**	-,067**	,292**	,281**	-,171**	-,171**	,555**
ROE	,029	-,212**	-,071**	-,203**	,027	,194**	,099**	-,117**	,118**	,659**	1	-,065**	-,060*	,045	,400**	-,007	,368**	,360**	,011	,024	,223**
LTD	-,010	,007	,150**	,257**	,026	-,058*	,036	,061	-,016	-,075**	-,065**	1	,087**	,188**	-,120**	-,090**	-,144**	-,147**	,130**	,120**	-,050*
FundingGAP	,071**	,034	,087**	,162**	,013	-,097**	,165**	-,020	-,059*	-,043	-,060*	,087**	1	-,017	-,060*	-,011	-,039	-,040	,053*	,030	-,010
LIQUID	,095**	-,137**	-,137**	,001	-,009	,075**	-,156**	-,026	,201**	,155**	,045	,188**	-,017	1	,194**	-,032	,153**	,155**	-,006	,001	-,046
PIB	,099**	-,267**	-,278**	-,415**	-,060*	,218**	,183**	-,075**	,260**	,356**	,400**	-,120**	-,060*	,194**	1	-,216**	,832**	,781**	-,093**	-,065**	,079**
Paro	,003	-,158**	-,066**	-,304**	-,221**	-,176**	-,101**	,012	,088**	-,067**	-,007	-,090**	-,011	-,032	-,216**	1	,172**	,258**	-,049*	-,030	-,139**
INT1	,092**	-,319**	-,283**	-,514**	-,170**	,124**	,094**	-,064**	,287**	,292**	,368**	-,144**	-,039	,153**	,832**	,172**	1	,986**	-,105**	-,075**	,023
INT2	,090**	-,330**	-,286**	-,534**	-,196**	,101**	,080**	-,070**	,295**	,281**	,360**	-,147**	-,040	,155**	,781**	,258**	,986**	1	-,108**	-,077**	,008
EMP	,129**	-,474**	,943**	,329**	-,058*	-,130**	,042	,251**	,143**	-,171**	,011	,130**	,053*	-,006	-,093**	-,049	-,105**	-,108**	1	,973**	-,262**
OF1	,202**	-,517**	,901**	,280**	-,065**	-,119**	,043	,267**	,166**	-,171**	,024	,120**	,030	,001	-,065**	-,030	-,075**	-,077**	,973**	1	-,331**
OF2	-,135**	,361**	-,176**	-,089**	,241**	,405**	,049*	-,279**	-,170**	,555**	,223**	-,050*	-,010	-,046	,079**	-,139**	,023	,008	-,262**	-,331**	1

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

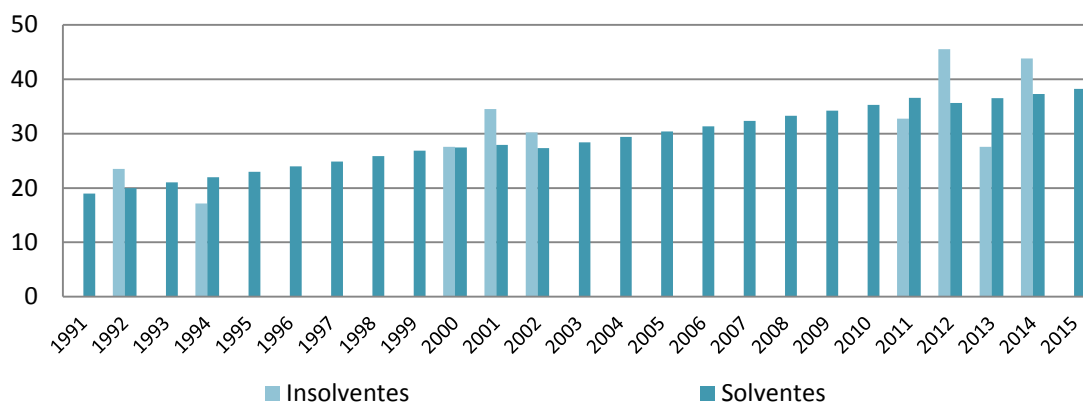
* La correlación es significativa en el nivel 0,05 (bilateral).

FUENTE: ELABORACIÓN PROPIA

En lo que respecta al comportamiento de cada una de las variables explicativas consideradas, a continuación se llevará a cabo un análisis univariante de cada una de ellas.

La Ilustración 10 recoge la antigüedad de las Cajas Rurales a lo largo del periodo objeto de estudio, observándose que en término medio las Cajas Rurales que cesaron actividad presentaron una mayor antigüedad que las que se mantuvieron operativas.

ILUSTRACIÓN 10: ANTIGÜEDAD MEDIA DE LAS CAJAS RURALES ESPAÑOLAS ENTRE 1991 Y 2015.



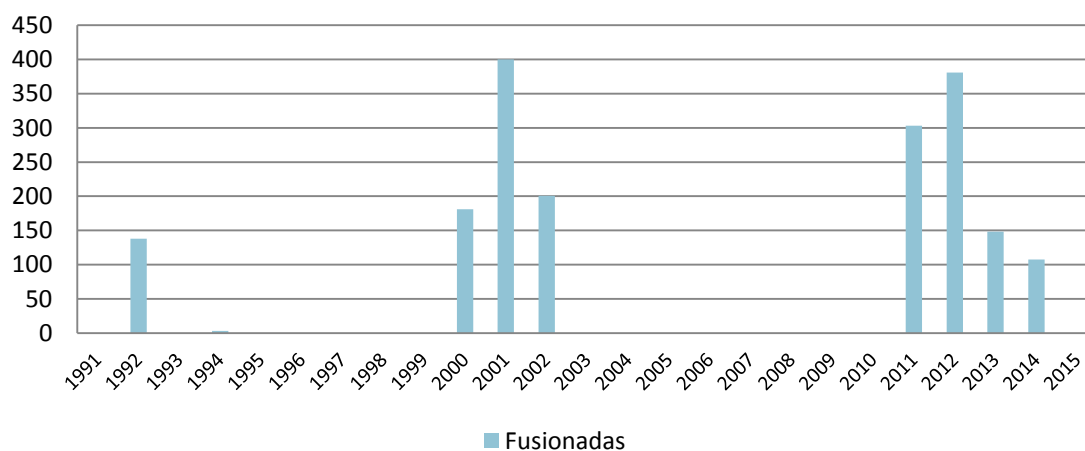
FUENTE: ELABORACIÓN PROPIA A PARTIR DE BANCO DE ESPAÑA.

La Ilustración 11 y la Ilustración 12 recogen la evolución del número medio de EMPLEADOS, diferenciando entre fusionadas y operativas a cierre de cada uno de los ejercicios.

Se puede observar que el grupo de las Cajas Rurales operativas mostró una tendencia creciente en el número de empleados, con cierto estancamiento en los últimos años del estudio.

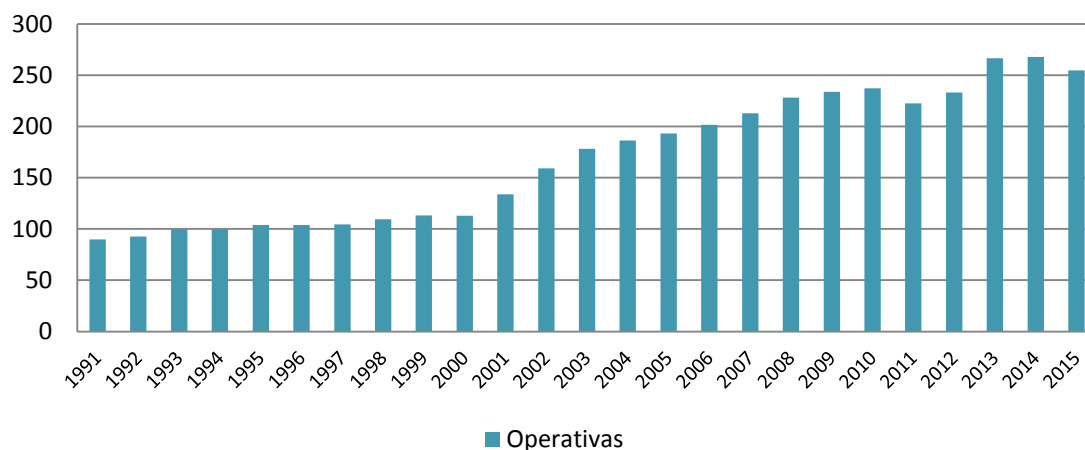
Por su parte, en las Cajas Rurales fusionadas el volumen medio de empleados se mantuvo por encima de la media sectorial en todo momento, hasta el punto de que en la última etapa los niveles presentados por estas entidades fueron los más elevados.

ILUSTRACIÓN 11: EVOLUCIÓN DEL NÚMERO MEDIO DE EMPLEADOS EN LAS CAJAS RURALES FUSIONADAS



FUENTE: ELABORACIÓN PROPIA

ILUSTRACIÓN 12: EVOLUCIÓN DEL NÚMERO MEDIO DE EMPLEADOS DE LAS CAJAS RURALES OPERATIVAS

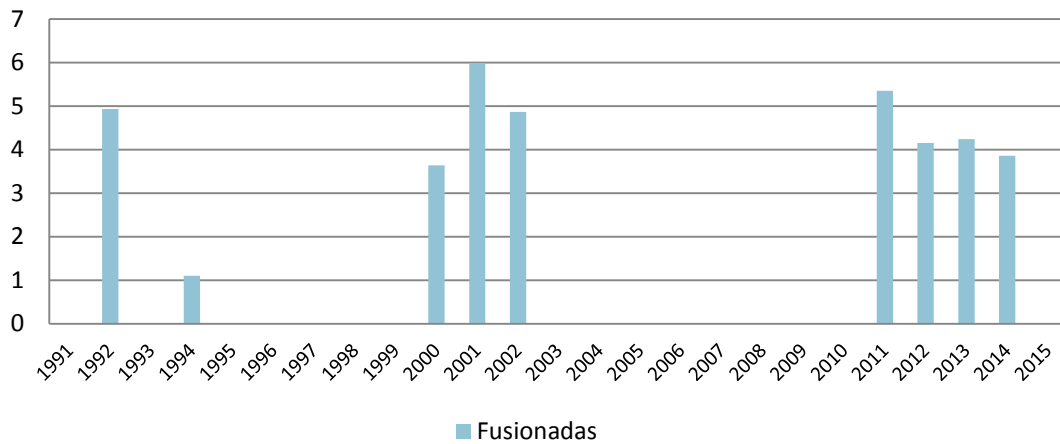


FUENTE: ELABORACIÓN PROPIA

Tendencia similar es la mostrada por el número medio de OFICINAS, creciente a lo largo de todo el periodo, tal y como se refleja en la Ilustración 13 y en la Ilustración 14.

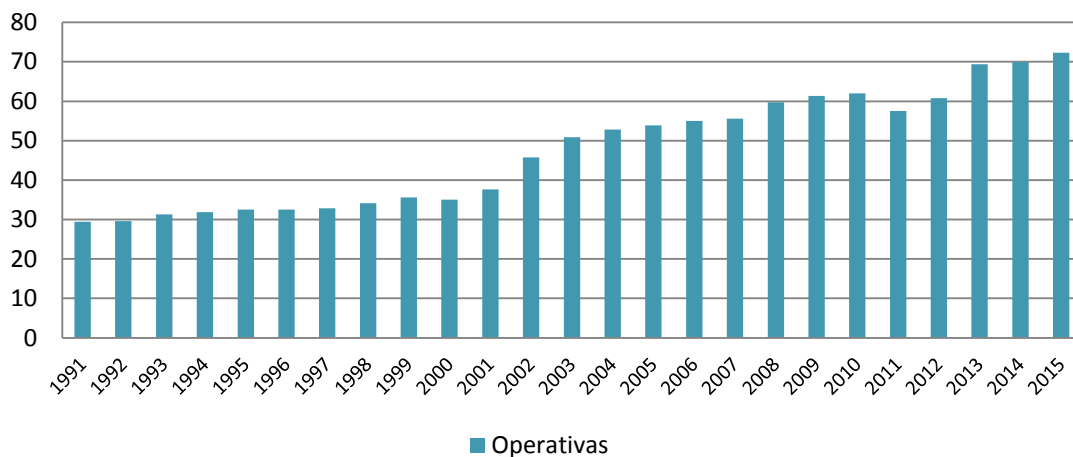
Las entidades fusionadas se caracterizaron por presentar un promedio de oficinas por debajo de la media del sector, con cierta contracción en la última etapa considerada, mientras que las entidades operativas mantuvieron una tendencia creciente en la apertura de oficinas, observándose dos periodos claramente diferenciados, entre 2001 y 2006, 2008 hasta 2009, con cierto estancamiento en la última fase del estudio.

ILUSTRACIÓN 13: EVOLUCIÓN DEL NÚMERO MEDIO DE OFICINAS DE CAJAS RURALES FUSIONADAS



FUENTE: ELABORACIÓN PROPIA

ILUSTRACIÓN 14: EVOLUCIÓN DEL NÚMERO MEDIO DE OFICINAS DE CAJAS RURALES OPERATIVAS

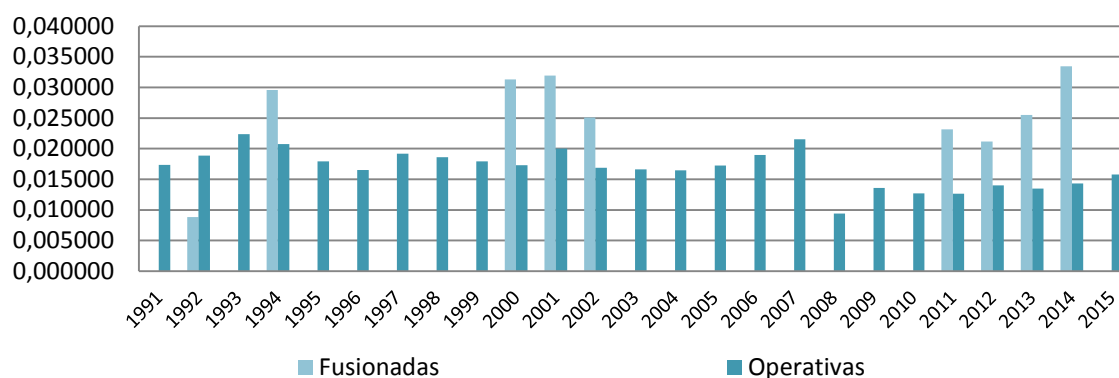


FUENTE: ELABORACIÓN PROPIA

La variable CAPITALIZACIÓN, que pondera el capital social sobre el activo total, es estable en el conjunto de entidades operativas a lo largo del periodo, con una ligera contracción en la última etapa que podría estar motivada por el efecto del incremento del negocio sobre el volumen del activo de estas entidades.

En este sentido se puede apreciar en la Ilustración 15 que las entidades fusionadas mantuvieron en todo momento niveles de capitalización superiores a los presentados por el conjunto de entidades operativas, muestra de que a pesar de incrementar sus niveles de capital para afrontar las pérdidas no les fue posible evitar la fusión con otras Cajas Rurales.

ILUSTRACIÓN 15: EVOLUCIÓN DE LA RATIO DE CAPITALIZACIÓN EN EL PERIODO DE ESTUDIO.



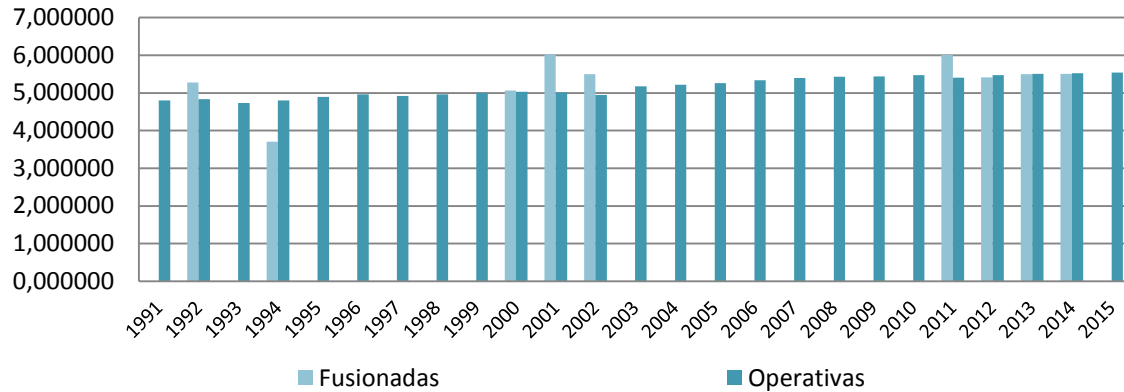
FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la calidad del activo, a continuación se analizará el comportamiento de cuatro ratios financieras: TAMAÑO, NEGOCIO, COSTE DEL RIESGO y ESFUERZO EN PROVISIONES.

La variable TAMAÑO mide el volumen del activo de la Caja Rural (en logaritmos neperianos para facilitar su comparación), mostrando la evolución que se ha recogido en la Ilustración 16, en la que se puede observar que en término medio presentó una tendencia creciente en el conjunto de Cajas Rurales, con cierto estancamiento en la última etapa del estudio.

Las Cajas Rurales fusionadas se caracterizan por su mayor tamaño en términos de activo, con un volumen medio que superó a la media del conjunto de Cajas Rurales a lo largo del periodo de estudio.

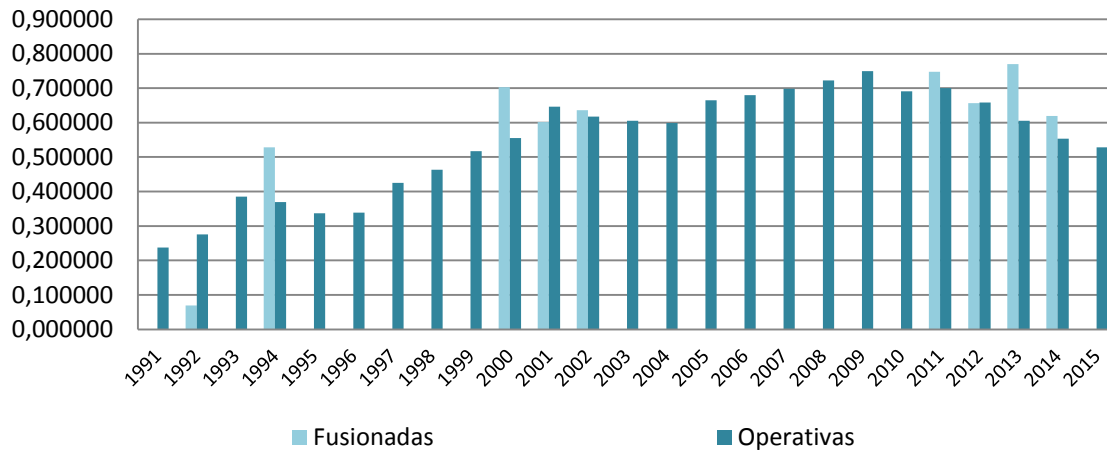
ILUSTRACIÓN 16: EVOLUCIÓN DE LA RATIO DE TAMAÑO A LO LARGO DEL PERIODO DE ESTUDIO.



FUENTE: ELABORACIÓN PROPIA

La variable NEGOCIO refleja la proporción del activo que se concentra en la cartera de crédito a la clientela. En la Ilustración 17 se recoge la evolución de esta variable a lo largo del periodo, observándose una tendencia creciente a partir del año 2000 y una contracción desde el año 2011 en adelante. En término medio el conjunto de Cajas Rurales fusionadas mostró una mayor concentración de su activo en el crédito a la clientela, síntoma de políticas más agresivas de inversión, cifras que en todo momento se mantuvieron por encima de la media sectorial, diferencias que también pone de manifiesto la prueba t ya comentada.

ILUSTRACIÓN 17: EVOLUCIÓN DE LA RATIO NEGOCIO A LO LARGO DEL PERIODO DE ESTUDIO.

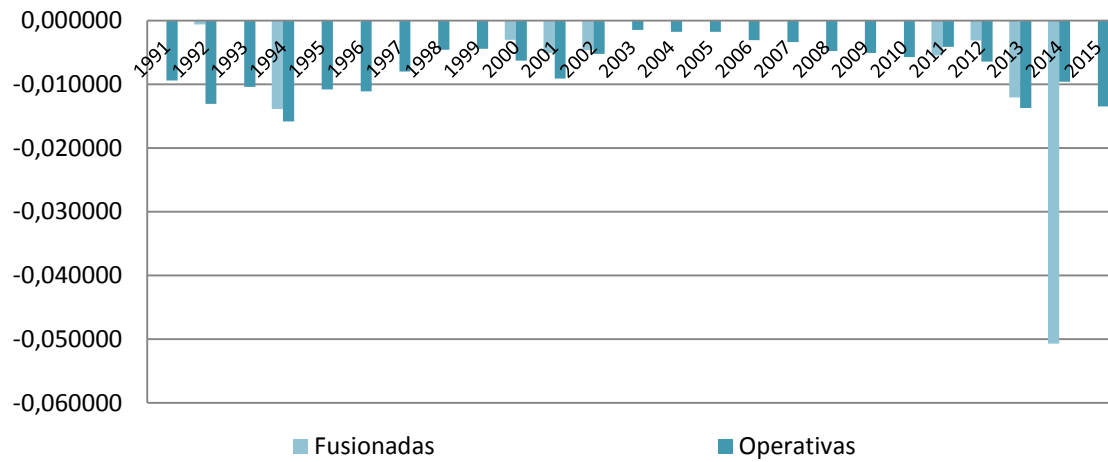


FUENTE: ELABORACIÓN PROPIA

Por su parte la variable **COSTE DEL RIESGO**, que mide las provisiones por insolvencia dotadas en relación a la cartera de inversiones crediticias, muestra una evolución muy similar en el conjunto de entidades operativas y fusionadas, tal y como se recoge en la Ilustración 18 y en la prueba *t* realizada, no observándose diferencias entre ambos grupos.

Se aprecia que, con la excepción de las dotaciones practicadas por una Caja Rural fusionada en el ejercicio de 2014 (muy elevadas), existen tres etapas claramente diferenciadas: un primer periodo que se alarga desde 1991 hasta 1996, en el que se reflejó la necesidad de practicar altos deterioros en la cartera de inversiones crediticias, previsiblemente como consecuencia de los efectos negativos de la crisis económica sobre la calidad crediticia de sus clientes; un segundo periodo comprendidos entre 1997 hasta el año 2005, en el que el volumen de provisiones dotadas (con la excepción de los años 2000, 2001 y 2002) se situaron por debajo de la media; y un tercer periodo iniciado en el año 2006 en el que nuevamente la crisis económica provocó serios deterioros en la calidad de las inversiones que nuevamente obligó a practicar mayores dotaciones, principalmente en los ejercicios de 2013 y 2014 en los que ambos grupos de entidades incrementaron el volumen de dotaciones dotadas.

ILUSTRACIÓN 18: EVOLUCIÓN DE LA RATIO COSTE DEL RIESGO A LO LARGO DEL PERIODO DE ESTUDIO

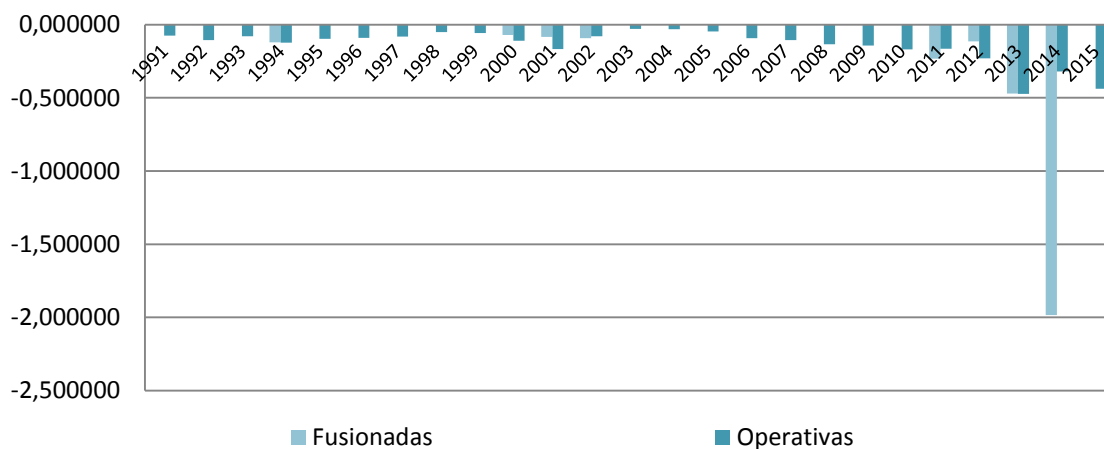


FUENTE: ELABORACIÓN PROPIA

Finalmente, y dentro del grupo de las ratios que miden la calidad del activo, se sitúa la variable ESFUERZO EN PROVISIONES, muy relacionada con la ratio anterior, si bien en esta ocasión se pone en relación las dotaciones por insolvencias de clientes sobre el margen bruto.

En la Ilustración 19 se recoge la evolución de la variable ESFUERZOPROV, observándose diferencias entre ambos grupos de entidades (confirmadas por la prueba *t*), sobre todo en el periodo comprendido entre 1991 y 2013, en el que las dotaciones medias practicadas por el conjunto de entidades fusionadas en relación a su respectivo margen bruto se situó por debajo de la media, con excepción de una entidad fusionada en el ejercicio de 2014 que recogió dotaciones muy superiores a la media del sector.

ILUSTRACIÓN 19: EVOLUCIÓN DE LA RATIO DE ESFUERZO EN PROVISIONES



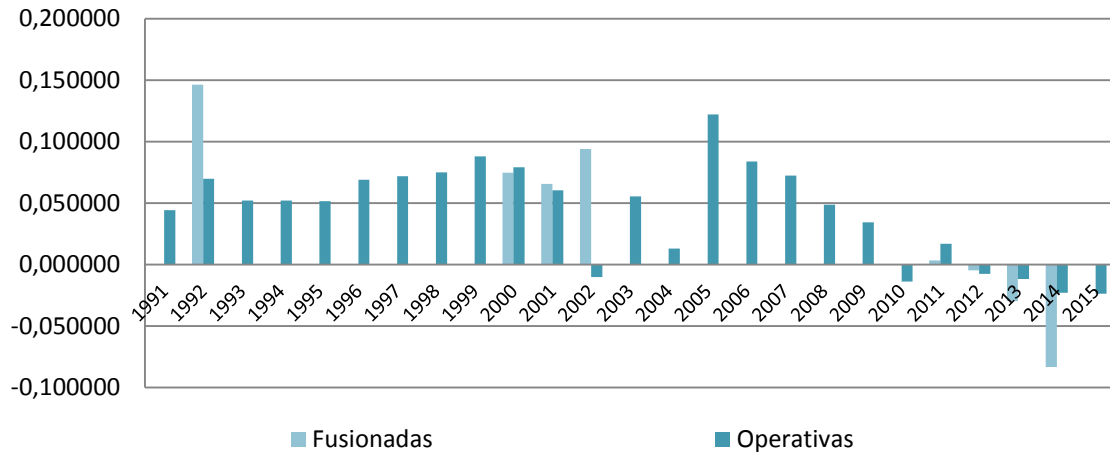
FUENTE: ELABORACIÓN PROPIA

Interesa determinar si existe influencia significativa del equipo gestor en la evolución de las respectivas Cajas Rurales, capacidad que se medirá a través de la ratio EQG1, que mide la evolución del crédito a la clientela de un ejercicio a otro, y la ratio EFICIENCIA, que calcula qué proporción del margen bruto va destinada a cubrir los gastos de explotación de la Entidad.

En la Ilustración 20 se ha recogido la evolución de la ratio EQG1, observándose que en general el conjunto de Cajas Rurales mantuvieron políticas expansivas entre 1991 y 2005, con tasas de crecimiento positivas del crédito a la clientela; una fase de estancamiento entre 2006 y 2011, en los que a pesar de continuar mostrando tasas de crecimiento positivas, estas se redujeron con respecto a los niveles de años anteriores; y una fase recesiva entre 2012 y 2015, en la que se produjo una reducción del crédito a la clientela.

Por tipología de entidad, las Cajas Rurales fusionadas mostraron las mayores tasas de crecimiento de la inversión en el periodo comprendido entre 1992 y 2012, mientras que las Cajas Rurales operativas los mostraron en los años previos a la crisis económica de 2007. No obstante el resultado de la prueba t recogido en la Tabla 14 no encuentra diferencias significativas entre ambos grupos de entidades.

ILUSTRACIÓN 20: EVOLUCIÓN DE LA RATIO EQG1 A LO LARGO DEL PERIODO DE ESTUDIO.

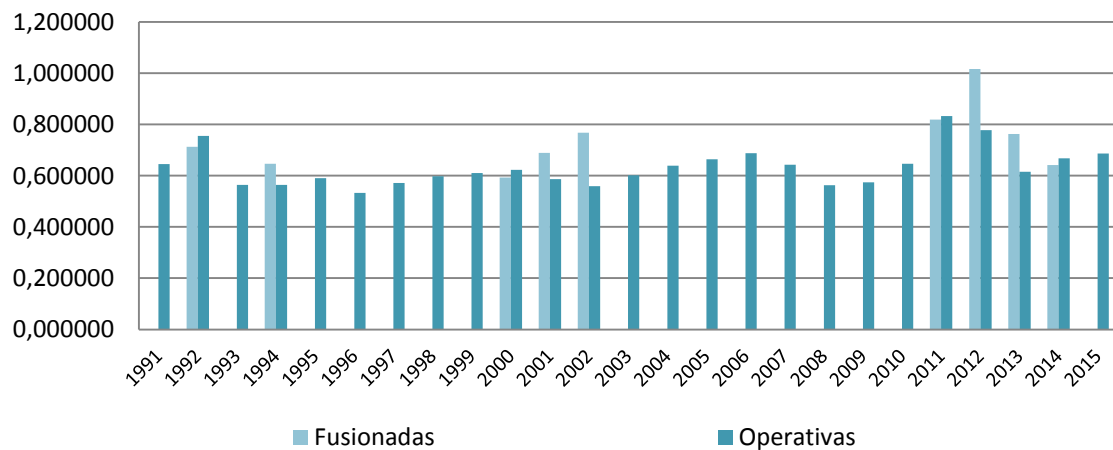


FUENTE: ELABORACIÓN PROPIA

La ratio EFICIENCIA, que relaciona los gastos de explotación en relación al activo total, se muestra diferente en ambos grupos de entidades, tal y como se puede observar en la Ilustración 21, lo que denota la menor eficiencia del conjunto de Cajas Rurales fusionadas en el año previo a la fusión.

Considerando ambos tipos de entidades, se puede observar como la ratio de eficiencia se ha mantenido prácticamente estable a lo largo del periodo de estudio, si bien se observa que en 2011 y 2012 estos gastos se mantuvieron por encima de la media del periodo total.

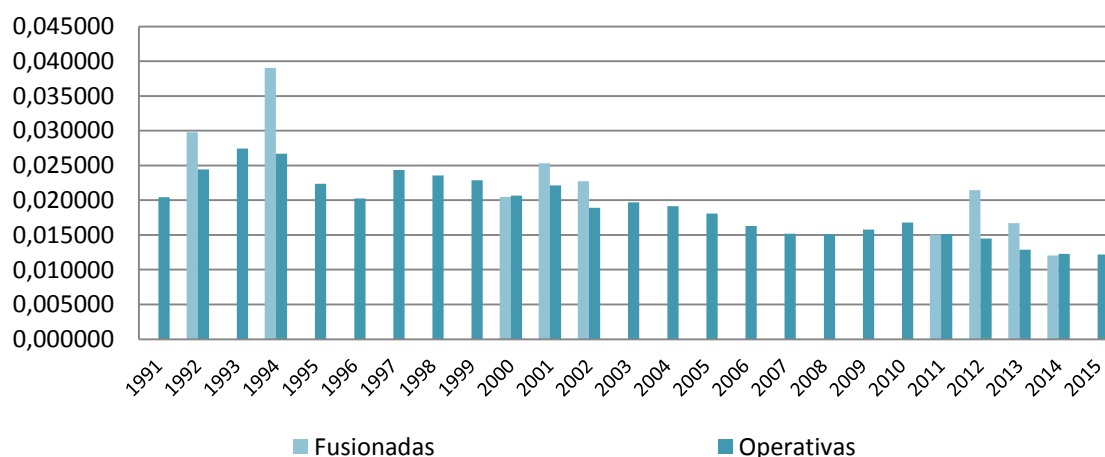
ILUSTRACIÓN 21: EVOLUCIÓN LA RATIO EFICIENCIA A LO LARGO DEL PERIODO DE ESTUDIO



FUENTE: ELABORACIÓN PROPIA

Considerando el conjunto de Cajas Rurales, en la Ilustración 22 se puede observar un ajuste de la estructura de gastos de todas ellas, con cierta mejoría en los últimos años del periodo, con un conjunto de Cajas Rurales fusionadas que aún continuarían mostrando un volumen de gastos de explotación superior a la media sectorial.

ILUSTRACIÓN 22: EVOLUCIÓN DE LA RATIO RENTAB A LO LARGO DEL PERIODO DE ESTUDIO

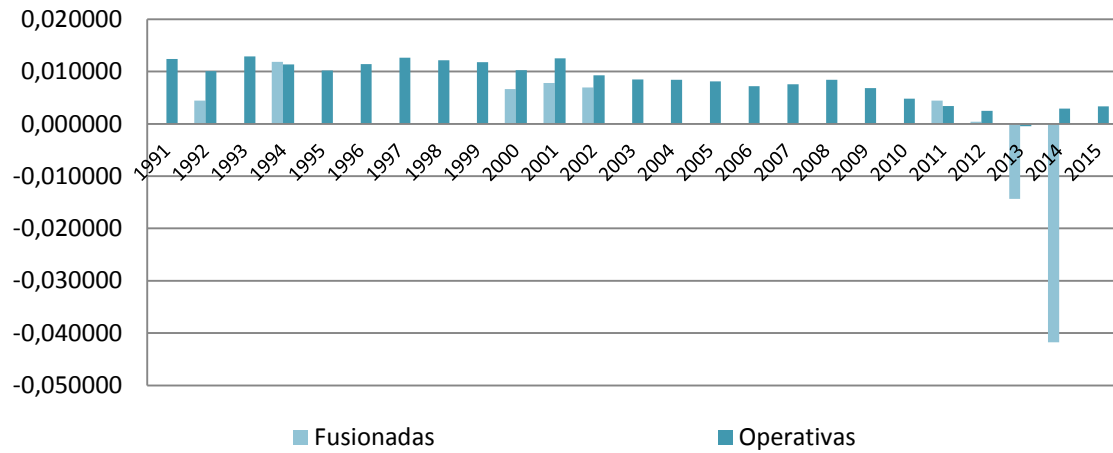


FUENTE: ELABORACIÓN PROPIA

En los que respecta a la rentabilidad del activo, medido en término del resultado sobre el activo total (ROA), en la Ilustración 23 se puede observar que en términos generales el sector de Cajas Rurales acusó una pérdida de rentabilidad a lo largo del periodo de estudio, principalmente a partir del año 2009.

En el caso de las Cajas Rurales fusionadas, esta situación se agrava, ya que en todo momento se situó por debajo de la medida sectorial, siendo especialmente significativa la mostrada en los años 2013 y 2014, en los que las Cajas Rurales fusionadas en ese ejercicio mostraron en término medio pérdidas en el ejercicio precedente.

ILUSTRACIÓN 23: EVOLUCION DE LA RATIO ROA A LO LARGO DEL PERIODO DE ESTUDIO

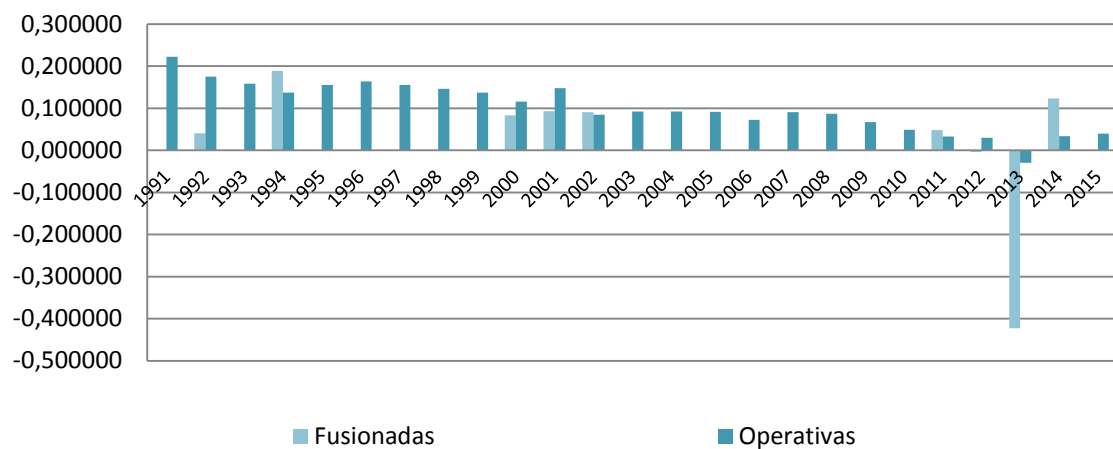


FUENTE: ELABORACIÓN PROPIA

La evolución de la rentabilidad de los fondos propios (ROE) se dibuja en la Ilustración 24 en la que también se aprecian diferencias entre los grupos de entidades operativas y fusionadas, como ya puso de manifiesto la prueba t efectuada en la Tabla 14.

En términos generales, se aprecia un ligero deterioro de la ratio ROE a lo largo del periodo de estudio, siendo especialmente significativo el comportamiento mostrado en los años 2010-2015, en el que los niveles de rentabilidad se mantienen por debajo de los presentados en años anteriores.

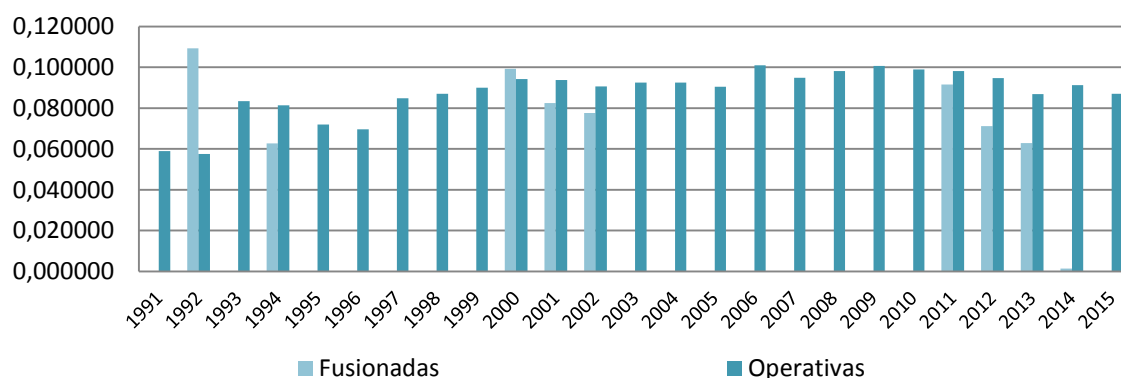
ILUSTRACIÓN 24: EVOLUCIÓN DE LA RATIO ROE A LO LARGO DEL PERIODO DE ESTUDIO



FUENTE: ELABORACIÓN PROPIA

En la Ilustración 25 se recoge el comportamiento de APALANCAMIENTO, que pondera fondos propios sobre activo total, observándose que las Cajas Rurales fusionadas mantuvieron las menores ratios de capitalización, aunque en sintonía con el nivel mínimo del 3% que Basilea III fija para esta ratio. Tan solo recordar que Basilea III además establece el requisito que la ratio TIER 1, que pondera fondos propios sobre activos ponderados por riesgo, debe situarse como mínimo en el 4,5%.

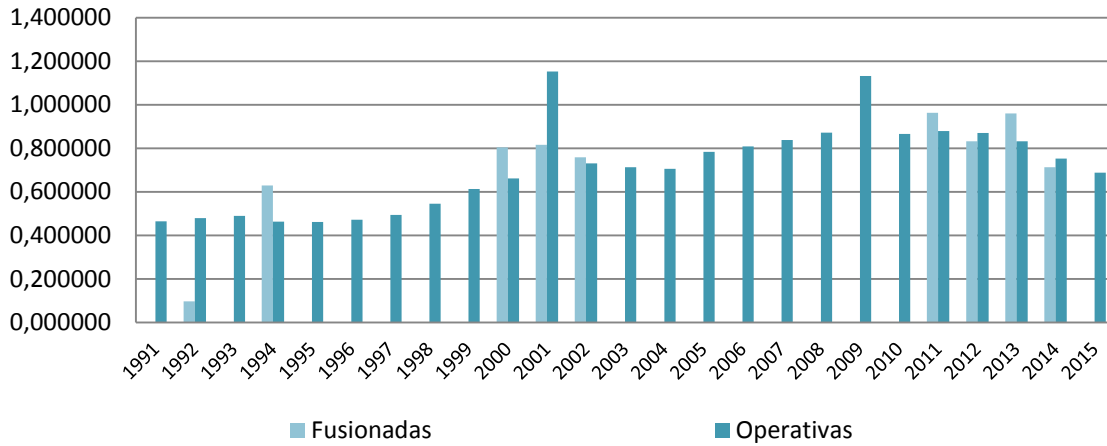
ILUSTRACIÓN 25: EVOLUCIÓN DE LA RATIO APALANCAMIENTO A LO LARGO DEL PERIODO DE ESTUDIO



FUENTE: ELABORACIÓN PROPIA

De la Ilustración 26 se desprende que en término medio LTD se mantuvo por debajo de 100% a nivel sectorial en todo momento, lo que es indicativo de que las Cajas Rurales consiguen financiar la mayor parte de sus inversiones crediticias con los depósitos a los clientes, con la excepción de los ejercicios de 2001 y 2009, en las que el recurso a la financiación interbancaria fue más acentuado.

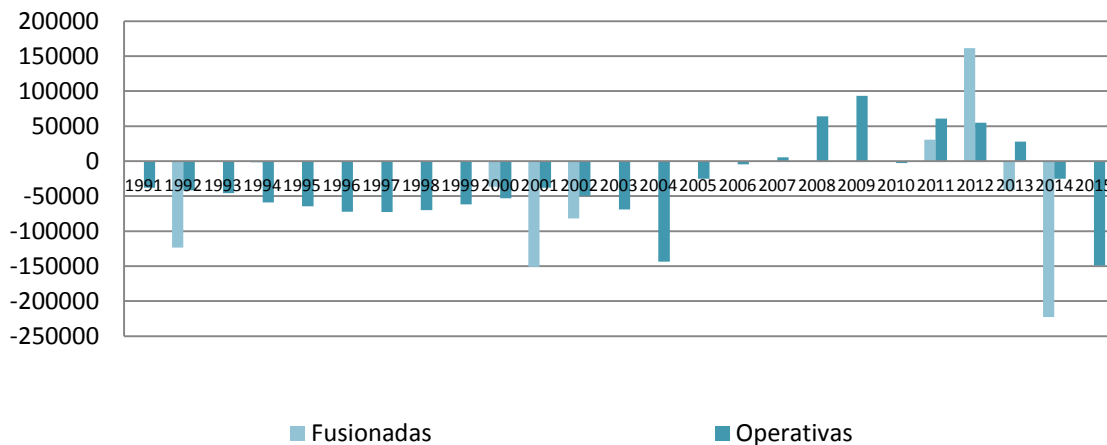
ILUSTRACIÓN 26: EVOLUCIÓN DE LA RATIO LTD A LO LARGO DEL PERIODO DE ESTUDIO



FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la variable FUNDING GAP, que mide la necesidad de financiación en los mercados interbancarios, en la Ilustración 27 se observa que en término medio las Cajas Rurales no precisaron de financiación ajena a los depósitos de sus clientes, ya que la ratio se negativa en todo momento, con la excepción de los últimos años del periodo en el que las necesidades de financiación se acentuaron en ambos grupos de entidades.

ILUSTRACIÓN 27: EVOLUCIÓN DE LA RATIO FUNDING GAP A LO LARGO DEL PERIODO DE ESTUDIO

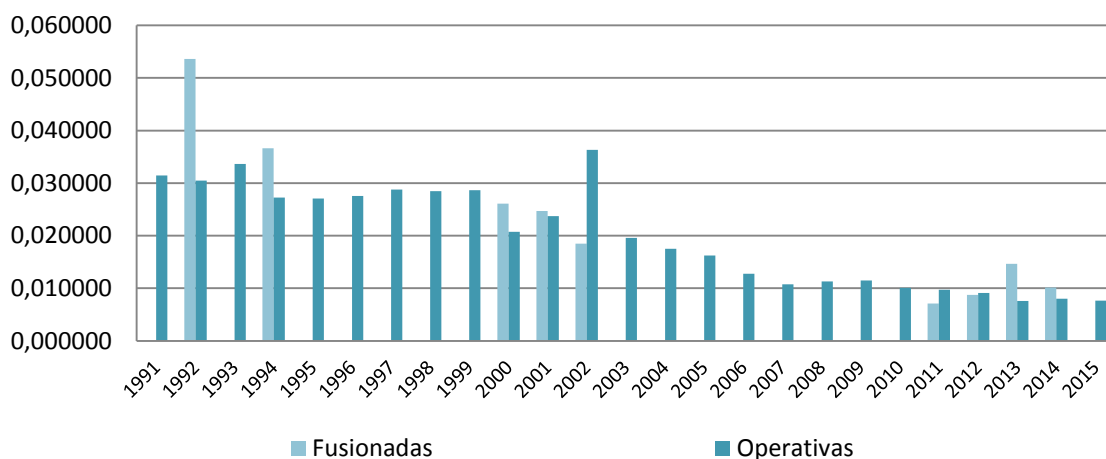


FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la capacidad para atender la salida de depósitos de sus clientes, en la Ilustración 28 se recoge la evolución de la ratio LIQUID, que mide las disponibilidades

líquidas en relación a los depósitos de clientes, observándose su contracción entre ambos grupos de entidades sobre todo a partir del año 2003, siendo muy similar el comportamiento en ambos grupos de entidades (operativas y fusionadas).

ILUSTRACIÓN 28: EVOLUCIÓN DE LA RATIO LIQUID A LO LARGO DEL PERIODO DE ESTUDIO



FUENTE: ELABORACIÓN PROPIA

4.3 SELECCIÓN DE VARIABLES EXPLICATIVAS

Cuando se dispone de un número elevado de variables explicativas para la búsqueda de un patrón que permita explicar el comportamiento de los individuos, cobra especial relevancia la selección de aquellas variables con mayor poder de explicación.

Fernández *et al* (2009) señala que “uno de los principales problemas de la investigación en materia de insolvencia empresarial es la inexistencia de un modelo restringido, universalmente aceptado, al que el investigador empírico pueda recurrir para la selección de las ratios financieras”, de ahí que en muchas publicaciones esa selección se haya realizado por procedimientos que en ocasiones hayan mermado la capacidad predictiva de los modelos, además de ocasionar problemas de multicolinealidad.

Como se explicará a continuación, existen diversas técnicas para la selección de las variables explicativas a utilizar en los diferentes modelos, si bien, la mayor parte de las investigaciones analizadas realizan de forma previa un análisis individualizado de cada una de las variables para conocer la influencia que ejercen sobre la variable dependiente (Anastasi *et al*, 1998; Porath, 2006; Fiordelisi *et al.*, 2013; y Carvalho *et al.*, 2015; entre otros), eliminando del estudio aquellas variables con menor poder predictivo.

En este sentido Trussel *et al.* (2012) ponen de manifiesto que el uso de un número elevado de ratios financieras podría conllevar un problema de parsimonia²², lo que implicaría una merma en el poder predictivo de los mismos.

De hecho, elaboran un modelo en el que con apenas seis ratios (a partir de variables utilizadas por otros autores) consiguen mejorar el poder de predicción de un modelo de supervivencia, al tiempo que trataron de cuantificar el coste de los errores tipo I y II para los inversores.

La prueba t realizada al principio de este epígrafe (Tabla 14) puso de manifiesto que tan solo las variables CAPITALIZACIÓN, APALANCAMIENTO, TAMAÑO, NEGOCIO, ESFUERZOPROV, EFICIENCIA, ROA, ROE, EMP, OF1 y OF2 eran diferentes en ambos grupos por lo que finalmente serán estas variables las seleccionadas.

Una vez determinadas las variables explicativas que pueden resultar útiles para la estimación de los modelos planteados en el capítulo tercero, se centrará la atención en el método a seguir para la selección de las variables explicativas con el fin de evitar los errores manifestados por Fernández *et al* (2009).

En este sentido en la literatura analizada se constata que, a pesar de existir cierta variedad de metodologías para ello, son los métodos tradicionales *stepwise*, *backward* y *forward* los

²² El principio de parsimonia invita a explicar los acontecimientos con el menor número de causas posibles.

más utilizados, previsiblemente por su facilidad de utilización al encontrarse implementados en los distintos paquetes informáticos estadísticos utilizados para la realización de las investigaciones (Pacheco *et al*, 2005).

No obstante, existen autores que proponen métodos alternativos por considerar que proporcionan resultados más eficientes.

En este sentido Mínguez (2005), Karacabey (2007) o Lopez *et al*. (2005) proponen el uso del análisis de los componentes principales para la selección de las variables explicativas, llegando incluso a proponer la estimación de los modelos a partir de los componentes principales calculados por contener información de todas las variables iniciales.

Otros autores proponen métodos alternativos basados en técnicas metaheurísticas más novedosas, como pueden ser los algoritmos genéticos (Acosta *et al*; 2007), la búsqueda tabú (Pacheco *et al*; 2005), algoritmos meméticos (Pacheco *et al*; 2007), entre otros, si bien estas nuevas metodologías se escapan del objetivo inicial de esta Tesis, abriendo futuras líneas de investigación al respecto.

En los siguientes subepígrafes se realizará una breve descripción de los procedimientos que finalmente serán utilizados para la selección de las variables explicativas que serán finalmente utilizadas estimación de los modelos propuestos.

4.3.1 Métodos tradicionales

Los métodos tradicionales se consideran procedimientos sencillos de selección de variables basados en criterios estadísticos como la Landa de Wilks, F de Fisher, etc (Pacheco *et al*, 2007). Se distinguen principalmente:

- a. El método *forward* o de selección hacia delante, que comienza eligiendo las variables con mayor poder de discriminación para continuar añadiendo variables y aceptándolas o rechazándolas en función del poder de discriminación conjunto observado.
- b. El método *backward* o de supresión hacia atrás. Este método es el proceso contrario del anterior. Se comienza considerando todas las variables explicativas en la estimación del modelo, para posteriormente eliminar una a una en función del poder de discriminación conjunto observado.
- c. El método *stepwise* o regresión por pasos, basado en los dos algoritmos anteriores. Este método permite que el investigador introduzca de forma manual las combinaciones de variables explicativas, llevando a cabo la selección a partir del poder de discriminación observado.

En el caso de las Cajas Rurales objeto de esta Tesis, se emplearán los tres métodos anteriores, siendo seleccionado finalmente aquella combinación de variables que ofrezca los resultados más eficientes.

4.3.2 Análisis factorial

Dentro del uso del análisis factorial para la selección de variables se pueden distinguir dos corrientes, las investigaciones que emplean esta técnica para seleccionar las variables explicativas a partir de las correlaciones que presentan estas variables con los componentes principales calculados, siendo finalmente seleccionadas aquellas variables que muestran los mayores coeficientes de correlación (Lopez *et al*, 2005) dentro de cada componente; o bien el empleo de esos componentes principales para la estimación de los modelos en sustitución de las variables explicativas iniciales (Karacabey, 2007).

En el caso de las Cajas Rurales, se empleará el método de los componentes principales tanto para la selección de las variables explicativas como para la estimación de los modelos a partir de los componentes principales calculados.

La técnica del análisis factorial permite reducir la dimensión de la información al máximo posible, eliminando aquella información que no sea significativa mediante la detección de la estructura subyacente de la serie de datos (Rúa *et al*; 2003).

Aunque esta técnica ha sido suficientemente explicada en los manuales de Comrey (1985), Cuadras (1996) y García *et al.* (2000), señalar que su mecanismo de funcionamiento se basa en reducir el conjunto de variables explicativas a un mínimo de factores que maximizan la varianza explicada de la muestra.

De forma previa a la aplicación del Análisis Factorial, se hace necesario comprobar la matriz de correlaciones entre las distintas variables con el fin de determinar si es adecuado llevar a cabo el proceso de factorización, ya que la presencia de niveles de correlaciones muy reducidos estaría confirmando la baja idoneidad de este procedimiento.

En este sentido en la Tabla 15 se calculan las correlaciones lineales para todas las variables explicativas (excluyendo aquellas que no resultaron ser significativas en los resultados de la prueba *t* recogidos en la Tabla 14), observando que en términos generales existe un grado de correlación reducido pero en niveles aceptables entre las variables que han resultado ser significativas según la prueba *t*, de forma que siguiendo a García *et al.* (2000) se podría considerar que las diferentes variables no son independientes entre sí y que, por tanto, es posible utilizar la técnica del análisis factorial para encontrar esos factores que maximizan la varianza explicada.

Existen dos pruebas adicionales que permiten validar los resultados, el test de esfericidad de Barlett (BTS) que estima la probabilidad de que las correlaciones observadas en la muestra procedan de la población de la que han sido extraídas; y el test de adecuación de muestreo de Kaiser-Meyer-Olkin (KMO).

Poveda (2006) afirma que *“para la realización del Test de Esfericidad de Bartlett se somete a contraste la hipótesis de que las variables no correlacionarían en la población de la que procede la muestra y de que las correlaciones observadas en la matriz de correlaciones difieren de cero sólo por azar”*. De hecho, *“si el valor del estadístico es alto y el nivel de significación asociado es pequeño, parece improbable que la matriz de correlación poblacional sea una identidad”*.

En la Tabla 16 se recogen los resultados del test BTO. Se puede observar que el estadístico asociado se sitúa en el valor de 630,782, con un nivel de significación de 0,0001 para 105 grados de libertad, lo que permite rechazar la hipótesis nula de ausencia de correlación entre las variables explicativas de la población de la que ha sido extraída la muestra.

TABLA 16: TEST DE KAISER-MEYER-OLKIN Y TEST DE LA ESFERICIDAD DE BARLETT

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,562
Prueba de esfericidad de	Aprox. Chi-cuadrado	630,782
Bartlett	gl	105
	Sig.	,000

FUENTE: ELABORACIÓN PROPIA

En lo que respecta al test de muestreo de Kaiser-Meyer-Olkin, Poveda (2006) indica que se trata de *“un índice que compara las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parciales”*, hasta el punto que *“valores pequeños de KMO cuestionan el empleo del modelo factorial dado que, en ese caso, las correlaciones entre los pares de variables no pueden explicarse por otras variables”*.

En la Tabla 16 se aprecia que el valor del KMO es de 0,562, dentro del límite aceptable por García *et al.* (2000) para aplicar el método de los componentes principales.

Una vez determinada la idoneidad del proceso de factorización, el siguiente paso es llevar a cabo la extracción de los factores a partir del procedimiento de los componentes principales. Los resultados han sido recogidos en la Tabla 17, observándose la existencia de cuatro factores que explicarían un 79,4% de la varianza total.

TABLA 17: TOTAL DE LA VARIANZA EXPLICADA POR LOS DISTINTOS FACTORES

COMPONENTE	AUTOVALORES INICIALES			SUMAS DE EXTRACCIÓN DE CARGAS AL CUADRADO			SUMAS DE ROTACIÓN DE CARGAS AL CUADRADO		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulad
1	4,766	31,772	31,772	4,766	31,772	31,772	3,545	23,634	23,634
2	3,590	23,931	55,704	3,590	23,931	55,704	3,487	23,250	46,884
3	2,200	14,665	70,369	2,200	14,665	70,369	3,277	21,844	68,728
4	1,365	9,102	79,471	1,365	9,102	79,471	1,612	10,744	79,471
5	,890	5,936	85,407						

FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la composición de cada uno de los factores, en la Tabla 18 se puede comprobar la existencia de un pequeño grupo que muestra las mayores correlaciones con los factores en los que se integran. En este sentido Gorsuch (1983) afirma que un factor se encuentra definido por todas aquellas variables que presenten pesos superiores a 0,35 (aunque cuanto mayor sea este valor, mayor será su importancia).

TABLA 18: MATRIZ DE COMPONENTES

	COMPONENTE			
	1	2	3	4
ROA	,850		,400	
EsfuerzoProv	,772		,533	
PIB	,757		-,434	
OF2	,733		,567	
INT1	,715	-,532		
Paro	-,692			
Tamaño		,887	-,385	
EMP		,820	-,427	
OF1	,436	,790	-,385	
Negocio	-,439	,617		-,373
INT2	,605	-,607		
Eficiencia			,527	,406
Capital				,599
Apalancamiento	,526		,427	-,574
ROE	,425			-,448

FUENTE: ELABORACIÓN PROPIA

No obstante García *et al.* (2000), refiriéndose a otros autores, indica la necesidad de rotar los factores con el fin de alcanzar una explicación factorial lo más sencilla posible, existiendo diferentes mecanismos para llevar a cabo estas transformaciones.

En la Tabla 19 se recoge la matriz de factores rotados para cada uno de los factores, ordenados por importancia y siendo eliminados aquellos ítems con pesos inferiores al 0,35. Lopez *et al.* (2005) propone seleccionar las variables explicativas a utilizar en el modelo a partir de las mayores correlaciones observadas, siendo finalmente seleccionadas las variables OF2, INT1, EMP y CAPITAL.

TABLA 19: MATRIZ DE COMPONENTES ROTADOS

	COMPONENTE			
	1	2	3	4
INT1	,945			
INT2	,940			
PIB	,867			
Negocio	-,750			
Paro	-,492			,385
OF2		,955		
EsfuerzoProv		,939		
ROA		,926		
Apalancamiento		,675		-,541
EMP			,974	
OF1			,962	
Tamaño			,953	
Capital				,619
Eficiencia				,555
ROE				-,544

FUENTE: ELABORACIÓN PROPIA

Por su parte Karacabey (2007) propone el uso exclusivo de los factores para la estimación del modelo. En este sentido, los cuatro factores hallados a partir del análisis de los componentes principales podrían ser definidos de la siguiente forma:

- a. Primer factor, denominado inversión por estar determinado por las variables INT1 (tipos de interés a corto plazo), INT2 (tipos de interés del mercado interbancario), PIB (tasa de crecimiento interanual del PIB), NEGOCIO (volumen del crédito a la clientela sobre activo total) y PARO.

Se observa que las variables macroeconómicas ejercen una influencia positiva sobre la inversión, corroborando que en las fases expansivas del ciclo económico se produce un incremento de la inversión de las Cajas Rurales. Sin embargo, el paro influye de forma negativa, poniendo de manifiesto que el incremento del desempleo está asociado a una

ralentización de la inversión, tal y como cabría esperar dada la alta presencia que las Cajas Rurales tienen sobre el crédito a las familias, lo que acentúa el riesgo de crédito.

Además de ello llama la atención el comportamiento mostrado por la variable NEGOCIO (que mide la proporción del crédito a la clientela sobre el activo total), ya que el signo negativo presentado podría ser indicativo de que las Cajas Rurales mantienen una política conservadora que les lleva a reducir la inversión en el momento en que se acentúa la concentración del activo en el crédito a la clientela.

- b. Segundo factor, al que se podría denominar rentabilidad por estar definido principalmente por las variables OF2 (beneficio neto por oficina), ESFUERZOPROV (provisiones dotadas sobre margen bruto), ROA (beneficio neto sobre activo total) y APALANCAMIENTO (fondos propios sobre activo total).

Se podría afirmar que cuanto mayor sean los beneficios y los fondos propios en función del activo, y el número de oficinas, mayor será la rentabilidad de la Caja Rural, mientras que la variable ESFUERZOPROV, por estar definida en signo negativo, indica que cuanto más pequeña sea la proporción de las provisiones en relación al margen bruto, mayor será la rentabilidad de la Caja Rural.

- c. Tercer factor, Dada la composición del mismo, se podría denominar a este factor dimensión, ya que considera tres variables que en definitiva están midiendo el tamaño de la Caja Rural, concretamente OF1 (número de oficinas), EMP (número de empleados) y TAMAÑO (volumen de activo). Este factor estaría indicando que cuanto mayor sea el número de empleados, oficinas o activos totales, mayor será la dimensión de la Caja Rural
- d. Cuarto factor, que se podría denominar necesidad de capital por estar determinado principalmente por las variables CAPITAL, EFICIENCIA (gastos de explotación sobre

ingresos netos) y ROE (beneficios sobre fondos propios), siendo destacable el hecho de que cuanto mayor sea el ROE, menor será la necesidad de capitalización de la Caja Rural. Por el contrario, cuanto mayor sean los gastos de explotación sobre los ingresos, mayor será la necesidad de capital.

4.4 EXPOSICIÓN DE LOS RESULTADOS OBTENIDOS DE LA ESTIMACIÓN DE LOS MODELOS LOGIT Y REGRESIÓN DE COX.

A continuación, se exponen los resultados obtenidos de la estimación de los modelos de respuesta binaria (logit y logit multiperiodo) y análisis de supervivencia (regresión de Cox).

A modo de recapitulación cada uno de los modelos ha sido estimado en tres ocasiones considerando los métodos de selección de variables explicativas expuestas en el epígrafe anterior:

- Selección de variables explicativas a partir de los procedimientos tradicionales (por pasos).
- Selección de variables a partir del método de componentes principales (seleccionando las que presentan mayor correlación con los factores calculados).
- Estimación de los modelos utilizando los cuatro factores calculados (en sustitución de las variables explicativas) combinados con las variables categóricas.

Así mismo, para la estimación de los modelos logit estático, logit multiperiodo y regresión de Cox se ha utilizado una muestra compuesta por 2/3 partes de la totalidad de entidades operativas y fusionadas, reservándose 1/3 parte para la validación final de cada uno de los modelos.

4.4.1 Estimación de la Regresión Logística Binaria (Logit estático)

De la bibliografía analizada no se desprende una uniformidad de criterio en la elaboración de la muestra a utilizar para la estimación del modelo logit, ya que unos autores consideran que es necesario emparejar las observaciones de forma que se disponga de una muestra uniforme de individuos solventes e insolventes para el establecimiento de puntos de corte óptimos en los modelos (Rodríguez *et al.*; 2014), mientras que otras investigaciones consideran la totalidad de individuos existentes en periodos de estudio más reducidos.

En lo que respecta al primer procedimiento, es habitual que el emparejamiento se realice de forma aleatoria, aunque respetando la dimensión temporal de las observaciones, si bien existen autores que introducen un componente arbitrario realizando el emparejamiento en función del tamaño o la ubicación de la Entidad.

En este sentido Laguillo (2016), en referencia a Zmijewski (1984), afirma que la utilización de muestras emparejadas por un criterio determinado no afecta a la capacidad predictiva del modelo, procedimiento que habría sido utilizado por Kumar *et al.* (1995), Borges *et al.* (1996), Magalhaes (2001), Li *et al.* (2011), Trussel *et al.* (2012) y Bologna (2011), entre otros.

Lane (1999) critica ese sistema de emparejamiento por distorsionar la calidad de la muestra, ya que elimina observaciones que podrían ser significativas en los modelos. Por ello propone usar de la totalidad de los individuos, pero reduciendo el alcance temporal de los periodos de estudio, línea seguida por Anastasi *et al.* (1998), Cole *et al.* (1998), Serra *et al.* (2002), Huang *et al.* (2010) y Apergis *et al.* (2013), entre otros.

Al no existir consenso entre las diferentes investigaciones, en esta Tesis se emplearán ambos para su posterior contrastación.

Se aplicará el primer procedimiento para una muestra homogénea que incluirá todas las Cajas Rurales fusionadas a lo largo del periodo de estudio, un total de 30 entidades que se emparejaran de forma aleatoria con otras Cajas Rurales que se encuentren dentro del mejor cuartil de la ratio de apalancamiento, de forma similar al procedimiento empleado para la definición de la variable dependiente. Del total de Cajas Rurales consideradas, se reservará un tercio para la posterior verificación del modelo.

En el caso del segundo procedimiento, se considerarán los periodos establecidos en el Capítulo Tercero, adaptados a la temporalidad de la muestra y a los años en los que se produjeron las fusiones de las Cajas Rurales:

- a. Desde 1992 hasta 1994. La muestra incorpora dos casos de bajas de Cajas Rurales sobre los ocho casos reales que se registran en Banco de España, apenas representativa del 25% del total de la población.
- b. Desde 2000 hasta 2002. La muestra incorpora un total de nueve casos de bajas sobre los 12 registrados en el Registro de Banco de España.
- c. Desde 2003 hasta 2007. Dadas las limitaciones de información comentadas, la muestra no incorpora ninguna de las dos observaciones que forman parte de la población.
- d. Desde 2011 hasta 2014. La muestra incorpora un total de diecinueve casos, uno más de los registrados en Banco de España por considerar una Caja Rural que se habría fusionado en 2.012 pero que, al ser la sociedad absorbente, no se habría registrado una variación en el código de entidad asignado.

Los resultados obtenidos para cada uno de los periodos han sido las siguientes:

4.4.1.1 Periodo 1: desde 1992 hasta 1994

Dada la poca representatividad de Cajas Rurales que habrían causado baja dentro del periodo comprendido entre 1992 y 1994, con solo dos fusiones, finalmente se decide excluir este periodo del estudio.

4.4.1.2 Periodo 2: desde 2000 hasta 2002

La muestra incorpora un total de 78 observaciones, 9 correspondientes a Cajas Rurales fusionadas y 69 Cajas Rurales operativas a diciembre de 2002. Para la posterior validación se utilizarán las observaciones correspondientes al periodo 2011-2014. La estimación del modelo logit se realiza con el software *IBM SPSS Statistics* en su versión 23.

En primer lugar es necesario codificar la variable dependiente (fusionada o no), en función de los valores cero y uno recogidos en la Tabla 20, codificación que posteriormente será utilizada para la estimación del resto de modelos.

TABLA 20: CODIFICACIÓN DE LA VARIABLE DEPENDIENTE

ESTADO	CODIFICACIÓN
Fusionada	1
Operativa	0

FUENTE: ELABORACIÓN PROPIA

Además de las variables cuantitativas seleccionadas en el epígrafe anterior, el modelo incluye una serie de variables categóricas que deben ser codificadas de forma similar a lo ya realizado con la variable dependiente, como se recoge en la Tabla 21.

TABLA 21: CODIFICACIÓN DE LAS VARIABLES INDEPENDIENTES CATEGÓRICAS

		FRECUENCIA	CODIFICACIÓN DE PARÁMETRO		
			(1)	(2)	(3)
Tipo	Provincial	41	1,000	,000	,000
	Local	1	,000	1,000	,000
	Comarcal	34	,000	,000	1,000
	Resto	2	,000	,000	,000
UB	Norte	16	1,000	,000	
	Centro-Sur	23	,000	1,000	
	Levante-Cataluña	39	,000	,000	
FGD	FGDno	72	1,000		
	FGDsi	6	,000		

FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la selección de las variables explicativas, en primer lugar se ha realizado por los métodos tradicionales, descartando aquellas variables que no resultaron ser significativas en la prueba t recogida en la Tabla 14

Posteriormente se ha vuelto a estimar el modelo seleccionando las variables explicativas mediante el método de los componentes principales, finalizando con una última estimación incluyendo los factores en sustitución de las variables explicativas iniciales.

La fortaleza inicial del modelo, sin incluir variables independientes, es decir, considerando el término constante, se mide a través del concepto de lejanía (denotada con $-2LL$), calculada a partir de la función de verosimilitud. Aunque en Silva *et al.* (2004) se puede encontrar una definición formal de este estadístico, simplemente indicar que cuanto más cercano sea a cero, mejor será el ajuste del modelo.

En la Tabla 22 se puede observar que el estadístico $-2LL$ para un modelo que solo considera el término constante toma el valor 55,790, alejado de cero, de ahí que se haga la reestimación del modelo incluyendo las variables explicativas consideradas.

TABLA 22: HISTORIAL DE ITERACIONES

ITERACIÓN		-2LL
Paso 0	1	58,035
	2	55,838
	3	55,790
	4	55,790
	5	55,790

FUENTE: ELABORACIÓN PROPIA

En la Tabla 23 se recogen las variables que han resultado ser significativas en la estimación del modelo mediante la selección por métodos tradicionales, observándose que TAMAÑO, EFICIENCIA, ROA, EMP, OF1 y OF2, son las que presenta un grado de significatividad individual inferior al nivel del 0,05.

TABLA 23: PARÁMETROS ESTIMADOS EN LA CONSTRUCCIÓN DEL MODELO

	B	ERROR ESTÁNDAR	WALD	GL	SIG.
Tamaño	63,521	24,790	6,566	1	,010
Eficiencia	23,658	11,641	4,130	1	,042
ROA	2299,123	992,478	5,366	1	,021
OF2	-,262	,112	5,432	1	,020
EMP	-8,968	3,831	5,481	1	,019
OF1	-17,578	7,778	5,107	1	,024
Constante	-259,851	101,708	6,527	1	,011

FUENTE: ELABORACIÓN PROPIA

Dentro de la Tabla 23 la columna “B” recoge el valor estimado para el parámetro asociado a la variable explicativa; la columna “error estándar” recoge el valor calculado para el error de la media poblacional calculado a partir de las medias muestrales; la columna “Wald” recoge el valor del estadístico utilizado para comprobar la significación o no de la variable explicativa dentro del modelo estimado; la columna “gl” recoge los grados de libertad asociados al estadístico de Wald; y finalmente la columna “Sig,” recoge la significación del estadístico.

Formalmente, se podría especificar el modelo logit a partir de los coeficientes estimados para cada una de las variables que han resultado ser significativas en la Tabla 23, siendo su definición la siguiente:

$$P(\text{Estado} = \text{Fusionada}) = \frac{1}{1 + e^{-(-259,8 + 63,5\text{TAMAÑO} + 23,6\text{EFICIENCIA} + 2299,1\text{ROA} - 17,6\text{OF1} - 0,3\text{OF2} - 8,9\text{EMP})}}$$

La mayor influencia sobre la probabilidad de fusión de una Caja Rural en el periodo 2000-2002 la ejercía la variable ROA, indica la existencia de una relación positiva entre la rentabilidad del activo y la probabilidad de fusión.

Situación similar es la mostrada por las ratios que miden su dimensión (TAMAÑO) y los gastos de explotación sobre activo (EFICIENCIA), con una relación que también influyó positivamente sobre la probabilidad de que una Caja Rural formase parte de un proceso de fusión.

Concretamente se puede observar que cuanto mayor sea el TAMAÑO, mayor será la probabilidad de fusión, ilustrativo de la existencia de un grupo de Cajas Rurales de mayor tamaño que buscaba una mayor integración entre ellas, y de entidades menores que conservaron su identidad propia.

Por su parte, las entidades menos eficientes, es decir, las que soportaban mayores gastos de explotación en relación a su activo, mostraban una mayor probabilidad de fusionarse.

Por el contrario, las variables relacionadas con la dimensión desde la óptica del volumen de oficinas o empleados influyeron negativamente, observándose que a mayor valor, menor probabilidad de fusión.

Igualmente, la variable que medía la rentabilidad por oficina mostraba una influencia negativa, determinando que las Cajas Rurales con oficinas menos rentables mostraron una

mayor probabilidad de iniciar un proceso de fusión en el periodo comprendido entre 2000 y 2002.

Una vez determinadas las variables que resultan significativas para la estimación del modelo se hace necesario evaluar la bondad del ajuste del mismo.

En este sentido, la bondad del ajuste de un modelo logit se puede medir no solo a partir del estadístico que mide la lejanía (-2LL), sino que también resulta útil el empleo de los estadísticos R-cuadrado de Cox y Snell²³ y R-cuadrado de Nagelkerke²⁴. En este sentido en la Tabla 24 se recogen los valores de ambos estadísticos, 0,31 y 0,61 respectivamente, en niveles que pueden considerarse aceptables. Además de ello, la lejanía del modelo se situaría en 26,25, inferior al valor de 55,79 observado para el modelo que no incluye ninguna variable explicativa.

TABLA 24: PRUEBA DE LA BONDAD DEL AJUSTE DEL MODELO

PASO	LOGARITMO DE LA VEROSIMILITUD -2	R CUADRADO DE COX Y SNELL	R CUADRADO DE NAGELKERKE
1	26,258 ^a	,315	,617

FUENTE: ELABORACIÓN PROPIA

Además de los procedimientos anteriores, también se puede evaluar la bondad del ajuste a partir de la Prueba de Hosmer y Lemeshow que consiste en calcular las probabilidades de cada observación para posteriormente compararlas con las probabilidades esperadas, previa agrupación en cuartiles o deciles. Cuanto mayor sea la aproximación, mejor será el ajuste del modelo.

²³ La R-Cuadrado de Cox y Snell básicamente calcula la proporción de la varianza de la variable dependiente que es explicada por las variables explicativas. Sus valores se encuentran entre cero y uno. (Fernández *et al.*; 2015)

²⁴ La R Cuadrado de Nagelkerke simplemente corrige la escala de la R-Cuadrado de Cox y Snell para que se encuentre entre cero y uno, porque en un ajuste perfecto el modelo no llegaría a tomar el valor de uno (Fernández *et al.*; 2015).

Los resultados de esta prueba están recogidos en la Tabla 25, en la que se puede observar la cercanía entre los valores observados y esperados, tanto en el Grupo de las Cajas Rurales operativas como en las Cajas Rurales fusionadas.

TABLA 25: TABLA DE CONTINGENCIA PARA LA PRUEBA DE HOSMES Y LEMESHOW

		VBLE. DEPENDIENTE = OPERATIVA		VBLE DEPENDIENTE = FUSIONADA		TOTAL
		Observado	Esperado	Observado	Esperado	
Paso 1	1	8	8,000	0	,000	8
	2	8	8,000	0	,000	8
	3	8	8,000	0	,000	8
	4	8	8,000	0	,000	8
	5	8	7,995	0	,005	8
	6	8	7,931	0	,069	8
	7	8	7,622	0	,378	8
	8	5	6,803	3	1,197	8
	9	6	5,029	2	2,971	8
	10	2	1,620	4	4,380	6

FUENTE: ELABORACIÓN PROPIA

En la Tabla 26 se recoge el estadístico chi-cuadrado asociado a la prueba de Hosmer y Lemeshow sobre el que se busca que no muestre significación. Concretamente se puede observar que el nivel de significación del estadístico es de 0,830. Al ser superior al nivel de confianza del 0,05 se confirma la ausencia de significación y se validando los resultados de las pruebas anteriores.

TABLA 26: PRUEBA DE LA BONDAD DEL AJUSTE. PRUEBA DE HOSMER Y LEMESHOW.

PASO	CHI-CUADRADO	GL	SIG.
1	4,293	8	,830

FUENTE: ELABORACIÓN PROPIA

Una vez comprobadas las condiciones de validación del modelo, en la Tabla 27 se recogen los resultados de su aplicación a partir de la muestra utilizada para su estimación,

observándose que en término medio el modelo consigue clasificar correctamente a un 91% de los individuos.

No obstante, esta capacidad de clasificación difiere entre el grupo de las Cajas Rurales operativas y fusionadas como se observa en la Tabla 27, ya que el modelo tan solo clasifica correctamente al 44% de las Cajas Rurales fusionadas, frente al 97% de las operativas. Sin embargo, en la muestra de validación la situación difiere notablemente, con un porcentaje de clasificación correcto del 84% en el grupo de las fusionadas, frente al apenas 31% de las entidades operativas.

TABLA 27: CLASIFICACIÓN DEL MODELO LOGIT PARA EL PERIODO 2000-2002

	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	97%	44%	91%	31%	84%	45%

FUENTE: ELABORACIÓN PROPIA

Se plantea ahora la reestimación del modelo seleccionando las variables explicativas por el método de componentes principales, buscando aquellas variables que presenten las mayores correlaciones con los factores, resultados que han sido recogidos en Anexo 1.

En primer lugar se han calculado los factores para la muestra del periodo comprendido entre 2000 y 2002. Tras verificar la existencia de coeficientes de correlación aceptables, así como el cumplimiento de las pruebas KMO y de esfericidad de Barlett, se observa que el conjunto de variables explicativas utilizadas en el periodo podría ser reducido a tres factores que explican conjuntamente el 72,7% de la varianza.

De todas ellas, las variables que muestran los mayores niveles de correlación con los factores calculados son OF1, ROA y ESFUERZOPROV, siendo utilizadas para la estimación del modelo.

Se observa que el modelo estimado tan solo determina la existencia de significatividad en niveles aceptables en la variable OF1, si bien la bondad del ajuste asociado se deterioraría

con respecto al modelo estimado siguiendo los criterios tradicionales de selección. En este sentido el estadístico R cuadrado de Nagelkerke empeoraría hasta el nivel del 0,17 (frente al 0,61 del modelo anterior). De hecho, el Error Tipo II (incorrecta clasificación de una Caja Rural fusionada) se elevaría hasta el 100% con este método de selección.

Finalmente se ha procedido a reestimar el modelo a partir de los tres factores calculados, excluyendo las variables explicativas, aunque combinados con las variables categóricas. Del modelo estimado se observa la existencia de significatividad tan solo en el factor denominado dimensión por presentar las mayores correlaciones con las variables TAMAÑO, EMP y OF1. No obstante, la bondad del ajuste empeoraría, con un R cuadrado de Nagelkerke del 0,168 y un Error Tipo II del 100%.

4.4.1.3 Periodo 3: desde 2002 hasta 2007

Se ha descartado la estimación del modelo logit en este intervalo porque la muestra no incorpora ninguna observación por las limitaciones de información ya indicadas.

4.4.1.4 Periodo 4: desde 2011 hasta 2014

En el periodo de 2011 a 2014 se incluyen 74 casos de los que 19 corresponden a Cajas Rurales que habrían cesado su actividad a lo largo del periodo.

Al igual que se ha realizado para el periodo 2000-2002, los datos económicos y financieros utilizados corresponden a los presentados por cada una de las Cajas Rurales en el ejercicio previo al inicio del periodo.

Tras la estimación del modelo logit mediante el empleo de los criterios de selección tradicionales (por pasos), se puede observar en la Tabla 28 que tan solo han resultado ser

significativa las variables TAMAÑO (logaritmo neperiano del volumen de activos) y OF1 (logaritmo neperiano del número de oficinas).

TABLA 28: PARÁMETROS ESTIMADOS EN EL MODELO LOGIT DEL PERIODO 2011-2014

		B	ERROR ESTÁNDAR	WALD	GL	SIG.
Paso 11 ^a	Tamaño	-3,090	1,522	4,122	1	,042
	OF1	1,451	,604	5,772	1	,016
	Constante	12,070	6,800	3,151	1	,076

FUENTE: ELABORACIÓN PROPIA

A diferencia de los resultados observados en el modelo estimado para el periodo 2000-2002, en este nuevo periodo han resultado ser significativas un número menor de variables explicativas, si bien ambos tendrían en común las mismas variables.

Llama la atención el hecho de que los signos asociados a los coeficientes de ambas variables son opuestos en ambos periodos, es decir, mientras que en el periodo 2000-2002 el TAMAÑO afectaría positivamente a los procesos de fusión, en el periodo 2011-2014 ejercería negativamente sobre la probabilidad de fusión, confirmándose que en este periodo la mayor parte de las integraciones respondieron a la necesidad de incrementar el tamaño de las Cajas Rurales menor dimensión, en consonancia con la situación mostrada por el sector financiero español en el contexto de la crisis económica iniciada en 2007.

Por su parte la variable OF1, que mide el volumen de oficinas operativas a cierre del ejercicio, afectaría positivamente sobre la probabilidad de fusión de una Caja Rural (a diferencia de lo determinado en el periodo 2000-2002 en el que la influencia era negativa), observándose que aquellas Cajas Rurales con mayor número de oficinas mostraban una mayor probabilidad de formar parte de un proceso de fusión, previsiblemente como consecuencia de la necesidad del sector de incrementar su ámbito territorial de actuación

ya que la posibilidad de incrementar el volumen de oficinas en sus territorios de origen era reducida dada la elevada dimensión de estas entidades.

Una vez determinadas las variables significativas, es necesario contrastar la bondad del ajuste del modelo estimado para este periodo.

En este sentido en la Tabla 29 se puede observar que la bondad, medida a partir del estadístico R cuadrado de Nagelkerke es reducida, y próxima al valor de 0,171, inferior al 0,617 observado en el modelo estimado para el periodo 2000-2002, por lo que el ajuste que estaría brindando el modelo no sería aceptable.

TABLA 29: PRUEBA DE LA BONDAD DEL AJUSTE DEL MODELO

PASO	LOGARITMO DE LA VEROSIMILITUD -2	R CUADRADO DE COX Y SNELL	R CUADRADO DE NAGELKERKE
11	75,186 ^b	,116	,171

FUENTE: ELABORACIÓN PROPIA

De hecho, en la Tabla 30 se puede apreciar como el poder de clasificación del modelo es bajo en el grupo de las Cajas Rurales fusionadas (apenas un 11%), contrastando con la muestra de validación, en el que poder de clasificación de las Cajas Rurales operativas se eleva hasta el 77%, aunque con un empeoramiento de la clasificación de las entidades operativas.

TABLA 30: TABLA DE CLASIFICACIÓN DEL MODELO LOGIT PARA EL PERIODO 2011-2014

	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	96%	11%	74%	58%	77%	60%

FUENTE: ELABORACIÓN PROPIA

Al igual que se ha realizado para el periodo 2000-2002 se plantea ahora el interrogante de si la selección de las variables explicativas a partir del método de componentes principales mejoraría el poder de clasificación del modelo, resultados recogidos en Anexo 3.

Para ello se calcula nuevamente el Análisis factorial a partir de las observaciones correspondientes a este periodo, y tras verificar los supuestos necesarios para la aplicación del análisis factorial, se constata la existencia de tres factores que explicarían conjuntamente el 84,171% de la varianza acumulada.

Se observa que las variables que muestran mayor correlación con los factores calculados son EFICIENCIA, OF1 y ESFUERZOPROV.

Tras la estimación del modelo empleando las tres variables que muestran mayor correlación se observa que tan solo ha resultado ser significativa la variable OF1, si bien se produciría un deterioro en la bondad del ajuste, con un R cuadrado de Nagelkerke que disminuiría hasta el valor de 0,08 (frente al 0,17 alcanzado mediante la selección tradicional) y un Error Tipo II del 100%.

Reestimando el modelo a partir de los tres factores calculados y sin utilizar las variables explicativas se observa que resultan ser significativos los factores que miden capitalización y dimensión, y aunque se produciría una mejora del valor del R cuadrado de Nagelkerke hasta rondar el 0,174, el Error tipo II continuaría siendo elevado y muy próximo al 79%.

4.4.1.5 Periodo completo: 1991-2015

Dada la ausencia de observaciones suficientes para estimar los modelos en los periodos de 1992 hasta 1994, y 2002 hasta 2007, así como los niveles de clasificación aportada por los

modelos estimados para el resto de periodos, se plantea ahora la posibilidad de estimar el modelo logit considerando la totalidad de individuos fusionados a lo largo del periodo de estudio (reservándose 1/3 parte para la posterior validación del modelo) y emparejándolos con Cajas Rurales que se habrían mantenido operativas a lo largo de todo el periodo de estudio.

El emparejamiento se ha realizado de forma aleatoria, si bien se habrían considerado aquellas entidades operativas que se encontrarían en el mejor cuartil de la ratio de APALANCAMIENTO.

Con este procedimiento se dispone de una muestra homogénea compuesta por un total de 40 observaciones, de las que 20 corresponderían a entidades fusionadas y 20 a entidades operativas.

En la Tabla 31 se observa que la significatividad de las variables TAMAÑO, ROA, OF1 y UB a un nivel de significatividad del 0,1.

TABLA 31: PARÁMETROS ESTIMADO DEL MODELO LOGIT PARA EL PERIODO 1991-2014

	B	ERROR ESTÁNDAR	WALD	GL	SIG.
Tamaño	-4,588	1,925	5,683	1	0,17
ROA	-196,085	95,479	4,218	1	0,040
OF1	2,046	0,804	6,479	1	0,011
UB			5,171	2	0,075
UB(1)	2,325	1,388	2,806	1	0,094
UB(2)	2,534	1,207	4,410	1	0,036
Constante	18,662	8,350	4,995	1	0,025

FUENTE: ELABORACIÓN PROPIA

En este sentido la variable TAMAÑO es la que ejerce mayor influencia sobre la probabilidad de fusión de una Caja Rural, indicando que cuanto mayor sea el tamaño, menor será la probabilidad de fusión. De forma similar la variable que mide la rentabilidad

del activo (ROA) estaría indicando que cuanto mayor sean los beneficios, menor será la probabilidad de que una Caja Rural sea integrada en otra entidad financiera. La variable OF1, que mide el volumen de oficinas de cada Caja Rural, pondría de manifiesto que cuanto mayor sea el número de oficinas, mayor será la probabilidad.

La variable categórica ubicación de la entidad (UB) indica una mayor probabilidad de fusión entre aquellas Cajas Rurales que se ubican en el Norte y Levante.

En la Tabla 32 se han recogido los resultados de los test que miden la idoneidad del modelo, observándose una mejora del estadístico R cuadrado de Nagelkerke hasta el valor del 0,628 muy superior a los niveles presentados en los modelos anteriores.

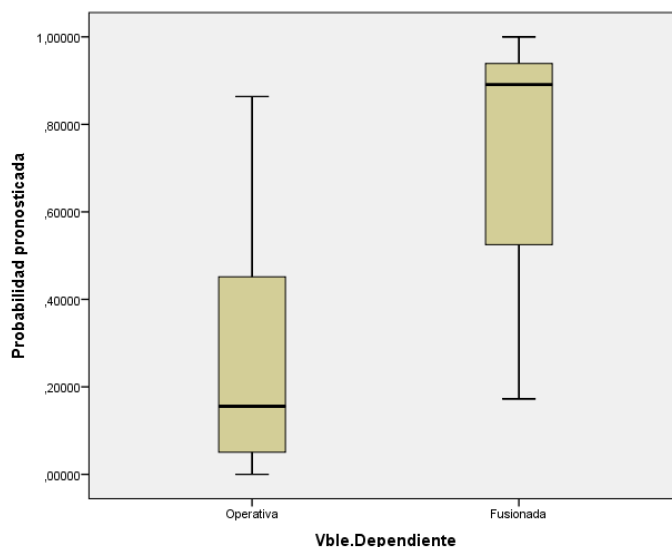
TABLA 32: BONDAD DEL AJUSTE DEL MODELO LOGIT

PASO	-2LL	R CUADRADO DE COX Y SNELL	R CUADRADO DE NAGELKERKE
9	29,998 ^e	,471	,628

FUENTE: ELABORACIÓN PROPIA

En la Ilustración 29 se han dibujado los percentiles en los que se ubican cada una de las Cajas Rurales utilizadas para la estimación del modelo, a partir de la probabilidad que les asigna y diferenciando entre las Cajas Rurales operativas y fusionadas. Se observa que el modelo asigna probabilidades inferiores al grupo de las Cajas Rurales operativas, mientras que la totalidad de las Cajas Rurales fusionadas se encuentran en los percentiles superiores (de mayor probabilidad).

ILUSTRACIÓN 29: CLASIFICACIÓN DEL MODELO LOGIT ESTÁTICO



FUENTE: ELABORACIÓN PROPIA

Para evaluar la presencia de los Errores Tipo I (falso positivo) y Error Tipo II (falso negativo) se establece un punto de corte en la probabilidad estimada (es decir, valor a partir del cual una Caja Rural puede ser considerada operativa o fusionada) siguiendo el criterio del punto de corte óptimo del Índice de Youden²⁵ (Fuentes; 2003)

En este sentido, en Tabla 33 se puede observar que el nuevo modelo propuesto clasificaría correctamente al 82% de las Cajas Rurales utilizadas para la estimación del mismo, minimizando la consecución del Error Tipo I al 20% y el Error tipo II al 15%.

TABLA 33: TABLA DE CLASIFICACIÓN DEL MODELO LOGIT ESTÁTICO.

	Muestra inicial			Muestra de validación		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	80%	85%	82%	73%	70%	72%

FUENTE: ELABORACIÓN PROPIA

²⁵ Con este índice se busca el valor del punto de corte que maximice los verdaderos positivos y los verdaderos negativos.

A pesar de que el modelo planteado presentaría una bondad del ajuste que se encontraría en niveles aceptables, se plantea observar si la utilización de un método diferente de selección de las variables explicativas mejoraría los resultados obtenidos.

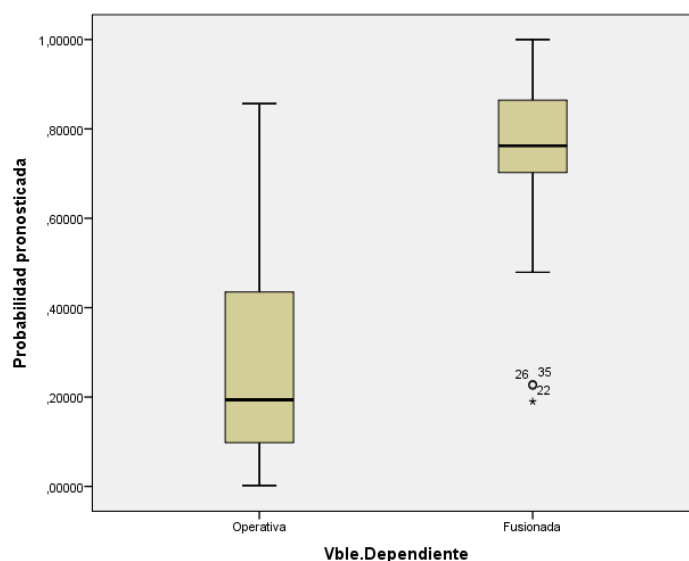
Para ello se considerará el método de los componentes principales, utilizándose para la estimación del modelo logit las variables que muestran las mayores correlaciones con los factores estimados. Posteriormente se repetirá el ejercicio incluyendo exclusivamente los factores en sustitución de las variables explicativas.

A modo de recordatorio, en el epígrafe segundo se constaba la existencia de cuatro variables altamente correlacionadas con los factores calculados, concretamente CAPITAL, OF2 y EMP. Se descarta el uso de la variable INT1 por su carácter dinámico.

Se ha procedido a reestimar el modelo logit estático se observa que el modelo logit estimado presentaría una bondad del ajuste buena, aunque inferior a la presentada por el modelo estimado a partir de los métodos de selección tradicionales, con un R cuadrado de Nagelkerke cercano al 0,52, si bien se produciría una mejora en la clasificación de las Cajas Rurales operativas y fusionadas.

De hecho, en la Ilustración 30 se puede apreciar como la se mejora la clasificación de las Cajas Rurales fusionadas, que se situarían en percentiles superiores a los mostrados por el modelo a partir de la selección de las variables explicativas a partir de la selección a partir de los métodos tradicionales. Por su parte, las Cajas Rurales operativas continuarían clasificadas en los percentiles inferiores.

ILUSTRACIÓN 30: CLASIFICACIÓN DEL MODELO LOGIT ESTÁTICO A PARTIR DE LA SELECCIÓN POR COMPONENTES PRINCIPALES.



FUENTE: ELABORACIÓN PROPIA

Sin embargo, en la Tabla 34 se pone de manifiesto que la aplicación de este modelo a la muestra de validación empeora la capacidad predictiva del modelo para las entidades que se han fusionado, ya que el porcentaje de acierto se reduce significativamente hasta apenas el 10% (con un Error Tipo II del 90%). Por el contrario, las Cajas Rurales operativas mejorarían su clasificación hasta el 94% (Error Tipo I del 6%).

TABLA 34: CLASIFICACIÓN DE LAS ENTIDADES A PARTIR DEL MODELO LOGIT CON SELECCIÓN DE VARIABLES POR COMPONENTES PRINCIPALES

	Muestra inicial			Muestra de validación		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	80%	80%	80%	94%	10%	52%

FUENTE: ELABORACIÓN PROPIA

Finalmente señalar que la reestimación del modelo a partir de los cuatro factores (en sustitución de las variables explicativas) combinados con las variables categóricas, no mejoraría sustancialmente los resultados del modelo logit estimado mediante selección por componentes principales, manteniendo el poder de clasificación de la muestra inicial.

4.4.2 Estimación de la regresión logística multiperiodo

Una vez confirmada que la bondad del ajuste ofrecida por el modelo logit es aceptable, con un alto grado de clasificación tanto en el grupo de las entidades operativas como fusionadas, se plantea a continuación comprobar si la versión del modelo logit que propuso Shumway (2001) mejora la calidad de las estimaciones.

Desde su propuesta, ha sido habitual el uso de esta versión extendida del modelo logit para la predicción de la insolvencia empresarial, si bien en el contexto de la insolvencia bancaria el volumen de investigaciones es inferior, siendo Cole *et al.* (2009) el primero en aplicarla para el estudio y predicción de la insolvencia bancaria en EE.UU. entre 1985 y 1993.

La principal diferencia entre el modelo logit clásico y la variación propuesta por Shumway (2001) se encuentra en la muestra a utilizar, ya que el carácter dinámico de esta nueva versión permite incluir todas las observaciones disponibles para cada Caja Rural a lo largo del periodo de estudio (1991-2014). Ejemplo de ello son las investigaciones de Nam *et al.* (2008), Cole *et al.* (2009) o Duda *et al.* (2010), entre otras.

En el caso de las Cajas Rurales españolas, se utilizará una muestra homogénea de entidades operativas y fusionadas que incorporará toda la información temporal disponible para cada una de ellas.

Para la estimación del modelo se han considerado un total de 755 observaciones, de las que 20 observaciones corresponden a Cajas Rurales fusionadas.

La variable dependiente mantiene la misma definición que la empleada en los modelos logit anteriores, una dicotómica que toma el valor 1 si la entidad se fusiona, y el valor 0 si

se mantiene operativa. De forma similar se mantiene la codificación de las variables explicativa categóricas.

Según Shumway (2001) para la estimación del modelo se sigue la metodología de máxima verosimilitud, aunque introduciendo la función de riesgo basal.

En este sentido Shumway (2001) propone utilizar como función de riesgo basal la edad, criterio que será empleado para la estimación del modelo propuesto.

Para seleccionar las variables explicativas a incorporar en el modelo se utilizará la metodología de selección basada en los criterios tradicionales (paso a paso) y la selección por el método de los componentes principales.

En la Tabla 35 se recogen las variables que han resultado ser significativas en la estimación del modelo, observándose que a un nivel de significación del 0,05 las variables NEGOCIO, PARO, OF1, OF2, EMP y FGD serían significativas.

TABLA 35: VARIABLES INTRODUCIDAS EN EL MODELO LOGIT MULTIPERIODO

	B	ERROR ESTÁNDAR	WALD	GL	SIG.
Paso 13 ^a Negocio	4,749	1,339	12,575	1	,000
ROA	-145,025	38,308	14,332	1	,000
Paro	12,829	5,309	5,838	1	,016
FGD(1)	-2,613	,797	10,737	1	,001
UB			14,642	2	,001
UB(1)	3,097	,845	13,446	1	,000
UB(2)	1,213	,711	2,914	1	,088
Constante	-6,859	1,601	18,357	1	,000

FUENTE: ELABORACIÓN PROPIA

La variable NEGOCIO influye positivamente en la probabilidad de que una Caja Rural forme parte de un proceso de fusión, indicando que cuanto mayor sea la concentración del activo en los créditos a la clientela, mayor será su probabilidad de fusión.

Por su parte la tasa de desempleo nacional (PARO) muestra un comportamiento positivo con la probabilidad de fusión, indicando que un incremento de las tasas de paro incrementa la probabilidad de que una Caja Rural se fusione. Esta influencia se puede justificar por el hecho de que la tasa de paro influye negativamente sobre la calidad de una cartera de crédito a la clientela mayoritariamente concentrada en el crédito a los hogares.

En lo que respecta a la ubicación geográfica, se observa una mayor probabilidad de fusión en las Cajas Rurales ubicadas en el Norte, Centro y Sur frente a sus homólogas ubicadas en el levante.

Por el lado contrario se encuentra las variables ROA (rentabilidad del activo) y FGD (intervención del fondo de garantía de depósitos), ya que ambas muestran una influencia negativa sobre la probabilidad de fusión.

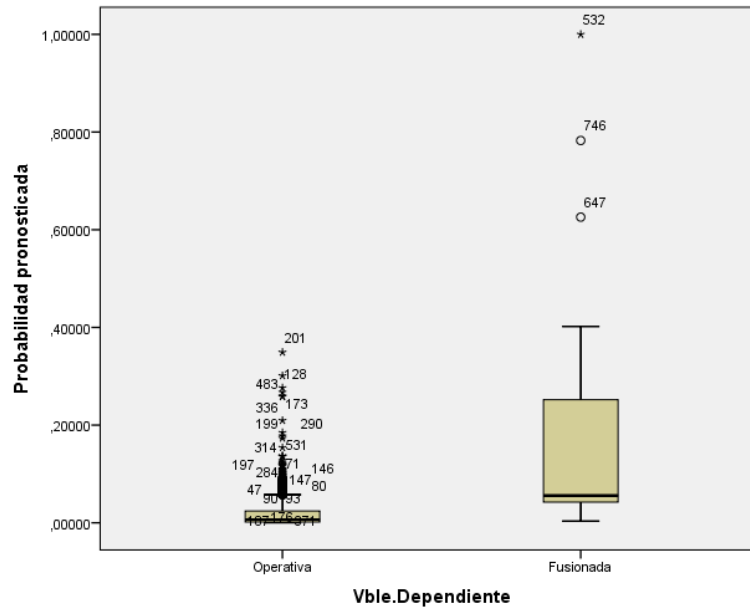
Para evaluar la bondad del ajuste del modelo se recurre al estadístico que mide la lejanía (-2LL), si bien Shumway (2001) indica la necesidad de realizar un ajuste sobre el mismo por la configuración de la muestra, en la que cada una de las observaciones no corresponden a un individuo diferente, proponiendo su división entre el número medio de observaciones por entidad incluidas en la muestra, obteniéndose un resultado de 7,079365, inferior al presentado en el epígrafe anterior por el modelo logit estático.

Para medir el poder de clasificación del modelo, Cole (2009) propone utilizar los percentiles, con el fin de observar si las probabilidades de fusión predichas por el modelo para cada uno de los individuos son superiores en el Grupo de las Cajas Rurales fusionadas, en comparación con el Grupo de las Cajas Rurales operativas.

En este sentido, en la Ilustración 31 se puede observar que el modelo logit multiperiodo estima las mayores probabilidades para el grupo de las Cajas Rurales fusionadas, que se

sitúan en los percentiles más altos. Por el contrario, el conjunto de cajas rurales operativas se encuentra clasificado en su mayoría en el percentil inferior.

ILUSTRACIÓN 31: CLASIFICACIÓN DEL MODELO LOGIT MULTIPERIODO

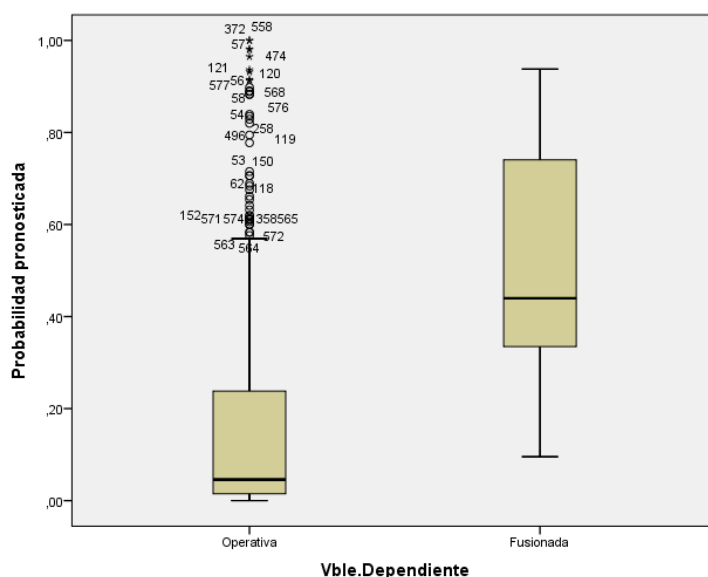


FUENTE: ELABORACIÓN PROPIA

La afinidad del ajuste es superior en el grupo de las entidades operativas, con una probabilidad de fusión que apenas se acerca al 5% para la mayor parte de ellas y que claramente se diferencian de las probabilidades presentadas por las Cajas Rurales fusionadas.

El modelo ofrece resultados diferentes para la muestra de validación, mejorando los anteriores. En este sentido, en la Ilustración 32 se puede observar que las probabilidades estimadas para las Cajas Rurales fusionadas se encuentran mayoritariamente comprendidas entre el 40% y el 80%, aunque empeorando la probabilidad de fusión para las Cajas Rurales que se mantuvieron operativas.

ILUSTRACIÓN 32: CLASIFICACIÓN DE LA MUESTRA DE VALIDACIÓN DEL LOGIT MULTIPERÍODO



FUENTE: ELABORACIÓN PROPIA

Para evaluar la presencia de los Errores Tipo I (clasificar incorrectamente a una entidad operativa) y Error Tipo II (clasificar incorrectamente a una Caja Rural fusionada) se establece un punto de corte en la probabilidad estimada (es decir, valor a partir del cual una Caja Rural puede ser considerada operativa o fusionada) siguiendo el criterio del punto de corte óptimo del Índice de Youden (Fuentes; 2003).

En la Tabla 36 se recoge la clasificación que efectúa el modelo considerando las Cajas Rurales incorporadas inicialmente para su estimación, y la muestra de validación. Se puede observar que el modelo funciona mejor clasificando las entidades operativas en la muestra inicial, que en la muestra de validación, si bien se mejoraría el porcentaje de aciertos en el caso de las Cajas Rurales realmente fusionadas a lo largo del periodo (minimiza el Error Tipo II).

No obstante, realizando un análisis pormenorizado de las Cajas Rurales que la muestra de validación considera operativas y que habrían sido mal clasificadas por el modelo, se observa que el 30% de las mismas corresponderían a observaciones de Cajas Rurales que varios años después formarían parte de un proceso de fusión, lo que podría ser señal de

que el modelo anticipa los procesos de fusión con varios años de antelación. Eliminando este efecto, el porcentaje de acierto del modelo en relación a las Cajas Rurales operativas se elevaría hasta el 77%.

TABLA 36: TABLA DE CLASIFICACIÓN MODELO LOGIT MULTIPERIODO

	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	98,5%	35%	96%	67%	90%	80%

FUENTE: ELABORACIÓN PROPIA

Se plantea ahora el interrogante sobre si con la selección de las variables explicativas a partir del método de los componentes principales podría mejorar la calidad en la clasificación del modelo.

Para ello se estimará nuevamente el modelo a partir de las variables explicativas mayoritariamente correlacionadas con los factores calculados en el epígrafe tercero de este Capítulo. Para ello se utilizarán las variables OF2, INT1, EMP y CAPITAL.

En la Tabla 37 se observa que finalmente han resultado ser significativas las variables OF2, INT1, FGD y UB.

TABLA 37: MODELO LOGIT MULTIPERIODO SELECCIONADO POR COMPONENTES PRINCIPALES

		B	ERROR ESTÁNDAR	WALD	GL	SIG.
Paso 4 ^a	OF2	-,007	,002	10,449	1	,001
	INT1	-29,828	11,817	6,371	1	,012
	FGD(1)	-2,431	,796	9,332	1	,002
	UB			8,096	2	,017
	UB(1)	1,955	,720	7,373	1	,007
	UB(2)	,656	,655	1,003	1	,316
	Constante	-,487	,971	,252	1	,616

FUENTE: ELABORACIÓN PROPIA

En la Tabla 38 se recogen la tabla de clasificación del modelo logit estimado a partir del método de selección de componentes principales, observándose que este sistema logra una ligera mejoría en la muestra de validación en el caso de las Cajas Rurales fusionadas

que, si bien, se encontraría alejado de los valores presentados por el modelo logit estimado por los métodos de selección tradicionales (paso a paso).

TABLA 38: TABLA DE CLASIFICACIÓN DEL MODELO LOGIT MULTIPERIDO SELECCIONADO POR COMPONENTES PRINCIPALES

	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Aciertos	98%	35%	97%	96%	40%	68%

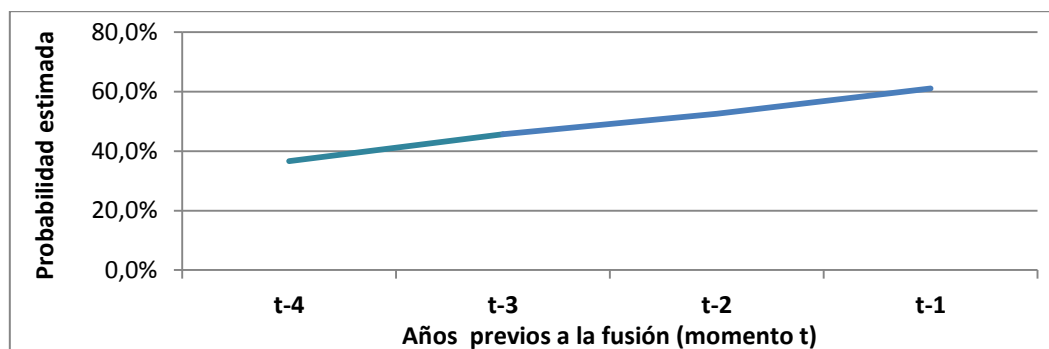
FUENTE: ELABORACIÓN PROPIA

Situación similar es la mostrada por el modelo estimado a partir del uso exclusivo de los cuatro factores y las variables categóricas, con una bondad del ajuste de apenas el 25% en la clasificación de las Cajas Rurales fusionadas.

Una de las ventajas del modelo logit multiperiodo subyace en el hecho de que al estar utilizando la totalidad de observaciones disponibles para una misma Caja Rural, permite conocer la anticipación con la que es posible prever su fusión por motivos de solvencia.

En este sentido en la Ilustración 33 se ha representado gráficamente las probabilidades medias estimadas por el modelo para las Cajas Rurales fusionadas, observándose que las mayores probabilidades se alcanzan utilizando los datos económicos y financieros de hasta dos años anteriores al proceso de fusión, si bien los datos de tres años anteriores también ofrecerían resultados aceptables, aunque inferiores a los anteriores.

ILUSTRACIÓN 33: PROBABILIDADES MEDIAS ESTIMADAS A PARTIR DE DATOS ECONÓMICOS Y FINANCIEROS DE HASTA CUATRO AÑOS ANTERIORES A LA FUSIÓN.



FUENTE: ELABORACIÓN PROPIA

4.4.3 Modelo de Riesgos Proporcionales o Modelo de Cox

Los modelos logit estimados han permitido conocer la probabilidad de que una Caja Rural forme parte de un proceso de fusión a partir de una serie de variables explicativas.

Se plantea ahora complementar las conclusiones de los modelos logit anteriores a partir del empleo de técnicas de supervivencia, concretamente el modelo de riesgos proporcionales de Cox, ya que esta técnica permite estudiar el tiempo de supervivencia y la influencia de las variables explicativas sobre la probabilidad de que el suceso acontezca en el siguiente intervalo de tiempo.

El único supuesto restrictivo del modelo de Cox es que la relación de riesgos entre dos individuos debe mantenerse estable a lo largo del tiempo., es decir, formalmente debe cumplirse que:

$$\frac{h(t, i)}{h(t, j)} = \frac{h_0(t)e^{\beta_1 X_1 + \beta_2 X_2}}{h_0(t)e^{\beta_1 X_1 + \beta_2 X_2}}$$

Siendo x las variables explicativas, i uno de los sujetos, j el otro sujeto y t el tiempo. Suponiendo que x_1 es similar para ambos casos y que x_2 toma el valor 1 en el caso del individuo i , y el valor 0 en el caso del individuo j , se verifica que:

$$\frac{h(t, i)}{h(t, j)} = \frac{h_0(t)e^{\beta_1 X_1 + \beta_2 1}}{h_0(t)e^{\beta_1 X_1 + \beta_2 0}} = e^{\beta_2}$$

Confirmándose así que la relación de riesgos es proporcional entre ambos individuos.

La variable dependiente a utilizar para la estimación del modelo de Cox no es otra que el tiempo que transcurre desde que se inicia el seguimiento del individuo hasta que sucede el evento (fusión) o se finaliza el estudio.

A las Cajas Rurales que se mantiene operativa a la finalización del mismo se las denomina censura por la derecha, ya que la supervivencia exacta de esas entidades más allá del estudio es desconocida (Carvalho *et al*, 2015).

Mínguez (2005), en su investigación sobre la insolvencia de empresas constructoras en España, puso de manifiesto que el uso la regresión de Cox a partir de covariables que varían con el tiempo, similares a las ratios financieras utilizadas en esta Tesis, podría suponer el incumplimiento de la hipótesis de riesgos proporcionales, proponiendo el uso de la versión extendida del modelo de Cox.

En el Manual de Lagares *et al*. (2014) se recoge una buena justificación formal del funcionamiento del modelo de Cox extendido, por ello simplemente señalar que su expresión formal es similar a la ya indicada en el capítulo tercero de esta Tesis, recogiendo el efecto de las covariables que varían con el tiempo ($\beta_k x_k(t)$):

$$h(t; x_t) = h_0(t)e^{[\sum \beta_j x_j + \sum \beta_k x_k(t)]}$$

En definitiva, el modelo de Cox extendido requiere que se divida el tiempo en subperiodos para los cuales la variable independiente se mantiene constante, de forma que para cada individuo se obtiene un número de observaciones similar a los distintos valores que toma la variable (Lagares *et al*, 2014).

Mínguez (2006) señala que con el uso de esta versión extendida se pierde la capacidad predictiva del modelo porque no es posible estimar la función de supervivencia, por ello el modelo solo puede ser utilizado para determinar que variables explicativas ejercen mayor influencia sobre la función de riesgo.

En esta línea Ayala *et al* (2007a) y posteriormente Ayala *et al* (2007b), en su análisis de la supervivencia de la banca comercial venezolana, confirman que un modelo estimado siguiendo este procedimiento cumplen con el supuesto de riesgos proporcionales, si bien

los autores tan solo utilizan esta técnica para determinar los factores asociados, no para llevar a cabo predicciones.

4.4.3.1 Modelo de Cox con covariables dependientes del tiempo.

Para la estimación del modelo de Cox aplicado a la supervivencia de las Cajas Rurales en España se utilizará la muestra homogénea empleada para la estimación del modelo logit multiperiodo.

La muestra a utilizar está integrada por un total de 756 observaciones de las que 20 corresponden a Cajas Rurales que se habrían fusionado.

Al igual que se realizó en el caso de los modelos logit, para la estimación del modelo de Cox se recurrirá inicialmente a los métodos de selección tradicionales entre las variables

En la Tabla 39 se han recogido las variables que han resultado ser significativas de la estimación del modelo de Cox, observándose que las variables EDAD, NEGOCIO, ROA, OF1, INT2, FGD y UB han resultado ser significativas a un nivel del 0,05.

TABLA 39: VARIABLES SIGNIFICATIVAS EN EL MODELO DE COX

	B	SE	WALD	GL	SIG.	EXP(B)	EXP(CXB)
Edad	,000	,000	8,941	1	,003	1,000	1,000
Negocio	3,321	1,618	4,213	1	,040	27,685	1,034
ROA	-94,205	46,911	4,033	1	,045	,000	,390
OF1	,392	,174	5,109	1	,024	1,481	1,004
INT2	66,205	32,613	4,121	1	,042	5,65*1028	1,939
FGD	-4,716	,987	22,804	1	,000	,009	,954
UB			23,996	2 ^a	,000		1,000
UB(1)	3,425	,799	18,374	1	,000	30,710	1,035

FUENTE: ELABORACIÓN PROPIA

La primera columna de la Tabla 39 recoge los coeficientes estimados de la regresión de Cox. Por su parte los valores de $\text{Exp}(B)$ representan los cambios en el riesgo de fusión al cambiar cada covariable en una unidad.

Para una mayor comprensión de los efectos de las covariables en la función de riesgos, dada la naturaleza de las ratios financieras utilizadas, Ayala *et al* (2007a) propone realizar la siguiente transformación para su interpretación:

$$\text{exp}(B) \rightarrow \text{exp}(C_x B)$$

Donde C_x representa una unidad razonable de cambio. Considerando que $C_x=0,01$, pueden interpretarse los cambios como la variación de las ratios en un 1%. Los resultados de esta modificación han sido recogidos en la columna $\text{Exp}(C_x B)$ de la Tabla 39. Para su interpretación se calcula la diferencia $1-\text{Exp}(C_x B)$, si el valor es superior a uno, la influencia es positiva, mientras que si es inferior a 1, la influencia es negativa.

- EDAD: la influencia de esta variable sobre la función de riesgo es poco significativa, ya que el valor tiende a cero.
- NEGOCIO: esta variable ejerce una influencia positiva sobre el riesgo de fusión, hasta el punto de que un incremento de un 1% en la concentración del activo en el crédito a clientes, está asociado un aumento del riesgo de fusión del 3,4%.
- ROA (rentabilidad del activo) ejerce una influencia negativa sobre el riesgo de fusión, hasta el punto de que un incremento en la rentabilidad del activo de una Caja Rural, está asociado a una menor probabilidad de fusión de la Caja Rural.
- OF1: cuando el volumen de oficinas se incrementa en un 1%, el riesgo de fusión se incrementa en un 0,4%.

- INT2: mide la evolución de los tipos de interés del mercado hipotecario a un año. También ejercen una influencia positiva sobre el riesgo de fusión de una Caja Rural. Concretamente, un incremento del 1% de los tipos, aumentan el riesgo de fusión en un 93%, situándose como la variable que mayor influencia ejerce en relación a la supervivencia de las Cajas Rurales consideradas en el estudio.
- UB: esta variable categórica está indicando que las Cajas Rurales que se ubican en el Norte de España presentan un riesgo de fusión que es un 3,5% superior al de resto de Cajas Rurales.

La validez del modelo de Cox se contrasta a través de la Hipótesis General del Modelo que busca determinar si alguna variable independiente tiene influencia estadísticamente significativa.

De forma genérica, la Hipótesis General del Modelo lleva a cabo el siguiente contraste de hipótesis sobre los parámetros poblacionales:

$$H_0: B_1 = B_2 = \dots = B_i = 0$$

$$H_1: B_i \neq 0, \text{ para algún } i$$

Si no se rechaza la hipótesis nula, el modelo de regresión de Cox no sería válido por implicar la existencia de influencia estadísticamente significativa de alguna variable independiente.

Para contrastar las hipótesis se recurre al estadístico Δ_{-2LL_0} , donde L indica el logaritmo neperiano y L_0 la verosimilitud.

Esta verosimilitud puede oscilar entre 0 y 1, mientras que LL_0 es el logaritmo neperiano de la función de verosimilitud. Interesa contrastar la evolución del logaritmo neperiano de la función de verosimilitud inicial y de la solución final. Este estadístico se denota mediante

el símbolo Δ_{-2LL0} y se distribuye mediante una distribución Chi-Cuadrado (X^2) con k grados de libertad, siendo k el número de variables independientes.

En la Tabla 40 se recoge que el valor del estadístico de contraste para la séptima iteración (después de eliminar todas las variables no significativas) es $-2LL_{0final} = 132,8$ que se distribuye mediante una *Chi-Cuadrado* con 9 grados de libertad que toma el valor 123,2 con una significación estadística próxima a cero, por lo que se acepta el cumplimiento de la hipótesis general del modelo.

TABLA 40: CONTRASTE DE LA HIPÓTESIS GENERAL DEL MODELO.

LOGARITMO DE LA VEROSIMILITUD	GLOBAL (PUNTUACIÓN)			CAMBIO RESPECTO A PASO ANTERIOR			CAMBIO RESPECTO A BLOQUE ANTERIOR			
	Chi- cuadrado	gl	Sig.	Chi- cuadrado	gl	Sig.	Chi- cuadrado	gl	Sig.	
-2	132,864	123,295	9	,000				60,097	9	,000

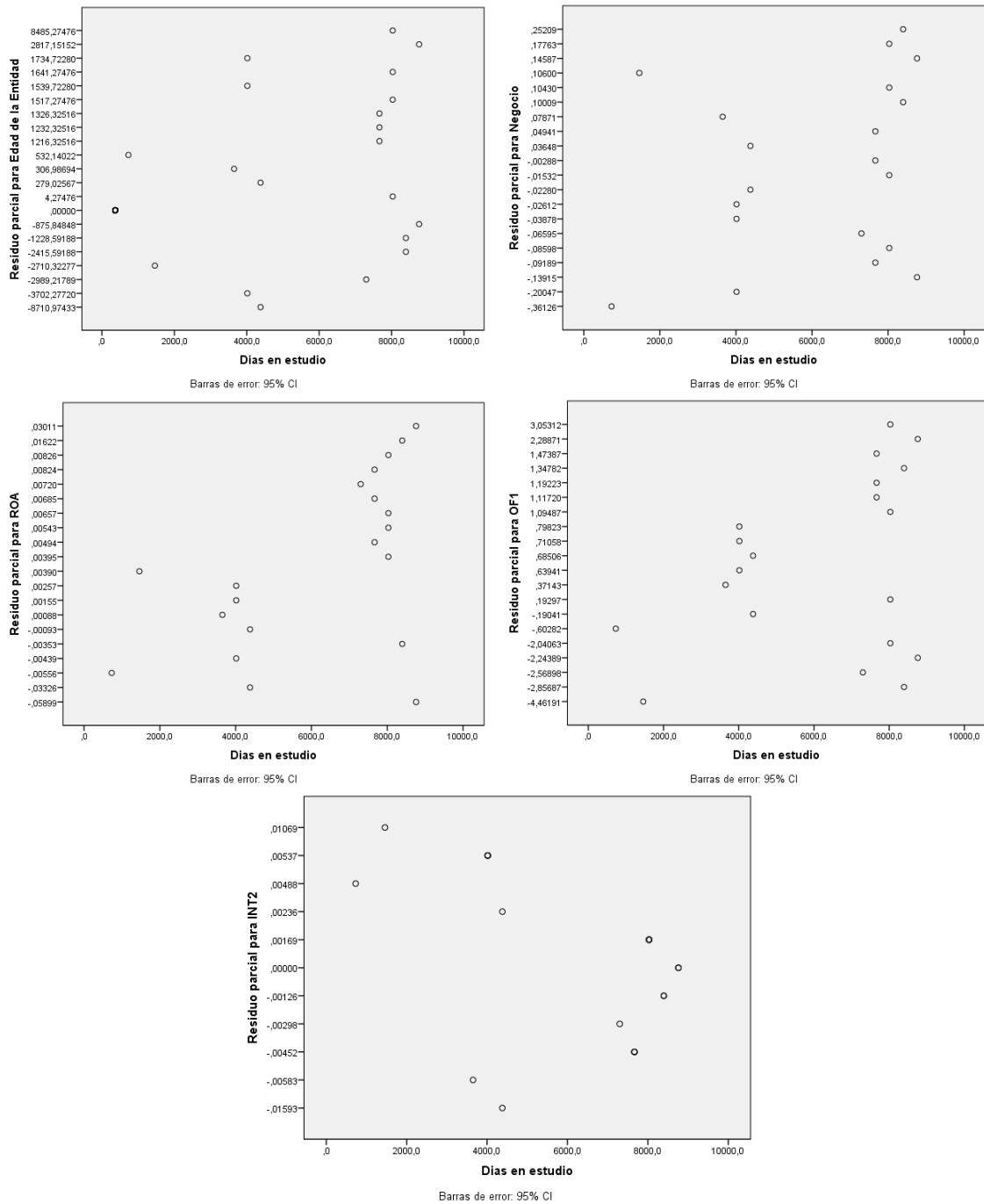
FUENTE: ELABORACIÓN PROPIA

Ayala *et al.* (2007a), Ayala *et al.* (2007b) y Lagares *et al.*, (2014) ponen de manifiesto que el modelo, a pesar de incorporar variables dependientes del tiempo, debe cumplir con la hipótesis de riesgos proporcionales para dar validez a las conclusiones obtenidas del análisis de los parámetros del modelo, si bien carecería de la capacidad de predicción.

Para su comprobación Ayala (2007a) sugiere representar gráficamente la evolución de los residuos de Schoenfeld (1982) a lo largo del tiempo de estudio, esperando una dispersión aleatoria de los mismos a ambos lados del valor cero.

En la Ilustración 34 se puede apreciar que en todas las variables la representación gráfica de los residuos de Schoenfeld se han distribuido de forma aleatoria, aceptándose el cumplimiento de la hipótesis de riesgos proporcionales.

ILUSTRACIÓN 34: COMPROBACIÓN DE LA HIPÓTESIS DE RIESGOS PROPORCIONALES MEDIANTE LA GRÁFICA DE LOS RESIDUOS DE SCHOENFELD.

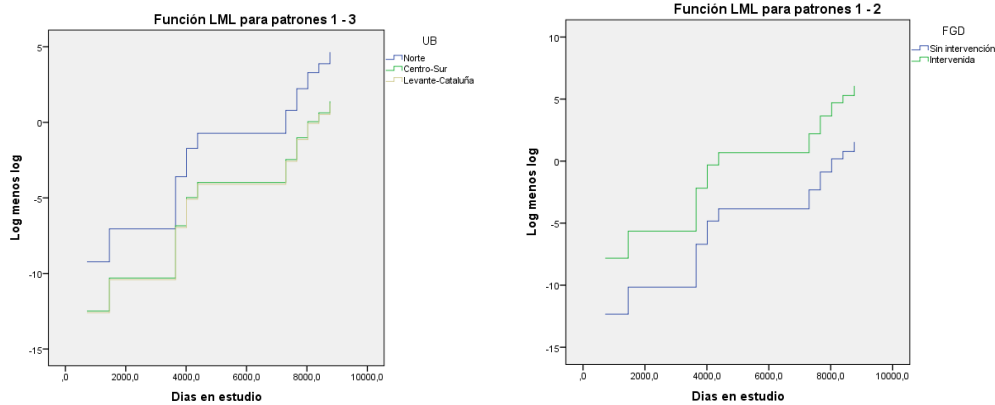


FUENTE: ELABORACIÓN PROPIA

Para validar el cumplimiento de la hipótesis de riesgos proporcionales en las variables categóricas se recurre a la representación del gráfico log-log de la curva de supervivencia, buscando que sea paralelo en todo momento.

En Ilustración 35 se ha recogido la representación del gráfico log-log para ambas variables categóricas (FGD y UB), observándose en ambos casos una distribución paralela, confirmándose así la hipótesis de riesgos proporcionales.

ILUSTRACIÓN 35: COMPROBACIÓN GRÁFICA DE LA CONDICIÓN DE RIESGOS PROPORCIONALES



FUENTE: ELABORACIÓN PROPIA

Por lo tanto, al cumplirse las hipótesis básicas del modelo, se puede aceptar las conclusiones del mismo, resultado que se encontrarían en línea con lo afirmado por Ayala *et al* (2007a).

La estimación del modelo de Cox modificado a partir de la selección por el método de componentes principales no ha mejorado la calidad de las predicciones, por lo que se rechaza el uso de esta metodología de selección.

4.4.3.2 Modelo de Cox estratificado

Como pone de manifiesto Carrasco (2001), la utilización de covariables que varían a lo largo del tiempo “*tiene una larga tradición en las ciencias biomédicas, siendo utilizados para describir hechos tales como el tiempo de supervivencia de pacientes con trasplantes de corazón*”

De la experiencia de ese tipo de investigaciones se observa que otra posible solución para el uso de variables dependientes del tiempo, como ocurre en Borges (2002) relativo a la supervivencia en paciente sometidos a diálisis renal, consistiría en realizar una estratificación de la muestra, procedimiento mediante el cual se aísla el componente dinámico de esas covariables, permitiendo el cumplimiento de la hipótesis de riesgos proporcionales dentro de cada uno de los estratos.

En este sentido Pol (1993), que recoge una justificación formal del funcionamiento del modelo de Cox estratificado, afirma que en este modelo los coeficientes de regresión son los mismos para todos los estratos, si bien la función de riesgo base puede ser diferentes para cada uno de ellos.

Para aplicar esta versión del modelo de Cox estratificado al estudio de la supervivencia de las Cajas Rurales españolas se considerará la misma muestra utilizada para la estimación del modelo de Cox en el punto anterior, siendo estratificados en función de los diferentes años, ya que los cambios en las covariables se producen anualmente de forma que se puede eliminar este efecto en estratos anuales.

En la Tabla 41 se recogen los estratos considerados para la estimación del modelo, un total de 9 estratos a lo largo de los cuales se distribuyen las Cajas Rurales operativas y fusionadas en cada momento.

Tabla 41: Estratos considerados para la estimación del modelo de Cox

ESTRATO	EVENTO	CENSURADO	PORCENTAJE CENSURADO
1.992	1	33	97,1%
1.994	1	32	97,0%
2.000	1	32	97,0%
2.001	1	29	96,7%
2.002	3	28	90,3%
2.011	4	24	85,7%
2.012	4	20	83,3%
2.013	3	20	87,0%
2.014	2	16	88,9%
Total	20	234	92,1%

FUENTE: ELABORACIÓN PROPIA

De la estimación del modelo de Cox estratificado se desprende la significatividad de un total de cuatro variables (ROA, OF2, FGD y UB) de las que tan solo las categóricas (FGD y UB) serían comunes con el modelo de Cox estimado en el subepígrafe anterior. Todas las covariables cumplirían con la hipótesis de riesgos proporcionales.

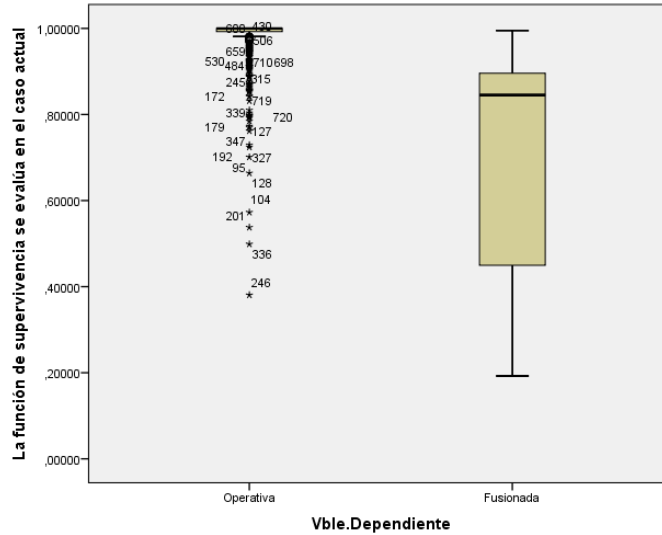
Cabe señalar que este procedimiento permite conocer la probabilidad de supervivencia y de riesgo para cada individuo, pero exclusivamente en cada estrato, perdiendo la capacidad de predicción futura.

En la Ilustración 36 se ha representado la probabilidad de supervivencia estimada a partir del modelo de Cox estratificado. En la misma se puede observar una clara diferencia en las probabilidades de supervivencia del grupo de Cajas Rurales operativas y fusionadas, ya que las primeras se encuentran concentradas en torno a 1, mientras que las Cajas Rurales fusionadas presentan probabilidades de supervivencia alejadas de ese valor.

Sin embargo, para el caso de las Cajas Rurales fusionadas cabe señalar que el modelo no recoge probabilidades de supervivencia cercanas al valor cero, de hecho se encuentran

comprendidas entre el 40% y el 80%, por lo que el ajuste que realiza no sería del todo adecuado.

ILUSTRACIÓN 36: PROBABILIDAD DE SUPERVIVENCIA ESTIMADA A PARTIR DEL MODELO DE RIESGOS PROPORCIONALES DE COX ESTRATIFICADO



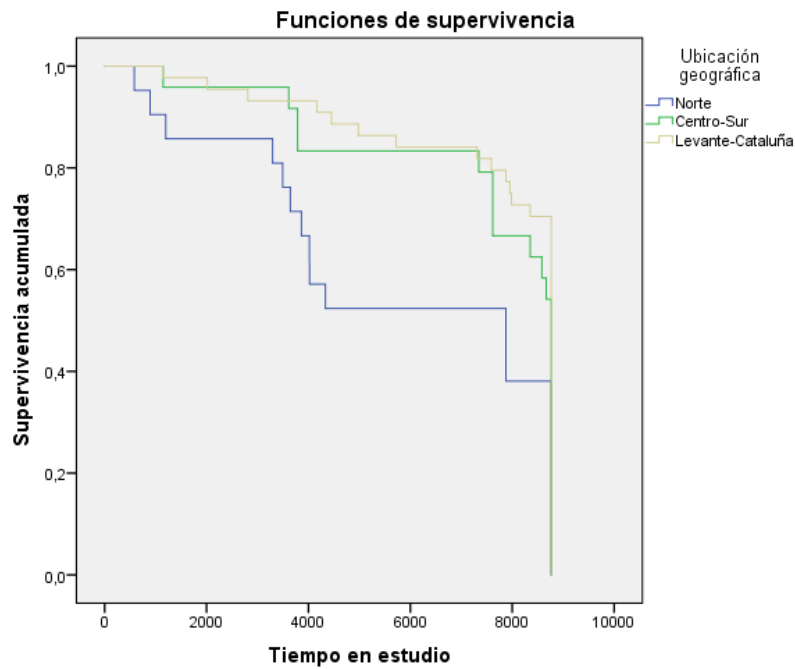
FUENTE: ELABORACIÓN PROPIA

Los resultados anteriores pueden ser complementados mediante el estimador de Kaplan-Meier, procedimiento no paramétrico que permite estimar la función de supervivencia de para cada uno de los individuos, sin considerar la influencia de las variables explicativas.

En este sentido en la Ilustración 37 se han recogido las funciones de supervivencia estimadas a partir del estimador de Kaplan-Meier, diferenciadas en función de la ubicación geográfica de la Caja Rural.

Se puede observar que las Cajas Rurales ubicadas en el Norte presentan una función de supervivencia más pronunciada, en comparación con las Cajas Rurales ubicadas en el Levante y Cataluña en las que la supervivencia ha sido superior a lo largo del periodo de estudio.

ILUSTRACIÓN 37: FUNCIÓN DE SUPERVIVENCIA A PARTIR DEL ESTIMADOR DE KAPLAN-MEIER



FUENTE: ELABORACIÓN PROPIA

4.5 CONCLUSIONES DEL CAPÍTULO

A lo largo de este capítulo se ha presentado los resultados obtenidos de la estimación de los diferentes modelos: logit para un periodo, logit estático para la muestra completa, logit multiperiodo y regresión de Cox.

De la estimación de esos modelos se puede observar que los modelos logit estático para la muestra completa y logit multiperiodo son los que presentan los mejores resultados, con un grado de clasificación elevado en ambos grupos de Cajas Rurales (operativas y fusionadas), y un mínimo error de clasificación.

Al inicio del capítulo se planteaba el interrogante sobre si el uso de un método de selección de variables basado en componentes principales podría mejorar la capacidad predictiva de

los modelos, confirmándose que en todos los casos la selección por métodos tradicionales muestra los mejores resultados.

En la Tabla 42 se han recogido las variables que han resultado ser significativas en la estimación de los diferentes modelos, observándose diferencias entre los modelos que consideran la muestra completa (1991-2014) y los modelos asociados a periodos concretos.

TABLA 42: VARIABLES SIGNIFICATIVAS EN LA ESTIMACIÓN DE LOS DIFERENTES MODELOS

Variables significativas	Logit (2000-2002)	Logit (2011-2014)	Logit estático (1991-2014)	Logit multiperiodo	Regresión de Cox
EDAD					X
TAMAÑO	X	X	X		
NEGOCIO				X	X
EFICIENCIA	X				
ROA	X		X	X	X
OF1	X	X	X		X
OF2	X				
EMP	X				
UB			X	X	X
FGD				X	X
PARO				X	
INT1					X

FUENTE: ELABORACIÓN PROPIA

En este sentido, los modelos logit para los periodos 2000-2002 y 2011-2014 encuentran como variables más significativas las que consideran el tamaño de la entidad en términos de activo y el volumen de oficinas, si bien el comportamiento de la variable TAMAÑO es diferente en ambos periodos de tiempo.

Concretamente en el primer periodo la variable TAMAÑO mostraba una influencia positiva sobre la probabilidad de fusión, señal de que los procesos de fusión acaecidos en esta primera época, entre otras cuestiones, buscaron consolidar el tamaño de las principales Cajas Rurales.

En lo que respecta al segundo periodo la influencia de la variable tamaño es negativa, indicando que, a mayor tamaño, menor probabilidad de ser absorbida, hecho que podría denotar el inicio de un proceso de consolidación del sector por la vía de la integración de

las entidades de menor tamaño, en línea con lo acontecido dentro del sistema financiero español con el resto de entidades financieras operativas.

Por su parte los modelos dinámicos (incluyen las observaciones correspondientes a todo el periodo de estudio), encuentran mayor significatividad en las variables que miden la calidad del negocio, medida tanto en términos de crédito a la clientela sobre activo, como de rentabilidad sobre activo total.

Cuestión especial es la referida a la importancia de la ubicación geográfica en los procesos de fusión de Cajas Rurales, además de haber percibido asistencia por parte del Fondo de Garantía de Depósitos en el pasado.

La función de supervivencia estimada por el método de Kaplan-Meier pone de manifiesto que las Cajas Rurales ubicadas en el norte y centro de España han mostrado una probabilidad de supervivencia menor en el periodo analizado.

En lo que respecta a la calidad en la predicción de los modelos, los modelos logit estático y dinámico muestran los mayores porcentajes de clasificación, tanto en la comprobación a partir de la muestra utilizada para su estimación, como en el posterior uso de la muestra de validación.

No obstante, el modelo logit estático (periodo completo) mostraría un peor ajuste en la muestra de validación, con una mayor consecución del Error Tipo I, situación que difiere a la mostrada por el modelo logit multiperiodo en el que la afinidad del ajuste es superior.

En este sentido en la Tabla 43 se ha comparado el poder de clasificación de los modelos logit para la muestra completa en sus versiones estática y dinámica, observándose que el modelo logit multiperiodo es el que presenta los mejores resultados para el caso de la muestra de validación, con un Error Tipo II de apenas el 10%.

En términos globales, el modelo logit multiperiodo muestra los mayores porcentajes de aciertos tanto en la muestra inicial como en la muestra de validación, de hecho esta última presentaría niveles muy superiores de aciertos en comparación con los mostrados por la muestra inicial, tal y como observó Shumway (2001) en la definición del modelo logit multiperiodo.

TABLA 43: COMPARACIÓN DEL GRADO DE CLASIFICACIÓN CORRECTA EN LA ESTIMACIÓN DE LOS MODELOS LOGIT.

	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Logit estático	80%	85%	82%	73%	70%	72%
Logit multiperiodo	98,5%	35%	96%	67%	90%	80%

FUENTE: ELABORACIÓN PROPIA

CAPÍTULO QUINTO:

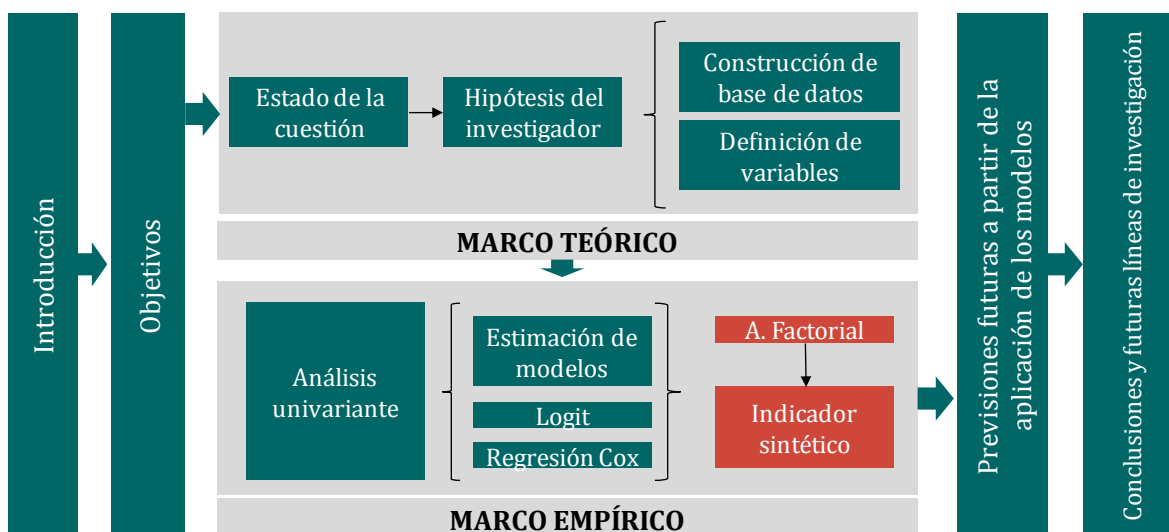
PROPUESTA DE UN ÍNDICADOR SINTÉTICO DE SALUD FINANCIERA DE LAS CAJAS RURALES

5.1 SINOPSIS DEL CAPÍTULO

Los modelos de respuesta binaria y supervivencia estimados en los capítulos anteriores han permitido confirmar que las fusiones de Cajas Rurales estuvieron motivadas por la sostenibilidad financiera, determinando los factores asociados y permitiendo conocer la probabilidad de que una Caja Rural experimente una situación similar en un futuro.

Módica *et al* (2012) afirma que “*el inconveniente de estos modelos radica en que la variable dependiente es un fenómeno difícil de establecer en forma dicotómica, debido a que empresas quebradas pueden identificarse perfectamente, si bien la definición de no quebrada es heterogénea porque el grado de salud financiera de una empresa no es directamente observable*”. Por ello se plantea a continuación la construcción de un indicador sintético específicamente diseñado para el sector de las Cajas Rurales, que permita evaluar la salud financiera de las Cajas Rurales a partir de una serie de variables explicativas que tienen en cuenta diferentes aspectos de la situación económica y financiera de estas entidades, siendo la ubicación del capítulo dentro de la sinopsis de la Tesis la reflejada en la Ilustración 38.

ILUSTRACIÓN 38: UBICACIÓN DEL CAPÍTULO QUINTO DENTRO DE LA SINOPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

5.2 METODOLOGÍA APLICABLE

Mondéjar *et al.* (2008), citando a otros autores, define un indicador sintético como una combinación de indicadores individuales que permite realizar estimaciones y mediciones con mayor sencillez que recurriendo al uso de esos componentes individuales.

El empleo de indicadores sintéticos es habitual en economía, y aunque no es objeto de esta Tesis llevar a cabo una recopilación de los principales trabajos existentes²⁶, entre otros destacar las publicaciones de Cabrero *et al* (1996), con la propuesta de un indicador sintético para predecir el comportamiento de la inflación; Perez *et al* (2009) con el que recurren a un indicador sintético para evaluar la sostenibilidad del turismo rural; o la propuesta de un indicador sintético que mide el bienestar social en las diferentes provincias españolas (Roget *et al*, 2005)..

Sin embargo en el ámbito de estudio de la insolvencia empresarial o bancaria son escasas las propuestas de indicadores sintéticos, tan solo habiendo sido localizada la investigación de Módica *et al* (2012) sobre un indicador sintético para predecir el fracaso empresarial.

A nivel general, Nardo *et al* (2005) señala que la construcción de un indicador sintético es un proceso minucioso que debe constar de las siguientes etapas:

- a. Desarrollo de un marco conceptual
- b. Selección de los indicadores
- c. Análisis multivariante
- d. Imputación de datos perdidos
- e. Normalización de los datos

²⁶ En el trabajo de Sánchez (2009) se encuentra una buena recapitulación de los principales trabajos existentes.

- f. Ponderación de la información
- g. Agregación de la información
- h. Análisis de la robustez y sensibilidad

El proceso de construcción del indicador sintético se inicia con la definición del marco conceptual, fase clave por determinar conceptualmente la finalidad del indicador, permitiendo un primer acercamiento a las variables explicativas que pueden ser necesarias para su consecución.

Enlazada con la etapa anterior se sitúa la fase de selección de las variables explicativas que finalmente serán agregadas mediante el indicador, siempre en el contexto del marco conceptual definido, ya que la calidad de un indicador sintético recae en las variables que lo definen (Schuschny *et al*, 2009).

En la fase de análisis multivariante se busca revisar las variables explicativas con el fin de encontrar patrones de comportamiento que puedan distorsionar el objetivo definido, determinando las relaciones entre variables y eliminando valores atípicos que puedan influenciar negativamente en la calidad del estimador, además de solucionar el problema de las observaciones ausentes.

La siguiente fase del proceso consiste en la normalización de los datos, es decir, expresar todas las variables explicativas en la misma unidad de medida, ya que es habitual que estas variables puedan estar midiendo diferentes conceptos que no son comparables entre ellas.

Una vez finalizadas las fases de preparación de datos y variables para la construcción del indicador sintético, se da paso a las dos etapas en las que se construye el indicador sintético como tal, y estas son la fase de ponderación, en la que se determina el peso que debe tener cada variable explicativa dentro del indicador sintético; y la fase de agregación,

en la que se agregan todas las variables explicativas (ponderadas a partir de los pesos establecidos) para la construcción final del indicador sintético.

Schuschny *et al* (2009) pone de manifiesto que no existe una metodología objetiva para establecer los pesos de cada variable en la construcción del indicador sintético, si bien estos autores detallan las de mayor tradición entre el conjunto de investigadores con la siguiente descripción:

1. Métodos subjetivos, basados en la experiencia del investigador, que asigna los pesos en función de la opinión de expertos o la experiencia observada en otros estudios.
2. Métodos técnicos, basados en distintas metodologías estadísticas, entre las que destacan por su alto grado de utilización la ponderación mediante regresiones, la determinación de los componentes principales o el análisis envolvente de datos, entre otros, tal y como se detallan a continuación:

- El cálculo de regresiones se basa en un modelo lineal que busca la relación entre la variable dependiente y una serie de variables explicativas, a partir de la cual estima una serie de parámetros asociados que son considerados los pesos de ponderación, siendo su principal limitación la suposición de relaciones lineales entre las variables.
- El análisis de componentes principales se basa en reducir la dimensión de las variables explicativas a un mínimo conjunto de factores que tratan de explicar la mayor parte de la varianza total.

Los fundamentos teóricos del análisis de componentes o análisis factorial fueron detallados en el capítulo cuarto de esta Tesis, no obstante indicar que la ponderación de los pesos de las variables explicativas se realiza a partir de las puntuaciones factoriales, corregidas en la proporción de la varianza explicada por cada uno de los factores.

- El análisis envolvente de datos es una *“técnica que permite identificar aquellas unidades de análisis que mejor desempeño tienen, permitiendo establecer un indicador global a partir del cual se evalúan las demás unidades”* (Schuschny et al, 2009).

El último paso en el proceso de construcción del indicador sintético lo constituye el método de agregación de las variables explicativas ponderadas.

Existen diversos procedimientos de agregación, siendo los más habituales la media aritmética ponderada y el promedio geométrico ponderado.

Una vez determinado el indicador sintético, la última etapa la constituye el análisis de la calidad del estimador a través del análisis de la varianza.

En el siguiente epígrafe se aplicará la metodología anterior para la construcción de un indicador sintético que permita medir la salud financiera de las Cajas Rurales.

5.3 CONSTRUCCIÓN DE UN INDICADOR SINTÉTICO PARA MEDIR LA SALUD FINANCIERA DE LAS CAJAS RURALES.

El objetivo de este capítulo es el de elaborar un indicador que condense en un único valor toda la información de un conjunto de variables relativas a diferentes aspectos económicos y financieros de las Cajas Rurales.

El marco conceptual que sirve de base para el indicador propuesto, plenamente desarrollado a lo largo de esta Tesis, busca determinar si existe relación entre la fusión de Cajas Rurales y la situación económica y financiera de estas entidades, todo ello a partir de una serie de modelos estadísticos (logit y regresión de Cox).

Los distintos modelos desarrollados en los capítulos anteriores han puesto de manifiesto que las variables TAMAÑO, NEGOCIO, EFICIENCIA, ROA, ESFUERZOPROV Y OF2,

construidas a partir de los estados financieros de las Cajas Rurales, muestran un elevado poder de discriminación entre el grupo de entidades operativas y fusionadas. A las anteriores habría que añadir las sectoriales OF1, EMP y FGD, así como las coyunturales INT1 y PARO.

Con el indicador sintético se pretende obtener puntuaciones a partir de la propia situación económica y financiera de la Caja Rural, por lo que se desecha el uso de variables coyunturales, así mismo tampoco es posible utilizar variables de tipo dicotómico.

Para la estimación del indicador sintético se utilizará la muestra completa de Cajas Rurales indicada en el Capítulo Tercero, es decir un total de 1.764 observaciones que incluyen 30 casos correspondientes a Cajas Rurales fusionadas.

Si en lugar de considerar la situación económica y financiera del año anterior a la fusión, se hubiera considerado la totalidad de observaciones disponibles para cada una de las Cajas Rurales fusionadas a lo largo de todo el periodo, el volumen de observaciones de este grupo de entidades se elevaría hasta las 485.

Las distintas variables están expresadas en diferentes unidades de medida, por ello es necesario normalizarlas, es decir, realizar un mapeo de un valor numérico a un nuevo valor, una vez que hayan sido eliminados los casos atípicos que pueden distorsionar los resultados.

Schuschny *et al* (2009) señala la existencia de diferentes procesos de normalización en función de la naturaleza de los datos, siendo el finalmente utilizado para la normalización de las variables relativas a las Cajas Rurales el método del reescalamiento, consistente en una transformación de los niveles de las variables para llevarlos al intervalo [0,100], de forma que la unidad de análisis de mayor desempeño tomará el valor 100, mientras que la variable de peor desempeño tomará el valor 0.

Tras la normalización de las variables, el siguiente paso ha consistido en el análisis de las variables. En este sentido en la Tabla 44 se confirma la existencia de coeficientes de correlación elevados, pero sin que lleguen a situarse en niveles en lo que pudiera existir un problema de multicolinealidad (fuerte correlación).

TABLA 44: COEFICIENTES DE CORRELACIÓN

	TAMAÑO	NEGOCIO	ESFUERZOPROV	EFICIENCIA	ROA	EMP	OF1	OF2
Tamaño	1	,476**	,257**	,302**	-,359**	,944**	,902**	-,265**
Negocio	,476**	1	,239**	,264**	-,283**	,400**	,346**	-,117**
EsfuerzoProv	,257**	,239**	1	,172**	-,317**	,207**	,196**	-,195**
Eficiencia	,302**	,264**	,172**	1	-,490**	,372**	,390**	-,542**
ROA	-,359**	-,283**	-,317**	-,490**	1	-,291**	-,289**	,525**
EMP	,944**	,400**	,207**	,372**	-,291**	1	,973**	-,406**
OF1	,902**	,346**	,196**	,390**	-,289**	,973**	1	-,514**
OF2	-,265**	-,117**	-,195**	-,542**	,525**	-,406**	-,514**	1

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

FUENTE: ELABORACIÓN PROPIA

Dado que el objetivo de este capítulo no es otro que la construcción de un indicador sintético que sea capaz de discriminar entre las diferentes Cajas Rurales en función de su situación económica y financiera, es necesario que los indicadores individuales considerados tengan ese poder de discriminación entre los dos grupos definidos por la variable dependiente definida en el capítulo tercero.

En este sentido, en la Tabla 45 se han recogido los resultados de la prueba *t* de igualdad de medidas, observándose que todas las variables son significativas a un nivel del 0,1, rechazándose la hipótesis nula de que las medias poblacionales son idénticas en ambos grupos.

TABLA 45: PRUEBA T DE IGUALDAD DE MEDIAS

	t	SIG.
Tamaño	-3,518	,000
Negocio	-2,795	,005
EsfuerzoProv	-1,734	,083
Eficiencia	-5,658	,000
ROA	4,570	,000
EMP	-3,148	,002
OF1	-3,697	,000
OF2	4,235	,000

FUENTE: ELABORACIÓN PROPIA

Una vez determinadas las variables a utilizar, el siguiente paso consiste en el cálculo de los pesos de ponderación. Para ello existen diferentes técnicas, si bien una de las más utilizadas es el análisis factorial a través del método de los componentes principales.

Los componentes principales son una combinación lineal de las variables observadas que mantienen la propiedad de ser variables aleatorias incorrelacionadas y cuyas varianzas son respectivamente máximas (Cabrero *et al*, 1996).

En la Tabla 46 se recogen los resultados del proceso de reducción de dimensiones, observándose tres factores que explican conjuntamente el 80,7% de la varianza total.

TABLA 46: VARIANZA TOTAL EXPLICADA

COMPONENTE	AUTOVALORES INICIALES			SUMAS DE EXTRACCIÓN DE CARGAS AL CUADRADO		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,120	51,502	51,502	4,120	51,502	51,502
2	1,315	16,434	67,936	1,315	16,434	67,936
3	1,021	12,768	80,705	1,021	12,768	80,705
4	,719	8,982	89,686			
5	,431	5,383	95,069			
6	,352	4,403	99,472			
7	,027	,339	99,811			
8	,015	,189	100,000			

FUENTE: ELABORACIÓN PROPIA

En la Tabla 47 se han recogido los resultados de la prueba KMO y de esfericidad de Bartlett (PBE), observándose el cumplimiento de los supuestos básicos para la aplicación del análisis factorial (KMO superior a 0,5 y PEB con significatividad inferior a 0,05).

TABLA 47: PRUEBA KMO Y BARTLETT

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,673
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	14457,489
	gl	28
	Sig.	,000

FUENTE: ELABORACIÓN PROPIA

Cabrero *et al* (1996) afirma que en la práctica el primer componente captura suficiente variación para ser una adecuada representación del conjunto inicial y, por tanto, suficiente para ser utilizado en la construcción del índice sintético, criterio que se utilizará para la construcción del indicador sintético de salud financiera de las Cajas Rurales, por ello serán utilizados exclusivamente los coeficientes de ponderación definidos por el primer factor.

Para la obtención de los coeficientes de ponderación se recurre a la matriz de puntuaciones reflejada en la Tabla 48, la cual representa la combinación lineal óptima de cada uno de los componentes para el conjunto de variables explicativas.

TABLA 48: MATRIZ DE COEFICIENTE DE PUNTUACIÓN DE COMPONENTE

	COMPONENTE		
	1	2	3
Tamaño	,211	,327	,026
Negocio	,130	,108	,540
Eficiencia	,157	-,391	-,201
ROA	-,154	,425	-,164
EMP	,217	,300	-,147
OF1	,218	,250	-,232
OF2	-,162	,363	,352
EsfuerzoProv	,109	-,157	,648

Método de extracción: análisis de componentes principales.

FUENTE: ELABORACIÓN PROPIA

Una vez determinados los factores y los pesos de ponderación, la construcción del índice sintético se realizará agregando el conjunto de variables explicativas seleccionadas y ponderadas en función de las puntuaciones y la varianza explicada por el primer componente (Estévez, 2002).

Existen diferentes métodos de agregación que pueden ser seguidos para la construcción del índice sintético, si bien Schuschny *et al* (2009) indicaba que la media aritmética ponderada y el promedio geométrico ponderado son los más utilizados. De hecho, Cabrero *et al* (1996) propone el uso de la media aritmética ponderada por su sencillez e interpretación.

Para la elaboración del indicador sintético de salud financiera de las Cajas Rurales se opta por la utilización del procedimiento de la media aritmética ponderada por la sencillez de interpretación puesta de manifiesto por Cabero *et al* (1996), por lo que formalmente se podría definir al indicador de salud financiera de las Cajas Rurales de la siguiente forma:

$$ISF_{CR} = V_1a_1 + V_2a_2 + V_3a_3 + V_4a_4 + V_5a_5 + V_6a_6 + V_7a_7 + V_8a_8$$

Donde ISF_{CR} denota al indicador de salud financiera para una Caja Rural, V_i denota a las variables explicativas seleccionadas, a_i denota el coeficiente de ponderación de la variable i dentro del primer componente.

Aplicando los resultados obtenidos del análisis factorial, la definición del indicador sintético de salud financiera calculado se define de la siguiente forma:

$$ISF_{CR} = 0,130 \cdot NEGOCIO + 0,211 \cdot TAMAÑO - 0,154 \cdot ROA + 0,157 \cdot EFICIENCIA \\ + 0,109 \cdot ESFUERZOPROV + 0,218 \cdot OF1 - 0,162 \cdot OF2 + 0,217 \cdot EMP$$

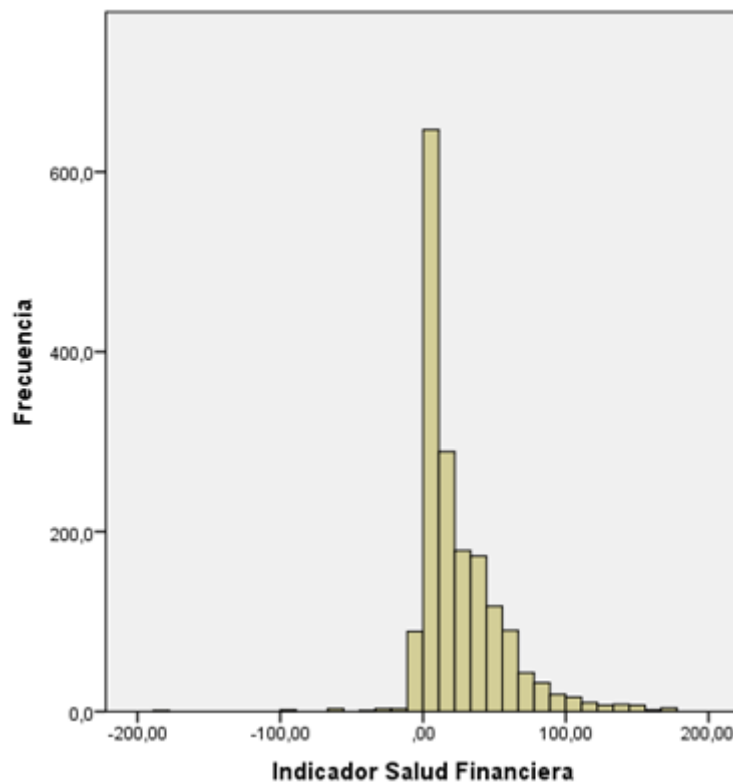
Una vez determinado el indicador sintético de salud financiera de una Caja Rural, en el epígrafe siguiente se procederá a realizar un análisis del mismo.

5.4 EXPOSICIÓN DE RESULTADOS.

Una vez expuesta la definición formal del indicador sintético, a continuación se aplicará el mismo sobre la muestra de Cajas Rurales utilizada para su determinación con el fin de observar el poder de discriminación del mismo entre el conjunto de Cajas Rurales operativas y fusionadas.

En la Ilustración 39 se ha reflejado la distribución de las puntuaciones otorgadas por el ISF_{CR} para cada una de las Cajas Rurales, observándose que la mayor parte de las puntuaciones se concentran en el intervalo inferior a los 50 puntos. Cabe tener en cuenta que cuanto mayor sea la puntuación del ISF_{CR}, mejor será la salud financiera de la Caja Rural.

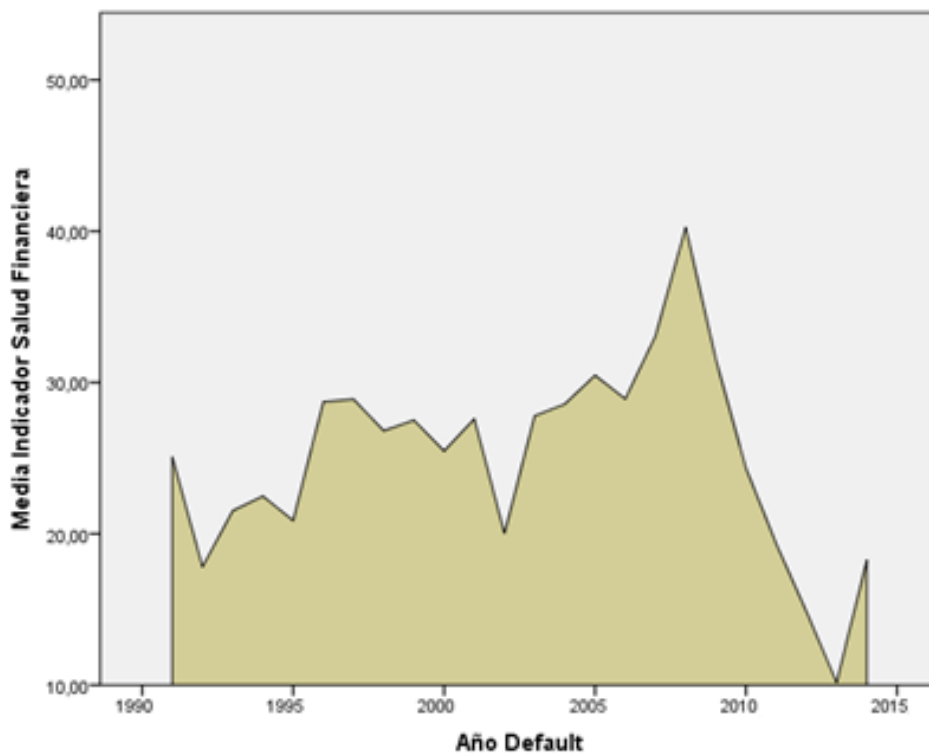
ILUSTRACIÓN 39: REPRESENTACIÓN GRÁFICA DE LAS PUNTUACIONES OBTENIDAS DEL INDICADOR SINTÉTICO



FUENTE: ELABORACIÓN PROPIA

En lo que respecta a la distribución temporal de las puntuaciones del ISF_{CR}, en la Ilustración 40 se ha recogido las puntuaciones medias otorgadas por el indicador sintético para cada uno de los años objetos de estudio. Se puede observar una tendencia alcista en la puntuación media del conjunto de Cajas Rurales, con un claro retroceso a partir del año 2009 previsiblemente por los efectos negativos de la crisis económica sobre este conjunto de entidades, con un ISF_{CR} que en término medio llega a situarse en niveles inferiores a los presentados en el año 1991.

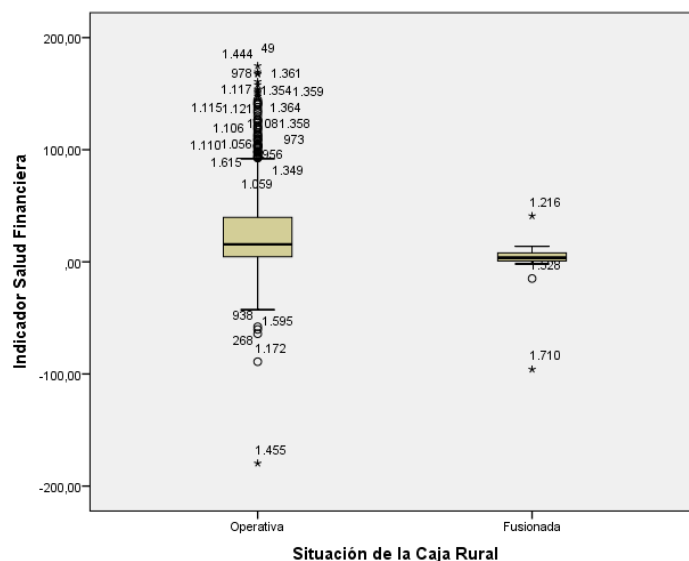
ILUSTRACIÓN 40: DISTRIBUCIÓN ANUAL DE LA PUNTUACIÓN MEDIA DEL INDICADOR SINTÉTICO.



FUENTE: ELABORACIÓN PROPIA

En función de las Cajas Rurales operativas y fusionadas, en la Ilustración 41 se puede observar una clara diferencia entre ambos grupos, encontrándose las puntuaciones más altas dentro del conjunto de Cajas Rurales que se mantuvieron operativas, si bien en la mayor parte de los casos estas puntuaciones oscilan por debajo del nivel de los 100 puntos.

ILUSTRACIÓN 41: GRÁFICO DE CARGAS PUNTUACIONES INDICADOR SINTÉTICO



FUENTE: ELABORACIÓN PROPIA

En la Tabla 49 se han recogido los resultados de la prueba ANOVA y los principales estadísticos asociados al indicador sintético. La Prueba ANOVA rechaza la hipótesis nula de que las medias poblacionales son idénticas en ambos grupos.

En lo que respecta a los principales estadísticos, en la Tabla 49 se pone de manifiesto que el grupo de Cajas Rurales operativas muestran una puntuación media del ISF_{CR} de 25,6 puntos, superior a la puntuación media de 1,8 puntos reflejada por el conjunto de Cajas Rurales fusionadas.

TABLA 49: ESTADÍSTICOS DESCRIPTIVOS ASOCIADOS AL INDICADOR SINTÉTICO

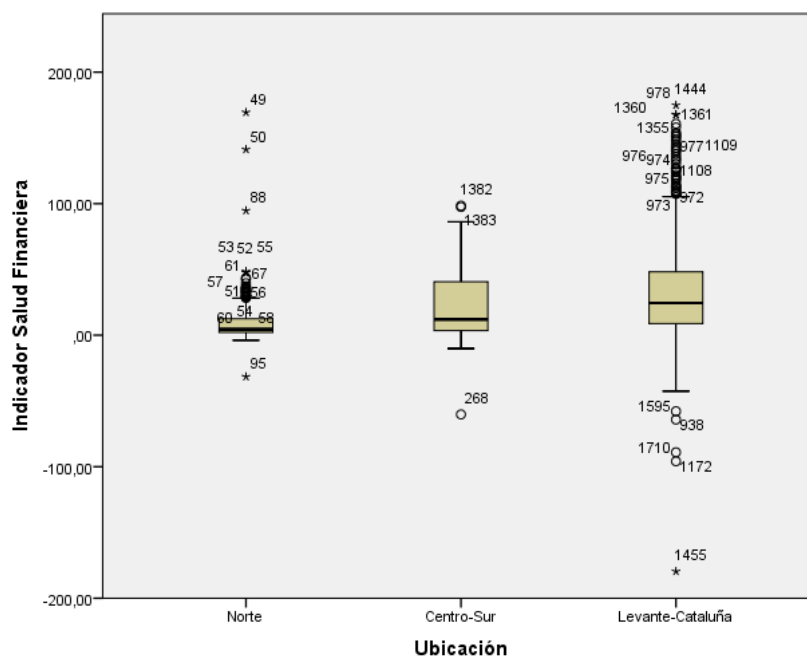
SITUACIÓN DE LA CAJA RURAL		ESTADÍSTICO
Operativa	Media	25,6515
	Mínimo	-179,63
	Máximo	174,85
Fusionada	Media	1,8268
	Mínimo	-95,9
	Máximo	40,97
Prueba t		
Entre grupos		t = 4,182 Sig. = 0,00000

FUENTE: ELABORACIÓN PROPIA

Una vez confirmado que el poder de discriminación del ISF_{CR} propuesto es aceptable, con poder de discriminación entre las Cajas Rurales operativas y fusionadas, a continuación se plantea observar si existen diferencias en la puntuación del indicador sintético en función de la ubicación geográfica de la Caja Rural.

Para ello, en la Ilustración 42 se han representado las puntuaciones del ISF_{CR} para cada una de las Cajas Rurales consideradas, diferenciadas en función de la ubicación geográfica de cada una de ellas. Se puede observar que el ISF_{CR} otorga una puntuación inferior al conjunto de Cajas Rurales ubicadas en el norte (con la excepción de tres entidades que presentarían puntuaciones superiores a la media de su Grupo), mientras que el conjunto de Cajas Rurales ubicadas en el centro y el levante mantienen puntuaciones más homogéneas, si bien este último conjunto de entidades presentaría un mayor grado de dispersión con Cajas Rurales que presentarían puntuaciones por encima de la puntuación media del grupo al que pertenecen.

ILUSTRACIÓN 42: RESULTADOS DEL INDICADOR SINTÉTICO EN FUNCIÓN DE LA UBICACIÓN GEOGRÁFICA



FUENTE: ELABORACIÓN PROPIA

Aunque el método gráfico confirma la existencia de diferencias en la puntuación del ISF_{CR} en función de la ubicación geográfica, este análisis ha sido complementado mediante una prueba ANOVA que indica la existencia de diferencias entre los tres grupos (norte, centro-sur y levante), prueba que posteriormente se ha repetido para confirmar diferencias dos a dos, resultados que han sido recogidos en la Tabla 50 siguiente.

TABLA 50: MEDIA DEL ISF Y PRUEBA T EN FUNCIÓN DE LA UBICACIÓN GEOGRÁFICA

ISF _{CR}		PRUEBA ANOVA
Ubicación	Media	Sig.*
Norte	9,9061	0,000
Centro-Sur	21,8887	
Levante-Cataluña	32,7584	

* Realizada par a par entre los tres grupos

FUENTE: ELABORACIÓN PROPIA

5.5 CONCLUSIONES DEL CAPÍTULO

A lo largo de este capítulo se ha elaborado un indicador sintético como agregación de una serie de variables explicativas que tienen en cuenta diferentes aspectos de la situación económica y financiera de las Cajas Rurales. Para ello, se han seleccionado aquellas variables explicativas que resultaron ser significativas en los modelos logit y supervivencia estimados en el capítulo anterior por mostrar un alto poder de discriminación entre el conjunto de Cajas Rurales operativas y fusionadas.

Tras la normalización de la muestra, para evitar posibles errores en los resultados del indicador, y el análisis de las correlaciones y prueba ANOVA entre las variables iniciales, se han seleccionado finalmente TAMAÑO, NEGOCIO, EFICIENCIA, ESFUERZOPROV, ROA y OF2 para formar parte del indicador sintético. En la construcción de un indicador sintético juega

un papel fundamental el establecimiento de los pesos de ponderación de cada variable explicativa dentro del indicador, así como el método de agregación seleccionado.

Tras analizar la bibliografía existente sobre el tema, se observa un uso mayoritario de la técnica de los componentes principales para el cálculo de los pesos de ponderación, y de la media aritmética ponderada, como método de agregación, de ahí que finalmente fuesen ambas técnicas las seleccionadas para la construcción del indicador sintético que mida la salud financiera de las Cajas Rurales.

La aplicación del indicador sintético a la muestra de Cajas Rurales pone de manifiesto la presencia de un elevado poder de discriminación entre el conjunto de Cajas Rurales operativas y fusionadas, hasta el punto de asignar puntuaciones medias superiores al conjunto de entidades operativas, si bien existen diferencias entre territorios, hasta el punto de que las Cajas Rurales ubicadas en el norte presentaron menores puntuaciones que sus homólogas ubicadas en el centro-sur y levante.

CAPÍTULO SEXTO:

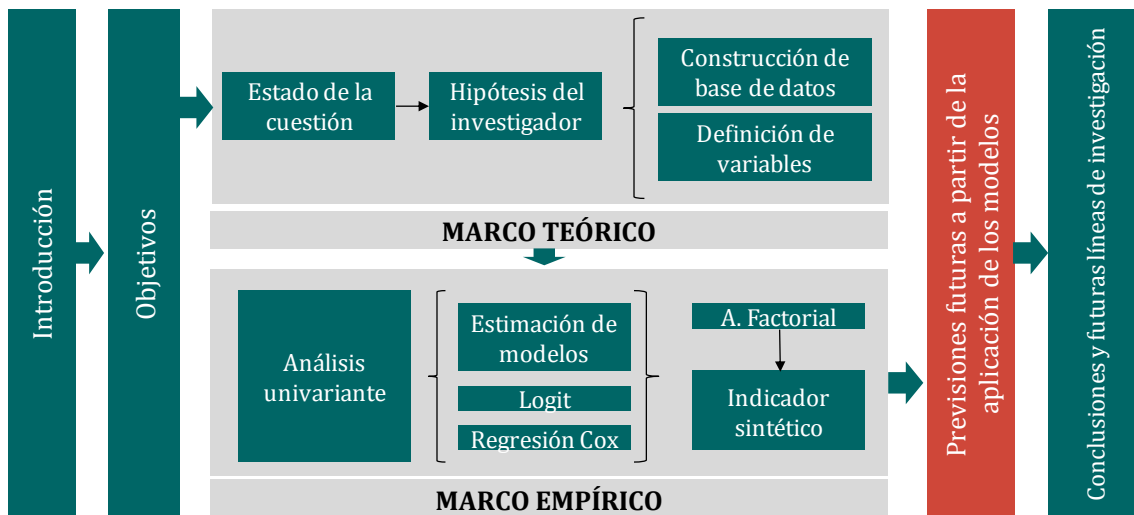
PREVISIONES FUTURAS SOBRE EL SECTOR DE CAJAS RURALES A PARTIR DE LOS MODELOS ESTIMADOS

6.1 SINOPSIS DEL CAPÍTULO

A partir de los modelos estimados en esta Tesis, a lo largo de este capítulo se realiza una aplicación práctica para conocer la situación de las Cajas Rurales que se encuentran operativas en la actualidad.

Para facilitar el seguimiento de la Tesis, en la Ilustración 43 se recoge la ubicación del capítulo dentro de la sinopsis de la Tesis.

ILUSTRACIÓN 43: UBICACIÓN DEL CAPÍTULO SEXTO DENTRO DE LA SINOPSIS DE LA TESIS



FUENTE: ELABORACIÓN PROPIA

6.2 PREVISIONES FUTURAS

Como se ha puesto de manifiesto en capítulos anteriores, el modelo logit multiperiodo presenta el mayor poder de discriminación entre Cajas Rurales operativas y fusionadas, con un porcentaje de aciertos en la clasificación de las entidades de la muestra de validación superior al mostrado por el resto de modelos.

Así mismo se ha observado una pauta similar de comportamiento en las puntuaciones otorgadas por el indicador sintético construido a partir de variables que miden la situación económica y financiera, mostrando poder de discriminación entre el conjunto de Cajas Rurales operativas y fusionadas.

Se plantea ahora la posibilidad de aplicar dichos modelos a las Cajas Rurales que se encuentran operativas en la actualidad, para conocer su salud financiera además de predecir la probabilidad de que cada una de ellas formen parte de futuros procesos de integración dentro del sector.

Para ello se dispone de una muestra de Cajas Rurales operativas a enero de 2016, con datos económicos, financieros, sectoriales y coyunturales del ejercicio anterior (diciembre de 2015).

Con dicha muestra se han estimado el modelo logit multiperiodo, siendo la expresión formal del mismo la recogida en EQ1:

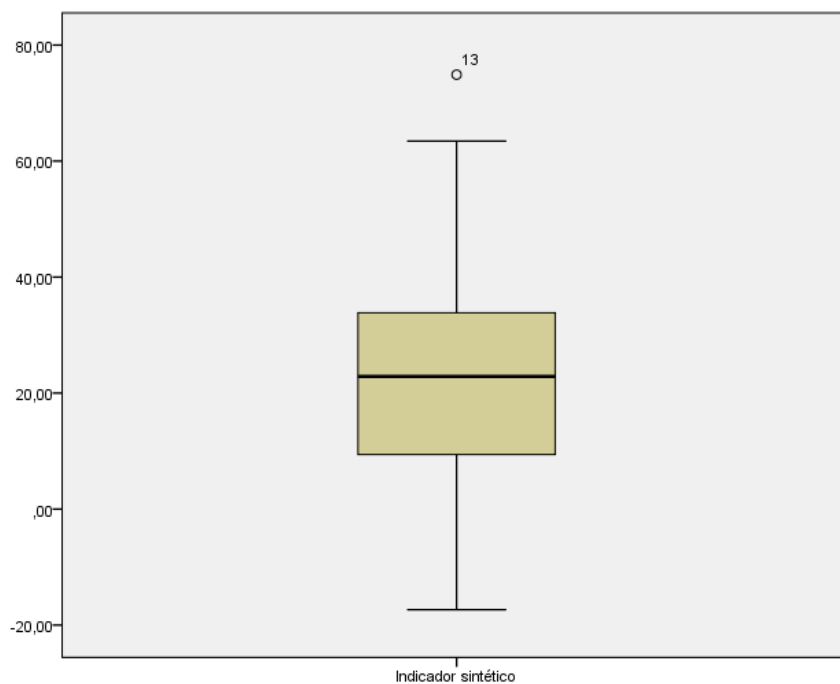
$$P(\text{Estado} = \text{Fusionada}) = \frac{1}{1 + e^{-(-6,85 + 4,74 \text{NEGOCIO} - 145,02 \text{ROA} + 3,09 \text{UB}(1) + 1,21 \text{UB}(2) - 2,6 \text{FGD} + 12,83 \text{PARO})}}$$

Igualmente se ha procedido a calcular el indicador sintético para la muestra de Cajas Rurales operativas a enero de 2016, siendo la definición formal del indicador la recogida en la EQ2:

$$\begin{aligned} ISF_{CR} = & 0,130 \cdot \text{NEGOCIO} + 0,211 \cdot \text{TAMAÑO} - 0,154 \cdot \text{ROA} + 0,157 \cdot \text{EFICIENCIA} \\ & + 0,109 \cdot \text{ESFUERZOPROV} + 0,218 \cdot \text{OF1} - 0,162 \cdot \text{OF2} + 0,217 \cdot \text{EMP} \end{aligned}$$

En lo que respecta a la medida de salud financiera de las Cajas Rurales que ofrece el indicador sintético, en la Ilustración 44 se recoge el gráfico de cajas de las puntuaciones obtenidos para cada una de ellas a partir de los datos económicos y financieros de 2015, observándose que en su mayor parte las puntuaciones obtenidas no superan los 30 puntos, con la excepción de una entidad que se aproximaría hasta los 80 puntos.

ILUSTRACIÓN 44: GRÁFICO DE CAJAS DE LAS PUNTUACIONES OTORGADAS POR EL INDICADOR SINTÉTICO A LAS CAJAS RURALES OPERATIVAS EN LA ACTUALIDAD.



FUENTE: ELABORACIÓN PROPIA

Considerando los valores máximos y mínimos mostrados por el indicador sintético para el conjunto de Cajas Rurales que fueron utilizados para su construcción en el capítulo anterior, en la Tabla 51 se confirma que las Cajas Rurales que en la actualidad se encuentran operativas presentan una puntuación media en el entorno de los 21 puntos, por debajo de los 25,6 puntos que en término medio presentaron las Cajas Rurales en años anteriores (Tabla 49 anterior).

TABLA 51: PUNTUACIONES MÁXIMAS Y MÍNIMAS OTORGADAS POR EL INDICADOR SINTÉTICO PARA LAS CAJAS RURALES OPERATIVAS EN EL AÑO 2016.

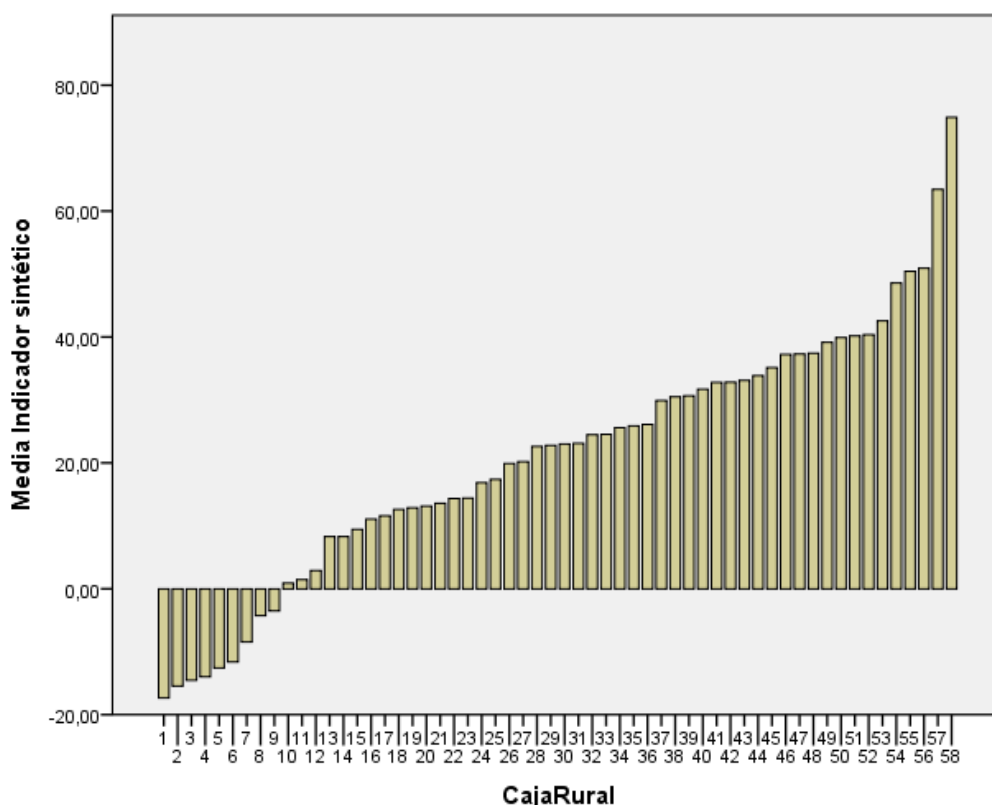
SITUACIÓN DE LAS CAJA RURALES EN 2016	
Media	21,00
Mínimo	-17,33
Máximo	74,90

FUENTE: ELABORACIÓN PROPIA

En la Ilustración 45 se recogen las puntuaciones otorgadas por el indicador sintético para cada una de las Cajas Rurales, ordenadas de menor a mayor. En dicha Ilustración se puede

observar que por regla general las Cajas Rurales operativas en la actualidad presentan puntuaciones superiores al valor medio de 1,82 puntos que presentaron las Cajas Rurales fusionadas con anterioridad, excepto diez Cajas Rurales que presentarían puntuaciones inferiores.

ILUSTRACIÓN 45: PUNTUACIONES OTORGADAS POR EL INDICADOR SINTÉTICO PARA CADA CAJA RURAL

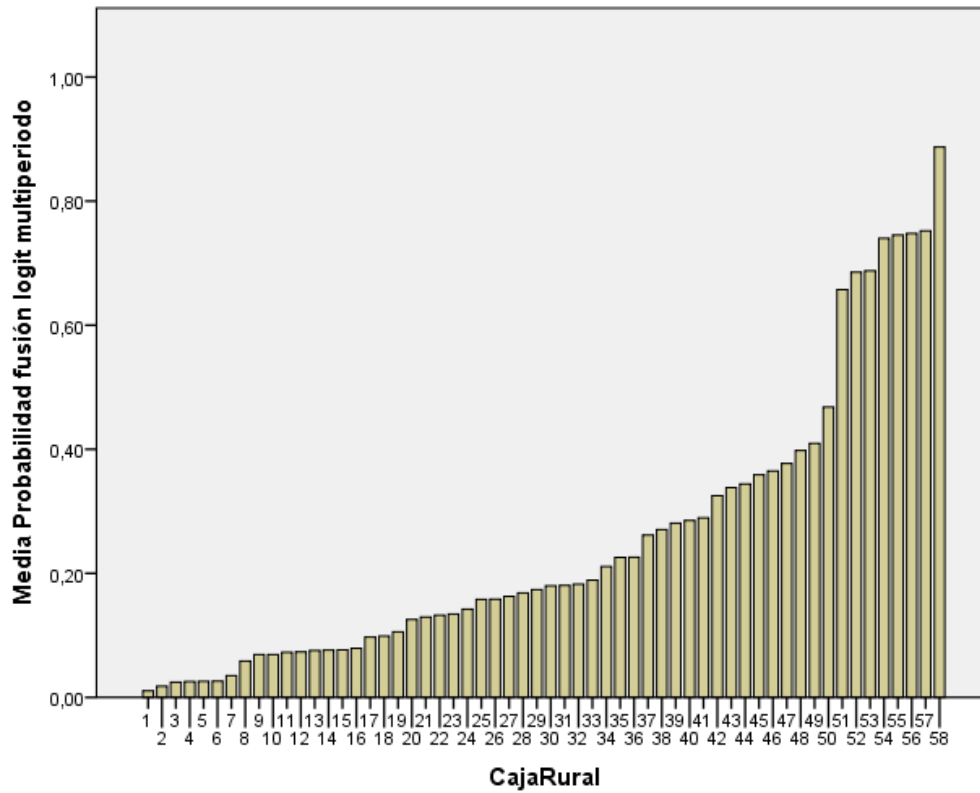


FUENTE: ELABORACIÓN PROPIA

En lo que respecta a los resultados de la aplicación del modelo logit multiperiodo, en la Ilustración 46 se han representado gráficamente y en orden ascendente la probabilidad de insolvencia calculada para cada una las Cajas Rurales.

Se puede observar que por regla general el modelo atribuye una reducida probabilidad de insolvencia para las Cajas Rurales que se encuentran operativas en la actualidad, si bien pone de manifiesto un total de ocho entidades que presentarían probabilidades que superarían el 50%.

ILUSTRACIÓN 46: DISTRIBUCIÓN DE LA PROBABILIDAD DE INSOLVENCIA ESTIMADA PARA LAS CAJAS RURALES OPERATIVAS EN LA ACTUALIDAD.

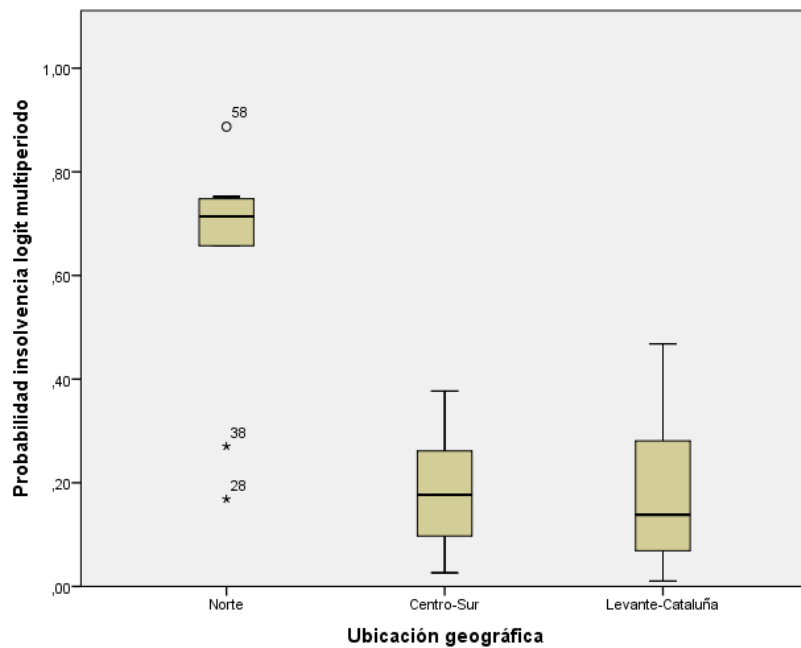


FUENTE: ELABORACIÓN PROPIA

En un análisis pormenorizado de las probabilidades calculadas por el modelo logit, y diferenciando en función de los territorios de las distintas Cajas Rurales, en la Ilustración 47 se puede observar que el modelo pone de manifiesto que las Cajas Rurales ubicadas en el norte de España presentan una mayor probabilidad de fusión que las ubicadas en el resto de regiones, si bien existirían dos casos de Cajas Rurales ubicadas en el norte con probabilidades inferiores.

Por su parte, las Cajas Rurales ubicadas en el centro-sur y levante presentarían en términos generales probabilidades de fusión inferiores, si bien las del levante serían algo superiores a las mostradas por sus homólogos del centro y sur peninsular.

ILUSTRACIÓN 47: PROBABILIDAD DE FUSIÓN PARA LA MUESTRA DE CAJAS RURALES OPERATIVAS EN LA ACTUALIDAD



FUENTE: ELABORACIÓN PROPIA

En definitiva, ambas técnicas ponen de manifiesto la fortaleza económica y financiera del sector de Cajas Rurales español, con un conjunto de entidades que estarían operando con una situación económica y financiera saneada.

Sin embargo, se observa la existencia de un reducido número de entidades que estarían experimentando ciertas dificultades que se encontrarían en sintonía con las ya presentadas con anterioridad por otras Cajas Rurales que finalmente formaron parte de procesos de fusión que, en ocasiones, implicó su desaparición.

Por todo ello, los modelos brindan a los equipos gestores de estas entidades la posibilidad de actuación sobre las mismas, por la vía de ciertas magnitudes de sus estados financieros, con el fin de mejorar la situación actual para evitar formar parte de procesos de integración dentro del sector que no sean necesarios por motivos diferentes a los de la propia situación económica y financiera intrínseca de la Caja Rural.

CAPÍTULO SÉPTIMO:

CONCLUSIONES

En el capítulo introductorio de esta Tesis se ha puesto de manifiesto que en los últimos años el sector financiero español se había mostrado muy activo en términos de fusiones de entidades financieras, principalmente por la bancarización de las cajas de ahorro, en parte obligadas por el supervisor dada su delicada situación económica y financiera.

Aunque a priori el sector de Cajas Rurales se ha mantenido ajeno a estos procesos de integración, un análisis pormenorizado del mismo pone de manifiesto la existencia de un proceso de concentración de estas cooperativas de crédito que busca la reestructuración del sector mediante fusiones con otras Cajas Rurales, acontecimiento que para autores como Porath (2006), Cabo *et al* (2010) y Lima (2012) están justificados por motivos de sostenibilidad financiera de las entidades explicadas.

El primer objetivo planteado en esta Tesis es el de determinar si los procesos de integración de Cajas Rurales acaecidos entre 1991 y 2015 pueden ser explicados por motivos de sostenibilidad de las entidades implicadas, corroborándose esta situación en la práctica totalidad de las Cajas Rurales fusionadas sobre las que se ha dispuesto de información económica y financiera suficiente para ello, cumpliéndose así la primera hipótesis planteada en esta Tesis.

En este sentido, se ha estimado un modelo logit y otro de supervivencia (regresión de Cox) a partir de una muestra de Cajas Rurales sobre la que se ha dispuesto de los estados financieros publicados anualmente en las Memorias de la UNACC, y con los que se ha podido construir un panel de datos desde 1990 hasta 2015.

El modelo logit es un modelo probabilístico de respuesta binaria que intenta explicar el comportamiento de una variable dependiente (de tipo dicotómico) a partir de una serie de variables explicativas.

En el caso de las Cajas Rurales, la variable dependiente ha tomado el valor 1 si la Caja Rural había sido dada de baja del Registro de Entidades Financieras de Banco de España y además se encontraba en el peor cuartil de una serie de ratios financieras; mientras que el valor 0 se ha asignado a aquellas Cajas Rurales que se encontraban operativas a cierre de cada uno de los ejercicios.

De la bibliografía académica analizada en el capítulo segundo no se desprende un consenso mayoritario entre el conjunto de investigadores en relación a la composición de la muestra a utilizar para la estimación del modelo logit, ya que algunos autores proponen el uso de muestras emparejadas (igual número de individuos que sobreviven o mueren) para cada uno de los años de estudio; si bien otros autores coinciden en usar la totalidad de la muestra de individuos, pero considerando plazos más cortos.

Así mismo también existen diferentes técnicas de selección de las variables que forman parte de los modelos, y aunque la mayor parte de los autores coinciden en utilizar los procedimientos tradicionales consistentes en la introducción unitaria y por pasos de las variables con el fin de observar la mejora en el poder de explicación del modelo, existen autores que proponen técnicas alternativas por mejorar la calidad de las predicciones, bien sean a través del método de componentes principales, o bien mediante procedimientos metaheurísticos basados en inteligencia artificial, siendo los finalmente utilizados en esta tesis aquellos basados en la selección tradicional y el método de componentes principales.

Adicionalmente, se ha aplicado una nueva versión del modelo logit para la inclusión de variables que varían a lo largo del tiempo, propuesta por Shumway (2001), y que aún no habría sido aplicada a esta tipología de entidades dentro de la investigación académica encontrada en España.

En la **Tabla 52** se puede observar que los diferentes modelos logit estimados presentan un alto poder de clasificación en la muestra de validación, principalmente mediante la selección de variables por los procedimientos tradicionales.

Entre todos los modelos, el modelo logit multiperiodo es el que presenta el mejor grado de clasificación en la muestra de validación, identificando al 90% de las entidades fusionadas, lo que implica un Error Tipo II (clasificar incorrectamente a una entidad fusionada dentro del grupo de las entidades operativas) del 10%, siendo esta versión la más indicada para la predicción de los procesos de integración.

TABLA 52: CLASIFICACIÓN CORRECTA DE LOS DIFERENTES MODELOS LOGIT

Selección tradicional	MUESTRA INICIAL			MUESTRA DE VALIDACIÓN		
	Operativa	Fusionada	Global	Operativa	Fusionada	Global
Logit (2000-2002)	97%	44%	91%	31%	84%	45%
Logit (2011-2014)	96%	11%	74%	58%	77%	60%
Logit (1991-2014)	80%	85%	82%	73%	70%	72%
Logit multiperiodo	98%	35%	96%	67%	90%	80%
Selección componentes principales	Operativa	Fusionada	Global	Operativa	Fusionada	Gobal
Logit (1.991-2.014)	80%	80%	80%	94%	10%	52%
Logit multiperiodo	98%	35%	97%	96%	40%	68%

FUENTE: ELABORACIÓN PROPIA

Todas las versiones de los modelos logit estimados han puesto de manifiesto la importancia de las variables económicas y financieras sobre la probabilidad de fusión de una Caja Rural, si bien las versiones que incorporan la muestra completa encuentran significativas las variables que miden aspectos coyunturales, principalmente la evolución de los tipos de interés y la tasa de paro, cumpliéndose así la tercera hipótesis planteada.

Las variables relacionadas con la dimensión, la calidad del negocio y la rentabilidad resultan ser las más significativas en el estudio de los procesos de fusión de Cajas Rurales,

lo que supone un cumplimiento parcial de la segunda hipótesis, ya que a diferencia de lo mostrado en otras investigaciones sobre cooperativas de crédito en un ámbito internacional, la variable capital no ha resultado ser significativa en la determinación de la sostenibilidad financiera de las Cajas Rurales españolas.

Se observa que las Cajas Rurales ubicadas en el norte peninsular muestran mayor sensibilidad a formar parte de un proceso de fusión que sus homólogas ubicadas en el resto de territorios, incumpléndose la cuarta hipótesis de esta Tesis, y aunque si bien es cierto que se esperaba que la ubicación fuese un factor determinante, la hipótesis inicial consideraba que las Cajas Rurales del sur y levante presentarían mayor riesgo de desaparecer.

Más concretamente, las curvas de supervivencia de Kaplan-Meier ponen de manifiesto que las funciones de supervivencia de las entidades del norte son más pronunciadas que las curvas presentadas por las entidades ubicadas en el resto de territorios.

Para tratar de determinar la influencia de las distintas variables explicativas consideradas sobre la supervivencia de las Cajas Rurales, se ha recurrido al modelo de riesgos proporcionales de Cox en su versión adaptada para la inclusión de variables que varían con el tiempo, ya que se trata de ratios financieras que por su propia naturaleza varían de forma anual, dado que la utilización de la versión tradicional del modelo no es posible por el incumplimiento del supuesto de riesgos proporcionales.

Este modelo de Cox confirma lo ya adelantado por los modelos logit anteriores, es decir, la influencia que las variables que miden la situación económica y financiera de la Caja Rural tienen sobre la probabilidad de supervivencia de estas entidades, principalmente aquel grupo de variables relacionadas con la dimensión, rentabilidad y ubicación de la Caja Rural, si bien es cierto que por sus propias restricciones formales no permite realizar predicciones a futuro.

Dada la relevancia de la situación económica y financiera puesta de manifiesto por los diferentes modelos estimados, se plantea la posibilidad de construir un indicador sintético que permitiera medir la salud financiera de las Cajas Rurales.

Para ello se ha dispuesto de la misma muestra utilizada para la estimación de los modelos logit y supervivencia, si bien se seleccionaron aquellas variables que resultan ser significativas en la estimación de dichos modelos con el fin de evitar la parsimonia y utilizar los indicadores que muestran el mayor poder de explicación, dado que la calidad de un indicador sintético recae en la capacidad de los indicadores individuales utilizados para su agregación.

Aunque el uso de indicadores sintéticos es habitual en economía para la medición y predicción de aspectos coyunturales, su uso en la predicción de situaciones económicas y financieras deficientes está poco extendido, no habiendo sido localizadas investigaciones referentes al ámbito de las entidades financieras.

Un aspecto fundamental en la construcción de un indicador sintético es la metodología a utilizar para la determinación de los pesos de ponderación que las variables explicativas tendrán dentro del indicador sintético, así como el procedimiento de agregación.

En esta Tesis se ha utilizado el criterio de los componentes principales, ampliamente utilizado en investigaciones sobre indicadores sintéticos, para la determinación de los coeficientes de ponderación, siendo el método de agregación propuesto la media aritmética ponderada, por su sencillez de uso y claridad de comprensión en sus resultados.

Tras la normalización de la muestra y el análisis de la significatividad individual de las variables inicialmente seleccionadas, se ha construido un indicador sintético de salud financiera de las Cajas Rurales que muestra un elevado poder de discriminación, asignando las mayores puntuaciones a aquellas entidades que se mantienen operativas a

lo largo del periodo de estudio, y las menores al grupo de Cajas Rurales fusionadas, cumpliéndose así la quinta hipótesis establecida al inicio de esta Tesis.

El indicador sintético pone de manifiesto marcadas diferencias entre las Cajas Rurales ubicadas en las diferentes regiones, señalando una vez más las peores puntuaciones a las Cajas Rurales ubicadas en el norte, frente a las Cajas Rurales del centro-sur y levante que mantendrían puntuaciones más o menos homogéneas.

Finalmente se han utilizado las técnicas anteriores para tratar de predecir el comportamiento actual del sector, observándose cierta sintonía en los resultados mostrados por el indicador sintético y el modelo logit multiperiodo (seleccionado por su mayor poder de clasificación en la muestra de validación), ejercicio que ha puesto de manifiesto la existencia de una serie de Cajas Rurales que presentarían una mayor probabilidad de desaparición, principalmente en el norte de España.

Por regla general el sector de las Cajas Rurales se encuentra en buena situación, con un conjunto de entidades que mostraría baja probabilidad de fusión y altas puntuaciones en el indicador sintético de salud financiera, si bien es cierta la existencia de algunas de ellas que previsiblemente se verán inmersas en procesos de concentración en un futuro próximo con el fin de preservar la sostenibilidad financiera de las entidades implicadas.

En definitiva, esta Tesis pone de manifiesto que los procesos de concentración de Cajas Rurales en España pueden ser explicados por motivos de sostenibilidad financiera de las entidades implicadas, ya que los modelos logit y supervivencia utilizados reflejan la debilidad de las Cajas Rurales que cesaron actividad (por fusión con otras Cajas Rurales) en aspectos clave como la dimensión, la calidad del negocio y la rentabilidad, siendo el modelo logit multiperiodo el que ha ofrecido los mejores resultados, anticipando estas situaciones hasta tres años antes de su desaparición, proponiendo un indicador sintético que permite futuras actuaciones sobre estas entidades para evitar su desaparición.

BIBLIOGRAFÍA

- Abad, J., y Gutiérrez, C. (2014). Evaluación de la solvencia bancaria: un modelo basado en las pruebas de resistencia de la banca española. *Estudios De Economía Aplicada*, 32(2), 593-616.
- Abrams, B. A., y Huang, C. J. (1987). Predicting bank failures: the role of structure in affecting. *Applied Economics*, 19(10), 1291-1302.
- Acosta, E., y Fernández, F. (2007). Predicción del fracaso empresarial mediante el uso de algoritmos genéticos. *Comunicación presentada en el X Encuentro de Economía Aplicada*. Logroño, España.
- Ahumada, A., y Budnevich, C. (1999). Indicadores Financieros y Clasificación de los Bancos: Un análisis estadístico multivariado. Banco Central de Chile, Gerencia de Análisis Financiero, División de Estudios, Santiago de Chile. http://www.cemla.org/old/pdf/red/CH_ahumada_budnevich.pdf.
- Ahumada, A., y Budnevich, C. (2001). Some measures of financial fragility in the chilean banking system: an early warning indicators application. *Banco Central De Chile Working Papers*, 117.
- Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal Of Finance*, 23(4), 589.
- Altman, E., Haldeman, R., y Narayanan, P. (1977). ZETA analysis A new model to identify bankruptcy risk of corporations. *Journal Of Banking y Finance*, 1(1), 29-54.
- Álvarez, J.A. (2008). La banca española ante la actual crisis financiera. *Estabilidad financiera*. 15, 21-38
- Anastasi, A., Burdisso, T., Grubisic, E., y Lencioni, S. (1998). ¿Es posible anticipar problemas en una entidad financiera? Argentina 1994-1997. *Banco De La República Argentina - Working Papers*, 7.
- Andersen, H. (2008). Failure prediction of Norwegian banks: a logit approach. *Norges Bank Working Papers*, 2.

- Annesi, I., Moreau, T., y Lellouch, J. (1989). Efficiency of the logistic regression and Cox proportional hazards models in longitudinal studies. *Statistics in medicine*, 8(12), 1515-1521.
- Apergis, N., y Payne, J. (2013). European Banking authority stress tests and bank failure: evidence from credit risk and macroeconomic factors. *Banking And Finance Review*, 5(2).
- Argun, A. (2007). Bank failure prediction using modified minimum deviation model. *International Research Journal Of Finance And Economics*, 12.
- Avilés, C., y Guzman, I. (2001). La rentabilidad en cajas de ahorro y cooperativas de crédito: análisis comparativo. *Mágina: Revista Universitaria*, 9, 31-52.
- Ayala, R. (1999). Modelos de alerta temprana para crisis financieras. El caso ecuatoriano: 1.994-1997. *Working Paper BCE*, 51
- Ayala, M. A., Borges, R. E., y Colmenares, G. (2007a). Análisis de supervivencia aplicado a la banca comercial venezolana, 1996–2004. *Revista colombiana de estadística*, 30(1), 97-113.
- Ayala, M. A., Borges, R. E., y Colmenares, G. (2007b). Verificación de los supuestos del modelo de Cox. Caso de estudio: Banca comercial venezolana 1996-2004. *Economía*, 32(23), 27-43.
- Aziz, Z. A., y Dato, T. S. (2006). Islamic Banking and Finance Progress and Prospects Collected Speeches: 2000-2006. *Working Paper*. Kuala Lumpur: Bank Negara Malaysia.
- Barron, D., West, E., y Hannan, M. (1994). A Time to Grow and a Time to Die: Growth and Mortality of Credit Unions in New York City, 1914-1990. *American Journal Of Sociology*, 100(2), 381-421.
- Beaver, W. (1968). Alternative accounting measures as predictors of failures. *The Accounting Review*, 43(1), 113-122.
- Begley, J., Ming, J., y Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Review Of Accounting Studies*, 1(4), 267-284.

- Bell, T. (1997). Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures. *International Journal Of Intelligent Systems In Accounting, Finance y Management*, 6(3), 249-264.
- Belmonte, L. J. (2007). El Sector de las Cooperativas de Crédito en España: un estudio por Comunidades Autónomas. Análisis de su eficiencia y dimensión (1995-2003). *Consejo Económico y Social de Andalucía*, Junta de Andalucía.
- Belongia, M., y Gilbert, R. (1990). The Effects of Management Decisions on Agricultural Bank Failures. *American Journal Of Agricultural Economics*, 72(4), 901.
- Beltrán, C., Martín-Aceña, P., y Pons, M. A. (2012). Financial crises in Spain: lessons from the last 150 years. *Revista de Historia Económica/Journal of Iberian and Latin American Economic History (New Series)*, 30(03), 417-446.
- Berger, A. N., y Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of banking & finance*, 21(7), 895-947.
- Bernhardsen, E. (2001). A model of bankruptcy prediction (Tesis de Master). Universitete i Oslo.
- Betz, F., Oprica, S., Peltonen, T., y Sarlin, P. (2014). Predicting distress in European Banks. *Journal Of Banking And Finance*, 45, 225-241.
- Boj, E. (2015). *El modelo de regresión de Cox*. Consultado el 19 de marzo de 2016 en <http://diposit.ub.edu/dspace/bitstream/2445/49070/6/El%20modelo%20de%20Cox%20de%20riesgos%20proporcionales.pdf>
- Bologna, P. (2011). Is there a role for funding in explaining recent US banks' failures?, *IMF Working Paper* No. 180, International Monetary Fund, Washington, DC
- Borges, A., y de Olivera, J. (1996). Risco bancário: modelo de previsao de insolvência de bancos no Brasil. *Revista De Administração*, 31(2), 19-28.
- Bovenzi, J. F., Marino, J. A., y McFadden, F. E. (1983). Commercial bank failure prediction models. *Economic Review*, 68(1), 14-26.

- Boyacioglu, M.A., Kara, Y. y Baykan, O.K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of saving deposit insurance funds (SDIF) transferred banks in Turkey. *Expert Systems with applications*, 36, 3355-3366.
- Braga, M., Fully, V., Colosimo, E., y Bressan, A. (2006). Investigating the solvency of brazilian credit unions using a proportional hazard model. *Annals Of Public And Cooperative Economics*, 77(1), 83-106.
- Bressan, V. G. F., Braga, M. J., Bressan, A. A., y de Andrade R. F. M. (2010). Avaliação de insolvência em Cooperativas de Crédito: uma aplicação do Sistema PEARLS. *Revista de Administração Mackenzie*, 12(2).
- Briones, J. L., Marín, J. L. M., y Cueto, M. J. V. (1991). La predicción de la quiebra bancaria: el caso español. *Revista española de financiación y contabilidad*, 151-166.
- Brookes, M., Dicks, M., y Pradhan, M. (1994). An empirical model of mortgage arrears and repossessions. *Economic Modelling*, 11(2), 134-144.
- Caballo, Á. (2011). *Modelos combinados de análisis de riesgo de crédito en empresas no cotizadas en bolsa* (Tesis Doctoral). Universidad Pontificia Comillas.
- Cabo, P., y Rebelo, J. (2010). The survival of Portuguese Credit Co-operatives: an econometric approach. In *Co-operatives contributions to a plural economy*. Lyon: ICA European Research Conference.
- Cabrero, A., y Delrieu, J. (1996). Elaboración de un índice sintético para predecir la inflación en España. *Banco De España - Working Papers*, 9619.
- Canbas, S., Cabuk, A., y Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, 166(2), 528-546.
- Carrasco, R. (2001). Modelos de elección discreta para datos de panel y modelos de duración: una revisión de la literatura. *Cuadernos Económicos Del I.C.E.*, 66, 21-49.

- Carvalho, F., Diaz, M., Bialoskorski Neto, S., y Kalatzis, A. (2015). Exit and Failure of Credit Unions in Brazil: A Risk Analysis. *Revista Contabilidade y Finanzas*, 26(67), 70-84.
- Casey, C. J. (1980). The usefulness of accounting ratios for subjects' predictions of corporate failure: Replication and extensions. *Journal of Accounting Research*, 603-613.
- Chamberlain, G. (2010). Binary Response Models for Panel Data: Identification and Information. *Econometrica*, 78(1), 159-168.
- Cihák, M., y Hesse, H. (2007). Cooperative Banks and Financial Stability. *IMF Working Papers*, 07(2).
- Cinca, C. S., y del Brío, B. M. (1993). Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales. *Revista española de Financiación y Contabilidad*, 153-176.
- Clarke, R. G. (1988). *Strategic financial management*. Irwin Professional Publishing.
- Climent, S. (2013). La reestructuración del sistema bancario español tras la crisis y la solvencia de las entidades financieras. Consecuencias para las cajas de ahorros. *Revista De Contabilidad - Spanish Accounting Review*, 16(2), 136-146.
- Coakley, J., y Brown, C. (1993). Artificial Neural Networks Applied to Ratio Analysis in the Analytical Review Process. *Intelligent Systems In Accounting, Finance And Management*, 2(1), 19-39.
- Cole, R., y Gunther, J. (1995a). A CAMEL rating's shelf life. *Working Paper*. <https://ssrn.com/abstract=1293504>
- Cole, R., Cronyn, B., y Gunther, J. (1995b). FIMS: a new monitoring system for banking institutions. *Federal Reserve Bulletin*, 81, 1-15.
- Cole, R., y Gunther, J. (1998). Predicting bank failures: a comparison of On-and-off-site monitoring systems. *Journal Of Financial Services Research*, 13(2), 103-117.
- Cole, R., y Wu, Q. (2009). Is Hazard or Probit More Accurate in Predicting Financial Distress? Evidence from U.S. Bank Failures. *SSRN Electronic Journal*.

Cole, R., y Wu, Q. (2009). Predicting bank failures using a simple dynamic hazard model. In *22nd Australasian Finance and Banking Conference*.

Comrey, A. L. (1985). *Manual de análisis factorial*. Cádiz: Cádiz.

Cooper, R., Goldstein, M., Kaminsky, G., y Reinhart, C. (2000). Assessing Financial Vulnerability: An Early Warning System for Emerging Markets. *Foreign Affairs*, 79(6), 176.

Cordón, E., y García, F. (2014). *Tratamiento de la supervivencia con datos de duración* (1ª ed.). [Huelva]: Servicio de Publicaciones, Universidad de Huelva.

Costa, M., y Thegeya, A. (2013). Financial Soundness indicators and Banking Crises. *IMF Working Paper*, WP/13/263.

Cox, D. (1972). Regression models and life-tables. *Journal Of The Royal Statistical Society*, 34, 187-220.

Cox, D R y Snell, E J (1989), *Analysis of Binary Data*, Second Edition, Chapman and Hall

Cramer, J S (1991), *The Logit Model – An Introduction for Economists*, Edward Arnold.

Crespo, J. (2011). CAMEL vs. discriminante, un análisis de riesgo al sistema financiero venezolano. *Ecos De Economía*, 15 (33).

Cuadras, C. (1996). *Métodos de análisis multivariante*. Barcelona: EUB

Dabós, M. (1996). Crisis Bancarias y medición del riesgo de default: métodos y el caso de los Bancos Cooperativos en Argentina. *Universidad De San Andrés - Working Paper*, 12.

Davis, E. P. (1995). Debt, financial fragility, and systemic risk. *Oxford University Press*.

De Miguel, L. J., Revilla, E., Rodríguez, J. M., y Cano, J. M. (1993). A comparison between statistical and neural network based methods for predicting bank failures. *Proceedings of the IIIth International Workshop on Artificial Intelligence in Economics and Management*, Portland (USA).

- Delgado, J., y Saurina, J. (2004). Riesgo de crédito y dotaciones a insolvencias. Un análisis con variables macroeconómicas. *Moneda y Crédito*, 219, 11-41.
- Demirguc-Kunt, A., y Detragiache, E. (1998). The Determinants of Banking Crises in Developing and Developed Countries. *Staff Papers - International Monetary Fund*, 45(1), 81.
- Demirgüc-Kunt, A., y Detragiache, E. (1999). Monitoring banking sector fragility: a multivariate logit approach. *IMF Working Paper*, WP/99/147.
- Detragiache, E., y Demirgüç-Kunt, A. (1998). Financial Liberalization and Financial Fragility. *IMF Working Papers*, 98(83), 1.
- DeYoung, R., y Torna, G. (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, 22(3), 397-421.
- Dimitras, A. I., Zanakis, S. H., y Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, 90(3), 487-513.
- Distinguin, I., Rous, P., y Tarazi, A. (2006). Market discipline and the use of stock market data to predict bank financial distress. *Journal of Financial Services Research*, 30(2), 151-176.
- Distinguin, I., Tarazi, A., y Trinidad, J. (2008). The use of accounting and stock market data to predict bank financial distress: the case of East Asian banks. Working Paper, Universite de Limoges, France
- Domínguez, M., Blancas, F., Guerrero, F., y González, M. (2011). Una revisión crítica para la construcción de indicadores sintéticos. *Revista De Métodos Cuantitativos Para La Economía Y La Empresa*, 11, 41-70.
- Doğanay, M. M., Ceylan, N. B., y Aktaş, R. (2006). Predicting financial failure of the Turkish banks. *Annals of Financial Economics*, 2(01), 0650005.
- Duda, M., y Schmidt, H. (2010). Bankruptcy Prediction: Static Logit Model versus Discrete Hazard Models Incorporating Macroeconomic Dependencies. Lund University, 1-60.

- Edison, H. (2003). Do indicators of financial crises work? An evaluation of an early warning system. *International Journal Of Finance y Economics*, 8(1), 11-53.
- Ekinci, A., y Erdal, H. İ. (2016). Forecasting Bank Failure: Base Learners, Ensembles and Hybrid Ensembles. *Computational Economics*, 1-10.
- Encinas, B. (2008). La crisis como oportunidad para la redefinición del modelo de negocio de las Cajas Rurales y su apuesta por la eficiencia en costes. *XII Jornadas de investigadores en Economía Social y Cooperativa*. Murcia.
- Espahbodi, P. (1991). Identification of problem banks and binary choice models. *Journal of Banking & Finance*, 15(1), 53-71.
- Estévez, J. (2002). La construcción de un Índice cuantitativo sobre educación superior utilizando la técnica de análisis de componentes principales. *Revista De La Educación Superior*, 31(121), 138-153.
- Estrella, A., Park, S., y Peristiani, S. (2000). Capital ratios as predictors of bank failures. *Federal Reserve Bank Of New York Economic Policy Review*, 6(2).
- Etemadi, H., Rostamy, A. A. A., y Dehkordi, H. F. (2009). A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications*, 36(2), 3199-3207.
- Fajardo, G. (2011). How viable are spanish credit cooperatives after recent bank capitalization and restructuring regulations. *CIRIEC Revista De Economía*, 73.
- Fantazzini, D., y Figini, S. (2009). Random survival forests models for SME credit risk measurement. *Methodology and Computing in Applied Probability*, 11(1), 29-45.
- Fernández, J. (1989). Análisis de las insolvencias bancarias en España: un modelo empírico. *Moneda Y Crédito: Revista De Economía*, 189, 187-227.
- Fernández, F., González, E. A., y Félix, J. A. (2009). Especificación de modelos econométricos utilizando minería de datos. *Rect@: Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, (10), 223-252.

- Fernández, M. (2015). Adecuación de los modelos de predicción de las crisis bancarias al caso español (Tesis Doctoral). Universidad de Sevilla.
- Fernández, I. C. P., Guillen, F. P., y Cibrán, M. Á. C. (2015). Factores Determinantes De La Rentabilidad De Los Accionistas En Plantas Fotovoltaicas En Galicia. *Revista Galega de Economía*, 24(2), 31-44.
- Fiordelisi, F., y Mare, D. (2013). Probability of default and efficiency in cooperative banking. *Journal Of International Financial Markets, Institutions And Money*, 26, 30-45.
- Fox, J. (2002). Cox proportional-hazards regression for survival data. An R And S-PLUS Companion To Applied Regression.
- Franganillo, A. R., y Álvarez, M. D. (1995). La concentración de la banca en España. *Revista de fomento social*, 50, 417-429.
- Frankel, A. B., & Palmer, D. E. (1996). The management of financial risks at German nonfinancial firms: The case of Metallgesellschaft. Working Paper. (No. 560). <https://ideas.repec.org/p/fip/fedgif/560.html>
- Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.
- Fuentelsaz, L., Gómez, J., y Polo, Y. (2004). Aplicaciones del análisis de supervivencia a la investigación en economía de la empresa. *Cuadernos De Economía Y Dirección De La Empresa*, 19, 81-114.
- Fuentes, L. (2003). Metodología para la elección del punto de corte óptimo para dicotomizar covariables continuas. *Revista Cubana de Genética Comunitaria* 7 (3), 36-42.
- Fuertes, A. M., y Kalotychou, E. (2006). Early warning systems for sovereign debt crises: The role of heterogeneity. *Computational statistics y data analysis*, 51(2), 1420-1441.
- Gaganis, C., Pasiouras, F., y Zopounidis, C. (2008). Regulations, supervision approaches and acquisition likelihood in the Asian banking industry. *Asia-Pacific Financial Markets*, 15(2), 135-154.

- Gaganis, C., Pasiouras, F. y Zopounidis, C. (2010). Multicriteria classification models for the identification of targets and acquirers in the Asian banking sector. *European Journal of Operational Research*, 204(2), 328-335.
- Gama, V., Braga, M., y de Lima, J. (2004). Análise de Insolvência das Cooperativas de Crédito Rural do Estado de Minas Gerais. *Estudos Econômicos*, 34(3), 553-585.
- Gama, V., Braga, M., Bressan, A., y Resende, M. (2011). Avaliação de insolvência em cooperativas de crédito: uma aplicação do sistema PEARLS. *Revista De Administração Mackenzie*, 12(2), 113-144.
- García, E., Gil, J. y Rodríguez, G. (2000). *Análisis Factorial*. Editorial La Muralla (Madrid).
- García, T. (2000), ¿Existe una acumulación de riesgo en las Cajas de Ahorro?. Actas del X Congreso Nacional de ACEDE
- García, B. H. (2004). La Supervisión de los bancos y el rol del Comité de Basilea para la Supervisión Bancaria. *Revista Contaduría y Administración*, (212), 1.
- García, F., Guijarro, F., y Moya, I. (2010). Factores financieros clave en la reorganización del sector de las Cajas Rurales. *CIRIEC Revista De Economía*, 68.
- Gaytan, A., y Johnson, C. (2002). A review of the literature on early warning systems for banking crises. *Banco Central De Chile - Working Papers*, 183.
- Glick, R. (1999). Thoughts on the origins of the Asian crisis: impulses and propagation mechanisms. *The Asian Financial Crisis: Origins, Implications, and Solutions*. 33-63. Springer US.
- Goldstein, M., Kaminsky, G. L., y Reinhart, C. M. (2000). Assessing financial vulnerability: an early warning system for emerging markets. Peterson Institute.
- Goddard, J., McKillop, D., y Wilson, J. O. (2009). Which credit unions are acquired?..*Journal of Financial Services Research*, 36(2-3), 231-252.
- Godlewski, C. J. (2003), Bank's Default Modelisation: An Application to Banks from Emerging Market Economies. SSRN: <https://ssrn.com/abstract=588181> or

Gorsuch, R.L. (1983). *Factor Analysis (2nd ed)* Hillsdale, NJ: Erlbaum

Greene, W H (1993), *Econometric Analysis*, Prentice Hall.

Guimarães, A. y Moreira, T. B. S. (2008), Previsão de Insolvência: Um Modelo Baseado em Índices Contábeis com Utilização da Análise Discriminante (Insolvency Predictor: A Model Based on Account Index with Application of Discriminant Analysis). *R. Econ. Contemp.*, V. 12, N. 1, pp. 151-178. <https://ssrn.com/abstract=2067235>

Gutiérrez, C., y Abad, J. (2014). ¿Permitían los estados financieros predecir los resultados de los tests de estrés de la banca española? Una aplicación del modelo logit. *Revista De Contabilidad*, 17(1), 58-70.

Halling, M., y Hayden, E. (2006). Bank Failure Prediction: A Two-Step Survival Time Approach. *SSRN Electronic Journal*.

Hamerle, A., Liebig, T., y Scheule, H. (2004). Forecasting credit portfolio risk (No. 2004, 01). Discussion Paper, Series 2: Banking and Financial Supervision.

Hanweck, G. A. (1977). Predicting bank failure (No. 19). Board of Governors of the Federal Reserve System (US).

Hardy, D. C., y Pazarbaşıoğlu, C. (1999). Determinants and leading indicators of banking crises: Further evidence. *IMF Economic Review*, 46(3), 247-258.

Henebry, K. L. (1997). A test of the temporal stability of proportional hazards models for predicting bank failure. *Journal of Financial and Strategic Decisions*, 10(3), 1-11.

Hesse, H. y Čihák, M. (2007), 'Co-operative Banks and Financial Stability', *IMF Working Paper*

Hirtle, B., y Lopez, J. (1999). Supervisory information and the frequency of bank examinations. *Federal Reserve Bank Of New York Economic Policy Review*, 5(1).

Horowitz, J., y Savin, N. (2001). Binary Response Models: Logits, Probits and Semiparametrics. *Journal Of Economic Perspectives*, 15(4), 43-56.

- Hosmer, D., y Lemeshow, S. (2013). *Applied Logistic Regression* (2nd ed.). New York: John Wiley y Sons, Inc.
- Huang, D., Chang, B., y Liu, Z. (2010). Bank failure prediction models: for the developing and developed countries. *Quality y Quantity*, 46(2), 553-558.
- Hutchison, M., y McDill, K. (1999). Are all banking crises alike? The Japanese experience in international comparison. *Journal of the Japanese and International Economies*, 13(3), 155-180.
- i Triay, C. S. (2014). Las crisis bancarias en España: una perspectiva histórica. *Estudios de economía aplicada*, 32(2), 1-24.
- Janot, M. M. (2009) Modelos de previsão de insolvência bancária no Brasil: aplicação de diferentes modelos entre 1995 e 1998. 1999. 94 p. (Tesis Master) Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1999
- Johnson, C. (1970). Ratio Analysis and the Prediction of Firm Failure. *The Journal Of Finance*, 25(5), 1166.
- Jordan, D.J., Rice, D., Sanchez, J., Walker, C. y Wort, D. H. (2010). Predicting Bank Failures: Evidence from 2007 to 2010. <https://ssrn.com/abstract=1652924>
- Karacabey, A. A. (2007). Bank failure prediction using modified minimum deviation model. *International Research Journal of Finance and Economics*, 12, 147-159.
- Karminsky, A., Kostrov, A. y Murzenkov, T. (2012). Comparison of Default Probability Models: Russian Experience. *Higher School of Economics Research Paper No. WP BRP 06/FE/2012*. <https://ssrn.com/abstract=2152384>
- Kiefer, N. (1988). Economic duration data and hazard functions. *Journal Of Economic Literature*, 26, 646-679.
- Klein, F. (2014). Estimación de la probabilidad de default: un modelo probit para los bancos argentinos. *Ensayos de política económica*. 2 (2).

- Kolari, J., Glennon, D., Shin, H., y Caputo, M. (2002). Predicting large US commercial bank failures. *Journal of Economics and Business*, 54(4), 361-387.
- Kumar, S., y Arora, S. (1995). A model for risk classification of banks. *Managerial And Decision Economics*, 16(2), 155-165.
- Laffarga, J., Martín, J., y Vazquez, M. (1987). Predicción de las crisis bancarias en España. Comparación entre el análisis logit y el análisis discriminante. *Cuadernos De Ciencias Económicas Y Empresariales*, 18, 49-57.
- Laffarga, J., Martín, J., y Vazquez, M. (1991). *La predicción de la quiebra bancaria: el caso español*. *Revista Española De Financiación Y Contabilidad*., 21, 151-166.
- Laitinem, T., y Kankaanpää, M. (1999). Comparative analysis of failure prediction methods: the Finnish case. *The European Accounting Review*, 8(1), 67-92.
- Lagares, E. y Ordaz, F. (2014). *Tratamiento de la supervivencia con datos de duración: aplicaciones con Excel y STATA*. Servicios de Publicaciones Universidad de Huelva.
- Laguillo, G. (2015). Predicción de insolvencias en los sectores económicos: un análisis comparativo (Tesis Doctoral). Universidad de Málaga.
- Lane, W. R., Looney, S. W., y Wansley, J. W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking y Finance*, 10(4), 511-531.
- Lane, T. (1999). The Asian financial crisis: what have we learned?. *Finance and Development*, 36(3), 44.
- Laurent, C. R. (1979). Improving the efficiency and effectiveness of financial ratio analysis. *Journal of Business Finance y Accounting*, 6(3), 401-413.
- Laviola, S., Marullo-Reedtz, P., y Trapanese, M. (1999). Forecasting bank fragility: the evidence from Italy. *Research in Financial Services: Private and Public Policy*, 11, 35-60.
- Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal Of Economics And Business*, 51(4), 347-364.

- Li, Y y Ma, W. (2010): Application of artificial neural networks in financial economics: a survey. *International Symposium on Computational Intelligence and Design*, 211-214
- Lima, D. (2012). Forecasting insolvencies of Portuguese Co-operative banks. Third Prize In The First EACB Award Young Researchers On Co-Operative Banks.
- Logan, A. (2001). The United Kingdom's small bank's crisis of the early 1990s: what were the leading indicators of failure?. *Bank Of England Working Papers*.
- López-Barajas, D. M., y Carrascosa, J. R. (2005). La evaluación de la docencia universitaria. Dimensiones y variables más relevantes. *Revista de Investigación Educativa*, 23(1), 57-84.
- López, F., y Sanz, I. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems With Applications*, 42(6), 2857-2869.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. New York, Cambridge University Press.
- Magalhaes, M. (2001). Modelos de Previsao de insolvência bancária do Brasil. Banco Central De Brasil - Working Paper.
- Maggiolini, P., y Mistrulli, P. (2005). A survival analysis of de novo co-operative credit banks. *Empirical Economics*, 30(2), 359-378.
- Männasoo, K. y Mayes, D. (2009). Explaining bank distress in Eastern European transition economies, *Journal of Banking & Finance*, 33, 244-253
- Marín, S., Gras, E., y Renart, M. (2011). Financial information and restructuring of spanish savings banks in a cotext of crisis. Changes in the regulation; content and evolution of FROB. *Revista De Economía Pública, Social y Cooperativa*, 73, 99-126.
- Mar-Molinero, C., y Serrano-Cinca, C. (2001). Bank failure: a multidimensional scaling approach. *The European Journal Of Finance*, 7(2), 165-183.
- Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of banking y finance*, 1(3), 249-276.

- Martín, P., y Nogues, P. (2012). Crisis bancarias en la historia de España. Del antiguo régimen a los orígenes del capitalismo moderno. *Asociación Española De Historia Económica - Working Paper*, 1201.
- Martínez, C., Sanz, F. y de la Cruz, M. (1989). Selección y explotación de los sistemas de alarma y prevención de quiebra. *Investigaciones Económicas* 13 (3), 465-484.
- Martinussen, T., y Scheike, T. H. (2007). *Dynamic regression models for survival data*. Springer Science y Business Media.
- Matias, A. B., y de Oliveira Siqueira, J. (1996). Risco bancário: modelo de previsão de insolvência de bancos no Brasil. *Revista de Administração da Universidade de São Paulo*, 31(2).
- Melián, A., Campos, V., y Sanchís, J. (2011). Análisis de las fusiones entre Cajas Rurales y su influencia en las cooperativas agrarias: El caso valenciano en España. *INNOVAR*, 21(41), 91-109.
- Melián, A., Palomo, R., Sanchís, J., y Soler, F. (2006). La investigación en materia de cooperativas de crédito y de grupos cooperativos. *CIRIEC Revista De Economía*, 56, 25-54.
- Messai, A., y Gallali, M. (2015). Financial Leading Indicators of Banking Distress: A Micro Prudential Approach - Evidence from Europe. *Asian Social Science*, 11(21).
- Meyer, P., y Pifer, H. (1970). Prediction of Bank Failures. *The Journal Of Finance*, 25(4), 853.
- Meza, L. (2003). El paradigma positivista y la concepción dialéctica del conocimiento. *Revista Digital Matemática*, 4 (2), 1-5.
- Min, S. H., Lee, J., y Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert systems with applications*, 31(3), 652-660.
- Mínguez, J. (2005). La información contable en la empresa constructora: factores identificativos del fracaso empresarial (Tesis Doctoral). Universidad de Valladolid.

- Ministerio de Asuntos Exteriores y de Cooperación. (2012). Memorando de entendimiento sobre condiciones de política sectorial financiera.
- Módica, A., Baixauli, J., y Álvarez, S. (2012). Propuesta de un indicador de salud financiera y su efecto en la predicción del fracaso empresarial. *Revista Internacional Administración Y Finanzas*, 5(3).
- Molina, C. A. (2002). Predicting bank failures using a hazard model: the Venezuelan banking crisis. *Emerging markets review*, 3(1), 31-50.
- Mondéjar, J., y Vargas, M. (2008). Indicadores sintéticos: una revisión de los métodos de agregación. *Economía, Sociedad y Territorio*, 8(27), 565-585.
- Mondragón, Z., Renart, M. A., y Marín, S. (2011). Análisis de la Información Financiera en el caso de la predicción de la crisis bancaria en México. Comunicación XV Congreso AECA. http://www.aeca1.org/pub/on_line/comunicaciones_xvcongresoaecca/cd/125a.pdf
- Montgomery, H., Santoso, W., Besar, D.S. y Hanh, T. (2005) *Coordinated Failure? A Cross-Country Bank Failure Prediction Model*. Asian Development Bank Institute Discussion Paper No. 32. <https://ssrn.com/abstract=1905857>
- Muñoz, J. (1999). Calidad de cartera del sistema bancario y el ciclo económico: una aproximación econométrica para el caso peruano. www.bcrp.gov.pe/espanol/publicaciones/Revista/Revjul99/tema5.pdf.
- Nam, C., Kim, T., Park, N., y Lee, H. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal Of Forecasting*, 27(6), 493-506.
- Nardo, M., Saisana, M., Saltelli, A. y Tarantola, S. (2005), *Tools for Composite Indicators Building*, European Communities. <http://farmweb.jrc.cec.eu.int/ci/bibliography.htm>.
- Navarro, A. M. (2004). El fenómeno concentratorio como estrategia de crecimiento en las Cajas Rurales españolas. *REVESCO: Revista de Estudios Cooperativos*, (82), 89-112.
- Núñez, E., Esteyerberg, E., y Núñez, J. (2011). Estrategias para la elaboración de modelos estadísticos de regresión. *Revista Española De Cardiología*, 64(6), 501-507.

- Nuxoll, D. (2003). The Contribution of Economic Data to Bank-Failure Models. SSRN Electronic Journal.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal Of Accounting Research*, 18(1), 109.
- Odom, M. y Sharda, R. (1990). A neural network model for bankruptcy prediction. Proceedings of the IEEE International Conference on Neural Networks, San Diego (California), 163-168
- Olmeda, I. y Fernández, E. (1997). Hybrid Classifiers for Financial Multicriteria Decision Making: the case of bankruptcy prediction, *Computacional Economics* (10), 317-335
- Orozco, V. (2013). Los criterios epistemológicos en la investigación económica. *Pensamiento Crítico*, 18(1), 63-79.
- Ozkan-Gunay, E., y Ozkan, M. (2007). Prediction of bank failures in emerging financial markets: an approach. *The Journal of Risk Finance*, 8(5), 465-480.
- Pacheco, J., y Casado, S. (2005). Solving two location models with few facilities by using a hybrid heuristic: a real health resources case. *Computers & Operations Research*, 32(12), 3075-3091.
- Pacheco, J., Casado, S., y Núñez, L. (2007). Algoritmos meméticos para selección de variables en el análisis discriminante. *Estadística Española*, 45(165), 333-347.
- Pacheco, J., Casado, S., y Núñez, L. (2009). A variable selection method based on Tabu search for logistic regression models. *European Journal of Operational Research*, 199(2), 506-511.
- Palacio, J.A. (1980). La investigación en materia de cooperativas de crédito y grupos cooperativos. *CIRIEC-España* (56), 25-64
- Palacio, J. R. S. (2003). Análisis estratégico de las cooperativas de crédito. Estudio empírico aplicado a las cajas rurales de la Comunidad Valenciana. *Información Comercial Española*, 145-170.

- Pasiouras, F., Gaganis, C., y Doumpos, M. (2007). A multicriteria discrimination approach for the credit rating of Asian banks. *Annals of Finance*, 3(3), 351-367.
- Pasiouras F., Tanna S., y Zopounidis, C. (2009), The impact of banking regulations on banks' cost and profit efficiency: Cross-country evidence, *International Review of Financial Analysis*, 18, 294-302
- Pelepu, K.G. (1986). Predicting takeover targets: a methodological and empirical analysis. *Journal of accounting and economics*, 8 (1), 3-35
- Palmer, A. (1993). Regresión de Cox: Ejemplo numérico del proceso de estimación de parámetros. *Psicothema*, 5(2), 387-402.
- Palomo, R. y Mateu J.L (1999). Análisis y perspectivas de las cajas profesionales y populares: una modalidad específica de relación banca empresa. *Banca y Fianzas* 50. 7-12.
- Palomo, R. y Valor, C. (2001) *Banca cooperativa. Entorno financiero y proyección social*. Madrid, UNACC.
- Palomo, R. y Sanchis, J. R. (2008a). Un análisis del crédito cooperativo en España: situación actual, expansión territorial y proyección estratégica. *Estudios de Economía Aplicada*, 26 (1), 89-132.
- Palomo, R.. (2008b). Co-operative banking groups in Europe: comparative analysis of the structure and activity. *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa* (62), 87-119
- Palomo, R., y Sanchís, J. (2010a). Efectos de las fusiones bancarias en los resultados. El caso de las cajas rurales en España durante la primera mitad de la década de 2000. *Cuadernos De Economía Y Dirección De La Empresa*, 13(44), 13-36.
- Palomo, R.; Sanchis, J.R., Soler, F. (2010b). Las entidades financieras de economía social ante la crisis financiera: un análisis de las cajas rurales españolas, *REVESCO-Revista de Estudios Cooperativos*, 100. 101-133.
- Pantalone, C.C. y Platt, M.B. (1987). Predicting commercial bank failure since deregulation. *New England Economic Review* (jul), 37-47

- Pasiouras, F., Tanna, S., y Zopounidis, C. (2007). The identification of acquisition targets in the EU banking industry: An application of multicriteria approaches. *International Review of Financial Analysis*, 16(3), 262-281.
- Perez, C. (2009). *Técnicas de análisis de datos con SPSS 15*. Pearson Prentice Hall
- Perez, V. E., Blancas, F. J., Gonzalez, M., Lozano, M., Perez, F., y Caballero, R. E. (2009). Evaluación de la sostenibilidad del turismo rural mediante indicadores sintéticos. *Investigación operacional*, 30(1), 40-52.
- Pille, P. (1998). Performance analysis of the Ontario Credit Unions (Tesis Doctoral). Universidad de Toronto.
- Pina, V. (1989). La información contable en la predicción de la crisis bancaria 1977-1985, *Revista Española de Financiación y Contabilidad*, 18, 309-334
- Poghosyan, T., y Čihak, M. (2011). Determinants of Bank Distress in Europe: Evidence from a New Data Set. *Journal Of Financial Services Research*, 40(3), 163-184.
- Pol, A. L. P. (1993). Modelo de Regresión de Cox: Ejemplo numérico del proceso de estimación de parámetros. *Psicothema*, 5(2), 387-402.
- Pons, M.A. (2015). Understanding Spanish Financial crises, 1850-2000: what determined their severity? *Working Paper en LSE Economic History Seminars*.
- Porath, D. (2006). Estimating probabilities of default for german saving banks and credit cooperatives. *Schmalenbach Business Review*, 58, 214-233.
- Poveda, R. (2006). Propuesta de una metodología de ayuda a la decisión para los procesos de dirección y gestión de proyectos. (Tesis Doctoral). Universidad Politécnica de Valencia.
- Puig, X. (2002). Análisis de los factores de éxito y fracaso de las secciones de crédito de las cooperativas en Catalunya (Tesis Doctoral). Universitat Pompeu Fabra.
- Putnam, B. H. (1983). Early Warning Systems and Financial Analysis in Bank Monitoring. *Federal Reserve Bank of Atlanta Economic Review*. Noviembre. 6-14.

- Quesada, F., Sánchez, J., Aceituno, E., Fernández, L., García, N., y Moreno, E. *et al.* (2011). Análisis financiero de cajas y bancos. Periodo 2006-2010 en España. *Revista Digital Del Instituto Internacional De Costos*, 8, 9-46.
- Ravi, V. y Pramodh, C. (2008): Threshold accepting trained principal component neural network and features subset selection: application to bankruptcy prediction in Banks, *Journal Applied Soft Computing*, 8 (4), 1539-1548.
- Ramajo, J., Márquez, M. A., y Nogales, L. (2002). *Econometría aplicada: técnicas y modelos básicos*. Badajoz: ICE, Universitas Editorial.
- Redondo, D., y Rodríguez, J. (2014). Crisis en las entidades de crédito españolas: un estudio mediante análisis discriminante. *Estudios De Economía Aplicada*, 32(2), 617-644.
- Rodero, A. (2003). Las cajas rurales españolas. *Revista Española de Estudios Agrosociales y Pesqueros*, 200(I), 601-615.
- Rodríguez, J.M. (1989). Análisis de las insolvencias bancarias en España: un modelo empírico. *Moneda y Crédito Revista de Economía* (189), 187-227
- Rodríguez, M., Piñero, C., y De Llano, P. (2014). Determinación del riesgo de fracaso financiero mediante la utilización de modelos paramétricos, de inteligencia artificial y de información de auditoría. *Estudios De Economía*, 41(2), 187-217.
- Roget, F. M., Fernández, M. P. M., y de Miguel, J. C. (2005). El análisis envolvente de datos en la construcción de indicadores sintéticos. Una aplicación a las provincias españolas. *Estudios de Economía Aplicada*, 23(3), 715-716.
- Rose, P. S., y Kolari, J. W. (1985). Early warning systems as a monitoring device for bank condition. *Quarterly Journal of Business and Economics*, 43-60.
- Rúa, A., Redondo, R., y del Campo, C. (2003). Distribución municipal de la realidad socioeconómica gallega. *Revista Galega De Economía*, 12(2), 1-20.
- Ruiz, C. (2012). Aplicación del análisis de componentes principales como técnica para obtener índices sintéticos de calidad ambiental. *UCV-SCIENTIA*, 4(2), 145-153.

- Sagarra, M., Mar-Molinero, C. y García-Cestona, M. (2011). Spanish saving Banks in the credit crunch: could distress have been predicted before the crisis? A multivariate statistical analysis. *Fundación de las Cajas de Ahorro, Documentos de trabajo*, 667, 1-20.
- Sahajwala, R., y Van den Bergh, P. (2000). Supervisory risk assessment and early warning systems. *Basel Committee Of Banking Supervisory Working Papers*, 4.
- Sajter, D. (2005) *Early Prediction of Bank Failures in the Republic of Croatia*. SSRN: <https://ssrn.com/abstract=963966>
- Salas, J. S. (2002). Solvencia bancaria, riesgo de crédito y regulación pública: el caso de la provisión estadística española. *Hacienda Pública Española*, 161 (2), 129-150
- Salas, V., y Saurina, J. (2002). Credit risk in two institutional regimes: Spanish commercial and savings banks. *Journal of Financial Services Research*, 22(3), 203-224.
- Salchenberger, L. M., Cinar, E., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences*, 23(4), 899-916.
- Sánchez, J. (2003). Dos aplicaciones empíricas de las redes neuronales artificiales a la clasificación y la predicción financiera en el mercado español. *Revista Asturiana De Economía*, 28, 61-87.
- Sarkar, S., y Sriram, R. S. (2001). Bayesian models for early warning of bank failures. *Management Science*, 47(11), 1457-1475.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 239-241.
- Schuschny, A., y Soto, H. (2009). *Guía metodológica. Diseño de indicadores compuestos de desarrollo sostenible*. Naciones Unidas - Cepal.
- Schmidt, H. y Duda, M. (2010). Bankruptcy prediction: static logit model versus discrete hazard models incorporating macroeconomic dependencies (Tesis Doctoral). LUND University. <http://lup.lub.lu.se/student-papers/record/1614029>

- Secrist, H. (1938). *National bank failures and non-failures: an autopsy and diagnosis*. Principia Press.
- Serra, C., y Zúñiga, Z. (2002). Identificando bancos en problemas.¿ Cómo debe medir la autoridad bancaria la fragilidad financiera?. *Estudios Económicos*. Banco Central de Reserva del Perú.
- Serrano, C., y Martín, B. (1993). Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales. *Revista Española De Financiación Y Contabilidad*, 23, 153-176.
- Shaffer, S. (2012). Bank failure risk: Different now?. *Economics Letters*, 116(3), 613-616.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal Of Business*, 74(1), 101-124.
- Silva, L. C., y Barroso, I. M. (2004). *Regresión logística*. Madrid: La Muralla.
- Simon, C.J. (1980). Predicting the failure of credit unions: an application of multivariate logit analysis. (Tesis de Master), Massachusetts Institute of Technology.
- Sinkey, J. (1975). A Multivariate Statistical Analysis of the Characteristics of Problem Banks. *The Journal Of Finance*, 30(1), 21.
- Sinkey, J. F. (1978). Identifying " Problem" Banks: How Do the Banking Authorities Measure A Bank's Risk Exposure?. *Journal of Money, Credit and Banking*, 10(2), 184-193.
- Soto, Á. P. M. (2003). Los orígenes del cooperativismo de crédito agrario en España, 1890-1934. *Ciriec-España*, (44), 57-104.
- Sriram, R. S y Etheridge, H. L. (1996). A neural network approach to financial distress analysis. *Advances in Accounting Information Systems*, 4, 201-222.
- Stuhr, D., y Van Wicklen, R. (1974). Rating the financial condition of banks: a statistical approach to Ald Bank supervision. *Federal Reserve Bank Of New York Working Paper*.
- Sudriá, C. (2014). La crisis bancaria en España: Una perspectiva histórica. *Estudios De Economía Aplicada*, 32(2), 473-496.

- Swicegood, P., y Clark, J. (2001). Off-site monitoring systems for predicting bank underperformance: a comparison of neural networks, discriminant analysis, and professional human judgment. *International Journal Of Intelligent Systems In Accounting, Finance y Management*, 10(3), 169-186.
- Tam, K., y Kiang, M. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science*, 38(7), 926-947.
- Tan, T., Quek, C., y Ng, G. (2007). Biological brain inspired genetic complementary learning for stocks market and bank failure prediction. *Computational Intelligence*, 23(2), 236-261.
- Tascón, M., y Castaño, F. (2011). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. *Revista De Contabilidad*, 15(1), 7-58.
- Tatom, J. y Houston, R. (2011). Predicting Failure in the Commercial Banking Industry. Networks Financial Institute Working Paper No. 2011-WP-27. SSRN: <https://ssrn.com/abstract=1969091>
- Taylor, R. A. (1979). Optimal reserve levels for credit unions. *Rivista Internazionale Di Scienze Economiche E Commerciali*, 26(10), 971-983.
- Thomson, J. (1989). An Analysis of bank failures: 1984 to 1989. Working Paper.
- Thomson, J. (1992), Modeling the bank regulator's closure option: A two-step logit regression approach, *Journal of Financial Services Research*, 6, 5-23
- Trussel, J., y Johnson, L. (2012). A parsimonious and predictive model of the recent bank failures. *Academy of Banking Studies Journal*, 11(1), 15.
- Trustorff, J., Konrad, P., y Leker, J. (2010). Credit risk prediction using support vector machines. *Review Of Quantitative Finance And Accounting*, 36(4), 565-581.
- Vargas, C. (2015). Situación y perspectivas del cooperativismo de crédito en España ante la futura e incierta reforma de su régimen legal. *Boletín De La Asociación Internacional De Derecho Cooperativo*, 48, 115-135.

- West , R. C. (1985). A factor analytic approach to bank condition. *Journal of banking and finance*, 9, 253-266
- Whalen, G. (1991). A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool. *Economic Review-Federal Reserve Bank of Cleveland*, 27(1), 21.
- Wheelock, D., y Wilson, P. (2000). Why do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions. *Review Of Economics And Statistics*, 82(1), 127-138.
- White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns. *Neural Networks*, 1988., IEEE International Conference on 24-27 julio 1988.
- Woon, L. F., Azizan, N. A., y Samad, M. F. A. (2011). A strategic framework for value enhancing enterprise risk management. *Journal of global business and economics*, 2(1), 23-47.
- Zmijewski, M.E. (1984). Methodological issues related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research*, 22, 59-82

ANEXOS

ANEXO 1: Resultados del análisis factorial en el periodo 2000-2002

TABLA 53: PRUEBA KMO Y BARTLETT PARA EL PERIODO 2000-2002

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,624
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado gl	963,382 55
	Sig.	,000

FUENTE: ELABORACIÓN PROPIA

TABLA 54: VARIANZA TOTAL EXPLICADA

Varianza total explicada

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,599	41,807	41,807	4,599	41,807	41,807	3,953	35,937	35,937
2	2,285	20,771	62,579	2,285	20,771	62,579	2,508	22,797	58,733
3	1,118	10,161	72,739	1,118	10,161	72,739	1,541	14,006	72,739
4	,949	8,624	81,363						
5	,750	6,817	88,180						
6	,550	5,001	93,181						
7	,427	3,882	97,063						
8	,267	2,426	99,489						
9	,038	,341	99,831						
10	,010	,092	99,923						
11	,008	,077	100,000						

Método de extracción: análisis de componentes principales.

FUENTE: ELABORACIÓN PROPIA

TABLA 55: MATRIZ DE COMPONENTES ROTADOS

	Componente		
	1	2	3
OF1	,882		
EMP	,868		,374
Tamaño	,866		
Apalancamiento	-,830		
ROE	,721	,548	
ROA		,884	
OF2	-,443	,754	
Eficiencia	,355	-,689	
Capital		-,509	
EsfuerzoProv			-,800
Negocio	,368		,700

FUENTE: ELABORACIÓN PROPIA

ANEXO 2: Resultados del modelo logit para el periodo 2000-2002 a partir de la selección por componentes principales y mediante el uso de los factores estimados.

TABLA 56: BONDAD DEL AJUSTE DEL MODELO A PARTIR DE LA SELECCIÓN POR COMPONENTES PRINCIPALES

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	48,797 ^a	,086	,168

FUENTE: ELABORACIÓN PROPIA

TABLA 57: TABLA DE CLASIFICACIÓN DEL MODELO A PARTIR DE SELECCIÓN POR COMPONENTES PRINCIPALES.

	Observado	Pronosticado		
		Dependiente		Porcentaje correcto
		Operativa	Fusionada	
Paso 1	Operativa	69	0	100,0
	Fusionada	9	0	,0
	Porcentaje global			88,5

FUENTE: ELABORACIÓN PROPIA

TABLA 58: VARIABLES SIGNIFICATIVAS EN EL MODELO LOGIT A PARTIR DE LA SELECCIÓN POR COMPONENTES PRINCIPALES

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	OF1	,552	,237	5,408	1	,020	1,737
	Constante	-3,637	,919	15,666	1	,000	,026

FUENTE: ELABORACIÓN PROPIA

TABLA 59: BONDAD DEL AJUSTE DEL MODELO UTILIZANDO LOS TRES FACTORES CALCULADOS.

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	48,786 ^a	,086	,168

FUENTE: ELABORACIÓN PROPIA

TABLA 60: TABLA DE CLASIFICACIÓN DEL MODELO UTILIZANDO LOS TRES FACTORES CALCULADOS

	Observado	Pronosticado		
		Dependiente		Porcentaje correcto
		Operativa	Fusionada	
Paso 1	Dependiente Operativa	68	1	98,6
	Fusionada	9	0	,0
	Porcentaje global			87,2

FUENTE: ELABORACIÓN PROPIA

TABLA 61: VARIABLES SIGNIFICATIVAS EN EL MODELO LOGIT UTILIZANDO LOS TRES FACTORES CALCULADOS.

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	FAC1_2	,698	,305	5,214	1	,022	2,009
	Constante	-1,590	,375	17,991	1	,000	,204

FUENTE: ELABORACIÓN PROPIA

ANEXO 3: Resultados del análisis factorial en el periodo 2011-2014

TABLA 62: PRUEBA KMO Y BARTLETT PARA EL PERIODO 2011-2014

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,606
Prueba de esfericidad de	Aprox. Chi-cuadrado	298,337
Bartlett	gl	55
	Sig.	,000

FUENTE: ELABORACIÓN PROPIA

TABLA 63: VARIANZA TOTAL EXPLICADA

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,582	41,658	41,658	4,582	41,658	41,658	4,045	36,769	36,769
2	3,503	31,849	73,506	3,503	31,849	73,506	3,493	31,757	68,526
3	1,173	10,665	84,171	1,173	10,665	84,171	1,721	15,645	84,171
4	,646	5,877	90,048						
5	,550	5,004	95,052						
6	,333	3,023	98,075						
7	,149	1,357	99,432						
8	,054	,494	99,926						
9	,004	,041	99,966						
10	,002	,022	99,988						
11	,001	,012	100,000						

Método de extracción: análisis de componentes principales.

FUENTE: ELABORACIÓN PROPIA

TABLA 64: MATRIZ DE COMPONENTES ROTADOS

	Componente		
	1	2	3
Eficiencia	-,956		
ROA	,936		
ROE	,918		
OF2	,877		
Negocio	,517	,437	,465
OF1		,977	
EMP		,968	
Tamaño		,941	
EsfuerzoProv			-,856
Capital	-,373		,699
Apalancamiento	,399	-,433	-,444

FUENTE: ELABORACIÓN PROPIA

ANEXO 4: Resultados del modelo logit para el periodo 2011-2014 a partir de la selección por componentes principales y mediante el uso de los factores estimados.

TABLA 65: BONDAD DEL AJUSTE DEL MODELO A PARTIR DE LA SELECCIÓN POR COMPONENTES PRINCIPALES

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	79,721 ^a	,060	,088

FUENTE: ELABORACIÓN PROPIA

TABLA 66: TABLA DE CLASIFICACIÓN DEL MODELO A PARTIR DE SELECCIÓN POR COMPONENTES PRINCIPALES.

	Observado	Pronosticado		
		Dependiente		Porcentaje correcto
		Fusionada	Operativa	
Paso 1	Dependiente ,0	54	1	98,2
	1,0	19	0	,0
	Porcentaje global			73,0

FUENTE: ELABORACIÓN PROPIA

TABLA 67: VARIABLES SIGNIFICATIVAS EN EL MODELO LOGIT A PARTIR DE LA SELECCIÓN POR COMPONENTES PRINCIPALES

	B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a OF1	,295	,142	4,325	1	,038	1,344
Constante	-1,867	,505	13,698	1	,000	,155

FUENTE: ELABORACIÓN PROPIA

TABLA 68: BONDAD DEL AJUSTE DEL MODELO UTILIZANDO LOS TRES FACTORES CALCULADOS.

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	73,890 ^a	,131	,193
2	75,012 ^a	,118	,174

FUENTE: ELABORACIÓN PROPIA

TABLA 69: TABLA DE CLASIFICACIÓN DEL MODELO UTILIZANDO LOS TRES FACTORES CALCULADOS

	Observado	Pronosticado		
		Dependiente		Porcentaje correcto
		Operativa	Fusionada	
Paso 2	Dependiente Operativa	53	2	96,4
	Fusionada	15	4	21,1
	Porcentaje global			77,0

FUENTE: ELABORACIÓN PROPIA

TABLA 70: VARIABLES SIGNIFICATIVAS EN EL MODELO LOGIT UTILIZANDO LOS TRES FACTORES CALCULADOS.

	B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 2 ^a FAC1_2	-,956	,515	3,447	1	,063	,385
FAC3_2	,698	,326	4,584	1	,032	2,009
Constante	-,655	,315	4,330	1	,037	,520

FUENTE: ELABORACIÓN PROPIA