A synthetic index for predicting financial distress: An empirical analysis of its effectiveness when applied to rural Spanish credit cooperatives

Antonio Madera del Pozo Ph.D.
Departamento de Riesgos. Banco Cooperativo Español
Natalia Cassinello Plaza Ph.D.
Profesor de finanzas Universidad Pontificia Comillas
Antonio Rua Vieites Ph.D
Profesor de Métodos cuantiativos Universidad Pontificia Comillas

Abstract

This paper proposes a new tool to predict the financial health of Spanish credit cooperatives by means of a synthetic index that allows us to forecast the likelihood of such credit cooperatives going out of business or remaining solvent. The index was constructed using only significant financial variables in binary response models logit, static logit and multi-period logit. The proposed synthetic indicator incorporates explanatory variables, rescaled and weighed its importance by using the weighted arithmetic mean of the scoring coefficients obtained from the factorial analysis through the principal components method. The factors that make up the synthetic index contain profitability ratios (2), efficiency (1), management (1) and size (3).

The index was applied to the case of rural Spanish credit cooperatives (a significant subsector of the credit cooperatives) from 1990 to 2016. The index shows that the proposed indicator discriminates the operating entities of the merged both ex post and ex ante, with degrees of success greater than 76%. In addition, significant differences have been obtained in the synthetic indicator scores according to the geographic location.

Keywords: Bankruptcy, synthetic index, rural credit cooperatives, credit cooperatives, binary response model, logit, multi period logit,

JEL: C51,G21,G33

1. INTRODUCTION

Since the findings by Beaver (1966) and Altman (1968), research carried out using economic and financial analysis on distressed financial institutions has been focused on the search for Early Warning System (EWS) which could anticipate the insolvency of financial institutions. In the financial sector, the objective of identifying EWS is to prevent disastrous economic hardships that banks might have to endure and possible economic recession in the affected countries as well as the high costs of restructuring and bailing out banks, when needed (Serrano 2003).

The origin of research on bankruptcy in banking is linked to the financial ratios that are used in research on corporate bankruptcy. Early research in this field started with descriptive studies that compared key ratios of companies that went bankrupt during the Great Depression of 1929, with those of companies that survived the crisis (Fitzpatrick 1930; Winakor and Smith 1935; Merwin 1942). Subsequently, a new era in predictive research emerged, spearheaded by the works of Beaver (1966, 1968) and Altman (1968), who employed univariate and multivariable analysis. These models are based on the belief that solvency of financial institutions is predictable through the use of accounting data as an independent variable.

It is rather common that these early warning system are developed from a basic framework of statistical models that use discriminant analysis, logit-probit, solvency, and neuronal networks or artificial intelligence. The models combine dichotomous dependent variables (1 for bankruptcy and 0 for solvency) and financial ratios that behave as explanatory variables. These models are often criticized due to the limitations of the sample. But the main objection to the statistical models is that the models do not sufficiently take into consideration the dynamic nature of the insolvency processes (Shumway 2001; Nam et al. 2008; Nuñez et al. 2011). Bankruptcy is the final result of financial decline over a period of time, and therefore synthetic indexes should be used to analyze it (Módica, A., Baixauli, J., and Álvarez, S. 2012).

The earliest publications with respect to the financial sector are from the first half of the XX century, in particular, those from Secrist (1938); Meyer and Pifer (1970). These authors tried to explain and anticipate financial distress in US financial institutions by using both financial ratios and linear regression.

Research in Spain on predicting bankruptcy in the banking sector based on discriminant analysis and logit models started with Laffarga, J., Martín, J. and Vazquez, M. (1987). Since then, the focus has been on the analysis of insolvencies in retail banks and savings banks, without any major attention paid to the subsector of rural credit cooperatives.

Research on insolvency of credit cooperatives is very limited (Madera 2017); Mainly, international research exist in this area and the research is focused on the development of early warning system that use logit techniques and survival analysis. Thus, no analysis has been performed employing synthetic index, as has been done on banks or financial stability. (Estrada and Morales, 2009; Morris 2010; Albulescu, 2010; Illing and Liu 2014).
In some other countries specific studies on credit cooperative insolvency have been carried out, such as the case of Porath (Germany 2006), Cabo et al (Portugal 2010) and Lima (Portugal 2012). There is therefore a field in literature on insolvency in rural cooperatives in Spain that has yet to be explored.

There are several aspects of credit cooperatives that make them worth studying independently from other financial institutions because there exists significant differences in their financial activity and their relationships with other cooperatives. Credit cooperatives, as such, are financial institutions subject to the banking regulation (Law 13/1989 and RD 84/1993). Nonetheless, they are also subject to special regulations pertaining to cooperatives, which affect governing bodies, management, the relationships with members of the cooperative and third parties as well as the requirements to build up special reserves. As a result, credit cooperatives can focus on lending to its members, mainly in rural areas, to guarantee sufficient funding to the rural sector (Palacios 2003; Navarro 2004; Palomo, R., and Sanchís, J. 2010). They also provide banking services to small towns that would not have access to branch banking, as they do not reach the critical mass for private banks to set up branches.

Insolvency of rural credit cooperatives is very rare in Spain. Nonetheless, it is quite normal to observe a reduction in the number of institutions due to mergers, which try to preserve the cooperatives´ financial stability (Madera 2017). This phenomenon was also observed by Porath in Germany, Cabo et al. and Lima in Portugal. As explained by Madera (2017), in Spain, mergers of rural credit cooperatives occur when such cooperatives are undergoing financial difficulties and are then absorbed by or merge with a more solvent institution. Therefore, our research analysis posits the question of whether we can use a synthetic index to predict the possibility of a merger of rural Spanish credit cooperatives into a more solvent institution.

This paper is structured as follows: The first part is an introduction to the object of study and a summary of the background. The second part is a review of the existing research on insolvency in credit cooperatives and the application of synthetic indices to predict banking crises. In the third section, we discuss the methodology used in the construction of our synthetic index to predict solvency and the sample data. In the fourth section, we present the results of our research on the insolvency of rural Spanish credit cooperatives from 1990 to 2016. Lastly, we discuss our conclusions of the research.

2. CONTEXT

Research on bank insolvencies is extensive, especially at the international level, and focused on the search for EWS that allow us to forecast insolvencies through discriminant analysis, logit-probit, transition or survival models and artificial intelligence.

At first, the methodology for the study of corporate bankruptcies focused on discriminant analysis that includes univariate analysis techniques as well as techniques based on the calculation of linear regressions. This methodology has been used in studies after the crisis of 29 in numerous studies summarized in Madera (2017), such as Secrist (1938) to date by many authors. Mainly, Altman (1968); Meyer and Pifer (1970); Stuhr and Van Wickle (1974); Sinkey (1975); Sinkey (1978); Ohslon (1980), Rose and Kolari (1985), Sriram and Etheridge (1996), Ahumada and Budnevich (1999), Swicegood and Clark (2001), Sajter (2005), Aziz and Dato (2006), Doganay, Ceylan and Aktas (2006), Ozkan-Gunay and Ozkan (2007), Pasiouras et al (2007), Gaganis et al (2008), Guimaraes et al (2008), Pasiouras, Tanna and Zopounidis (2009), Gaganis, Pasiouras and Zopounidis (2010), Jordan, Rice, Sanchez, Walker and Wort (2010), Crespo (2011).

The main limitations of this methodology exposed in the literature are due to the assumptions underlying the existence of linearity between the variables and the hypothesis of normality of the accounting ratios. For this reason, studies based on logit / probit models are the most used in the literature. The largest number of works since 1975 have applied the logit and probit methodologies summarized in Madera (2017), such as Sinkey (1975), Martin (1977); Hanweck (1977), Ohslon (1980), Maddala (1983), Bovenzi et al. (1983); West (1985) Pantalone and Plat (1987); Abrams and Huang (1987), Kolari, Glennon, Shin and Caputo (1987), Cox and Snell (1989), Belongia and Gilbert (1990), Espahbodi (1991), Whalem (1991), Thompson (1992); Cramer (1991), Greene (1993), Cole and Gunther (1995a), Cole, Cronyn and Gunther (1995b), Borges and Olivera (1996); Siriam and Etheridge (1996), Frankel and Palmer (1996), Cole and Gunther (1998), Demirgüc et al. (1998); Hardy and Pazarbasioglu (1999); Hutchinson and McDill (1999); Glick (1999); Demirgüc et al. (1999); Estrella, Park and Peristiani (2000), Logan (2001), Maghalanes (2001), Kolari et al (2002), Godlewski (2003), Sajter (2005), Montgomery, Santoso, Besar and Hanh (2005), Distinguin et al (2006), Pasiouras et al (2007), Pasiouras et al (2007), Distinguin, Tarazi and Trinidad (2008), Gaganis, Pasioras and Zopounidis (2008), Gaganis et al (2010), Tatom and Houston (2011) , Poghosyan et al. (2011); Karminsky, Kostrov and Murzenkov (2012), Shafer (2012), DeYoung (2013).

This paper uses the logit and probit models to choose the explanatory variables of a possible bankruptcy. The originality of the work is the combination of using a static logit and a multi-period logit to incorporate the effect of time.

Other researchers have followed survival models summarized in Madera (2017), such as Lane, Looney and Wansley (1986), Whalen (1991), Henebry (1997), Anastasi (1998); Wheelock and Wilson (2000); Magalhaes (2001), Molina (2002). And since the late 90s, the application of other methodologies based on neural networks and artificial intelligence has begun (Odom and Sharda (1990), Espahbodi (1991), Tam et al. (1992), Salchenberger, Cinar and Lash (1992). ; Serrano (1993), Kumar and Arora (1995), Bell (1997), Swicegood and Clark (2001), Ravi and Pramodh (2008), Boyacioglu, Kara and Baykan (2009), Etemadi, Rostamy and Dehkordi (2009), Messai and Gallali (2015).

Of the extensive literature on business insolvencies, only a small group of authors focus on bankruptcy forecasting of credit cooperatives, summarized in Madera 2017, see Simon (1980) in the United States; Dabos (1996) in Argentina; Pille (1998) in Canada; Gama et al., (2004), Braga, Fully, Colosimo and Bressan (2006), Gama et al (2011) and Carvalho, Diaz, Bialoskorski and Kalatzi (2015) in Brazil; Porath (2006) in Germany; Cabo y Rebelo (2010) and Lima (2012) in Portugal; and Maggiolini and Mistrulli (2005), and Fiordelisi and Mare (2013) in Italy.

In the Spanish case, this line of research is scarce, highlighting the publication of Redondo and Rodríguez (2014) that studies the insolvencies of Spanish financial institutions between 2008 and 2010, including credit cooperatives.

In relation to the definition of the dependent variable, this work is in agreement with previous research in other countries that link the possibility of bankruptcy of credit cooperatives to the probability of merger. Within this sector it is usual for credit cooperatives to be assisted by other members, either through mergers or acquisitions, or through direct aid that masks the true situation of the cooperative involved. Maggiolini and Mistrulli (2005), Porath (2006), Cabo and Rebelo (2010), Lima (2012).

Although the proposed models, they are correct in the classification between solvents and insolvent in high levels, among others 89% in Dabos (1996), 98% in Gama et al. (2004) or 79% in Braga et al. (2006), the use of this type of models is not exempt from criticism for ignoring the dynamic nature of the financial economic structures of the companies and entities analyzed (Shumway, 2001, Nam et al., 2008, and Nuñez et al., 2011).

For this reason, the synthetic indicator is prepared according to the hypothesis of the need to take into account that bankruptcy is a process over time. (Nam et al., 2008., Modica et al, 2012., and Albulescu, 2010).

However, the proposals of synthetic indicators for the study of bank insolvencies are scarce, non-existent in Spain, highlighting Estrada and Morales (2009) with their index of financial stability in Colombia; Morris (2010) with an indicator for Jamaica; Albulescu (2010) on the financial stability system of Romania; Illing and Liu (2014) on an indicator of financial stability in Canada; and Sere, Udom, Saliju, Atoi and Yaaba (2014) on a bank stability index in Nigeria.

With our proposal, we contribute to the line of research based on the use of synthetic indicators applied to the Spanish financial sector in the case of credit cooperatives. Using the statistical models Logit, Static Logit and Logit multi period for the determination of the variables that will conform the same. The novelty is in the combination of the methods when in the literature the most used is the use of probabilistic regressions (logit-probit) and discriminant analysis.

3. METHODOLOGY

Our hypothesis is that it is possible to construct a synthetic indicator that allows us to identify those rural credit cooperatives ("Cajas Rurales") ceased activity due to financial sustainability. Following Mondéjar and Vargas (2008), we can define a synthetic indicator as a combination of explanatory variables with the characteristics of simplicity of use and reliability of results.

The construction of a synthetic indicator is a process that begins with the selection of individual indicators (Nardo, M., Saisana, M., Saltelli, A. and Tarantola, S., 2005). After being normalized and weighted, they are added to construct the synthetic indicator.

1. Selection of individual indicators (or explanatory variables)

We have not found previous researches focus on the use of synthetic indicators to measure and predict financial soundness of Spanish banks. Therefore, the selection of the individual indicators has been made on the basis of the explanatory variables used in the proposals of EWS applied to credit cooperatives. In those researches the explanatory variables have been grouped around capitalization, assets, profitability, indebtedness, liquidity and macroeconomic environment. Authors have chosen the explanatory variables following two criteria:

a. Variables that previously have been statistically significant in other investigations, for example in Braga et al. (2006), Porath (2006) and Cabo and Rodero, (2010).

b. Following PEARLS System (Gama et al., 2011), CAMEL (Lima, 2012, and Fiordelisi et al., 2013) or other non-academic techniques (Gama et al., 2004).

For the construction of the synthetic indicator, we have chosen 25 explanatory variables (that have been significant in previous EWS) grouped in assets, profitability, indebtedness, liquidity and macroeconomic environment (Annex 1).

To avoid problems of parsimony, the initial selection has been subjected to a new selection process under the criterion of being significant in four logit models estimated in Madera (2017)[1] as EWS proposals for the Spanish credit cooperatives (see Table 1). Where ROA is Return on Assets; ROB is return on branches; C/I is cost to income; TA is Ln Total Assets; FTEs is Ln Full time equivalents and Bchs is Ln Branches

TABLE 1: SIGNIFICANT VARIABLES IN MADERA (2017).

VARIABLES	Logit (2000-2002)	Logit (2011-2014)	Logit estátic (1991-2014)	Logit multiperiod
TA	X	X	X	
MNGT				X
C/I	X			
ROA	X		X	X
Branches	X	X	X	
ROB	X			
FTEs	X			
UB			X	X
DGS				X
Unemployment				X

SOURCE: MADERA (2017)

[1] Madera (2017) proposes four logit models as EWS of the Spanish Cajas Rurales, differentiated into two types:

a) Static logit models, which match the number of operational Rural Savings Banks and those that ceased activity due to reasons of economic and financial sustainability. To do so, they use financial information from each Caja Rural corresponding to the year immediately prior to the phenomenon studied. These logit models follow the following general expression:

$$P = \frac{1}{1 + e^{-\alpha - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_k X_k}}$$

Where P is the probability of lose of activity of the Caja Rural; X_i the explanatory variables used (financial economic ratios); and β_i the estimated parameters for each of the explanatory variables. In Madera (2017) the authors use this model for a sample restricted to the period 2000-2002, 2011-2014 and 1991-2014.

b) Multi-period logit models proposed by Shumway (2011). It differs from the static version in that those models use all the economic and financial information available to each individuals, so that the number of cases available increases. The formal expression of these models is very similar, with the difference of including explanatory variables that depend on time:

$$P = \frac{1}{1 + e^{-(\beta'_1 x_t + \beta'_2 x_{it})}}$$

Where x_t is a time-dependent variable (baseline risk function, which for Shumway (2001) and Wood (2017) corresponds to the logarithm of the individual's survival time). i_t corresponds to the individual's own characteristics, variables that also depend on time.

We have chosen this selection process for its originality and also for the high percentages of success shown both in the estimation and in the validation samples (see Table 2).

TABLA 2: PERCENTAGES OF SUCCESS. LOGIT MODEL IN MADERA (2017).

	ESTIMATION SAMPLE			VALIDATION SAMPLE		
Seleccion	**Operating**	**Merged**	**Global**	**Operating**	**Merged**	**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%

SOURCE: MADERA (2017)

Finally, for the construction of our synthetic indicator we have chosen explanatory variables that measure size (Total assets (TA), Branches (Bchs) and employees (FTEs)), management (Management (MNGT) and Cost to income (C/I)) and profitability (Retur on assets and on branches,(ROA and ROB)). Variables related to capitalization, indebtedness and liquidity have not been significant. Qualitative variables (location (UB) and Deposit guarantee scheme (DGS)) have not been considered because formal restriction of the factorial analysis used to calculate the weights of each variable within the synthetic indicator. Neither the macroeconomic variables (Unemployment) has been chosen, because that variable would act as constants within the indicator by assuming similar values for all Credit cooperatives.

In Table 3 we have showed the expression of each explanatory variables used for the construction of the synthetic indicator. The first column reflect its denomination and the second its calculation procedure.

TABLE 3: EXPLANATORY VARIABLES.

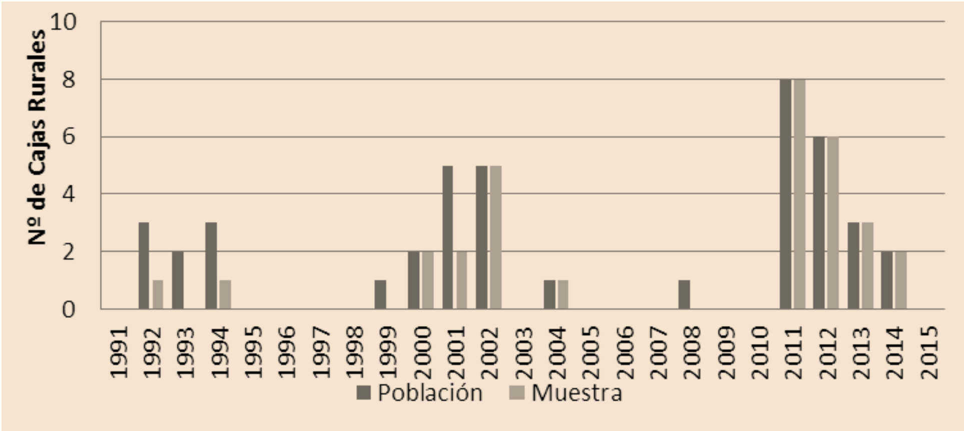
DENOMINATION	RATIO
TA	log(total assets)
MNGT	$\frac{\text{Loans to customers}}{\text{Total assets}}$
C/I	$\frac{\text{Operating expenses}}{\text{Interest and fees}}$
ROA	$\frac{\text{Net results}}{\text{Total assets}}$
Branches	ln (branches)
ROB	$\frac{\text{Net results}}{\text{Branches}}$
FTEs	ln (employees)

SOURCE: OWN ELABORATION

2. The sample

We have used a data panel from 1991 to 2015 of 74 Cajas Rurales (at 1991). A total of 1,745 observations of financial statements have been used (of the year before). A Caja Rural is a type of credit cooperative that combines characteristics of financial entities and cooperative societies. Its main difference with other credit cooperatives is its agrarian character. Our sources were UNACC (National Association of Credit Cooperatives) and Cajamar Cooperative Group. We stablished the hypothesis that a synthetic indicator could discriminate between Cajas Rurales for reasons of financial sustainability based on a series of explanatory variables. Therefore, it is necessary to identify those Cajas Rurales that disappeared during the period in order to construct the sample. We draw on Bank of Spain public data. It has been possible to identify a total of 30 Cajas Rurales that disappeared during the period considered by mergers with other Cajas Rurales. We reject the rest of the cases shown in Figure 1 due to the lack of financial statements.

FIGURE 1: MERGERED CAJAS RURALES BETWEEN 1991 AND 2015.



SOURCE: OWN ELABORATION

Finally, the sample has been divided into two parts. The first one from 1991 to 2014 (with financial economic data from the previous year) for the estimation; and the second part, from 2015 to 2016, used to validate the results.

3. Construction of the synthetic indicator

The explanatory variables are expressed in different units of measurement. After eliminating atypical cases (it has meant the loss of two cases corresponding to Caja Rural that ceased activity), the variable have been standardized (rescaling levels [0,100], with the maximum of 100).

At first, we have verified if the individual indicators can differentiate between Cajas Rurales (in order of its financial sustainability) throw a t-test of equality of means between both groups.

After confirming that it exist correlation between variables and that there is no multicollinearity problem, the t test confirms the significance of all the variables (at a level of 0.05), so the null hypothesis is rejected. In this sense we can confirm that the population means are not identical in both groups and therefore they are apt to be able to discriminate.

Once the variables to be used are determined, the next step consists of calculating the weights.

There are different techniques, although one of the most used is the use of the scoring coefficients obtained from the Factorial Analysis (Principal Component Analysis), because they are optimal linear combinations of the variables inside the factors.

The technique of Factorial Analysis allows reducing the dimension of the information to the maximum possible and eliminating the information that is not significant, detecting the underlying structure of the data series (Rúa et al., 2003) with the identification of the proportion of the explained variance by a latent variable called factor.

For its formal definition, and following the García et al (2000) manual, we start from a set of random variables with zero mean, on which we look for its reduction to a set of k factors (Fk) defined in the following way:

$$X_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1k}F_k + e_1$$

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \ldots + a_{pk}F_k + e_p$$

Where a_{p1} indicate the correlation between the variable and the corresponding common factor and are called factor loads, being useful to identify the function of each variable for the definition of the factors; while e_p are model errors (also defined as specific factors).

In a matrix form, the expression of the factorial model is defined as follows:

$$X = \Lambda F + U$$

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \Lambda = \begin{bmatrix} F_{11} & \cdots & F_{1k} \\ \vdots & \ddots & \vdots \\ F_{p1} & \cdots & F_{pk} \end{bmatrix} F = \begin{bmatrix} F_1 \\ \vdots \\ F_k \end{bmatrix} U = \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix}$$

Where X is the data matrix, Λ the factorial loading matrix and F and U the unobservable variables.

For calculation of the factors, García et al (2000) indicate various procedures (principal components, main factor, minimum residues, maximum likelihood, among others) although all of them try to find a matrix of factorial weights such that multiplying it by its transpose, it is obtained the reduced correlation matrix.

Table 4 shows the results of the dimension reduction process, observing the existence of two factors that jointly explain 75.17% of the total variance, with a KMO greater than 0.5 and a PEB below 0.05.

In Table 4 shows the scoring coefficients of each variable in both Factors. Those coefficients will be used as weight of each variable for the construction of the synthetic indicator.

TABLE 4: SCORE MATRIX, EXPLAINED VARIANCE. KMO AND PEB TEST.

	COMPONENT	
	1	2
TA	,367	-,132
MNGT	,192	-,041
C/I	-,115	,445
ROA	,116	-,428
FTEs	,348	-,091
Branches	,315	-,043
ROB	,099	-,428
% VARIANCE	56,598	18,581
Kaiser-Meyer-Olkin (KMO)	0,653	
Bartlett (PEB)		
	Chi-squared	14054,193
	gl	21
	Sig.	,000

SOURCE: OWN ELABORATION

Cabrero et al (1996) affirm that the first component captures enough variance to be an adequate representation of the explanatory variables. We follow this criterion for the construction of our synthetic index.

Finally the synthetic indicator is the sum of all explanatory variables weighted according to the scores (Table 4) following Estévez (2002).

There are different aggregation methods, although Schuschny et al (2009) indicates that the weighted arithmetic mean and the weighted geometric average are the most used. We chose the arithmetic mean method because it is simpler (Cabero et al, 1996).

The formal definition of the synthetic indicator (ISF_{CR}) is:

$$ISF_{CR} = a_1V_1 + a_2V_2 + a_3V_3 + a_4V_4 + a_5V_5 + a_6V_6 + a_7V_7$$

Where ISF_{CR} denotes the financial health indicator for a Rural Savings Bank; Vi denotes the selected explanatory variables, a_i denotes the weighting coefficient of variable i within the first component.

Applying the scores obtained from the factor analysis (Table 5), the ISF_{CR} is defined as follows:

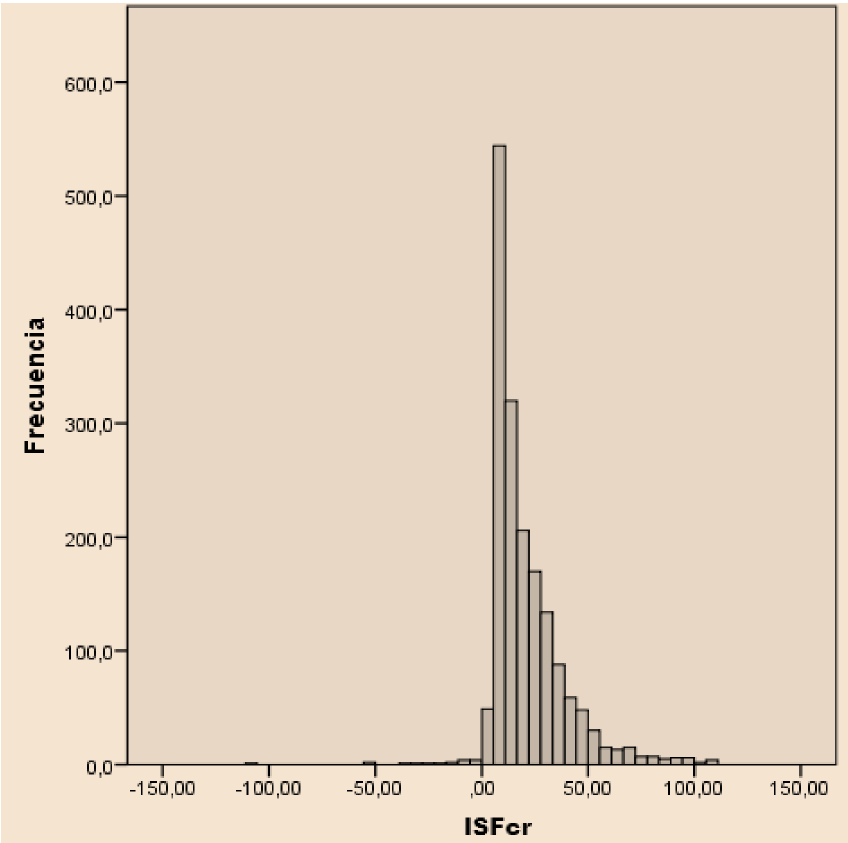
$$ISF_{CR} = 0{,}116_1 ROA_1 + 0{,}099_2 ROB_2 - 0{,}115_3 C/I_3 + 0{,}192_4 MNGT_4 + 0{,}367_5 TA_5 + 0{,}348_6 FTEs_6 + 0{,}315_7 Branches_7$$

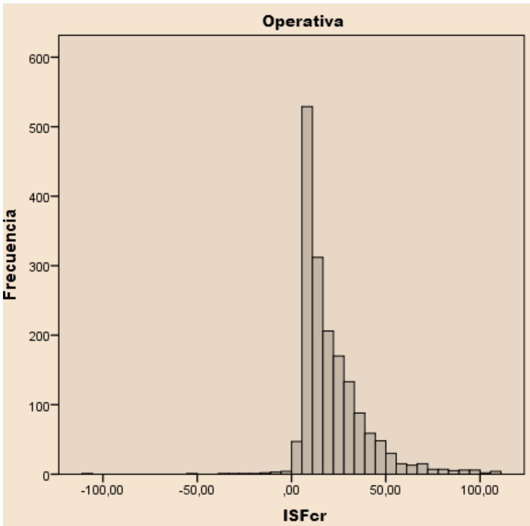
4. ANALYSIS RESULTS

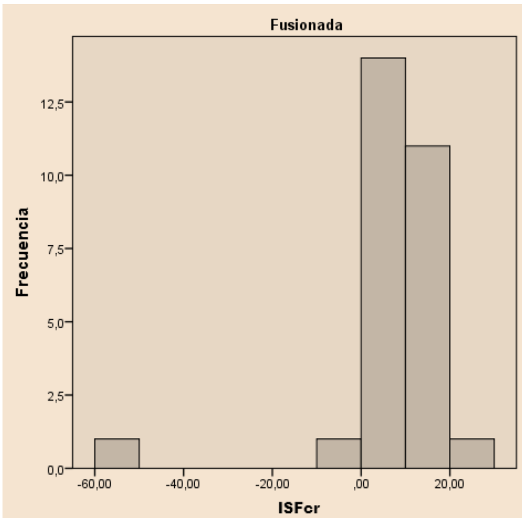
The functioning of the ISF_{CR} is simple and consists of assigning a score to each Rural bank according to the values taken by the explanatory variables in the financial statements of the previous year, taking into account that the higher the score obtained, the better the financial health of the Rural Bank analysed.

After calculating the ISF_{CR} for the set of Rural Credit cooperatives that make up the validation sample, Figure 2 shows a histogram of the scores of all of them (upper histogram), of the operational Rural Credit cooperatives (lower left histogram) and of the Rural Banks that ceased activity (lower right histogram) with concentration in the lower range of 50 points. On the other hand, scores below zero and those above 100 points are the least significant.

FIGURE 2: DISTRIBUTION OF THE SCORES ASSIGNED BY THE ISF_{CR}





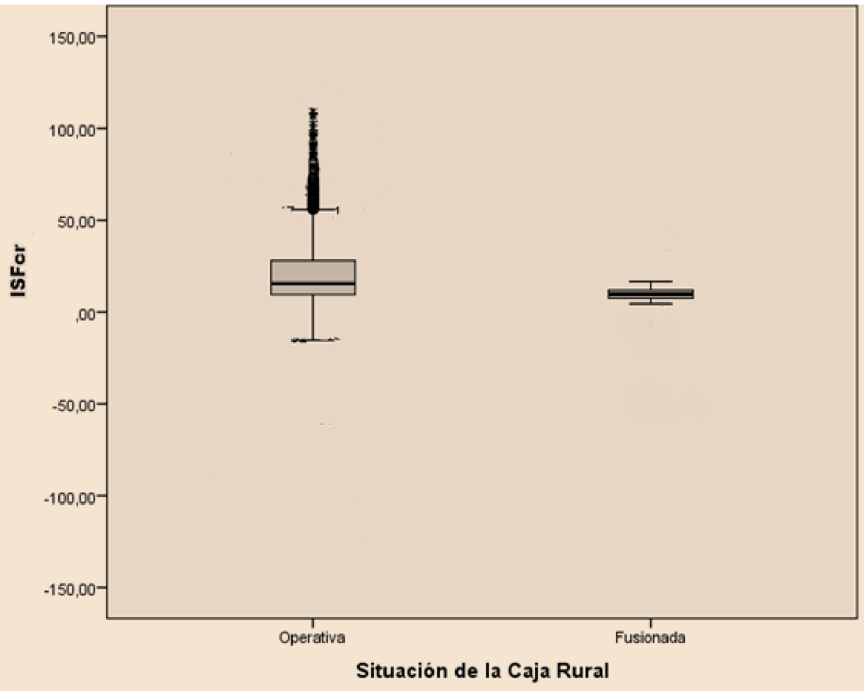


SOURCE: OWN ELABORATION

The question arises as to whether the ISF_{CR} has sufficient capacity to differentiate the operating Rural Credit cooperatives from those that ceased activity due to solvency reasons, therefore a graphic analysis and a t-test of equality of means will be carried out.

Regarding the graphical analysis, Figure 3 shows a box-plot of the ISF_{CR} for the operational Rural Credit cooperatives (first column) and those that ceased activity (second column), being able to observe how this second group presented, in term medium, ISF_{CR} scores, lower and further than those obtained by the group of operations, with less dispersion, which a priori allows to affirm the power of discrimination of the ISF_{CR} between both groups of Rural Credit cooperatives.

FIGURE 3: DISTRIBUTION OF THE SCORES ASSIGNED BY THE ISF_{CR} DIFFERENTIATING BETWEEN RURAL OPERATING BANKS AND RURAL BANKS THAT CEASED ACTIVITY.



SOURCE: OWN ELABORATION

In order to check the results of the graphical analysis, a t-test of equality of means was then carried out between the group of operative Rural Credit cooperatives and the group that ceased activity. Given that the p-value is less than 0.05 (Table 5), the null hypothesis that the population means are identical in both groups is rejected, confirming the discrimination power of the ISF_{CR}, observed in the graphical analysis.

TABLE 5: TEST AND DESCRIPTIVE STATISTICS

SITUATION OF THE CAJA RURAL		ESTATISTIC
Operative	Average	21,14

	Minimum	-105,75
	Maximum	110,27
	95% confidence interval (lower limit)	20,3186
	95% confidence interval (upper limit)	21,9775
Merged	Average	7,93
	Minimum	-51,79
	Maximum	29,13
	95% confidence interval (lower limit)	2,8943
	95% confidence interval (upper limit)	12,9732

t-Test	t	Sig.
Between groups (operative/merged)	-3,972	0,00000

SOURCE: OWN ELABORATION

In fact, on average, Table 5 shows that the operating Rural Credit cooperatives reached an ISF_{CR} score of 21.1 points, much higher than the 7.9 points collected by the set that ceased activity, although with a minimum score in the group of operations that was much higher (-105.6 points) than that recorded by the group of Rural Credit cooperatives that ceased activity (-51.7).

The confidence intervals in both groups are different. Specifically, in the case of the Rural Credit cooperatives that remained operative, the population value is within the limits of 20.3 and 21.9 points at 95% probability. On the contrary, in the case of the Rural Credit cooperatives that ceased activity, the confidence interval is delimited by the lower limit of 2.89 and the upper limit of 12.9.

To assess the predictive capacity of the ISF_{CR}, we used the confidence intervals, so that if a Rural Credit cooperative presents an ISF_{CR} within the confidence interval delimited for the operating Rural Credit cooperatives, we classify it within this group, and vice versa.

From this we observe that the ISF_{CR} correctly classifies 82% of the Rural Credit cooperatives that ceased activity due to solvency reasons, since it assigns them scores below 13 points that would fall within the confidence interval delimited for this group.

On the contrary, a priori the ISF_{CR} correctly classifies 58% of the Rural Credit cooperatives, assigning them scores higher than 13 points. However, analysing these cases in detail, we observe that a large part of the Rural Credit cooperatives that have been classified outside the operational group correspond effectively to Rural Credit cooperatives that years later effectively ceased activity. In fact, eliminating this group of Rural Credit cooperatives the degree of success in the classification of operating entities improves up to 76%.

Albulescu (2010) showed that one of the advantages of using synthetic indicators is the ability to make comparisons between different time intervals.

For this reason, Figure 4 shows the time evolution of the average ISF_{CR} scores obtained by the set of Rural Credit cooperatives in each of the years under study. This figure shows a growing trend in the average score awarded by the ISF_{CR} until 2007, when it reached a maximum of 32 points. However, as of that year, there was a steep fall to the minimum of 13 points recorded in 2013, and subsequently a timid recovery with values that were still far from those collected in the nineties of the 20th century.

In fact, it can be observed that the ISF_{CR} shows a behaviour very similar to that of the economic crises registered by the Spanish economy in the period 1991-2015, to the point that the setbacks of the ISF_{CR} are simultaneous to the different periods of recession, highlighting the crisis of 1992, the one of the year 2000 and the one initiated in 2008.

FIGURE 4: TEMPORARY DISTRIBUTION OF THE ISF_{CR}

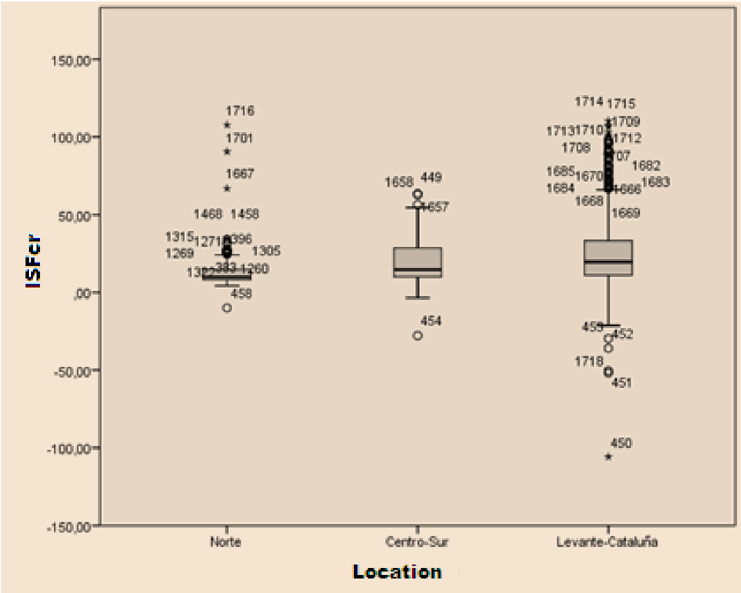
SOURCE: OWN ELABORATION

The logit models used for the selection of the explanatory variables highlighted the importance of a qualitative variable that considers the geographic location of each of the Rural Credit cooperatives. Therefore, we have considered it convenient to observe if, in the case of the ISF_{CR}, the geographic location is a differentiating factor in the scores awarded.

Figure 5 shows the ISF_{CR} scores of the Rural Credit cooperatives as a function of geographic location, with the first column being the score obtained by the Rural Credit cooperatives located in the North; the second column the scores of the Rural Credit cooperatives located in the Center-South; and the third column the score obtained by the Rural Credit cooperatives located in the Levante-Catalonia.

Figure 5 shows a higher ISF_{CR} in the set of Rural Credit cooperatives located in the north (with the exception of three entities that have scores higher than the average of their group), while the set of Rural Credit cooperatives located in the center and the release the scores are more homogeneous, although in the last group the dispersion would be more accentuated.

FIGURE 5: RESULTS OF THE SYNTHETIC INDICATOR ACCORDING TO GEOGRAPHICAL LOCATION



SOURCE: OWN ELABORATION

Although the graphical method confirms the existence of differences in the ISF_{CR} according to geographic location, this analysis has been completed with an ANOVA test among the three groups defined by geographical location, and whose results have been collected in Table 6 that the p-value is lower than the confidence level of 0.05, the existence of differences depending on the location is confirmed. Pair to pair between both groups

TABLE 6: ISF_{CR} AND TEST DEPENDING ON GEOGRAPHICAL LOCATION

ISF_{CR}		PRUEBA ANOVA
Ubicación	**Media**	**Sig.***
Norte	12,8030	0,000
Centro-Sur	19,2415	
Levante-Cataluña	24,8489	

(*) Pair to pair between both groups

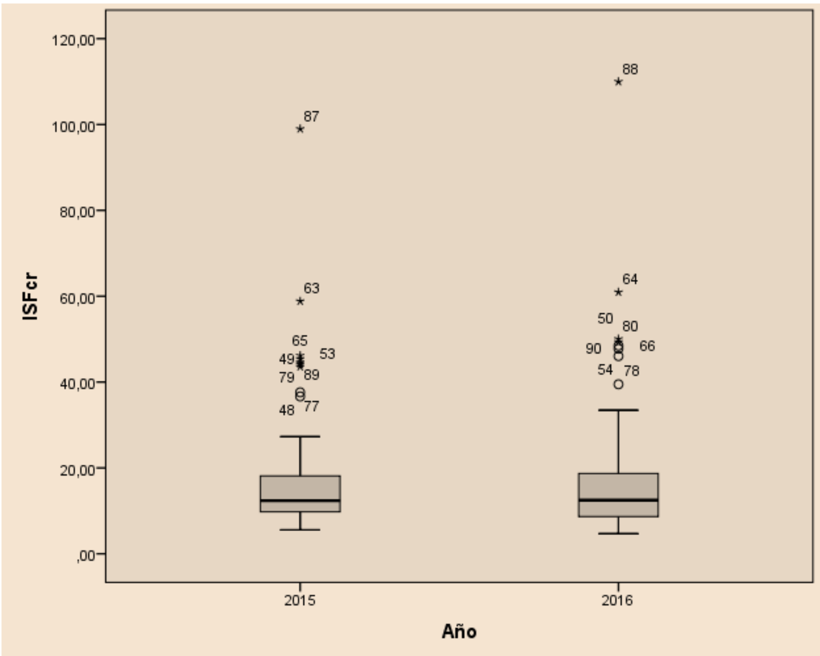
SOURCE: OWN ELABORATION

In fact, Table 6 shows that on average, the Rural Credit cooperatives located in the North presented an average score of 12.8 points, far from the average value of the Rural Credit cooperatives located in the Center-South and Levante-Catalonia, and of the historical average of 21 points collected in Table 6, the highest score being reached by the set of Rural Credit cooperatives located in Levante-Cataluña.

Once the results of the estimation sample have been analysed, the ISFCR will then be applied to the validation sample, composed of the operational Rural Credit cooperatives in 2015 and 2016, for which the ISF_{CR} has been calculated from the financial statements of the previous exercises.

According to FIGURE 6, in general terms the Rural Credit cooperatives presented in 2015 and 2016 an ISFCR close to 18 points, lower than the average of 22 points presented in the estimation sample, although with some recovery with respect to the data of the year of 2014.

FIGURE 6: ISF_{CR} IN THE VALIDATION SAMPLE



SOURCE: OWN ELABORATION

During this period there were no merger processes between the Rural Credit cooperatives, however, considering the average score that the Rural Credit cooperatives that ceased activity presented in the estimation sample, we observed that there is a group of sixteen Rural Credit cooperatives in which the score of the ISF_{CR} would be in the environment of the indicated average value.

From a detailed analysis of the Rural Credit cooperatives that present that low score in the ISF_{CR}, we observed that fourteen Rural Credit cooperatives are already part of an Institutional Protection System (SIP) among Rural Credit cooperatives. On the two remaining entities, one of them would have received financial assistance from other Rural Credit cooperatives without becoming part of a SIP.

5. DISCUSSION AND CONCLUSION

Our study proves that it is possible to measure and predict the financial health of rural Spanish credit cooperatives by means of a synthetic index which uses variables that measure profitability, efficiency, management performance and size.

Variables have been selected using relevant financial economic information drawn from Logit binary static and multi-period response models as proposed by Shumway (2001). Identifying the ratios has been possible thanks to the research done on bankruptcies in the banking sector, in accordance with the CAMEL and PEARL models as well as the key ratios applied in the sector. The variables have been normalized and aggregated to build the index as per the studies of Illing and Liu (2006) and Albulescu (2010).

The resulting synthetic index is comprised of seven components: profitability ratios (2), efficiency ratios (1), management ratios (1) and size ratios (3).

$$ISF_{CR} = 0{,}116_1 ROA_1 + 0{,}099_2 ROB_2 - 0{,}115_3 C/I_3 + 0{,}192_4 MNGT_4 + 0{,}367_5 TA_5 + 0{,}348_6 FTEs_6 + 0{,}315_7 Branches_7$$

These ratios are in line with the literature on bankruptcies in the financial sector regarding profitability ratios (Cabo 2010), efficiency (Lima 2012), management and business performance (Muñoz 1999), and size (Goddard et al. 2009; Garvalho et al. 2015). From our study, we conclude that size is of significant importance, particularly regarding assets, but also the number of employees and branches.

Our work differs from previous literature on the need for capital level and liquidity ratios to discriminate. Nonetheless, they are indirectly present as part of the management ratio relating credit assets to total assets, which affect both liquidity and solvency ratios. This ratio has been kept relatively low as rural credit cooperatives have maintained a strong client deposit base, and have received limited funding from the interbank market or the European Central Bank. Additionally, the fact that clients of rural credit cooperatives are at the same time "shareholders", has allowed for an agile capital base management.

We have defined the dependent variable as "merger process" in which an institution undergoing difficulties is absorbed or liquidated by other more solid institutions. Existing international literature on the rural credit cooperatives sector analyzes insolvency using "mergers" as a variable, which is the most common way to restructure the sector (Porath 2006; Cabo et al. 2010; Lima 2012). Our research has been based on 74 rural credit cooperatives that were in existence in 1991 and from which we gathered 1,745 financial statements from 1990 to 2016. Of these 74 institutions, 30 had financial problems and merged with other cooperatives, and the other 44 remained operative.

Our research shows that the synthetic index is able to successfully distinguish, ex post, between the rural credit cooperatives that merged due to financial difficulties (with 82% accuracy) from those that remained operative (with 76% accuracy).

Cooperatives that remained operational reached an average of 21.1 points on the ISFCR, substantially higher than those that needed a merger, with 7.9 points. 82% of cooperatives that ceased activities reached less than 13 points, which is the cut off point to detect problems.

The synthetic index shows that it is highly effective at distinguishing solvent cooperatives from those with financial problems, even when significant geographical differences are a factor. The index also corroborates the strong financial health of rural Spanish credit cooperatives (2015-2016).

In summary, this paper examines a new tool to predict the future financial health of rural Spanish credit cooperatives with a high level of accuracy which employs a synthetic index that predicts the probability that a business will succeed or disappear. It provides managers, supervisors and regulators with a practical tool to monitor the financial health of rural credit cooperatives on an ongoing basis, and allows them to act accordingly to prevent potential problems in the future.

REFERENCES

Albulescu, C. T. (2010). Forecasting the Romanian financial system stability using a stochastic simulation model. Romanian journal of economic Forecasting, 13(1), 81-98.Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal Of Finance*, *23*(4), 589-609.

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal Of Finance*, *23*(4), 589-609.

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of accounting research, 71-111.

Beaver, W. (1968). Alternative accounting measures as predictors of failures. *The Accounting Review*, *43*(1), 113-122.

Braga, M., Fully, V., Colosimo, E., & 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.

Cabo, P., & 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.,& Delrieu, J. (1996). Elaboración de un índice sintético para predecir la inflación en España. *Banco De España - Working Papers*, *9619*.

Carvalho, F., Diaz, M., Bialoskorski Neto, S., & Kalatzis, A. (2015). Exit and Failure of Credit Unions in Brazil: A Risk Analysis. *Revista Contabilidade y Finanças*, *26*(67), 70-84.

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*.

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.

Estrada, D. & Morales, M. A., (2010). A financial stability index for Colombia. Annals of Finance, 6(4), 555-581.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).

Fiordelisi, F., & Mare, D. (2013). Probability of default and efficiency in cooperative banking. *Journal Of International Financial Markets, Institutions And Money*, 26, 30-45.

Gama, V., Braga, M., & 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., & Resende, M. (2011). Avaliaçao de insolvência em cooperativas de crédito: uma aplicaçao do sistema PEARLS. *Revista De Administração Mackenzie*, 12(2), 113-144.

García, E., Gil, J. & Rodríguez, G. (2000). *Análisis Factorial.* Editorial La Muralla (Madrid).

Illing, M., & Liu, Y. (2003). An index of financial stress for Canada. Ottawa: Bank of Canada.

Laffarga, J., Martín, J., &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.

Lima, D. (2012). Forecasting insolvencies of Portuguese Co-operative banks. Third Prize In The First EACB Award Young Researchers On Co-Operative Banks.

Madera, A (2017). 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 (Doctoral dissertation, Universidad Pontificia Comillas).

Maggiolini, P.,& Mistrulli, P. (2005). A survival analysis of de novo co-operative credit banks. *Empirical Economics*, 30(2), 359-378.

Meyer, P., & Pifer, H. (1970). Prediction of Bank Failures. *The Journal Of Finance*, 25(4), 853.

Módica, A., Baixauli, J., & Álvarez, S. (2012). Propuesta de un indicador de salud financiera y su efecto en la predicción del fracaso empresarial. Re*vista Internacional Administración Y Finanzas*, 5(3).

Mondéjar, J., y Vargas, M. (2008). Indicadores sintéticos: una revisión de los métodos de agregación. *Economía, Sociedad y Territorio*, 8(27), 565-585.

Morris, V. (2010). Measuring and forecasting financial stability: The composition of an aggregate financial stability index for Jamaica. Bank of Jamaica.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.

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. & 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.,& Núñez, J. (2011). Estrategias para la elaboración de modelos estadísticos de regresión. *Revista Española De Cardiología*, 64(6), 501-507.

Palacios, 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.

Palomo, R., & Sanchís, J. (2010). 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.

Perez, C. (2009). *Técnicas de análisis de datos con SPSS 15*. Pearson Prentice Hall

Pille, P. (1998). Performance analysis of the Ontario Credit Unions (Tesis Doctoral). Universidad de Toronto.

Porath, D. (2006). Estimating probabilities of default for german saving banks and credit cooperatives. *Schmalenbach Business Review*, 58, 214-233.

Roget, F. M., Fernández, M. P. M., & 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.

Rúa, A., Redondo, R., y del Campo, C. (2003). Distribución municipal de la realidad socioeconómica gallega. *Revista Galega De Economía*, 12(2), 1-20.

Schuschny, A., & Soto, H. (2009). *Guía metodológica. Diseño de indicadores compuestos de desarrollo sostenible*. Naciones Unidas - Cepal.

Secrist, H. (1938). *National bank failures and non-failures: an autopsy and diagnosis*. Principia Press.

Sere-Ejembi, A., Udom, I. S., Salihu, A., Atoi, N. V., & Yaaba, B. N. (2014). Developing banking system stability index for Nigeria. CBN Journal of Applied Statistics, 5(1), 49-

77.Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal Of Business*, 74(1), 101-124.

Tascón, M., & 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.

Secrist, H. (1938). *National bank failures and non-failures: an autopsy and diagnosis*. Principia Press.

Simon, C.J. (1980). Predicting the failure of credit unions: an application of multivariate logit analysis. (Tesis de Master), Massachusetts Institute of Technology.