Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA. **Rect@**
Volumen 20(2019). Páginas 131 a 152.
DOI: 10.24309/recta.2019.20.2.02

BAYESIAN NETWORKS TO PREDICT FINANCIAL DISTRESS IN SPANISH BANKING

JESSICA PAULE-VIANEZ
jessica.paule@urjc.es
Universidad Rey Juan Carlos, Departamento de Economía de la Empresa
Paseo de los Artilleros s/n 28032, Madrid (Spain)

JOSÉ PABLO ARIAS-NICOLÁS
jparias@unex.es
Universidad de Extremadura, Departamento de Matemáticas
Avenida de la Universidad s/n 10071, Cáceres (Spain)

JOSÉ LUIS COCA-PÉREZ
jlcocap@unex.es
Universidad de Extremadura, Departamento de Economía Financiera y Contabilidad
Avenida de la Universidad s/n 10071, Cáceres (Spain)

Recibido (28/04/2019)
Revisado (29/07/2019)
Aceptado (14/10/2019)

RESUMEN: Este trabajo desarrolla un modelo predictivo a corto plazo del financial distress en el sistema bancario español con redes bayesianas. Como las quiebras de bancos han sido escasas, este documento ha considerado también otros problemas financieros, agrupados bajo el término financial distress, como son el incumplimiento de sus obligaciones, la necesidad de intervención de organismos externos, la ayuda estatal, las fusiones y adquisiciones con problemas, y las liquidaciones. Las variables utilizadas para predecir el financial distress en el sistema bancario español han sido variables financieras, clasificadas siguiendo el sistema de calificación de CAMELS, y variables económicas, cuya repercusión en la salud de estas entidades ha sido demostrada por diversos trabajos previos. Con una muestra de 148 instituciones bancarias, la alta tasa de aciertos obtenida demuestra que las redes bayesianas constituyen una metodología prometedora para predecir el financial distress a corto plazo en el sector bancario español.

Palabras Clave: Financial Distress, Modelo Bayesiano, Sector Bancario, Predicción, CAMELS, España.

ABSTRACT: This paper develops a short-term predictive model of financial distress in Spanish banking system with Bayesian networks. As bank failures have been scarce, this document has also considered other financial problems, encompassed under the term financial distress, such as non-compliance with its obligations, the need for intervention by external agencies, state aid, mergers and acquisitions with problems, and liquidations. The variables used to predict financial distress in the Spanish banking system have been financial variables, classified according to the CAMELS rating system, and economic variables, whose impact on the health of these entities has been demonstrated by several previous studies. With a sample of 148 banking institutions, the high success rate obtained shows that the Bayesian networks constitute a promising methodology for predicting short-term financial distress in the Spanish banking sector.

Keywords: Financial Distress, Bayesian model, Banking industry, Prediction, CAMELS, Spain.

1. Introduction

The financial crisis of 2007 and the bursting of the real estate bubble in Spain have greatly harmed the Spanish banking system, showing the need for greater control of these entities to avoid the consequences suffered by them in recent years, especially after the financial rescue of 24th July 2012, where the big banks of the country were affected.

The need to maintain a healthy banking system has motivated the establishment of various measures to improve its solvency, highlighting the Basel Agreements[1], at international level. In Spain highlight the Fund for Orderly Bank Restructuring (FROB), which was established through Royal Decree-Law 9/2009[2] on the restructuring and reinforcement of the own resources of credit institutions. The objective of this institution is to enable the restructuring process when the reinforcing of their own resources is not possible through private financing or the Deposit Guarantee Fund (DGF) for institutions in difficulty, and to enable voluntary mergers and to allow Saving Banks to become banks to establish viable entities.

The study of financial problems in the Spanish banking system, known as financial distress, has become a matter of vital importance. Avoiding the kind of events that occurred in recent years and strengthening the banking system belong to the key issues for the Spanish economic recovery. Therefore, in this paper we decided to focus our study on financial distress in credit institutions. Its early warning is a crucial research field for corporate finance whose core is predicting financial problems (Sun et al., 2014[3]).

Predicting methods of financial distress can be classified into statistical methods and artificial intelligence (AI) methods. In this work we have focused on the methods of artificial intelligence because several works have demonstrated, in their practical application, a greater predictive power of these methods over statistics for the prediction of financial distress (Bell et al., 1990[4]; Odom and Sharda, 1990[5]; Serrano and Martín, 1993[6]; Jo and Han, 1997[7]; McKee and Greenstein, 2000[8]; Shin et al., 2005[9]; Johnson, 2005[10]; Min and Lee, 2005[11]; Van Gestel et al., 2006[12]; Angelini et al., 2007[13]; Lin, 2009[14]; Chen and Du, 2009[15]; Etemadi et al., 2009[16]; Li et al., 2011[17]; Rafiei et al., 2011[18]; Lee and Choi, 2013[19]; Slavici et al., 2016[20]). Specifically, in this study we opted for the use of Bayesian networks because it is a promising method for predicting financial problems (Sarkar and Sriram, 2001[21]; Sun and Shenoy, 2007[22]; Aghaie and Saeedi, 2009[23]).

The objective of this work is to study the Bayesian networks application in predicting financial distress for Spanish credit institutions, including under the term of financial distress, various situations, which, as a whole, have not been collected by any previous study, being the same: bankruptcy, non-compliance of payment, intervention of the Deposit Guarantee Fund (FGD), the absorption of the entity or part of its assets, the merger with problems and state economic aid. To predict these problems, two different groups of variables have been used: financial variables based on CAMELS[1] system (Thomson, 1991[24]; Cole and Gunther, 1995[25],1998[26]; Bongini et al., 2001[27]; Poghosyan and Cihak, 2009[28]; Roman and Sargu, 2013[29]; Betz et al., 2014[30]; Wanke et al., 2015[31]; Constantin et al., 2018[32]) and macroeconomic variables (González-Hermosillo, 1999[33]; Curry et al., 2007[34]; Betz et al., 2014[30]; Constantin et al., 2018[32]). The period taken into account for this study goes from 2012 to 2016.

To carry out our purpose, this work has the following structure:

In the first place, we will analyse the literature based on financial difficulties and the different methodologies applied to predict the distress from its initial stages. After that, we will explain the

[1]This framework has its origin in the CAMEL rating system, created in 1979 in the United States by the banking regulatory agencies to assess the soundness and safety of banks and which included as variables the capital adequacy, the asset quality, the management, profits and liquidity, evolving this system towards CAMELS in 1996 incorporating risk sensitivity.

chosen methodology and the data used, to then perform the empirical test and finalize with the conclusions obtained.

2. Financial distress definition

There are several definitions for financial distress including different financial situations. During the last 50 years, numerous investigations has been focused on this topic, which in general can be said to be the situation in which a company is certain of some type of financial difficulty (Sun et al., 2014[3]). Beaver (1966)[35] stated that financial distress is based on the theoretical framework of cash flow models, by performing a comparison between water deposits and companies in financial difficulties, as resources that were draining away. Carmichael (1972)[36] defined financial distress as the situation where a company cannot meet its obligations. He includes within that frustration: insufficient liquidity, insufficient capital, non-payment of debt and insufficient liquid capital. Foster (1986)[37] defines this concept as a real liquidity problem that only must be solved through a wide-scale restructuring of operations or the economic entity structure.

Before the 90s, financial distress definitions referred to the inability of paying off and its consequences. Doumpos and Zopounidis (1999)[38] added to this definition the situation where a company has a negative net asset value. Ross et al. (1999)[39] put all these definitions together establishing four kinds of financial problems: business failure, legal bankruptcy, technical bankruptcy and accounting bankruptcy. However, Bose (2006)[40] deviates from the previous definitions and explains financial distress as the point where a listed company provides a quote lower than 10 USD cents. This definition is also supported by Ravisankar et al. (2010)[41]. Altman and Hotchkiss (2006)[42] based the definition of financial distress on concepts like failure, insolvency and bankruptcy. Lin (2009)[14] described financial distress as the company's inability to meet its financial obligations as they grow. The following situations determine if a company finds itself in this situation: bankruptcy, default, overdrafts of bank deposits, corporate events that may not allow the company to pay their debts at maturity, being under bankruptcy proceedings or a significant decline of the asset value below the minimum required. Chen and Du (2009)[15] explained financial distress occurs when a company has to face significant losses or when it must declare bankruptcy with much higher liabilities than its assets.

Focusing on the definition for financial distress in the banking sector, Betz et al. (2014)[30] also enclose similar situations like bankruptcy, default and the condition in which the company cannot meet its obligations to its creditors. In addition, Betz et al. (2014)[30] included another situation in which a company cannot perform its activity independently and they established a company is under financial distress when it receives a capital injection from the State, participates in a relief program or if it merges with another company. This definition is supported by Constantin et al. (2018)[32].

Based on the previous studies, financial distress can be defined as the solvency problems of a company in different levels, preventing it from exercising its activity without external aid and reducing its value, reaching bankruptcy in the most extreme case and, consequently, speeding up its market exit.

In this study, we used the previous definition of financial distress, which is essential to collect all these financial problems that a credit company has suffered or can suffer in the future. The State plays a very important controlling role in the banking system because of its influence on the economy. Unlike other sectors, this situation often leads to State's intervention in order to avoid big banks entities to face bankruptcy (too big to fail) and that is why there are only few banking entities went bankrupt. Thus, it is necessary to study a concept broader than bankruptcy in order to measure the banking entity "health".

3. Methodologies applied to financial distress predicting

First studies conducted on financial distress have focused on the classification of companies as "healthy" or "with failures" with descriptive methods based on a series of financial ratios (Fitzpatrick, 1932[43]; Smith and Winakor, 1935[44]; Merwin, 1942[45]).

In the mid-1960s, the predictive methods began to appear. Beaver (1966)[35] applied the univariate analysis in order to predict credit risk assuming some cut-off values for specific return, liquidity and solvency variables. From the contribution of Beaver (1966)[35], applying different methodologies for predicting financial problems began to become widespread highlighting multivariate statistical. Altman (1968)[46] was the first to predict bankruptcy through the multiple discriminant analysis with its Z-score model. This model showed to obtain a better prediction than the one of the univariate models. Deaking (1972)[47] aligns itself with this model by using the ratios of Beaver (1966)[35] and confirming this methodology is suitable for predicting business failure up to three years in advance. Considering the dependent variable for the multiple discriminant analysis assumed to be a continuous one, there is a conflict because the probability of a company presenting financial problems is between 0 and 1 (Sun et al., 2014[3]). That is why new models emerged like logit and probit models applied by Ohlson (1980)[48] and Zmijewski (1984)[49] respectively. The main difference between both models is the functional form for the probability.

In the 90s, with computer technology advancing so rapidly, some artificial intelligence methods started to be popular for predicting financial problems. Bell et al. (1990)[4] was the first applying this methodology for predicting financial distress using artificial neural networks.

Artificial neural networks is a methodology that replicate information processing mechanisms in nervous system of biological organisms and they should be effective and efficient (Johnson, 2005[10]). Bell et al. (1990)[4] compared the obtained results with the artificial neural networks and the logistic regression showing the first methodology was superior. Odom and Sharda (1990)[5] also applied this methodology for predicting financial problems and they compared neural networks with the multiple discriminant analysis, confirming the artificial neural networks have a higher capacity for predicting bankruptcy. Other authors that implemented this methodology were, among others, Serrano and Martín (1993)[6] and Lee and Choi (2013)[19].

Another technique used to predict financial distress is the data mining, also known as "found knowledge in databases". It consists of showing significant patterns in large databases (Han and Kamber, 2001[50]). Chen and Du (2009)[15] state that together with artificial neural networks, data mining techniques present satisfactory results in predicting financial distress. However, neural network 'back-propagation' gets better predictions than data mining. Decision trees are important techniques in predicting financial distress considered within data mining. Frydman et al. (1985)[51] initiated this methodology in this area using a recursive partitioning method, which compared to the multiple discriminant analysis and concluded the best option was to combine both methods. McKee and Greenstein (2000)[8] also supported the recursive partitioning method. They compared it with the logit model and the neural networks concluding the decision trees was the best method. Another technique to be considered within artificial intelligence methods is the expert systems. These technique aims to emulate the decision-making ability of an expert in order to solve problems. One of the most important expert methods is the case-based reasoning. It has the benefit in the prediction of financial distress of being easy to understand and it has a good performance when there is not enough data. Jo and Han (1997)[7] carried out a study that compared the case-based reasoning method, neural networks and the multiple discriminant analysis. This study concluded the neural networks are more efficient than the other methods although there were no significant differences in the results obtained with neural networks method and case-based reasoning method. Park and Han (2002)[52], Li and Sun (2008) and Li et al. (2011)[53] also used this method for predicting financial distress. Other kind of expert systems methods used for predicting financial distress were the predicting models based on the Bayesian learning model. Sarkar and Sriram (2001)[21] introduced the Bayesian networks in predicting banking failure using financial ratios to get the early warning of the failure. Two models were evaluated: a simple one with more assumptions than the other one that was more complex. They concluded both are effective in predicting an early warning of banking failures and the complex model could be a useful technique to audit these financial entities. Sun and Shenoy (2007)[22] proposed to use Bayesian networks for predicting bankruptcy and showed their suitability for the decision-making process. Other authors that used Bayesian networks for predicting financial problems are Aghaie and Saeedi (2009)[23] who compared Bayesian networks with logistic regression in predicting bankruptcy. They concluded Bayesian networks obtained better results. Heuristic techniques based on cooperative work and on GDSS systems were also developed for predicting financial distress (Sun and Li, 2009[54]) within decision support systems.

The evolutionary algorithms are another methodology used in artificial intelligence for predicting financial distress. This method uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination and selection. Varetto (1998)[55] was the first that applied evolutionary algorithms to extract lineal functions with no statistical restrictions and its respective discriminant rules. However, this method did not show better results than the multiple discriminant analysis in predicting financial distress. Shin and Lee (2002)[56] used evolutionary algorithms to establish cut points of financial ratios to obtain rules that can predict business failure. Etemadi et al. (2009)[16] showed how evolutionary algorithms were better than multiple discriminant analysis through the McNemar test (McNemar, 1947[57]). Rafiei et al. (2011)[18] studied the application of genetic algorithms, neural networks and multiple discriminant analysis for the prediction of financial distress. They found genetic algorithms obtained better results in predicting financial distress with genetic algorithms than with multiple discriminant analysis. However, the most accurate results were obtained with the neural networks method.

The rough set is another artificial intelligence method that has been used for predicting financial distress. Dimitras et al. (1999)[58], McKee (2000)[59] and Bose (2006)[40] applied this method. Another relatively new artificial intelligence method is the support vector machine that try to solve classification problems based on risk minimization. Shin et al. (2005)[9] studied bankruptcy comparing support vector machine with artificial neural networks and showed the first method was better in smaller samples. Min and Lee (2005)[11] compared the support vector machine with the multiple discriminant analysis, the logit model and the artificial neural networks, and concluded support vector machine were the most appropriate method for predicting bankruptcy. Van Gestel et al. (2006)[12] used the least squares classifiers of the support vector machine known as Fisher kernel discriminant analysis. This function is based on Bayesian inference and its results are better than the ones obtained with discriminate analysis or logistic regression when predicting credit risk.

Other highlighted models are based on the theory of fuzzy set theory and fuzzy logic (Slowinski and Zopounidis, 1995[60]; McKee and Lensberg, 2002[61]), and the data envelopment analysis (DEA) (Paradi et al., 2004[62]; Sueyoshi and Goto, 2009[63]; Wanke et al., 2015[31]; Li et al., 2017[64]).

4. Methodology

The chosen methodology for predicting financial distress in the banking system in this work is the Bayesian networks due to its big capacity to establish relationships between different variables, being able to report these in probabilistic terms. This methodology does not need the fulfilment of previous hypotheses for its application. Bayesian networks are commonly used because they provide significant support in the decision making process showing the conditional dependencies of the variables involved in the problem (Castillo et al., 1997[65]; Neapolitan, 2004[66]; Korb and Nicholson, 2003[67]; Jensen, 2001[68]; Calle, 2014[69]).

The concept "Bayesian network" was introduced by Pearl (1985)[70] and it was included in artificial intelligence within the expert systems defined by Stevens (1984, p.40)[71] like: 'machines with thinking and reasoning skills like an expert would do in a certain filed [...] a real expert system not only processes large amounts of data but it also manages the data so that the obtained results are intelligible and useful to answer questions even they have not been determined yet'. According to Castillo et al. (1997, p.3)[65] 'an expert system is similar to a computer system because they also emulate human experts in an area of specialization'.

Bayesian networks represent a distribution function of a finite set of variables (Tuya et al., 2007[72]) and that is why this methodology is very useful not only due to the relational structure obtained but also because it provides probability distribution. This allows us to calculate marginal probabilities and update them depending on the provided evidences to the network (Calle, 2014[69]). Thus, Bayesian networks provide an inference system where new information on the variables is updated in its probability tables and this also spreads to other variables if they are interconnected (Tuya et al., 2007[72]).

Bayesian networks are divided into two parts:

- The qualitative part is represented with a Directed Acyclic Graph (DAG) that describes the variables and their dependencies.
- The quantitative part is defined by parameters establishing the relationships through conditional probabilities that represent the problem's uncertainty.

In order to understand how Bayesian networks work, it is essential to mention Bayes theorem.

Bayes theorem gives a solution to the obtaining of a posteriori probability, for this we will define the following expressions (Beltrán et al., 2014[73]):

- $P(y)$: prior probability that the hypothesis y is fulfilled without taking into account other observed data x.
- $P(y/x)$: posterior probability that the hypothesis y is fulfilled once the data x is known. It represents the influence of the data on the hypothesis y.
- $P(x/y)$: probability of the data x has been observed once the hypothesis y is true, known as the likelihood function.

If we consider the following probability distribution:

$$P(y \cap x) = P(y) \cdot P(x/y) = P(x) \cdot P(y/x) \tag{1}$$

Posterior probability is:

$$P(y/x) = \frac{P(y \cap x)}{P(x)} = \frac{P(y) \cdot P(x/y)}{P(x)} \tag{2}$$

This would be the function used in case there are two nodes (variables), being x the predicting variable and y the response variable.

Figure 1 (the graph) shows how nodes represent random variables of the set $x_1, x_2, \ldots, x_5$ the explanatory or predicting variables and y the response variable. The arcs represent the dependency relationships between the variables, relationships in which, the parent variables send the information to the child variables.

Bayesian networks assume nodes depend direct from its parent nodes and each node is linked to a conditional probability table that defines the probability of each variable's status given their possible parent nodes status (Tuya et al., 2007[72]).

$$P(x_1, x_2, \ldots, x_n, y) = P(y/parent(y)) \prod_{i=1,2,\ldots,n}^{n} P(x_i/parent(x_i)).$$

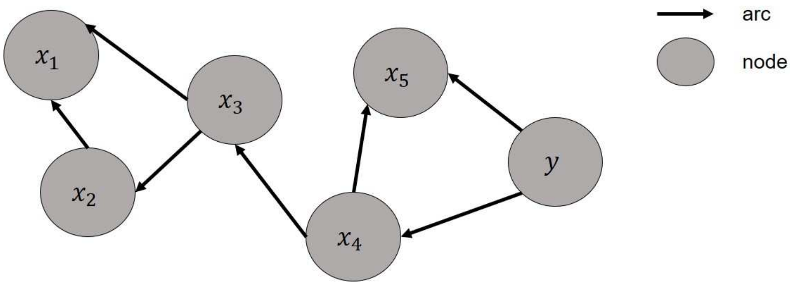


Figure 1. Bayesian network structure.

In order to get the optimal network, we should collect all possible networks, however, firstly we should differentiate two types of learning (Castillo et al., 1997[65]):

- Structure learning: learning the graphic structure (dependency).
- Parameter learning: learning the parametric structure (probabilities).

These two types of learning are composed of two elements (Castillo et al., 1997[65]):

- Quality measure: a set of Bayesian networks can be classified according to its quality. The quality of its network will be measured (graphic and parametric structure).
- Search algorithm: the best network from a high quality Bayesian networks subset will be selected.

Once the Bayesian network is selected and in order to predict the probability of an event according to a data set outside its network construction, the so-called classifiers come into play to obtain the most probable prediction for that data set.

Coming back to the Bayes theorem and having several predicting variables $(x_1, x_2, \ldots, x_n)$ the posterior probability is:

$$P(y/x_1, x_2, \ldots, x_n) = \frac{P(y)P(x_1, x_2, \ldots, x_n/y)}{P(x_1, x_2, \ldots, x_n)}. \tag{3}$$

The most accurate hypothesis for a data set will be the one presenting a higher posterior probability once the predicting variables are given (maximum a posteriori estimation MAP). The function is (Beltrán et al., 2014[73]):

$$y_{MAP} = \arg\max_{y \in \Omega_y} P(y/x_1, x_2, \ldots, x_n) = \arg\max_{y \in \Omega_y} \frac{P(y)P(x_1, x_2, \ldots, x_n/y)}{P(x_1, x_2, \ldots, x_n)} \tag{4}$$

Where Ω_y represents the data set that can be taken by variable y.

$P(x_1, x_2, \ldots, x_n)$ will be removed from the function because it remains unchanged for all categories of the response variable. Thus, we obtain the optimum network for our model representing the maximum a posteriori probability like:

$$y_{MAP} = \arg\max_{y \in \Omega_y} P(y)P(x_1, x_2, \ldots, x_n/y) \tag{5}$$

5. Empirical test

5.1. *Model variables*

The response variable represents whether a financial institution is or is not under financial distress in the short term. That is why it is essential to determine when can we classify a financial entity under financial distress. In this work we establish a credit institution will be under financial distress when the following situations occur:

(i) **The financial entity has declared bankruptcy.** This is one of the most important financial problems that can affect a company. This is the most studied situation in predicting financial distress in an independent way (Serrano and Martín, 1993[6]; Bongini et al., 2001[27]; Betz et al., 2014[30]; Constantin et al., 2018[32]).

(ii) **The financial institution has not met its payment obligations, or it has delayed them.** Default payment is a clear sign of poor liquidity to meet its obligations (Angelini et al., 2007[13]; Curry et al., 2007[34]; Betz et al., 2014[30]; Constantins et al., 2018[32]).

(iii) **The financial entity needs to be rescued by the Deposit Guarantee Fund (DGF).** This is a clear sign that shows the entity is unable to face its obligations to its clients (Laffarga et al., 1985[74]; Pina, 1989[75]). Bell et al. (1990)[4], Thomson (1991)[24] and Cole and Gunther (1995, 1998)[25,26] include insured banks requiring payments from the Federal Deposit Insurance Corporation (FDIC).

(iv) **The financial institution has been partly or totally acquired by another institution.** When a company is absorbed by another company or part of its assets have been absorbed, it means the institution does not work properly independently or presents serious liquidity problems. Pina (1989)[75], Bongini et al. (2001)[27] and González-Hermosillo (1999)[33] include banks that were absorbed by other bank or banks. Bell et al. (1990)[4] included banks whose deposits were taken over by other banks.

(v) **The financial institution has been merged with other with coverage ratio under 0.** A useful ratio in order to determine whether a company has merged with difficulties or not is the coverage ratio. (González-Hermosillo, 1999[33]). This shows if the merge has been forced because of financial problems or not. According to this situation, we determine a company is under financial distress if the coverage ratio is under 0 the year before the merge. The coverage ratio means the proportion between capital and loan reserves excluding impaired loans and total assets (Betz et al., 2014[30]; Constantin et al., 2018[32]). Other authors that include mergers are Bell et al. (1990)[4], Bongini et al. (2001)[27] and Curry et al. (2007)[34].

(vi) **The financial entity has received state aid in many ways.** State aid for restructuring (mainly by the FROB) or a company's rescue is a clear consequence of financial problems and a clear sign of being unable to thrive independently (Bell et al., 1990[4]; Bongini et al., 2001[27]; Betz et al., 2014[30]; Constantin et al., 2018[32]).

Financial distress can be defined as the financial problems a company can suffer that make it unable to meet its obligations independently which derives into the need of external aids in order to continue performing its activity. These solutions can be mergers, acquisitions, interventions by consumer protection organizations or state aid, the most extreme case of financial distress being the bankruptcy of the credit institution. In order to get the needed data to establish the different situations of financial distress that have been previously explained, we needed to go to the following sources (Table 1).

Based on this, the response variable can take two different results: the institution is under financial distress if one of the previous situations occurs or the institution is not under financial distress if none of the previous situations occur.

Regarding the explanatory variables of the model, we have selected several financial ratios within the CAMELS framework whose parameters are indicators to assess the financial strength of a bank (Roman and Sargu, 2013[29]). Many authors have used these variables in their research (Thomson, 1991[24]; Cole and Gunther, 1995[25],1998[26]; Bongini et al., 2001[27]; Poghosyan and Cihak, 2009[28]; Roman and Sargu, 2013[29]; Betz et al., 2014[30]; Wanke et al., 2015[31]; Constantin et al.,

Table 1. Sources used to define the response variable.

Source	Evidence
Orbis Bank Focus	Bankruptcy, absorption, mergers and coverage ratios
Datastream	Mergers and acquisitions
Deposit Guarantee Fund	DGF intervention
National Securities Market Commission	Deferred and unpaid coupons
European Commission	Public aid
Bank of Spain	Public aid

2018[32]), however, in addition to these variables, we have included macroeconomic variables in the model, since their influence on the financial problems of banking entities is proven (González-Hermosillo, 1999[33], Curry et al., 2007[34], Betz et al., 2014[30]; Constantin et al., 2018[32]). With all this, we have selected 52 explanatory variables to predict financial distress (Table 2).

5.2. *Sample*

In order to predict financial distress in Spanish credit institutions and to create a global model for the system, we have selected credit institutions from the three groups: banks, saving banks and credit unions.

Due to the availability of information we have done the study for the period 2012-2016, that is, we have studied the financial problems that have occurred in Spanish credit institutions in this period.

All banking institutions have been obtained from the database Orbis Bank Focus. We have restricted entities that are not classified as credit institutions according to the Bank of Spain and credit institutions for which data were not available. The sample we used had 148 banking institutions: 59 banks, 16 saving banks and 73 credit unions.

Given that financial distress may exist at different times and that the aim of the study was to predict it in the short term, in designing the prediction model, the sample was organized by taking the explanatory variables as at 31 December of the previous year to that in which the entity was in a situation of financial distress, in order to predict whether or not it would be in financial distress during the next 12 months. For entities that were not in financial distress, we used the explanatory variables for the last year in which information was available. How some entities have had more than one financial distress situation, we obtained a total of 151 observations of the independent variable, of which 32 showed a situation of financial distress.

5.3. *Modeling*

Once the variables and the sample have been defined, we can proceed to construct the Bayesian network.

Because we have lost values in the sample, we decided to use the k-nearest neighbour method with the statistic program R 3.4.1. This method fills the missing data by means of a value obtained from related cases in the record set, presenting the advantage of its simplicity, ease of understanding and relatively high precision (Zhang, 2012[76]), where the number of nearest neighbours selected is 5.

In order to construct the network, as the variables used are continuous, we discretize the explanatory variables in 5 intervals of equal amplitude. The discretization of the variables we use Genie 2.2 software.

To avoid the over adjustment in the model, we used the k-fold cross-validation. This divides the data set into equal k parts using random parts for the training $k-1$ and leaving the last part

Table 2. Explanatory variables of the model.

Type of variable	Variables used	Source
Capital	Common Equity Tier 1 capital ratio %	Orbis Bank Focus
	Ratio Tier 1 %	Orbis Bank Focus
	Total capital ratio %	Orbis Bank Focus
	Net equity/ total assets %	Orbis Bank Focus
	Ordinary capital/ tangible assets %	Orbis Bank Focus
	Common CET1 growth %	Orbis Bank Focus
Assets	Loans / total assets %	Orbis Bank Focus
	Total asset growth %	Orbis Bank Focus
	Total loan growth %	Orbis Bank Focus
	Impaired loans / gross loans %	Orbis Bank Focus
	Impaired loans + mortgaged assets / gross loans + mortgaged assets %	Orbis Bank Focus
	Impaired loans overdue / gross loans in preceding year %	Orbis Bank Focus
	Impaired loans (including restructured loans and potentially difficult loans) / gross loans %	Orbis Bank Focus
	Loan losses reserve/ impaired loans %	Orbis Bank Focus
	Provisions for losses / net earnings from interest %	Orbis Bank Focus
	Provisions for losses on loans / average gross loans %	Orbis Bank Focus
	Charges for impaired loans and securities / operating profit prior to impairment %	Orbis Bank Focus
	Net charges / average gross loans %	Orbis Bank Focus
	Impaired loans / net equity %	Orbis Bank Focus
	Impaired loans without reserves / net equity %	Orbis Bank Focus
Management	Cost / income ratio %	Orbis Bank Focus
	Average cost of assets ratio %	Orbis Bank Focus
	Client deposit interest expenses / average client deposits %	Orbis Bank Focus
	Interest expenses / average interest accrued on liabilities %	Orbis Bank Focus
Earnings	Financial return %	Orbis Bank Focus
	Operating profit / average net equity %	Orbis Bank Focus
	Economic return %	Orbis Bank Focus
	Ongoing earning capability / average total assets %	Orbis Bank Focus
	Net interest margin %	Orbis Bank Focus
	Earnings from interest / average interest earning assets %	Orbis Bank Focus
	Earnings without interest / operating revenue %	Orbis Bank Focus
	Earnings from interest / average gross loans %	Orbis Bank Focus
	Earnings from interest / operating revenue %	Orbis Bank Focus
Liquidity	Liquid assets / total assets %	Orbis Bank Focus
	Loans with less then a 1-year maturity / total loans %	Orbis Bank Focus
	Deposits with less than a 1-year maturity / total deposits %	Orbis Bank Focus
	Client loans / client deposits %	Orbis Bank Focus
	Interbank assets / interbank liabilities %	Orbis Bank Focus
	Minimal risk assets / total deposits and finance %	Orbis Bank Focus
	Liquid assets / deposits and finance %	Orbis Bank Focus
	Client deposits / total finance without derivatives %	Orbis Bank Focus
	Wholesale finance / total finance without derivatives %	Orbis Bank Focus
Sensibility	Asset reasonable value / total assets %	Orbis Bank Focus
	Level 3 assets / total securities %	Orbis Bank Focus
	Level 3 assets 3 / CET1 %	Orbis Bank Focus
	Earnings from commercial transactions / total operating revenue %	Orbis Bank Focus
Macro-economic variables	Yield on long-term Government bonds %	Datastream
	Unemployment rate %	INE [2]
	General price index variation %	INE
	Housing price index %	INE
	Mortgages on total property %	INE
	Gross Domestic Product %	INE

to be evaluated. This process will be repeated k times. In this work we have selected $k = 10$ so we have used 9/10 parts to be trained and 1/10 part to be evaluated.

To obtain the Bayesian network, we have decided to use the learning algorithm proposed by Friedman et al. (1997)[77] known as the Tree Augmented Naive (TAN).This algorithm results from an adaptation of the general Bayesian search algorithm proposed by Chow and Liu (1968)[78] beginning its application with a Naive Bayes structure to which connections are added between variables (except for the class variable that is the father of all) to take into account the possible dependence between them. According to Hernández et al. (2004)[79] this algorithm ensures the network structure obtained will have the maximum likelihood. With this network obtained with Genie 2.2 software, we explained the capacity to predict financial distress in the short term with different variable groups that belong to the CAMELS system as well as the macroeconomic variables. This will be carried out in each variable group in isolation so that we can evaluate the capacity of predicting the model with all variables globally.

5.4. *Results*

Firstly, we evaluated the capacity of predicting financial distress with Bayesian networks in the short term for Spanish credit institutions through the different characteristics. In Figure 2 you can see the networks obtained for each group of variables.

Relationship between the different variables for each group can be seen in Table A in Appendix section.

With the networks obtained, the ability to predict financial distress with Bayesian networks for each group can be seen in Table 3.

Table 3. Prediction for each variable group.

Type of variable	Percentage of correct prediction Financial Distress	Percentage of correct prediction No Financial Distress	Percentage of correct prediction Global
Capital	53.13%	88.24%	80.80%
Assets	78.13%	90.76%	88.08 %
Management	53.13%	94.12%	85.43%
Earnings	43.75%	89.08%	79.47%
Liquidity	50.00%	90.76%	82.12%
Sensibility	25.00%	88.24%	74.83%
Macroeconomic variables	56.25%	98.32%	89.40%

We observed how the quality of assets and macroeconomic variables had the highest hit rate in predicting financial distress, being the variables for the quality of assets the most important. We can confirm the quality of assets represents a key factor for the health of credit institutions, which means, bad quality of granted loans is decisive for a financial entity to be under financial distress in the short term. We observed this situation with the subprime loans which is a clear sign that providing these loans could be the main reason of the problems that occurred in credit institutions in the last years.

Regarding the macroeconomic variables, we can determine the economic environment has a strong impact on the financial situation of credit institutions, which means, the institutions that were under financial distress could have reached this situation not only because of internal activities but also because of its environment.

Sensitivity and Earnings are the features with the lowest capacity to predict financial distress in the short term. Risk exposure has had a low influence on credit institutions if we understand such risk as systematic risk, in other words, independent and not controlled by the company. External risk exposure has not been a relevant factor in the financial problems that credit institutions have suffered during the last years.

It is important to highlight how profit, although it has commonly used to predict financial problems, have had the second lower hit rate in this test. This means, results obtained from credit institutions are not relevant for their financial health situation.

We have evaluated the capacity of predicting financial distress in isolation, and that is why it is necessary to assess the capability of predicting financial distress in the short term with all variables from the different sets.

If we apply the TAN Bayesian network in order to predict financial distress in the short term in credit institutions in Spain with all the variables, we obtain the network in the Figure 3.

Relationships between the different variables can be seen in Table B in Appendix section. We can see how the explanatory variables stem from the response variable and how all explanatory variables are interconnected.

The obtained results in the network can be seen in the Table 4.

Table 4. Results from the global Bayesian network.

Observed	Prediction		
	No Financial Distress	Financial Distress	Percentage of correct prediction
No Financial Distress	116	3	97.48%
Financial Distress	5	27	84.38%
Global	80.13%	19.87%	94.70%

We observe how this network provide a higher hit rate, so that interconnections among the different variables of the different sets allow improving prediction.

The network that has been built has allowed us to obtain a global success probability of 94.70%, hitting on a 97.48% in predicting which entities will not be under financial distress in the short term and predicting correctly the 84.38% of all entities that will be under financial distress. It draws attention the increase of hits in this group, which shows how explanatory variables from different categories in financial entities allow predicting negative events for them.

The model has made the mistake type 1 (false positive), in other words, it has classified a entity as it will have no financial problems in the short term but that it will be under financial distress in the short term, with Caja de Ahorros and Monte de Piedad from Zaragoza, Aragón and Rioja (Ibercaja) obtained a 100% of certainty in prediction and for Cajamar Caja Rural S.C.C. and Montes de Piedad Caja de Ahorros of Ronda, Cádiz, Almería, Málaga, Antequera and Jaén (Unicaja) obtained a 95.6%.

The mistake type 2 (false negative), represents the classification of a entity under financial distress in the short term as 'healthy'. This occurred with Caja de Ahorros de Vitoria and Álava-Caja Vital and in UBS Bank with a certainty in prediction of 100% and also for Banco de Crédito Social Cooperativo with a 99.8% of certainty, for Caja de Ahorros de la Inmaculada de Aragón (Caja Inmaculada) with 97.8% and in Banco de Madrid with 67.3%.

In addition, and in order to demonstrate the validity and reliability of the models obtained by group and global, we use the area under the ROC curve and the Kappa statistic (Table 5).

Considering the area under the ROC curve, it is observed that in all cases it is above 60%, which demonstrates a good discriminative capacity of the Bayesian network models built, especially in the Global model, in which the low area the ROC curve is 96.6%. This means that the network presents the ability to differentiate companies that will be under financial difficulties and those that will not be with a probability of 96.61%.

Regarding the accuracy of the proposed models, it is observed following the Kappa statistic, how the most accurate models in their predictions are the Assets, Management and Macroeconomic

Table 5. Kappa statistic and area ROC of the proposed models.

Model	Kappa statistic	Area ROC
Capital	0.418	0.845
Assets	0.659	0.913
Management	0.520	0.866
Earnings	0.348	0.751
Liquidity	0.432	0.828
Sensibility	0.149	0.632
Macroeconomic variables	0.632	0.861
Global	0.838	0.966

and the Global model, reaching the latter a value of the Kappa statistic of 0.838.

The results of these indicators show that the short-term predictive model of financial distress in Spanish banking system built with Bayesian networks not only has a high predictive capacity (94.70%), but also has a high discriminatory power (0.966) and high accuracy in its predictions (0.838).

6. Conclusions

Financial distress prediction has been a key objective in the studies of credit institutions worldwide. After the events occurred during the last years, it is essential to create a methodology to anticipate the consequences a problem like that could have in credit institutions in a certain country.

Focusing on Spain and using a broad range of entities considered credit institutions, we have defined the following situations as determinants under financial distress: bankruptcy, payment default, the intervention by the Deposit Guarantee Fund, acquisitions, merger with problems and state aid, aiming to include the maximum situations in this concept.

Based on the literature, we have used as explanatory variables of this concept several ratios within the CAMELS system and also other macroeconomic variables due to its strong impact on the financial situation in credit institutions.

We have used Bayesian networks for predicting financial distress in the Spanish banking system. This is a promising technique for predicting processes and for decision-making processes, which is very popular nowadays.

With Bayesian networks TAN, we can confirm the explanatory variables of the quality of assets and the macroeconomic variables have been the most accurate in predicting financial distress in isolation. This shows the influence that real estate bubble had in credit institutions thanks to the subprime loans. On the other hand, explanatory variables based on sensibility to risk and profits had a lower capacity of predicting being this last group the most studied one in predicting financial distress from the very beginning.

With this methodology we have developed a network for all variables globally obtaining a hit rate of 94.70%. This confirms the interdependence between the variables of different groups manages to provide more information to predict the financial distress than the information provided by each of them separately. In addition, we have validated the network analyzing its discriminatory capacity and its precision, obtaining an area under the curve of 0.966 and a value of the Kappa statistic of 0.838.

This model provides a significant contribution to financial distress literature since it is the first model of Bayesian networks applied to financial problems in credit institutions in Spain. Furthermore, another significant contribution lies in how we studied all financial distress situations, which have not been contemplated in any previous research. This helps credit institutions to avoid those situations that make difficult for them to meet their obligations, so they can take measures to reduce the impact in advance or avoid it. Therefore, we believe the network we obtained constitutes a very useful and global technique for predicting financial distress in the short term.

References

1. Basel Committee on Banking Supervision (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. Bank for International Settlements.
2. Real Decreto-ley 9/2009, de 26 de junio, sobre reestructuración y reforzamiento de los recursos propios de las entidades de crédito. Boletín Oficial del Estado. Madrid, 28 de junio de 2009, núm 155.
3. J. Sun, H. Li, Q-H. Huang and K-Y. He, Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches, *Knowledge-Based Systems* **57** (2014) 41-56.
4. T. B. Bell, G. S. Ribar and J. Verchio, Neural nets versus logistic regression: a comparison of each models ability to predict commercial bank failures, in *X Auditing Symposium Deloitte & Touche* (Kansas: University of Kansas, 1990).
5. M. Odom and R. Sharda, A neural networks model for bankruptcy prediction, in *IJCNN International Joint Conference on neural networks* (San Diego, United States, 1990).
6. C. Serrano and B. Martín (1993). Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales, *Revista Espaola de Financiación y Contabilidad - Spanish Journal of Finance and Accounting* **23**(74) (1993) 153-176.
7. H. Jo and I. Han, Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis, *Expert Systems with Applications* **13**(2) (1997) 97-108.
8. T. E. McKee and M. Greenstein, Predicting bankruptcy using recursive partitioning and a realistically proportioned data set, *Journal of Forecasting* **19**(3) (2000) 219-230.
9. K-S. Shin, T. S. Lee and H-J. Kim, An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications* **28**(1) (2005) 127-135.
10. C. A. Johnson, Modelos de Alerta Temprana para pronosticar crisis bancarias: desde la extracción de seales a las redes neuronales, *Revista de Análisis Económico* **20**(1) (2005) 95-121.
11. J. H. Min and Y-C. Lee, Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters, *Expert Systems with Applications* **28**(4) (2005) 603-614.
12. T. Van Gestel, B. Baesens, J. Suykens, D. Poel, D. Baestaens and M. Willekens, Bayesian kernel based classification for financial distress detection, *European Journal of Operational Research* **172** (2006) 979-1003.
13. E. Angelini, G. Di Tollo and A. Roli, A neural network approach for credit risk evaluation, *The Quarterly Review of Economics and Finance* **48**(4) (2007) 733-755.
14. T-H. Lin, A cross model study of corporate financial distress prediction in Taiwan: multiple discriminant analysis, logit, probit and neural networks models, *Neurocomputing* **72**(16) (2009) 3507-3516.
15. W-S. Chen and Y-K. Du, Using neural networks and data mining techniques for the financial distress prediction model, *Expert System with Applications* **36**(2) (2009) 4075-4086.
16. H. Etemadi, A. A. A. Rostamy and H. F. Dehkordi, A genetic programming model for bankruptcy prediction: Empirical evidence from Iran, *Expert Systems with Applications***36**(2) (2009) 3199-3207.
17. H. Li, Y. C. Lee, Y. C. Zhou and J. Sun, The random subspace binary logit (RSBL) model for bankruptcy prediction, *Knowledge-Based Systems* **24**(8) (2011) 1380-1388.
18. F. M. Rafiei, S. M. Manzari and S. Bostanian, Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications* **38**(8) (2011) 10210-10217.
19. S. Lee and W. S. Choi, A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis, *Expert System with Applications* **40**(8) (2013) 2941-2946.
20. T. Slavici, S. Marris and M. Pirtea, Usage of artificial neural networks for optimal bankruptcy forecasting. Case study: Eastern European small manufacturing enterprises, *Quality & Quantity* **50**(1) (2016) 385-398.
21. S. Sarkar and R. S. Sriram, Bayesian models for early warning of bank failures, *Management Science* **47**(11) (2001) 1457-1475.
22. L. Sun and P. Shenoy, Using Bayesian networks for bankruptcy prediction: Some methodological issues, *European Journal of Operational Research* **180**(2) (2007) 738-753.
23. A. Aghaie and A. Saeedi, Using bayesian networks for bankruptcy prediction: Empirical evidence from

Iranian companies, in *2009 International Conference on Information Management and Engineering, ICIME* (Kuala Lumpur, Malaysia, 2009), pp. 450-455.
24. J. Thomson, Predicting Bank Failures in the 1980s, *Economic Review of the Federal Reserve Bank of Cleveland* **27**(1) (1991) 9-20.
25. R. A. Cole and Gunther, Separating the Likelihood and Timing of Bank Failure, *Journal of Banking and Finance* **19** (1995) 1073-1089.
26. R. Cole and Gunther, Predicting Bank Failures: A Comparison of On- and Off- Site Monitoring Systems, *Journal of Financial Services Research* **13**(2) (1998) 103-117.
27. P. Bongini, S. Claessens and G. Ferri, The Political Economy of Distress in East Asian Financial Institutions, *Journal of Financial Services Research* **19**(1) (2001) 5-25.
28. T. Poghosyan, and M. Cihak, Distress in European Banks: An Analysis Based on a New Data Set, *Journal of Financial Services Research* **40** (2009) 163-184.
29. A. Roman and A. C. Sargu, Analysing the Financial Soundness of the Commercial Banks in Romania: An Approach Based of the Camels Framework, *Procedia Economics and Finance* **6** (2013) 703-712.
30. F. Betz, S. Opric?, T. A. Peltonen and P. Sarlin, Predicting distress in European banks, *Journal of Banking and Finance* **45** (2014) 225-241.
31. P. Wanke, C. P. Barros and J. R. Faria, Financial distress drivers in Brazilian banks: A dynamic slacks approach, *European Journal of Operational Research* **240**(1) (2015) 258-268.
32. A. Constantin, T. A. Peltonen and P. Sarlin, Network linkages to predict bank distress, *Journal of Financial Stability* **35** (2018) 226-241.
33. B. González-Hermosillo, Determinants of ex-ante banking system distress: A macro-micro empirical exploration of some recent episodes, *IMF Working Paper, WP/99/33* (Washington, DC: International Monetary Fund, 1999).
34. T. J. Curry, P. J. Elmer and G. S. Fissel, Equity market data, bank failures and market efficiency, *Journal of Economics and Business* **59**(6) (2007) 636-559.
35. W. Beaver, Financial ratios as predictors of failure, *Journal of Accounting Research* **4** (1966) 71-111.
36. D. R. Carmichael, *The meaning and implementation of the Fourth Standard of reporting* (American Institute of Certified Public Accountants, AICPA, New York, 1972).
37. G. Foster, *Financial statement analysis* (Englewood Cliffs, N.J.: Prentice Hall, 1986)
38. M. Doumpos and C. Zopounidis, A multicriteria discrimination method for the prediction of financial distress. The case of Greece, *Multinational Finance Journal* **3**(2) (1999) 71-101.
39. S. A. Ross, R. Westerfield and J. F. Jaffe, *Corporate finance (2 Ed.)* (Homewood IL, New York, 1999).
40. I. Bose, Deciding the financial health of dot-coms using rough sets, *Information & Management* **43**(7) (2006) 835-846.
41. P. Ravisankar, V. Ravi and I. Bose, Failure prediction of dotcom companies using neural network-genetic programming hybrids, *Information Sciences* 180(8) (2010) 1257-1267.
42. E. I. Altman and E. Hotchkiss, *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt (3 Ed.)* (John Wiley & Sons, Inc, Hobokek, NJ, 2006).
43. P. J. Fitzpatrick, A comparison of ratios of successful industrial enterprises with those of failed companies. *Certified Public Accountant* 6 (1932) 727-731.
44. R. Smith, and A. Winakor, Changes in financial structure of unsuccessful industrial companies. *Bulletin N 51, Bureau of Business Research*, University of Illinois, IL (1935).
45. C. Merwin, *Financing small corporations in five manufacturing industries*, 1926-36 (National Bureau of Economic Research, New York, 1942).
46. E. I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* **23**(4) (1968) 589-609.
47. E. B. Deakin, A discriminant analysis of predictors of business failure, *Journal of Accounting Research* **10**(1) (1972) 167-179.
48. J. A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* **18**(1) (1980) 109-131.
49. M. E. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research* **22** (1984) 59-82.
50. J. Han and M. Kamber, *Data mining: concepts and techniques* (Morgan Kaufmann, San Francisco, CA, 2001).
51. H. Frydman, E. I. Altman and D. L. Kao, Introducing recursive partitioning for financial classification: the case of financial distress, *The Journal of Finance* **40**(1) (1985) 269-291.

52. C. S. Park and I. Han, A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction, *Expert Systems with Applications* **23**(3) (2002) 255-264.
53. H. Li and J. Sun, Ranking-order case-based reasoning for financial distress prediction, *Knowledge-Based Systems* **21**(8) (2008) 868-878.
54. J. Sun, and H. Li, Financial distress early warning based on group decision making, *Computers & Operations Research* **16** (2009) 885-906.
55. F. Varetto, Genetic algorithms applications in the analysis of insolvency risk, *Journal of Banking & Finance* **22**(10) (1998) 1421-1439.
56. K-S. Shin and Y-J. Lee, A genetic algorithm application in bankruptcy prediction modelling, *Expert Systems with Applications* **23**(3) (2002) 321-328.
57. Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika* **12**(2) (1947) 153-157.
58. A. I. Dimitras, R. Slowinski, R. Susmaga and C. Zopounidis, Business failure prediction using rough sets, *European Journal of Operational Research* **114**(2) (1999) 263-280.
59. T. E. McKee, Developing a bankruptcy prediction model via rough sets theory, *Intelligent Systems in Accounting and Finance and Management* **9** (2000) 159-173.
60. R. Slowinski and C. Zopounidis, Application of the rough set approach to evaluation of bankruptcy risk, *Intelligent Systems in Accounting, Finance and Management* **4**(1) (1995) 27-41.
61. T. E. McKee and T. Lensberg, Genetic programming and rough sets: A hybrid approach to bankruptcy classification, *European Journal of Operational Research* **138**(2) (2002) 436-451.
62. J. C. Paradi, M. Asmild and P. C. Simak, Using DEA and Worst Practice DEA in Credit Risk Evaluation, *Journal of Productivity Analysis* **21**(2) (2004) 153-165.
63. T. Sueyoshi and M. Goto, Can R&D Expenditure Avoid Corporate Bankruptcy? Comparison Between Japanese Machinery and Electric Equipment Industries Using DEADiscriminant Analysis, *European Journal of Operational Research* **196**(1) (2009) 289-311.
64. Z. Li, J. Crook and G. Andreeva, Dynamic prediction of financial distress using Malmquist DEA, *Expert Systems with Applications* **80** (2017) 94-106.
65. E. Castillo, J. M. Gutiérrez and S. A. Hadi, *Sistemas expertos y modelos de redes probabilísticas* (Monografías de la Academia de Ingeniería, Madrid, 1997).
66. R. E. Neapolitan, *Learning Bayesian Networks* (Prentice Hall, Illinois, 2003).
67. K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence* (CRC press, Florida, 2003)
68. F. Jensen, *Bayesian Networks and Decision Graphs* (Springer-Verlag., New York, 2001).
69. F. Calle, Técnicas Bayesianas de apoyo a la toma de decisiones y sus aplicaciones, Ph. D. tesis (Universidad de Extremadura, 2014).
70. J. Pearl, Bayesian networks: A model of self-activated memory for evidential reasoning, in *Proceeding of the 7th Conference of the Cognitive Science Society* (California, 1985).
71. L. Stevens, *Artificial Intelligence. The Search for the Perfect Machine* (Hayden Book Company Hasbrouck Heights, N.J., 1984)
72. J. Tuya, I. Ramos and J. Dolado, *Técnicas cuantitativas para la gestión en la ingeniería del software* (NetBiblio, La Corua, 2007).
73. M. Beltrán, A. Muoz and Á. Muoz, Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica, *Cuadernos de Economía* **37**(104) (2014) 73-86.
74. J. Laffarga, J. L. Martín and M. J. Vázquez, El análisis de la solvencia en las instituciones bancarias: Propuesta de una metodología y aplicaciones de la Banca espaola, *Esic Market* **48** (1985) 51-73.
75. V. Pina, La información contable en la predicción de la Crisis Bancaria 1977-1985, *Revista Espaola de Financiación y Contabilidad - Spanish Journal of Finance and Accounting* **18** (58) (1989) 309-338.
76. S. Zhang, Nearest neighbour selection for iteratively kNN imputation, *Journal of Systems and Software* **85**(11) (2012) 2541-2552.
77. N. Friedman, D. Geiger and M. Goldszmidt, Bayesian Networks Classifiers, *Machine Learning* **29** (1997) 131-163.
78. C. K. Chow and C. N. Liu, Approximating Discrete Probability Distributions with Dependence Trees, *IEEE Transactions on Information Theory* **14**(3) (1968) 462-467.
79. J. Hernández, *Introducción a la minería de datos* (Pearson-Prentice Hall, Madrid, 2004)

Anexo A. Appendix

Table A. Relationship between variables of each group in the Bayesian networks.

Type of variable	Variables	Child variables	Parent variables
Capital	Common Equity Tier 1 capital ratio %		State Total capital ratio %
	Ratio Tier 1%	Net equity/total assets % Total capital ratio %	State
	Total capital ratio %	Common Equity Tier 1 capital ratio % Common CET1 growth %	State Ratio Tier 1%
	Net equity/ total assets %	Ordinary capital/tangible assets %	State Ratio Tier 1%
	Ordinary capital/tangible assets %		State Net equity/ total assets %
	Common CET1 growth %		State Total capital ratio %
Assets	Loans/total assets %	Charges for impaired loans and securities/operating profit prior to impairment %	State Impaired loans/net equity %
	Total assets growth%	Total loan growth%	State Impaired loans without reserves/ net equity %
	Total loan growth %		State Total assets growth%
	Impaired loans/gross loans %		State Impaired loans (including restructured loans and potentially difficult loans)/gross loans %
	Impaired loans + mortgaged assets/gross loans + mortgaged assets %	Net charges/average gross loans % Impaired loans without reserves/ net equity %	State
	Impaired loans overdue/gross loans in preceding year %	Impaired loans (including restructured loans and potentially difficult loans)/gross loans %	Impaired loans without reserves/ net equity %
	Impaired loans (including restructured loans and potentially difficult loans)/gross loans %	Impaired loans/net equity % Impaired loans/gross loans %	State Impaired loans overdue/gross loans in preceding year %
	Loan losses reserve/impaired loans %		State Impaired loans/net equity %
	Provisions for losses/net earnings from interest %	Provisions for losses on loans/ average gross loans %	State Charges for impaired loans and securities/operating profit prior to impairment %
	Provisions for losses on loans/ average gross loans %		State Provisions for losses on loans/ average gross loans %
	Charges for impaired loans and	Provisions for losses/net	State

	securities/operating profit prior to impairment %	earnings from interest %	Loans/total assets %
	Net charges/average gross loans %		State Impaired loans+mortgaged assets/gross loans+mortgaged assets %
	Impaired loans/net equity %	Loans/total assets % Loan losses reserve/ impaired loans %	State Impaired loans (including restructured loans and potentially difficult loans)/gross loans %
	Impaired loans without reserves/ net equity %	Total assets growth % Impaired loan overdue/gross loans in preceding year %	State Impaired loans+mortgaged assets/gross loans+mortgaged assets %
Management	Cost/income ratio %		State Average cost of assets ratio %
	Average cost of assets ratio %	Cost / income ratio % Interest expenses/average interest accrued on liabilities %	State
	Client deposit interest expenses/ average client deposits %		State Interest expenses/average interest accrued on liabilities %
	Interest expenses/average interest accrued on liabilities %	Client deposit interest expenses/ average client deposits %	State Average cost of assets ratio %
Earnings	Financial return %	Ongoing earning capability/ average total assets % Operating profit/average net equity %	State Economic return %
	Operating profit/average net equity %		State Financial return %
	Economic return %	Financial return %	State
	Ongoing earning capability/ average total assets %	Net interest margin %	State Financial return %
	Net interest margin %	Earnings from interest/average interest earning assets % Earnings from interest/operating revenue %	State Ongoing earning capability/ average total assets %
	Earnings from interest/average interest earning assets %	Earnings from interest/average gross loans %	State Net interest margin %
	Earnings without interest/ operating revenue %		State Earnings from interest/operating revenue %
	Earnings from interest/average gross loans %		State Earnings from interest/average interest earning assets %
	Earnings from interest/operating revenue %	Earnings without interest/ operating revenue %	State Net interest margin %
Liquidity	Liquid assets/total assets %	Liquid assets/deposits and finance %	State Loans with less then a 1-year maturity/total loans %

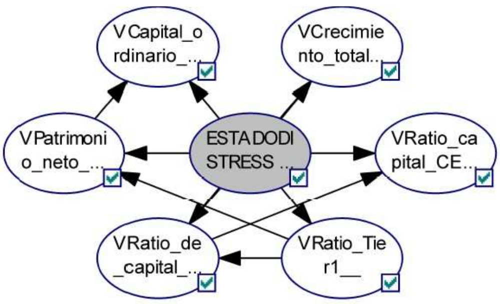
	Loans with less then a 1-year maturity/total loans %	Liquid assets/total assets %	State Client loans/client deposits %
	Deposits with less then a 1-year maturity/total deposits %	Client deposits/total finance without derivatives %	State
	Client loans/client deposits %	Loans with less then a 1-year maturity/total loans % Minimal risk assets/total deposits and finance %	State Client deposits/total finance without derivatives %
	Interbank assets/interbank liabilities %		State Wholesale finance/total finance without derivatives %
	Minimal risk assets/total deposits and finance %		State Client loans/client deposits %
	Liquid assets/deposits and finance %		State Liquid assets/total assets %
	Client deposits/total finance without derivatives %	Wholesale finance/total finance without derivatives % Client loans/client deposits %	State Deposits with less then a 1-year maturity/total deposits %
	Wholesale finance/total finance without derivatives %	Interbank assets/interbank liabilities %	State Client deposits/total finance without derivatives %
Sensibility	Asset reasonable value/total assets %		State Level 3 assets 3/CETI %
	Level 3 assets/total securities %		State Earnings from commercial transactions/total operating
	Level 3 assets/CETI %	Asset reasonable value/total assets %	State Earnings from commercial transactions/total operating
	Earnings from commercial transactions/total operating revenue %	Level 3 assets/total securities % Level 3 assets/CETI %	State
Macro–economic variables	Yield on long-term Government bonds %		State Gross Domestic Product %
	Unemployment rate %	Gross Domestic Product %	State
	General price index variation %		State Gross Domestic Product %
	Housing price index %		State Gross Domestic Product %
	Mortgages on total property %		State Gross Domestic Product %
	Gross Domestic Product %	Yield on long-term Government bonds % General price index variation % Housing price index % Mortgages on total property %	State Unemployment rate %

Table B. Relationship between variables in the Bayesian networks.

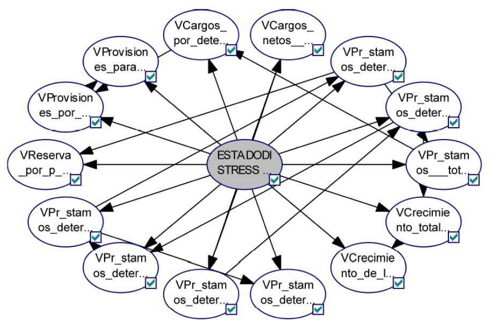
Variables	Child variables	Parent variables
Common Equity Tier 1 capital ratio %	% Minimal risk assets / total deposits and finance %	State Total capital ratio %
Ratio Tier 1%	Total capital ratio %	State Impaired loans + mortgaged assets/ gross loans + mortgaged assets %
Total capital ratio %	Common CET1 growth % Common Equity Tier 1 capital ratio % Net charges / average gross loans %	State Ratio Tier 1%
Net equity/ total assets %	% Ordinary capital/tangible assets %	State Client deposits/total finance without derivatives %
Ordinary capital/tangible assets %		State Net equity/ total assets %
Common CET1 growth %		State Total capital ratio %
Loans/total assets %	Liquid assets/deposits and finance % Charges for impaired loans and securities/ operating profit prior to impairment % Interest expenses/average interest accrued on liabilities % Client loans/client deposits %	State Impaired loans/net equity %
Total asset growth%	Total loan growth%	State Wholesale finance/total finance without derivatives %
Total loan growth %		State Total assets growth%
Impaired loans/gross loans %		State Impaired loans (including restructured loans and potentially difficult loans)/gross loans %
Impaired loans + mortgaged assets/gross loans + mortgaged assets %	Ratio tier 1%	State Impaired loans without reserves/ net equity %
Impaired loans overdue/gross loans in preceding year %		State Impaired loans (including restructured loans and potentially difficult loans)/gross loans %
Impaired loans (including restructured loans and potentially difficult loans)/gross loans %	Impaired loans/gross loans % Impaired loans/net equity % Impaired loans overdue/gross loans in preceding year %	State Yield on long-term Government bonds %
Loan losses reserve/impaired loans %	Level 3 assets/total securities %	State Impaired loans/net equity %
Provisions for losses/net earnings from interest %	Provisions for losses on loans/ average gross loans %	State Charges for impaired loans and securities/operating profit prior to impairment %
Provisions for losses on loans/ average gross loans %	Cost / income ratio %	State Provisions for losses/net earnings from interest %
Charges for impaired loans and securities/operating profit prior to impairment %	Provisions for losses/net earnings from interest % Economic return %	State Loans/total assets %

Net charges/average gross loans %		State Total capital ratio %
Impaired loans/net equity %	Earnings from commercial transactions/ total operating revenue % Loans/total assets % Loan losses reserve/impaired loans %	State Impaired loans (including restructured loans and potentially difficult loans)/ gross loans %
Impaired loans without reserves/ net equity %	Impaired loans+mortgaged assets/gross loans+mortgaged assets %	State Client deposit interest expenses/ average client deposits %
Cost/income ratio %		State Provisions for losses on loans/average gross loans %
Average cost of assets ratio %	Earnings from interest/operating revenue %	State Ongoing earning capability/average total assets %
Client deposit interest expenses/ average client deposits %	Deposits with less than a 1-year maturity/ total deposits % Impaired loans without reserves / net equity %	State Interest expenses/average interest accrued on liabilities %
Interest expenses/average interest accrued on liabilities %	Client deposits/total finance without derivatives % Client deposit interest expenses/average client deposits % Earnings from interest/average interest earning assets %	State Loans / total assets %
Financial return %	Operating profit/average net equity %	State Economic return %
Operating profit/average net equity %		State Financial return %
Economic return %	Financial return %	State Charges for impaired loans and securities/operating profit prior to impairment %
Ongoing earning capability/ average total assets %	Average cost of assets ratio %	State Net interest margin % Earnings from interest/average interest earning assets %
Net interest margin %	Ongoing earning capability/average total assets % Earnings from interest/operating revenue %	State Earnings from interest/average interest earning assets %
Earnings from interest/average interest earning assets %	Earnings from interest/average gross loans % Net interest margin %	State Interest expenses/average interest accrued on liabilities %
Earnings without interest/operating revenue %		State Earnings from interest/operating revenue %
Earnings from interest/average gross loans %		State Earnings from interest/average interest earning assets %
Earnings from interest/operating revenue %	Earnings without interest/operating revenue %	State Net interest margin % Average cost of assets ratio %
Liquid assets/total assets %	Loans with less then a 1-year	State

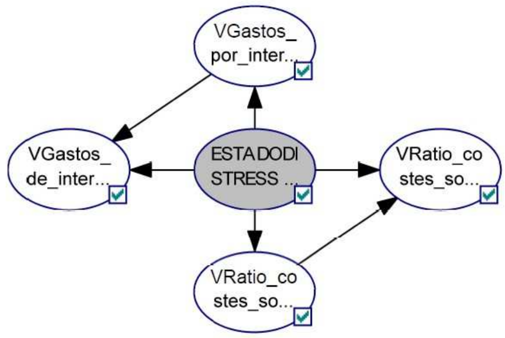
	maturity/total loans %	Liquid assets/deposits and finance %
Loans with less then a 1-year maturity / total loans %		State Liquid assets/total assets %
Deposits with less than a 1-year maturity/total deposits %	Level 3 assets 3/CET1 %	State Client deposit interest expenses/ average client deposits %



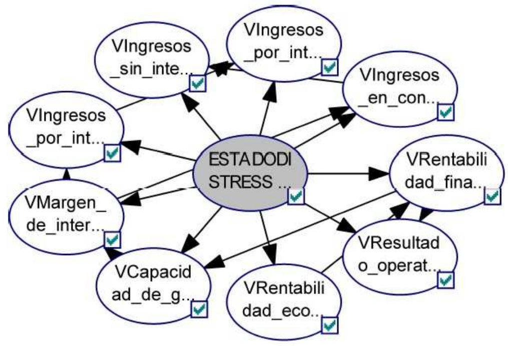
(a) Capital



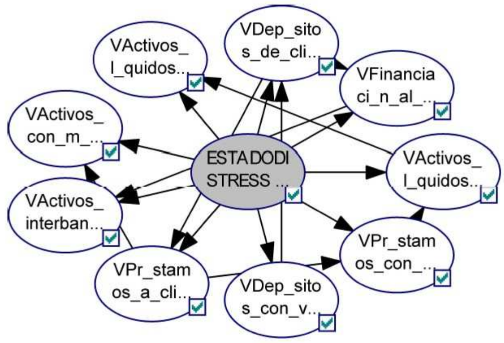
(b) Assets



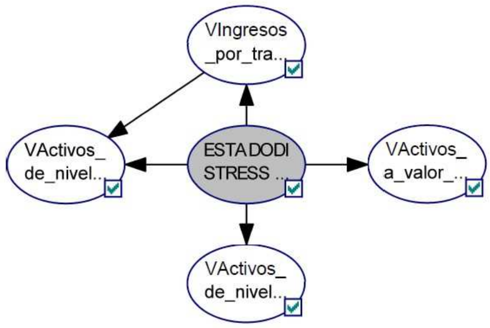
(c) Management



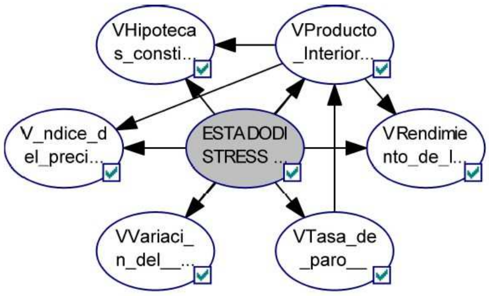
(d) Earnings



(e) Liquidity



(f) Sensibility



(g) Macroeconomic variables

Figure 2. Bayesian networks for each group of variables.

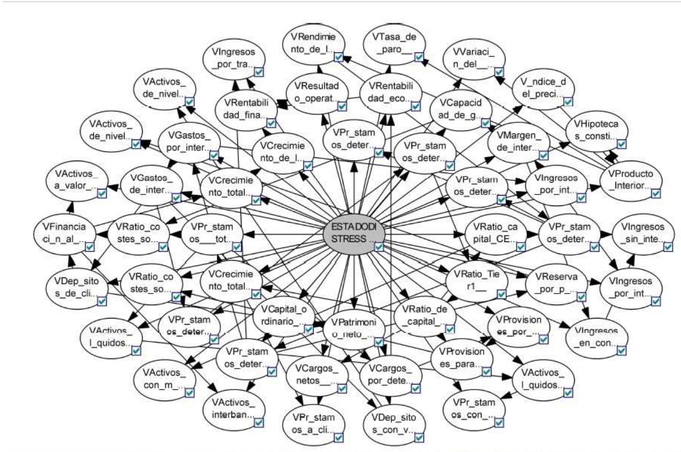


Figure 3. Bayesian networks for each group of variables.