Comput Econ
DOI 10.1007/s10614-017-9737-x

Predicting Corporate Financial Failure Using Macroeconomic Variables and Accounting Data

Eduardo Acosta-González[1] · **Fernando Fernández-Rodríguez**[1] · **Hicham Ganga**[1]

Accepted: 30 August 2017

Abstract Recent studies of the prediction of corporate financial failure have taken into account many factors, mostly corresponding to financial ratios derived from firms' annual accounts. Nevertheless, the current crisis and the consequent exponential increase in rates of insolvency have made it clear that the phenomenon of bankruptcy cannot be explained without reference to macroeconomic variables; thus, the overall condition of the economy, and not just the internal financial ratios of firms, must be addressed. In this paper, focusing on the Spanish construction sector from 1995 to 2011, we analyse selected econometric models for predicting bankruptcy, in which both macroeconomic variables and financial ratios are employed. In view of the large number of variables with these characteristics, which are frequently correlated with each other, and the consequent enormous number of models that would be obtained, we decided to focus on just five optimal econometric models for predicting the financial failure of firms, at 1, 2, 3, 4 and 5 years in advance, with a limited number of explanatory factors, to be selected by an automatic statistical procedure, guided solely by the data and based on a genetic algorithm. The empirical results obtained show that these econometric models are capable of achieving high rates of predictive success, both for in-sample and for out-of-sample predictions. In the latter case, failure and non-failure firms were classified with success rates of 98.5 and 82.5%, respectively,

✉ Fernando Fernández-Rodríguez
fernando.fernandez@ulpgc.es

Eduardo Acosta-González
eduardo.acosta@ulpgc.es

Hicham Ganga
hganga2@hotmail.com

1 Faculty of Economics, Management and Tourism, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

Published online: 06 September 2017

1 year in advance. This predictive quality is maintained at 2, 3 and even 4 years in advance.

Keywords Forecasting financial failure · Genetic algorithms · Real estate · Bankruptcy · Financial distress

1 Introduction

Bankruptcy forecasting is a recurring theme in financial literature. Traditionally, bankruptcy prediction studies have sought to characterise corporate health according to financial differences among firms, measured using accounting data. Thus, various papers in the field of financial analysis have attempted to show the utility of prediction models based on firms' annual published accounts (see Dimitras et al. 1996 for a general perspective). Although the business cycle obviously affects the state of firms, microeconomic factors were initially considered to be most responsible for financial failure, because it was held that external macroeconomic factors affected all firms equally.

Many studies have sought to establish, from the statistical point of view, an empirical relation between bankruptcy and corporate accounting data, with special reference to financial ratios; this has been so since the 1960s, when discriminant analysis came into widespread use in this type of research, with Altman (1968) being a pioneer in this field. In the 1970s, linear econometric probability models and the LOGIT and PROBIT models were introduced. The LOGIT model, first used by Ohlson (1980), has come to be one of the most widely employed statistical tools for predicting the financial failure of firms, perhaps only surpassed by the use of multiple discriminant analysis (Altman 1968). Nevertheless, the LOGIT model presents certain advantages that make it clearly superior to discriminant analysis (see Balcaen and Ooghe 2004 for a general view on the advantages and inconvenience of these kinds of models).

With the increasingly powerful calculating ability of computers, by the mid-1980s mathematical programming techniques were being applied to predict the failure of firms, a field in which Frydman et al. (1985) did pioneering work. Subsequently, in the 1990s, Johnsen and Melicher (1994) introduced multinomial LOGITs, and a few years later Wilson (1997) employed advanced econometric techniques based on time series and integrated LOGIT models. In addition Expert Systems (Messier and Hansen 1988), nonparametric models such as Multivariate Adaptive Regression Splines (Friedman 1991), Artificial Neural Networks (Tam and Kiang 1992), Hybrid Classifiers for combining previous procedures (Olmeda and Fernández 1997) and mixed LOGIT models (Jones and Hensher 2004), have also been proposed.

Nevertheless, the wide-ranging empirical experience that has been historically accumulated shows that some of the causes of bankruptcy are macroeconomic or external. Accordingly, researchers have sought to identify other variables that may be relevant to financial failure, apart from the financial ratios. This search has given rise to a diverse body of literature, arising from Altman's seminal 1968 paper, in which detailed economic reasoning is offered of the explanatory variables involved; another groundbreaking paper on corporate default was the theoretical study conducted by

Merton (1974). Thus, a specific sub-category of research has emerged, combining internal financial information with macroeconomic information, taking into account the empirical fact that not all economic sectors and firms are equally capable of resisting a given impact during a macroeconomic crisis.

Inspired by previous research into the business cycle several papers have examined cases in which variables such as profits, investment, money, credit, interest rates and assets are analysed to account for the financial failure of firms. In his seminal paper on the use of discriminant analysis in the prediction of business failure, Altman (1968) implemented a multiple regression model in which the explanatory variables included the change in GDP lagged by a quarter, the S&P 500 Index and money supply M1.

In the same area, Rose et al. (1982) selected a wide variety of lagged macroeconomic variables, including the S&P 500 Index, the base interest rate, the 3-month US Treasury Bonds rate, the ratio of gross domestic private investment to GDP, and the ratio of retail sales to GDP, all of which were found to be significant. Altman (1983) employed a regression model with distributed lags, based on numerous macroeconomic variables, notably the changes in the S&P 500 Index, the formation of new entities and money supply M2 with several lags, but found the change in GDP to be nonsignificant. Fama (1986) observed signs of the influence of the business cycle on bankruptcy probability, after examining variations in quarterly premiums with the business cycle. Levy and Bar-Niv (1987) conjectured that income and the level of price fluctuation adversely affect corporate performance, and measured a positive correlation between the number of annual bankruptcies, the variance in GDP, with respect to 24 terms, and the GDP deflator; they also recorded a negative correlation with the covariance between the GDP and its deflator. Melicher and Heart (1988) used credit conditions to explain aggregate business failure, concluding that aggregate failure activity lags behind the volatility of interest rates, the cost of short-term credit (3-month T-bill) and the availability of short-term credit (free bank reserves). Lane and Schary (1989) explained the percentage of financial failure by reference to 21 macroeconomic variables, plus the age of the firm and the year in which it was founded. In this respect, too, Hol (2001) developed a model of non-performing loans for Norwegian banks, in which the following parameters were found to be relevant: the lagged GDP, the interest rate on loans, the industrial production index and money supply M1. Cressy (1992) focused on the macroeconomic effects on small-firm bankruptcy, including annual dummy variables as proxies in the macroeconomic environment. In their model, Burn and Redwood (2003) included explanatory variables such as profitability, interest coverage, indebtedness, liquidity, size of the company and the industry, growth of GDP and subsidiary dummies. Benito et al. (2004) showed that the growth of GDP in real terms and financing costs are both significant predictor variables of bankruptcy probability for Spanish firms. Hol (2007) combined financial statement analysis with an analysis of the macroeconomic environment, and concluded that GDP gap, production index and money supply M1, in combination with certain financial health indicators for individual firms, were significant predictors of default for Norwegian firms during both recovery and expansion in the 1990s. Another important model was proposed by Duffie et al. (2007), in which distance-to-default (a volatility adjusted measure of leverage) was shown to be almost a sufficient statistic for accounting variables. Besides, there are more recent studies using more powerful prediction approaches; in this sense it is worth mentioning, among others, the following papers and the references therein: Hernandez-Tinoco and Wilson (2013), which show the utility of combining accounting, market and macro-economic data to explain corporate credit risk; Lua et al. (2015), on the effects and applicability of financial media reports on corporate default ratings; Fejér-Király (2015), which surveys the historic evolution of Bankruptcy Prediction; Mousavia et al. (2015), that study the performance evaluation of competing bankruptcy prediction models in a multi-criteria assessment framework; Kumar and Rao (2015), proposing to improve the prediction accuracy of Altman's Z-score by developing a multivariate nonlinear model for computing it; Tkáč and Verner (2016), that surveyed the recent research of Artificial Neural Networks in business, especially the one aimed at financial distress; Bauweraerts (2016), which investigates the relative importance of various bankruptcy predictors commonly used; Altman et al. (2016), that assesses the classification performance of the Z-Score model in predicting bankruptcy with the goal of examining the model's usefulness for all parties, especially banks that operate internationally and need to assess the failure risk of firms.

All of the above models, both those that employ information on the accounting status of the company and also mixed models that combine accounting and macroeconomic factors, suffer from multicollinearity problems in their estimations, due to the redundancy of the information provided by the variables. In this paper, we implement an econometric model, which reacts well to the presence of multicollinearity, for predicting bankruptcy in the Spanish construction sector, with particular reference to the effects of the most recent episodes of the economic crisis. Our methodology is guided solely by data and incorporates two kinds of the above-mentioned explanatory variables: financial ratios on the accounting status of the companies concerned (as analysed previously in other studies) and macroeconomic variables. Among the macroeconomic variables considered are the interest rate term structure, the volatility of stock markets, the country risk premium, the inflation level, the unemployment rate, credit fluctuations and the default rate.

The remainder of this paper is organized as follows. Section 2 introduces the methodology used for selecting the optimal set of bankruptcy explanatory variables, describing a procedure for selecting variables in a LOGIT model via a genetic algorithm; this section also considers the multiple imputation methodology developed by King et al. (2001) as a solution to the missing values problem in financial failure data bases; in addition, the data base used is described. Section 3 then shows the empirical results for the Spanish construction industry data base. Finally, Sect. 4 presents the main conclusions drawn.

2 Variable Selection and Sample Preparation

The first step in any empirical study on financial failure is to select the explanatory variables to be incorporated: corporate financial ratios, macroeconomic variables or a combination of both. Groups of financial ratios often share the same numerator or denominator, and many macroeconomic variables are strongly correlated, and so an important problem that arises in this first step of any research is the redundancy of the information provided by the ratios and the correlation between macroeconomic variables, which can produce multicollinearity problems in the estimations. It is very well known that multicollinearity produces high standard errors in the estimated coefficients of econometric models, and that this has a negative effect on the precision of estimations. Nevertheless, no logical procedure exists by which we can identify the ratios containing the most complete information, and at the same time avoid its duplication.

Most studies in this field begin with a large initial set of financial ratios and macroeconomic variables, which are subsequently reduced to a final set of factors on the basis of statistical considerations. A first consequence of this procedure is that a very large number of possible sub-models are obtained from this set of variables. For K initial financial ratios, there are 2^{K} possible sub-models. For $\mathrm{K} = 40$, hence, there are 1,099,511,627,776 possible models, and the explicit computation of each and every one of these models would be prohibitively expensive. In order to resolve this intractable problem, several heuristic methods have been proposed, restricting the focus to a smaller number of potential subsets of regressors. Among the most popular of these methods are stepwise procedures, such as forward selection or backward elimination, which sequentially include or exclude variables based on t-ratio statistic considerations (see Miller 2002 for a review of subset selection in regression). However, the stepwise approach tends to over-identify models, and there is a very high probability of introducing false significant variables into the model (see Lovell 1983). Accordingly, other procedures have been proposed for improving the selection of regressors, by authors such as Sala-i-Martin (1997), Hoover and Perez (1999), Hendry and Krolzig (2001), Perez-Amaral et al. (2003) and Acosta-González and Fernández-Rodríguez (2007).

In this paper, we describe a computational search procedure for forecasting the financial failure of firms by careful selection of the variables to be included in the econometric model. More specifically, given the variety of possible financial ratios and of other internal variables such as size and age, as well as macroeconomic variables (with redundant information) that can be used in constructing a LOGIT model, we apply an algorithmic procedure for selecting models, choosing a parsimonious model with a limited number of financial and macroeconomic factors. For this purpose, a recently developed method by Acosta-González and Fernández-Rodríguez (2007, 2014) was applied. This procedure is based on a genetic algorithm (GA, henceforth)[1] guided by the Schwarz Information Criterion (SIC, henceforth) which is designated by the acronym GASIC. Finally, McKee and Lensberg (2002) conducted an important study, using accounting data from 291 US public companies to construct a model with four explanatory variables previously selected using the sophisticated methodology of "rough sets", whereby the functional form of the model is selected by means of genetic programming. In our own approach, on the other hand, a LOGIT model is employed in which the variables are selected by means of the GA from a large set of financial ratios and macroeconomic variables.

[1] A genetic algorithm is a metaheuristic procedure, introduced by Holland (1975), which generates solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover. See Acosta-González and Fernández-Rodríguez (2007, 2014), and the references therein, for a simple, but comprehensive introduction to GA.

Table 1 Detailed compositions of samples for the five models under study

Models	Total sample		Estimation subsample		Validation subsample	
	Failed	Healthy	Failed	Healthy	Failed	Healthy
One year before bankruptcy	2200	2200	2000	2000	200	200
Two years before bankruptcy	2200	1100	2000	1000	200	100
Three years before bankruptcy	2090	2090	1900	1900	190	190
Four years before bankruptcy	1870	1870	1700	1700	170	170
Five years before bankruptcy	1430	1430	1300	1300	130	130

Our sample includes building firms, with data from 1995 to 2011. The accounting data were obtained from the SABI[2] database. When a failed firm was selected for inclusion in the sample, we compiled all its annual reports available from 1995 to 2011. Since our propose is to predict bankruptcy in advance, more specifically, 1, 2, 3, 4 and 5 years before it occurs, we prepared five databases of failed firms with information from their balance sheets 1, 2, 3, 4 and 5 years before bankruptcy took place. In addition to these samples of failed firms we also selected, at random, a similar number of healthy firms, also from the SABI database.[3] Table 1 shows the size of the samples for the five models under study.

The macroeconomic variables were collected from the Thomson Reuters Datastream database, the sectorial variables from INE[4] and the variable "Price of land" from the Spanish Government Ministerial Department for Infrastructure.

There appears to be no consensus concerning which ratios are the most appropriate for predicting company failure, and so in general there is a great disparity in the classification of the economic and financial ratios employed for this purpose. In the literature, these ratios are usually classified into categories or groups sharing similar characteristics. The financial ratios included in this work are grouped into seven categories, each of which describes a particular aspect of the firm. Those categories are: Liquidity/Solvency, Profitability, Activity, Indebtedness, Equilibrium, Cash Flow and Asset Structure. Table 2 details all these financial ratios and the other internal variables included. The macroeconomic variables are listed in Table 3.

Once the raw data are obtained from the SABI database, the first problem to be addressed is that of the missing data. In practice, in order to produce a data matrix with no missing values, a carefully chosen set of observations (firms) and variables (usually ratios) is usually eliminated. But in many cases this procedure provokes sample selection bias, especially when the missing data process has a causal structure. Therefore, to overcome the missing data problem we propose the technique of

[2] Sistema de Análisis de Balances Ibéricos (Iberian Balance Sheet Analysis System).

[3] This matching is frequent in the studies of financial distress because the whole data set for healthy firms would unbalance the prediction problem because the number of healthy firms is overwhelmingly higher than the failed firms (more than 99% in the SABI database). In this sense, as Maddala (1983, p. 91) points out, in the general context of LOGIT models which we have used, such sampling procedures only affect the constant term in the estimations.

[4] Instituto Nacional de Estadística (Spanish Institute of Statistics).

Table 2 Initial set of financial ratios and other internal variables for explaining company failure

Group	No.	Ratio	Description
Liquidity/solvency	1	General liquidity	Current assets/current liabilities
	2	Immediate liquidity	Cash and cash equivalent/current liabilities
	3	Liquidity ratio	(Current assets-stocks)/current liabilities
	4	Shareholders' liquidity ratio	Shareholders' funds/non-current liabilities
	5	Shareholders' funds/invested capital (%)	(Shareholders' funds/(shareholders' funds + non-current liabilities)) * 100
	6	Payback capacity	(Long term debt + current liabilities)/(sales + amortization + variation in provisions)
	7	Solv1	Current liabilities/total assets
	8	Solv2	Financial debts/cash flow
	9	Solv3	Cash/total liabilities
Profitability	10	Economic profitability (%)	Net income/total assets
	11	Financial profitability (%)	(Net income/shareholders' funds) * 100
	12	Return on capital employed (%)	EBIT/(shareholders' funds + non-current liabilities)) * 100
	13	Return on total assets (%)	(Earnings before income taxes/total assets) * 100
	14	Profit margin (%)	(Earnings before income taxes/operating revenue) * 100
	15	Net assets productivity	Operating revenue/(shareholders' funds + non-current liabilities)
	16	Interest cover	Operating income/interest paid
	17	Financial expenses (%)	Financial and similar expenses + variation in provisions for financial investment/sales
Activity	18	Net sales growth (%)	((Sales (t) − Sales (t − 1))/Sales (t − 1)) * 100
	19	Total assets rotation	Sales/total assets
	20	Fixed assets rotation	Sales/fixed assets
	21	Stock rotation	Operating revenue/stocks
Indebtedness	22	Ind1	Long term debt/total assets
	23	Ind2	Shareholders' funds/total liabilities
	24	Ind3	Long term debt/shareholders' funds
	25	Ind4	Long term debt/current liabilities
	26	Ind5	Total liabilities/total assets
	27	Ind6	Total liabilities/shareholders' funds
	28	Ind7	EBIT/financial expenses
	29	Ind8	Fixed assets/shareholders' funds
	30	Ind9	(Shareholders' funds + long term debt)/total assets

Table 2 continued

Group	No.	Ratio	Description
	31	Ind10	(Shareholders' funds + long term debt)/current liabilities
	32	Ind11	(Shareholders' funds + long term debt)/total liabilities
	33	Ind12	Cash flow/total liabilities
Equilibrium	34	Working capital (Th.)	Shareholders' funds + provision for risks and expenses + long-term debt-fixed and other noncurrent assets
	35	Working capital requirement (Th.)	[Accounts receivable + inventory + prepaid expenses] − [accounts payable + accruals]
	36	Equilibrium	(Shareholders' funds + other non-current liabilities + long term debt)/fixed assets
	37	Eq1	Working capital/total assets * 100
	38	Eq2	(Shareholders' funds − capital)/total assets
	39	Eq3	(Shareholders' funds + long term debt)/total assets
	40	Eq4	Current assets/total sales
Cash flow	41	CF1	Cash flow/total assets
	42	CF2	Cash flow/total sales
	43	CF3	Cash flow/shareholders' funds
	44	CF4	Cash flow/current liabilities
Asset structure	45	Fixed assets weight	Fixed assets/total assets
	46	Current assets weight	Current assets/total assets
Other internal variables		Size	Company's total assets
		Age	Years elapsed since the creation of the company

multiple imputations, developed by King et al. (2001). This methodology has been successfully employed in the field of bankruptcy by Acosta-González and Fernández-Rodríguez (2014), enabling the imputation of values to missing observations. The multiple imputation procedure may be summarized in two steps. First, the missing values are considered as parameters which are estimated in a maximum likelihood framework (see King et al. 2001) based on all the available data. Second, in order to avoid introducing an erroneous precision into the estimation based on the imputed values, several random samples of these values are acquired and averaged in order to derive the final estimations. Thus, only the firms with a large amount of missing data (over 10%) are discarded. Once the missing data have been estimated, the samples employed for the selection of factors are divided into an estimation subsample and a validation subsample. Table 1 shows the details of these samples.

Table 3 Macroeconomic variables and trending process employed for explaining financial failure

Macroeconomic variables	Description
Interest rates term structure (IRTS)	Spread of IRTS (long-term interest rate-short-term interest rate)
Interest rate	12 month Euribor, date: December of each year
Volatility of the stock market	Standard deviation of the IBEX-35 returns during the last 60 days of each year
Country risk premium	Country risk premium for Spain at 31 December each year
Government debt	Annual government debt
Inflation	Annual percentage change in Consumer Price Index
Unemployment rate	Annual unemployment rate
Sector's share of GDP	Sector GDP/national GDP
Credit to construction companies	Total amount of credit granted by banks to construction companies
Credit to householders	Total amount of credit granted by banks to householders
Bank arrears	Annual bank arrears
Land price	Price per urbanized square metre
Trending process	**Description**
Time (t)	t = 1, 2, 3, …, N (N = 17 years)
t^2	Squared time

3 Results

Table 4 shows the models selected by the GASIC method. Henceforth, Model 1 refers to the model that predicts bankruptcy 1 year in advance; Model 2, 2 years in advance, and so on until Model 5 which predicts bankruptcy 5 years in advance. Table 4 shows the data for each model, with the explanatory variables of bankruptcy probability in the Spanish construction sector during the period 1995–2011. Tables 5 and 6 show the frequency of appearance of each variable in each model and the standardized model coefficients,[5] respectively.

Table 4 shows that, for each model, all the coefficients of the explanatory factors selected by GASIC are statistically significant and present the sign (positive or negative) that is to be expected. In accordance with the main purpose of this work, the explanatory variables selected by GASIC are not only financial ratios; other internal variables such as the firm's size and age, together with macroeconomic factors, also play a major role in accounting for bankruptcy in the Spanish construction sector.

[5] The variables selected by GASIC in the five models are expressed in different units. Therefore, in order to determine the variables that exert most influence in bankruptcy, standardized coefficients are obtained for each model. To that end, we follow the methodology proposed by Menard (2001) with respect to logistic regression.

Table 4 Statistical significance of the explanatory factors selected by GASIC for predicting bankruptcy 1, 2, 3, 4, 5 years before it happens (Models 1, 2, 3, 4, 5, respectively)

Selected variables	Beta	t-Student	P-value
Model 1			
Constant	−10.0842	−20.33824	0
Shareholders' funds/non-current liabilities	−7.20E−05	−2.69096	0.003
Total liabilities/total assets	0.0064	9.16593	0
Company size	5.50E−08	10.10552	0
Interest rate term structure	68.3242	17.18692	0
Credit to householders	−1.16E−08	−14.97317	0
Land price	0.1504	18.1184	0
Model 2			
Constant	−20.7924	−17.124	0
Total liabilities/total assets	0.0191	10.842	0
Current assets/total sales	1.0663	11.063	0
Bank arrears	1.79E−08	8.652	0
Credit to construction companies	−8.04E−08	−12.066	0
Sector's share of GDP	259.265	14.362	0
Volatility of the IBEX-35	0.0021	15.206	0
Model 3			
Constant	−34.683	−16.253	0
Current liabilities/total assets	1.637	11.133	0
Long term debt/shareholders' funds	0.01047	7.578	2.18E−14
Shareholders' funds/invested capital	−7.41E−05	−5.581	1.27E−08
Company size	1.45E−05	6.312	1.53E−10
Age	0.043	6.69	1.27E−11
Interest rate	364.799	12.865	0
Interest rate term structure	230.959	9.766	0
Credit to construction companies	−1.88E−07	−17.003	0
Land price	0.357	17.484	0
Volatility of the IBEX-35	40.039	16.57	0
Model 4			
Constant	−13.94852	−18.58215	0
Cash flow/total liabilities	−1.45424	−10.26263	0
Interest rates	181.36606	12.01215	0
Interest rate term structure	213.13811	11.92032	0
Sector's share of GDP	74.93369	17.54474	0
Volatility of the IBEX-35	−8.07212	−8.83678	0
Model 5			
Constant	−1.275996	−7.48028	5.04E−14
Fixed Assets/total Assets	0.031183	6.95544	2.22E−12
Total liabilities/total assets	0.017378	8.85405	0
Size	−4.18E+08	−8.44273	0

Table 5 Summary of variables selected by our genetic-algorithm methodology for predicting the financial failure 1, 2, 3, 4, 5 years before it happens (Models 1, 2, 3, 4, 5, respectively)

Category of variables	Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Total
Financial ratios	Indebtedness	•	•	•	•	•	5
	Solvency	•		•			2
	Asset rotation		•				1
	Liability structure					•	1
	Financial structure			•			1
Off-balance variables	Age			•			1
	Size	•		•		•	3
Macroeconomic variables	Interest rates			•	•		3
	IRTS	•		•	•		3
	Credit granted	•	•	•			3
	Bank arrears		•				1
	Sector's share of GDP		•		•		2
	IBEX 35 volatility		•	•	•		3
	Land price	•		•			2

Two explanatory factors were selected in almost all the models: the indebtedness ratio and the macroeconomic variable credit fluctuation. Specifically, in the short term (Models 1 and 2), company indebtedness is evaluated by the Total liabilities/Total assets ratio. In Model 3, the indebtedness ratio is the Long term debt/Equity ratio. Finally, in Model 5 the same indebtedness ratio that was selected in Models 1 and 2 (Total liabilities/Total assets) is again employed. In all cases, the signs of the coefficients of these ratios were found to be positive, as expected; thus, an increase in indebtedness increases the probability of default. As can be seen in Table 6, the indebtedness ratio is relatively low 1 year before company failure, but its effect is three times greater in Models 2 and 5, which shows that indebtedness is most damaging in the middle and long term. Models 3 and 4 include two new indebtedness ratios (Long term debt/Shareholders' funds and Cash flow/Total liabilities, respectively); the latter represents a firm's debt repayment capacity and has a negative sign, which means that the higher the ratio, the lower the probability of the firm going bankrupt.

After the indebtedness ratios, the ratios most commonly included are those referring to solvency, which appear in Models 1 and 3. Finally, a cash flow ratio is included in Model 4 and one related to the asset structure in Model 5.

There are more macroeconomic variables than financial ratios in these models, reflecting the importance of the economic situation to the survival of construction firms. Among all the macroeconomic variables in the database, GASIC selected four which are directly or indirectly related to the banking and financial sectors: credit for the housing sector, bank arrears, interest rates and the interest rate term structure (IRTS), highlighting the relationship between these sectors and the building industry. Although the banking sector plays a crucial role in overall economic activity, in the case

Table 6 Standardized coefficients for Models 1, 2, 3, 4 and 5

Explanatory variables	Standardized coefficients
Model 1	
Shareholders' funds/non-current liabilities	−0.3003
Total liabilities/total assets	0.0978
Company size	0.2899
Interest Rate Term Structure	0.2741
Credit to construction companies	−0.4517
Land price	0.5767
Model 2	
Total liabilities/total assets	0.2814
Current assets/total sales	0.2561
Bank arrears	0.2426
Credit to construction companies	−0.8641
Sector's share of GDP	0.8759
IBEX-35 Volatility	0.355
Model 3	
Current liabilities/total assets	0.1522
Long term debt/equity	0.1291
Shareholders funds/investment capital	−0.2466
Size	0.2631
Age	0.1018
Interest rates	1.3143
Interest Rate Term Structure	0.9007
Credit to householders	−2.6008
Land price	2.9263
IBEX-35 volatility	0.6947
Model 4	
Cash flow/total liabilities	−0.1665
Interest rates	0.5899
Interest Rate Term Structure	0.5581
Sector's share of GDP	0.4225
IBEX-35 volatility	−0.1752
Model 5	
Fixed assets/total assets	0.4370
Total liabilities/total assets	0.5484
Size	−4.1621

As the independent variables in each model are measured in different unities, in order to know which of them have a greater effect on the financial failure, the Standardized coefficients for Models 1, 2, 3, 4 and 5 are presented

of the construction sector in particular this relationship is of particular significance: banks are the primary creditors of construction firms and an essential resource in financing their investments. Furthermore, the outputs of these firms (flats, houses, tourist accommodation, etc.) usually require bank financing for customers to be able to purchase them. Among the macroeconomic variables used, those appearing most frequently are credit, IRTS and IBEX-35 volatility. However, neither the variable t (time) nor t^2 were selected by GASIC in any model, which means that the models selected are not time dependent for the period considered (1995–2011).

For Model 5, the macroeconomics variables become irrelevant and are omitted by GASIC; in this case, the first variable concerns the firms' asset structure, expressed as the Fixed assets/Total assets ratio, which is relevant to the firm's long term prospects. The second variable concerns the firms' level of indebtedness, expressed as the Total liabilities/Total assets ratio (which also appears in Models 1 and 2). In Model 5, contrary to all other models, the coefficient of firm size is negative, indicating that in the long run greater size is associated with a lower probability of company failure. Thus, from Tables 4, 5 and 6, we conclude that in the short and medium term (1–3 years), large construction firms are more likely to go bankrupt than small and medium-sized ones, but in the long term, the latter are more prone to failure. The fact that the variable "Size of the firm" has a different sign in the models according to the time remaining to failure suggests that the effect of this variable should be considered with caution, because it depends on the time horizon of bankruptcy.

Finally, Table 6 shows that the accounting ratios have lower weights in the models than the macroeconomic variables, which have the greatest influence in bankruptcy, and which in the first four models have higher standardized coefficients than the accounting ratios. Among these macroeconomic variables, "Credit availability" is especially significant in the first three models, followed by "Price of land", in Models 1 and 3, and by "Interest Rates" in Models 2 and 3. Among the accounting ratios selected by GASIC, "Solvency" is the most influential in the short term (1 year) survival, while "Indebtedness" is the most influential ratio for Models 2, 3, 4 and 5. This finding highlights the importance of solvency for short-term survival and of indebtedness for medium to long-term survival.

3.1 Sensitivity Analysis

Figure 1 shows the sensitivity to change of the failure probability when one explanatory variable varies and the others remain constant.[6] It can be seen, for example, that greater solvency reduces the probability of bankruptcy from 0.9 to 0, but an increase in the indebtedness ratio raises the probability from 0.35 to 0.7. On the other hand, rises in the price of land and in the volume of credit extended to householders do not produce an effect until a given threshold is exceeded (80 euros/m^2 and 40 billion euros per year, respectively). It should also be noted that the impact of company size on the probability of bankruptcy is minimal, as greater size only raises the probability from a minimum value of 0.47 to a maximum of 0.51.

For the sake of simplicity, no figures are shown for the sensitivity to change of the failure probability for models 2, 3, 4 and 5, although these figures are available to

[6] To draw the following figures, the variables in question are ordered from lowest to highest, and the value of the corresponding median is assigned to the remaining variables. The probability of failure is then calculated using the Logit function.

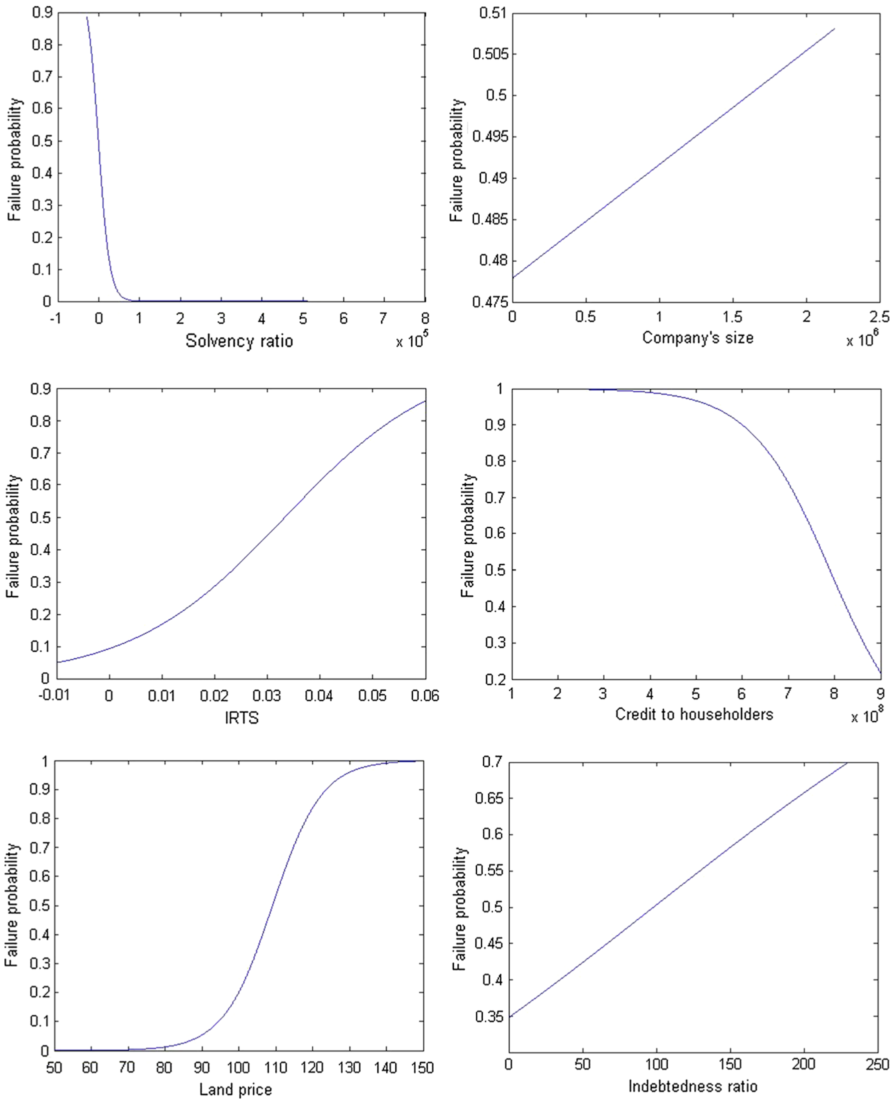

Fig. 1 Evolution of failure probability for Model 1. Sensitivity to change of the failure probability when one explanatory variable varies and the others remain constant. To draw the following figures, the variables in question are ordered from lowest to highest, and the value of the corresponding median is assigned to the remaining variables. The probability of failure is then calculated using the Logit function

the reader on request. They can be summarised as follows. For Model 2, the failure probability is greatly increased when the sector's share of GDP is 9–11%. For all models, greater credit available to the sector, above a threshold of 80 billion euros, is associated with a lower probability of company failure. For Model 3, a rise in the price of land is prejudicial to companies' chances of avoiding bankruptcy above a threshold of 100 euros/m^2. Greater credit availability is associated with a lower probability of bankruptcy when this credit provision exceeds 100 billion euros. A rise in the 1 year Euribor rate increases the probability of bankruptcy when this rate exceeds 2%, but reduces it when the rate exceeds 4%. In Model 4, rising payback capacity has a notable inverse effect on the chances of company failure, which decrease from a maximum of 0.9–0. On the other hand, although IBEX 35 volatility has an inverse relation with the probability of bankruptcy, its influence is slight, as the probability only falls from a maximum of 0.64 to a minimum of 0.35. For Model 5, the variables that have most impact on bankruptcy probability are the two asset structure ratios and indebtedness; however, variations in the size of the firm only reduce the probability from 0.57 to 0.47.

In short, the probability of corporate bankruptcy in the construction sector is more sensitive to changes in macroeconomic and sector variables than in those reflecting internal factors (accounting ratios), which corroborates the results presented in Table 7 with respect to standardized coefficients. The most influential variables are the price of land, credit availability, interest rates and the share of GDP. These results confirm the characteristics and functioning of a sector that has achieved growth on the basis of speculation on land prices, easy credit and low interest rates. However, these factors, which contributed so much to past growth, could become the cause of business failure in a situation of economic crisis.

3.2 Comparing Predictive Capability of Models

To test the in-sample and out-of-sample predictive power of the models, three different procedures were used: Prediction/Outcome cross tables, the Receiver Operating Characteristic (ROC) and the Cumulative Accuracy Profile (CAP) curve (see Sobehart et al. 2000, as a general reference on ROC and CAP curves).

When using a LOGIT model, the classification of firms as failed or healthy depends on the cut-off point C established in the predicted bankruptcy probability $\hat{y}_i$. Thus, given firm i, if $\hat{y}_i > C$ it is classified as failed, and if $\hat{y}_i \leq C$, it is classified as healthy. Accordingly, a failed firm ($y_i = 1$) is correctly classified when $\hat{y}_i > C$, while a healthy one ($y_i = 0$) is correctly classified when $\hat{y}_i \leq C$. Two types of error may arise:

- Type I error: when a failed firm is classified as healthy. In this case ($y_i = 1$) and $\hat{y}_i \leq C$.
- Type II error: when a healthy firm is classified as failed. In this case ($y_i = 0$) and $\hat{y}_i > C$.

Depending on the context of the analysis, one type of error could be more or less harmful than the other. For instance, when the aim of the company failure model is to evaluate the potential risk of lending money to consumers and of losses due to bad debt, a type I error is more harmful because classifying a failed firm as healthy may have serious consequences, while a type II error produces only opportunity cost.

The cut-off point established for classifying the out-of-sample firms is defined as the proportion of failed firms in the sample. Table 7 presents the in-sample and out-of-sample cross table of hit rates for the predictions from all models.

Table 7 In-sample and out-of-sample cross table of hit rates for the predictions from Models 1, 2, 3, 4 and 5 years before bankruptcy

Real	Predicted										Total (%)
	Model 1		Model 2		Model 3		Model 4		Model 5		
	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	
In-sample											
Non failed	68.30	31.70	60.00	40.00	71.32	28.68	70.71	29.29	49.31	50.69	100
Failed	25.55	74.45	9.15	90.85	13.37	86.63	16.18	83.82	12.85	87.15	100
Out-of-sample											
Non failed	98.50	1.50	91.00	9.00	82.11	17.89	85.29	14.71	53.07	46.92	100
Failed	17.50	82.50	7.50	92.50	7.89	92.11	10.59	89.41	20.76	79.23	100

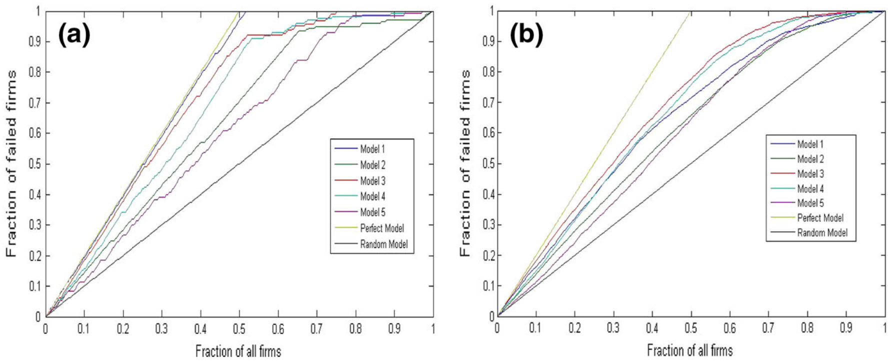

Fig. 2 In-sample and out-of-sample CAP curves. CAP analyses enable us to visualize the predictive capability of failure models comparing them with two bench marks. In the random model, which coincides with the diagonal, the failure probabilities are assigned randomly; the closer the model's CAP curve to this line, the worse the model. On the contrary, the perfect model produces an ideal CAP curve, and the closer the model's CAP curve to this line, the better the model. **a** Out sample CAP curve, **b** In sample CAP curve

Table 7 shows that, with the exception of Model 5, high predictive success rates are achieved, both for in-sample and out-of-sample predictions. An outstanding result is the out-of-sample success rates of Model I, of 98.5% for failed firms and 82.5% for healthy ones. This predictive quality is maintained to a certain extent in Models 2 and 3, but declines as the time horizon increases.

ROC and CAP analyses enable us to visualize the predictive capability of failure models. Figure 2a, b show the in-sample and out-of-sample CAP curves for each model. In order to interpret these figures, two benchmarks should be taken into account; the Perfect Model and the Random Model. Both of these are represented in Fig. 2a, b as straight lines, which in the case of the Random Model coincides with the diagonal. The Perfect Model produces an ideal CAP curve. The closer the model's CAP curve to this line, the better the model. On the other hand, in the Random Model the failure probabilities are assigned randomly, thus representing a very poor discrimination power. The closer the model's CAP curve to this line, the worse the model.

As can be seen, Model 1 is the most closely aligned with the Perfect Model, especially for out-of-sample predictions. Nevertheless, Models 2 and 3 also present CAP curves that are reasonably well in line with the Perfect Model.

In addition to the CAP curve, but derived from it, the Accuracy Ratio (AR) can be defined. The AR is given by the ratios of two areas: the area delimited by the Perfect Model and the Random Model divided by the area delimited by the model being evaluated and the Random Model; therefore, the higher the AR, the better the predictive performance of the model. According to Sobehart et al. (2000), the AR is a very useful measure, making it possible to summarize the predictive power of the model in a single value.

Table 8 presents the out-of-sample AR for the five models, showing that the most accurate is Model 1 (AR = 97.77%), followed by Model 3 (AR = 86.05%); as is to be expected, the worst is Model 5 (AR = 42.72%).

Table 8 Accuracy rate of the results obtained

	Model 1	Model 2	Model 3	Model 4	Model 5
Accuracy rate	0.9777	0.8112	0.8605	0.7510	0.4272

In the context of CAP curve in Fig. 2, the accuracy rate is given by the ratios of two areas: the area delimited by the perfect model and the random model divided by the area delimited by the model being evaluated and the random model. The higher the AR, the better the predictive performance of the model

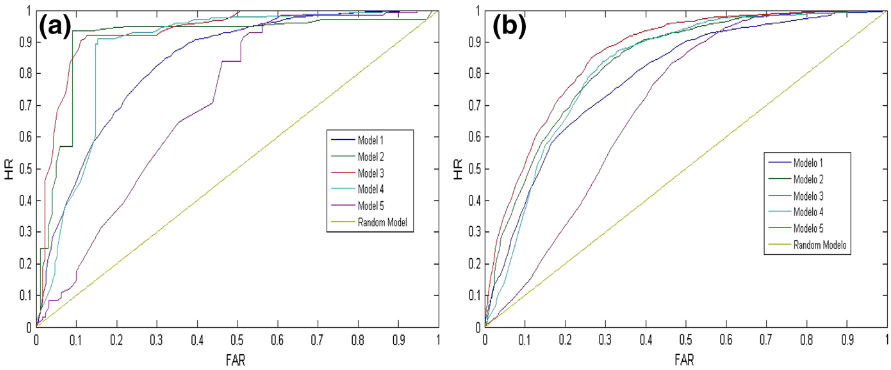

Fig. 3 In-sample and out-of-sample ROC curves. The ROC curve analyses enable us to visualize the predictive capability of failure models through two benchmarks: the perfect model is embodied by a *vertical line* from (0, 0) to (0, 1), followed by a *horizontal line* from (0, 1) to (1, 1), while the random model is the *diagonal line*. **a** Out sample ROC curve, **b** In sample ROC curve

Figure 3a, b show the in and out-of-sample ROC curves for each model. As with the CAP curve, the ROC curve presents two benchmarks, representing the Perfect Model and the Random Model. In this case, the Perfect Model is represented by a vertical line from (0, 0) to (0, 1), followed by a horizontal line from (0, 1) to (1, 1), while the Random Model is the diagonal line shown in Fig. 3a, b.

Figure 3a, b show that, both in and out-of-sample, the ROC curves for the five models are closer to the Perfect Model than to the Random Model. The best results are obtained for Models 2, 3 and 4, which present similar patterns of behaviour, and which are all very close to the Perfect Model; in contrast, Model 5 is the furthest from the Perfect Model. It can also be seen that the out-of-sample behaviour for Models 2, 3, and 4 is better than that of the in-sample behaviour.

In summary, the results of the CAP and ROC curves show that all models, for both in and out-of-sample prediction, are more or less distant from the Random Model. In the case of the ROC curves, Models 2, 3 and 4 are the closest to the Perfect Model, while for the CAP curves, Model 1 is the closest to the Perfect Model.

Finally, we compare the performance obtained in each case using the GASIC procedure with the outcome when we apply the widely-used Z-score model (Altman 1968) which is a classic in the literature on the choice of ratios. Altman's original model was only applicable to public traded entities (since one of its variables requires stock price data), but it was later adapted (Altman 1993) for use with private firms. The resulting equation is

$$Z = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5$$

where

$$X_1 = \frac{\text{Current Asset} - \text{Current Liabilities}}{\text{Total Assets}}$$
$$X_2 = \frac{\text{Retained Earnings}}{\text{Total Assets}}$$
$$X_3 = \frac{\text{Earnings before Interest and Taxes}}{\text{Total Assets}}$$
$$X_4 = \frac{\text{Book Value of Equity}}{\text{Total Equity}}$$
$$X_5 = \frac{\text{Sales}}{\text{Total Assets}}$$

In this equation the decision rule is

$$\begin{aligned} Z &> 2.90 : \text{Safe Zone} \\ 1.23 < Z &< 2.90 : \text{Grey Zone of uncertainty} \\ Z &< 1.23 : \text{Distress Zone} \end{aligned}$$

In order to compare this model with GASIC, which classifies all the firms because it does not have an uncertainty zone, we have transformed Altman's decision rule into the following

$$\begin{aligned} Z &> 2.065 : \text{Safe Zone} \\ Z &< 2.065 : \text{Distress Zone} \end{aligned}$$

where 2.065 is the class mark of the interval [1.23, 2.90].

To compare the Z-score and GASIC methods for out-of-sample predictions, Table 9 shows the contingency table of the real and forecasted states of the firms, from Models 1–5.

In Table 9, it is quite clear that the prediction rates obtained by GASIC are significantly higher than those obtained with the Z-score method. In fact, except for Model 1 (for which the two methods obtained similar results, with 82.5% correct predictions for GASIC and 85.5% for Z-score), the GASIC procedure always proved superior to the Z-score technique.

To compare the GASIC and Z-score results in greater detail, we also applied the CAP and ROC statistical procedures. In both cases, the superiority of GASIC was apparent, producing curves that were closer to the Perfect Model. For the sake of simplicity these curves are not shown here, but they are available from the authors on request. Table 10 shows the AR for both methods; it can be seen that the GASIC values are significantly higher than those obtained with the Z-score procedure.

Table 9 Comparison of out-of-sample predictions between GASIC and Z-score methods by means of hit rates

Real	Predicted										Total (%)
	Model 1		Model 2		Model 3		Model 4		Model 5		
	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	
GASIC											
Non failed	82.50	17.50	92.50	7.50	92.11	7.89	89.41	10.59	79.23	20.76	100
Failed	1.50	98.50	9.00	91.00	17.89	82.11	14.71	85.29	46.92	53.07	100
Z-Score											
Non failed	85.50	14.50	60.00	40.00	53.16	46.84	59.41	40.59	61.54	38.46	100
Failed	55.00	45.00	62.00	38.00	41.05	58.95	70.00	30.00	88.46	11.54	100

Table 10 Accuracy rate for GASIC and Z-score methods

Methods	Accuracy rates				
	Model 1	Model 2	Model 3	Model 4	Model 5
GASIC	0.9777	0.8112	0.8605	0.7510	0.4272
Z-score	0.0273	0.1439	0.0194	0.2150	0.3428

Table 11 Schwarz information criterion of goodness of fit values for models without and with cross variables

	SIC (models without cross variables)	SIC (models with cross variables)
Model 1	1.166497242	1.099948085
Model 2	0.942165111	0.960997163
Model 3	0.951996739	0.90780317
Model 4	1.025846	1.00747043

3.3 Using Cross-Variables to Predict Company Failure

In this section, we consider the possibility of interaction between the different explanatory variables of business failure, to determine whether the status of one variable may be influenced by that of another. For instances, if the indebtedness ratio is influenced by the level of credit fluctuation, the product of these two variables could provide new information to predict the probability of failure. In the case in question, the use of a general model including the product of pairs of all financial ratios, together with the macroeconomic and sector variables, as the GA search space, would produce an intractable computational burden, and so the problem has been simplified, considering as initial candidate variables those that were previously selected by GASIC in the five models obtained previously. This means that we only need calculate the products of the pairs derived from the eight accounting ratios plus the macroeconomic and sector variables considered; thus, GASIC was derived from all the variables selected in the previous models plus the following cross-variables:

- Each financial ratio * each macroeconomic variable.
- Size * Age.
- Size * each macroeconomic variable.
- Age * each macroeconomic variable.

Hence, 145 candidate variables were obtained with which to construct new models.

The first noteworthy results to arise from this new situation are shown in Table 11, which presents the Schwarz Information Criterion of goodness of fit from models with and without the iteration of variables. As can be seen, with the exception of Model 2, the introduction of cross-variables improves the SIC.

Table 12 presents the new variables selected by GASIC, in the presence of cross-variables. All of these variables are highly significant. Perhaps the most striking fact is that accounting ratios now only appear in the model when they interact with macroeconomic or sector variables, but never alone. Specifically, Model 1 consists

Table 12 Variables selected and estimation of Models 1, 2, 3, 4, 5 with cross variables using GASIC

Variables	Beta	t-Student	P-value
Model 1			
Constant	−21.41196	−14.699	0
Size	1.18196	5.2585	7.64E−08
IRTS	116.15056	12.6416	0
Inflation	2.13026	8.8232	0
Credit to householders	−1.85327	−15.4736	0
Land price	19.3886	18.1347	0
IBEX35 volatility	48.1791	8.9955	0
Size * Age	−0.00144	−7.7316	6.66E−15
IRTS * Total liabilities/total assets	0.23313	9.3065	0
Risk premium * Shareholders' funds/non-current liabilities	−0.00098	−4.6955	1.37E−06
Risk premium * Size	−0.17198	−6.7782	6.98E−12
Credit to householders * Size	0.19368	6.4309	7.09E−11
Land price * Size	−2.07027	−5.7994	3.58E−09
Land price * Age	0.03359	6.7854	6.64E−12
Model 2			
Constant	−41.7016	−10.53194	0
Interest rate	2.51E+02	6.66E+00	1.57E−11
Bank arrear	4.69E−02	4.82E+00	7.54E−07
Credit to construction companies	−0.444958	−9.868146	0
Land price	0.682419	10.563795	0
IBEX35 volatility	8.966459	12.080051	0
Credit to householders * Age	6.47E−03	7.13E+00	6.04E−13
Sector's share of GDP * Fixed Assets/total Assets	−7.71E+00	−4.26E+00	1.05E−05
Government debt * Total liabilities/total assets	0.000397	11.02689	0
Model 3			
Constant	−31.6457	−15.1203	0
Interest rate	362.4046	12.8971	0
IRTS	235.9849	10.0959	0
Credit to construction companies	−188.9257	−16.9193	0
Land price	354.7108	17.3259	0
IBEX 35 volatility	29.1432	10.3849	0
IRTS * Size	−4.1778	−9.2762	0
Risk premium* Shareholders' funds/invested capital	−6.80E−05	−3.7032	0.000107
Risk premium * Age	6.75E−04	6.3502	1.20E−10
Land price * Size	1.219	9.3253	0

Table 12 continued

Variables	Beta	t-Student	P-value
Sector's share of GDP * Fixed assets/total assets	−15.54402	−10.4223	0
IBEX 35 volatility * Total liabilities/total assets	1.22E−01	6.6356	1.84E−11
Model 4			
Constant	−14.0978	−18.5997	0
Interest rate	156.1828	9.9872	0
IRTS	213.5195	11.8561	0
Sector's share of GDP	73.8019	16.9364	0
IBEX 35 volatility	−8.743	−9.3573	0
Interest rate * Total liabilities/total assets	2.71E−01	6.5034	4.50E−11
Sector's share of GDP * Fixed assets/total assets	−13.0791	−9.1546	0
Sector's share of GDP * Age	0.3853	6.3547	1.18E−10

of macroeconomic variables and the interaction between macroeconomic variables and ratios, and this can be interpreted straightforwardly: in the case of "IRTS * Total liabilities/Total assets", a high level of indebtedness that coincides with an increase in the level of interest rate spread could further increase the risk of bankruptcy. The variable "Company size" distorts the estimation of "Risk Premium", Credit to Householders" and "Price of Land" in the form of cross products; therefore company size has a significant influence on the effect of these variables on company failure.

With respect to Model 2, Table 12 shows that the variables selected are almost the same as those examined when no interactions were considered; among the cross-variables, "Government debt * Total liabilities/Total assets" has a positive coefficient; this variable reflects the fact that company indebtedness affects the probability of its failure, depending on the level of public indebtedness. In Model 3, of the 10 variables selected without interactions (Table 4), 7 appear in the model with interactions (Table 12). A new and very interesting cross-variable is "Volatility of IBEX 35 * Indebtedness", which has a positive coefficient, showing that indebtedness is dangerous when it coincides with adverse macroeconomic situations manifested by an increased volatility in financial markets, which is risky for the survival of firms. For Model 4, a remarkable fact highlighted in Table 13 is that, as in Model 4 without cross-variables (Table 4), and contrary to the case of Models 1, 2 and 3, the coefficient of this variable is negative and its effect on the probability of failure is negative. Also noteworthy is the emergence in Model 4 of the variable "Interest Rates * Indebtedness", which informs us that a high level of debt in a scenario of higher interest rates will increase financing costs, thus aggravating the risk of failure.

Finally, for Model 5, the GASIC algorithm was unable to achieve convergence in order to select an optimal model. This could be due to the fact that in the initial model without interactions (Table 4) the algorithm does not select macroeconomic variables,

Table 13 In-sample and out-of-sample prediction/outcome cross-table with cross-variables

Real	Predicted								Total (%)
	Model 1		Model 2		Model 3		Model 4		
	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	Non failed (%)	Failed (%)	
In-Sample									
Non failed	73.50	26.50	87.40	12.60	73.84	26	72.76	27.24	100
Failed	26.00	74.00	15.45	84.55	15	84.89	17.18	82.82	100
Out-sample									
Non failed	97.50	2.50	77.00	23	80.53	19.47	85.29	14.71	100
Failed	24.00	76.00	16	84.50	8.42	91.58	10.59	89.41	100

nor does it select crossed variables with them; this outcome leads us to imagine that, 5 years ahead, macroeconomic variables are not good predictors of business failure.

3.4 Sensitivity Analysis When the Cross-Variables Act as Explanatory Variables

In order to obtain the sensitivity to change of the probability of company failure when a change takes place in one explanatory variable, while the others remain constant, we estimated this probability with respect to a change in each variable. Figure 4 illustrates the evolution of bankruptcy probability versus a gradual increase in each of the explanatory variables of Model 1.

For the sake of simplicity, we do not present the figures showing the sensitivity to change of the probability of bankruptcy, in reaction to changes in each explanatory variable, for Models 2, 3 and 4, but they are available to the reader on request. In summary, these figures show that most variables exert an influence similar to that corresponding to the models without interactions (Fig. 1), with the exception of the following variables:

- In Model 1, the variable "Price of Land" has a negligible impact on the probability of company failure, due to the negative effect of the variable "Size". The land price variable is distorted because there is a cross-variable in common, namely "Price of Land * Size", which has a negative coefficient.
- The variable "Share of GDP" presents a constant, though weak, effect on bankruptcy probability in Model 3. Nevertheless an increase in this variable in Model 4 could raise the probability of bankruptcy from 0.05 to 0.8.
- "Indebtedness" and "Interest Rates" have a stronger effect on the bankruptcy probability in models with interactions. For instance, while an increase in the Total liabilities/Total assets ratio raises the probability of company failure from 0.35 to 0.70 in Model 1 without cross-variables, in the models with interactions this variable could increase the failure probability from 0.2 to 1. On the other hand, the variable "Interest Rate" could raise the bankruptcy probability to 1 when this rate is around 3% in Model 2 with interactions; on the other hand, in the models without interactions, the maximum bankruptcy probability is reached when interest rates exceed 5%. Figure 5a, b illustrate these differences.

Table 13 shows Prediction/Outcome cross-tables, both in-sample and out-of-sample, for the models described in Table 12.

As can be seen in Table 13, in all cases the models with cross-variables present high rates of accuracy, both in-sample and out-of-sample. For Model 1 the out-of-sample accuracy is 97.5% for healthy firms and 76% for failed firms; for Model 4, this rate rises to 85.29% for healthy firms and 89.41% for failed ones. These rates are similar to those obtained for the models without interactions. The predictive performance of the models, both in and out-of-sample, were also evaluated using CAP and ROC curves, as illustrated in Fig. 6a–d. These figures show that all models are closer to the Perfect Model than to the Random Model. Finally, Table 14 shows the AR obtained in Fig. 6.

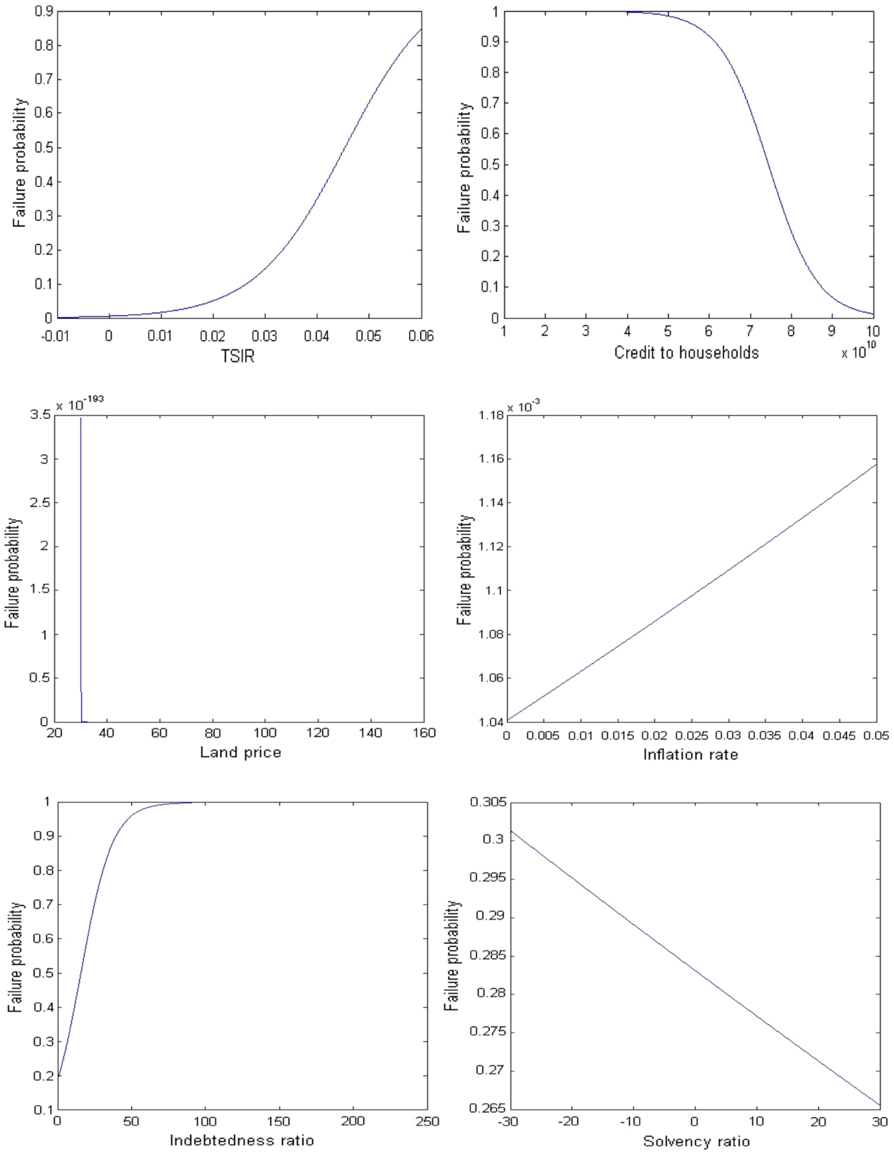

Fig. 4 Evolution of bankruptcy probability in Model 1. Evolution of bankruptcy probability versus a gradual increase in each of the explanatory variables of Model 1

4 Conclusions

This study has two main goals: first, to identify the factors that explain the probability of failure among Spanish firms in the construction sector. Second, to make a reasonable prediction of the probability of this company failure. To address the first goal,

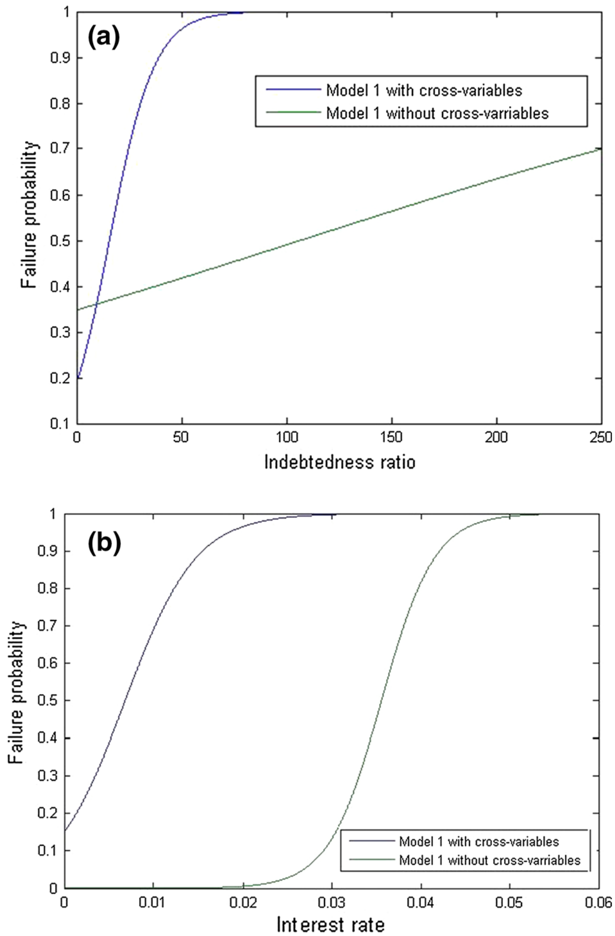

Fig. 5 **a** Comparison of the effect of the variable "Indebtedness" on bankruptcy probability with and without interactions. While an increase in the Indebtedness ratio raises the probability of company failure from 0.35 to 0.70 in Model 1 without cross variables, in the models with interactions this variable could increase the failure probability from 0.2 to 1. **b** Comparison of the effect of the variable "Interest Rates" on bankruptcy probability with and without interactions. The variable "Interest Rate" could raise the bankruptcy probability to 1 when this rate is around 3% in Model 2 with interactions; on the other hand, in the models without interactions, the maximum bankruptcy probability is reached when interest rates exceed 5%

two kinds of explanatory factors were applied: financial ratios and other internal variables describing the status of the firms, and macroeconomic variables, which are of special importance because they may reflect changes in the economic cycle, the general evolution of the sector and changes in regulatory policies. Thus, without needing an additional theoretical framework on business failure, this work provides empirical evidence about the macroeconomic variables underlying this question. Our methodology, guided solely by data and based on a genetic algorithm, consists of a selection process of explanatory variables. Starting with an initial set of financial ratios with redundant information, together with macroeconomic and sectorial variables (all of

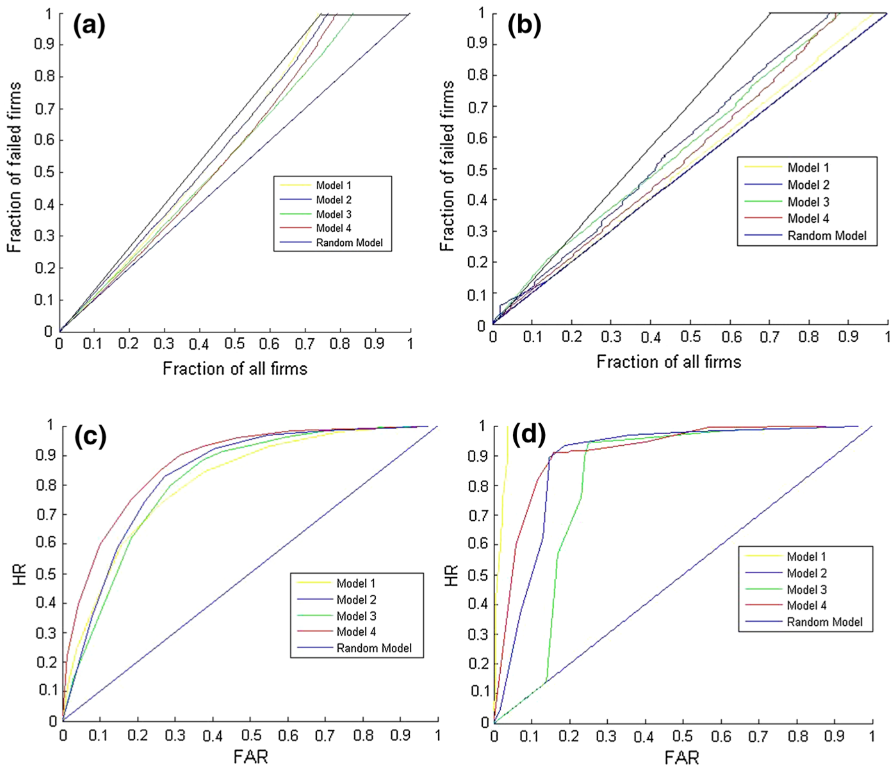

Fig. 6 In-sample and out-of-sample CAP and ROC curves with cross-variables. As explained in the text and in the captions from Figs. 2 and 3, these curves are used for visualizing and organizing forecasting models in terms of their performance. **a** In sample CAP curves, **b** Out of sample CAP curves, **c** In sample ROC curves, **d** Out of sample ROC

Table 14 Accuracy ratios for cross-variable models

	Model 1	Model 2	Model 3	Model 4
Accuracy ratio	0.9672	0.6326	0.8428	0.7727

which are candidates to explain business failure), these variables enable us to construct econometric models which incorporate both kinds of variables and predict company failure in the construction sector in Spain.

Although the firms sampled experienced bankruptcy in different years (between 1995 and 2011), the models analysed in this paper correctly classified most of them, regardless of the year corresponding to their failure; besides, the fact that the variable "Time" and its square, although included in the initial set of variables, were never selected in the final model means that the probability of company failure is not subject to a trending process. On the other hand, the explanatory variables of bankruptcy that were selected for inclusion in the models present coefficients that coincide with the effect expected. Furthermore, models with interactions among their explanatory variables were selected in order to improve their predictive accuracy. Analysis of CAP and ROC curves confirmed the conditionality of financial ratios to the macroeconomic and sectorial situation.

Up to 4 years ahead, our models are capable of classifying failure and no-failure firms with a success rate of over 80% in most cases, thus improving on the classical benchmarks of Altman's Z-score model.

It is apparent that companies in the Spanish construction sector are more sensitive to changes in macroeconomic factors than to their own accounting ratios. The ratios that were found to be most significant were those corresponding to categories of solvency and indebtedness. Solvency ratios are most evident in models that predict failure in the short term, while those referring to indebtedness are present in most models. The most important macroeconomics variables are credit and land price fluctuations, two variables that are closely linked to the sector. Finally, it should be noted that beyond 4 years our ability to forecast company failure declines considerably. In fact, no macroeconomic variable was selected when the genetic algorithm attempted to specify a model capable of predicting failure 5 years ahead, and the accounting ratios examined have little explanatory power.

The method presented in this paper enables the early detection of a risk of company failure, and thus offers firms in this sector the opportunity to take the preventive and corrective actions necessary for them to stay in business.

References

Acosta-González, E., & Fernández-Rodríguez, F. (2007). Model selection via genetic algorithms illustrated with cross-country growth data. *Empirical Economics*, *33*, 313–337.

Acosta-González, E., & Fernández-Rodríguez, F. (2014). Forecasting financial failure of firms via genetic algorithms. *Computational Economics*, *43*, 133–157.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*, 589–609.

Altman, E. (1983). Why businesses fail. *Journal of Business Strategy*, *3*, 15–21.

Altman, E. (1993). *Corporate financial distress and bankruptcy* (2nd ed.). New York: Wiley.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2016). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*,. doi:10.1111/jifm.12053.

Balcaen, S., & Ooghe, H. (2004). *35 Years of studies on business failure*. Working paper No. 248. Faculteit Economie en Bedrijfskunde.

Bauweraerts, J. (2016). Predicting bankruptcy in private firms: Towards a stepwise regression procedure. *International Journal of Financial Research*, *7*(2), 147–153.

Benito, A., Delgado, F., & Pages, J. (2004). *A synthetic indicator of financial pressure for Spanish firms*. Working paper No. 411. Banco de España.

Burn, P., & Redwood, V. (2003). *Company accounts based modelling of business failures and the implications for financial stability*. Working paper 210. Bank of England.

Cressy, R. (1992). UK small firms bankruptcy predictions: A logit analysis of industry, trend and macro effects. *Journal of Small Business Finance*, *1*, 233–253.

Dimitras, A., Zanakis, S., & Zopounidis, C. (1996). A survey of business failures with an enphasis on prediction methods and industrial applications. *European Journal of Operation Research*, *90*, 487–513.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, *83*, 635–665.

Fama, E. (1986). Term premiums and default premiums in money markets. *Journal of Financial Economics*, *17*, 175–196.

Fejér-Király, G. (2015). Bankruptcy prediction: A survey on evolution, critiques, and solutions. *Acta Universitatis Sapientiae, Economics and Business*, *3*, 93–108.
Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*, 1–141.
Frydman, H., Altman, E., & Kao, D. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *Journal of Finance*, *40*, 269–291.
Hendry, D. F., & Krolzig, H. M. (2001). *Automatic econometric model selection using PcGets 1.0*. London: Timberlake Consultants Press.
Hernandez-Tinoco, M., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394–419.
Hol, S. (2001). Explanations of losses on loans with macroeconomic variables: Norway 1991–1999. *Medium Econometric Applications*, *9*, 15–18.
Hol, S. (2007). The influence of the business cycle on bankruptcy probability. *International Transactions in Operational Research*, *14*(3), 75–90.
Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, *2*, 167–191.
Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: The University of Michigan Press.
Johnsen, T., & Melicher, R. (1994). Predicting corporate bankruptcy and financial distress: Information value added by multinomial logit models. *Journal of Economics and Business*, *46*, 269–286.
Jones, S., & Hensher, D. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review*, *79*(4), 1011–1038.
King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, *95*, 49–69.
Kumar, M., & Rao, V. S. H. (2015). A new methodology for estimating internal credit risk and bankruptcy prediction under Basel II Regime. *Computational Economics, 46*(1), 83–102.
Lane, S., & Schary, M. (1989). *The macroeconomic component of business failure, 1959–1988*. Working paper 89-31. Boston University.
Lovell, M. C. (1983). Data mining. *Review of Economic and Statistics*, *65*, 1–12.
Levy, A., & Bar-Niv, R. (1987). Macroeconomic aspects of firm bankruptcy analysis. *Journal of Macroeconomics*, *9*, 407–415.
Lua, Y., Weib, Y., & Chang, T. (2015). The effects and applicability of financial media reports on corporate default ratings. *International Review of Economics & Finance*, *36*, 69–87.
Maddala, G. S. (1983). *Limited-dependent and qualitative variables in economics*. New York: Cambridge University Press.
McKee, T. E., & Lensberg, T. (2002). Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research*, *138*, 436–451.
Melicher, R., & Heart, D. (1988). A time series analysis of aggregate business failure: A macroeconomic perspective. *Journal of Accounting, Auditing and Finance*, *6*, 20–31.
Menard, S. (2001). *Applied logistic regression analysis. Sage University papers series on quantitative applications in the social sciences, 07-106*. Thousand Oaks, CA: Sage.
Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*, 449–470.
Messier, W., & Hansen, J. (1988). Inducing rules for expert system development: An example using default and bankruptcy data. *Management Science*, *34*(12), 1403–1415.
Miller, A. (2002). *Subset selection in regression*. London: Chapman & Hall/CRC.
Mousavia, M. M., Ouennichea, J., & Xub, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, *42*, 64–75.
Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*, 109–131.
Olmeda, I., & Fernández, E. (1997). Hybrid classifiers for financial multicriteria decision making: The case of bankruptcy prediction. *Computational Economics*, *10*(4), 317–335.
Perez-Amaral, T., Gallo, G. M., & White, H. (2003). A flexible tool for model building: The relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and Statistics*, *65*, 821–838.

Rose, P., Andrews, W., & Giroux, G. (1982). Predicting business failure: A macroeconomic perspective. *Journal of Accounting, Auditing and Finance*, *6*, 20–31.
Sala-i-Martin, X. X. (1997). I just ran two million regressions. *American Economic Review*, *87*, 178–183.
Sobehart, J. R., Keenan, S. C., & Stein, R. (2000). Validation methodologies for default risk models. *Credit*, *1*(4), 51–56.
Tam, K., & Kiang, M. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, *38*(7), 926–947.
Tkáč, M., & Verner, R. (2016). Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, *38*, 788–804.
Wilson, T. (1997). *Credit risk modelling: A new approach*. New York: McKinsey Inc.