DEFAULT PREDICTION OF SPANISH COMPANIES.
A LOGISTIC ANALYSIS

Concepción BARTUAL

Universitat Politècnica de València
E-mail: conbarsa@esp.upv.es

Fernando GARCIA

Universitat Politècnica de València
E-mail: fergarga@esp.upv.es

Francisco GUIJARRO

Universitat Politècnica de València
E-mail: fraguima@esp.upv.es.

Ismael MOYA

Universitat Politècnica de València
E-mail: imoya@esp.upv.es

Abstract. In the field of credit risk management, the calculation of the probability of default of companies plays a key role. For that reason, bankruptcy prediction of companies has generated extensive research in the past decades. This paper applies one of the most popular techniques, the logistic regression. This technique is extensively used both by professionals and academics and is employed in many studies as a benchmark. Here we will apply it on a vast data base of the Spanish companies and a statistical analysis of the robustness of the model will be undertaken, with very satisfactory results.

Introduction

Credit risk analysis is one of the most important tasks to be undertaken by financial institutions. The lack of a correct methodology to calculate the probability of default of the clients may lead to high losses in the banks, create systemic risk, and affect the whole economy of a country. An example of such an event can be seen in the Spanish

case, where the economy is suffering because of the high default rates of the credits and the huge losses of the credit institutions from 2010. It has become obvious that the credit risk management undertaken by the Spanish banks in the previous decade has been inadequate.

Forecasting bankruptcy risk of enterprises is one of the first uses given to financial accounting information. In fact, the study of the firms' probability of default and the development of credit ratings had already gained relevance in 1909, when the US rating agencies started analysing the financial situation of the railway companies. Nevertheless, the origin of the bankruptcy prediction models lies in the 1960's decade. Beaver (1966) showed using 30 accounting ratios that the value of some of those ratios varied significantly between distressed and non-distressed companies. Furthermore, Altman (1968) applied discriminant analysis on several financial ratios in a multivariate context and generated a model to predict business failure. That was the starting point of a field of research that has become increasingly important, where new, more accurate, prediction models with new methodologies and bigger data bases are still developed.

Many techniques have been used after the studies by Beaver (1966) and Altman (1968) to predict corporate financial distress employing the economic and financial information from the accounting system.

Following Ravi Kumar and Ravi (2007), the methods used to analyse corporate credit risk can be divided into two big groups: statistical methods and artificial intelligence techniques. The quantity of studies undertaken in this field is enormous, which is a proof of how important it actually is in the present times to manage credit risk appropriately, and the interest showed by academics and practitioners.

Without being exhaustive, we can refer to the following methodologies and techniques: some of the most widespread statistical methods are discriminant analysis (Yim and Mitchell, 2004); the probit model (Ginoglou and Agorastos, 2002) and the logistic regression (Ohlson, 1980), which is the one to be applied in this study. Therefore, we will comment on it later. Within the artificial intelligence techniques we can count neuronal networks (Ravi and Pramodh, 2008), decision trees (Korol, 2013), rough sets (Tay and Shen, 2002; McKee, 2003), data envelopment analyses (Cielen *et al.*, 2004), support vector machines (Kim and Sohn, 2010) and genetic algorithms (Etemadi *et al.*, 2009), and goal programming (García *et al.* 2013) are the most common ones and some recent works.

Although many new techniques have been applied in the last years to predict companies' distress, the logit model is still widely accepted among researchers and practitioners. Some of the most prestigious rating agencies use logit models to generate their company classifications. And many recent academic researches apply the logistic regression to predict corporate default (Joo-Ha and Taehong, 2000; Kolari *et al.*, 2002; Lin and Piesse, 2004; Jones and Hensher, 2004; Canbas *et al.*, 2005; Chen and Zhang, 2006; Altman and Sabato, 2007; Pang-Tien *et al.* 2008; Psillaki *et al.*, 2010; Hernández and Wilson, 2013; Zaghdoudi, 2013) or as a benchmark to compare the new credit risk prediction methods (Min and Jeong, 2009; Kim and Sohn, 2010; Su and Huang, 2010; Chen, 2011; Li *et al.*, 2011; Chaudhuri, 2013). One of the biggest challenges faced by researchers when applying the logistic regression is to benefit from a large and comprehensive data

base with high quality accounting information that guarantees for the robustness of the model obtained and the accuracy of the predictions (Bartual *et al.*, 2012a).

However, one must be aware of certain robustness problems that may arise when using logit, especially in relation to the composition of the sample used to estimate the model. Researchers should pay close attention to three factors: the choice of variables to be used in the model, the influence of the sample on the model results and the cutoff point.

Whatever the variables used, the logit model finally obtained will depend on the sample on which the model is based. This means that only some of the preselected variables will actually be used in the model, since both the selection and the weighting of the variables will depend on the sample of companies.

Furthermore, whatever the chosen cutoff point, even though it will not modify neither the selected variables nor their weights, this cutoff point will affect the discrimination process and thus also the percentage of correct and incorrect predictions.

The aim of the present study is to obtain a model to predict corporate default for the Spanish manufacturing companies applying the logistic regression model. The goal is to help solving an important issue in the Spanish economy, immersed as it is in a profound economic and financial crisis. Not only financial institutions, but also clients, suppliers, the public administration and other economic agents are in the need to know the probability of default of the companies with which they interact. In order to construct the model, a database has been created with the accounting information from more than 2,000 companies for the year 2010. At that year, the effects of the financial crisis emerged and many firms were negatively affected. It is a large database, much bigger than the ones used in similar works, which makes it possible to obtain a robust model capable of accurate predictions.

The remainder of the paper is structured as follows. Section 2 describes the database and the selected explanatory variables. Section 3 presents the main results and the sensitivity analyses. Finally, section 4 is devoted to the conclusions.

1. Description of the Database and the Selected Explanatory Variables

In this section, the database employed is presented. The selection of the group of companies to analyse and the quality and veracity of the financial information employed are critical steps towards obtaining a robust model. Furthermore, in order to implement the sensibility analysis it is necessary to count with a high number of data. Finally, it is important to include in the database an elevate amount of defaulted companies. In fact, if the number of distressed companies in the subsamples were too low, a solution stating that all companies are solvent would be very difficult to beat.

After taking all of these factors into consideration, the database SABI-Informa was consulted. This database has got accounting and financial information of almost all non-financial firms in Spain, supplied by the companies themselves. In order to guarantee the veracity and accuracy of the information and to work with a homogeneous sample of companies, only firms with accounting information of 2010 were considered. This year

was chosen because it was the first one in which the Spanish economy felt the impact of the financial crisis, increasing the number of defaulted companies. It is noteworthy to mention that in the precedent years the bankrupt firms within the selected sample were almost inexistent, making any corporate distress prediction analysis virtually impossible.

Out of the 2,783 selected companies, 736 were identified as insolvent (26.5% of the sample). A firm is classified as being in financial distress when its net worth is negative or has formally defaulted on its obligations, that is, when its legal status is defined as suspended or in liquidation. Among the 736 companies that failed, 251 had a negative equity value and the remaining 485 companies had formally defaulted.

The variables employed in the solvency analysis were obtained from the balance sheets and the income statements of the companies: total assets, current assets, cash, equity, current liabilities, total sales, cost of raw materials, labour costs, operating income, financial income and pre-tax profits. Using these 11 accounting inputs, several financial ratios were calculated:

1. ET: Equity / Total assets
2. CL: Cash / Current liabilities
3. LS: Labour costs / Sales
4. OS: Operating income / Sales
5. OT: Operating income / Total assets

Altogether, 16 explanatory variables were used to obtain the corporate default prediction model. The selection of the variables was undertaken following Bartual *et al.*, 2012 b.

Summary statistics of explanatory financial and ratio variables can be found in Table 1.

Table 1. Descriptive statistics

Variable	Minimum	Maximum	Mean	Median	Standard deviation
Total assets	11	3,065,534	6,339.8	687	76,298.0
Current assets	2	1,749,916	3,205.0	391	36,195.9
Cash	0	33,991	183.9	28	1,021.2
Equity	-26,264	873,721	2,186.4	150	24,734.9
Current liabilities	1	392,810	2,095.2	288	13,422.2
Total sales	5	1,476,753	4,922.0	640	40,729.6
Cost of raw materials	0	873,897	2,954.2	286	25,195.9
Labor costs	0	125,211	814.2	210	4,122.3
Operating income	-26,019	285,172	153.1	6	6,394.7
Financial income	-42,043	22,769	-43.3	-6	1,033.8
Pre-tax profits	-18,244	873,897	2,954.2	286	5,161.0
ET	-16.657	0.998	0.231	0.266	0.659
CL	0.000	105.000	0.511	0.087	3.020
LS	0.000	15.571	0.404	0.331	0.641
OS	-37.000	33.862	-0.131	0.013	1.323
OT	-39.363	0.956	-0.077	0.013	0.807

Note: Balance sheet and income accounts are expressed in thousands of euros.
Source: The authors.

The proposed logit model makes it possible to estimate the likelihood of a firm to belong to the group of solvent companies (if the obtained value is 1) or to the group of insolvent companies (if the value is 0) using the 16 independent variables introduced above. In order to identify the most significant and robust model, the stepwise method has been applied and optimised by the Akaike index (AIC). The selected model is presented in Table 2.

Table 2. Result of the stepwise logistic regression

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.837e-01	8.698e-02	4.411	1.03e-05 ***
Current assets	-1.842e-04	5.047e-05	-3.649	0.000263 ***
Equity	1.577e-04	5.179e-05	3.046	0.002320 **
Current liabilities	1.764e-04	4.920e-05	3.585	0.000337 ***
Financial income	1.648e-03	4.108e-04	4.011	6.06e-05 ***
Pre-tax profits	8.911e-04	1.832e-04	4.864	1.15e-06 ***
Equity / Total assets	5.281e+00	3.222e-01	16.391	< 2e-16 ***
Cash / Current liabilities	8.275e-01	2.184e-01	3.790	0.000151 ***
Operating income / Total assets	5.910e+00	5.226e-01	11.309	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 3219.4 on 2782 degrees of freedom				
Residual deviance: 1632.7 on 2774 degrees of freedom				
AIC: 1650.7				
Number of Fisher Scoring iterations: 9				

Source: The authors.

The process has converged after 9 iterations and the coefficients of following explanatory variables are significant with a level of confidence of 99%: Current assets, equity, current liabilities, financial result, pre-tax profits, equity / total assets, cash / current liabilities, y operating income / total assets. The value obtained by the Akaike index is 1,650.7.

It is possible to draw some interesting conclusions by simply looking at the sign of the coefficients. The coefficient with the highest z-value is the one for the ratio equity / total assets, with a positive sign. This is due to the importance of having an elevate equity value compared with the liabilities, that is, having a reduced financial leverage, in order to avoid bankruptcy. The second coefficient with the highest z-value, having a positive sign as well, is the ratio operating income / total assets. This underlines the necessity of having high benefits compared with the size of the firm to guarantee the survival of any company. The ratio cash / current liabilities is remarkable as well. Its positive sign shows that it is imperative for the companies to have enough cash to pay the money back to creditors and settle the debts. If this condition is not met, the company will be threatened with insolvency. The remaining variables have the coefficients that share a similar economic interpretation as the variables commented above.

When the model is applied on the total database, the results shown on Table 3 are obtained. Out of the 738 defaulted corporations, the model correctly predicts the

situation of 573 (77.6%). The rest (22.4%) is incorrectly assigned to the group of solvent companies. In the case of the solvent companies, out of the 2,045 firms, 1,880 are correctly defined as solvent (91.9%), whereas only 165 (8.1%) are wrongly predicted to be in bankruptcy.

If the total number of companies in the sample is to be considered regardless their solvency / insolvency state, the model can correctly assess the credit risk in 88.1% of the cases. A naïve model would predict that all the firms are solvent[1], obtaining a success rate of 73.5%; equal to the percentage of solvent firms in the sample. Therefore we can confirm that our model beats by almost 15% the results obtained by the naïve model. This performance of our model can be considered as a good result.

Table 3. Solvency state prediction applying the logistic regression model

Observed solvency	Predicted solvency		
	0	**1**	**Row Total**
0	573	165	738
	727.386	262.499	
	0.776	0.224	0.265
	0.776	0.081	
	0.206	0.059	
1	165	1880	2045
	262.499	94.731	
	0.081	0.919	0.735
	0.224	0.919	
	0.059	0.676	
Column Total	738	2045	2783
	0.265	0.735	

Source: The authors.

2. Results and Sensitivity Analyses

The results offered in the previous section clearly indicate the kind of accounting variables that can be included in a model for the prediction of corporate insolvency in Spain. They also show the success rate of the model for each of the groups into which the sample was divided: solvent and insolvent firms. The average success rate is 88.1%. Nevertheless, the analysis might be biased as the same sample has been employed for the estimation of the model and for the prediction of the dependent variable.

In order to avoid this kind of bias, the complete simple has been divided into two subsamples, the first one including 80% of the companies and the second one the remaining 20%. The first, bigger subsample has been employed as training set to estimate the

---

[1]   A naive model considers all the firms to be solvent. So, if there are 90% of solvent firms in the sample, the naïve model would make correct predictions in 90% of the cases. But all insolvent firms would be wrongly assigned to the group of solvent companies.

prediction model. The second sample is the test set. The model obtained using the big subsample is applied on the test set and the rate of success is calculated.

Furthermore, this process has been simulated 1,000 times by randomly selecting the companies to be assigned to the training set and to the test set. Doing this, the model obtained in each simulation is slightly different, and so was the rate of success.

Figure 1 represents the histogram of the variable "success rate". This variable shows the percentage of companies belonging to the set test for which their solvency / insolvency estate was correctly predicted. We can observe that the distribution of this variable is similar to the normal distribution, with a mean value of 87.96% and a standard deviation of 0.01479. Clearly, the success rate obtained for the complete sample of companies (88.1%) can be considered to be within the range obtained for the mean success rate with a confidence level of 99%.

In conclusion, once the simulation process is undertaken, it becomes obvious that the logit model is a robust methodology to predict corporate default and that a high success rate might be expected (88.1%) for the case of the Spanish manufacturing companies when using the standard accounting variables.
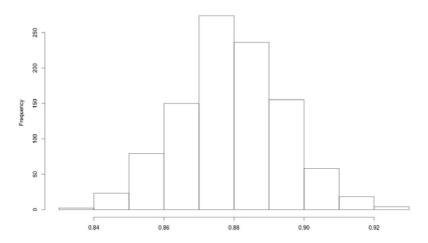
Figure 1. Histogram of the variable "success rate"

Source: The authors.

Conclusion

The present work shows the application of a multivariate statistical model, the logit model, to predict corporate distress. As predictive variables, commonly-used accounting information has been used from a database including 2,783 Spanish manufacturing firms.

In our sample, 73.5% of the firms are solvent, while 26.5% have been defined as defaulted. The logit model obtained combines information obtained from the balance sheet and from the income statement of the companies, which has been used in order to calculate several economic and financial ratios. When the model is applied on the complete sample, we obtain a correct prediction of the solvency/insolvency of the companies in 88.1% of the cases. Therefore, we can state that the logit model beats the result that would be obtained by a naïve model, which would be a success rate of only 73.5%.

With the aim of avoiding the bias that can appear when the same simple is used to generate the model and to calculate the rate of success, 1,000 simulations have been calculated. In the simulations, 80% of the companies of the sample were randomly assigned to the training set and the remaining 20% of the companies to the test set. The results make evident that the logit model is a robust methodology to explain and predict corporate bankruptcy.

For future lines of research we can propose the use of other techniques from the artificial intelligence field to be compared with the logit model, which is the traditional benchmark for credit risk evaluation. Such techniques include neuronal networks, and support vector machines. The comparison of the logit model applied in the present study and these alternative methodologies will probably reveal interesting differences regarding the number of independent variables in the models and the rate of success.

References

1. Altman, E. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23. 589-609.
2. Altman, E., Sabato, G. (2007) Modelling risk for SMEs –evidence from US market. Abacus 43 (3), 332-356.
3. Beaver, W. (1966) Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111.
4. Bartual. C., García, F., Giménez, V., Romero, A. (2012a) Credit risk analysis: Reflections on the use of the logit model. *Journal of Applied Finance and Banking* 2 (6), 1-13.
5. Bartual, C., García, F., Guijarro, F., Romero-Civera, A. (2012b) Probability of default using the logit model: The impact of explanatory variable and data base selection. 2nd International Scientific Conference Whither Our Economies. Conference Proceedings. Mykolas Romeris University.
6. Canbas, S., Cabuk, A., Kilic, S. B. (2005) Prediction of comercial bank failure via multivariate statistical analysis of financial structure: The Turkish case. *European Journal of Operational Research* 166, 528-546.
7. Chaudhuri, A. (2013) Bankruptcy prediction using bayesian, hazard, mixed logit and rough Bayesian models: A comparative analysis. *Computer and Information Science*, 6, 103-125.
8. Chen, J., Zhang, J. (2006) Financial distress prediction in China. *Review of Pacific Basin Financial Markets and Policies*, 9 (2), 317-336.

9. Chen, M. Y. (2011) Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38, 11261-11272.

10. Cielen, A., Peeters, L., Vanhoof, K. (2004) Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, 154, 526-532.

11. Etemadi, H., Rostamy, A., Dehkordi, H. (2009) A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications*, 36 (2), 3199-3207.

12. García, F., Giménez, V., Guijaro, F. (2013) Credit risk management: A multicriteria approach to assess creditworthiness. *Mathematical and Computer Modelling*, 57, 2009-2015.

13. Ginoglou, D., Agorastos, K. (2002) Predicting corporate failure of problematic firms in Greece with LPM, logit, probit and discriminant analysis models. *Journal of Financial Management and Analysis*, 15 (1), 1-15.

14. Hernández, M., Wilson, N. (2013) Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*.

15. Jones, S., Hensher, D. A. (2004) Predicting firm financial distress: A mixed logit model. *Acounting Review* 79 (4), 1011-1038.

16. Joo-Ha, N. Taehong, J. (2000) Bankruptcy prediction- evidence from Korea listed companies during the IMF crisis. *Journal of International Financial Management and Accounting*, 11 (3), 178-197.

17. Kim, H. S., Sohn, S. Y. (2010) Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, 201, 838-846.

18. Kolari, J., Glennon, D., Shin, H., Caputo, M. (2002) Predicting large US commercial bank failures. *Journal of Economics and Business* 54 (32), 318-324.

19. Korol, T. (2013) Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, 31, 22-30.

20. Li, H., Adeli, H., Sun, J., Han, J.G. (2011) Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction. *Computers and Operations Research*, 38, 409-419.

21. Lin, L., Piesse, J. (2004) Identification of corporate distress in UK industrials – a conditional probability analysis approach. *Journal of Applied Financial Economics*, 14. 73-82.

22. McKee, T. E. (2003) Rough sets bankruptcy prediction models versus auditor signalizing rates. *Journal of Forecasting*, 22, 569-589.

23. Min, J. H., Jeong, C. (2009) A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36, 5256-5263.

24. Ohlson, J. A. (1980) Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 109-131.

25. Pang-Tien, L., Chin-Wen, L., Hui-Fun, Y. (2008) Financial early-warning models on cross-holding groups. *Journal of Industrial Management and Data Systems*, 108 (8), 1060-1080.

26. Psillaki, M., Tsolas, I. E., Margaritis, D. (2010) Evaluation of credit risk based on firm performance. *European Journal of Operational Research*, 201, 873-881.

27. Ravi, V., Pramodh, C. (2008) Threshold accepting trained principal component neuronal network and feature subset selection: Application to bankruptcy prediction in banks. *Applied Soft Computing*, 8 (4), 1539-1548.

28. Ravi Kumar, P., Ravi, V. (2007) Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180, 1-28.

29. Su, E. D., Huang, S. M. (2010) Comparing firm failure predictions between logit, KMV, and ZPP models-. Evidence from Taiwan's electronics industry. *Asia-Pacific Financial Markets*, 17, 209-239.

30. Tay, F. E. H., Shen, L. (2002) Economic and financial prediction using rough set model. *European Journal of Operational Research*, 141 641-659.

31. Yim, J. Mitchell, H. (2004) A comparison of Japanese failure models – hybrid neuronal networks, logit models and discriminant analysis. *International Journal of Asian Management*, 3, 103-120.

31. Zaghdoudi, T. (2013) Bank failure prediction with logistic regression. *International Journal of Economics and Financial Issues*, 3 (2), 537-543.

ISPANIJOS KOMPANIJŲ ĮSIPAREIGOJIMŲ NEĮVYKDYMO NUMATYMAS. LOGISTINĖ ANALIZĖ

Santrauka. Straipsnyje nagrinėjamas kredito rizikos valdymas bei pateikiami kompanijų įsipareigojimų neįvykdymo tikimybiniai skaičiavimai, kurie labai svarbūs rizikos valdymui. Dėl šios priežasties pastaraisiais dešimtmečiais kompanijų bankroto prognozėms buvo skirta daug mokslinių tyrimų. Šiame straipsnyje pasitelkta populiariausia metodika – logistinė regresija. Ši metodika populiari tiek tarp profesionalų, tiek tarp akademinės bendruomenės atstovų ir taikoma kaip standartinė procedūra daugelyje studijų. Šiame straipsnyje logistinė analizė, pritaikyta didelei ispanų kompanijų duomenų bazei, bei atlikta modelio patikimumo statistinė analizė davė labai gerus rezultatus.

Concepción BARTUAL – PhD, Department of Economy and Social Sciences, Faculty of Business Administration of Universidad Politecnica de Valencia. Research fields: finance, valuation, risk management.

Concepción BARTUAL – Valensijos politechnikos universiteto Verslo administravimo fakulteto Ekonomikos ir socialinių mokslų katedra, mokslų daktaras. Moksliniai interesai: finansai, vertinimas, rizikos valdymas.

Fernando GARCÍA – PhD, Department of Economy and Social Sciences, Faculty of Business Administration of Universidad Politecnica de Valencia. Research fields: finance, valuation, risk management, portfolio management.

Fernando GARCÍA – Valensijos politechnikos universiteto Verslo administravimo fakulteto Ekonomikos ir socialinių mokslų katedra, mokslų daktaras. Moksliniai interesai: finansai, vertinimas, rizikos valdymas, portfelių valdymas.

Francisco GUIJARRO – PhD, Department of Economy and Social Sciences, Faculty of Business Administration of Universidad Politecnica de Valencia. Research fields: finance, valuation, risk management, mathematical modelling, decision making.

Francisco GUIJARRO – Valensijos politechnikos universiteto Verslo administravimo fakulteto Ekonomikos ir socialinių mokslų katedra, mokslų daktaras. Moksliniai interesai: finansai, vertinimas, rizikos valdymas, matematinis modeliavimas, sprendimų priėmimas.

Ismael MOYA – PhD, Department of Economy and Social Sciences, Faculty of Business Administration of Universidad Politecnica de Valencia. Research fields: finance, risk management, company valuation, default prediction.

Francisco GUIJARRO – Valensijos politechnikos universiteto Verslo administravimo fakulteto Ekonomikos ir socialinių mokslų katedra, mokslų daktaras. Moksliniai interesai: finansai, rizikos valdymas, kompanijų vertinimas, prognozavimas.