Journal of Scientific & Industrial Research
Vol. 78, November 2019, pp. 733-737

Predicting Sovereign Debt Crises with Fuzzy Decision Trees

David Alaminos[1], Sergio M. Fernández[1], Paulo Magalhães Neves[2] and José António C Santos[3]*
[1] University of Malaga, PhD Program in Mechanical Engineering and Energy Efficiency, 29071, Malaga, Spain
[2]Universidade do Algarve, 8005-139 Faro, Portugal
[3] Universidade do Algarve, ESGHT and CINTURS Research Centre, 8005-139 Faro, Portugal

Received 15 August 2018; revised 10 May 2019; accepted 14 September 2019

Considering the great capacity of data mining techniques to extract useful information from large databases and to manage heterogeneous variables, this paper uses Fuzzy C4.5 Decision Trees for the prediction of sovereign debt crises. To this end, prediction models have been constructed for different regions, and another global model for the whole world. The results obtained show that Fuzzy C4.5 Decision Trees method overcomes the predictive power of existing models in the previous literature and provides more explanatory information on the reasons that cause sovereign debt crises.

Keywords: Sovereign debt crisis, C4.5 algorithm, Fuzzy decision trees, Default prediction

Introduction

A sovereign debt crisis occurs when a country does not have the capacity to pay its public debt. Due to the economic and social impact of these crises, which directly affect the efficient use of public resources and threaten private companies[1], especially banks[2], the development of models capable of predicting these events has increased in recent years[3]. In this way, existing models on sovereign debt crises prediction have shown that the best predictors are the volume of external debt, the growth of foreign exchange reserves and the ability of export revenues to pay off debt[4]. However, the precision results have been poor, and the current literature demands more research to solve this limitation [3,5]. The aim of this study is to contribute to increase the accuracy of the prediction models of sovereign debt crises. To do this, new models have been constructed using the C4.5 computational algorithm in its diffuse variant, which has obtained excellent prediction results in previous studies related to economics and finance[6].

Methods

Fuzzy C4.5 decision trees

The Fuzzy C4.5 model is based on the C4.5 algorithm. This algorithm can be used to establish a decision tree according to the attributes that are divided into smaller subsets, where the process of forming a decision tree or rule depends on the decision to obtain a value from the information[6,7]. In general, the decision tree algorithm C4.5 is formed fulfilling the following order: a) select the attributes as root; b) create a branch for each value; c) process the cases of each branch; d) repeat the process for each branch until all the cases of the branches have the same class. The highest gain is used for the selection of attributes as the root, according to equation (1).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} = Entropy(S) \quad \ldots (1)$$

where S is the set of cases, A is the attributes, n is the partition number of the attribute A, and S_i is the number of cases in the *i-th* partition. For its part, the Entropy value is calculated according to equation (2).

$$Entropy(S) = \sum_{i=1}^{n} -p_i * log_2 * p_i \quad \ldots (2)$$

where S represents the set of cases, n is the number of partitions of S, and $p_i$is the proportion of S.

Fuzzy decision trees allow data to simultaneously follow multiple branches of a node with different degrees of satisfaction in the interval (0.1)[8,9]. Fuzzy decision trees differ from traditional decision trees because they use division criteria based on fuzzy constraints, their inference procedures are different, and the fuzzy sets representing the data must be defined. In addition, the induction of diffuse decision tree has two main components, referred to a procedure for the creation of a fuzzy decision tree and an inference procedure for decision making.

*Author for Correspondence
E-mail: jasantos@ualg.pt

Sensitivity analysis

The objective of the sensitivity analysis is to determine the relative importance of the independent variables in relation to the dependent variable[10]. A variable is considered more important than another if it increases the variance, compared to the set of variables in the model. To this end, applying the Sobol' method[11], the total variance $V(Y)$ is decomposed according to equation (3).

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>1} V_{ij} + \ldots + V_{12\ldots k} \quad \ldots (3)$$

where $V_i = V(E(Y|X_i)$ y $V_{ij} = V(E(Y|, X_j)) - V_i - V_j$.

For its part, the sensitivity indexes are determined by $S_i = V_i/V$ y $S_{ij} = V_{ij}/V$, where S_{ij} indicates the effect of interaction between two factors. The decomposition of Sobol allows the estimation of a total sensitivity index S_{Ti}, which measures the sum of all the sensitivity effects involved in the independent variables.

Data and Variables

The database used in the present study includes 30 explanatory variables with an annual frequency for the period 1970-2017 and a set (unbalanced) of 50 developed and emerging countries. This database has been constructed from the sample used by [1], to which the time range and countries have been extended, and new variables have been included on credit rating and political conditions of the countries. The countries' Macroeconomic variables have been extracted from four different databases: IMF International Financial Statistics, World Bank Development Indicators, World Economic Outlook, and the World Bank Global Financial Database; the credit rating indicators come from the statistics of Fitch Ratings (Sovereign Rating History); and the political variables of the Polity IV of Centre for Systemic Peace project. Table 1 shows the six categories in which the independent variables used in this study can be classified. On the other hand, the dependent variable has been defined according to the

Table 1 — Independent variables

Attribute	Factors	Concept	Exp.Sign
DebtExposure	TotalDebt	Gross external debt as % of GDP	+
	IMF Credit	Loans from IMF as % of GDP	+
	Global Interest	Global lending interest rate	+
External Sector	ForeignExch Reserves	Total reserves (without gold) as % of GDP	-
	TradeOpenness	Ratio of exports plus imports to GDP	+/-
	ExportGrowth	Annualexportsgrowthrate	-
	CurrentAccount	Current account balance as % of GDP	-
	FDI	Net FDI inflows as % of GDP	-
Domestic Macroeconomic Factors	Real GDP Growth	Annual growth of real GDP	-
	Inflation	Rate of change in CPI	+
	M2/Reserves	Ratio of M2 to foreign exch. reserves	+
	REER Overall	Deviation of real effective exchange rate from 5-year rolling mean	-
	GovSpending	General government final spending as % of GDP	+/-
	NationalSaving	Ratio to GDP	-
	Contagion	Event of sovereign debt crisis in any country of the same region (t-1)	+
Banking Sector	DomesticCredit	Ratio of domestic credit to GDP	+/-
	Bank Assets	Ratio of bank assets to GDP	-
	Gov Bank Loans	Net bank claims on central government	+
Credit Rating Indicators	SovCredit Local Rating	Sovereign long-term credit quality step in local currency	+/-
	SovCreditForeign Rating	Sovereign long-term credit quality step in foreign currency	+/-
	Sov Bond Spread	Interest rate paid on 10-year sovereign bond	+
	Credit Default Swap	Sovereign default risk as pricing	+
Political Factors	Fragment	Polity fragmentation score (regional/ethnic tensions)	+
	Polity	Combined polity score (autocracy score minus democracy score)	+/-
	Durable	Regime durability (control variable of Polity)	+
	Persist	Polity persistence (control variable of Polity)	+
	Civil War	Magnitude score of episodes of civil warfare involving that state (year)	+
	Scoup	Successfulcoupsd'etat (binary)	+
	SFI	Statefragilityindex	+
	Eco Effectiv	GDP per capita	-

criteria used by [3]. For the case of emerging economies, the dependent variable is denoted as 1 if any of the following four events occurs and zero otherwise: a) accumulated interest and / or capital arrears exceed 5% of the outstanding debt; b) receive a loan from the IMF that exceeds 100% of the country's quota; c) the accumulated credit obtained from the IMF increases above 200% of the quota; d) participate in a debt restructuring that involves more than 20% of the outstanding debt. For developed countries, in addition to the two events involving IMF loans, the dependent variable is also denoted as 1 if the outstanding public debt exceeds 150% of the nominal value of GDP. All the variables considered have been referred to a selection of countries covering four regions: Africa and the Middle East, South and East Asia, Latin America and Western Europe.

Results and Discussion

Table 2 reports on the accuracy and the significant variables in the global and regional prediction models obtained in the present study. These models have been developed using 500 random data sets, to which 10-fold cross-validation was applied, randomly dividing and mutually exclusive, the available set of samples by 70% for training sample, and 30% for testing sample. The accuracy rates reached with the training sample reach 100% for the models of Africa & Middle East, South & East Asia and developed countries, while for the cases of Latin America and global model they reach 99.22% and 98.95%, respectively. With the testing sample, the accuracy achieved is 100% for Africa & Middle East, 98.85% for Latin America, 96.82% for Asia, 99.76% for developed countries, and 97.80% for the global model shows the goodness of the estimated models offering a graphic representation of their corresponding ROC curves. Table 3 shows the sensitivity of the variables considered as possible predictors in this study.

The results obtained indicate a set of significant variables that are repeated in practically all the estimated models. These variables are Total Debt,

Table 2 — Results of accuracy evaluation

Method	Classification (%) Training Testing	RMSE Training Testing	Model Selection AIC BIC	ROC Curve	Significant Variables
Africa & Middle East	100.00 100.00	0.06 0.07	85.26 88.31	0.99	Total Debt, IMF Credit, Real GDP Growth, M2/Reserves, Gov Spending, Sov Foreign Rating, Sov Bond Spread, Polity
Latin America	99.22 98.85	0.15 0.19	126.11 133.72	0.95	IMF Credit, Global Interest, Trade Openness, REER Overall, Inflation, Gov Spending, Sov Local Rating, Sov Bond Spread, SFI
South & East Asia	100.00 96.82	0.09 0.26	101.04 106.68	0.92	Total Debt, Current Account, Real GDP Growth, REER Overall, Gov Spending, National Savings, Sov Local Rating, Polity
Developed	100.00 99.76	0.08 0.14	42.85 49.21	0.97	Total Debt, Global Interest, Current Account, M2/Reserves, Gov Spending, National Savings, Credit Default Swap, Eco Effectiv
Global	98.95 97.80	0.16 0.22	366.17 387.28	0.94	Global Interest, ForeignExch Reserves, Trade Openness, Inflation, Sov Foreign Rating, Sov Bond Spread, Credit Default Swap, SFI

RSME: Root of the Mean Square Error; AIC: Akaike Information Criteria; BIC: Bayesian Information Criteria.

List of Countries:

Africa & Middle East: Algeria, Angola, Central African Republic, Egypt, Jordan, Lebanon, Morocco, Mozambique, Namibia, Tunisia, Nigeria, Seychelles, South Africa.

LatinAmerica: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, DominicanRepublic, Ecuador, Mexico, Panama, Paraguay, Peru, Uruguay, Venezuela.

South & East Asia: Bangladesh, Indonesia, Philippines, China, India, Malaysia, Thailand, South Korea, Singapore, Vietnam

Developed: Austria, Belgium, Denmark, Finland, Germany, Greece, Ireland, Italy, Japan, Portugal, Spain, Sweden, Switzerland, United Kingdom.

Table 3 — Variable importance values of factors for Sovereign Debt Crisis

Factors	Africa&Middle East	LatinAmerica	South & East Asia	Developed	Global
TotalDebt	0.493	0.000	0.428	1.351	0.000
IMF Credit	1.370	0.427	0.000	0.000	0.000
Global Interest	0.000	0.317	0.000	0.087	0.114
ForeignExch Reserves	0.000	0.000	0.000	0.000	0.752
TradeOpenness	0.000	1.145	0.000	0.000	0.875
ExportGrowth	0.000	0.000	0.000	0.000	0.000
CurrentAccount	0.000	0.000	1.157	0.168	0.000
FDI	0.000	0.000	.0.000	0.000	0.000
Real GDP Growth	0.275	0.000	0.185	0.142	0.000
Inflation	0.000	0.227	0.000	0.000	0.375
M2/Reserves	0.275	.0.000	.0.000	1.145	0.000
REER Overall	0.000	0.329	0.626	0.000	0.000
GovSpending	0.452	0.175	0.000	0.325	0.000
NationalSaving	0.000	0.000	0.502	0.124	0.000
Contagion	0.000	0.000	0.000	0.000	0.000
DomesticCredit	0.000	0.000	0.000	0.000	0.000
Bank Assets	0.000	0.000	0.000	0.000	0.000
Gov Bank Loans	0.000	0.000	0.000	0.000	0.000
SovCredit Local Rating	0.000	0.452	0.155	0.000	0.000
SovCreditForeign Rating	0.355	0.000	0.000	0.000	0.452
Sov Bond Spread	0.184	0.175	0.000	0.000	0.217
Credit Default Swap	0.000	0.000	0.000	0.137	0.000
Fragment	0.000	0.000	0.000	0.000	0.000
Polity	0.452	0.000	0.317	0.000	0.000
Durable	0.000	0.000	0.000	0.000	0.000
Persist	0.000	0.000	0.000	0.000	0.000
Civil War	0.000	0.000	0.000	0.231	0.000
Scoup	0.000	0.000	0.000	0.000	0.000
SFI	0.000	0.625	0.000	0.000	0.375
Eco Effectiv	0.000	0.000	0.000	0.175	0.000

Global Interest, and Gov Spending, all of them related to the increase in debt levels. These results show the importance of the level of indebtedness in the probability of default. Our results also confirm that other significant variables such as Sov Local Rating and Sov Bond Spread have not been tested by the previous literature. These variables show that a worsening of the country's credit rating and an increase in the interest paid make it more difficult to obtain financing and pay the debt, increasing the risk of default. Finally, the most significant political variables in our models have been: SFI and Polity. These variables are related to the ability to implement public policies and the level of democracy in the country, respectively. While Polity has been shown to be significant in previous literature, SFI has not been previously.The results of this study show, therefore, that the models developed with the diffuse C4.5 increase the capacity to predict sovereign debt crises, in comparison with previous studies. Mainly, it is worth highlighting the case of the global model, which achieves an accuracy of 97.8%, higher than the 87.1% obtained by [3] using logistic regression. It also improves the results obtained by [5], which reached 87.0% accuracy with regression trees for emerging economies.

Conclusions

The sovereign debt crises are a worldwide phenomenon that has been the focus of concern for researchers and public policy makers in the last decade. Our results show that decision trees, specifically the Fuzzy C4.5 algorithm, improve the accuracy of sovereign debt crisis prediction models, as well as showing a unique set of variables and with differences to the factors shown by studies.

Previously, they offer a greater amount of information for the policymakers of the regions considered that require empirical tools to mitigate and resolve the risk of non-payment of the debt and its negative effects. Also, our models can be of special relevance to financial institutions, such as rating agencies and central banks, that need to control the risk of a possible imminent crisis. Finally, this study has used both variables usually used by the previous literature and new variables such as credit rating and policies. These new variables have turned out to be significant in the detection of sovereign debt crises and contribute to a greater precision of the prediction models.

References

1 Fernández-Gámez M A, Diéguez-Soto J, Santos José António C & Rosa J M, Bankrupt cyprediction of family firms using combined classifiers. *J Sci Ind Res*, **78**(5) (2019), 269-273.

2 Martins, A I, Efficiency determinants in Portuguese banking industry – an application through fractional regression models. *Tour & Manage Stud*, **14**(2)(2018), 63-71.

3 Dawood M, Horsewood N & Strobel F, Predicting sovereign debt crises: An early warning system approach. *J Financ Stabil*, **28**, (2017), 16-28.

4 França A. & Rua A, Relationship between intangible resources, absorptive capacities and export performance, *Tour & Manage Stud*, **14**(1)(2018), 94-107.

5 Chakrabarti A & Zeaiter H, The determinants of sovereign default: a sensitivity analysis. *Int Rev Econ Finc,* **33** (2014), 300–318.

6 Rawal B & Agarwal R, Improving accuracy of classification based on c4.5 decision tree algorithm using big data analytics. Computational intelligence in data mining. *Adv Intell Sys Comput*, **711**(2019), 203-211.

7 Naseer T & Asghar S, *An evaluation of C4.5 and Fuzzy C4.5 with effect of pruning methods. Artificial intelligence: concepts, methodologies, tools, and applications*, IGI Global, (2017).

8 Lee, T.C, Chung, P.H, & Shyu, Performance evaluation of medical device manufacturers using a hybrid Fuzzy MCDM. *J Sci Ind Res*, **76**(1) (2017), 28-31.

9 Prashanth K D, Parthiban P & Dhanalakshmi R, Evaluation and ranking of criteria affecting the supplier's performance of a heavy industry by Fuzzy AHP Method. *J Sci Ind Res*, **77**(5) (2018), 268-270.

10 Groen E A, Bokkers E A M, Heijungs R& de Boer I.J.M, Methods for global sensitivity analysis in life cycle assessment. *Int J Life Cycle Ass*,**22**(7)(2017), 1125–1137.

11 Iooss B & Lemaître P, Uncertainty management in simulation-optimization of complex systems: Algorithms and applications. *Rev Glob Sensitiv Anal Method*,(2015), 101–122.