SYRTO

SYstemic Risk TOmography
Signals, Measurements, Transmission Channels, and Policy Interventions

A Data-Driven Explanation of Country Risk: Emerging Markets vs. Eurozone Debt Crises

Roberto Savona, Marika Vezzoli, Enrico Ciavolino

SYRTO WORKING PAPER SERIES
Working paper n. 17 | 2015

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 320270.

This documents reflects only the author's view. The European Union is not liable for any use that may be made of the information contained therein.

A Data-Driven Explanation of Country Risk: Emerging Markets vs. Eurozone Debt Crises†

Roberto Savona♣, Marika Vezzoli♣♣ and Enrico Ciavolino♣♣♣

† The authors was supported by the European Union's Seventh Framework Programme (FP7-SSH/2007-2013) for research, technological development and demonstration under grant agreement no 320270 – SYRTO.

♣ Department of Economics and Management, University of Brescia. Address: Dipartimento di Economia e Management, Università degli Studi di Brescia, c/da S. Chiara 50, 25122 Brescia (Italy). E-mail: roberto.savona@unibs.it.

♣♣ Department of Molecular and Translational Medicine, University of Brescia. Address: Dipartimento di Medicina Molecolare e Traslazionale, Università degli Studi di Brescia, Viale Europa 11, 25123 Brescia (Italy). E-mail: marika.vezzoli@med.unibs.it.

♣♣♣ Department of History, Society and Human Studies, University of Salento. Address: Palazzo Parlangeli, Via V. M. Stampacchia 45, 73100, Lecce (Italy). E-mail: enrico.ciavolino@unisalento.it.

1. Introduction

As pointed out by Reinhart and Rogoff (2013), even after the global financial crisis of 2007-2010, the common view within policy circles is that developed countries economies are "completely different animals from their emerging market counterparts". The reason is because advanced economies can implement countercyclical policy measures more easily than their developing counterparts, and thus they do not need the standard toolkit that, instead, emerging markets use to manage financial crises. This question clearly involves the ex-post phase of a crisis, when the objective is stabilization, but also inspecting the ex-ante phase of the crisis, prevention, is important as well in order to take the right measures to mitigate, first, and restore, then, economies that suffered a financial crisis.

Existing literature on the determinants of debt crises is abundant and mainly focused on solvency and liquidity risk dimensions also including the willingness to pay the debt. The statistical evidence on the forecasting ability of some leading indicators came up with a list of factors that help predict impending debt crises (see for e.g. Savona and Vezzoli, 2013, and the references herein):

- A first class of *solvency-based leading indicators* include international reserves, capital flows, foreign direct investment, real exchange rate, current account balance, exports and imports, as well as public foreign debt, total foreign debt, short-term debt and debt servicing. Together, these proxies act as signals of potential current and capital account problems and, thus, of debt sustainability.
- *Liquidity-based indicators* are very closed to the first class, while they are more focused on the liquidity side of the solvency profile of a country. As a result, the most used predictors are short-term debt over reserves or M2 and debt service over reserves.
- Finally, *willingness-to-pay-based indicators* includes both macroeconomic variables, such as real GDP growth, inflation, exchange and interest rates, and political and institutional variables, such as financial liberalization, political instability and political rights.

The key question we face here is whether these factors, which have been scrutinized massively in emerging market debt crises, are also important for advanced economies. In this perspective, the Eurozone sovereign debt crisis is an interesting laboratory in light of the surge in the government bond spreads shown by Greece, Ireland, Portugal, Spain (GIPS) in 2010-2011, and the political debate that followed about the fiscal austerity vs. growth and the loop between banking and sovereign risks.

We do this by using data from emerging markets and GIPS over the period 1975-2010 with the end to select the more important crisis predictors for the two groups of countries (emerging markets and GIPS) and estimate their contribution in explaining the sovereign defaults occurred in the years 1975-2010, for emerging markets, and in the 2007-2010 global financial crisis, for European countries.

In our empirical analysis we proceed, firstly, by implement a novel Regression Tree-based algorithm used in Vezzoli and Savona (2013) together with Random Forrest technique (Breiman, 2001). Secondly, we focus on the GIPS sub-sample and inspect how the selected variables by the Random Forest are related to the sovereign default probabilities, also considering interactions among single predictors. Our empirical findings prove that developing and developed countries differ in terms of the inner sources of debt crises. For emerging markets, we confirm the results in Savona and Vezzoli (2013) proving that sovereign defaults are essentially driven by: (1) default history (number of past defaults), (2) US interest rates, (3) short-term debt to reserves, (4) contagion (number of other countries that experience a default in the same year). Instead, for GIPS we find that sovereign defaults can be explained (and hopefully forecasted) by: (1) inflation, (2) public debt over GDP, (3) export growth, (4) real GDP growth; (5) US Treasury Bill rates. Moreover, by linearly mapping these predictors in a OLS regression also including their main interactions, we prove that inflation and public debt over GDP are significant only when interacted one each other and together with real GDP growth and export growth; finally, export growth is the unique leading indicator showing a significant impact both as stand-alone variable and interacted with other leading indicators.

The structure of the chapter is the following. Section 2 discuss data and methods, section 3 present empirical results, and section 4 provides conclusions.

2. Data and Methods

Understanding the economic process underlying sovereign defaults is particularly challenging essentially because debt crises exhibit different and complex reasons conditional on time, idiosyncratic and systematic factors. Furthermore, such a multidimensional nature is complicated by the fact that, as pointed out by many authors (Bulow and Rogoff, 1989; Eaton and Gersovitz, 1981), a sovereign default is endogenously triggered by political decisions, which in turn imply a trade-off between the debt payment costs and unwillingness debt payment costs, for e.g. costs connected to reputation or international trade impediment.

Many authors forced the default process to be model-based, suggesting in many cases linear relationships between the crisis and a set of informative variables. Although these approaches are intellectually appealing and easy to estimate, usually they are not able to explain the intricate and non-linear relationship between sovereign defaults and economic predictors. This is the reason why we rely on data mining techniques, thereby letting the data to speak about a puzzling and partly unknowable process in which a priori theory could give us wrong explanations about the real reasons underlying the sovereign default process.

Our methodology is mainly based on Regression Trees which has been recently used in financial crisis studies and that proven to be extremely useful to detect the most important predictors, also explaining the complex and non-linear nature of defaults of banks (Manasse, Savona and Vezzoli, 2013), sovereigns (Manasse and Roubini, 2009) and corporations (Lin and McClean, 2001).

Regression Trees analysis, introduced in Breiman et al. (1984), is a recursive predictor space partition by a series of subsequent nodes that collapse into distinct and homogeneous partitions (final nodes or regions). Computationally, for each covariate in the sample, the algorithm partitions the overall sample in two sub-samples based on the best values that guarantee the maximum homogeneity within the two regions. The homogeneity degree of the sub-samples is measured through the Gini index for classification trees (when the dependent variable is categorical), or by the sum of squared errors for Regression Trees (when the dependent variable is continuous). The variable showing the maximum homogeneity (measured by the Gini Index or the sum of squared errors) is the first indicator on the top of the tree which splits the overall sample in two sub-samples, based on the corresponding splitting value (threshold). The same procedure is next re-computed for each sub-sample using the same covariates ending up in a collection of binary splits in the form of "$x_{ji} \leq s_j$" and "$x_{ji} > s_j$" for each *j* predictor *x* relative to the corresponding threshold value *s* of the *i*-th observation.

More formally, let denote by T a tree with $m = 1, \dots, M$ terminal nodes, i.e. the disjoint regions $\tilde{T}_m$, and by $\Theta = \theta_1, \dots, \theta_M$ the parameter that associates each m-th θ value with the corresponding node, then a generic dependent variable Y conditional on Θ assumes the distribution $f(Y|\Theta)$, and according to whether the Y is quantitative or qualitative the model is called Regression Tree or Classification Tree, respectively. Mathematically,

$$f(y_i|\Theta) = \sum_{m=1}^{M} \theta_m I\left(\mathbf{X} \in \tilde{T}_m\right)$$

where θ_m represents a specific $\tilde{T}_m$ region and I denotes the indicator function that takes the value of 1 if $\mathbf{X} \in \tilde{T}_m$. This signifies that predictions are computed by the average of the Y values within the terminal nodes, i.e.

$$\hat{y}_i = \hat{\theta}_m \Rightarrow N_m^{-1} \sum_{\mathbf{x}_i \in \tilde{T}_m} y_i$$

with $i = 1, \dots, N$ the total number of observations and N_m the number within the m-th region. Computationally, the optimal tree is obtained by minimizing the following loss function[1]

$$\underset{\Xi = \{T, \Theta\}}{\arg\min}\, L = \left(Y - f(Y|\Theta)\right)^2,$$

which entails selecting the optimal number of regions and corresponding splitting values.
Let s^* be the best split value and $R(m) = N_m^{-1} \sum_{\mathbf{x}_i \in \tilde{T}_m} \left(y_i - \hat{\theta}_m\right)^2$ be the measure of the variability within each node, the fitting criterion is given by

$$\Delta R\left(s^*, m\right) = \max_{s^*} \Delta R(s, m)$$

with

$$\Delta R(s, m) = R(m) - \left[R(m_1) + R(m_2)\right].$$

The procedure is run for each predictor then ranking all of the best splits on each variable according to the reduction in impurity achieved by each split. The selected variables and corresponding split points are those that most reduce the loss function in each partition.

Another interesting feature of Regression Trees is that they are conceived with the end to improve the out-of-sample predictability. The estimation process is indeed based on the cross-validation, through which the data are partitioned into subsets such that the analysis is initially performed on a single subset (the training sets), while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis (the validation or testing sets).

[1] In See Hastie et al. (2009) for technical details.

In our empirical analysis we used a novel Regression Tree-based approach (CRAGGING) together with the Random Forest algorithm in order to select the more relevant predictors, finally running a pooled OLS regression with the end to assess the significance of selected indicators, their interactions and the relationship between defaults and predictors as a whole.

Table 1: List of Countries

Emerging Market (A)	# of Defaults	Emerging Market (B)	# of Defaults	GIPS	# of Defaults
Algeria	1	Madagascar	2	Greece	1
Argentina	3	Malawi	3	Ireland	1
Bangladesh	1	Malaysia	0	Portugal	0
Bolivia	2	Mali	0	Spain	0
Botswana	0	Mauritius	2		
Brazil	3	Mexico	4		
Burkina Faso	1	Moldova	3		
Burundi	0	Morocco	2		
Cameroon	2	Nicaragua	1		
Chile	1	Niger	1		
Costa Rica	1	Nigeria	1		
Czech Rep.	0	Oman	0		
Dominican Rep.	1	Pakistan	3		
Ecuador	3	Papua New Guinea	0		
Egypt, Arab Rep.	2	Paraguay	2		
El Salvador	1	Peru	3		
Estonia	0	Philippines	4		
Ethiopia	2	Poland	1		
Gabon	3	Romania	2		
Haiti	1	Senegal	0		
Honduras	2	Sierra Leone	3		
Hungary	3	Slovak Republic	0		
India	1	South Africa	3		
Indonesia	2	Sri Lanka	1		
Jamaica	4	Thailand	3		
Jordan	2	Trinidad and Tobago	1		
Kazakhstan	1	Tunisia	1		
Kenya	4	Turkey	5		
Korea, Rep.	3	Ukraine	2		
Latvia	1	Uruguay	5		
Lebanon	0	Venezuela	3		
Lesotho	0	Zambia	3		
Lithuania	1	Zimbabwe	4		

The table reports the list of countries used in the empirical analysis splitting between the 66 Emerging Markets (A and B), and the 4 GIPS. For each country the table report the number of sovereign debt crises included in our crisis dataset over the period 1975-2010, which contains 120 debt crises episodes for emerging markets and 4 episodes for GIPS.

The dataset used in our study comes from Savona and Vezzoli (2013), who collected data from S&P's, World Bank's Global Development Finance (GDF), IMF, Government Finance Statistics database (GFS), and Freedom House, for 66 emerging markets and the 4 European countries that experienced a crisis episode and/or exhibited a large surge in government bond spreads over the period 1975–2010 (see Table 1). In total we used 21 variables reported below in Table 2 including capital, current account, and debt variables, liquidity and macroeconomic variables, also using a proxy for contagion measured as the number of other debt crises occurring in the same year. Sovereign defaults are classified whether at least one of the following three conditions is met:

1. the S&P's default classification;

2. the country had access to a large nonconcessional IMF loan in excess of 100 per cent of quota;
3. the country that had access to the Emergency Financing Mechanism (EFM) used during the global financial crisis of 2008–2010.

All predictors were lagged 1 year, since our aim is to predict a default entry rather than a continuing default. In total we analyzed 120 defaults for emerging markets and 2 cases for Europe (Greece and Ireland in 2010).

Table 2: Candidate Predictors for Sovereign Debt Crisis

Variable	Symbol
Contagion tot	cont
Current account balance (% of GDP)	cab_gdp
Debt service on external debt, long-term to Reserves	dt_ser
Def History	def_h
Exchange Rate residual over linear trend	over
Exports of goods and services (annual % growth)	exp_gr
Exports of goods and services (BoP, current BILLION US$)	exp
External debt stocks (% of exports of goods, services and income)	dt_exp
External debt stocks (% of GDP)	dt_gdp
Foreign direct investment, net inflows (BoP, current US$) (% Change)	fgn
Foreign direct investment, net inflows (BoP, current US$) to GDP (current in US$)	fgn_gdp
Imports of goods and services (annual % growth)	imp
Inflation, consumer prices (annual %)	inf
Money and quasi money (M2) to total reserves ratio	m2
Openess	open
Public Debt to GDP	pub_dt_gdp
Real GDP Growth (%)	gdp
Reserves growth (%)	res_gr
Short-term debt (% of total reserves)	std_res
Short-Term Debt to GDP	std_gdp
US Treasury Bill	ust

The table reports the list of 21 potential predictors used in the empirical analysis. All variables were computed on annual basis and lagged 1 year except for Contagion.

2.1. Probability of Sovereign Default

The first step we run in our empirical analysis was intended to come up with an estimation of the probability of default for each country conditional on more relevant economic leading indicators. To do this we used the CRAGGING (CRoss-validation AGGregatING) algorithm introduced in Vezzoli and Stone (2007), and recently implemented in Vezzoli (2011), Savona and Vezzoli (2013). These papers proved that such a new algorithm is better suited for panel data and other types of structured data, since the common Regression Tree approach assumes that covariates are i.i.d. within each region (node) and independent across regions, when instead autocorrelations and other latent dependencies could play a major role especially in panel data.

To describe how the CRAGGING works, let's denote by:

- (Y, $\mathbf{X}$), with Y dependent variable and $\mathbf{X}$ the covariates, a panel data with N observations for $j = 1, \ldots, j, \ldots, J$ countries in each time t, with $t = 1, \ldots, t, \ldots, T$;
- $\mathcal{L} = \{1, \ldots, j, \ldots, J\}$ the set of units, and by $x_{jt-1} = (x_{1jt-1}, \ldots, x_{rjt-1}, \ldots, x_{Rjt-1})$ the vector of predictors of country j observed at time $t-1$ where $j \in \mathcal{L}$.

The algorithm proceeds by randomly partition $\mathcal{L}$ into $\mathcal{L}_V$ test sets and one of these is taken out of the observations used for estimation and reserved for testing. The corresponding training (estimation) set, we denote by $\mathcal{L}_V^c = \mathcal{L} - \mathcal{L}_V$, is used for estimation by repeatedly removing one country per time and testing the corresponding Regression Tree on the same test set. This type of perturbation, conceived with the end of maintaining the hierarchical structure of the panel data, is repeated for all the $\mathcal{L}_V$ test sets. As a result, for all the countries in the sample we provide multiple predictions which are next averaged thus obtaining the CRAGGING probabilities[2].

2.2. Debt Crisis "Markers"

As in medical studies, where researchers scrutinize gene expressions to detect abnormalities associated to specific diseases (e.g. tumors), similarly in this study we are searching for the most important "disease detectors" among possible economic leading indicators. As for biomarkers in medicine, which are measurable indicators of the severity or presence of some disease state, these economic indicators can be conceived as a sort of "debt crisis markers", since their role is to measure the start and the progress of the "economic disease", so as to help provide early detection of a possible default state. To detect the more influential debt crisis markers we used Random Forrest (RF) algorithm, which are collection of many Regression Trees using different combinations of variables and samples in order to make predictions more stable and less prone to estimation errors. In a nutshell, the RF proceed as follows:

- From the total sample of observations, a random sub-sample (usually around 60%) is selected to grow a Regression Tree by using some randomly selected predictors (in our case the number of randomly selected variables was set at 5).
- The previous point is carried out many times, thereby obtaining a large numbers of bootstrap samples ("out-of-bag" data) to be used to grow Regression Trees based on some randomly selected predictors. In our empirical analysis we set the number of bootstrap samples, and thus the number of Regression Trees, at 3,000 since it was heuristically proven that the accuracy of RF converges around 3,000 trees.
- The accuracy of a RF's prediction, obtained by averaging all realized trees, is estimated from these out-of-bag (OOB) simply computing the mean squared error (MSE), namely:

  where denotes the average prediction for the i*th* observation from all trees for which this observation has been OOB.

[2] See Savona and Vezzoli (2013) for more technical details.

- Using the MSE one can finally assess the importance attributed at each variable based on the MSE reduction (Breiman, 2003) and it is computed for the generic Regression Tree *tree*, by calculating the MSE over all the OOB observations as:

  where are the predictions of the tree, and *i* are its observations over OOB data only, and is the number of OOB observations in the same tree. To assess the importance of regressors **X**, one looks at how each regressor impacts on predictions in terms of MSE reduction: if a specific regressor does not play a significant contribution in predicting *Y*, it should not make a difference if the values for the predictor are randomly permuted in the OOB data before the predictions are generated. Hence, one can compute the MSE reduction by comparing the MSE with and without permuted thereby obtaining the following Variable Importance measure (*VI*):

  .

  Through the *VI* measure it is next possible to rank all predictors from the more influential variable (with the highest MSE reduction value) to the lowest variables (with the lowest MSE reduction values). Finally, to make results on single variable importance metrics more comparable, a relative measure of the *VI* is computed by simply dividing the *VI* of each variable over the highest value, namely

  .

2.3. Parametric-Based Representation

Based on the results obtained through the *RVI* we next tried to map the more important predictors we selected by choosing the variables with , going to mean that we focus only on predictors which show a MSE reduction all above the 50% relative to the *VI* of the best indicator. In other terms, we tried to reconcile and confront the results of RF with those from traditional parametric approaches, namely an OLS model using the covariates selected by the RF with also including their interactions.

This approach is intended to give an answer to the question about the statistical significance of the best leading indicators, and whether they act as leading indicators as a stand-alone predictors or in conjunction with others. Analytically, we run a pooled OLS stepwise regression using as dependent variable the predicted **y** and as covariate the RF showing together with their (first-layer) interactions, namely by considering their multiplication effects. More precisely, starting with *R* potential predictors and selecting indicators based on the 0.5 *RVI* threshold, the number of covariates used in the stepwise OLS regression is . The model can be formalized as follow:

where **y** is a vector of N observations, which represents the standardized random forest previsions, **C** is the standardized matrix of covariate with N rows and columns, is the vector of dimension for the error term and the **B** is the vector, , of the coefficients to be estimated.

Having already the results from Savona and Vezzoli (2013) about the leading indicators for emerging markets and GIPS together, here we only focus on the 4 Euro countries being interested to know whether the reasons underlying sovereign defaults are the same for GIPS and emerging markets or they differ.

3. Results

3.1. Economic Leading Indicators: EMs vs. GIPS

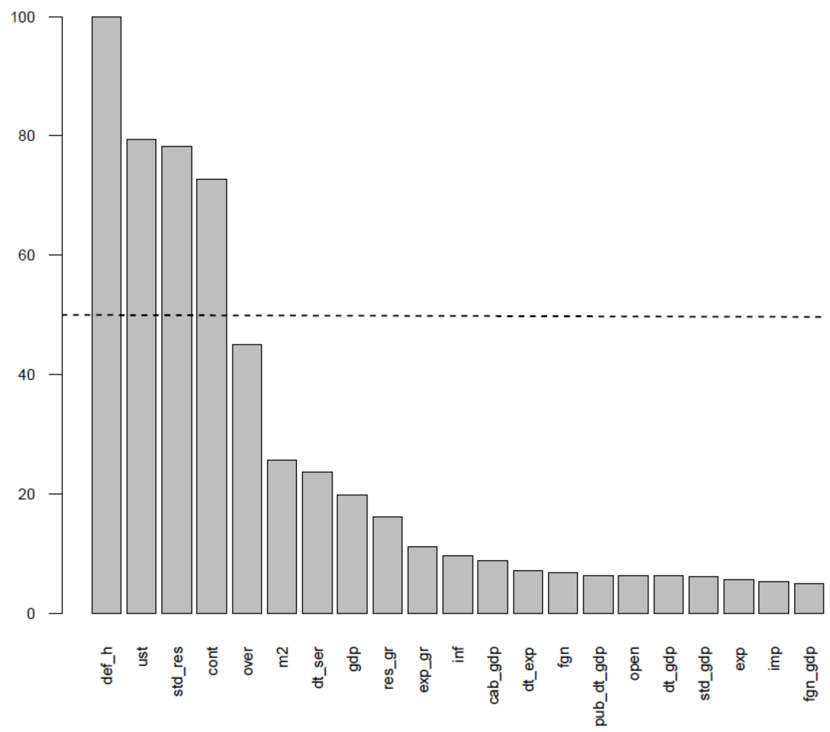
After having computed the CRAGGING algorithm along the lines discussed in the previous section 2.1, we next run the RF using as dependent variable the CRAGGING probabilities and as covariates the battery of the 21 potential predictors reported in Table 2. Next, we computed the *RVI* as described in section 2.3 obtaining a variable ranking based on their contribution in making predictions more accurate (MSE reduction). The results are depicted in Figure 1 and Figure 2 for emerging markets and GIPS, respectively.

The variable importance for emerging markets led to the following more relevant economic indicators (listed in order):

1. default history;
2. US Treasury Bill;
3. short-term debt to reserves;
4. contagion.

These four factors all exhibit an and confirm the findings in Savona and Vezzoli (2013) who proven that short-term debt to reserves and default history are the most significant variables in predicting a debt crisis, together with US interest rates and real GDP growth. As commented in Manasse and Roubini (2009), high interest rates in U.S., which reflects tight monetary conditions, could reduce capital flows to emerging markets, thereby leading to debt servicing difficulties. This is exactly what happened during the crisis of 1980-1983, when debt problems were connected to high interest rates, although the debt levels were not so high. Short-term debt to reserves is instead indicative of illiquidity problems, and it is also related to large reversals of capital flows as proven by Radelet and Sachs (1998). Bad default history is another important predictor as proven by Reinhart and Rogoff (2004), who shown that among debtor countries, serial default on debts tends to recur like clockwork in some countries (serial defaulters). Finally, contagion has clearly a pervasive effect not only in emerging markets, but also in developed countries as it was the case for the 2010-2011 Eurozone debt crisis. In sum, these 4 variables selected by our procedure strongly match previous empirical findings thus confirming the importance of these leading indicators in predicting sovereign default in emerging markets.

Figure 1: Relative Variable Importance – Emerging Markets



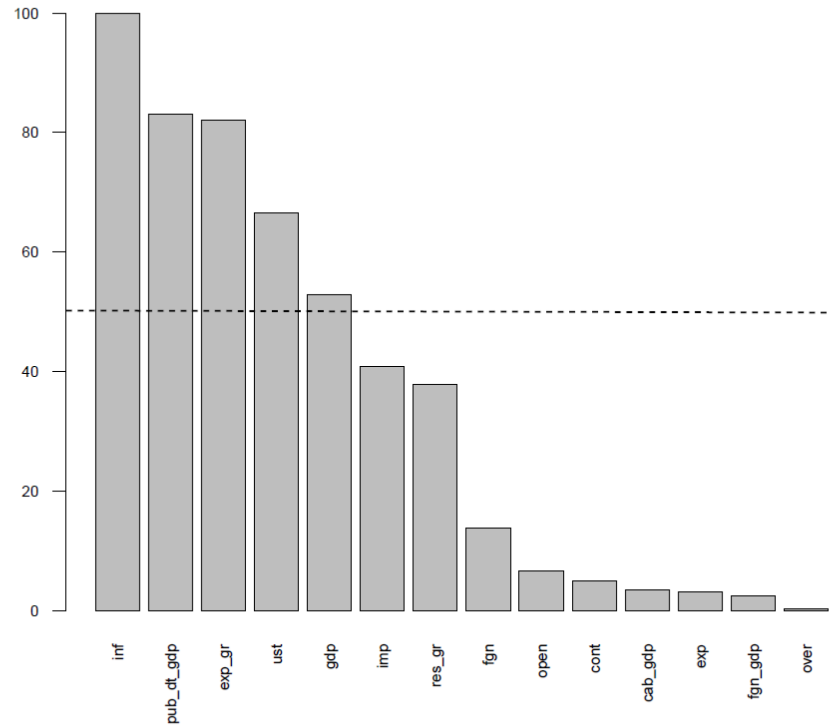
The figure depicts the Relative Variable Importance obtained by running the Random Forest algorithm on the emerging market sub-sample. For each variable (x-axis) the corresponding bin shows the value in percentage form (y-axis) of the *RVI*.

On the other hand, by running the CRAGGING-RF procedures for GIPS we obtained different results with the following more relevant leading indicators (listed in order):

1. inflation;
2. public debt over GDP;
3. export growth;
4. US Treasury Bill rates;
5. real GDP growth.

Only US interest rates appear as the most relevant factor for both emerging markets and GIPS, while different variables were selected as main default drivers. Indeed, inflation plays the leading role followed by "austerity-based" (public debt over GDP) and "growing-based" (export growth and real GDP growth) factors, together with the level of interest rates, which of course led to high indebtedness since they were at a very low levels in particular after the 2000s. The fact that the US Treasury Bill rates have played a common relevant role for both emerging markets and GIPS is interesting not only because they were related to overabundance of cheap credit especially for southern Europe and Ireland, thereby fostering their level of public and private debt, but also because they are associated with the business cycle, since short-term interest rates rise in expansions and fall in recessions. And indeed, with low interest rates real GDP and export growth were both negative for all GIPS in the year before the Greek crisis of 2010.

Figure 2: Relative Variable Importance – GIPS



The figure depicts the Relative Variable Importance obtained by running the Random Forest algorithm over the GIPS sub-sample. For each variable (x-axis) the corresponding bin shows the value in percentage form (y-axis) of the *RVI*.

3.2. Exploring Interactions for GIPS Leading Indicators

Once selected the more relevant "debt crisis markers" for GIPS through a "democratic" data mining technique, with which the data were the only responsible for results no matter about a priory theory and/or a priori beliefs, we also inspected how such indicators are related to the sovereign defaults. By focusing on GIPS only, we complement the empirical analyses in most of the literature on sovereign defaults, which is essentially based on debt crises occurred in emerging markets. This is, in a sense, the novelty of our study, which should be read by contrasting the results coming from the analyses contained herein with the empirical findings in particular obtained by Savona and Vezzoli (2013) and Manasse and Roubini (2009).

As discussed in section 2.3, we did this by running a pooled OLS stepwise regression using the 5 leading indicators and their interactions thus resulting in covariates, we demeaned and standardized to eliminate the intercept, hence obtaining scale-independent coefficient estimates. The results are reported in Table 3. Over 15 covariates, 7 were retained based on their statistical significance, of which only export growth was selected as a significant stand-alone indicator. Moreover, this indicator shows a negative coefficient which is coherent with economic expectation, being the negative variations in exports associated with higher debt crisis probabilities. All other variables were selected in terms of their interaction.

While inflation was selected as the most important indicator according to its *VI* measure, it seems essential the variable should be monitored in conjunction with other factors, namely: (a) public debt over GDP; (b) export growth; (c) real GDP growth. All these interactions were highly statistically significant proving the importance to look at indicator interactions other than single indicators.

Public debt over GDP is of course relevant as leading indicator but any possible forecast regarding impending sovereign debt crisis is possible if scrutinized together with: (a) export growth; (b) real GDP growth. This is in line with Reinhart and Rogoff (2009, 2011) who argued that total external debt is an important crisis indicator.

Finally, Export growth interacted with US Treasury Bill rates has a significant positive impact on sovereign debt crisis probability: negative export variations was indeed experienced by all GIPS in 2009, i.e. 1 year before the Greek and Irish crises, during a downward trend in interest rates thus reflecting on a positive sign of the term (exp_gr*ust).

Table 3: Stepwise OLS - GIPS

Variable	Coefficient	t-Statistic
exp_gr	-2.1942	-9.8165***
(inf*pub_dt_gdp)	-0.3151	-4.9374***
(inf*exp_gr)	1.0783	8.2838***
(inf*gdp)	-0.2155	-3.3856***
(pub_dt_gdp*exp_gr)	1.0355	6.6435***
(pub_dt_gdp*gdp)	-0.3397	-4.0932***
(exp_gr*ust)	0.8145	10.4637***
Adjusted R-squared	0.595	
Sum squared residual	55.8691	
Log likelihood	-136.1574	
Durbin-Watson statistic	1.901	
Akaike info criterion	1.9883	

The table reports results from the stepwise OLS regression on GIPS sub-sample with demeaned and standardized dependent and independent variables. *** denotes significance at 0.001 level.

The implications coming from these results are relevant for both policy and country risk analysts. From the policy side, what the data could tell us is that although public debt level is of course a main risk signal and needs to be contained relative to the GDP, such an austerity rule should be matched together with the growing path of countries especially in terms of GDP trends and exports contributions. Furthermore, inflation is another core factor to be explored in terms of its relationships with public debt, GDP and exports. All that seems suggest that the best policy recipe to prevent sovereign defaults in Europe should include liability- as well as asset-based targets, namely the level of public debt together with export growth jointly with the GDP trends. On the top of this, inflation and interest rates (monetary variables) strongly influence the behavior of such asset-and-liability real variables.

Our findings are important also for investors, country risk analysts and in general who are massively concerned with estimating risk/return profiles of sovereign bonds. Indeed, after the Greek crisis of 2010, all international investors are showing increased sensitivity towards macroeconomic conditions of countries. Therefore, the macro-economic variables we found as significant debt predictors focusing on "physical default probabilities", are expected to play a substantial role in forming sovereign credit spreads as well (De Grauwe and Ji, 2012). Furthermore, having proven that interactions among leading indicators are so significant is a confirmation of the complex and non-linear nature of the sovereign risk. In other terms, when forming estimations on future sovereign risk priced by the market, one should pay extremely attention on how macroeconomic conditions could move together instead of looking at stand-alone leading variables.

4. Conclusions

In this chapter we focused on the Eurozone sovereign debt crisis by inspecting Greece, Ireland, Portugal, Spain (GIPS) fundamentals with the objective to verify whether the reasons underlying a sovereign debt crisis are common between developing and developed countries. Using data from emerging markets and GIPS over the period 1975-2010 we selected the more important "debt crisis markers", by approaching to the sovereign defaults as if they were a human disease such as cancer. And as in medical studies, we used data mining techniques, which are extremely useful to detect complex and non-linear relationship within large datasets, with the end to select the more important crisis predictors for 66 emerging markets, on one hand, and 4 European countries (GIPS), on the other. We found different economic leading indicators under the sovereign default process for GIPS, compared to the leading indicators exhibited by emerging markets. Therefore we are in line with the view of Reinhart and Rogoff (2013), who consider developed countries economies as "different animals from their emerging market counterparts", while our main finding relates to the risk signals that can help predict an impending crisis. But of course, the risk signals represent the key variables of policy measures toolkit that countries should implement to mitigate the risk impacts and restore economies that suffered a sovereign debt crisis.

Inflation seems to be the best (statistical) indicator but needs to be decoded together with public debt over GDP, export growth, real GDP growth. The policy side of our results suggest to strongly monitor liability- (public debt over GDP) and asset-based (export growth and real GDP variations) country fundamentals. Export growth, in particular, seems to play a key role since it appears to be statistically significant as a stand-alone leading indicator and interacted with other selected predictors.

References

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, 5–32.

Breiman, L. (2003), "Manual on Setting Up, Using, and Understanding Random Forests V4.0", unpublished manuscript, University of California, Berkeley, Department of Statistics.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984), *Classification and Regression Trees*, Belmont, CA: Wadsworth.

Bulow, J. and K. Rogoff (1989) "Sovereign Debt: Is to Forgive to Forget?" *American Economic Review* 79: 43-50.

De Grauwe, P. and Ji, Y. (2012), "Mispricing of Sovereign Risk and Multiple Equilibria in the Eurozone", CEPS Working Document, 361.

Eaton, J. and M. Gersovitz (1981). "Debt with potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* (48): 289-309.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning: Data mining, Inference, and Prediction*, Springer, Berlin.

Lin, F., and S. McClean (2001), "A Data Mining Approach to the Prediction of Corporate Failure", *Knowledge-Based Systems*, 14(3-4), 189–195.

Manasse, P. and N. Roubini (2009), "Rules of Thumb for Sovereign Debt Crises", *Journal of International Economics* 78 (2), 192–205.

Radelet, S. and J. Sachs (1998), "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects" Brooking Papers on Economic Activity 1, 1–90.

Reinhart C., and K. Rogoff (2009), "Serial Default and the "Paradox" of Rich to Poor Capital Flows", NBER Working Paper No. 10296.

Reinhart C., and K. Rogoff (2009), *This Time Is Different: Eight Centuries of Financial Folly*, Princeton, New Jersey: Princeton University Press.

Reinhart C., and K. Rogoff (2011), "From Financial Crash to Debt Crisis," *American Economic Review*, 101(5), 1676–1706.

Reinhart C., and K. Rogoff (2013), Financial and Sovereign Debt Crises: Some Lessons Learned and Those Forgotten, IMF Working Paper WP /13/266.

Savona, R. and M. Vezzoli (2015), "Fitting and Forecasting Sovereign Defaults Using Multiple Risk Signals", *Oxford Bulletin of Economics and Statistics* 77(1), 66–92.

Vezzoli, M. (2011), "Exploring the facets of overall job satisfaction through a novel ensemble learning", *Electronic Journal Of Applied Statistical Analysis,* 4(1), 23–38.

Vezzoli, M. and C. Stone (2007). "CRAGGING", manuscript, Department of Statistics, University of California, Berkeley also published in Book of Short Papers CLADAG 2007, University of Macerata, September 12-14, 2007.