Predicting sovereign debt crises

Dawood, Mary; Horsewood, Nicholas; Strobel, Frank

DOI:
10.1016/j.jfs.2016.11.008

Predicting Sovereign Debt Crises: An Early Warning System Approach

Mary Dawood

Alexandria University, Egypt

Nicholas Horsewood and Frank Strobel[1]

University of Birmingham, UK

November 18, 2016

[1]Corresponding author; email: f.strobel@bham.ac.uk

Abstract

In light of the renewed challenge to construct effective "Early Warning Systems" for sovereign debt crises, we empirically evaluate the predictive power of econometric models developed so far across developed and emerging country regions. We propose a different specification of the crisis variable that allows for the prediction of new crisis onsets as well as duration, and develop a more powerful dynamic-recursive forecasting technique to generate more accurate out-of-sample warning signals of sovereign debt crises. Our results are shown to be more accurate compared to the ones found in the existing literature.
Keywords: sovereign debt crisis; early warning system; logit; dynamic signal extraction; dynamic-recursive forecasting
JEL Classification: F34, F37, C53

1 Introduction

In the aftermath of the 2008 global financial crisis, which hit the major advanced economies and affected many emerging and developing countries, governments were forced to bail out and recapitalize their failing banking systems. Such interventions resulted in large fiscal deficits at the same time as their economies slowed after the burst of the property bubble. As a consequence, several European nations, in particular Greece, Portugal, Ireland and Spain, faced a prolonged debt crisis, unable to repay or refinance their sovereign debt and having to rely on the assistance of other Eurozone countries, the IMF and the ECB. Considering the economic and social effects of sovereign debt crises at both national and international levels, it has become increasingly important to construct financial monitoring tools that can forewarn the build-up of such financial turmoil. The main purpose of such systems is to provide policymakers with some lead time to take corrective actions that would help avert, or at least mitigate, the damage associated with an approaching crisis.

Since the late 1990s, several studies have attempted to develop a framework for such Early Warning Systems (EWS) using various econometric models.[1] However, the forecasting performance of these EWS was not generally satisfactory, especially in predicting out-of-sample crisis incidents (Berg *et al.*, 2005). The challenge of designing an effective EWS escalated even further when the pre-2008 models failed to foresee the severity and international span of this recent global crisis (Candelon *et al.*, 2014). As a result, several modified econometric methods have recently been introduced in the literature, which appear to outperform the traditional techniques in forecasting a specific type of financial crisis, or crises in a specific type of economy. However, no study has attempted to cross-evaluate the performance of these recent methods in forewarning sovereign debt crises in different regions.

The present study attempts to contribute to the literature in several ways. First, given the distinct nature of national economies, their vulnerability to shocks and the effectiveness of their institutions and policy responses, the causes and associated leading indicators of sovereign debt crises can reasonably be expected to differ across countries. Yet, until recently, the focus of modeling EWS for sovereign defaults was on developing countries

[1] See e.g. Frankel and Rose (1996); Kaminsky *et al.* (1998); Demirguc-Kunt and Detragiache (1998); Peter (2002); Manasse *et al.* (2003).

only[2], usually pooled into a single group. Our study, on the other hand, investigates the possibility of signaling indicator differences between developed and developing countries, and between different regions; our results support the notion of regional heterogeneity of forewarning indicators. Next, we evaluate and contrast the predictive performance of two recently developed econometric methods, namely the multinomial logit regression and the dynamic signal extraction approach vis-a-vis our own, novel specification of the binary logit model, in which the crisis variable accounts for all periods in which a country suffered a debt crisis as individual crisis episodes. In addition, we develop and apply a new dynamic-recursive forecasting technique to generate more accurate out-of-sample warning signals. We find that our binary logit specification significantly outperforms that of the multinomial logit and the traditional binary logit models prevalent in the literature, and to some extent also that of the dynamic signal extraction model.

The remainder of the paper is then structured as follows: section 2 surveys the findings of the previous literature, while section 3 summarizes the data and performs a preliminary quantitative analysis of the potential EWS indicators. The econometric methods and their results are then outlined in section 4, the warning indicators and the results of the "horse-race" are presented in section 5, while section 6 concludes the paper.

2 Previous Literature

Empirical studies that focus on constructing EWS for financial crises have mostly relied on one of two main approaches. Kaminsky *et al.* (1998) developed the (static) signal extraction approach, a non-parametric method that entails identifying and monitoring certain variables that tend to behave in an unusual manner in the build-up to financial or economic distress. This model is designed so as to signal an impending crisis if these indicators exceed a certain threshold value, calculated as a specific percentile of each indicator's sample distribution. More recently, Casu *et al.* (2012) proposed a dynamic (non-sample-specific) choice of the threshold that focuses more on the volatility of the indicators. For this, they specified the threshold as a certain number of standard deviations away from the variable's long-run mean. Whereas the static approach was developed in the context of currency

[2] This is mainly due to the fact that there were previously no major concerns about governments in developed countries not being able to meet their obligations to an extent that would progress into a serious debt crisis.

crises, and the dynamic one for the detection of banking distress, neither specification has been used for the modeling of an EWS for sovereign defaults, with the exception of Savona and Vezzoli (2015).

Frankel and Rose (1996) alternatively proposed the utilization of logit or probit regression models to estimate the probability of an approaching currency crisis. Manasse *et al.* (2003) and Fuertes and Kalotychou (2006) analogously applied pooled logit models to examine debt crises in emerging economies. Manasse *et al.* (2003) argued that logit models tend to perform better than probit ones when the dependent variable is not evenly distributed between the two outcomes, i.e. crisis and no crisis; this is usually the case as crisis events are not too common. More recently, Jedidi (2013) attempted to predict sovereign debt crises using a fixed-effects logit model while including a number of developed countries, whereas Pescatori and Sy (2007) and Lausev *et al.* (2011) applied a random-effects model instead.

It is important to note that EWS that are based on binary dependent variable models, where the crisis variable assumes the value of one for the periods a country is hit by a crisis and zero otherwise, have an inherent endogeneity problem. This is due to the fact that the behavior of the indicator variables is affected both by the crisis itself and the policies undertaken to mitigate it. Furthermore, the signaling indicators can be reasonably expected to behave differently during tranquil times as compared to post-crisis periods, where the economy is undergoing an adjustment process to recover from a crisis. Hence, combining observations of tranquil periods with those of post-crisis ones into a single (zero) group can lead to a form of bias; Bussiere and Fratzscher (2006) referred to this as “post-crisis bias”. To avoid this pitfall, several authors (e.g. Fuertes and Kalotychou, 2007; Savona and Vezzoli, 2015) dropped the post-crisis observations from their sample, however thereby suffering from loss of information, while others (e.g. Peter, 2002; Manasse *et al.*, 2003) used a dummy variable to allow for different coefficients in the post-crisis periods. Bussiere and Fratzscher (2006), on the other hand, suggested the use of a multinomial crisis variable instead that reflects all three states of the economy. Ciarlone and Trebeschi (2005), employing an earlier (2002) version of Bussiere and Fratzscher, investigated its performance in predicting debt crisis episodes in the case of emerging economies.

Several other less common methods were proposed in the literature. Fuertes and Kalotychou (2007) used the K-means clustering approach, which entails assigning every observation to the cluster with the nearest mean vector so as to maximize within-cluster similarity and between-cluster discrepancy. However, their results showed that the binary logit regression outperforms this approach in the out-of-sample period. Moreover, Manasse *et al.* (2003) and Manasse and Roubini (2009) used regression tree analysis, while Fioramanti (2008) applied artificial neural network models to predict sovereign defaults. However, the author noted that despite its better ability to predict crises than probit regressions, neural network models do not give any marginal effects interpretation of the individual signaling indicators, and thus are less useful as a policy tool.

There appears to be a widespread consensus in previous studies regarding significant indicators that could act as explanatory variables for debt crises. In particular, several ability-to-pay indicators are emphasized, such as the external debt ratio, growth in foreign exchange reserves and export earnings, reflecting the ability to service debt. In addition, often highlighted is the importance of current account deficits as a measure of illiquidity risk, and other macroeconomic indicators that affect a country's capacity to meet its obligations. Further indicators, such as trade openness and measures of macroeconomic stability, were also suggested by the willingness-to-pay approach, pioneered by Eaton and Gersovitz (1981); here defaults are modeled as an event where a sovereign chooses to repudiate its debt if the perceived costs of defaulting are less than the benefits. Additionally, the survey of Reinhart (2002), covering about 60 countries over the period 1979-1999, conveyed that 84% of the sampled debt crises were preceded by a currency crisis. Hence, variables that are well-suited for predicting currency crisis could also be expected to have some explanatory power in EWS for sovereign defaults. Chakrabarti and Zeaiter (2014) carried out a recent comprehensive review[3] regarding these issues, summarizing the empirically significant factors and their observed effect on the probability of sovereign default.

3 Data and Preliminary Analysis

Our panel consists of 38 advanced and emerging economies during the period 1980-2012. We rely on an annual frequency of the data, as sovereign debt crises tend to last for prolonged periods and show persistence (Manasse *et al.*, 2003). For the construction of the EWS and the estimation of our models we only use the sub-sample 1980-2005, whereas the seven-year period from 2006 to 2012 is used to evaluate the out-of-sample forecasts. This is a challenging exercise given the limited occurrence of previous sovereign debt problems

[3] See Table 1 in Chakrabarti and Zeaiter (2014).

in advanced countries, which makes the training of the EWS rather difficult. The selection of countries is guided mainly by data availability; it covers four main regions[4]: Africa and the Middle East, South and East Asia, Latin America and Western Europe. The list of countries considered in each region, along with details on each crisis incident, is outlined in Appendix A.

3.1 Sovereign Defaults and their Indicators

To capture both actual and potential defaults on sovereign debt, Manasse *et al.* (2003) defined a country to be in crisis either if it is rated by Standard & Poor's as being in default (i.e. is failing to meet its external obligations) or if it receives a loan from the IMF in excess of 100% of its quota as an extensive rescue package. The same definition was later applied by Fioramanti (2008), Manasse and Roubini (2009), Savona and Vezzoli (2015) and Jedidi (2013). Ciarlone and Trebeschi (2005) included other events as well: in addition to the ones mentioned above, they also consider a country to be in crisis if the amount of overdue interest or principal payments is more than 5% of its outstanding external debt, or if it engaged in any restructuring or rescheduling schemes.

To make our results comparable to the ones found in the previous literature, we employ the same crisis definitions. Hence, in the case of emerging economies, the dependent variable (DC_{it}) assumes unity if any of the four following events occurs, and is zero otherwise: (1) accumulated interest and/or principal arrears exceed 5% of the outstanding debt; (2) receiving a loan from the IMF in excess of 100% of the country quota; (3) cumulative credit obtained from the IMF increases above 200% of the quota; (4) engaging in a debt restructuring (buybacks or reductions) or rescheduling scheme that involves more than 20% of the outstanding debt. With respect to developed countries, we use a slightly different rule[5] due to the lack of reported details on the arrears and the amounts involved in restructuring and rescheduling programs. Therefore, for developed countries, in addition

[4]We do not include countries from Eastern and Central Europe, as their data are only available from 1995 onwards. This leaves only ten observations per country to make out-of-sample forecasts of seven years, which is clearly not enough of a training period for the EWS, especially given that those countries experienced a very limited number of debt crises during this period (see to Table 1 in Manasse and Roubini, 2009).

[5] As a robustness check, we use this alternative rule to define debt crises in emerging economies as well. A simple correlation test reveals that the dependent variable using this rule is 81.6% linearly correlated with the debt crisis variable using the definition prevalent in the literature. We find that estimation results are qualitatively the same for all three emerging regions, except for some changes in the significance of the

to the two events involving loans from the IMF, the crisis index is also set to one if the outstanding government debt exceeds 150% of the nominal value of GDP[6].

Regarding the indicator variables that could be used to provide warning signals of a forthcoming crisis, Table 1 illustrates that these can be grouped into four main categories. The first group reflects the exposure of a country to sovereign debt problems. A higher stock of external debt and/or IMF credit compared to the country's GDP increases the chances of unsustainable debt. Moreover, to measure the burden of servicing external debt, we consider the GDP-weighted average of the bank lending interest rates in seven major developed countries. The health of the country's external sector is captured in the second group, where the erosion of foreign exchange reserves is expected to raise the likelihood of sovereign default. On the other hand, a stronger current account balance, growth of export revenues, and net inflows of FDI reduce the country's financial need for acquiring foreign debt. A less clear impact on the probability of default is that of the change in trade openness: a low degree can have an adverse effect on trade surpluses and make the country more willing to repudiate its debt, whereas increased free trade can make the economy more vulnerable to external shocks.

With respect to the third group, domestic macroeconomic variables can reasonably be expected to show some deterioration prior to a debt crisis. Specifically, lower growth of real GDP and reduced national savings can reduce the country's ability to meet its obligations. Furthermore, a rise in the rate of inflation and the ratio of M2 to foreign exchange reserves reduce external competitiveness and reflect the extent of unbacked implicit government liabilities. This may lead to a confidence crisis, as lenders suspect that the government is attempting to inflate away the value of its external debt. We also consider the overvaluation of the real exchange rate[7] to capture the effect of an approaching currency crisis on the ability to meet external obligations. While larger government expenditures can increase the likelihood of a debt crisis, governments usually undergo austerity measures during times of crisis. Thus, higher public spending can also be associated with tranquil periods, where the likelihood of a debt crisis is minimal. Finally, we include three variables to investigate the

---

variables; however, the forecast performance of the models is poorer. Detailed results are available from the authors upon request.

[6]This particular ratio is chosen following IMF estimates that the median maximum sustainable debt level ranges between 100-190% of GDP; see IMF (2011) for further details.

[7]This variable is measured as the negative deviation of REER (measured in domestic currency) from its long-run trend.

possibility of spillover from the banking sector. Whereas growing bank assets and a higher ratio of domestic credit can reflect the development of the banking industry, the latter can also increase the vulnerability of the banking sector to macroeconomic shocks. In addition, we also consider net bank claims (loans minus deposits) on the central government.

The last column in Table 1 summarizes the previous studies that considered each potential variable in their analysis. The data on the indicators are collected from four main databases: IMF International Financial Statistics, World Bank Development Indicators, World Economic Outlook, and the World Bank Global Financial Database.

3.2 Quantitative Analysis

We conduct a preliminary analysis to investigate how the candidate variables tend to behave around default episodes compared to tranquil periods, and thus whether they are expected to perform well as signaling indicators. Accordingly, Table 2 depicts the respective mean of each variable in the global sample during normal vs. crisis years, along with a *t*-test of the population means being the same for the two periods. The results are presented over the global sample and in each country group separately[8].

It is evident from this table that there is a tangible difference across the regions with respect to the candidate EWS indicators. Specifically, the external sector variables seem to behave significantly different around crises in Asia and Latin America, while the domestic macroeconomic conditions seem to play the major role in Africa. In developed countries only the debt exposure variables appear to be potentially important signaling indicators of debt crises. Nonetheless, a small set of variables appears as good crisis indicators in most regions. Primarily, a rise in the global lending rate significantly increases the cost of servicing external debt and magnifies the likelihood of sovereign defaults in general. Likewise, the erosion of foreign exchange reserves can act as a potentially good indicator of debt problems. In the case of emerging economies, two additional variables seem to play

[8] We also conduct a *t*-test of the population variances of the indicator variables to check whether their variability is significantly different in normal vs. in crisis periods. The results of the test are reported in Appendix B. It can be noted that most variables that had significant mean changes across both periods have also experienced significant changes in their variability, except for the global interest rate, the current account, and national savings. However, as the number of observations in the normal and crisis periods are different (crisis periods account for only 23% of the sample), comparisons of the variance of the indicators need to be treated with caution as they do not necessarily allow for comparing like with like.

Table 1: Signaling Indicators of Sovereign Debt Crises

Symptoms	Indicators	Measurement	Exp.Sign	Literature
Debt Exposure	Total Debt	gross external debt as % of GDP	+	Ciarlone and Trebeschi (2005); Fuertes and Kalotychou (2007); Lausev *et al.* (2011)
	IMF Credit	loans from IMF as % of GDP	+	Fuertes and Kalotychou (2006)
	Global Interest	global lending interest rate[1]	+	Ciarlone and Trebeschi (2005)
External Sector	ForeignExch Reserves	as % of GDP	-	Jedidi (2013)
	Trade Openness	ratio of exports plus imports to GDP	+ / -	Fuertes and Kalotychou (2007, 2006); Jedidi (2013)
	Export Growth	annual exports growth rate	-	Ciarlone and Trebeschi (2005); Lausev *et al.* (2011)
	Current Account	current account balance as % of GDP	-	Fioramanti (2008); Manasse and Roubini (2009)
	FDI	net FDI inflows as % of GDP	-	Peter (2002)
Domestic Macroecon. Conditions	Real GDP Growth	annual growth of real GDP	-	Fioramanti (2008); Savona and Vezzoli (2015)
	REER Overval	deviation of real effective exchange rate from 5-year rolling mean	-	Manasse and Roubini (2009); Pescatori and Sy (2007)
	Inflation	rate of change in CPI	+	Manasse *et al.* (2003); Manasse and Roubini (2009)
	M2/Reserves	ratio of M2 to foreign exch. reserves[2]	+	Lestano *et al.* (2003)
	National Saving	ratio to GDP	-	Lestano *et al.* (2003); Jedidi (2013)
	Gov Expenditures	as % of GDP	+ / -	Lausev *et al.* (2011); Jedidi (2013)
Banking Sector	Domestic Credit	ratio of domestic credit to GDP	+ / -	Fuertes and Kalotychou (2006, 2007)
	Bank Assets	ratio of bank assets to GDP	-	Lestano *et al.* (2003)
	Gov Bank Loans	net bank claims on central gov.	+	Manasse and Roubini (2009)

Notes: (1) GDP-weighted bank lending interest rates in seven developed countries: USA, Canada, UK, Germany, France, Italy, and Sweden. (2) In Euro-zone countries, M2 represents the national contribution to the Euro area, while the foreign exchange reserves are those held by the national central banks and monetary authorities.

an important role in the possibility of crises, namely foreign capital flows and national savings.

To act as an effective forewarning indicator of sovereign defaults, it is not sufficient for a variable to act differently during times of distress, but also before trouble starts building up. In order to highlight the candidate indicators that can signal an approaching crisis, Figure 1 illustrates how the mean of each variable changes on average from normal periods to pre-crisis years, during crisis episodes, and after the crisis hits the economy. According to this graph, factors like foreign exchange reserves, M2, export growth, current account balance, trade openness, and national savings have a distinctive behavior during pre-crisis periods compared to after the crisis hits the economy. Other factors, like total external debt, global interest rate, FDI, real exchange rate, and bank claims show a sharp change in behavior before the crisis hits the economy and only change slightly afterwards, but in the same direction. Unlike the two previous types of factors, which are expected to prove significant in predicting debt crises, there is another group of factors that only changes behavior markedly after the onset of the crisis. These are mainly: inflation, IMF credit, government expenditures, domestic credit, and bank assets. This group of variables is not expected to perform well as EWS signaling indicators (although IMF credit does increase well before crisis onsets, but only slightly compared to afterwards).

4 Methodology of EWS

In order to formally test the significance of the variables and their performance as signaling indicators of sovereign debt crises in the different regions, we apply three recently developed econometric techniques, namely the dynamic signal extraction approach and binary and multinomial logit regressions. This section outlines these methods and their results.

4.1 Dynamic Signal Extraction Approach

This approach entails setting critical threshold levels for the crisis indicators, such that if any variable crosses its specified threshold (above or below, depending on whether the variable increases or decreases the probability of debt crises), it is said to signal an imminent crisis over a given time period h, called the *crisis window*. For each variable over seven different multiples of standard deviations (ranging from 0.5 to 3) and for four types of

Table 2: Quantitative Analysis of Debt Crisis Indicator Means

	Full Model			Regional Models			
Indicator	$\overline{x}_{NoCrisis}$	$\overline{x}_{Crisis}$	t-stat	Dev.	Asia	Latin	Africa
Debt Exposure							
Total Debt	69.1	96.7	-1.3	×	×	✓	×
IMF Credit	0.3	1.7	-6.1*	✓	×	✓	×
Global Interest	8.4	11.0	-5.1*	✓	✓	✓	✓
External Sector							
ForeignExch Reserves	16.4	6.2	9.3*	✓	✓	✓	✓
Export Growth	5.9	3.0	1.8	×	×	×	×
Current Account	-0.7	-4.1	4.9*	×	✓	✓	×
Trade Openness	81.8	63.5	3.7*	×	✓	✓	×
FDI	3.3	1.6	5.9*	×	✓	✓	✓
Macroeconomic Condition							
Real GDP Growth	4.1	2.0	2.4*	✓	×	✓	×
Inflation	6.8	28.9	-3.9*	×	×	✓	✓
M2/Reserves	10.0	15.0	-1.2	×	×	✓	✓
REER Overval	2.2	-10.5	5.4*	×	×	✓	✓
Gov Expenditures	15.7	15.0	0.7	×	×	×	×
National Saving	23.7	17.2	6.1*	×	✓	✓	✓
Banking Sector							
Domestic Credit	69.6	60.3	1.2	×	×	✓	×
Bank Assets	71.0	56.0	2.5*	✓	×	✓	×
Gov Bank Loans	12.7	16.6	-1.4	×	×	✓	×

Notes: The first two columns depict the full-model variable means during tranquil and crisis periods, respectively. Column three shows the Welch adaptation of the t-test of mean differences to account for unequal variances and sample sizes of the two economic states. The result of the mean-difference t-test in each region separately is illustrated in the last four columns. Both * and ✓denote statistical significance of the mean differences at the 95% level, while × denotes no significant difference in the means.

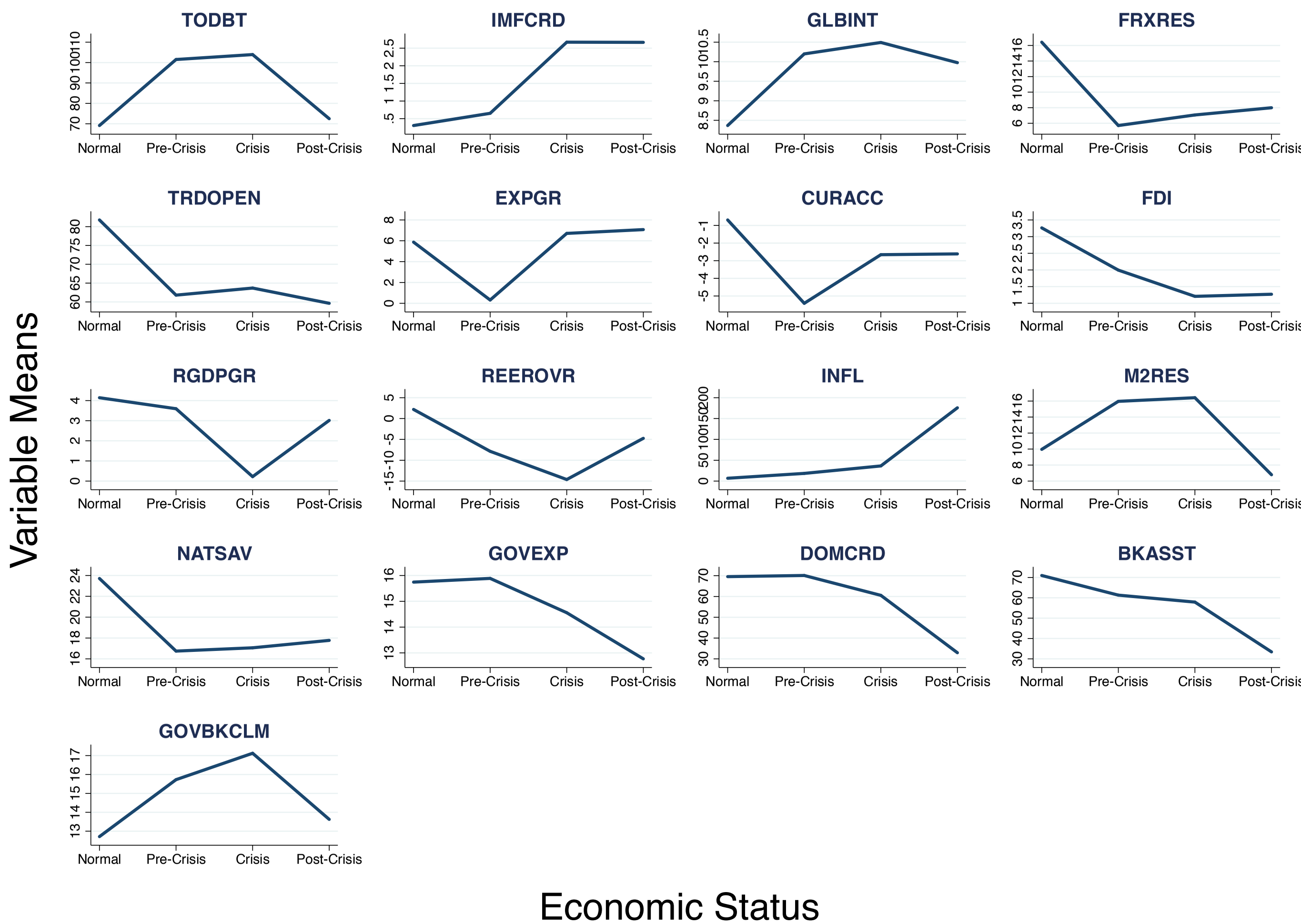

Figure 1: Behavior of Candidate Variables around Debt Crisis Episodes

long-run means (fixed and rolling), a grid search is applied in order to identify the optimal threshold that simultaneously minimizes the noise-to-signal ratio ($NTSR$) and maximizes Youden's J-statistic.[9]

We define the forward-looking response variable DCs_{it} to capture the incidents of approaching debt crises in country i over the period t within a specific crisis window h:

$$DCs_{it} = \begin{cases} 1 & if \ \exists \ k = 1, \ldots, h \ s.t. \ DC_{i,t+k} = 1 \\ 0 & otherwise \end{cases} \tag{1}$$

where DC_{it} refers to the binary crisis index (defined below). Three different crisis windows h are included in the grid search (one, two and four years)[10]. The results indicate that the two-year specification is preferable, as it improves the $NTSR$ of the majority of indicators compared to the four-year specification, and does not cause significant losses compared to the one-year specification. Therefore, we set $h = 2$. After identifying the optimal thresholds, the indicators can be evaluated and ranked according to three different criteria: (a) the percentage of crises correctly forewarned, (b) the optimal $NTSR$, and (c) the average lead time of the signals (i.e. the average number of periods in advance of the crisis when the first signal occurs).

The results of the grid search are reported in Table 3 for the global sample and for each region. It can be noticed from this table that the majority (about 75%) of the variables have a $NTSR \leq 0.5$ in Latin America and in Africa and the Middle East, whereas in the developed countries and in South and East Asia only few indicators can provide reliable signals of approaching debt crises. The most prominent indicators are, generally, IMF credit, global interest rate, foreign exchange reserves, current account balance and domestic credit, which have the lowest $NTSR$ ratios.

The overall lead time of the signals (column 3) is not very long, though. Only four variables tend to issue their first signals two years in advance, namely IMF credit, global interest rate, the current account and foreign exchange reserves. The signals of four other variables have a lead time of more than 18 months; these are export growth, FDI, ratio of

[9] The $NTSR$ is the ratio of false alarms (noise) to the correct signals issued by the model, whereas the J-statistic is calculated as the difference between the correct signals and the false alarms. See subsection 5.2 for full details of the calculation of these measures.

[10] These windows are chosen in line with the bulk of the literature (see e.g. Peter, 2002; Ciarlone and Trebeschi, 2005; Fuertes and Kalotychou, 2006; Gourinchas and Obstfeld, 2012). One year is the shortest possible forecast window, since the data are annual; the window is then increased from there to ideally allow policymakers more time to take pre-emptive measures.

Table 3: Results of Grid Search on Individual Indicators

	Global			Developed		SE-Asia		Latin America		Africa	
	NTSR	Onsets	Lead	NTSR	Onsets	NTSR	Onsets	NTSR	Onsets	NTSR	Onsets
Total Debt	1.20	28.6	1.1	0.18	100.0	0.38	50.0	0.64	22.2	0.38	33.3
IMF Credit	0.21	54.3	1.8	0.05	50.0	0.26	50.0	0.24	72.2	0.32	22.2
Global Interest	0.33	48.6	1.8	0.62	50.0	0.18	66.7	0.19	50.0	0.36	66.7
ForeignExch Reserves	0.36	54.3	1.7	0.94	50.0	0.78	16.7	0.25	55.6	0.10	77.8
Trade Openness	0.41	31.4	1.3	0.64	50.0	0.53	66.7	0.33	33.3	0.17	55.6
Export Growth	0.67	34.3	1.4	1.22	50.0	0.54	50.0	0.64	38.9	0.34	55.6
Current Account	0.33	48.6	2.0	0.71	50.0	0.22	83.3	0.20	55.6	0.30	77.8
FDI	0.48	31.4	1.5	0.63	100.0	0.67	50.0	0.31	38.9	0.59	22.2
Real GDP Growth	0.75	28.6	1.3	0.58	50.0	0.48	33.3	0.54	27.8	0.37	33.3
Inflation	0.68	25.7	1.1	0.35	50.0	0.30	50.0	0.50	22.2	0.46	33.3
M2/Reserves	0.42	42.9	1.6	0.50	50.0	0.64	16.7	0.38	50.0	0.20	55.6
REER Overval	0.89	22.9	0.6	1.08	50.0	0.74	16.7	0.50	27.8	0.16	44.4
Gov Expenditures	0.61	28.6	1.3	0.15	100.0	0.60	50.0	0.32	44.4	0.57	22.2
National Saving	0.42	34.3	1.4	0.23	100.0	0.57	33.3	0.38	44.4	0.25	66.7
Domestic Credit	0.33	25.7	0.8	0.51	50.0	0.38	33.3	0.77	38.9	0.58	55.6
Bank Assets	1.29	17.1	0.9	0.77	50.0	0.74	33.3	0.59	27.8	0.71	22.2
Gov Bank Loans	0.55	25.7	0.9	0.31	50.0	2.25	16.7	0.32	33.3	0.14	66.7

Source: Authors' calculations.

Notes: The table reports the results of the grid search of the dynamic signal extraction approach in the full sample and in each region separately. The columns labeled "NTSR" depict for each variable the optimal $NTSR$ (*i.e.* the lowest $NTSR$ that maximizes J-statistic), while the columns labeled "Onsets" show the percentage of crisis onsets correctly forewarned. The "Lead" column reports the average lead time of the signals (*i.e.* how early the first warning signal is usually issued).

M2 to reserves and national savings. The rest of the indicators considered start signaling an approaching debt crises only one year in advance. The shortest average lead time is that of the warnings issued by overvaluation of the domestic currency; this is not surprising given the high degree of volatility in exchange rates.

Taking a closer look at the separate regions, and consistent with the primary t-tests conducted in the previous section, the debt exposure variables are the major signaling indicators in developed countries, having the lowest $NTSR$ ratios. They are also important forewarning indicators in South-East Asia, while neither the domestic macroeconomic variables nor the banking sector seem to act as significant indicators. On the other hand, the external sector appears to provide more accurate warning signals of debt crises in Latin America and Africa. Thus, it can be noted that, save for the debt exposure variables that appear to issue good warning signals in all regions, there is a distinct set of indicators that performs best in each region, which supports the notion of regional heterogeneity of the signaling variables.

4.2 Binary Logit Model

In order to be able to construct an EWS that can predict the likelihood of an upcoming crisis as well as its duration, and to avoid post-crisis bias without having to drop potentially valuable observations from the sample, we include all periods in which a country suffered a debt crisis as individual crisis episodes. Thus, the binary dependent variable DC_{it} is set to one for all crisis periods (as outlined in Appendix A) and is zero only during tranquil times. We also consider a multinomial specification of the crisis variable, which accounts for all three economic states (normal, crisis, post-crisis), as discussed in subsection 4.3.

The logit model estimates the probability of a crisis using the logistic distribution function

$$Pr\left(Y_{it}=1\right)=F\left(X_{it-h}\beta\right)=\frac{e^{X_{it-h}\beta}}{1+e^{X_{it-h}\beta}} \tag{2}$$

where $F(\cdot)$ is the cumulative logistic distribution, X_{it-h} is the vector of h-period lagged explanatory variables[11], Y_{it} denotes the binary crisis variable DC_{it}, and β is the vector of coefficients.

[11]Henceforth, the lags will be suppressed for simplicity, but are implied in all following equations.

Maximum likelihood estimation is then undertaken to obtain the parameters, where the log-likelihood function is written as:

$$\log \mathcal{L} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[Y_{it} \ln F(X_{it}\beta) + (1 - Y_{it}) \ln \left(1 - F\left(X_{it}\beta\right)\right) \right] \tag{3}$$

We report the marginal effects, rather than the raw beta coefficients of the logged odds ratio, in all results tables for simplicity of interpretation. Furthermore, the Huber-White robust variance estimator of the covariates is calculated and reported to account for country-specific variances in all regression models (see Manasse *et al.*, 2003, p. 19).

We examine the fit of five models: the global model that incorporates all countries together, both developed and emerging, and four separate regional models. The two-year-lagged marginal effect of each indicator on the probability of a debt crisis is displayed in the upper panel of Table 4. In addition to these pooled regressions, Table 5 estimates these five models using fixed- and random-effects panel regressions to account for possible country-specific heterogeneity. The lower panel of the tables reports the corresponding McFadden's pseudo R^2, the log-likelihood ratio and the BIC criteria of each model, along with the in-sample percentage of correct crisis signals.

While subsection 5.1 discusses the general significant warning indicators of debt crises using the three econometric methods, a quick glance at the results of the binary logit tables shows that credit from the IMF, foreign exchange reserves, public spending, and domestic credit have the major effects on the probability of an approaching debt crisis in South and East Asia. The debt exposure and macroeconomic variables seem to be playing the important role in Latin America, while in Africa and the Middle East total debt, foreign exchange reserves, and national savings have the lead in anticipating sovereign default problems. Finally, total debt, IMF credit, and the balance of the current account are the main contributors to the probability of a debt crisis in West European countries.

In line with the results of Fuertes and Kalotychou (2006), who aimed to identify the most accurate parametrization of a logit regression model, we also find that the fixed-effects models that control for unobserved heterogeneity across countries describe the data better than the pooled and the random-effects models. However, when it comes to forecasting performance, the pooled logit model with full country homogeneity tends to significantly outperform the more complex specifications. In fact, the pooled global models correctly

Table 4: Binary Logit Models of Sovereign Defaults

	(1) Global	(2) Asia	(3) Latin	(4) Africa	(5) Developed
Total Debt	0.028**	-0.006	0.024**	0.016	1.001**
IMF Credit	0.300**	1.064**	0.088	0.164	22.638**
Global Interest	0.154**	0.076	0.083**	0.000	
ForeignExch Reserves	-0.119**	-0.346**	0.005	-0.136**	-1.974**
Trade Openness	0.010*	0.004	-0.004	0.017	0.735*
Current Account	-0.060**		-0.038*		-6.186**
FDI	-0.472**		-0.230**	-0.088	-7.418
Real GDP Growth	-0.078**	0.055	-0.038*	-0.037	-4.439
Inflation	0.001	0.182		0.046*	
M2/Reserves	0.005		0.164**	0.037	
REER Overval	-0.025**	0.004	-0.008*	-0.039**	
Gov Expenditures	-0.121**	-1.217**	-0.069**	0.065	-6.157
National Saving	-0.099**	-0.059	-0.032*	-0.136**	0.378
Domestic Credit	0.010*	0.086**	-0.010**	-0.008	
Bank Assets	-0.016			-0.040*	
Gov Bank Loans	0.038**		0.011	0.031	
Asia	3.450**				
Latin	5.225**				
Africa	4.610**				
N	912	192	288	216	216
Pseudo R^2	0.584	0.595	0.665	0.622	0.947
Log-Likelihood	-227.9	-27.6	-66.7	-54.7	-1.1
BIC	592.1	118.4	218.5	195.4	55.8
% of Correct Crises	87.1	90.9	93.4	94.1	100.0

Notes: * $p < 0.05$, ** $p < 0.01$. The marginal effects, rather than the logit coefficients of the five binary logit regressions, are reported in this table. For developed countries total external debt is proxied by gross government debt due to data availability. The lower panel depicts the number of observations used in each regression (N) over the in-sample period 1980-2005, the Pseudo McFadden's R^2, the log-likelihood ratio, the Schwarz-Bayesian information criterion (BIC), and the percentage of in-sample crises correctly signaled by each model, global and regional.

Table 5: Binary Panel Logit Models using FE and RE

	Global		Asia		Latin		Africa		Developed	
	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
Total Debt	0.07**	0.02**	-0.02	0.16*	0.22**	0.14**	0.09**	0.07**	0.85*	0.68**
IMF Credit	0.81**	0.91**	5.32*	3.98**	0.95**	0.96**	0.37	0.29	12.77*	15.56*
Global Interest	0.18**	0.12*	-0.60	-0.41	0.01	0.03	0.04	-0.01		
ForeignExch Reserves	-0.21**	-0.18**	-2.43*	-1.17*	-0.34*	-0.18	-0.33**	-0.23*	1.51	-0.41
Trade Openness	0.02	0.02*	-0.06						-1.34	
Current Account	0.02	-0.03		-0.65*	0.04	-0.01	0.10	0.15*		-0.98
FDI	-0.60**	-0.33**	-2.01*	-0.13	-0.94**	-0.96**	-0.74	-0.51	3.90	0.69
Real GDP Growth	-0.05	-0.07**	0.70*	0.18	0.01	-0.05	0.08	0.08	-2.94	-2.87
Inflation	0.01	0.02*	0.84*	0.23*			0.14	0.16**		-0.43
M2/Reserves	0.11*	-0.04	-2.03	-1.02	0.54*	0.64**	0.11	0.11		
REER Overval	-0.02*	-0.03**	0.02	0.05	-0.03	-0.04	-0.03	-0.03	-0.79	0.07
Gov Expenditures	-0.06	-0.11**			-0.19	-0.24*	0.33*	0.22		-1.37
National Saving	0.01	-0.03	-0.42	0.38	-0.19*	-0.13	0.12	-0.07	-1.65	0.25
Domestic Credit	0.01	0.02*	0.65*	0.16*	-0.02	-0.03	0.01	-0.01		
Bank Assets	-0.01	-0.05**			-0.11	-0.08	-0.06	-0.07		
Gov Bank Loans	0.02	0.04**	0.17	-0.08						
N	624	912	96	192	288	288	192	216	50	225
Pseudo R^2	0.618	0.460	0.796	0.733	0.824	0.719	0.752	0.625	0.858	0.814
Log-Likelihood	-120.4	-230.6	-8.7	-16.4	-28.1	-54.4	-25.4	-49.2	-1.5	-3.4
BIC	343.8	583.9	76.7	110.9	129.8	193.7	124.4	184.4	34.4	71.9
% of Correct Crises	43.9	70.0	18.2	77.3	77.3	28.3	90.8	73.8	80.0	75.0

Notes: * $p < 0.05$, ** $p < 0.01$. The marginal effects, rather than the logit coefficients of the five fixed-effects (FE) and five random-effects (RE) binary logit regressions, are reported in this table. FE have significantly smaller number of observations as the model excludes all countries that did not experience a debt crisis over the in-sample period. For developed countries total external debt is proxied by gross government debt due to data availability. The lower panel depicts the number of observations used in each regression (N) over the in-sample period 1980-2005, the Pseudo McFadden's R^2, the log-likelihood ratio, the Schwarz-Bayesian information criterion (BIC), and the percentage of in-sample crises correctly signaled by each model, global and regional.

forewarned slightly less than 90% of the crisis episodes that occurred in the sample countries, while the fixed-effects model did not improve over a naive random guess (i.e. less than 50%). Furthermore, it is evident from the highly significant coefficients of the regional dummies included in the pooled global model, as well as the basic goodness-of-fit measures and the in-sample predictions depicted in the lower panels of all three tables, that the regional models consistently outperform the global one. Our results, hence, support the assertion of Fuertes and Kalotychou (2006) that heterogeneity seems to be regional rather than country-specific.

4.3 Multinomial Logit Model

Whereas the binary logit regression estimates the effect of a set of explanatory variables on a binary crisis variable, the multinomial logit model studies their effect on a multinomial variable DCm_{it} that allows for three states. The crisis index is defined as follows in this case

$$DCm_{it} = \begin{cases} 0 & if\ DC_{it} = 0 \\ 1 & if\ DC_{it-1} = 0\ and\ DC_{it} = 1 \\ 2 & otherwise \end{cases} \tag{4}$$

where DC_{it} denotes the binary crisis variable, and the value of zero reflects tranquil periods, the value of one denotes the first year of the crisis (see also Fuertes and Kalotychou, 2007; Savona and Vezzoli, 2015), and two refers to the post-crisis periods until the country returns to the normal state.

The maximum likelihood estimation procedure is utilized to regress the multinomial dependent variable (Y_{it}) on the lags of the proposed economic indicators (X_{it}) using the cumulative logistic distribution function

$$\begin{aligned} Pr\left(Y_{it}=0\right) = F\left(X_{it}\beta\right) &= \frac{1}{1+e^{X_{it}\beta^1}+e^{X_{it}\beta^2}} \\ Pr\left(Y_{it}=1\right) = F\left(X_{it}\beta\right) &= \frac{e^{X_{it}\beta^1}}{1+e^{X_{it}\beta^1}+e^{X_{it}\beta^2}} \\ Pr\left(Y_{it}=2\right) = F\left(X_{it}\beta\right) &= \frac{e^{X_{it}\beta^2}}{1+e^{X_{it}\beta^1}+e^{X_{it}\beta^2}} \end{aligned} \tag{5}$$

where β^1 measures the effect of a change in the indicators on the probability of entering into a crisis, while β^2 measures their effect on the probability of being in the post-crisis period. To make the reported coefficients comparable to those of the binary logit regression, we report the marginal effects of the indicators. Furthermore, we continue to use the Huber-White robust variance estimator to allow for country-specific variances.

A country usually undergoes two types of development in its economic state during post-crisis periods. If the crisis deepens after it originally hit the economy, the economic indicators will be expected to worsen. On the other hand, as the authorities undertake corrective policies, the economic indicators may improve as the economy recovers from the crisis and returns to its normal state. However, it is not possible *ex-ante* to identify which development will take place or prevail. Therefore, we cannot form reasonable expectations regarding the signs of the coefficients during the post-crisis periods. Table 6 presents the results of the global and regional multinomial logit regressions, where the upper panel depicts the marginal effects of the variables on the probability of entering into a new crisis, with the lower panel focusing on the probability of being in a post-crisis period.

With respect to the Asian countries, the variables reflecting the extent of debt exposure and the health of the banking system appear to have a significant effect on the probability of going into crisis, as well as on being in one, with IMF credit having the largest marginal effect. In Latin America, the results show that the indicators are only able to explain the post-crisis periods rather than crisis onsets. This implies that the debt situation in these countries tends to worsen after the entry year, which is also evident from the higher marginal effects of the indicators in the post-crisis period compared to the crisis onset periods. Turning to Africa and the Middle East, the overvaluation of the domestic currency and diminishing national savings are both associated with crisis and post-crisis periods. On the other hand, IMF credit, global interest rate and FDI can only explain the onset of sovereign defaults, while increasing debt, the erosion of foreign exchange reserves and rising rates of inflation are more observed during periods of recovery. Finally, the estimated model for Western Europe shows that most indicators are statistically significant, where their marginal effects are usually higher during the post-crisis periods, indicating the deepening of the crises after their onset. The two indicators with the highest marginal effects are credit from the IMF and government expenditures.

We proceed in the next section to discuss further the estimation results of all three econometric methods together in order to identify the variables that can act as warning

Table 6: Multinomial Logit Models

		(1) Global	(2) Asia	(3) Latin	(4) Africa	(5) Developed
Crisis Period $DCm_{it} = 1$	Total Debt	0.020*	0.742**	-0.001	0.012	0.176*
	IMF Credit	-0.357	-7.290**	-0.888	-1.825*	2.262**
	Global Interest	0.212**		0.154	0.465*	
	ForeignExch Reserves	-0.114	-0.056	-0.027	-0.079	-1.989*
	Trade Openness	-0.009	0.110	-0.024		0.120*
	Current Account	-0.149**		-0.080		
	FDI	-0.184		0.778*	-1.147*	
	Real GDP Growth	-0.067*				-0.102
	Inflation	-0.012	-0.409		-0.247	
	M2/Reserves	-0.049		0.002		-1.018*
	REER Overval	-0.026*	0.110	-0.001	-0.088**	
	Gov Expenditures	-0.027		0.058	0.094	-2.343*
	National Saving	-0.026			-0.134*	-0.738**
	Domestic Credit	0.010	0.464*	-0.004	-0.003	
	Bank Assets	0.019	-0.778**		0.009	
	Gov Bank Loans	0.021	1.017*	0.003	0.009	
Post-Crisis Period $DCm_{it} = 2$	Total Debt	0.032**	0.335**	0.047**	0.025*	0.197**
	IMF Credit	0.421**	-0.050	0.226*	0.181	4.133**
	Global Interest	0.148**		0.083*	-0.019	
	ForeignExch Reserves	-0.139**	0.184	-0.018	-0.178**	-1.092**
	Trade Openness	0.016**	-0.057*	-0.007		0.100**
	Current Account	-0.031		-0.065*		
	FDI	-0.579**		-0.555**	0.061	
	Real GDP Growth	-0.091**				-0.470*
	Inflation	0.001	0.132		0.072**	
	M2/Reserves	0.019		0.263**		-1.600*
	REER Overval	-0.026**	-0.083*	-0.009	-0.040**	
	Gov Expenditures	-0.150**		-0.096*	0.100	-1.502**
	National Saving	-0.116**			-0.148**	-0.809**
	Domestic Credit	0.012*	0.223**	-0.017**	-0.010	
	Bank Assets	-0.032*	-0.349**		-0.036	
	Gov Bank Loans	0.047**	0.382*	0.017	0.021	
	N	912	192	288	216	216
	Pseudo R^2	0.566	0.725	0.638	0.612	0.786
	Log-Likelihood	-275.2	-21.2	-89.9	-64.6	-5.8

Notes: * $p < 0.05$, ** $p < 0.01$. The marginal effects, rather than the logit coefficients of the five multinomial logit regressions, are reported in this table. For developed countries total external debt is proxied by gross government debt due to data availability. The lower panel depicts the number of observations used in each regression (N) over the in-sample period 1980-2005, the Pseudo McFadden's R^2 and the log-likelihood ratio.

indicators of sovereign debt crises in general. Furthermore, we evaluate the constructed EWS using a number of criteria to assess their relative performance.

5 Discussion and Evaluation

After detailing the various econometric methods used to construct EWS, this section further examines and discusses their results, and conducts a horse-race between the constructed EWS to evaluate their performance from a policy-maker's point of view.

5.1 Warning Indicators of Sovereign Defaults

Considering the estimated coefficients of the three econometric techniques in more detail, we find that the variables suggested by economic theory and the preliminary quantitative analysis are able to provide a good measure of the likelihood of an approaching debt crisis. The estimation results show in particular that the debt exposure variables (ratio of external debt to GDP and credit acquired from the IMF) are significant indicators in all regions, which was also previously reported by Lausev *et al.* (2011) and Jedidi (2013). However, the multinomial logit model shows that IMF credit is low before crisis onsets and high afterwards in the case of emerging countries, but is high before and after for more advanced economies. A probable explanation for this phenomenon is that developed countries may have easier and quicker access to IMF funds, while emerging-country governments could apply for a loan before the onset of a crisis and only obtain the funds after the crisis has hit the economy. When using the binary logit estimation, the positive after-crisis impact of the IMF credit appears to be dominant.

Another general finding that is consistent among all the regions is that governments tend to keep their expenditures low before and during times of crises. Arguably, public spending is increased only during tranquil times when the finances are available and there is no serious threat of compounding unsustainable debt. In addition to these variables and in line with the previous literature (Peter, 2002; Lausev *et al.*, 2011), rising FDI inflows, current account improvements, and growth of national savings tend to signal a reduced need for external credit, and thus less pressure on government debt in Latin America. As for the countries in Africa and the Middle East, inflation causes external debt servicing to be more expensive, overvaluation of the domestic currency drains the required foreign reserves to service maturing sovereign debts, and trade openness seems to be doing more harm than good by making the African economies more vulnerable to foreign shocks. These results are also consistent with those found by Manasse and Roubini (2009) and Savona and Vezzoli (2015). With respect to South-East Asia, the accumulation of foreign reserves increases the ability of the government to service its external obligations (as also reported by Jedidi, 2013), whereas banking sector distress and increased pressure on the real exchange rate tend to contribute to debt problems, leading to twin or even triple crises. Finally, in developed countries, and consistent with Peter (2002) and Savona and Vezzoli (2015), the rate of real GDP growth, the ratio of national savings to nominal GDP, and the banking sector variables (domestic credit and bank assets growth) have a major influence on the likelihood of debt crises.

5.2 EWS: A Horse-Race

The previous detailed discussion of the statistical significance of the proposed indicators of sovereign debt problems, though important for policymakers, is not sufficient to conclude whether the estimated models can act as an effective EWS. It is, therefore, imperative to test and compare the forecasting performance of the different econometric methods. This requires selecting a cut-off probability such that, if the predicted probability[12] of any model exceeded that threshold, the model is said to issue a signal of a forthcoming crisis. By comparing these signals to the actual crisis episodes defined by the dependent variable, the following contingency table can be constructed:

	Crisis	No Crisis
Signal	A	B
No Signal	C	D

[12]To obtain the predicted probabilities for the dynamic signal extraction models, the signals S^j_{t-h} generated by the most reliable indicators j (with $NTSR < 0.5$), in each region r, are summarized into a single composite crisis index I_{rt} as follows: $I_{rt} = \sum_{j=1}^{n} \frac{S^j_{rt-h}}{NTSR_{rj}}$. The values of each composite indicator are converted into a series of conditional probabilities of approaching crises in each country i. These are calculated as the ratio between the number of times I_{rt} falls within a lower and an upper bound (exogenously determined over the in-sample period for each region separately) and a crisis did occur over the crisis window, and the total number of periods it falls within this interval in general.

where outcomes A and D reflect "good" signals of crisis and tranquil periods, respectively, while outcome C depicts the failure to predict an actual crisis (i.e. "missed crisis") and outcome B denotes a "false alarm" as the warning signal was not followed by a crisis within the specified crisis window.

Clearly, choosing a low (high) cut-off probability increases the probability of false alarms (missed crises). In practice, Fuertes and Kalotychou (2007) argued that false alarms are less important to policymakers than missed crises, since the actual costs of adopting pre-emptive policies are usually less severe than the significant economic and social losses of unanticipated crises. On the other hand, Savona and Vezzoli (2015) warned against trivializing the costs associated with false alarms, as they tend to trigger negative market sentiments and affect international reputation. Furthermore, it should be noted that false alarms are not always *mistakes* caused by the predictive failure of the EWS, but could simply be the result of undertaking suitable policy actions that were successful in mitigating or avoiding a crisis that would have hit otherwise. In addition, due to the way the models are designed, a signal issued "too early" (i.e. outside the crisis window) is also counted as a false alarm, even though being followed by an actual crisis.

Therefore, most studies select the cut-off threshold so as to minimize a joint error measure, that is, the in-sample $NTSR$, or to maximizes Youden's J-statistic. The former is calculated as the ratio of bad to good signals:

$$NTSR \;=\; \frac{P\left(B|B \cup D\right)}{P\left(A|A \cup C\right)} \tag{6}$$

while the latter is defined as the hit rate (HR) minus the false alarm rate (FAR):

$$J \;=\; HR - FAR = P(A|A \cup C) - P(B|B \cup D) \tag{7}$$

We follow the recommendation of Savona and Vezzoli (2015) in selecting the optimal cut-off probability that maximizes the J-statistic rather than the one that minimizes the $NTSR$, as they found that the J-statistic is quite robust to extreme errors, whereas the $NTSR$ could lead to extreme thresholds causing close-to-zero FAR and negligible HR.

Hence, to assess the predictive power of the models considered, we calculate three measures, namely the percentage of crisis onsets and crisis periods (duration) correctly forewarned, along with the ratio of false alarm signals. These are calculated based on in-sample predictions, as well as the more policy-relevant out-of-sample predictions. With respect to the parametric methods, we calculate two types of out-of-sample predictions. First, the models are estimated once over a sub-sample of the data, and the regular h-step-ahead forecast is calculated over the h most recent held out observations. We also implemented a new recursive forecasting technique that allows for dynamic predictions. In particular, our *dynamic-recursive forecasting* technique estimates the model several times, each time adding one further out-of-sample observation ("recursive") along with the predicted probability of the previous period ("dynamic"), and generating a 1-step-ahead forecast.

While a rolling-window forecast approach[13] drops early observations and adds new ones, keeping the forecast window constant, the recursive method is more of an incremental approach in that it adds one additional out-of-sample observation each time without dropping older ones, then iteratively updates the model. This recursive approach was applied quite recently in the context of constructing EWS for debt crises by Savona and Vezzoli (2015). However, the originality of our approach stems from including a dynamic dimension to the forecast process in that we incorporate the previous period's predicted probability in the following years recursive estimation. This dynamic updating, along with the recursive iteration, makes the maximum amount of information available to the model during the forecasting process, and hence can improve its predictive power. To the best of our knowledge, this dynamic-recursive technique has not been implemented before in the context of forecasting financial crises (be they currency, banking, or sovereign debt).

As a consequence, the sub-sample 1980-2005 is used to generate predictions of 2006; then the sub-sample 1980-2006 and the predictions of 2006 are used to forecast the crisis probability in 2007; and so on. Incorporating new information in the EWS as it becomes available can reasonably be expected to improve significantly its forecasting performance. The results are summarized in Table 7, where the first panel focuses on the in-sample forecasts, the second panel depicts the regular out-of-sample, and the last panel presents the dynamic-recursive forecasting performance.

Starting with the in-sample performance, the multinomial logit estimation does not emerge as an appropriate method to construct effective EWS. In fact, it seems better in predicting tranquil periods than crisis episodes, where the models are able to predict correctly more than 90% of the tranquil periods in Latin America and Africa, and about 100% in Asia and developed countries. However, the HR of crisis episodes only range

[13] This is also referred to as the style rotation strategy by Levis and Liodakis (1999).

Table 7: Evaluating the Performance of the EWS

	Global			S-E Asia			Latin America			Africa & ME			Developed		
	SA	BL	ML	SA	BL	ML	SA	BL	ML	SA	BL	ML	SA	BL	ML
In-sample Forecasts															
Optimal Cut-off	35	35	–	10	15	–	50	45	–	50	30	–	5	30	–
Total Crises	263	263	263	22	22	22	152	152	152	85	85	85	4	4	4
% of Correct Crises	20.2	**87.1**	79.5	40.9	**90.9**	72.8	58.6	**93.4**	56.7	62.4	**94.1**	68.6	100.0	**100.0**	75.0
Detected Onsets	6	14	9	2	2	2	13	8	3	6	3	3	2	2	1
Total Onsets	26	26	26	3	3	3	15	15	15	6	6	6	2	2	2
% of Correct Onsets	23.1	53.9	34.6	66.7	66.7	66.7	*86.7*	53.3	20.0	*100.0*	50.0	50.0	100.0	100.0	50.0
% of False Alarm	10.8	8.6	4.1	28.2	8.8	1.2	26.5	12.5	8.1	22.9	12.2	9.2	25.9	0.0	0.0
Regular Out-of-sample Forecasts															
Cut-off Prob.	–	–	–	10	15	–	50	45	–	50	30	–	5	30	–
Total Crises	–	–	–	0	0	0	4	4	4	1	1	1	7	7	7
% of Correct Crises	–	–	–	–	–	–	**100.0**	0.0	25.0	0.0	100.0	100.0	**85.7**	71.4	0.0
Detected Onsets	–	–	–	0	0	0	1	0	0	0	0	0	2	2	0
Total Onsets	–	–	–	0	0	0	1	1	1	0	0	0	3	3	3
% of Correct Onsets	–	–	–	–	–	–	100.0	0.0	0.0	–	–	–	66.7	66.7	0.0
% of False Alarm	–	–	–	39.3	1.8	16.1	22.5	5.0	3.7	19.4	9.7	9.7	23.2	12.5	0.0
Dynamic-Recursive Forecasts															
% of Correct Crises	–	–	–	–	–	–	–	**100.0**	25.0	–	**100.0**	100.0	–	71.4	28.6
Detected Onsets	–	–	–	–	0	0	–	1	0	–	0	0	–	2	1
Total Onsets	–	–	–	–	0	0	–	1	1	–	0	0	–	3	3
% of Correct Onsets	–	–	–	–	–	–	–	100.0	0.0	–	–	–	–	66.7	33.3
% of False Alarm	–	–	–	–	0.0	7.2	–	12.5	3.7	–	6.5	6.5	–	10.7	0.0

Note: This table presents the in- and out-of-sample predictive performance of all three econometric methods used to construct EWS for debt crises. SA denotes Signal Approach, BL Binary Logit models, and ML Multinomial Logit models. Optimal cut-off is defined as the cut-off probability that maximizes J-statistic, % of correct crises is the percentage of total crisis periods (duration) correctly forewarned, and % of correct onsets is the percentage of correctly predicted new crisis onsets. "-" stands for "n.a.", arising as: (a) ML does not use cut-off probability to classify observations into the three states (normal, crisis, post-crisis), but it calculates for each observation the probability of being in each state; the state with the highest probability is chosen as the most probable one. (b) We do not report out-of-sample forecasts for the global model, since it was found inferior to the regional models in the in-sample. (c) The dynamic-recursive technique only applies to both parametric methods (BL and ML). (d) There were no out-of-sample crisis periods in South-East Asia, and no new onsets in Africa & the Middle East.

between 55-75% in the different regions, which lies below the other two methods. On the other hand, the binary logit method is shown to outperform significantly in its forecasting performance. Particularly, in developed countries, it correctly predicts *all* onsets and crisis periods without generating any false signals, compared to the higher FAR of 25% of the dynamic signal extraction technique. Similar results are found in South-East Asia, where all models predict correctly two of the three crisis onsets, but the binary logit model stands out with lower FAR and a much higher HR of crisis periods of 90%. On the other hand, with respect to Latin America and Africa, the dynamic signal extraction approach is able to forewarn 87% of the crisis onsets in the former region and 100% in the latter. However, this relatively high HR compared to the other two methods comes at the expense of a higher FAR of over 25%. In addition, the dynamic signal extraction is only able to detect half of the crisis episodes in these regions, while the binary logit models can forecast over 90% of the episodes with a more reasonable average of 10% FAR. In addition, the regional heterogeneity suggested by the goodness-of-fit measures is further confirmed here, as the HR of crisis onsets and episodes of the regional models significantly outperform those of the global one. Therefore, we exclude the global model from our further analysis.

Compared to Savona and Vezzoli (2015), the only paper that previously used the signal approach, we find that the dynamic version which takes the regional heterogeneity of the indicators into consideration significantly outperforms the static version where developed and emerging economies are pooled together. More specifically, their model was able to predict correctly about 80% of the in-sample crises at a FAR of 45%. Our models, on the other hand, have a collective HR of about 90% (being able to correctly predict 23 out of the 26 crisis onsets) and generate almost half as many false alarms (25% on average) as the static version. Our binary logit models that account for the entire crisis period also appear to improve over the traditional models in the previous literature. In particular, Ciarlone and Trebeschi (2005) generated 36% false signals while only correctly predicting 72% of the in-sample crisis episodes, whereas the model estimated by Pescatori and Sy (2007) had a sensitivity of 86% and a false alarm rate of 14%. Even Manasse *et al.* (2003), who were able to issue about 5% false signals, could only foresee 75% of the crisis episodes. Furthermore, Savona and Vezzoli (2015), as the only study that included developed countries, albeit pooled with emerging markets, had an in-sample hit rate of 77% with a false alarm rate of 16%.

Although in-sample forecasts are important for evaluating the performance of EWS, the more relevant test for policymakers is that of out-of-sample forecasts. In this respect, the figures in the lower two panels of Table 7 imply that, even when applying our novel dynamic-recursive forecasting technique, the multinomial logit method continues to perform relatively poorly. It only predicted one of the three crisis onsets that occurred in developed countries, and none of those that hit Latin America. On the other hand, the dynamic signal extraction models correctly forewarned two of the three crisis onsets in developed countries and 86% (six of the seven) of their entire crisis years. Furthermore, in Latin America *all* crisis onsets and default periods are correctly forewarned at a slightly lower FAR of around 20%. In Africa and the Middle East, where no new crises occurred during the holdout period 2006-2012, the FAR remains around the in-sample range of 20%, while in South-East Asia it doubled to almost 40%.

However, the warning signals generated by the South-East Asian composite index cannot be considered as real false alarms, but as indicators of an alarming debt situation that did not progress into a full-fledged crisis. In fact, although no actual debt crises occurred in these countries, it is evident that their sovereign debt condition was rather worrisome. A study conducted by Jiang and Xu (2014) reported the alarming rapid growth of government debt and argued that the outbreak of a debt crisis is a possibility in China. Moreover, a recent report by Moody's (2014) highlighted that India has a high fiscal deficit and a large government debt burden that could become unsustainable if the current low GDP growth and high inflation persist over the medium term. With respect to South Korea, the ratio of government debt to GDP has also grown significantly over the holdout period.

Regarding the binary logit models, the regular forecasts are unable to detect either the crisis onset or any of the four crisis periods in Latin America. In developed countries, five out of the seven episodes (70%) and two of the three (67%) onsets are correctly forewarned. These results are in line with the small number of papers that reported out-of-sample forecasts, with Ciarlone and Trebeschi (2005) able to predict two out of five (40%) crisis episodes in emerging economies, and Manasse *et al.* (2003) correctly forecasting 45% of sovereign defaults. Nevertheless, our findings improve substantially when applying the dynamic-recursive forecasting technique, proving the superiority of this method over the regular forecasts. Particularly, in Asia, where no crises occurred during the holdout period, all tranquil periods are captured without issuing any false alarms. In Latin America and Africa, all crisis periods are correctly signaled while generating FAR of around 10%.

Moreover, the estimated model in developed countries is able to forewarn the debt crises at a lower false alarm rate than the one generated by the regular forecasting technique. These ratios outperform to a great extent the most accurate EWS constructed so far. Between Fuertes and Kalotychou (2006) and Savona and Vezzoli (2015), a maximum of 75% of the out-of-sample crisis episodes was forewarned with a false alarm rate of 15-30%. Fuertes and Kalotychou (2007) had a better sensitivity ratio of 82%, but at a relatively high FAR of 23%.

Thus, the overall results of the horse-race highlight the superiority of our specification of the binary logit model over the more traditional ones, as well as over the multinomial specification of the dependent variable. Furthermore, our dynamic-recursive forecasting technique improves significantly on the regular out-of-sample forecasts, enabling the binary logit models to closely match the performance of the dynamic signal approach, but with a much reduced occurrence of false alarms.

6 Conclusion

This study investigates the performance of several econometric techniques (binary logit, multinomial logit, dynamic signal extraction) recently developed to construct more effective EWS for sovereign debt crises in different developed and developing country regions. We contribute to the literature by, for the first time, designing a separate EWS for developed countries, improving the specification of the binary logit models by treating the entire crisis period as individual episodes, and developing a more accurate forecasting technique.

Our models show that, in order to construct an effective EWS for sovereign debt crises, it is crucial to include variables that allow for the possibility of spillover from the banking sector and the foreign exchange market. Furthermore, the predictive performance of the EWS is significantly improved when using simple pooled models that account for the regional heterogeneity of the signaling indicators, and using the dynamic-recursive forecasting technique to generate out-of-sample forecasts. Regarding the in-sample forecast, the dynamic signal approach can predict more crisis onsets, while the binary logit model outperforms in generating significantly lower false alarms and correctly forewarning crisis duration. As for out-of-sample performance, the binary logit model using our novel dynamic-recursive forecasting technique is able to forecast correctly most of the out-of-sample crises onsets and periods, while generating half as many false signals as the dynamic signal extraction technique. Multinomial logit models, on the other hand, fall behind the other two econometric techniques in all cases.

Thus, in conclusion, an EWS based on our binary logit model can be recommended to policymakers in the different regions considered, particularly when the avoidance of negative market sentiments and damage to international reputation, as potentially triggered by false signals of sovereign debt problems, are high on their agenda. Possible extensions to our work could allow for indicators with a forward-looking perspective in order to capture the possibility of self-fulfilling crises. These might include credit default swaps, sovereign bond spreads and other variables that can be used to assess sovereign credit ratings; these would, however, need to be made available in developing countries on a more timely basis. In addition, future research could attempt to include variables that express the possibility of contagion from neighboring countries. Unfortunately, the annual frequency of the data would rule out the use of same-year information on defaulting sovereigns for assessing the possibility of contagion to other economies.

Acknowledgments

M.D. acknowledges the financial support of the Egyptian Ministry of Higher Education. We are also grateful to very constructive comments by the editor and an anonymous referee; all remaining errors are ours.

References

Berg, A., Borensztein, E., Pattillo, C., 2005. Assessing Early Warning Systems: How Have They Worked in Practice? IMF Staff Papers 52 (3), 462–502.

Bussiere, M., Fratzscher, M., 2006. Towards a New Early Warning System of Financial Crises. Journal of International Money and Finance 25 (6), 953–973.

Candelon, B., Dumitrescu, E.-I., Hurlin, C., 2014. Currency Crises Early Warning Systems: Why They should be Dynamic. International Journal of Forecasting 30, 1016–1029.

Casu, B., Clare, A., Saleh, N., 2012. Towards a New Model for Early Warning Signals for Systemic Financial Fragility and Near Crises: An Application to OECD Countries, *MPRA Working Paper* No. 37043, University Library of Munich, Germany.

Chakrabarti, A., Zeaiter, H., 2014. The Determinants of Sovereign Default: A Sensitivity Analysis. International Review of Economics and Finance 33, 300–318.

Ciarlone, A., Trebeschi, G., 2005. Designing an Early Warning System for Debt Crises. Emerging Markets Review 6 (4), 376–395.

Demirguc-Kunt, A., Detragiache, E., 1998. The Determinants of Banking Crises in Developing and Developed Countries. IMF Staff Papers 45 (1), 81–109.

Eaton, J., Gersovitz, M., 1981. Debt with Potential Repudiation: Theoretical and Empirical Analysis. Review of Economic Studies 48 (2), 289–309.

Fioramanti, M., 2008. Predicting Sovereign Debt Crises using Artificial Neural Networks: A Comparative Approach. Journal of Financial Stability 4 (2), 149–164.

Frankel, J., Rose, A., 1996. Currency Crashes in Emerging Markets: An Empirical Treatment. Journal of International Economics 41 (3-4), 351–366.

Fuertes, A.-M., Kalotychou, E., 2006. Early Warning Systems for Sovereign Debt Crises: The Role of Heterogeneity. Computational Statistics & Data Analysis 51, 1420–1441.

Fuertes, A.-M., Kalotychou, E., 2007. Optimal Design of Early Warning Systems for Sovereign Debt Crises. International Journal of Forecasting 23 (1), 85–100.

Gourinchas, P.-O., Obstfeld, M., 2012. Stories of the Twentieth Century for the Twenty-First. American Economic Journal: Macroeconomics 4 (1), 226–265.

IMF, 2011. Modernizing the Framework for Fiscal Policy and Public Debt Sustainability Analysis, *Policy Paper*, International Monetary Fund.

Jedidi, O., 2013. Predicting Sovereign Debt Crises: A Panel Data Approach Using Composite Indices, *CREM-CNRS Working Paper*, University of Rennes.

Jiang, L., Xu, Y., 2014. Will China's Local Debt Crisis Break Out? Evidence from Local Government Financing Practive. International Journal of Business and Commerce 3 (10), 32–42.

Kaminsky, G., Lizondo, S., Reinhart, C., 1998. Leading Indicators of Currency Crises. IMF Staff Papers 45 (1), 1–48.

Lausev, J., Stojanovic, A., Todorovic, N., 2011. Determinants of Debt Rescheduling in Eastern European Countries. Economic Annals 56 (188), 7–31.

Lestano, L., Jacobs, J., Kuper, G., 2003. Indicators of Financial Crises Do Work! An Early Warning System for Six Asian Countries, *CCSO Working Paper* No. 200313, Centre for Economic Research, University of Groningen.

Levis, M., Liodakis, M., 1999. The Profitability of Style Rotation Strategies in the United Kingdom. The Journal of Portfolio Management 26 (1), 73–86.

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

Manasse, P., Roubini, N., Schimmelpfennig, A., 2003. Predicting Sovereign Debt Crises, *IMF Working Paper* No. 221, International Monetary Fund.

Moody's, 2014. India's Government Debt Structure mitigates Macroeconomic Imbalances, *Global Credit Research*, Moody's Investors Service.

Pescatori, A., Sy, A., 2007. Are Debt Crises Adequately Defined? IMF Staff Papers 54 (2), 306–337.

Peter, M., 2002. Estimating Default Probabilities of Emerging Market Sovereigns: A New Look at a Not-So-New Literature, *HEI Working Paper* No. 06/2002.

Reinhart, C., 2002. Default, Currency Crises, and Sovereign Credit Ratings. World Bank Economic Review 16 (2), 151 – 169.

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

A Appendix: Crisis Episodes by Country

Country	Crisis Episodes	Comment
Latin America		
Argentina	1983-1992	Arrears exceeded 20% of total debt
	2001-2005	Credit from IMF exceeded 500% of quota
Brazil	1983-1994	Massive rescheduling and restructuring schemes
	1999	IMF credit exceeded 200% of country quota
	2002-2003	IMF credit reached more than 600%
Mexico	1982-1990	40% of the debt was rescheduled or forgiven
	1995-1996	IMF loans increased to more than 600%
Chile	1983-1990	Rescheduling of 20% of debt
Paraguay	1986-1990	Defaults on 20% of debt
Dominican Rep.	1983-1999	IMF loans increased to about 250% of quota
	2003-2005	Rescheduling 20% of outstanding debt
	2010-2011	IMF credit reached 400% of country quota
Ecuador	1983-1995	Arrears reached 40% of total debt
	1999-2000	Rescheduling 30% of the debt
Venezuela	1989-1996	Rescheduling over 50% of the debt
Bolivia	1980-1985	Arrears increased to 20% of outstanding debt
	1986-1994	IMF credit reached 200% of quota
Peru	1980-1997	Arrears increased to 50% of outstanding debt
Panama	1983-1997	Arrears reached about 60% of outstanding debt
Costa Rica	1981-1991	IMF credit reached over 200% of quota
South and East Asia		
Indonesia	1998-2003	IMF credit of about 400% of quota
Philippines	1981-1990	10% of debt was restructured
China	–	No significant external debt problems
India	–	No significant external debt problems
Malaysia	–	No significant external debt problems

Thailand	1981-1982	IMF loans increased to 280% of quota
	1997-1999	IMF loans reached 400% of quota
South Korea	1980-1982	IMF loans reached 400% of quota
	1997-1998	IMF credit accounted to over 1500% of quota
Singapore	–	No significant external debt problems
Middle East and Africa		
Egypt	1980-1991	Arrears increase to more than 20% of debt
Jordan	1989-1994	Default on more than 10% of the debt
South Africa	1985-1989	Failure to meet about 50% of debt obligations
Lebanon	1985-1991	Arrears reached 12% of total debt
Morocco	1981-1989	Loans from IMF reached about 400% of quota
Tunisia	1986-1991	IMF credit reached 150% of country quota
Algeria	1990-1996	Rescheduling over 10% of debt principal
Nigeria	1988-1999	Defaulting on 60% of outstanding debt
Central Africa	1981-2006	Arrears increased to more than 30% of debt
Advanced Europe		
Greece	2010–	Debt reached 160% of GDP, IMF credit is over 1700% of quota
Portugal	1986	IMF credit amounted to 150% of quota
	2011–	IMF credit of 1700% of quota
Spain	–	No significant external debt problems
Ireland	2011–	IMF credit reached 1300% of quota
Italy	–	No significant external debt problems
Belgium	1992-1994	Debt increased to over 140% of GDP
Sweden	–	No significant external debt problems
Germany	–	No significant external debt problems
UK	–	No significant external debt problems

Source: Authors' calculations.

B Appendix: Debt Crisis Indicator Variability

	Full Model			Regional Models			
Indicator	$\sigma_{NoCrisis}$	σ_{Crisis}	F-stat	Dev.	Asia	Latin	Africa
Debt Exposure							
Total Debt	88.7	176.4	0.3*	✓	×	✓	×
IMF Credit	0.8	2.0	0.2*	✓	✓	✓	×
Global Interest	3.3	4.3	0.6*	×	×	✓	×
External Sector							
ForeignExch Reserves	19.7	7.5	6.9*	×	✓	✓	✓
Trade Openness	66.9	37.9	3.1*	×	✓	×	×
Export Growth	11.3	13.0	0.8	×	×	×	×
Current Account	6.7	5.5	1.5*	×	✓	×	×
FDI	4.4	2.1	4.4*	×	✓	✓	✓
Macroeconomic Condition							
Real GDP Growth	4.1	7.6	0.3*	×	✓	×	✓
Inflation	8.6	48.6	0.0*	×	✓	✓	✓
M2/Reserves	41.2	34.5	1.4	×	×	✓	×
REER Overval	14.1	19.5	0.5*	✓	×	×	✓
Gov Expenditures	5.1	8.5	0.4*	×	×	×	✓
National Saving	9.3	8.7	1.2	✓	×	×	×
Banking Sector							
Domestic Credit	45.0	66.6	0.5*	✓	×	✓	✓
Bank Assets	40.2	50.5	0.6*	✓	×	✓	×
Gov Bank Loans	19.1	23.1	0.7*	×	✓	×	×

Notes: The first two columns depict the standard deviations of the variables using the full model during normal and crisis periods, respectively. Column three shows the F-statistic of the variance comparison test using the full model, while the last four columns illustrate the result of the F-test in each region separately. Both * and ✓denote statistical significance of the variance differences at the 95% level, while × denotes no significant difference in the variances.