Theoretical and Applied Economics
Volume XXII (2015), No. 1(602), pp. 15-34

Predicting balance of payments crises for some emerging economies

Archana KULKARNI
University of Hyderabad, Hyderabad, India
kulkarni.arch@gmail.com
Bandi KAMAIAH
University of Hyderabad, Hyderabad, India
kamaiahbandi@gmail.com

Abstract. *The study aims at developing an Early Warning System for predicting balance of payments crises for 17 emerging economies, which constitute a relatively homogenous group, over the period 1975-2012. We construct an index of exchange market pressure, based on monthly depreciations of the nominal exchange rate and declines in reserves, to identify crisis episodes. To construct the index we propose a new weighting scheme using principal components analysis, as an improvement over the conventionally used precision-weighting scheme. Probit regressions are used to identify key macroeconomic indicator variables that can predict the onset of a crisis. These include the ratio of M2 to reserves, short-term debt to reserves, export growth, ratio of total reserves to external debt, change in reserves, openness and overvaluation of the real exchange rate. From alternative specifications, we identify the best model based on various accuracy measures. We use criteria such as area under the Receiver Operating Characteristic, Quadratic Probability Score, Pseudo R2 and Kuiper's Score to evaluate model performance. Empirical results show the warning system exhibits a high degree of accuracy and performs well. The variables identified show significant ability to signal vulnerability of the external sector of the economy. Policymakers can use the early warning system as the core of a larger set of variables on their radar to take pre-emptive measures to avoid crises or dampen their effects.*

Keywords: Early warning system, Foreign exchange, Balance of payments, Crises, Model evaluation.

JEL Classification: C52, C53, F31, F47.

1. Introduction

Balance of payments crises have always been a matter of concern for emerging economies given the history of their development; with economic upheaval accompanying financial liberalization. The Tequila Crisis and Asian Financial Crisis were two events in particular that testified to the need for an early warning system that could help economists and policymakers identify a looming crisis and either adopt pre-emptive measures, or, if far advanced and inevitable, to define a strategy to mitigate its effects. A few pioneering models were proposed around the time of these crises that proved to be seminal works of research. Subsequent studies represented efforts to test the forecasting performance of these models and improve upon these systems by exploring alternative techniques and indicator variables of a crisis.

The present study attempts to contribute to the vast literature on predicting balance of payments crises by testing an early warning system for a group of 17 emerging economies that constitute a relatively homogenous group. A balance of payments crisis can be understood as a sharp devaluation of a country's currency in the face of speculative attacks or the inadequacy of foreign exchange reserves to meet a country's external obligations and import needs. It is important to note that the crisis definition adopted for the paper means that both successful and failed speculative attacks are designated as crises. This can translate into a number of false alarms of crises. However, as pointed out by Berg et al. (2005), these may simply be cases where a crisis was averted by sheer chance or external circumstances that proved beneficial to the country's situation despite it exhibiting greater vulnerability. Therefore, these situations could still provide useful information to policymakers looking to address any signs of vulnerability shown by an economy.

Capturing such vulnerability is thus essential to our study. Therefore, in order to identify crisis episodes, we construct an index of 'exchange market pressure' for each country. The index combines information on the depreciation of the exchange rate and decline in foreign exchange reserves. We use a new weighting scheme for the index, wherein principal components analysis is used to arrive at the weights for each variable. This approach improves on the conventional practice of using precision weights that has been criticized in recent times by introducing statistical and economic logic in the selection of weights. An increase in the index above a certain critical threshold indicates significant stress in the foreign exchange market and that the economy is experiencing a crisis.

In our study, a probit model is used to predict the probability of a crisis based on information from key indicator variables, real and financial, that have a bearing on the external sector. The paper tests alternative specifications of the model, arrives at the best fit, determines accuracy using various measures and evaluates model performance. Based on the empirical results we are able to identify key indicator variables such as the ratio of M2 to reserves, short-term debt to reserves, export growth, change in reserves, ratio of total reserves to external debt, openness and overvaluation of the real exchange rate, that are able to signal the onset of a crisis. These variables can form the core of a set of variables on the radar of policymakers that can provide insights into the external health of the economy.

The outline of the paper is as follows. Section 2 provides a brief review of the literature on early warning systems. Section 3 provides details of the data and methodology used for the study, elaborating on the selection of weights for the exchange market pressure index. Section 4 presents the empirical results and section 5 presents some concluding remarks.

2. Literature Review

In the mid-to-late 1990s, the pioneers of the use of Early Warning Systems (EWSs) to predict balance of payments and currency crises were Frankel and Rose (1996) and Kaminsky et al. (1998). A group of IMF researchers working under the aegis of the Developing Countries Studies Division (DCSD) then extended and further built upon these models to predict the probabilities of crises using probit models. They sought to study the efficacy of the earlier EWSs in predicting crises in the aftermath of the Asian Financial Crisis and then propose their own EWS that could improve upon previous efforts.

Berg and Patillo (1999 and 1999 b) and Berg et al. (1999) worked extensively on the subject and it is the approach adopted by them that guides our study. Three models were evaluated by these researchers – Kaminsky et al. (1998) who used a signals approach, Frankel and Rose (1996) who used a probit model and Sachs et al. (1996) who studied the Tequila Crisis using data from a cross-section of countries. Kaminsky et al. (1998) defined an index of “exchange market pressure”, using monthly percentage depreciations of the exchange rate and percentage declines in reserves, that signals a looming crisis for a sample of 20 developed and developing countries. 15 indicator variables were examined that were expected to provide signals over a 24 month period prior to the crisis and individual thresholds for each of these were identified based on the minimum value of the noise-to-signal ratio. They found that 8 indicators were informative – deviations of the real exchange rate from trend, growth of M2 as a fraction of reserves, export growth, change in international reserves, “excess” M1 balances, growth of domestic credit as a share of GDP, the real interest rate and growth in terms of trade.

Frankel and Rose (1996) examined 100 developing countries using annual data over the period 1971-1992. The definition of a crisis adopted was a depreciation of the nominal exchange rate of at least 25 percent, and which exceeds the change in the exchange rate over the previous year by at least 10 percent. Informative indicators for the study included high foreign interest rates and domestic credit growth, overvaluation of the real exchange rate relative to the average level, large current account and fiscal deficits as a share of GDP, low levels of FDI as a proportion of external debt, low external concessional debt and lack of openness.

The Sachs et al. (1996) definition of a crisis was the same as Kaminsky et al. (1998); however, they examined the severity of a crisis across countries, given the expected timing of a crisis while Kaminsky et al. (1998) and Frankel and Rose (1996) looked to predict the timing of crises. The three models were evaluated based on goodness of fit and correlations between the predicted probabilities of a crisis and actual incidence. The evaluation of these models revealed that the Kaminsky et al. (1998) model enjoyed mixed success, the Frankel and Rose (1996) model was not able to predict probabilities of crises in a useful manner and the Sachs et al. (1996) model performed poorly. Based on these findings, Berg et al. (1999) proposed a multivariate probit model including explanatory variables that were highlighted in the literature on the Asian Financial Crisis. Some of the variables expected to be informative were high domestic credit growth, an overvalued real exchange rate relative to trend, high ratio of M2 to reserves and a large CAD. The multivariate probit showed a definite improvement over previous EWSs; however the authors cautioned that even the best models have limited predictive power.

Berg et al. (1999) also compared EWSs with forecasts made and models developed by private analysts such as Goldman Sachs' GS-Watch, Credit Suisse First Boston's Emerging Markets Risk Indicator and the Deutsche Bank Alarm Clock. This involved comparing EWSs to bond spreads, agency ratings and risk scores. The study found that while the Kaminsky et al. (1998) model performed well out of sample, the DCSD model performance deteriorated in this context. Private sector forecasts performed poorly out of sample. Therefore they concluded that while EWSs are useful, they need to be supplemented with other available information.

The DCSD models trace their development to the work of Berg et al. (1999 a and 1999 b). While the Kaminsky et al. (1998) model assumes that the probability of a crisis "…is a step function of the value of the indicator" (Berg et al., 1999: pp. 18) and a crisis signal is issued by the indicator variable upon crossing a particular threshold, the DCSD model assumes that crisis probabilities are linked in a linear manner with changes in the indicator variables. Thus, the probit model was used which identified six key variables – deviations of the real exchange rate from trend, current account to GDP ratio, export growth, reserve growth, level of M2 to reserves and ratio of short-term debt to reserves. Another addition to the EWS literature was made by the IMF in the form of the Policy Development and Review (PDR) model developed by Mulder, Perrelli and Rocha (2002). The PDR model added data available from the corporate sector such as balance sheet data to the exercise of predicting crises.

Kaminsky (1998) used a composite indicator of crises by calculating a weighted average of the indicators. The weights used were the inverse of each indicator's noise-to-signal ratio. The probability of crisis was estimated for each value of the index by examining how frequently a given value of the index is followed by a crisis within 24 months. The role of contagion was also examined by some studies such as Eichengreen et al. (1996) and J. P. Morgan (1998) where the variable assumes a value of zero or one based on whether there was a crisis elsewhere in the world or measures the number of recent crises in other countries.

Glick and Hutchinson (1999) studied banking and currency crises over the period 1975-97 for 90 industrial and developed countries using a multivariate probit model. They found that banking crises are useful leading indicators of currency crises although the converse does not hold true. The index used by the authors differed from the one proposed by Kaminsky et al. (1998) in that percentage depreciations of the real exchange rate were considered rather than the nominal exchange rate. Informative indicator variables were found to be overvaluation of the real exchange rate, export revenue growth and ratio of M2 to reserves. Other variables included in the study were current account by GDP ratio, nominal and real M2 growth, nominal and real domestic credit, M2 by reserve money multiplier and ratio of the budget surplus to GDP. The accuracy of predictions was tested using pseudo R2 statistics, quadratic probability scores (QPS) and log probability scores (LPS).

For India in particular, two studies have attempted to predict balance of payments crises. The first by Callen and Cashin (1999) followed the approach of Berg and Patillo (1999 a, b) and Berg at al. (1999), predicting crises using a multivariate probit model based on monthly data for 23 developing economies. The independent variables used in the model included the overvaluation of the real exchange rate, the current account deficit as a percentage of GDP, reserve and export growth and the ratio of short-term debt to reserves. They found that crisis probabilities for India have been low since 1991 and that the model was able to predict with accuracy and identify contributing factors to crisis. The second study by Goyal (2012) used a sample of 9 select economies – Brazil, Chile, India, Korea, Malaysia, Mexico, Russia, South Africa and Turkey over the period 1970-2010. Informative indicator variables were found to be composition of liabilities in terms of equity and debt, proportion of short-term debt, size of foreign exchange reserves, per capita income, policy environment (proxy used to capture this was the Index of Government Effectiveness). Variables found to be non-informative were exchange rate misalignment - measured by deviations of the REER from the base level, global liquidity conditions measured by US bond spreads and the government deficit to GDP ratio. Crisis probabilities were primarily estimated to identify the sustainable level of the current account (im)balance for India based on the threshold level of external liabilities associated with a crisis. The sustainable level of current account deficit for India was found to be between 2.4 to 2.8 percent of GDP based on certain assumptions about the growth of the economy and cost of financing external debt.

The literature on current account sustainability also provides several useful insights into potential indicator variables that may be used in EWSs. For instance, Adedji and Handa (2008), in their study of the Nigerian economy, identified some such indicator variables – size of external debt, changes in the real exchange rate, economic growth, trade openness, fiscal deficits, foreign exchange reserves, political instability and uncertainty. Milesi-Ferretti and Razin (1996) listed several structural and macroeconomic factors that contribute to external vulnerability. These include low savings and investment, economic and export growth, degree of openness, composition of debt, financial structure of the economy, weak banking systems, openness of the capital account which can render the economy vulnerable to sudden stops of capital flows, persistent appreciation of the real exchange rate, fiscal balance, political stability and credibility, market expectations reflected in bond and interest rate spreads and foreign exchange reserves. Amongst these, key variables, based on case studies of individual countries, were found to be the ratio of external debt to GDP, interest payments on debt, ratio of exports to GDP, appreciation of the real exchange rate, fiscal balances and two variables that are difficult to measure quantitatively – the health of the banking system and political stability.

A recent contribution to the literature on EWSs is that of Catao and Milesi-Ferretti (2013) who defined external crises based on events of default on debt or rescheduling, as well as large multilateral or IMF assistance to overcome debt related difficulties. The study was based on a sample of 70 countries that included 41 emerging market economies over the period 1970-2011. They examined questions such as: Are there thresholds beyond which build-up for net external liabilities raises the risk of crises? Does composition of external liabilities matter for crisis risk? The independent variables used in the probit model included ratios of net foreign assets to GDP, net external debt assets to GDP, net external portfolio equity to GDP, net FDI to GDP, foreign exchange reserves to GDP, relative per capita income, current account balance to GDP, REER gap, global stock market volatility proxied by the VIX index, changes in lagged credit growth to GDP, current account gap, public debt to GDP, US corporate credit spread, trend output growth, foreign exchange regime, openness, capital controls, institutional quality, external debt assets to GDP, external debt liabilities to GDP and an outlier FDI dummy. Crisis risk was found to increase sharply when NFL/GDP crossed 50 percent and whenever the NFL/GDP ratio crossed the country-specific mean by 20 percentage points. Debt liabilities increased the risk of crisis. Current account deficits were found to be powerful predictors of crises and reserves reduced crisis risk by more than other asset holdings. The parsimonious probit model was found to have substantial predictive power, in and out-of sample, particularly with respect to the 2008-11 crises. An interesting feature of their study was the use of the Receiver Operating Characteristic (ROC) technique as a model selection criterion and use of treatment effects models to compare the pre-crisis and non-crisis behavior of indicator variables.

An important issue that has emerged in recent times is the construction of the Exchange Market Pressure index used to detect crises. The original index proposed by Girton and Roper (1977), used equal weights for the two components – the depreciation of the nominal exchange rate and reserve losses – for aggregation. Eichengreen et al. (1994) added an interest rate component to the index, as monetary authorities could look to stave off a crisis through interest rate hikes and the behavior of this variable could reveal pressures in the foreign exchange market. They used a weighting scheme known as the precision weighting scheme for the weighted average of the three components, where the weights were the inverse of each component's variance. This technique was adopted so as to avoid assigning undue importance to any component that exhibited greater volatility; thereby more volatile components received a lower weight compared to less volatile components. This practice became the norm in subsequent research, although some indices dropped the interest rate component due to non-availability of data.

In recent times, this weighting scheme has received some criticism for being imprecise and ignoring the fact that reserve fluctuations could capture both genuine market-determined behavior as well as policy responses of the monetary authorities(1). This leads to only market-determined variance being taken into account in the precision weighting scheme. As this distinction is difficult to make in practice, the solution that remains is to assign equal weights to these variables when calculating a weighted average. This implies that both the exchange rate and reserves are seen as instruments of policy, likely to be a reasonable assumption for most emerging economies. Therefore, only policy-related variance is taken into account in this case and if this variance component is closer to the total variance of the variable, we may assume that the market-related variance component is zero. Even the opposite case will hold true in this respect; at the other extreme, if the market related variance dominates the movement of the variable, it would be fair to assume it accounts for the total variance of the variable. The equal weighting scheme is thus an improvement over the precision weighting scheme. The ideal weights, acknowledged by Eichengreen et al. (1994) as well, are the elasticities of the variables in the foreign exchange market. Owing to the difficulties in estimating these parameters, researchers have advocated the equal weighting scheme as the second best method (Li et al., 2006 and Willett et al., 2005). As an alternative, our study attempts to make a contribution by putting forward a weighting scheme where principal components analysis is used to determine the weights of variables that make up the index, thereby employing statistical and economic logic in deciding the weights for the index.

Although several studies have been conducted to test the efficacy of Early Warning Systems, the heterogeneity of countries on which these are based may lead to generalizations being made about the utility of different variables in signaling the onset of a crisis. The present study attempts to address this issue by examining a relatively homogenous group of 17 emerging economies in order to develop a EWS for predicting balance of payment crises. Our study also looks to improve upon existing weighting schemes in the construction of the exchange market pressure index. Thus, the present study seeks to develop an Early Warning System for predicting balance of payments crises using a probit model for a group of 17 emerging economies over the period 1975 to 2012.

3. Data and methodology

3.1 Data sources and the model

The present study proposes an Early Warning System (EWS) to predict balance of payments crises for 17 emerging economies using annual data over the period 1975 to 2012. The sample of emerging economies used for the study consists of Argentina, Brazil, Colombia, Costa Rica, Hungary, India, Indonesia, Malaysia, Mexico, Morocco, Philippines, Romania, South Africa, Thailand, Turkey, Ukraine and Venezuela(2). The choice of countries was restricted by constraints of data availability. Data for the study was taken from the International Financial Statistics, World Bank, Bank for International Settlements and the Reserve Bank of India.

In order to detect a crisis an index of 'exchange market pressure' is constructed using a weighted average of monthly data on percentage depreciations of the nominal exchange rate (vis-à-vis the US dollar) and the percentage decrease in reserves. We do not use the interest rate component due to non-availability of data for some countries in the sample. The weights for the two variables are determined by using principal components analysis for each series; the first principal component for each series is used as the factor loading or weight. This is a contribution of the paper as it marks a departure from the usual practice of using the inverse of variances of the variables as weights – a practice which has been criticized for being imprecise.

Principal components analysis is a statistical data reduction technique that strips the information content in a variable to the most essential components. It uses an eigen value decomposition of the correlation matrix of variables to yield an uncorrelated linear combination of the variables that accounts for a dominant part of their variance. This preserves all the information in the original variable, with the first principal component accounting for most of the variance of a variable. This technique was developed through the work of Pearson (1901) and Hotelling (1933).

By using the first principal component of each variable as the weight, although we fail to decompose the variance into market-related and policy-related components, we are able to use only the dominant part of the variance in our analysis. The first principal component accounts for approximately 55 per cent of the variance of both variables for almost all the countries in the sample. On average, we find that the two variables in the index receive equal weights in nearly all cases. This ensures the neither variable dominates the index in the aggregation process and so information from both variables is incorporated harmoniously in the construction of the index. In the absence of data on relative elasticities of variables in the foreign exchange market, our method coincidentally yields equal weights for the variables for almost all the countries in our sample which was identified as the second best solution.

After constructing the index using a weighted average of the variables, we adopt the conventional threshold for detecting crisis events. Any observation from the index which exceeds the mean by three standard deviations is considered as signaling the onset of a crisis within the next 12 months. In this way, a binary variable is created as the dependent variable for the probit model based on the values of the index with a value of 0 being assigned to a non-crisis or tranquil year and 1 to a crisis year. If the index signals a crisis during any month of the year, that year is demarcated as a crisis year. The study does not exclude consecutive crisis years if the signals from the index persist, as is done in other studies, in order to gauge the behavior of indicator variables in cases where a crisis remains unresolved after a year of onset. Thus an index of exchange market pressure is prepared for each country to identify crisis episodes.

In the tradition of Berg and Patillo (1999a, b) and Berg et al. (1999 and 2005), a probit model is used to predict the onset of a balance of payments crisis. This has the advantage of identifying key indicator variables that can signal a crisis based on the significance of coefficients from the probit regression. The utility of employing a probit model lies in the nature of the dependent variable, which is a qualitative, binary variable and in the underlying logic of the model. Both probit and logit models, in their formulation, consider a latent or unobservable variable that determines the outcome or value of the dependent variable (which in this context represents the probability of it being a crisis or non-crisis year). This latent variable could be determined by any or all of the independent variables and when this variable crosses a particular (unobservable) threshold a crisis ensues. If the predicted probability equals or exceeds this threshold, a crisis will follow. In our study, each of the indicator variables is relevant to the external health of the economy. When these variables cross a particular threshold, they are able to signal that the economy is showing significant signs of distress that demands immediate attention from policymakers. Although the threshold for each variable is unobservable, we can gauge the importance of each variable based on its statistical significance and also obtain a numerical magnitude of its contribution to the probability of a balance of payments crisis. Thus, these variables act as leading indicators of a crisis.

The choice between a probit and logit model is often considered arbitrary and unimportant. The dependent variable in the logit model is the log of the odds ratio, which represents the odds of success or the event occurring to that of failure or the event not occurring. The dependent variable in the probit model has a less intuitive interpretation and can be understood as a z-score or probability index. Besides differences in interpreting coefficients from the models, the logistic distribution has thicker tails, relative to the distribution for the probit. Studies have shown that for sample sizes less than 500 observations the probit model yields more accurate results (3). Therefore, given that our sample size is less than 500, the probit model seems the natural choice for our study.

The variables considered in the study as indicator variables are the ratio of M2 to reserves, ratio of short-term debt to reserves, ratio of external debt to gross national income, growth of GDP, export growth, ratio of total reserves to external debt, change in reserves, external balance on goods and services as a percentage of GDP, openness (measured as ratio of the sum of exports and imports to GDP) and REER overvaluation. All variables are measured in terms of percentages except for the ratio of M2 to reserves, measured as the natural log of the ratio and REER overvaluation. REER overvaluation is measured as the deviation of REER (Real Effective Exchange Rate) from trend and this cyclical component is extracted using the Hodrick-Prescott filter. The choice of variables was constrained by availability of data for all the countries during the study period.

The ratio of M2 to reserves is a variable highlighted in the literature on the Asian Financial Crisis as being a useful indicator of vulnerability. For economies that are open to capital flows, the ratio of reserves to imports is no longer considered an informative indicator of vulnerability. Instead, the ratio of M2 to reserves is thought to be a better indicator as it signifies the proportion of domestic liquid liabilities that may be converted to foreign currencies as the domestic exchange rate depreciates in a crisis, putting greater pressure on foreign exchange reserves. Following a flight of capital, the adequacy of reserves in terms of monetary aggregates becomes important owing to a loss of confidence in the domestic currency. For open economies, one of the most important indicators of vulnerability has been found to be the ratio of short-term debt to reserves, as creditors may refuse to roll-over debt leading to a self-fulfilling crisis, of the kind described by second-generation theories of crises.

The ratio of external debt to gross national income indicates whether the country can use its productive capacity to divert more production to exports and use these revenues to finance its external debt. GDP growth and export growth also have similar implications for the ability to prevent or tide over a crisis. Trade deficits are widely thought to be an important indicator of stress as they represent the outflow of foreign exchange and a drain on reserves that are the first line of defense of an economy experiencing a balance of payments crisis. REER overvaluation symbolizes the misalignment of the domestic currency, generating expectations that a correction will ensue to bring it back to equilibrium. Openness is another variable that has been highlighted in the literature on crises. Studies have found that relatively open economies are more vulnerable to exogenous shocks and the vagaries of foreign demand, however a diversified export base can help an economy show greater resilience to crises. We also examine the influence of different Exchange Rate Arrangements (ERA) on crises (4).

Thus, the probit model takes the general form of equation (3.1.1) below:

$$\begin{aligned} C_t = {} & f(\beta_0 + \beta_1 \ln(M2/Reserves) + \beta_2 STD/Reserves + \beta_3 Ext.Debt/GNI + \\ & + \beta_4 GDPgrowth + \beta_5 Exportgrowth + \beta_6 TotalDebt/Reserves + \\ & + \beta_7 TradeBalance/GDP + \beta_8 Openness + \beta_9 REERoverval + \beta_{10} ERA + u_t) \end{aligned} \tag{3.1.1}$$

Where the dependent variable C_t represents the probability of the event, conditional on all the explanatory variables i.e. Pr ($C_t = 1 | X_i$), and equals

$C_t = 1$; i.e. $C_t \geq C^*$; if there is a crisis (C* represents the unobservable latent variable)

$= 0$ *i.e.* $C_t < C^*$; indicates a tranquil/non-crisis year (3.1.2)

The probability that $C^* \leq C_t$ can be calculated from the cumulative normal distribution function as follows:

$$\Pr(C_t = 1 | X_i) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k) \tag{3.1.3}$$

where Pr is the probability that a crisis occurs and Φ represents the standard normal CDF. The inverse of the normal CDF is used to arrive at the beta coefficients in the probit model as follows:

$$C_t = \Phi^{-1}(C_t) = \Phi^{-1}(\Pr) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{3.1.4}$$

The probit model is estimated using maximum likelihood estimation. The interpretation of coefficients is not as direct and intuitive as for linear regression models. The coefficient β_1, for instance, represents the change in the z-score or probability index for a one unit change in the explanatory variable X_1. The signs of the coefficients, however, can be interpreted directly; a positive coefficient indicating an increase in the probability of a crisis and a negative coefficient indicating a decrease in probability.

3.2. Prediction accuracy and performance evaluation

Alternative model specifications are examined to arrive at the model that provides the best fit. The accuracy of different models is examined using an array of criteria (5). For the study, the optimal cut-off probability is determined based on the minimum value of the noise to signal ratio. We find that an optimal cut-off of 0.5 minimizes the noise-to-signal ratio and yields the most accurate predictions. This implies that if the predicted probability of a crisis is equal to or exceeds 0.5 the observation is classified as a crisis observation, while predicted probabilities less than the cut-off value are classified as non-crisis or tranquil periods. The observations and empirical predictions are categorized in the manner defined in the table below to determine the accuracy of the model.

	True Value			
Predicted Result	Event	Crisis	No Crisis	Total
	Crisis	True Positive	False Positive	All Predicted Crises
	No crisis	False Negative	True Negative	All Predicted Non-Crises
	Total	All True Crises	All True Non-Crises	Total Sample Size

Reproduced from Candelon et al. (2012).

In the table above, the diagonal elements represent the number of observations correctly classified as Crisis and Non-Crisis respectively, at the chosen cut-off (i.e. True Positive (TP) and True Negative (TN)). The off-diagonal elements represent the misclassified observations (False Positive (FP) and False Negative (FN)) at the chosen cut-off. The row totals represent the actual number of crises and non-crises or tranquil periods. The column totals represent the predicted number of crises and non-crises events.

In order to select the best model that provides the most accurate predictions we use the following criteria:

1) $$\text{Noise to Signal Ratio} = \frac{\text{Number of False Positives}}{\text{Number of True Positives}} \qquad (3.2.1)$$

2) $$\text{Sensitivity or the Hit Rate} = \frac{\text{True Positives}}{\text{All True Crises}} \qquad (3.2.2)$$

3) $$\text{1- Specificity or the False Alarm Rate} = \frac{\text{False Positives}}{\text{All True Non- Crises}} \qquad (3.2.3)$$

4) $$\text{Total Accuracy} = \frac{\text{True Positives + True Negatives}}{\text{Total Sample Size}} \qquad (3.2.4)$$

5) Younden Index = Hit Rate – False Alarm Rate (i.e. Sensitivity – (1-Specificity)) (3.2.5)

6) Total Misclassification Errors

$$TME = FN^* L_{fn} + FP * L_{fp} \qquad (3.2.6)$$

where L_{fn} = 1/All True Crises and L_{fp} = 1/All True Non-Crises

7) Weighted Misclassification Errors

$$WME = \frac{(FN + W_{fn}) + (FP + W_{fp})}{\text{Total sample size}} \quad (3.2.7)$$

Where $W_{fn} = L_{fn}/(L_{fn} + L_{fp})$ and $W_{fp} = L_{fp}/(L_{fn} + L_{fp})$

The weights for TME and WME represent the losses that arise from wrongly classifying crisis periods as non-crisis and vice-versa, with WME weights capturing relative losses as opposed to absolute losses captured by TME weights. The hit rate represents the number of crisis periods which are correctly identified, while the false alarm rate represents the number of tranquil periods which are incorrectly identified as crisis periods.

These criteria are used to select the best models in terms of accuracy and then these models are evaluated using four criteria that are widely used in the credit-scoring literature. The Receiver Operating Characteristic is one such criterion, used primarily in engineering and more recently in epidemiology. The ROC curve shows the trade-off between sensitivity and 1-specificity. In other words, it examines the tradeoff between making a type I error and a type II error, at a given cut-off. Lower cut-offs favor identifying crises correctly at the cost of more misclassifications of tranquil periods as crisis events i.e. lower type I errors but higher type II errors. This can be visualized graphically as a curve; the area under the curve ranges from 0 to 1, with one indicating perfect accuracy in making predictions. Models with a higher area under the ROC curve (AUROC) indicate a better fit, indicating that the model is able to correctly assign higher probabilities to a crisis relative to a non-crisis event and thereby distinguish between the two events.

The other criteria include Kuiper's Score which measures the difference between the hit rate and false alarm rate with a positive value indicating that the model performs well. If the forecasts are random in nature, Kuiper's Score, on average, will be zero. We also use the Quadratic Probability Score which is analogous to a mean square error measure and part of the business cycle literature. The QPS ranges from 0 to 2, with 0 indicating perfect accuracy. The measure is defined as follows:

$$QPS = 2/T \sum (P_t - C_t)^2 \quad (3.2.8)$$

Here P_t represents the predicted probability of a crisis or non-crisis event at time t and C_t represents the actual occurrence of a crisis or non-crisis event at time t.

Pseudo R^2 is a goodness of fit measure, analogous to the R^2 measure that is used to represent the percentage of variation of the dependent variable that is explained by the explanatory variables in ordinary least squares regression. Pseudo R^2, however, cannot be interpreted in the same way as the R^2 measure as it is based on estimation by maximum likelihood. There are several types of Pseudo R^2s. We use McFadden's R^2 for our study, which can be interpreted as describing the improvement from a model without explanatory variables to the fitted model or as the square of the correlation between the observed and predicted values from the regression. Therefore, in this sense, a higher

Pseudo R^2 value indicates good performance of the model, when comparing alternative models for the same data. McFadden's R^2 has the following formula:

$$\text{Pseudo } R^2 = 1 - [\ln \hat{L}(M_{Full}) / \ln \hat{L}(M_{Intercept})] \tag{3.2.9}$$

Where $\hat{L}$ is the estimated likelihood, M_{Full} represents a model with explanatory variables and $M_{Intercept}$ represents the intercept-only model (without predictors). The numerator can be understood as the sum of squared errors and the denominator as total sum of squares. As the likelihood ranges between values of 0 to 1, the log of a likelihood will be less than or equal to zero. Unlikely models will have larger log likelihood values while models which are more likely will have low log likelihood values. As a result, if the ratio above is small then McFadden's Pseudo R^2 indicates that the full model is a substantial improvement over the intercept only model.

The study thus uses various accuracy measures to identify the optimal cut-off and best model and then evaluates the performance of the model using methods such as the area under the ROC curve, Kuiper's score, Quadratic Probability Score and Pseudo R^2. The results of the empirical analysis are given in the section that follows.

4. Empirical results and analysis

The study uses annual data from 17 emerging economies over the period 1975 to 2012 to develop an Early Warning System (EWS) to predict balance of payments or external crises. Alternative specifications are tested through probit regressions to identify key indicator variables, listed in Section 3, that have the potential to signal the onset of a crisis and distinguish such an event from a tranquil or non-crisis period. Probit regressions are used to estimate the probability of a crisis (6). The dependent variable is a binary/dichotomous variable constructed using a particular threshold for an index of exchange market pressure as explained in section 3. Whenever the index signals the onset of a crisis, the observation assumes a value of 1 and a value of 0 is assigned to tranquil or non-crisis periods. In all, the sample consists of 25 crises out of a total of 432 observations, indicating an unconditional probability of crisis of approximately 0.06 (7).

The probit regressions reject the panel specification in favor or ordinary probit regressions, finding no significant difference between the panel estimator and the pooled estimator. This supports the contention of the study that the countries selected in the sample form a homogenous group and do not exhibit considerable panel level variation. Therefore, we employ ordinary probit regressions for our study (as done by a majority of other researchers) and clustered robust standard errors to take into account any heteroskedasticity across the country errors.

Alternative specifications of the model are tested to arrive at the best fit based on the value of the area under the Receiver Operating Characteristic (AUROC). The larger the area under the ROC curve, the better the fit of the model. The range for the AUROC is between 0 and 1, with a value of 1 indicating a perfect fit. We find that even a single indicator variable like REER overvaluation has statistically significant predictive power in an Early Warning System; the AUROC value is high at approximately 0.6323 indicating that the model performs well. In testing alternative specifications of the model we adopt the following rule - variables are retained in the model if they are found to be statistically significant and if the coefficient conforms to the expected sign, otherwise they are dropped. The variables that drop out if this manner are the trade balance, exchange rate arrangement (ERA) and external debt to gross national income for being statistically insignificant and GDP growth for its coefficient not conforming to the expected sign.

The model that has the best fit has an AUROC of 0.8904. Therefore, this model is chosen as our Early Warning System. The indicator variables that are part of this system are the ratio of M2 to reserves, short-term debt to reserves, change in reserves and REER overvaluation at the first lag, export growth at the second lag and the change in the ratio of total reserves to external debt and change in openness. The empirical results from estimation of the Early Warning System are presented in Table 1. We find that a one unit increase in the ratio of M2 to reserves leads to a statistically significant increase in the z-score or probability index of crisis by 0.58. Similarly, we find that an increase in the ratio of short-term debt to reserves, openness and REER overvaluation lead to a statistically significant increase in the probability of a crisis. On the other hand, an increase in export growth, the ratio of total reserves to external debt and an increase in reserves lead to a statistically significant decrease in the probability of a crisis.

Table 1. *Shows the results from the estimation of the Early Warning System*

Variables	Coefficients
$\ln(M2/Reserves)_{t-1}$	0.5822* (0.2138)
Short-term debt/Reserves $_{t-1}$	0.0010* (0.0004)
Export growth $_{t-2}$	-0.0163* (0.0066)
Δ Total Reserves/ External debt	-0.0478* (0.0155)
Change in Reserves $_{t-1}$	-0.0074** (0.0044)
Δ Openness	0.0288** (0.0164)
REER overvaluation $_{t-1}$	3.6239* 1.76455
Constant	-2.7270* (0.3999)

Note: ** indicates statistical significance at the 10 per cent level and * at the 5 per cent level. Standard errors are given in parentheses.

We use an array of accuracy measures to test the precision of predictions from the EWS. Based on the minimization of the noise-to-signal ratio at varying cut offs, we identify 0.5 as the optimal cut-off for predicting a crisis. This implies that predictions that equal or exceed 0.5 are classified as crisis events while those lesser than 0.5 are classified as non-crisis events (8). The results from the accuracy measures are presented in Table 2.

Table 2. *Shows the results of various accuracy measures for the probit model*

Accuracy Measure	Value
Noise-to-Signal Ratio (NSR)	0.2000
Sensitivity	0.2000
1-Specificity	0.0025
Total Accuracy	0.9514
Younden Index	0.1975
Total Misclassification Errors (TME)	0.8025
Weighted Misclassification Errors (WME)	0.0509

Note: The optimal cut-off for crisis identification is 0.5.

At the optimal cut-off of 0.5, the EWS is able to correctly identify 95.14 per cent of the observations, with 20 percent of crises being correctly identified and 99.75 per cent of tranquil periods being correctly classified. Although crisis identification appears low, it is imperative to remember that at lower cut offs such as 0.1, although a higher percentage of crises are correctly identified (72 per cent) an efficiency cost is associated with this predictive gain as the number of false alarms increases. This implies a decline in the percentage of the observations being called correctly at 87.73 per cent- representing a drop of approximately 7.4 percentage points in accuracy. This loss of efficiency is captured more effectively by the noise-to-signal ratio. With 0.5 as the cut off the noise-to-signal ratio is low at 0.2, while at the cutoff of 0.1 it increases dramatically to 2.56. At higher cut-offs such as 0.75, we find that zero crises are identified. This indicates that efforts to reduce the number of false alarms by setting unrealistically high cut offs can prove detrimental to the objective of identifying signals of stress in the economy. The details of predictive performance of the EWS at different cut offs are presented in Table 3. Empirical results from the noise-to-signal ratio measure at various cut-offs are presented in Table 4.

Table 3. *Shows the goodness of fit of the probit model at various cut off probabilities*

Model	Goodness of Fit at Various Cut-off Probabilities	
EWS	10% cut off	
	% of observations correctly called	87.73%
	% of crises correctly called	72.00%
	% of non-crises correctly called	88.70%
	25% cut off	
	% of observations correctly called	93.29%
	% of crises correctly called	44.00%
	% of non-crises correctly called	96.31%
	30% cut off	
	% of observations correctly called	93.06%
	% of crises correctly called	32.00%
	% of non-crises correctly called	96.81%
	50% cut off	
	% of observations correctly called	95.14%
	% of crises correctly called	20.00%
	% of non-crises correctly called	99.75%
	75% cut off	
	% of observations correctly called	93.98%
	% of crises correctly called	0.00%
	% of non-crises correctly called	99.75%

Table 4. *Shows the accuracy of the probit model at various cut off probabilities in terms of Noise-to-Signal Ratios*

Cut-off probability	Noise-to-Signal Ratio (NSR)
0.08	3.1111
0.10	2.5556
0.25	1.3636
0.30	1.6250
0.50	0.2000
0.75	Not a Number

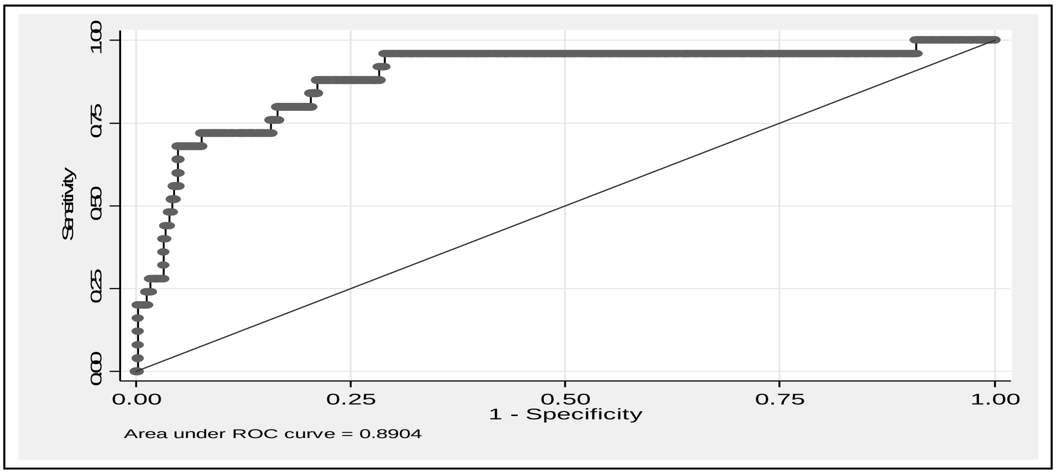
Note: At the 0.75 cut off, 0 crises are identified by the model and since division by zero is not meaningful, the NSR this cut off cannot be measured.

Results from measures to evaluate the performance of the model are presented in Table 5. For this purpose we use four measures, details of which are provided in section 3. We find that Kuiper's Score is positive; indicating that the model generates more hits relative to false alarms and this is an important finding indicating efficiency from the perspective of economic policy. The Quadratic Probability Score is low and close to zero which represents perfect accuracy; thus the models perform reasonably well in terms of predictive ability. The pseudo R^2 is 0.3024, which indicates that the model fits the data well. The most important criterion remains the area under the ROC curve which is 0.8904, indicating good performance by the model. Figure 1 shows the area under the ROC curve for the EWS. The correlation between the observed values and the predicted values is also found to be positive and statistically significant at 0.4804.

Table 5. *Shows the results from measures to evaluate model performance*

Model Evaluation Criteria	Value
Area under ROC curve (AUROC)	0.8904
Kuiper's Score	0.1975
Quadratic Probability Score (QPS)	0.1025
Pseudo R^2	0.3024

Figure 1. *Shows the area under the Receiver Operating Characteristic curve (AUROC) for the Early Warning System*



For India, we find that crisis probabilities have remained low throughout the sample period, never approaching the optimal cut off value of 0.5 during the entire period. This finding is significant as India's current account deficit has been a cause for concern in recent times, leading to fears of an external crisis for the economy.

Thus, we find that key indicator variables such as the ratio of M2 to reserves, short-term debt to reserves, export growth, ratio of total reserves to external debt, change in reserves, openness and REER overvaluation are able to signal the onset of a crisis and act as leading indicators of a balance of payments crisis. An early warning system designed with these variables is able to correctly predict crisis and non-crisis events in excess of 90 per cent of the time. Various measures used to evaluate model performance also indicate that the model performs reasonably well. Therefore, policymakers in emerging market economies can monitor these variables to identify significant signs of external stress displayed by an economy and take suitable remedial action.

5. Conclusions

Early Warning Systems (EWS) have witnessed significant development since they were first proposed in the mid- to- late nineties. The present study seeks to use these significant advances to propose a relatively parsimonious model using data that is easily available in the public domain to monitor a nation's external health. This model enables us to predict the probability of balance of payments crises for 17 emerging economies based on the statistical significance of a small set of indicator variables from a probit regression. The paper differs from a majority of previous studies as it focuses solely on 17 emerging economies which comprise a relatively homogenous group compared to previous studies which examine advanced, emerging and developing economies collectively.

We also propose a new weighting scheme to construct an index of 'exchange market pressure' which signals the onset of a crisis within a year whenever the index exceeds its mean by three standard deviations. Previous studies used the inverse of variances of the two variables which comprise the index – depreciation of the nominal exchange rate and change in reserves – as the weights to construct a weighted average for the index, a procedure which has been criticized in recent times. Instead, we use principal component analysis for the two variables and use the first principal component as the weight in order to construct the index. This procedure retains the information from each variable, while capturing the dominant part of the variance of each variable. This procedure coincidentally yields equal weights for the two variables for nearly all the countries in the sample, which has been dubbed the second best solution for the weighting scheme in the absence of data on relative elasticities which is the ideal solution.

We test alternative specifications of the model and arrive at the best fit based on the area under the Receiver Operating Characteristic curve (AUROC). The key indicator variables that are able to predict the onset of a crisis are found to be the ratio of M2 to reserves, short-term debt to reserves, export growth, ratio of total reserves to external debt, change in reserves, openness and the overvaluation of the Real Effective Exchange Rate (REER). The optimal cut-off for predicting the probability of a crisis is found to be 0.5 based on minimization of the noise to signal ratio. We also use other accuracy measures such as sensitivity, specificity and total accuracy among others to test the precision of the predictions. The performance of these models is evaluated using the area under the ROC curve, Kuiper's score, the Quadratic Probability Score (QPS) and the Pseudo R^2. All these measures indicate that the model performs reasonably well in predicting crises and in the ability to make a distinction between crisis and tranquil periods. An important aspect to note is that predicted crisis probabilities have remained low for India and much below the optimal cut off probability in recent times despite concerns about the current account and fiscal deficits for the economy.

Future research in this direction can focus on the nature of debt and its composition, data that is harder to obtain, to examine how crisis probabilities change based on these variables for the emerging economies and supplement these using forward looking variables such as bond spreads.

Notes

(1) See Li et al. (2006) who show how the practice of precision weighting leads to extreme weights in the case of the two extremes of foreign exchange regimes – fixed and freely floating, and the biases it leads to in case of failed speculative attacks on a fixed exchange rate.

(2) Data availability varies for individual countries across the time period 1975-2012; in that sense the group of countries constitutes an unbalanced panel.

(3) For instance, see Cakmakyapan, S. and Goktas, A. (2013).

(4) This takes the form of an indicator/qualitative variable with four levels, each level representing fixed, crawling, managed floating and freely floating exchange rates respectively.

(5) We follow the approach of Candelon et al., 2012 in this regard, who apply these accuracy measures, classifying observations in the same format for their EWS. See their paper for a detailed discussion of the same.

(6) Estimation using logit models was also carried out and was found to yield similar results.

(7) The total number of crises in the sample is 29 out of a total of 432 observations. This means that the unconditional probability of a crisis is 0.0617. However, due to the use of lags in the probit regression, some missing values are generated for the indicator variables, leaving 25 crisis observations in the sample.

(8) It is important to note that the optimal cut offs for individual countries are likely to vary; however for the entire sample of emerging economies the cut off is found to be 0.5. Individual country cutoffs are difficult to estimate due to inadequate sample size when individual countries are examined in isolation.

References

Adedeji, O., Handa, J. (2008). "The size and sustainability of Nigerian current account deficits", *The Journal of Developing Areas*, 41(2), pp. 1-25

Berg, A., Pattillo, C. (1999a). "Are currency crises predictable?: a test", *IMF Staff Papers*, 46(2), pp. 107-138

Berg, A., Pattillo, C. (1999b). "Predicting currency crises: the indicators approach and an alternative", *Journal of International Money and Finance*, 18, pp. 561-586.

Berg, A., Borensztein, E., Milesi-Ferretti, GM., Pattillo, C. (1999c). "Anticipating balance of payments crises: The role of early warning systems". *Occasional Paper 186*, International Monetary Fund, Washington DC

Berg, A., Borensztein, E., Pattillo, C. (2005). "Assessing early warning systems: how have they worked in practice?", *IMF Staff Papers*, 52(3), pp. 462-502

Cakmakyapan, S. Goktas, A. (2013). "A comparison of binary logit and probit models with a simulation study", *Journal of Social and Economic Statistics*, 2(1). Retrieved from <http://jses.ase.ro/downloads/Vol2NO1/211goktas.pdf>

Candelon, B., Dumitrescu, E., Hurlin, C. (2012). "How to evaluate an early warning system: towards a unified statistical framework for assessing financial crises forecasting methods", *IMF Economic Review*, 60(1), pp. 75-113

Callen, T., Cashin, P. (1999). "Assessing external sustainability in India". *IMF Working Paper,* 181. International Monetary Fund, Washington DC

Catao, L., Milesi-Ferretti, GM. (2013). "External liabilities and crises". *IMF Working Paper,* 113. International Monetary Fund, Washington DC

Eichengreen, B., Rose, A.K. Wyplosz, C. (1994). "Speculative attacks on pegged exchange rates: an empirical exploration with special reference to the European monetary system". *NBER Working Paper* No. 4898

Eichengreen, B., Rose, A., Wyplosz, C. (1996). "Contagious currency crises". *CEPR Working Paper,* No. 1453, Centre for Economic Policy Research, London.

Girton, D., Roper, L. (1977). "A monetary model of exchange market pressure applied to the post-war Canadian experience". *The American Economic Review*, 67, pp. 537-548

Glick, R., Hutchinson, M. (1999). "Banking and currency crises: how common are twins?". *Proceedings*, Federal Reserve Bank of San Francisco, issue Sep. 7, 1999

Goyal, R. (2012). "Sustainable level of India's current account deficit". *RBI Working Paper Series,* 16/2012

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 24(6), Sep. 1933, pp. 417-441

Morgan, JP. (1998). *Event risk indicator handbook*, London

Kaminsky, G. (1999). "Currency and banking crises: the early warnings of distress". *IMF Working Paper,* 178, International Monetary Fund, Washington DC

Li, J., Rajan, R.S., Willet, T. (2006). "Measuring currency crises using exchange market pressure indices: the imprecision of precision weights". *Mimeo,* the Central University of Finance, Economics and Business, Beijing

Milesi-Ferretti, GM., Razin, A. (1996). "Current account sustainability". *Princeton Studies in International Finance,* 81

Mulder, C., Perrelli, R., Rocha, M. (2002). “The role of corporate, legal, and macroeconomic balance sheet indicators in crisis detection and prevention”. *IMF Working Paper,* 59, International Monetary Fund, Washington DC

Pearson, K. (1901). “On lines and planes of closest fit to systems of points in space”, *Philosophical Magazine*, 2(6), pp. 559-572

Willett, T., Nitithanprapas, E., Nitithanprapas, I., Rongala, S. (2005). “The Asian crises reexamined”, *Asian Economic Papers*, 3, pp. 32-87