Comparing the Performance of Logit and Probit Early Warning Systems for Currency Crises in Emerging Market Economies

Fabio Comelli[1]

International Monetary Fund, USA
fcomelli@imf.org

ABSTRACT

We compare how logit (fixed effects) and probit early warning systems (EWS) predict in-sample and out-of-sample currency crises in emerging markets (EMs). We look at episodes of currency crises that took place in 29 EMs between January 1995 and December 2012. Stronger real GDP growth rates and higher net foreign assets significantly reduce the probability of experiencing a currency crisis, while high levels of credit to the private sector increase it. We find that the logit and probit EWS out-of-sample performances are broadly similar, and that the EWS performance can be very sensitive both to the size of the estimation sample, and to the crisis definition employed. For macroeconomic policy purposes, we conclude that a currency crisis definition identifying more rather than less crisis episodes should be used, even if this may lead to the risk of issuing false alarms.

JEL Classification: F30, F32, F37

Keywords: Early warning systems, currency crises, out-of-sample performance.

I. INTRODUCTION

The global financial crisis of 2008–2009 has revived the interest of professional economists in designing and assessing the performance of early warning systems (EWS), a class of models employed to quantify the likelihood of observing financial crisis episodes in the short term. In this context, the goal of this study is to compare how two competing parametric limited dependent variable (fixed effects logit and random effects probit) EWS predict in-sample and out-of-sample currency crises in emerging market economies (EMs).[2] What makes our empirical analysis interesting is that we use a rich panel dataset which includes macroeconomic and external vulnerability indicators for 29 EMs, with monthly data between January 1995 and December 2012.

---

[1]   Corresponding author: Fabio Comelli, International Monetary Fund, 700 19th Street N.W., Washington, D.C. 20431, USA. Telephone: +1/202/623-5403, Fax: +1/202/589-5403.

[2]   Ideally one would compare the performance of a fixed effects logit estimator with that of a fixed effects probit estimator. However, as in Wooldridge (2002), the estimation of unobserved country-specific effects along with the estimation of the explanatory variables coefficients leads to obtain inconsistent estimates of the latter, particularly if the length of the panel is small.

In the EWS literature, currency crises are usually defined as large depreciations of the nominal exchange rate and/or extensive losses of foreign exchange reserves over a 24–month forecast horizon. Specifically, a currency crisis is said to occur when the exchange market pressure index – a weighted average of one-month changes in the exchange rate and foreign exchange reserves – is two or three (country-specific) standard deviations above its (country-specific) mean.[3] In this context, a relevant question is: should a currency crisis be defined as a situation when the exchange rate pressure index is two or three standard deviations above its mean? Since there is no clear consensus in the EWS literature about which crisis definition should be used, we attempt to fill this gap by taking an agnostic approach. We use two definitions of currency crisis. According to one definition, a currency crisis occurs when the exchange rate pressure index is *two* standard deviations above its mean, while according to the other a crisis occurs when the index is *three* standard deviations above its mean. For each EWS, we are interested to establish how the performance changes if we use one crisis definition or the other.

In addition, this study contributes to the EWS literature as follows. We conduct a cross-country empirical analysis to compare how two competing limited dependent variable EWS perform in predicting out-of-sample currency crises episodes. As in Candelon and others (2012) and Comelli (2013), we assess in-sample and out-of sample EWS performance by calculating optimal cut-off values for the estimated crisis probability, while in most of the EWS literature those cut-off values are selected arbitrarily. This matters because the cut-off value for the crisis probability determines the total misclassification error of a EWS.[4] Selecting cut-off values arbitrarily implies that the quantification of the total misclassification error is also arbitrary.

We find that stronger real GDP growth rates and higher net foreign assets significantly reduce the probability of experiencing a currency crisis, while high credit to the private sector increases it. By contrast, the current account balance and the measure of real exchange rate misalignment are not always statistically significant. Overall, the logit and probit EWS out-of-sample performances are broadly similar. The logit EWS is able to classify correctly between 42% and 66% of the total out-of-sample observations (e.g. crisis and tranquil periods), while the probit EWS is able to classify correctly between 41% and 64% of the total out-of-sample observations. We also find that the EWS performance is sensitive to the size of the estimation sample, and to the crisis definition used. In particular, both EWS perform better when a crisis episode is defined as a situation when the country–specific exchange market pressure index is two standard deviations above its mean.

The results offer two macroeconomic policy conclusions. First, as is common in the EWS literature, the EWS out-of-sample performance can be very sensitive to the size of the estimation sample. Specifically, the EWS total misclassification error and the probability of experiencing a currency given a crisis alarm can vary considerably if a particular year with many outlying observations (this is the case for the year 2008) is included in the estimation sample. This suggests that the total misclassification error and the probability of observing a currency crisis may crucially depend on new economic and financial data. Second, the results imply that selecting a crisis definition as a situation when the exchange rate pressure index is two standard deviations above its average value reduces the EWS total misclassification error. Therefore, for macroeconomic policy purposes, a currency crisis definition identifying more rather than less crisis episodes should be employed in a EWS, even if this may lead to the risk of issuing false alarms.

This paper is organized as follows. Section 2 reviews the relevant literature, while section 3 discusses the methodology used in this study. Section 4 presents the results obtained with the logit (fixed effects) and probit EWS, while in section 5 we compare the out-of-sample performance of the logit and probit EWS. Concluding remarks are presented in section 6.

---

[3] See IMF (2002) and Bussiere and Fratzscher (2006).

[4] The total misclassification error of an early warning system (EWS) is the sum between the percentages of missed crisis episodes and of false alarms issued by the EWS.

2. LITERATURE REVIEW

Following the episodes of severe financial distress in Mexico (1994–95) and Asia (1997–98), economists became interested in thinking about frameworks that could help policymakers anticipating episodes of financial crises, whose economic costs are well documented (Cerra and Saxena, 2008).

We divide the EWS literature contributions relevant for this study in two groups. The first group includes those studies that propose parametric (i.e. regression-based) and non-parametric (i.e. crisis signal extraction) EWS and assess in-sample and out-of-sample performances of different EWS. Kaminsky, Lizondo and Reinhart (KLR) look at the evolution of those indicators which exhibit an unusual behavior in periods preceding financial crises. When the indicator exceeds a given threshold then that indicator is issuing a signal that a crisis could take place within the next 24 months. They find that exports, measures of real exchange rate overvaluation, GDP growth, the ratio between the money stock and foreign exchange reserves and equity prices have the best track record in terms of issuing reliable crisis signals. Berg and Pattillo (1999) test the KLR model out-of-sample and show that their regression-based approach tends to produce better forecasts compared to the KLR model.

Bussiere and Fratzscher (2006) develop a multinomial logit regression-based EWS, which allows distinguishing between tranquil periods, crisis periods and post-crisis periods. They show that the multinomial logit model tends to predict better than a binomial logit model episodes of financial crisis in emerging market economies. Beckmann and others (2007) compare parametric and non-parametric EWS using a sample of 20 countries during the period included between January 1970 and April 1995. They find that the parametric EWS tends to perform better than non-parametric EWS in correctly calling financial crisis episodes. However, as noted by Candelon and others (2012), in these studies the choice of the crisis probability cut-off value is arbitrarily made and not optimally derived. Comelli (2013) compares the performance of parametric and non-parametric EWS for currency crises in 28 emerging market economies and finds that the parametric EWS achieves superior out-of-sample results compared to the non-parametric EWS.

The second group of relevant EWS literature contributions for this study includes studies that discuss the significance of the various macroeconomic indicators to explain crisis incidence. Berkmen and others (2012) looked at the change in growth forecasts by professional economists before and after the global financial crisis. They found that countries with more leveraged domestic financial systems and rapid credit growth tended to suffer larger downward revisions to their growth forecasts, while international reserves did not play a significant role. Similarly, Blanchard and others (2010) do not find a significant role played by reserves in explaining unexpected growth, which is defined as the forecast error for output growth in the semester from October 2008 until March 2009. Rose and Spiegel (2012) find that the only robust predictor of crisis incidence in the 2008 global financial crisis is the size of the equity market prior to the crisis. They are unable to link most of the other commonly cited causes of the global financial crisis to its incidence across countries. By contrast, Gourinchas and Obstfeld (2011) look at financial crisis episodes in advanced and emerging economies from 1973 until 2010. They find that for both advanced and emerging market economies, the two most robust predictors are domestic credit growth and real currency appreciation. In addition, they find that in emerging market economies the country's level of foreign exchange reserves is a significant factor in determining the probability of future crises. Borio and Drehmann (2008) build indicators to quantify financial imbalances (based on credit, equity prices and property prices). These indicators are informative about financial strains in a given country but, since they do not take into account cross-border exposures of banking systems, are unable to anticipate crisis episodes associated with losses on foreign portfolios, even when the domestic economy does not exhibit credit or asset price booms.

Llaudes and others (2010) find that foreign exchange reserve holdings helped to mitigate the growth collapse in EMs provoked by the global financial crisis.

Frankel and Saravelos (2012) estimate the crisis incidence of the 2008–2009 global financial crisis. They surveyed the existing literature on early warning indicators to see which leading indicators were the most reliable in explaining the crisis incidence. They find that foreign exchange reserves, the real exchange rate, credit growth, real GDP growth and the current account balance as a percentage of GDP are the most reliable indicators to explain crisis incidence and conclude that the large accumulation of foreign exchange reserves has played an important role in reducing countries' vulnerability during the global financial crisis. The results obtained in this study are in line with the notion that the stock of foreign exchange reserves is significantly negatively related with our measure of crisis incidence. Finally, Goldman Sachs (2013) estimates the probability of a sudden stop in capital inflows across emerging market economies. They find that an increase of 25 basis points in the ratio between foreign exchange reserves and short-term external debt has the same impact as a one percentage point improvement in the current account balance (as a percentage of GDP).

3. METHODOLOGY

We build two competing early warning systems (EWS) and compare their ability to correctly predict in-sample and out-of-sample episodes of currency crises in EMs in the period between January 1995 and December 2012.[5] We proceed as follows. We build an exchange rate pressure index from which we derive a crisis variable (or crisis incidence) that identifies episodes of currency crisis in EMs. The crisis variable is binary, as it assumes the value of one if a currency crisis takes place within the next 24 months, and 0 otherwise.

Once defined the crisis variable, we proceed to construct the logit and probit EWS, where the crisis variable is regressed on a set of selected external vulnerability indicators of EMs. A crisis probability is then calculated with the coefficient estimates obtained from the regression. Afterwards, we describe how we select the optimal cut-off value for the estimated crisis probability.

Formally, we follow Bussiere and Fratzscher (2006) and assume that that there are N countries, $i = 1, 2, ..., N$, that we observe during T periods $t = 1, 2, ..., T$. For each country and month, we observe a forward-looking crisis variable Y_{it} that can assume as values only 0 (non-crisis) or 1 (crisis). To derive the crisis binary variable, we follow Kaminsky and others (1998) and build an exchange rate pressure index.[6] The exchange rate pressure index for country i at time t ($ERPI_{it}$) is defined as a weighted average between the monthly change in the nominal exchange rate and that in the stock of foreign exchange reserves:

$$ERPI_{it} = \frac{e_{it} - e_{it-1}}{e_{it-1}} - \left(\frac{s_{e_i}}{s_{fxr_i}}\right) \frac{fxr_{it} - fxr_{it-1}}{fxr_{it-1}} \qquad (1)$$

where e_{it} is the nominal exchange rate of country i's currency against the U.S. dollar at time t, and fxr_{it} is the stock of foreign exchange reserves held by country i at time t. Finally, s_{e_i} and s_{fxr_i} are the standard deviations of the nominal exchange rate and the stock of foreign exchange reserves, respectively.

---

[5] The emerging market economies that we consider are: Argentina, Brazil, Bulgaria, Chile, China, Colombia, Croatia, Czech Republic, Egypt, Hungary, India, Indonesia, Kazakhstan, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Romania, Russia, South Africa, Taiwan, Thailand, Turkey, Ukraine, Uruguay and Vietnam.

[6] For a discussion on exchange rate pressure indices, see Eichengreen and others (1995).

As a next step, we define a currency crisis hitting country i at time t, CC_{it}, as a binary variable that can assume either 1 (when the *ERPI* is above its mean by a number η of standard deviations) or 0 (otherwise):

$$CC_{it} = \begin{cases} 1 \ \text{if} \ ERPI_{it} \geq \overline{ERPI}_i + \eta \sigma_{ERPI_i} \\ 0 \ \text{otherwise} \end{cases} \qquad (2)$$

where $\overline{ERPI}_i$ denotes the (country–specific) mean of the exchange rate pressure index, σ_{ERPIi} denotes its standard deviation, which is multiplied by the weight η. Condition (2) states that a currency crisis is observed if the exchange rate pressures index of country i at time t is equal or larger than a country–specific threshold. The threshold is calculated as the sum between the mean of the (country–specific) exchange rate pressure index and the product between a coefficient η and the standard deviation of the (country–specific) exchange rate pressure index. In the EWS literature, η typically assumes the values of three.[7] The choice of values to assign to η determines the position of the exchange rate pressure index threshold: if $\eta=3$, condition (2) implies that the threshold of the exchange rate pressure index is higher than when $\eta=2$. Because of the higher threshold, the index identifies less crisis episodes compared to when $\eta=2$. The choice of η involves a trade-off. With a low crisis threshold, an early warning system may miss few crisis episodes but, at the same time, issue several false alarms. By contrast, with a high threshold, an early warning system may issue few false alarms, but may miss several crisis episodes. As an illustration, figure 1 plots the exchange rate pressure index for selected EMs, and the thresholds for the index when $\eta=2$ and when $\eta=3$.

Figure 1
Exchange Rate Pressure Index and Thresholds: January 1995–December 2012.
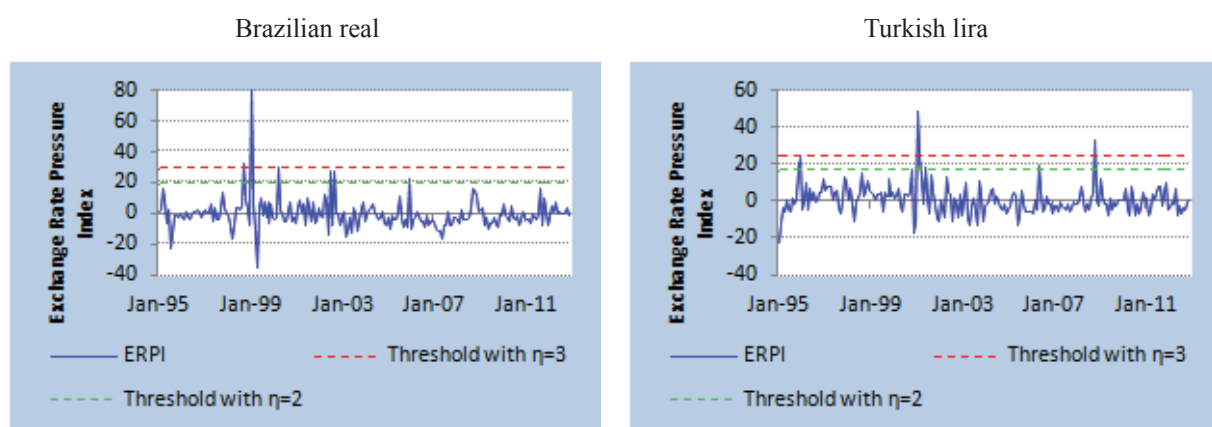
Source: International Financial Statistics and Author's calculations.

Unlike most of the EWS empirical literature, we let η assuming both values (two and three), and then compare how the EWS perform when η assumes one value or the other. Put differently, we are interested to see how the choice of the value to assign to η affects the EWS in-sample and out-of-sample performance. When $\eta=2$, the exchange rate pressure index identifies 191 crisis episodes across the set of emerging market economies, between January 1995 and December 2012. When $\eta=3$, the exchange rate pressure index threshold is higher and the index identifies only 77 crisis episodes in the panel.

Next, the variable CC_{it} is converted into the forward-looking crisis variable Y_{it} which is defined as follows

---

[7] See IMF (2002).

$$Y_{it} = \begin{cases} 1 & \text{if } \exists k = 1, 2, ..., 24 \text{ s.t. } CC_{it} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The forward-looking crisis variable Y_{it} is equal to 1 if within the next 24 months a currency crisis is observed in country i, and to 0 otherwise. As in Bussiere and Fratzscher (2006), the crisis definition adopted in this study allows capturing both successful and non-successful speculative attacks to a given currency. Finally, conditions (2) and (3) imply that the crisis variable Y_{it} will also depend on the choice of η. Since we allow η to assume the value of two or three, as a result we will have two crisis variables, one defined when $\eta=2$, and one when $\eta=3$.

We define $\Pr(Y_{it}=1)$ as the probability of country i to experience a currency crisis at time t. We estimate the probability of a currency crisis following two approaches. In the first approach, we estimate the probability of currency crisis using a fixed effects logit model. In the second approach we estimate the probability of currency crisis with a probit model. More formally, in each model the probability of a currency crisis is expressed as a non-linear function of a given set of explanatory variables X:

$$(\Pr Y_{it} = 1) = \Lambda(X'\beta) = \frac{e^{X'\beta}}{1 + e^{X'\beta}} \tag{4}$$

$$(\Pr Y_{it} = 1) = \Phi(X'\beta) = \int_{-\infty}^{X'\beta} j(z)dz \tag{5}$$

where $\Lambda(X'\beta)$ denotes the cumulative distribution function (cdf) of the logistic distribution, while $\Phi(X'\beta)$ denotes the cdf of the normal distribution. Conditions (4) and (5) express the conditional probabilities that country *i* experiences a currency crisis at time *t* as a function of selected external vulnerability indicators, denoted by X. The crisis binary variable Y_{it} is regressed on the external vulnerability indicators X in the period January 1995–December 2012, using logit (fixed effects) and probit estimation techniques.

The explanatory variables that we use in (4) and (5) are external vulnerability indicators. Following Goldman Sachs (2013), we select the following external vulnerability indicators: the ratio between the stocks of foreign exchange reserves and short-term external debt, the current account balance as a percentage of nominal GDP, the real GDP growth rate and a measure of the real effective exchange rate misalignment. In addition, we use the stocks of net foreign assets and credit to the private sector, both expressed as percentages of nominal GDP.

Once obtained the logit (fixed effects) and probit coefficient estimates, we derive the estimated probability of experiencing a currency crisis from conditions (4) and (5). Then, we choose a cut-off value for the estimated crisis probability in order to assess the performance of the logit and probit EWS. The cut-off value is chosen such that the total misclassification error (TME) is minimized. The TME is calculated as the sum between the percentage of missed crisis episodes (expressed as the ratio between missed crisis calls over the total number of crisis called) and the percentage of false alarms (expressed as the ratio between false alarms over the total number of tranquil period called). Formally:

$$\text{TME} = \text{Type 1 error} + \text{Type 2 error} \tag{6}$$

Where

$$\text{Type 1 error} = \frac{\text{Total missed crisis episodes}}{\text{Total crisis episodes}} \tag{7}$$

$$\text{Type 2 error} = \frac{\text{Total false alarms}}{\text{Total non-crisis episodes}} \tag{8}$$

Finally, note that in (6), type 1 and type 2 errors are equally weighted. In Comelli (2013), type 1 and type 2 errors are allowed to assume different weights in the TME. Changing the weights of type 1 and type 2 errors affects the EWS in-sample and out-of-sample performances.

4. RESULTS

We use a panel containing monthly observations of external vulnerability indicators for 29 emerging market economies, for the period included between January 1995 and December 2012.[8] The indicators are the ratio between the stocks of foreign exchange reserves and short-term external debt, the current account balance as a percentage of nominal GDP, the real GDP growth rate, a measure of real effective exchange rate misalignment, the stocks of net foreign assets and credit to the private sector, both expressed as a percentages of nominal GDP. The dependent variable is the crisis binary variable, which is regressed on the lagged indicators using logit (fixed effects) and probit estimation techniques. Since we allow the coefficient η in the exchange rate pressure index (section III) to assume either the value of two or three, we have two different dependent variables: one where the crisis variable is defined with $\eta=2$, and another where the crisis variable is defined with $\eta=3$. Tables 1 and 2 report the coefficient estimates obtained with logit and probit panel regressions. In each table, panel A reports the estimates obtained when $\eta=2$, while panel B reports the estimates when $\eta=3$.

For each estimation technique, seven different EWS specifications have been estimated.[9] The coefficient estimates across the seven EWS specifications tend to have the correct sign. Stronger real GDP growth rates and higher net foreign assets as a percentage of nominal GDP tend to reduce significantly the probability of experiencing a currency crisis episode. By contrast, in most of the specifications a higher stock of credit to the private sector as percentage of nominal GDP is significantly associated to a higher currency crisis incidence. The ratio between foreign exchange reserves and short–term external debt is correctly signed but is statistically significant only when $\eta=2$, in which case there are 191 crisis episodes identified in the panel. Conversely, the ratio loses significance when $\eta=3$, in which case there are only 77 crisis episodes identified in the panel. Put differently, the statistical significance of the ratio between foreign exchange reserves and short–term external debt depends on the choice of η, hence on the number of crisis episodes identified. The current account balance as a percentage of nominal GDP has the correct sign – e.g. a higher current account balance is associated with a decline in crisis incidence – but the estimates are not significant. Similarly, the measure of real exchange rate misalignment (deviations from the 3-year moving average) has the correct sign – a systematically overvalued currency raises the likelihood of experiencing a currency crisis – but the estimates are not always significant.[10]

---

[8]　See annex for a description of the variables.

[9]　We also run regressions for the seven EWS specifications using a logit random effects estimator. For some (but not all) of the EWS specifications, the Hausman Specification Test leads to reject the null hypothesis according to which the logit random effects estimator is efficient.

[10]　As an alternative measure of real exchange rate misalignment, in the logit and probit regressions we tried to include among the explanatory variables the real exchange rate deviations from a deterministic time trend. The coefficient estimates turned out not to be significant across the logit and probit specifications.

Table 1
Logit Fixed Effects Regression: Coefficient Estimates

	Panel A: $\eta=2$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
FXR/STED	−0.006*** (0.002)	−	−0.006*** (0.002)	−0.006*** (0.002)	−0.003*** (0.002)	−0.007*** (0.002)	−0.006*** (0.001)
CAB/Y	−0.029 (0.037)	−0.057 (0.036)	−	−0.011 (0.034)	−0.022 (0.036)	−0.043 (0.035)	−0.057 (0.035)
ΔY	−0.107** (0.040)	−0.098** (0.040)	−0.098** (0.038)	−	−0.118*** (0.039)	−0.136*** (0.036)	−0.136*** (0.037)
REERM	0.002 (0.002)	0.003** (0.001)	0.002 (0.001)	0.003** (0.001)	−	0.002 (0.001)	0.001 (0.001)
NFA/Y	−0.028 (0.020)	−0.057*** (0.018)	−0.028 (0.020)	−0.051*** (0.018)	−0.029 (0.020)	−	−0.027 (0.019)
PRCR/Y	0.022** (0.009)	0.026*** (0.009)	0.023** (0.009)	0.031*** (0.009)	0.018* (0.009)	0.014* (0.008)	−
Observations	4029	4029	4029	4029	4029	4255	4110
ROC Statistics	0.668	0.624	0.638	0.638	0.677	0.665	0.691
	Panel B: $\eta=3$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
FXR/STED	−0.003 (0.003)	−	−0.004 (0.003)	−0.003 (0.003)	−0.004 (0.003)	−0.008** (0.003)	−0.005 (0.003)
CAB/Y	−0.077 (0.061)	−0.089* (0.060)	−	−0.036 (0.053)	−0.067 (0.058)	−0.096* (0.057)	−0.118** (0.058)
ΔY	−0.124* (0.068)	−0.121* (0.067)	−0.093 (0.062)	−	−0.141** (0.067)	−0.188*** (0.061)	−0.192*** (0.062)
REERM	0.004* (0.002)	0.004* (0.002)	0.003 (0.002)	0.005** (0.002)	−	0.002 (0.002)	0.001 (0.002)
NFA/Y	−0.067* (0.036)	−0.082** (0.033)	−0.061* (0.035)	−0.094*** (0.032)	−0.066* (0.004)	−	−0.048 (0.003)
PRCR/Y	0.046** (0.020)	0.049** (0.019)	0.051*** (0.019)	0.059*** (0.019)	0.035** (0.017)	0.012 (0.013)	−
Observations	2961	2961	2961	2961	2961	3176	3042
ROC Statistics	0.649	0.616	0.615	0.596	0.699	0.723	0.753

***: Significant at 1%, **: Significant at 5%, *: Significant at 10%.

Sources: Joint External Debt Hub (www.jedh.org), Author's calculations based on International Financial Statistics.

Table 2
Probit Regression: Coefficient Estimates

	Panel A: $\eta=2$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
FXR/STED	−0.001** (0.000)	−	−0.001*** (0.000)	−0.001*** (0.000)	−0.001** (0.000)	−0.001*** (0.000)	−0.001** (0.000)
CAB/Y	−0.010 (0.012)	−0.019* (0.011)	−	−0.005 (0.011)	−0.007 (0.012)	−0.012 (0.011)	−0.012 (0.011)
ΔY	−0.035** (0.015)	−0.040*** (0.014)	−0.033** (0.014)	−	−0.040*** (0.014)	−0.045*** (0.013)	−0.043*** (0.014)
REERM	0.002** (0.001)	0.002** (0.000)	0.002* (0.000)	0.002*** (0.000)	−	0.002* (0.001)	0.001 (0.000)
NFA/Y	−0.012* (0.006)	−0.017*** (0.006)	−0.013** (0.006)	−0.017*** (0.006)	−0.012* (0.006)	−	−0.007 (0.006)
PRCR/Y	0.008*** (0.002)	0.007*** (0.002)	0.008*** (0.002)	0.009*** (0.002)	0.007*** (0.002)	0.006*** (0.002)	−
C	−1.835	−1.985	−1.798	−1.911	−1.750	−1.849	−1.483
Observations	4161	4161	4161	4161	4161	4387	4242
ROC Statistics	0.659	0.634	0.655	0.654	0.664	0.640	0.684
	Panel B: $\eta=3$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
FXR/STED	−0.001 (0.001)	−	−0.001* (0.000)	−0.001* (0.000)	−0.001 (0.001)	−0.002** (0.001)	−0.001 (0.001)
CAB/Y	−0.024 (0.015)	−0.031** (0.014)	−	−0.017 (0.015)	−0.023 (0.015)	−0.022 (0.014)	−0.028* (0.016)
ΔY	−0.037** (0.019)	−0.044** (0.017)	−0.031* (0.018)	−	−0.044** (0.008)	−0.044** (0.017)	−0.043** (0.019)
REERM	0.002* (0.001)	0.002* (0.001)	0.002 (0.001)	0.002** (0.001)	−	0.001 (0.001)	0.001 (0.001)
NFA/Y	−0.017** (0.008)	−0.020*** (0.008)	−0.017** (0.008)	−0.021*** (0.007)	−0.015* (0.008)	−	−0.010 (0.007)
PRCR/Y	0.008*** (0.002)	0.007** (0.003)	0.007*** (0.002)	0.008*** (0.002)	0.007** (0.003)	0.005 (0.002)	−
C	−2.316	−2.421	−2.209	−2.341	−2.247	−2.231	−2.021
Observations	4161	4161	4161	4161	4161	4387	4242
ROC Statistics	0.716	0.694	0.713	0.713	0.719	0.698	0.748

***: Significant at 1%, **: Significant at 5%, *: Significant at 10%.

Sources: Joint External Debt Hub (www.jedh.org), Author's calculations based on International Financial Statistics.
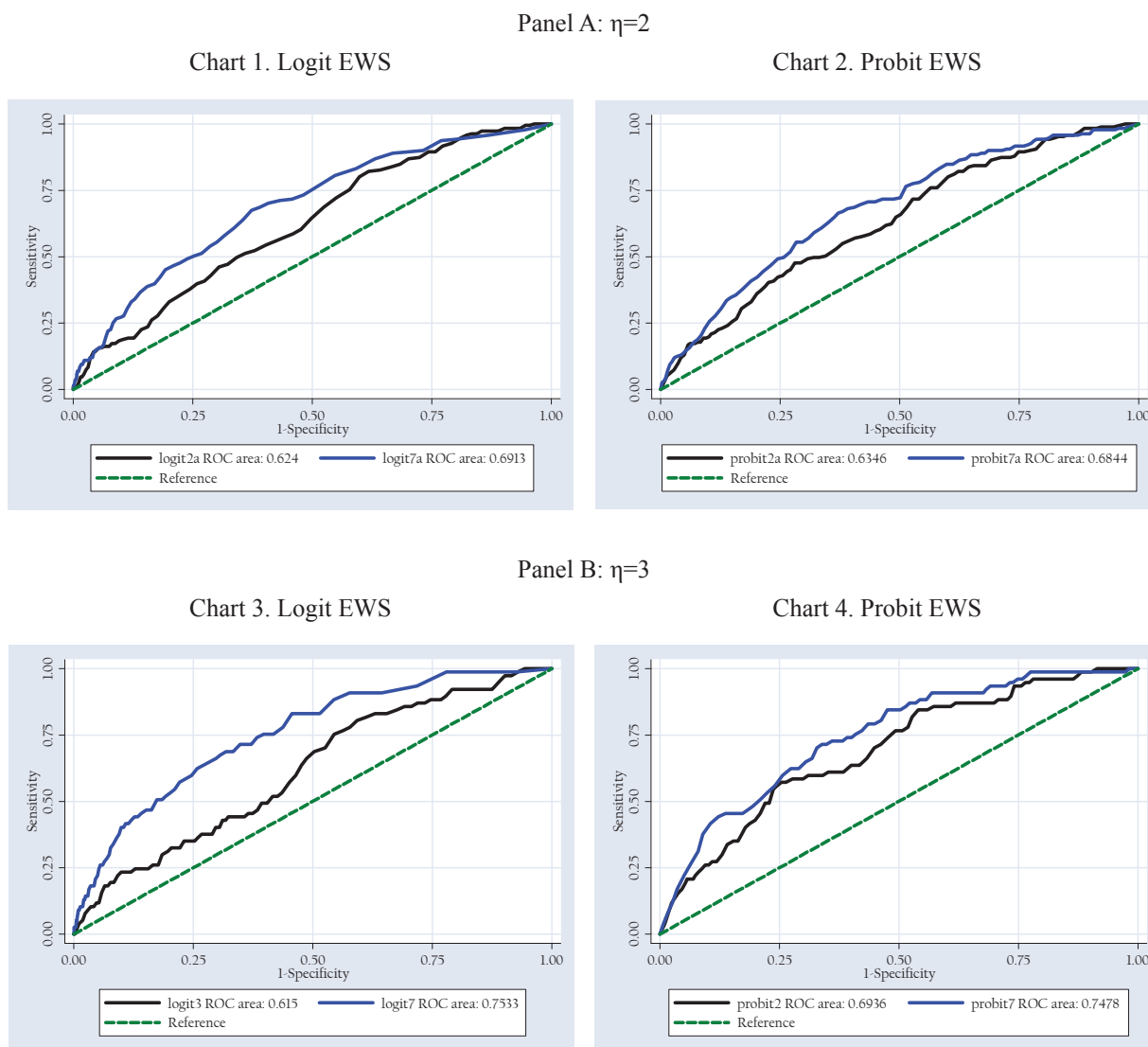
At this stage, following Minoiu and others (2013), for each estimation technique employed, we select that EWS specification whose Receiver Operating Characteristic (ROC) statistics is the largest. ROC analysis provides a quantitative measure of the accuracy of diagnostic tests to discriminate between two states or conditions (e.g. crisis and non-crisis).[11] For our analysis we use the ROC curve, which depicts the relationship between the fractions of positive cases correctly

---

[11]   See STATA Press (2013).

classified (true positives) and that of positive cases incorrectly classified (false positives), for a range of probability thresholds (see figure 1). The fraction of positive cases that are correctly identified (the true-positive rate) is also called *sensitivity*, while *specificity* is the true-negative rate. The measure *1–specificity* is also called the false-positive rate. Therefore, for every cut-off value the ROC curve measures the trade-off between the true positive rate (sensitivity) and the false-positive rate (1–specificity).

The ROC curve is interpreted as follows. If the curve lies above the 45-degree line, for every cut-off value the true-positive rate is higher than the false-positive rate and the model generates crisis predictions that are superior to random guessing. By contrast, along the 45-degree line, sensitivity is equal to (1–specificity), meaning that for every cut-off value, the true-positive rate is exactly equal to the false-positive rate. The larger the ROC statistics, the most accurate is the diagnostic test to discriminate between crises and non-crises. For each estimation technique, the ROC curves of the logit and probit EWS specifications having the lowest and highest ROC statistics are plotted in figure 1. The ROC curves in panel A have been obtained when $\eta=2$, while those in panel B have been generated when $\eta=3$.

Figure 2
Areas Under Receiver Operating Characteristic (ROC) Curves

Panel A: η=2

Chart 1. Logit EWS — logit2a ROC area: 0.624; logit7a ROC area: 0.6913; Reference

Chart 2. Probit EWS — probit2a ROC area: 0.6346; probit7a ROC area: 0.6844; Reference

Panel B: η=3

Chart 3. Logit EWS — logit3 ROC area: 0.615; logit7 ROC area: 0.7533; Reference

Chart 4. Probit EWS — probit2 ROC area: 0.6936; probit7 ROC area: 0.7478; Reference

Sources: Author's calculations based on International Financial Statistics.

We select those logit and probit specifications having the largest ROC statistics. For each estimation technique and for each definition of the crisis variable used, specification (7) is the one with the largest ROC statistics. On this basis, we select the logit and probit EWS with specification (7) and assess their ability in correctly predicting in-sample and out-of-sample crisis and non-crisis episodes.[12]

5. THE PERFORMANCE OF THE SELECTED EARLY WARNING SYSTEM

We measure the performance of the two competing logit and probit EWS by looking primarily at their out-of-sample total misspecification error (TME). For each EWS, tables 3 and 4 report the following measures: the TME, the percentages of crisis episodes and tranquil periods correctly called, the percentages of missed crisis episodes and false alarms, the probability of observing a crisis within the next 24 months given a crisis alarm, the probability of observing a crisis if no prior alarm has been issued and the estimated probability cut-off values. As usual, in each table there are two panels. Panel A reports the EWS in-sample and out-of-sample performance results obtained when $\eta=2$, while panel B reports the results when $\eta=3$.

To assess the EWS in-sample performance, we proceed as follows. We restrict the original sample period (January 1995– December 2012) to the period January 1995–December 2006, which is the new estimation sample. For each of the two competing EWS, we obtain coefficient estimates, derive the probability of experiencing a currency crisis and calculate the TME. We begin by considering the estimation sample corresponding to the period January 1995–December 2006. Then, we gradually extend the estimation sample by one year at a time, for the following four years. We stop the procedure when the estimation sample corresponds to the period January 1995–December 2010. Then, we look at the one-year-ahead out-of-sample performance of each of the two competing EWS, as we are primarily interested to assess how the EWS perform in the first year outside the estimation sample.

The results in tables 3 and 4 show that the logit and probit EWS out-of-sample performances deteriorate compared to their in-sample performances, as reflected by the higher out-of-sample TME scores and the higher percentages of missed crisis episodes.[13] Both EWS tend to perform out-of-sample reasonably well until 2008 as indicated by the TME scores. From 2009 onwards, however, the TME scores of both EWS rise considerably, as well as the probability of experiencing a currency crisis given a crisis alarm. Summing up, including 2008 in the estimation sample – the year with most disruptive episodes of market turbulence observed during the global financial crisis – considerably deteriorates the EWS out-of-sample performance.

The data in table 3 suggest that when $\eta=2$ the logit EWS TME out-of-sample scores vary between 68 and 111, meaning that the logit EWS is able to classify correctly between 44% and 66% of the total out-of-sample observations. When $\eta=3$ – hence when less crisis episodes (77) are identified in the panel – the logit EWS out-of-sample performance deteriorates, as the TME out-of-sample scores vary between 83 and 116, meaning that the EWS correctly classifies only between 42% and 58% of the total out-of-sample observations.

---

[12] We estimated the EWS specification (7) with a logit random effects estimator. The Hausman Specification Test rejects the null hypothesis of random effects efficiency for the EWS specification (7). Therefore, we limit ourselves to compare the EWS specification (7) estimated with logit fixed effects and probit estimating techniques.

[13] See also charts in the Annex.

Table 3
Logit EWS: In-sample and Out-of-sample Performances

Sample Windows	Prob. cut-off value	Crisis Episodes Correctly Called (%)	Non-Crisis Episodes Correctly Called (%)	Missed Crisis Episodes (%)	False Alarms (%)	Crisis Prob. Given Alarm	Crisis Prob. Given No Alarm	TME
Panel A: η=2								
In-sample								
1995–06	0.06	83.3	53.5	16.7	46.5	0.26	0.06	63
1995–07	0.08	73.4	61.0	26.6	39.0	0.25	0.07	66
1995–08	0.07	63.4	61.4	36.6	38.6	0.29	0.13	75
1995–09	0.11	60.9	62.7	39.1	37.3	0.28	0.13	76
1995–10	0.09	60.8	59.5	33.2	40.5	0.27	0.11	74
Out-of-sample								
2007	0.06	44.8	86.8	55.2	13.2	0.74	0.35	68
2008	0.08	52.1	78.1	47.9	21.9	0.63	0.30	70
2009	0.07	30.0	58.9	70.0	41.1	0.04	0.07	111
2010	0.11	51.6	77.4	48.4	22.6	0.19	0.06	71
2011	0.09	47.8	76.0	52.2	24.0	0.13	0.05	76
Panel B: η=3								
In-sample								
1995–06	0.06	66.5	65.9	33.5	34.1	0.28	0.09	68
1995–07	0.05	66.3	68.3	33.7	31.7	0.27	0.08	65
1995–08	0.08	46.7	73.2	53.3	26.8	0.30	0.15	80
1995–09	0.18	36.6	85.7	63.4	14.3	0.38	0.15	78
1995–10	0.24	35.1	86.9	64.9	13.1	0.37	0.14	78
Out-of-sample								
2007	0.06	23.4	86.8	76.6	13.2	0.60	0.43	90
2008	0.05	35.0	81.6	65.0	18.4	0.58	0.36	83
2009	0.08	30.0	63.6	70.0	36.4	0.05	0.07	106
2010	0.18	0.0	83.6	100.0	16.4	0.00	0.11	116
2011	0.24	0.0	95.5	100.0	4.5	0.00	0.07	105

Sources: Joint External Debt Hub (www.jedh.org), International Financial Statistics and Author's calculations.

Table 4 reports the in-sample and out-of-sample performance results of the probit EWS. When η=2, the TME out-of-sample scores vary between 72 and 97, meaning that the EWS correctly classifies between 51% and 64% of the total out-of-sample observations. By contrast, when η=3, the probit EWS out-of-sample performance deteriorates, as the out-sample TME scores vary between 87 and 117. This implies that the EWS correctly classifies only between 41% and 53% of the total out-of-sample observations.

Overall, the logit and probit EWS out-of-sample performances are broadly similar. When η=2, no clear hierarchy emerges among the two EWS: the logit EWS out-of-sample TME scores vary between 68 and 111, while the probit out-of-sample TME scores vary between 72 and 97. When η=3, the logit EWS performs slightly better than the probit EWS, as the logit EWS out-of-sample TME scores are included between 83 and 116, while the probit EWS out-of-sample TME scores are included between 87 and 117. In addition, the results show that the EWS performance is sensitive to the crisis variable definition. Specifically, both EWS perform better (e.g. the TME scores are lower)

when the ERPI is defined by setting $\eta=2$ – when more currency crisis episodes are identified in the panel – then when setting $\eta=3$ – when less currency crisis episodes are identified in the panel.

Finally, the results show that crisis alarms issued out-of-sample are not always reliable, as the conditional probability of observing a crisis given an alarm declines considerably once that the estimation sample includes the year 2008. From a macroeconomic policy perspective, these results offer two implications. First, these results underscore the importance that similar early warning exercises having the goal to estimate the likelihood of experiencing financial crises should be run at least once every year. This is motivated by observing that the probability of experiencing a currency crisis may crucially depend on new incoming economic and financial data. Second, the results confirm that while running an EWS today can help identifying past crisis episodes with some accuracy (in-sample performance), predicting crisis episodes outside the estimation sample is much more challenging because of the presence of uncertainty, as the available information set outside the estimation sample is much more limited than within the estimation sample.

Table 4
Probit EWS: In-sample and Out-of-sample Performances

Sample Window	Prob. cut-off value	Crisis Episodes Correctly Called (%)	Non-Crisis Episodes Correctly Called (%)	Missed Crisis Episodes (%)	False Alarms (%)	Crisis Prob. Given Alarm	Crisis Prob. Given No Alarm	TME
Panel A: $\eta=2$								
In-sample								
1995–06	0.51	92.1	51.5	7.9	48.5	0.45	0.06	56
1995–07	0.48	86.9	53.2	13.1	46.8	0.43	0.09	60
1995–08	0.62	53.4	68.7	46.6	31.3	0.49	0.27	78
1995–09	0.52	67.8	54.9	32.2	45.1	0.45	0.24	77
1995–10	0.63	57.3	67.3	42.7	32.7	0.48	0.25	75
Out-of-sample								
2007	0.51	31.5	81.8	68.5	18.2	0.83	0.70	87
2008	0.48	49.8	72.9	50.2	27.1	0.80	0.60	77
2009	0.62	47.6	64.3	52.4	35.7	0.31	0.21	88
2010	0.52	45.8	56.5	54.2	43.5	0.37	0.35	98
2011	0.63	39.6	88.5	60.4	11.5	0.60	0.23	72
Panel B: $\eta=3$								
In-sample								
1995–06	0.75	67.9	63.7	32.1	36.3	0.27	0.09	68
1995–07	0.63	69.3	65.7	30.7	34.3	0.27	0.08	65
1995–08	0.63	55.0	66.2	45.0	33.8	0.29	0.14	79
1995–09	0.74	48.8	74.1	51.2	25.9	0.31	0.14	77
1995–10	0.80	39.1	83.0	60.9	17.0	0.34	0.14	78
Out-of-sample								
2007	0.75	9.7	96.7	90.3	3.3	0.71	0.44	94
2008	0.63	30.7	82.1	69.3	17.9	0.55	0.38	87
2009	0.63	30.0	58.2	70.0	41.8	0.04	0.07	112
2010	0.74	3.2	79.3	96.8	20.7	0.02	0.11	117
2011	0.80	0.0	95.8	100.0	4.2	0.00	0.07	104

Sources: Joint External Debt Hub (www.jedh.org), International Financial Statistics and Author's calculations.

6. CONCLUSIONS

In this study we compared the performance of logit (fixed effects) and probit EWS in correctly predicting in-sample and out-of sample currency crisis in selected emerging market economies.

We found that stronger real GDP growth rates and higher net foreign assets significantly reduce the probability of experiencing a currency crisis, while high credit to the private sector increases it. By contrast, the current account balance and the measure of real exchange rate misalignment are not always statistically significant. The ratio between foreign exchange reserves and short–term external debt has the correct sign but it is significant only when the currency crisis is defined as a situation where the (country–specific) exchange rate pressure index is two standard deviations above its mean. The ratio is no longer significant if the exchange rate pressure index is three standard deviations above its mean.

The logit and probit EWS out-of-sample performances are broadly similar. The logit EWS is able to classify correctly between 42% and 66% of the total out-of-sample observations (e.g. crisis and tranquil periods), while the probit EWS s able to classify correctly between 41% and 64% of the total out-of-sample observations. We also find that the EWS performance is sensitive to the size of the estimation sample, and to the crisis definition used. In particular, both EWS perform better when a crisis episode is defined as a situation when the (country–specific) exchange market pressure index is two standard deviations above its mean. All in all, the results show that the model's ability to produce reliable out-of-forecasts currency crises prediction is limited. However, the model can be a useful framework to identify which vulnerabilities may significantly affect the probability of experiencing a currency crisis.

The results offer two macroeconomic policy conclusions. First, the EWS out-of-sample performance can be very sensitive to the size of the estimation sample. Specifically, the EWS total misclassification error and the probability of experiencing a currency given a crisis alarm can vary considerably if a particular year with many outlying observations is included in the estimation sample. Second, the results imply that selecting a crisis definition as a situation when the exchange rate pressure index is two standard deviations above its average value reduces the EWS total misclassification error. Therefore, the results obtained in this study suggest that a crisis definition identifying more rather than less currency crisis episodes should be employed when setting up an EWS model, even if this may lead to the risk of issuing several false alarms.

Finally, the analysis in this study can be extended in a number of ways. First, the two EWS employed in this study rely mainly on the information conveyed by standard macroeconomic and external vulnerability indicators. It would be interesting to assess if and how the EWS out-of-sample performance changes if indicators quantifying cross–country contagion, spillover effects or cross-border financial linkages were included in the EWS. Second, it would be interesting to modify the analysis in this study to estimate the probability of sudden stops in capital inflows in emerging market economies, and to check the model out–of–sample performance. Thirdly, as regards the definition of currency crisis adopted in this paper, η (the coefficient that multiplies the standard deviation of the exchange rate pressure index in the definition of currency crisis) was chosen arbitrarily. As an extension, it would be interesting to find out what is the value that η should assume in order to minimize the total misclassification error of the EWS.

References

Abiad, A. (2003), Early Warning Systems: A Survey and a Regime-Switching Approach. IMF Working Paper, No. WP/03/32, (Washington: International Monetary Fund).
Beckmann, D., Menkhoff, L., Sawischlewski, K. (2006), Robust Lessons About Practical Early Warning Systems. *Journal of Policy Modeling*, Vol. 28, pp. 163–193.

Berg, A, Pattillo, C. (1999), Predicting Currency Crises: The Indicators Approach and an Alternative. *Journal of International Money and Finance*, Vol. 18, pp. 561–586.

Berkmen, S.P., Gelos, G., Rennhack, R., Walsh, J. (2012), The global financial crisis: Explaining cross-country differences in the output impact. *Journal of International Money and Finance*, Vol. 31, pp. 42–59.

Blanchard, O., Das, M., Faruqee, H. (2010), The Initial Impact of the Crisis on Emerging Market Countries. *Brooking Papers on Economic Activity*, Spring 2010, pp. 263–323.

Borio, C., Drehmann, M. (2008), Assessing the Risk of Banking Crises – Revised. *BIS Quarterly Review*, March 2009, pp. 29–46 (Basel: Bank for International Settlements).

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

Candelon, B., Dumitrescu, E., Hurlin, C. (2012), How to Evaluate an Early-Warning System: Toward a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods. *IMF Economic Review*, Vol. 60, No.1 (Washington: International Monetary Fund).

Cerra, V., Saxena, S.C. (2008), Growth Dynamics: The Myth of Economic Recovery. *American Economic Review*, Vol. 98, No.1 (March 2008), pp. 439–457.

Comelli, F. (2013), Comparing Parametric and Non-parametric Early Warning Systems For Currency Crises in Emerging Market Economies. IMF Working Paper, No. WP/13/134, (Washington: International Monetary Fund).

Eichengreen, B., Rose, A.K., Wyplosz, C. (1995), Exchange Market Mayhem. *Economic Policy*, Vol. 10, No. 21, pp. 249–312.

Frank, N., Hesse, H. (2009) Financial Spillovers to Emerging Markets during the Global Financial Crisis. IMF Working Paper, No. WP/09/104 (Washington: International Monetary Fund).

Frankel, J., Saravelos, G. (2012), Can Leading Indicators Assess Country Vulnerability? Evidence From the 2008-09 Global Financial Crisis. *Journal of International Economics*, Vol. 87, pp. 216–231.

Goldman Sachs (2013), 'Sudden–stops' in Capital Inflows and The Role of FX Reserves. *CEEMEA Economics Analyst*, Issue No: 13/25.

Gourinchas, P.-O., Obstfeld, M. (2011), Stories of The Twentieth Century For The Twenty-first. NBER Working Paper No. 17252, (Cambridge, Massachusetts: National Bureau of Economic Research).

International Monetary Fund (2002), Global Financial Stability Report. March 2002, Chapter 4 (Washington: International Monetary Fund).

Kaminsky, G. (1998), Currency and Banking Crises: The Early Warning of Distress. International Finance Discussion Paper No. 629, (Washington: Board of Governors of the Federal Reserve System).

Kaminsky, G., Lizondo, S., Reinhart, C. (1998), Leading Indicators for Currency Crisis. *IMF Staff Papers*, Palgrave Macmillan Journals, 45(1).

Kaminsky, G., Reinhart, C. (1999), The Twin Crises: The Causes of Banking and Balance-of-Payments Problems. *American Economic Review*, Vol. 89, No. 3, pp. 473–500.

Llaudes, R., Salman, F., Chivakul, M. (2010), The Impact of the Great Recession on Emerging Markets. *International Monetary Fund Working Paper*, No. WP/10/237 (Washington: International Monetary Fund).

Milesi-Ferretti, G., Tille, C. (2011), The great retrenchment: international capital flows during the global financial crisis. *Economic Policy*, Vol. 26, No. 66, pp. 289–346.

Minoiu, C., Kang, C., Subrahmanian, V.S., Berea, A. (2013), Does Financial Connectedness Predict Crises? *IMF Working Paper*, No. WP/13/267, (Washington: International Monetary Fund).

Rose, A., Spiegel, M. (2012), Cross-country Causes and Consequences of the 2008 Crisis: Early Warning. *Japan and the World Economy*, Vol. 24, pp. 1–16.

Stata Press (2013), Stata Base Reference Manual, College Station, TX: StataCorp LP.

Wooldridge, J.M. (2002), Econometric Analysis of Cross Section and Panel Data, Cambdrige, MA: MIT Press.

ANNEX

A1. COUNTRY LIST

Argentina, Brazil, Bulgaria, Chile, China, Colombia, Croatia, Czech Republic, Egypt, Hungary, India, Indonesia, Kazakhstan, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Romania, Russia, South Africa, Taiwan, Thailand, Turkey, Ukraine, Uruguay and Vietnam.

A2. DESCRIPTION OF THE VARIABLES

(1) Ratio between foreign exchange reserves and short term external debt, FXR/STED: This is calculated as the ratio between the stocks of foreign exchange reserves and short-term external debt (i.e. maturing within one year). Both numerator and denominator are expressed in U.S. dollars. It is a reserve adequacy ratio which is often used in early warning exercises. Quarterly data have been interpolated in order to have monthly time series. Source: Joint External Debt Hub (www.jedh.org).

(2) Current account balance as a percentage of GDP, CAB/Y. Ratio between the current account balance and nominal GDP. Both numerator and denominator are expressed in U.S. dollars. Annual data have been interpolated in order to have monthly time series. Source: World Economic Outlook Database, International Monetary Fund.

(3) Real GDP growth, ΔY: Annual percentage change in real GDP. Annual data have been interpolated in order to have monthly time series for real GDP growth. Source: World Economic Outlook Database, International Monetary Fund.

(4) Real effective exchange rate misalignment, REERM. The series has been obtained by taking the real effective exchange rate (REER) deviation from the three-year moving average. Monthly data. Source: International Financial Statistics, International Monetary Fund.

(5) Ratio between the stock of net foreign assets and nominal GDP, NFA/Y. Both numerator and denominator are expressed in U.S. dollars. Source: International Financial Statistics, International Monetary Fund.

(6) Ratio between private credit and nominal GDP, PRCR/Y. Available for most emerging economies only from January 2001 onwards. Source: International Financial Statistics, International Monetary Fund.

Figure 3

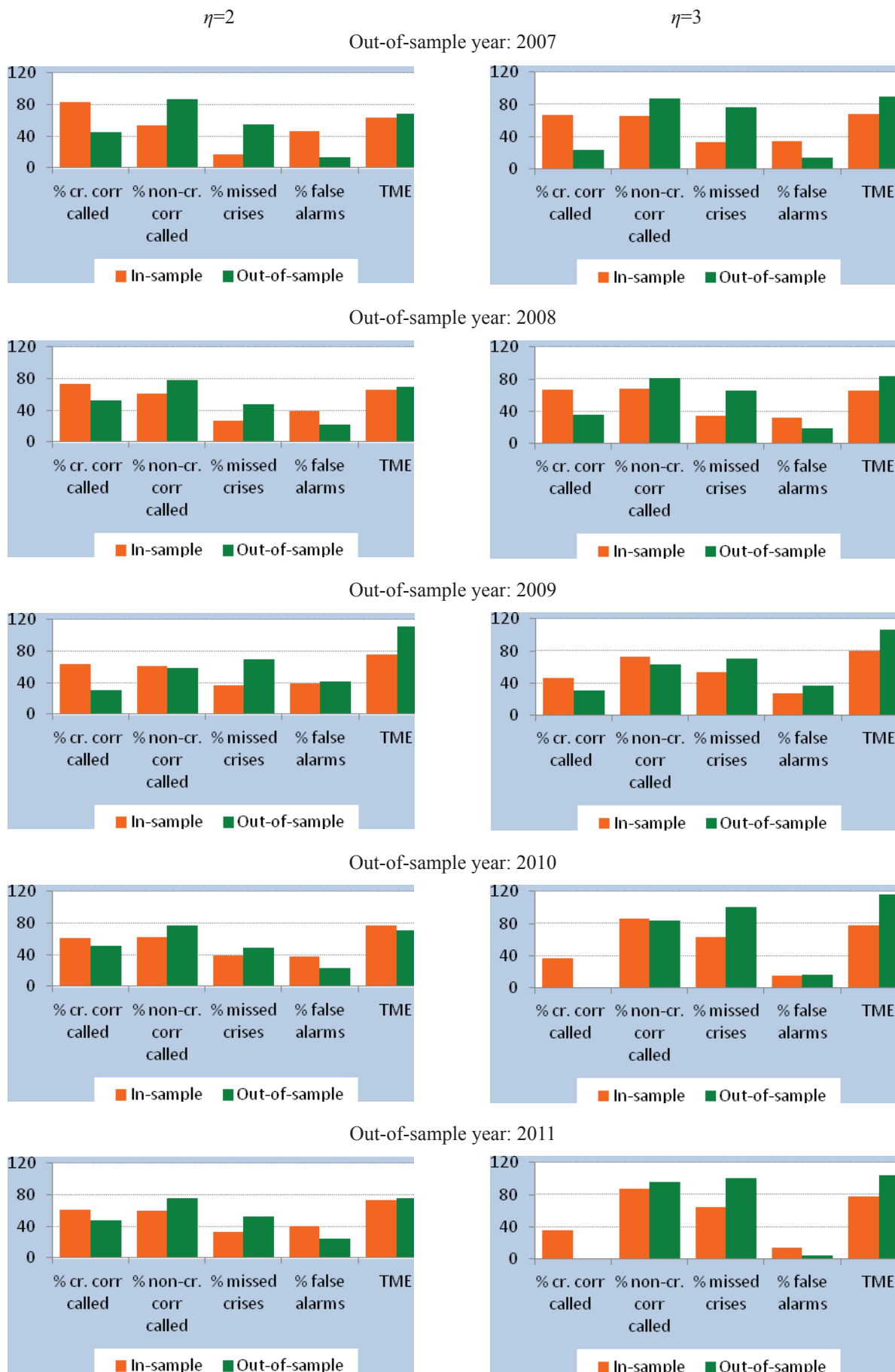Logit In-sample and Out-of-sample Performances *(Percentages)*

$\eta=2$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\eta=3$

Out-of-sample year: 2007

Out-of-sample year: 2008

Out-of-sample year: 2009

Out-of-sample year: 2010

Out-of-sample year: 2011

Figure 4
Probit In-sample and Out-of-sample Performances *(Percentages)*