Banco de México

Documentos de Investigación

Banco de México

Working Papers

N° 2017-16

Early Warning Systems with Real-Time Data

Tjeerd M. Boonman
Banco de México

Jan P.A.M. Jacobs
University of Groningen, University of Tasmania, CAMA and CIRANO

Gerard H. Kuper
University of Groningen

Alberto Romero
Banco de México

September 2017

Documento de Investigación
2017-16

Working Paper
2017-16

Early Warning Systems with Real-Time Data*

Tjeerd M. Boonman†
Banco de México

Jan P.A.M. Jacobs‡
University of Groningen, University of Tasmania,
CAMA and CIRANO

Gerard H. Kuper§
University of Groningen

Alberto Romero**
Banco de México

Abstract: This paper investigates the performance of early warning systems in real-time, using forecasts of indicators that were available at the moment predictions are to be made. The study analyzes currency crises in eight Latin American and Central and Eastern European countries, distinguishing an estimation period 1990-2009 and a prediction period 2010-2014. We apply two varieties of early warning systems: the signal approach and the logit models. For both methods we find that using forecasts of the indicators worsens the predictive ability of early warning systems compared to using the most recently available information (ex post).
Keywords: Real-time data, Early warning system, Signal approach, Logit model, Emerging economies
JEL Classification: E47, G01, F31, C23, E58.

Resumen: En este trabajo se investiga el desempeño de los sistemas de alerta temprana en tiempo real, utilizando los pronósticos de indicadores que se encontraban disponibles al momento de hacer las predicciones. El estudio analiza crisis cambiarias en ocho países de América Latina y de Europa Central y Oriental, separando el período de estimación, 1990-2009, del período de predicción, 2010-2014. Se aplicaron dos métodos de sistemas de alerta temprana: el método de señales y el de regresión logística. Con ambos métodos encontramos que el uso de pronósticos de los indicadores reduce la capacidad predictiva de los sistemas de alerta temprana en comparación con el uso de la información disponible más reciente (ex post).
Palabras Clave: Datos en tiempo real, sistemas de alerta temprana, método de señales, regresión logística, economías emergentes

*The opinions expressed here are those of the authors and do not necessarily represent Banco de México's or its board of governors' opinions. We thank Georgia Bush, Gabriel Cuadra, Mariela Dal Borgo, Juan Pablo Graf, Arturo Lamadrid, Fabrizio López-Gallo, Lorenzo Menna, participants at the EEFS 2017 annual meeting in Ljubljana, Slovenia, and seminar participants at the University of San Diego, San Diego State University, EGADE Guadalajara and Banco de México for their valuable comments and suggestions. We also thank Diego Cid and Stefano Lord for their excellent research assistance.

† Dirección General de Estabilidad Financiera. Email: tboonman@banxico.org.mx.

‡ University of Groningen, University of Tasmania, CAMA and CIRANO. Email: j.p.a.m.jacobs@rug.nl.

§ University of Groningen. Email: g.h.kuper@rug.nl.

** Dirección General de Estabilidad Financiera. Email: alberto.romero@banxico.org.mx.

1 Introduction

The high costs of financial crises for the public sector as well as private investors has led to a large number of empirical studies that intend to detect crises in a timely manner, through the use of Early Warning Systems (EWSs). EWSs are quantitative models that predict extreme events based on statistical information and mechanisms from the past.[1] EWSs for financial crises have been criticized for several reasons, one of which is that they are only useful for detecting past crises (Frankel and Saravelos, 2010).[2] An important reason is that the data used by forecasters to construct the EWS are not available in a timely manner. Another reason is that indicators are selected with the benefit of hindsight. EWSs are typically tested out-of-sample using the most recent data, known as current-vintage data. When the EWS detects crises using current-vintage data but fails to detect crises for the out-of-sample prediction period, then the EWS provides a false sense of security.

We focus on a particular type of financial crisis, the currency crisis, which is defined as a large, sudden depreciation or devaluation of the exchange rate, or an episode with high pressure on the exchange rate that may result in large losses of international reserves and/or a hike in domestic interest rates to defend the exchange rate (Berg, Borensztein and Patillo, 2005). We set up two types of early warning systems for currency crises, the signal approach and the logit model. We apply each EWS to a panel of four Latin American and four Central and Eastern European countries, in the period 1990–2014. We first estimate our EWSs on the in-sample period (1990–2009), and then predict currency crises for the out-of-sample period (2010–2014) with two types of real-time data, with and without employing information that is not available at the moment predictions have to be made.

[1] EWSs are applied to predict financial crises, but also for natural disasters such as tsunamis, earthquakes, droughts and epidemics (Choo, 2009).

[2] Other reasons are that the out-of-sample performance of EWSs has been rather poor—even for prominent papers (Berg and Patillo, 1999). A more fundamental problem with early warning signals is that these are meant to predict crises with the aim to avoid their occurrence; but if these signals are used for policy purposes, the predicted crises will be avoided, which means that model predictions will not be accurate any longer. Bussière (2013) refers to this contradiction as an "impossibility theorem". A second problem is similar in essence, but leads to the opposite conclusion: when Early Warning Systems predict a crisis this may lead to self-fulfilling prophecies by market participants.

Figure 1: Real-time data trapezoid

	Vintage years									
	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
2000	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
2001	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9
2002	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8
2003	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1	8.1
2004	*1.5*	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
2005	*0.4*	*0.4*	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
2006	*7.1*	*6.7*	*10.1*	*10.6*	10.9	10.9	10.9	10.9	10.9	10.9
2007	*6.8*	*8.3*	*10.1*	*10.7*	*10.4*	10.5	10.5	10.5	10.5	10.5
2008		*9.3*	*8.6*	*1.6*	*0.5*	*0.4*	0.1	0.1	0.1	0.1
2009			*9.2*	*4.3*	*8.1*	*8.5*	*9.1*	*10.4*	10.6	10.6
2010				*2.2*	*7.8*	*7.4*	*6.7*	*6.0*	5.1	5.1
2011					*6.6*	*6.1*	*4.9*	*3.5*	*2.9*	1.7
2012						*4.5*	*0.2*	*-2.3*	*-7.2*	*-6.9*
2013							*0.0*	*-1.8*	*-3.3*	*-4.2*
2014								*-1.3*	*-2.8*	*-3.0*

	Data set	
	1	**2**
2000	8.0	
2001	6.9	
2002	7.8	
2003	8.1	
2004	1.6	
2005	0.3	
2006	10.9	*7.1*
2007	10.5	*8.3*
2008	0.1	*8.6*
2009	10.6	*4.3*
2010	5.1	*7.8*
2011	1.7	*6.1*
2012	*-6.9*	*0.2*
2013	*-4.2*	*-1.8*
2014	*-3.0*	*-2.8*

Notes: Example of a real-time data trapezoid (left panel). Columns show vintages, and rows subsequent observations for years. Italicized numbers are subject to revisions. The cells with the darkest backgrounds represent first estimates, the cells with lighter backgrounds are second estimates, third, and fourth estimates of the data. Data that are not italicized represent data which are no longer revised. The right panel shows data concepts that we use in our analysis, current-vintage data (data set 1) and second estimates (data set 2).

In this paper, we investigate the performance of EWSs for currency crises with real-time data, which reflects the notion that data are subject to revision over the course of time. Real-time data are typically displayed in the form of a real-time data trapezoid. Figure 1 provides an example. Vintages are in columns and observations are in rows. As we move across columns, from left to right, later vintages are displayed and we observe how indicators are revised over time. To illustrate the real-time data trapezoid, let's take the values for the year 2012, which corresponds to the row 2012. The first estimate of the variable in 2012 is a forecast made one year earlier in 2011 (4.5). This forecast is adjusted in 2012 (0.2), while in 2013 the first official data observation is published -2.3, which is revised to -7.2 in 2014, followed by another data revision in 2015, which brings the value at -6.9.

We assess the sensitivity of EWS currency crisis predictions to the choice of two alternative data sets shown in Figure 1. As our first data set, we employ **current-vintage**

data which implies using information of the most recent vintage, as shown in the first column of the right panel of the figure. The second data set that we employ contains **second estimates** (Right panel, data set 2), for this we take forecasts for year t published in the beginning of year t. We find that using second estimates worsens the ability of EWSs in signalling crises compared to using currently available data. This is in line with the critique that EWSs perform well to predict past crises, but do poorly in predicting future crises.

To review the performance of alternative EWSs on forecasting currency crises in emerging economies in the late 1990s Berg, Borensztein and Patillo (2005) use only information that was available at the time, no information about actual outcomes was used in the forecasts. They use the internal IMF July 1999 forecasts and subsequent internal forecasts to compare the performance of alternative EWSs. Gunther and Moore (2003) compare the performance of an EWS for banking crises in the period 1996–1998 using first releases (first official publication of the data after the period has ended) versus current vintage data that was available in May 2000 and find that the EWS with current vintage data performs better than the EWS with first releases. Reagle and Salvatore (2005) test the robustness of predicting the Asian 1997–1998 financial crisis without and with data revisions. They estimate a probit model for a cross-section of a group of 54 emerging economies with three sets of data: the original, unrevised 1996 World Bank data, and the 1999 and 2004 updates of the 1996 World Bank data. They conclude that data revisions lead to significant changes in the model's estimates, which presents a problem for researchers that should be recognized and addressed.

The importance of using early estimates in EWSs has been acknowledged in the literature, but real-time data have not been implemented on a widespread scale. Frankel and Saravelos (2012) comment that predictions issued in real time would be impressive, but also especially difficult. Lo Duca and Peltonen (2013) remark that real-time data sets that contain information on the revisions of data after the first publication do not exist yet for several countries in their sample of developed and emerging economies. Data availability issues seem to be the major reason why implementation has not been widespread; early estimates for emerging economies are not publicly available.

Alessi and Detken (2011) use *quasi real-time* data. Although it sounds similar to real-time data, the authors actually mean current-vintage data. In their own words:

> The caveat is that we use the most recent vintage of data and not a true real time data set with unrevised data. Nevertheless, we use conservative lags to proxy for standard publication lags and thus real time data availability. [...] The performance of real variables could possibly be worse in a true real time setting compared to a quasi real time setting, as real variables can be heavily revised (i.e., the quality of current vintage data can be much better than the quality of real time vintages). [page 524]

Lo Duca and Peltonen (2013) also use quasi real-time data, in their case to predict financial stress events. Holopainen and Sarlin (2016) use quasi real-time data to predict banking crises. Their contribution to the crises EWS literature is to assume that the forecaster uses all available accounting and market-based information, with a lag. However, an important limitation still holds as data revisions may occur after the first estimate.

Our approach differs from Berg, Borensztein and Patillo (2005) in several ways. First, we explicitly focus on the difference between second estimates and current-vintage data when making out-of-sample predictions. Second, we use the consensus forecasts available in Haver Analytics, whereas they used internal predictions from the IMF, which are not publicly available. Third, we consider a longer period. As a consequence, our sample includes more and different crises, both from the 1990s (e.g. Mexico 1994-1995, Russia 1998, Brazil 1999) and the next decade (e.g. Argentina 2001–2002, and the Global Financial Crisis that hit most emerging economies in 2008–2009). Including more crises makes our analysis more complete, because of the variety in the crises (Kaminksy, 2006). Compared to Gunther and More (2003) and Reagle and Salvatore (2005) we use second estimates to predict currency crises whereas they use first releases to predict banking crises.

Our results can be useful for policymakers as it draws attention to the issue of data availability. Since current-vintage data is not available on time, a realistic out-of-sample evaluation of EWSs requires early estimates of the indicators. Our results show that the crisis prediction results are worse when using early estimates, which demonstrates the necessity to dedicate more efforts in producing forecasts of the indicators themselves.

The remainder of this paper is structured as follows. Section 2 reviews definitions of currency crises and EWSs for currency crises. Section 3 describes the two types of EWSs we employ. After the description of the data in Section 4, we present empirical results in Section 5. Section 6 concludes.

2 Currency Crises and Early Warning Systems

Currency crises not only occur in countries fixed exchange rate regimes, but also in countries with flexible exchange rates which, in principle, should be more resistant to currency crises. One would expect continuous market adjustment to limit the buildup of pressures leading to extreme currency overvaluation and subsequent large discrete currency declines as may occur under fixed exchange rate regimes.

Pegged and intermediate exchange rate regimes are indeed associated with greater susceptibility to currency crises, particularly in developing and emerging market countries with more open capital accounts (Ghosh, Ostry and Tsangarides, 2010). However, many countries with floating exchange rates have experienced currency crises. A possible explanation is the fact that countries reporting their currencies as on a floating rate regime are often quite reluctant to allow their currencies to float due to so-called fear of floating behavior (Calvo and Reinhart, 2000), and de facto follow a pegged exchange rate regime (Glick and Hutchison, 2011).

2.1 Currency Crisis Definition

How to identify currency crises has been debated since the mid 1990s. Two approaches can be distinguished: the successful attack approach and the speculative pressure approach. In the **successful attack approach** a currency crisis is identified when a currency depreciates significantly. Frankel and Rose (1996) identify a currency crisis when two conditions are met: (i) the depreciation of the nominal exchange rate of the currency is larger than 25% in a year, and (ii) the rate of the nominal depreciation must be 10 percentage points larger than it was in the previous year. Variations of this approach have emerged, with differences in the threshold and sample frequency.

The second approach, known as the **speculative pressure approach**, is inspired by Girton and Roper (1977) and later used by Eichengreen, Rose and Wyplosz (1995) and many others for currency crises. This approach does not only take into account actual devaluation or depreciation of the currency, but also includes periods of great stress of the exchange rate. The latter occurs when the monetary authorities avoid a devaluation or depreciation through the use of its international reserves or by increasing interest rates. Although the 'currency attack' was unsuccessful, one may argue that periods of exchange rate pressure should be considered a crisis. The measure for speculative pressure is an index of the weighted average of changes in the exchange rate, reserves and interest rate. The index is known as the Exchange Market Pressure Index (EMPI). Many variations have been proposed, with the adjusted definition from Kaminsky, Lizondo and Reinhart (1998) being the most common for emerging economies. The index consists of the weighted average of monthly changes in the nominal exchange rate versus the US dollar and the monthly percentage change of the international reserves, measured in US dollars. In contrast to Eichengreen, Rose and Wyplosz (1995), they do not include the interest rate in the definition. Their argument is that in emerging economies interest rate spreads are not always available or useful. Hawkins and Klau (2000) comment that there are periods where interest rates were controlled rather than market-determined. However, the omission of interest rates in the EMPI is also recognized as a shortcoming in identifying crises (Berg, Borensztein and Patillo, 2005). Two often cited examples are the attacks on the Argentinian peso and the Hong Kong dollar in 1995 and 1997 respectively, which were deterred by a rapid increase in domestic interest rates (Hawkins and Klau, 2000, and Klaassen and Jager, 2011). In Section 4.1 we explain how we identify currency crises.

2.2 Early Warning Systems for Currency Crises

Currency crises can be costly, particularly when they lead to sovereign debt or banking crises. It is therefore important to signal currency crises in a timely manner such that a crisis can be avoided, or the impact can be reduced. Early warning systems are models that send a signal well in advance of a crisis. Over the years dozens of EWSs have been developed which differ widely in the definition of a currency crisis, the estimation period and the countries included in the database, the inclusion of indicators, the forecast horizon and the statistical or econometric method. For overviews see Kaminsky, Lizondo and Reinhart (1998) or Abiad (2003). The differences between the EWSs make it hard to compare the studies.

3 Methodology

This section introduces two of the most used types of EWS to predict currency crises, the signal approach and the discrete choice model. The approaches will be described in the next subsections. The final subsection compares the out-of-sample currency crisis prediction performance using current-vintage data and using second estimates.

3.1 Signal Approach

Eichengreen, Rose and Wyplosz (1995) and Frankel and Rose (1996) introduce the event study graph to analyze and predict currency crises. The method involves a graphical comparison of the performance of indicators in times of crisis versus their performance in tranquil periods. Kaminsky, Lizondo and Reinhart (1998) extend this methodology to what is known as the signal approach. This approach consists of two stages. In the first stage the indicators that are expected to play a role in the crisis, such as inflation, debt as a percentage of GDP and the current account, are selected. Typically, a visual inspection of the event study graph determines whether the indicator shows a special, extraordinary behavior before a crisis. This helps to restrict the number of potential crisis indicators. In the second stage a threshold is determined for each indicator by minimizing the probability of not signalling a crisis that occurred (type I error) and the probability of signalling a crisis that did not occur (type II error). If the variable exceeds a pre-established threshold, then a crisis is signaled and the value of 1 is assigned to the binary variable, and zero otherwise. For each threshold we construct a contingency table as in Table 1. A represents the number of observations in which the model signals a crisis that actually took place (correct crisis signals); B corresponds to the number of observations in which the model signals a crisis that did not take place (false alarms, also known as type II errors); C is the number of observations in which the model does not signal a crisis that actually took place (missed crises, a.k.a type I errors); and D is the number of observations in which the model does not signal a crisis that did not take place (correct non-crisis signals).

Table 1: Contingency table of crisis realizations and signals.

	Realization	
Signal	Crisis	No crisis
Crisis	A	B
No crisis	C	D

The main advantages of the signal approach are that the method does not impose any parametric structure on the data, and that the method is more accessible and informative than tables of coefficient estimates. The main disadvantage is that the approach is intrinsically univariate as we analyze the individual contribution instead of the marginal contribution conditional on other variables (Frankel and Rose, 1996).

Determining the Level of the Threshold

The higher the threshold, the less likely it is for the indicator to send a crisis signal. This will result in less false alarms (type II errors), but also in more missed crises (type I errors). A lower threshold leads to less missed crises, but more false alarms. The optimal threshold depends on the relative costs of the two error types. There are several ways to determine the optimal threshold. We use the noise-to-signal ratio and the loss function.[3] The letters (A, B, C, D) used in the formulas below refer to the categories in Table 1.

Noise-to-Signal Ratio

The first method is used in the original signal approach model of Kaminsky, Lizondo and Reinhart (1998). The noise-to-signal (NtS) ratio is defined as:

$$NtS = \frac{B/(B+D)}{A/(A+C)} \tag{1}$$

The lower the NtS, the better the variable identifies the actual crisis. Indicators with an NtS equal to or greater than 1 should be discarded, since these do not have intrinsic predicting power. According to this criterium false alarms and missed crises are treated equally.

Loss Function

Alessi and Detken (2011) define an alternative criterium, which allows taking into account the policy maker's risk aversion. The loss function of the policy maker is defined as:

$$L = \theta\frac{C}{A+C} + (1-\theta)\frac{B}{B+D}, \tag{2}$$

where θ reflects the policy maker's relative risk aversion between missed crises (type I error) and false alarms (type II error). θ can take any value between 0 and 1. A θ smaller

[3] We also applied the ROC curve methodology, which provides similar results as the loss-function method. Results are available upon request.

than 0.5 implies that the policy maker is less risk averse towards missing a signal for a crisis than for receiving a false alarm. The costs of a false alarm are the costs of taking preventive actions, the risk of a self-fulfilling prophecy and the loss of trust in the policy makers when false alarms become frequent. Generally, policy makers and practitioners will prefer to be 'better safe than sorry'. For them missed crises are far more important than false alarms. Bussière and Fratzscher (2006) mention two arguments for the relative importance of missed crises versus false alarms. First, false alarms are less costly from a welfare perspective than missed crises. Second, false alarms may have been caused by appropriate policy initiatives that were taken when the fundamentals were so weak that a crisis was predicted.

A central banker can always realize a loss of min[θ; $1-\theta$] by disregarding the indicator. If θ is smaller than 0.5, the benchmark is obtained by ignoring the indicator, which amounts to never having any signals issued so that A = B = 0. The resulting loss according to Equation (2) is θ. If θ exceeds 0.5, the benchmark indicates that there is always a crisis, i.e. assuming a signal is always issued so that C = D = 0. The resulting loss is $1-\theta$. An indicator is useful to the extent that it produces a loss lower than min[θ; $1-\theta$] for a given θ. The usefulness of an indicator can then be defined as:

$$U = \min[\theta; 1-\theta] - L, \tag{3}$$

where the maximum value is to be determined.

3.2 Logit Models

Logit and probit models are widely used in EWSs for financial crises, including currency crises. Compared to the signal approach, the logit and probit models have advantages. First, the methods consider all the variables simultaneously (Kaminsky, Lizondo and Reinhart, 1998), and second, the independent variables can have a nonlinear effect on the probability of a crisis, which is appropriate because of the presence of strong nonlinear effects in currency crises mechanisms (Bussière, 2007).

3.2.1 Binary Logit Model

In the binary logit model the dependent variable is dichotomic and takes the value of 1 if the event occurs and 0 otherwise. In this setup, Y_{it} represents a binary variable for country $i \in \{1, \ldots, N\}$ at time $t \in \{1, \ldots, T_i\}$ where T_i denotes the number of time periods considered for the i^{th} country. The probability of an event is characterized by the logistic distribution. That is, for each country, the probability of the event is given by:

$$P(Y_{it} = 1) = \frac{\exp(X_{it}\beta)}{1 + \exp(X_{it}\beta)}, i = 1, \ldots, N; \tag{4}$$

where X_{it} denotes a vector of exogenous variables and β the vector of slope parameters.

The odds ratio, which is useful for interpretations, is determined as

$$\frac{P(Y_{it} = 1)}{1 - P(Y_{it} = 1)} = \exp(X_{it}\beta). \tag{5}$$

A common alternative for discrete choice models is the probit model. However, in our setup the logit model is preferred over the probit model because a crisis event has a relatively low frequency (as is the case in currency crises and sovereign debt crises). The reason for which the logit model is preferred is because the logistic distribution (logit model) has heavier tails than the normal distribution (probit model) (Manasse, Roubini and Schimmelpfennig, 2003; Bussière, 2007). However, differences are small (Cornelli, 2014).

3.3 Out-of-Sample Performance

We estimate an Early Warning System for the period 1990–2009, with current-vintage data. Then we compare its out-of-sample prediction for the period 2010–2014, with (i) current-vintage data, and (ii) original, second estimates. We include the currency crises that occur in the last half of 2008 (the fall of Lehmann Brothers) in the in-sample period, because it has elements not seen in earlier crises such as the role of advanced economies. According to Frankel and Saravelos (2012) leading indicators that most frequently appeared in earlier reviews are not statistically significant indicators in the Global Financial Crisis.

Measures for the Out-of-Sample Performance

The methods to determine the optimal threshold, the Noise-to-Signal or loss function approach (see Section 3.1), can also be used to measure the out-of-sample performance of the signal approach and the binary logit model. For the logit model an additional measure is available which is the quadratic probability score (QPS) proposed by Diebold and Rudebusch (1989) to evaluate out-of-sample forecasts. This measure indicates how close, on average, the predicted probabilities P_t and the observed realizations Z_t are. The QPS is given by

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2(P_t - Z_t)^2.$$

The QPS ranges from 0 to 2, with a score of 0 corresponding to perfect accuracy (well-predicted crisis, or a well-predicted tranquil period), and a score of 2 corresponding to a perfect false signal (missed crisis or false alarm).

4 Data

We focus on emerging economies, in particular on two regions: Latin America and Central and Eastern Europe (CEE) for which we have second estimates available. Countries in these regions implemented market reforms in the 1990s after a period of domestically-oriented economic policies. All countries have experienced political and institutional changes, changes in exchange rate regimes and at least one currency crisis since the early 1990s. In terms of total GDP and GDP per capita the regions are comparable, as shown in Table 2. For Latin America we select Argentina, Brazil, Mexico and Venezuela. These countries are the largest economies in terms of GDP and share economic and institutional features, such as the importance of commodities, and a history of changes in exchange rate regimes, and political and institutional changes. For CEE we select the four largest economies in the region: Russia, Poland, Czech Republic and Hungary.

Table 2: Comparison GDP and GDP per capita, 2014

Country	GDP	Ranking	GDP per capita	Ranking
Argentina	332.6	26	7,956	64
Brazil	1,206.1	11	5,970	76
Mexico	1,067.1	13	8,626	60
Venezuela	186.9	42	6,057	75
Czech Rep.	157.1	45	14,945	44
Hungary	117.2	54	11,888	50
Poland	429.5	22	11,305	54
Russia	999.8	14	6,844	69

Notes: total real GDP in billions of constant 2005 USD. Real GDP per capita in constant 2005 USD.
Ranking refers to world ranking (authors' calculations).
Source: World Bank (2014).

4.1 Currency Crisis Dating

To identify currency crises we combine the speculative pressure approach (EMPI) and the successful attack approach (large depreciation). An advantage of the speculative pressure approach is that the method works well under both fixed and floating exchange rate regimes. Frankel and Saravelos (2010) state that the inclusion of reserves is particularly relevant for countries with fixed exchange rate regimes, because capital flight and crisis incidence are present through larger drops in reserves rather than exchange rate weakness. This is very useful for our panel that starts in the early 1990s, when various emerging economies had their currencies pegged. A disadvantage of the speculative pressure approach is that the threshold is based on the standard deviation of the variables. When the movements in a particular currency crisis are relatively strong then these may prevent other depreciations that should be identified as currency crises to become visible, in other words: crises with very large EMPI values 'push out' other crises. The successful attack approach is not based on the standard deviation and is therefore used as a complementary crisis identification method.

For the speculative attack approach we choose for the EMPI with two components, changes in nominal exchange rate and changes in reserves. We do not include the interest rate for three reasons. First, the number of observations reduces, as interest rates are not available for all countries and years. Second, some interest rates are highly unrealistic. For example, Venezuela has a nominal rate in the period 2011-2014 lower than 1% annual in most months (source: IFS). Third, as Angkinand, Li and Willett (2006) observed when comparing crisis dating alternatives:

> ... while the three-component indices with interest rates may be able to pick up the exchange market pressure from interest rate hikes, they might miss the crisis periods when authorities intervenes the market with reserves only. In terms of picking up the mild EMP from selling reserves, the two-component indices without interest rates seem to be superior to the three-component indices." [page 16].

The EMPI is standardized with the mean and standard deviation for the in-sample period (in our case: 1990–2009) and not the entire period (in-sample and out-of-sample), to keep the out-of-sample exercise as pure as possible. For each country an index is constructed and a crisis is identified when the index exceeds a threshold, for which Kaminsky, Lizondo and Reinhart (1998) use three standard deviations of the index. For the successful attack approach we use the Frankel and Rose (1996) definition of a crisis. We refine the definition by also including higher frequency time frames: monthly, three-months and six-months periods. The reason is that several crises (in particular in the 2008–2009 period) are not identified because the depreciation takes place during two calendar years (and in none of the years is large enough to classify as a currency crisis), or the depreciation is (partially) reversed in the same year.

We exclude periods with hyperinflation as they should be categorized as inflation crises instead of currency crises. We use Cagan (1956)'s hyperinflation definition: a monthly increase in consumer prices of 50% or more. We define a calender year with hyperinflation if there is at least one month of hyperinflation.

We label the year prior to the crisis as a crisis run-up year. This run-up year is used since we are interested in an early warning system, so we want to detect a currency crisis before it actually occurs. Additionally, we use a window exclusion period of 12 months, which implies that a crisis that takes place within 12 months after a previous crisis is not considered a separate crisis, but a continuation of the previous crisis.

The resulting currency crises are shown in Table 3. The fifth column of the table shows there have been 17 currency crisis episodes in the period 1990–2009, and the last column shows there have been seven currency crisis episodes in the out-of-sample period 2010–2014.

4.2 Explanatory Variables

As explanatory variables we use 10–15 economic and financial indicators that correspond to different types of currency crises (Kaminsky, 2006), and that are typically used in studies on EWSs for currency crises. The following variables are available as second estimates for the out-of-sample period 2010–2014: current account balance as a percentage of GDP, import cover, import growth, real GDP growth in the US, a selection of commodity price indices (agricultural goods, metals, energy), inflation, real GDP growth, and gross fixed capital investments. The second estimates are taken from the Focus Economic Consensus of Haver Analytics (HA). The database offers monthly reported forecasts for economic indicators. Current account balance and inflation are not available in HA, so we take these from World Economic Outlook (WEO)—in a semi-annual vintage frequency (April and October). We take the second estimates from HA as published in January for the year in course, and from WEO as published in October of the previous year. This is an ad hoc choice, that is a trade-off between accuracy and timeliness. Forecasting over a long time

Table 3: Identification of currency crises, 1990–2014

Country	**Starting year**	**Definition 1: EMPI**	**Definition 2: Depreciation**	**Combined 1990–2009**	**Combined 2010–2014**
Argentina	1992	2002	2002, 2013, 2014	2002	2013, 2014
Brazil	1995	1999	1999, 2001-2002, 2008	1999, 2001-2002, 2008	
Mexico	1990	1994-1995, 2008	1994-1995, 2008-2009	1994-1995, 2008-2009	
Venezuela *	1990	1994, 1995-1996, 2002-2003, 2009, 2011, 2013	1993-1996, 2002, 2011, 2013	1993-1996, 2002-2003, 2009	2011, 2013
Czech Republic	1993		1997, 2008-2009	1997, 2008-2009	
Hungary	1990	1993, 2008-2009	1996, 2008-2009, 2011	1993, 1996, 2008-2009	2011
Poland	1991		2008-2009, 2011	2008-2009	2011
Russia **	1994	1995, 1998	1994-1995, 1998-1999, 2008-2009, 2014	1998-1999, 2008-2009	2014

Notes:
Starting year: We excluded years with hyperinflation, and years without available data.
Definition 1: a crisis is identified when the EMPI is greater than three times the standard deviation.
Definition 2: a crisis is identified when: (i) annual depreciation is greater than 25%, (ii) annual depreciation 10 percentage points higher than previous year's annual depreciation. Additionally, a crisis is identified in the year in which depreciation of monthly, three-months or six-months periods exceeds 25%.
Combined: combines definitions 1 and 2, with the implementation of a 12 months window exclusion period. We distinguish the in-sample period (1990–2009) in the penultimate column from the out-of-sample period (2010–2014) in the ultimate column.
* There is no data available on the reserves in 2014 in Venezuela, so no EMPI was calculated.
** Currency crises that occur at the start of the in-sample period are excluded, because the run-up period is not part of the in-sample period. The 1994-1995 crisis in Russia is therefore excluded.
Data source: IFS.

period typically leads to greater variation relative to the actual outcome, while forecasting over a shorter period may lead to more precision, but also leaves less time for corrective action. Berkmen et al. (2009) use consensus growth forecast changes, which has the advantage of pooling across various forecasters and potentially suffering from less bias than the WEO. For the same argument we prefer to use the consensus data from HA, and only include WEO predictions when these data series are not available in HA.

The following variables are not available as second estimates, but are essential to include in any EWS for currency crises: general government gross debt as a percentage of GDP, domestic credit from financial institutions, broad money growth, foreign direct investments, portfolio investments, and changes in the US interest rate. For these variables and for current-vintage data we use International Financial Statistics (IMF), Haver Analytics and World Development Indicators. Details on the variables, their sources, frequency and availability are shown in the Appendix. All explanatory variables have been standardized, using the mean and standard deviation of the in-sample period (1990-2009) to keep the out-of-sample prediction exercise as pure as possible.

5 Empirical Results

Since early warning systems are only useful when they send an *early* warning, it is common to use indicators from year $t-1$ to detect a possible crisis in year t. In other words, the model links the dependent variable (the crisis dummy) with a selection of indicators from the year prior to the crisis entry on the dependent variable (the crisis dummy).

5.1 Signal Approach

In the signal approach we determine the optimal threshold separately for each indicator. For each of the thirteen selected variables we analyze the performance according to two criteria: the Noise-to-Signal ratio and the loss function (the latter with a relatively high penalty for missed crises, i.e. $\theta \geq 0.5$). We select the five indicators that are in top positions according to these two criteria: Current Account to GDP ratio, import cover, M2 growth, domestic credit by financial institutions and changes in food prices. Of these, M2 growth and domestic credit by financial institutions are not available as real-time data, that is second estimates are not available, which makes them not suited for our comparison. We discard changes in food prices, because we consider the countries produce and consume (export and import) different foods, where the prices are not uniformly increasing and decreasing (the commodity price lottery as documented by Blattman, Hwang and

Williamson, 2007). We continue our analysis with the other two indicators, change in import cover and the Current Account to GDP ratio.

Import cover, defined as the ratio of reserves to imports, is a signal of currency stability. The ratio is typically expressed in months and can be interpreted as the number of months that a country can continue to import financed by international reserves only. A deterioration in the import cover can be caused by a decrease in reserves and/or an increase in imports. A decrease in the import cover ratio is associated with a higher probability of a currency crisis. We use two values for the thresholds of this indicator, -0.7σ and -0.1σ. Since all indicators have been standardized, the threshold can be expressed simply as -0.7 and -0.1. The first (-0.7) is the optimal threshold according to the noise-to-signal ratio and the loss function with a relatively mild penalty for missed crises. For stronger penalties for missed crises, the optimal threshold is -0.1. For both thresholds we show the contingency table, in Table 4. With a threshold of -0.7 the approach generates only 18 false alarms, but at the cost of having a high number of missed crises. With a lower threshold in absolute terms (-0.1), the approach identifies more crises (11), but also generates more false alarms (73). The latter threshold would be preferred by a policy maker that is more averse to missed crises than to false alarms.

Table 4: Contingency tables for change in import cover

Threshold	**−0.7**	**−0.1**
Correct crisis signal	6	11
Missed crisis (type I error)	11	6
False alarm (type II error)	18	73
Correct non-crisis signal	101	46

The pattern of the indicator, the currency crises and the threshold is shown in Figure 2. Using a threshold of -0.7 six crises are correctly identified: Argentina in 2002, Brazil in 1999 and 2001, Venezuela in 1993 and 2002, and Czech Republic in 1997. With a threshold of -0.1 eleven crises are correctly identified. Apart from these six crises, also the crises of Mexico in 2008, Czech Republic in 2008, Hungary in 2008, and Russia in 1998 and in 2008 are identified correctly.

Another indicator that is commonly used in EWSs for currency crises is the current account balance, which is a signal of an economy's external trade position, and indirectly the capital account.[4] A worsening current account deficit puts the exchange rate under pressure and may result in a depreciation of the currency, and/or depletion of international reserves to finance the deficit. As with the previous indicator, we use several threshold values, -1.4, -0.3 and 0.8. The first threshold is optimal when missed crises are not punished stronger than false alarms (optimal threshold according to Noise-to-Signal ratio), the second threshold is optimal when missed crises are punished relatively mildly (optimal threshold according to loss function with equal risk aversion for missed crises and false alarms), and the third threshold is optimal when missed crises are punished relatively strongly (optimal threshold according to loss function with higher risk aversion for missed crises than for false alarms). For both thresholds we show the contingency tables in Table 5. The pattern of the indicator, the currency crises and the thresholds are shown in a time line in Figure 3.

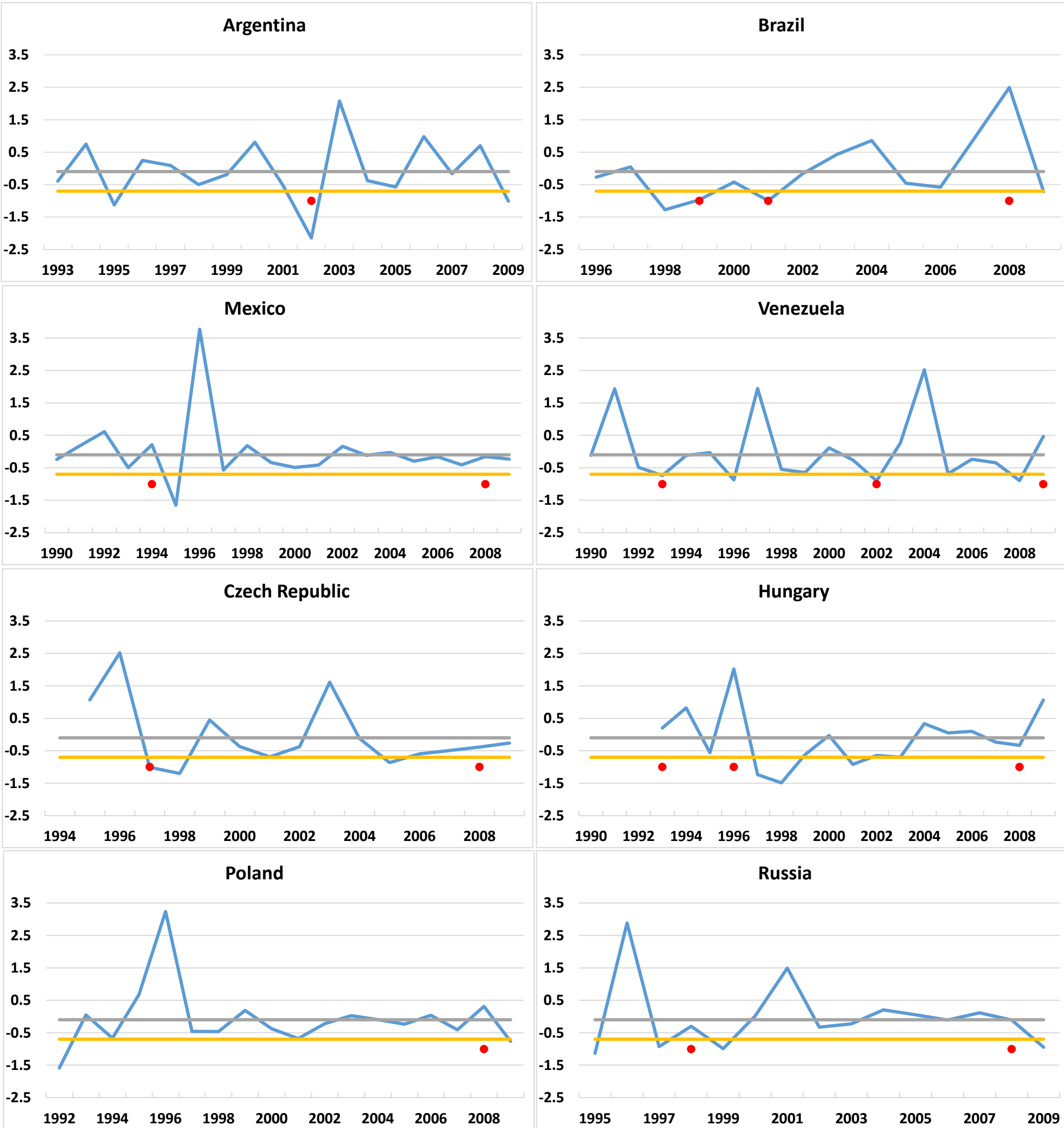
Table 5: Contingency tables for current account balance to GDP ratio

Threshold	-1.4	-0.3	0.8
Correct crisis signal	4	12	16
Missed crisis (type I error)	13	5	1
False alarm (type II error)	5	50	89
Correct non-crisis signal	114	69	31

The performance of the signal approach using the change in import and the current account balance is very similar.

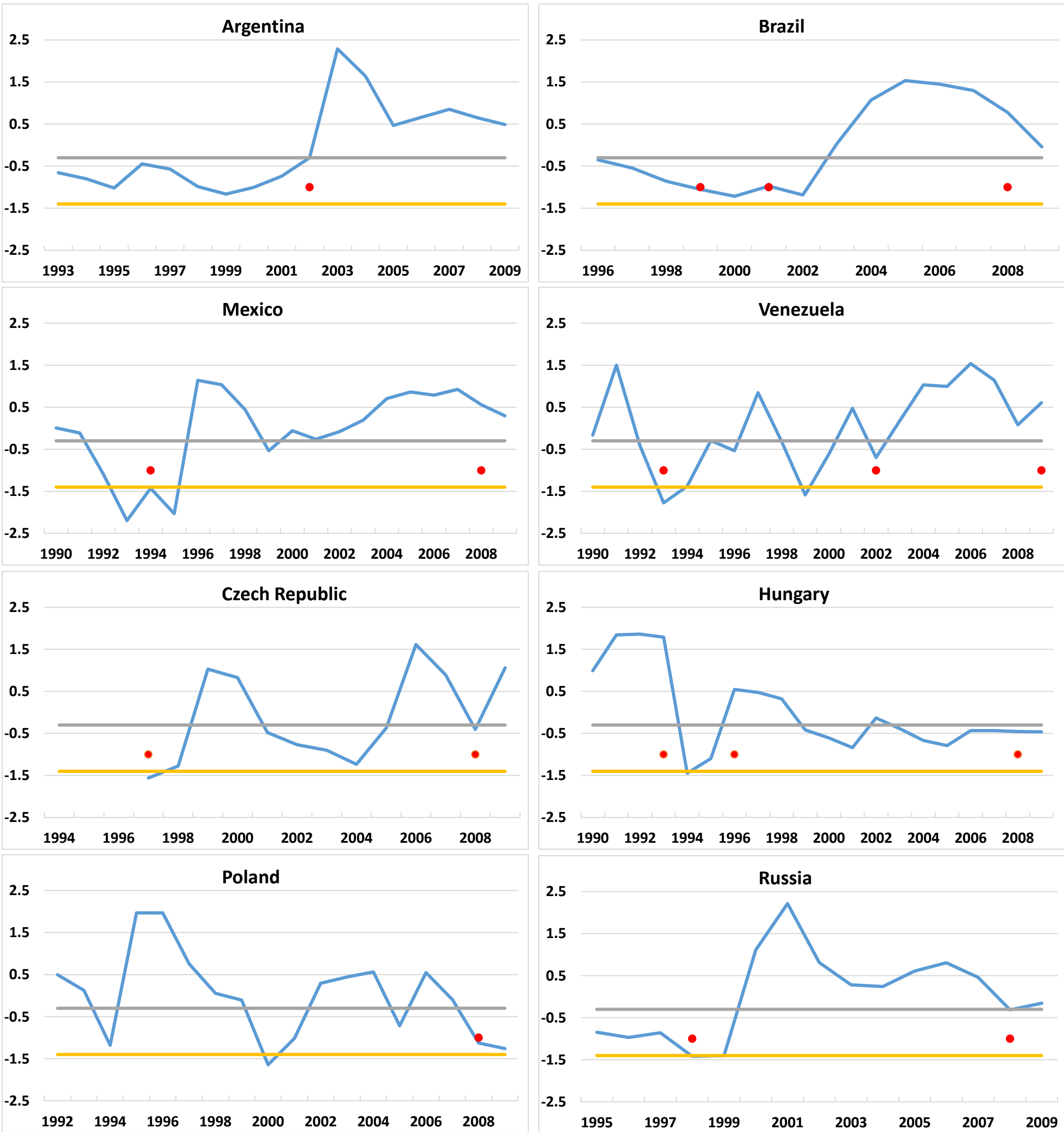
[4]A current account deficit (surplus) is typically accompanied by a capital account surplus (deficit).

Figure 2: Signal approach for the change in import cover in eight countries during 1990–2009.



Notes: The dots represent the realization of the binary variable that takes the value of 1 if there was a crisis and zero otherwise. The two horizontal lines represent the thresholds, the bottom line the threshold of -0.7 and the top line the threshold of -0.1. The solid line is the standardized movement of the change in the import cover, with a 1 year lag.

Figure 3: Signal approach for the current account balance to GDP ratio.



Notes: This figure shows the signal approach for the current account balance as a percentage of GDP for eight countries from 1990 to 2009. The dots represent the realization of the binary variable that takes the value of 1 if there was a crisis and zero otherwise. The two horizontal lines represent the thresholds: the bottom line is the threshold of -1.4 and the top line is the threshold of -0.3. The solid line is the standardized movement of the current account to GDP ratio, with a one-year lag.

Out-of-Sample Performance

We compare the performance of using second estimates versus current-vintage data in the predictions. We use the thresholds that were estimated in the in-sample period (1990–2009), to predict crises in the out-of-sample period 2010–2014.

Table 6: Signal approach: out-of-sample results for the change in import cover.

		Threshold -0.7		***Threshold -0.1***	
Country	**Actual crisis**	**Current Vintage**	**2nd Estimates**	**Current Vintage**	**2nd Estimates**
Argentina	2013–2014	2011–2013	2011, 2012	2011–2014	2011, 2012, 2014
Brazil	–	2011, 2014	2011, 2012	2011, 2013, 2014	2011, 2012
Mexico	–	–	–	2011, 2013, 2014	2011–2013
Venezuela	2011, 2013	2011, 2012, 2014	–	2011–2014	2011, 2012
Czech Rep.	–	2012	–	2011, 2012	2011, 2012, 2014
Hungary	2011	–	–	2011, 2013, 2014	2011, 2012, 2014
Poland	2011	–	–	2011, 2012, 2014	2011, 2012, 2014
Russia	2014	2011, 2012	–	2011–2014	2011, 2012, 2014
# correct crises		2	0	6	4
# missed crises		4	6	0	2
# false alarms		9	4	19	17
# correct non-crises		17	22	7	9
Noise to Signal ratio		1.04	N/A	0.73	0.98
Usefulness of the loss function à la Alessi and Detken (2011) with:					
$\theta = 0.5$		0.08	-0.04	0.32	0.17
$\theta = 0.6$		-0.08	-0.24	0.22	0.04
$\theta = 0.7$		-0.24	-0.43	0.15	-0.07
$\theta = 0.8$		-0.39	-0.62	0.08	-0.17

Notes: Current-vintage refers to the data vintage as available in June 2015. Second estimates refer to the use of forecasts compiled by Haver Analytics in the month January for the current year. The top section of the table contains the crisis years, as took place according to our definition (column 2), and according to the predictions with the Signal Approach, using current-vintage data (columns 3 and 5) and second estimates (columns 4 and 6). The middle section of the table summarizes the model's performance. The observations in the out-of-sample period are divided over the four possible categories: correctly predicted crises, missed crises, false alarms and the correctly predicted tranquil years. The bottom section of the table shows two criteria that we used to measure and compare the out-of-sample performance.

For the change in the import cover as shown in Table 6 we compare column 3 versus 4, and column 5 versus 6. We can see that the current-vintage data provides better prediction results than the second estimates, because more crises are correctly picked up, the noise-to-signal ratio is lower and the usefulness is higher for all θ. Monetary authorities will prefer the threshold with the highest value (-0.1), because this will lead to a higher number of correctly predicted crises, reflected in a higher usefulness (comparing columns 3 and 5 of Table 6, e.g. for θ = 0.5: 0.10 $\geq$ -0.10).

For the other indicator, current account balance as a percentage of GDP, we discard one threshold (-1.4), because no crisis signal was sent, neither when using current-vintage data nor when using second estimates. Therefore in Table 7 we show the performance with thresholds -0.3 and 0.8.

Table 7: Signal approach: out-of-sample results for current account balance as a percentage of GDP.

		Threshold -0.3		***Threshold 0.8***	
Country	Actual crisis	Current Vintage	2nd Estimates	Current Vintage	2nd Estimates
Country	actual crisis	Current vintage	2^{nd} estimate of forecast	Current vintage	2^{nd} estimate of forecast
Argentina	2013-2014	—	—	2011-2014	2012-2014
Brazil	—	2013-2014	2012-2014	2011-2014	2011-2014
Mexico	—	—	—	2012-2014	2011-2014
Venezuela	2011, 2013	2011, 2013-2014	—	2011-2014	2011-2014
Czech Rep.	—	—	—	2011	2011, 2013
Hungary	2011	—	—	—	—
Poland	2011	2011-2012	2013	2011-2013	2011-2014
Russia	2014	2011-2014	2011-2014	2011-2014	2011-2014
# correct crises		4	1	5	5
# missed crises		2	5	1	1
# false alarms		7	7	18	20
# correct non-crises		19	19	8	6
Noise to Signal		0.49	1.96	0.79	0.89
Usefulness, based on loss function with:					
θ = 0.5		0.22	0.00	0.26	0.24
θ = 0.6		0.08	-0.18	0.15	0.13
θ = 0.7		-0.06	-0.36	0.06	0.04

Notes: see Table 6.

We find very similar results for this second indicator. Current-vintage data generates better results in terms of lower noise-to-signal ratios, and higher usefulness values. The difference is marginal when using the threshold of 0.8. For the threshold of -0.3 the currency crises in Venezuela in 2011 and 2013 as well as the crisis in Poland 2011 are picked up when we use current-vintage data, but not with second estimates. Monetary authorities that are averse to missed crises would prefer the highest threshold (0.8), because this will lead to less missed crises. However, our results do not capture the possible actions taken when a crisis signal arrives and the costs, including the loss in credibility.

The predictions based on current-vintage data are better than the ones based on second estimates. More crises are predicted correctly and higher usefulness figures are obtained for all levels of θ.

5.2 Binary Logit Model

We estimate the model for the in-sample period (1990–2009) of the pooled dataset for eight countries. We use the Hausman test to decide between fixed or random effects. In addition, we use a regional dummy to distinguish the four Latin American countries from the four Central and Eastern European countries. We include all potentially relevant variables, even when there could be multicollinearity among variables. Under multicollinearity the estimates are unbiased and consistent, but the standard deviations may be inflated. This does not affect the predictions.

The outcomes of the logit model are shown in Table 8, for the pooled observations of the eight countries for the period 1990–2009. The dependent variable is a dummy variable with value 1 if the country experiences a currency crisis. A value of 1 is also assigned to the pre-crisis year, the run-up to the currency crisis.

We now turn to the performance in the out-of-sample period, by comparing the generated probabilities of a currency crisis with the actual outcome. The out-of-sample performance is shown graphically in Figure 4. We perform the out-of-sample predicting with (i) current-vintage data, and (ii) second estimates. In Figure 4 we observe that predictions

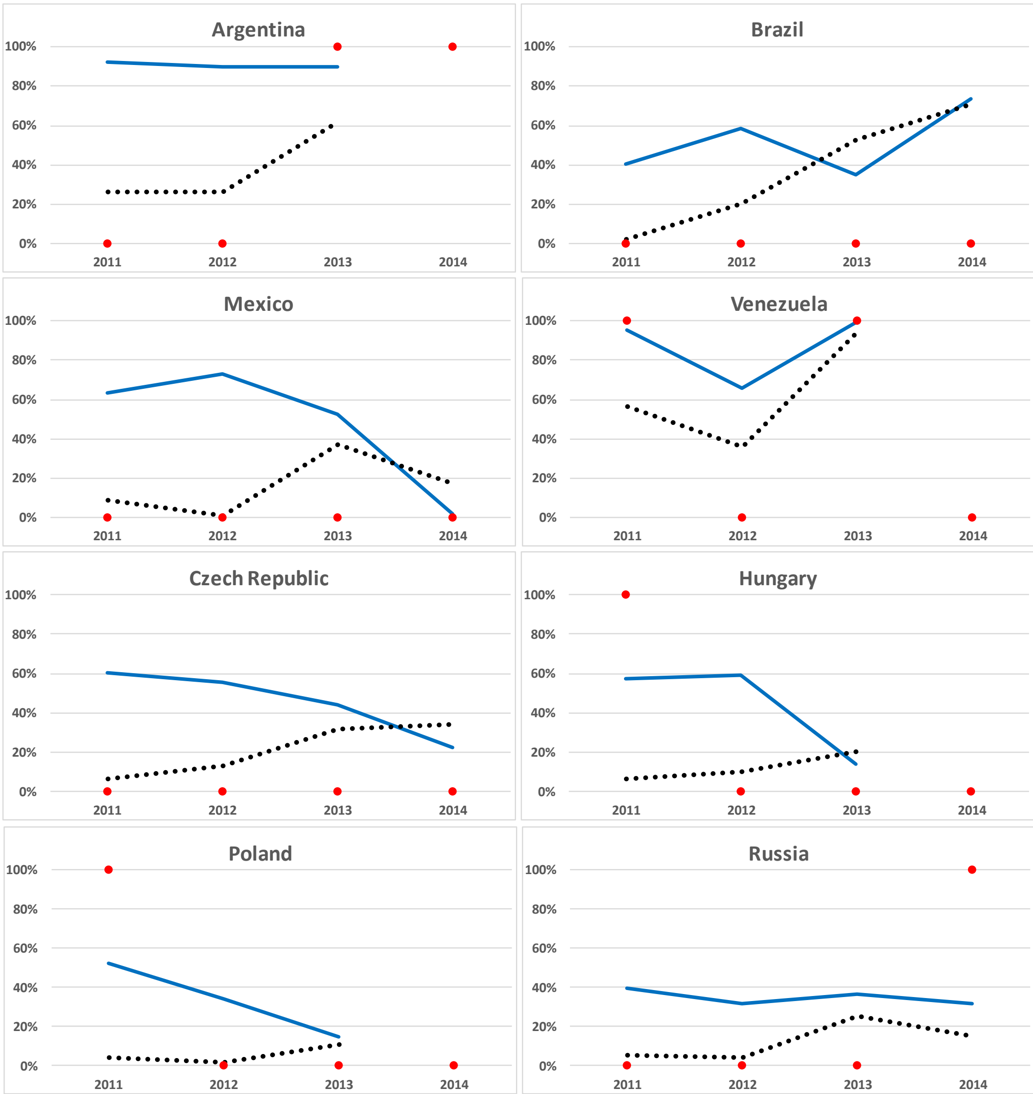
Table 8: Binary logit regressions with currency crisis dummy as the dependent variable

Variables	**Coefficient**	**Std. Error**
CAGDP	-0.520	(0.470)
ΔIMPCOVER	-0.643	(0.611)
INFLAT	$1.213*$	(0.666)
GDPGROW	$1.513*$	(0.892)
IMPGROW	$-2.759***$	(0.937)
ΔUSGROW	-0.262	(0.275)
AGRI	$0.908***$	(0.350)
METAL	$-1.032**$	(0.429)
ENERGY	0.137	(0.328)
FDI	0.288	(0.350)
ΔGDEBTGDP	0.698	(0.615)
ΔCREDFIN	0.678	(0.590)
ΔPFINV	0.141	(0.468)
ΔM2GROW	0.046	(0.526)
ΔUSINT	0.373	(0.325)
REGION	$-0.846**$	(0.406)
REGION $\times$ CAGDP	0.507	(0.634)
REGION $\times\Delta$IMPCOVER	0.427	(0.880)
REGION $\times$ INFLAT	-0.711	(0.741)
REGION $\times$ GDPGROW	-1.714	(1.148)
REGION $\times$ IMPGROW	1.723	(1.156)
REGION $\times$ FDI	0.208	(0.531)
REGION $\times\Delta$GDEBTGDP	-1.009	(0.808)
REGION $\times\Delta$CREDFIN	-0.401	(0.737)
REGION $\times\Delta$PFINV	-0.004	(0.546)
REGION $\times\Delta$M2GROW	0.092	(0.714)
Log likelihood		-57.428
McFadden pseudo R^2		0.366
Total observations		132
Observations of crisis		46

Notes: Countries in the sample: Argentina, Brazil, Mexico, Venezuela, Czech Republic, Hungary, Poland and Russia; annual data from 1990 to 2009. CAGDP = Current Account to GDP ratio; ΔIMPCOVER = percentage change in import cover (reserves / imports); INFLAT = consumer price inflation; GDPGR = Real annual GDP growth; IMPGROW = imports growth; ΔUSGROW = change in US real GDP growth; AGRI, METAL, ENERGY are the annual changes in the composite price indices for agricultural products, metals and minerals, and energy, respectively; FDI = Foreign Direct Investment to GDP ratio; GDEBTGDP = Gross central government debt to GDP ratio; CREDFIN = domestic credit given by financial institutions; PFINV = portfolio investments to GDP ratio; M2GROW = Annual M2 growth; USINT = interest rate on 10 year US treasury bonds; REGION = dummy variable with value 1 for the (four) CEE countries. All variables in this in-sample estimation are current vintage data.

based on current-vintage data produce a higher probability of crises than predictions based on second estimates. The predictions based on current-vintage data yield a low number of missed crises, but at the cost of a high number of false alarms. In contrast predictions based on second estimates do not pick up many crises and issue few false alarms.

Figure 4: Probability of currency crises: binary logit regressions



Notes: The dots represent the realization of the binary variable that takes the value of 1 if there was a crisis and zero otherwise. The solid line represents the probability of a crisis when using the current-vintage data. The dotted line represents the probability of a crisis when using the second estimates.

To compare the out-of-sample performance we determine a cut-off probability to distinguish crisis signals from non-crisis signals. Through a grid search we find that the

Table 9: Binary logit model: out-of-sample performance.

Country	**Actual Crisis**	**Current Vintage**	**2nd Estimates**
Argentina	2013-2014	2011, 2012, 2013	2013
Brazil	–	2012, 2014	2013, 2014
Mexico	–	2011, 2012, 2013	–
Venezuela	2011, 2013	2011, 2012, 2013	2011, 2013
Czech Republic	–	2011, 2012	–
Hungary	2011	2011, 2012	–
Poland	2011	2011	–
Russia	2014	–	–
# correct crises		5	3
# missed crises		1	3
# false alarms		11	2
# correct non-crises		12	21
Noise to Signal		0.57	0.17
Usefulness for the loss function à la Alessi and Detken (2011) with:			
$\theta = 0.5$		0.80	0.73
$\theta = 0.6$		0.79	0.68
$\theta = 0.7$		0.78	0.63
$\theta = 0.8$		0.79	0.59

Notes: Current-vintage refers to the data as available in June 2015. 2nd estimates refers to the use of forecasts compiled by Haver Analytics in the month January for the current year for HA data, and October of the previous year for WEO-data.
The top section of the table contains the crisis years according to our definition (column 2), and according to the predictions with the logit model, using current-vintage data (column 3) and second estimates (column 4).
The middle section of the table summarizes the model's performance. The 29 observations in the out-of-sample period are divided over the four possible categories: correctly predicted crises, missed crises, false alarms and the correctly predicted tranquil years.
The bottom section of the table shows two criteria that we used to measure and compare the out-of-sample performance.

Table 10: Quadratic probability score for binary logit regressions: out-of-sample results 2011–2014.

	Current-Vintage	***2nd Estimates***
All countries	0.413	0.433
Latin America	0.542	0.453
Central Europe	0.285	0.414
Argentina	0.515	0.570
Brazil	0.910	0.375
Mexico	0.537	0.212
Venezuela	0.084	0.763
Czech Republic	0.289	0.053
Hungary	0.310	0.662
Poland	0.148	0.624
Russia	0.364	0.431
Countries with crises	0.289	0.599
Countries without crises	0.578	0.213

Notes: Current-vintage refers to the data as available in June 2015. 2nd estimates refers to the use of estimates compiled by Haver Analytics in the month January for the year in course. The QPS score ranges between 0 and 2, where a lower score reflects a better out-of-sample performance.

optimal cut-off (in terms of the lowest Noise-to-Signal ratio, and the highest values for usefulness for the loss function a la Alessi and Detken, 2011) is 50%. Thus, when the probability of a crisis exceeds 50% then the logit model calls a crisis. The results are shown in Table 9. We observe that using current-vintage data identifies 5 crises (versus 3 when using second estimates), although at the cost of more false alarms (11 versus 2). Note that according to the Noise-to-Signal ratio the second estimates are considered better for crisis predictions. We give more attention to the usefulness because we consider the costs of missed crises higher than the cost of false alarms; the Noise-to-Signal ratio does not make any distinction.

A more formal statistic for the performance of the predictions, the Quadratic Probability Score, is presented in Table 10. Here we can observe that current-vintage time data do not always perform better than second estimates. In particular for countries where no crisis took place (Brazil, Mexico and Czech Republic), the QPS for predictions based on second estimates is lower than the QPS for predictions based on current-vintage data, indicating that predictions based on second estimates are better. For countries where a currency crisis took place the reverse holds: the QPS is lower for predictions based on current-vintage data than for predictions based on second estimates.

6 Conclusion

In this paper we focus on the use of real-time data for early warning systems for currency crises. EWSs have received many critiques, one of which is related to data availability. The use of realized data for EWSs is unrealistic and not feasible in practice, since these are not available when predictions are made. We select eight emerging economies from two regions: Argentina, Brazil, Mexico and Venezuela from Latin America and Czech Republic, Hungary, Poland and Russia from Central and Eastern Europe (CEE). These countries form a more or less homogeneous sample in terms of size of the economy, comparable economic history since the 1990s including financial crises, and switches in exchange rate regimes. We analyze these countries in a pooled data set, covering the period from 1990 to 2014.

The signal approach is often used as EWS for currency crises. Based on the in-sample data and two criteria we determine two critical thresholds for each indicator separately. In the out-of-sample period we use these thresholds to identify crises. Comparison of current-vintage time data and second estimates shows that the latter perform worse in signalling crises. This conclusion also holds for logit models, the second EWS that we analyze. Predictions with current-vintage yields higher probabilities of crises than predictions with second estimates; fewer crises are missed, but more false alarms are generated.

We conclude that in our models current-vintage data perform better than second estimates in terms of predicting currency crises. In other words, based on second estimates a high number of crises will not be signaled in time. Some possible explanations are that the second estimates (which are based on consensus expectations) tend to smooth out extremes, and that the realizations were more dramatic than predicted in crisis episodes. A limitation of our work is the small number of crises in the out-of-sample prediction period, which implies that the results should be taken with care. Future research will consist of expanding the number of countries. In addition we aim at investigating the impact of using other real-time data, like the descriptive statistics of the information in early estimates.

Predicting financial crises, including currency crises, is notoriously hard, even in retrospect. Taking properly account of the information that is available at the moment a researcher has to prepare predictions, makes it even more difficult. For supervisors, who assume that the costs of missed crises are much higher than the costs associated with false alarms, this is not a comforting finding, because current-vintage data is not available on time for making genuine out-of-sample predictions. The EWS literature has focused mainly on applying different methodologies on current vintage data sets. Given our findings it would be better to dedicate more resources in producing forecasts of the indicators themselves.

References

Abiad, A. (2003), "Early warning systems: A survey and a regime-switching approach", *IMF Working Paper* 32, International Monetary Fund, Washington, DC.

Alessi, L. and C. Detken (2011), "Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity", *European Journal of Political Economy*, **27**, 520–533.

Angkinand, A. P., J. Li, and T. Willett (2006), "Measures of currency crises: A survey", mimeo Claremont Graduate University, Claremont, CA.

Berg, A., E. Borensztein, and C. Pattillo (2005), "Assessing early warning systems: How have they worked in practice?", *IMF Working Paper* 52, International Monetary Fund, Washington, D.C.

Berkmen, Pelin, Gaston Gelos, Robert Rennhack, and James P. Walsh (2009), "The global financial crisis: Explaining cross-country differences in the output impact", IMF Working Paper WP/09/280, International Monetary Fund.

Blattman, C., J. Hwang, and J. G. Williamson (2007), "The impact of the terms of trade on economic development in the periphery, 1870–1939: Volatility and secular change", *Journal of Development Economics*, **82**, 156–179.

Bussière, M. (2007), "Balance of payment crises in emerging markets. How early were the 'early' warning signals?", *ECB Working Paper* 713, European Central Bank, Frankfurt am Main, Germany.

Bussière, M. (2013), "In defense of early warning systems", *Document de Travail* 420, Banque de France, Paris.

Bussière, M. and M. Fratzscher (2006), "Towards a new early warning system of financial crises", *Journal of International Money and Finance*, **25**, 953–973.

Cagan, P. (1956), "The monetary dynamics of hyperinflation", in M. Friedman, editor, *Studies in the quantitative theory of money*, University of Chicago Press, Chicago.

Calvo, G. A. and C. M. Reinhart (2000), "Fear of floating", *NBER Working Papers* 7993, National Bureau of Economic Research, Cambridge, MA.

Choo, C. W. (2009), "Information use and early warning effectiveness: Perspectives and prospects", *Journal of the American Society for Information Science and Technology*, **60**, 1071–1082.

Cornelli, F. (2014), “Comparing the performance of logit and probit Early Warning Systems for currency crises in emerging market economies”, *IMF Working Paper* 65, International Monetary Fund, Washington, DC.

Diebold, F. X. and G. D. Rudebusch (1989), “Scoring the leading indicators”, *The Journal of Business*, **62**, 369–391.

Eichengreen, B., A. K. Rose, and C. Wyplosz (1995), “Exchange rate mayhem: The antecedents and aftermath of speculative attacks”, *Economic Policy*, **21**, 251–312.

Frankel, J.A. and A. K. Rose (1996), “Currency crashes in emerging markets: An empirical treatment”, *Journal of International Economics*, **41**, 351–366.

Frankel, J.A. and G. Saravelos (2010), “Are leading indicators of financial crises useful for assessing country vulnerability? Evidence from the 2008-2009 global crisis”, NBER Working Paper 16047, National Bureau of Economic Research.

Frankel, J.A. and G. Saravelos (2012), “Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis”, *Journal of International Economics*, **87**, 216–231.

Ghosh, R., J. D. Ostry, and C. Tsangarides (2010), “Exchange rate regimes and the stability of the international monetary system”, *IMF Occasional Paper* 270, International Monetary Fund, Washington, D.C.

Girton, L. and D. Roper (1977), “A monetary model of exchange market pressure applied to the postwar Canadian experience”, *American Economic Review*, **4**, 537–548.

Glick, R. and M. Hutchison (2011), “Currency crises”, *FRBSF Working Paper series* 22, Federal Reserve Bank of San Francisco, San Francisco, CA.

Gunther, J. W. and R. R. Moore (2003), “Early Warning Models in real time”, *Journal of Banking and Finance*, **27**, 1979–2001.

Hawkins, J. and Klau M. (2000), “Measuring potential vulnerabilities in emerging market economies”, *BIS Working Papers* 91, BIS, Basel, Switzerland.

Holopainen, M. and P. Sarlin (2016), “Toward robust early-warning models: A horse race, ensembles and model uncertainty”, *ECB Working Paper* 1900, European Central Bank, Frankfurt, Germany.

Kaminsky, G. L. (2006), “Currency crises: Are they all the same?”, *Journal of International Money and Finance*, **25**, 503–527.

Kaminsky, G. L., S. Lizondo, and C. M. Reinhart (1998), “Leading indicators of currency crisis”, *IMF Staff Papers* 45/1, International Monetary Fund, Washington, DC.

Klaassen, F. and H. Jager (2011), “Definition-consistent measurement of exchange market pressure”, *Journal of International Money and Finance*, **30**, 74–95.

Lo Duca, M. and T. A. Peltonen (2013), “Assessing systemic risks and predicting systemic events”, *Journal of Banking and Finance*, **37**, 2183–2195.

Manasse, P., N. Roubini, and A. Schimmelpfennig (2003), “Predicting sovereign debt crises”, *IMF Working Paper* 221, International Monetary Fund, Washington, DC.

Reagle, D. and D. Salvatore (2005), “Robustness of forecasting financial crises in emerging market economies with data revisions—a note”, *Open Economies Review*, **16**, 209–216.

A Variables: Definitions and Sources

	Indicator	Definition	Source and vintage	Source and realized data
1	Inflation	Annual change in the consumer price inflation index	WEO: 1999–2020	WEO: 1990–2014
			HA: 2006–2020	WDI: 1990–2014, except Arg, Ven
2	Broad money growth	Annual change in M2	N/A	WDI: 1990–2014
3	Real GDP growth	GDP in local currency, in constant prices; annual growth	WEO: 1999–2020	WEO: 1990–2014
			HA: 2006–2020	HA: 1990–2014 WDI: 1990–2014
4	Gross fixed capital formation	Annual growth of real gross fixed capital formation	HA: 2006–2020	HA: 1993–2014, WDI: 1990–2014
		Idem, % of GDP	WEO: 2011–2015	WEO: 1990–2014, WDI: 1990–2014
5	Government gross debt to GDP	Government gross debt as a percentage of GDP	N/A	WEO: 1990–2020
6	Short term debt	Short term debt as a % of total debt	N/A	WDI: 1990–2013
7	Domestic credit	Domestic credit provided by financial sector, as % of GDP	N/A	WDI: 1990-2014
8	Portfolio investments	Portfolio investments (bonds and equity) inflows	N/A	WDI: 1990–2013/2014
9	FDI inflows	Foreign Direct Investments inflows, as % of GDP	N/A	WDI: 1990–2014
10	Current Account balance	Current Account balance, as % of GDP	WEO: 2004–2015	WEO: 1990–2014
11	Imports growth	Annual change in value of imports of goods (USD)	HA: 2006–2015	HA: 1995–2014, WDI: 1990–2014
12	Import cover	Calculation: International reserves / (Imports / 12)	HA: 2006–2015	HA: 1995–2014
12B	International reserves	International reserves (excluding gold), in billions of USD	HA: 2006–2015	HA: 1995–2014 WDI: 1990–2014
12C	Value of imports of goods	Value of imports of goods; billions of USD	HA: 2006–2020	HA: 1995–2014

Continues on the next page.

	Indicator	Definition	Source and vintage	Source and realized data
13	US interest rate	10-year US government bond yield	N/A	WDI: 1990–2014
14	US real GDP growth	Quarterly real GDP in USA; converted into annual growth rate	HA: 2009–2015	WDI: 1990–2014
			HA: 2009–2015	HA: 2009–2015
15	Agriculture price index	Agriculture products price index (2000 = 100), incl. vegetable oils, soy beans, rice, wheat, maize, meat, bananas, seafood, sugar, coffee, tea and cocoa	WB: 2002–2015	WB: 1991–2014
16	Metals price index	Metals and minerals price index (2000 = 100), incl. aluminum, copper, iron ore, lead, nickel, steel, tin and zinc	WB: 2002–2015	WB: 1990–2014
17	Energy price index	Energy price index (2000 = 100), consists of coal, crude oil and natural gas	WB: 2002–2015	WB: 1990–2014
18	Regional dummy	Distinguish Latin America and CEE		

Data sources:
WEO: World Economic Outlook. Semi-annual reports.
WDI: World Development Indicators. Annual reports.
HA: Haver Analytics. Monthly reports.