Journal of Banking & Finance 37 (2013) 2183–2195

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

# Assessing systemic risks and predicting systemic events

Marco Lo Duca, Tuomas A. Peltonen *

*European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany*

ARTICLE INFO

*Article history:*
Available online 7 July 2012

*JEL classification:*
E44
E58
F01
F37
G01

*Keywords:*
Early warning system
Systemic risk
Financial stress
Financial crisis
Macro-prudential policy

ABSTRACT

The paper develops a framework for assessing systemic risks and for predicting systemic events, i.e. periods of extreme financial instability with potential real costs. It contributes to the literature on the prediction of financial crises mainly in two ways: first, it uses a Financial Stress Index for identifying the starting date of systemic financial crises. Second, it uses discrete choice models that combine both domestic and global indicators of macro-financial vulnerabilities to predict systemic financial crises. The performance of the models is evaluated in a framework that takes into account policy maker's preferences between missing crises and issuing false alarms. Our analysis shows that combining indicators of domestic and global macro-financial vulnerabilities substantially improves the models' ability to forecast systemic financial crises. Our framework also displays a good out-of-sample performance in predicting the ongoing Global Financial Crisis.



## 1. Introduction

The Global Financial Crisis that started in the United States in 2007 has demonstrated the importance of understanding and measuring systemic risks and predicting systemic events, i.e. events *when financial instability becomes so widespread that it impairs the functioning of the financial system to the extent that economic growth and welfare suffer materially*.[1]

This paper develops a framework for assessing systemic risks and for predicting (out-of-sample) systemic events, i.e. periods of extreme financial instability with potential real costs.

The prediction of financial crises has been the subject of a large number of studies since the mid 1990s. In one of the earliest contributions, Frankel and Rose (1996) study the determinants of currency crashes in 100 developing countries from 1971 to 1992. They evaluate the predictive power of several indicators by looking at each indicator separately and at set of indicators jointly using a probit model. Their findings suggest that currency crashes tend to occur when FDI inflows dry up, when foreign exchange reserves are low, when domestic credit growth is elevated, when the real exchange rate is overvalued and when the "northern" interest rate rise.[2]

While the paper of Frankel and Rose (1996) is an important contribution to the early warning system (EWS) literature, it has two limitations. First, it focuses on currency crises only. Second, the paper lacks a clear framework to assess the leading properties of the indicators and to issue early warning signals.[3] These limitations are taken care of in Kaminsky and Reinhart (1999) who extend the analysis of Frankel and Rose to a wider set of crises, including banking and balance of payment crises that occurred in the 1990s. Kaminsky and Reinhart find that both types of crises are closely linked to the aftermath of financial liberalisation, which activates boom/bust cycles with banking crises preceding a currency collapse. An important contribution of the paper is the introduction of the so-called "signal" approach to evaluate the leading properties of indicators. In the approach, a variable signals an incoming crisis when it exceeds a predefined threshold. Correct signals (signals followed by a crisis) and wrong signals (signals not followed by a crisis or "noise") are collected and thresholds assigning signals to classes are optimised by

* Corresponding author. Tel.: +49 69 1344 8705; fax: +49 69 1344 6950.
*E-mail addresses:* marco.lo_duca@ecb.europa.eu (M.Lo Duca), tuomas.peltonen@ecb.europa.eu (T.A. Peltonen).

[1] See the definition of the concept of systemic risk in the ECB Financial Stability Review, December 2009 (ECB, 2009b). For a review of the concept of systemic risk see De Bandt and Hartmann (2000).

[2] Other papers document the "anomalous" behaviour of a number of variables in the periods preceding financial crises. See for example, Gavin and Hausmann (1996), Sachs et al. (1996), Mishkin (1996), Calvo (1996) and Honohan (2000).

[3] The paper simply presents a graphical analysis of the indicators in a time interval around crisis periods, while, regarding the probit model, it simply evaluates the significance of the coefficients.

0378-4266/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jbankfin.2012.06.010

minimising a noise to signal ratio. Finally, the indicators are ranked according to the noise to signal ratio.

The study of Kaminsky and Reinhart (1999) has, however, two limitations. First, in predicting crises, it does not use a multivariate framework that combines the information of the different indicators, as for example, a discrete choice model.[4] Second, due to the limited number of crises, there is not much scope for testing the out-of-sample performance of the leading indicators. Berg and Pattillo (2000), Edison (2003) and Berg et al. (2005) use the methodology of Kaminsky and Reinhart (1999)[5] and a more general probit model for the out-of-sample prediction of the Asian crisis with encouraging results. The key limitation of all these studies, however, is that they do not adopt a structured approach for the in-sample and out-of-sample evaluation of the early warning properties of the probit models.[6]

The evaluation of the performance of discrete choice models is addressed in Demirgüc-Kunt and Detragiache (2000), who use a multivariate logit model for the prediction of banking crises.[7] The main contribution of their paper is to show that considering policy maker's relative preferences between missing crises (Type I errors) and false alarms (Type II errors) is crucial to evaluate early warning models. Their paper shows that optimising early warning thresholds on the basis of the noise to signal ratio as in Kaminsky and Reinhart (1999) could lead to sub-optimal results under some preference schemes.[8] Therefore, the authors propose to select thresholds by minimising a loss function that takes into account policy maker's preferences between Type I and Type II errors.[9] Finally, Demirgüc-Kunt and Detragiache (2000) apply this approach to select optimal early warning thresholds for the crisis probability estimated with a discrete choice model.

Other features were introduced to early warning models in subsequent years. Bussière and Fratzscher (2006) show that binomial discrete-dependent-variable models are subject to a so called post-crisis bias. This bias arises when no distinction is made between tranquil periods, when economic fundamentals are largely sound and sustainable, and crisis/post-crisis periods, when economic variables go through an adjustment process before returning to a more sustainable level or growth path. The authors show that the performance of early warning models improves when correcting this bias.[10]

In a recent paper, Alessi and Detken (2011) use the signal approach to test the leading properties of real and financial variables in predicting costly asset price boom/bust cycles in a framework that takes into account policy maker's preferences between Type I and Type II errors. The main contribution of their paper is the signal evaluation framework and analysis of the role of global variables in predicting financial crises. The authors' results show that global measures of liquidity, in particular a global private credit gap, outperform domestic variables.

This paper builds upon the above studies and extends the existing literature on predicting financial crises mainly in two ways. First, we adopt an alternative approach for the identification of the starting date of systemic financial crises, which is crucial for the calibration of early warning models. By doing so, we first note that the approach normally employed in the literature relies on qualitative information and judgement. In Laeven and Valencia (2008), for example, a systemic banking crisis is defined as a period when defaults are widespread, non-performing loans increase and the capital of the banking system is exhausted.[11] While this definition is indeed a good description of the symptoms of a banking crisis, it leaves to judgement the identification of the starting date of the crisis.

The paper proposes to overcome this problem by using a composite index measuring the level of stress in the financial system of one country to identify the starting date of a systemic financial crisis in a more objective way. The start of the crisis coincides with the Financial Stress Index exceeding a predefined threshold, which in the past, anticipated real economic downturns with output losses.[12] Our approach to identify systemic events can be seen as an extension of Eichengreen et al. (1995, 1996), who use an index of exchange market pressure to identify currency crises. Compared to Eichengreen et al. (1995, 1996) our Financial Stress Index is broader than the exchange market pressure index, because it includes also other market segments. This enables us to identify episodes that are truly systemic, in the sense that many market segments are affected, and not specific to a single market segment. In addition, we define systemic financial crises or systemic events as episodes of extreme financial stress with potential real economic consequences. In this way, we focus on financial crises that are relevant for policy makers, who want to avoid real economic costs. The real cost dimension is absent in Eichengreen, Rose and Wyplosz, where a simple statistical rule is used to identify crisis periods.[13]

The second contribution of this paper is that, in predicting systemic events, we combine domestic and global indicators of macro-financial vulnerabilities in multivariate discrete choice models. While recent studies show that global variables are important determinants of domestic financial instability (Borio and Drehmann, 2009; Alessi and Detken, 2011), approaches in the earlier literature looked only at the leading properties of domestic indicators (Kaminsky and Reinhart, 1999; Demirgüc-Kunt and Detragiache, 2000; Berg and Pattillo, 2000; Borio and Lowe, 2002, 2004; Edison, 2003; Berg et al., 2005; Bussière and Fratzscher, 2006; Schularick and Taylor, 2011 and Jordá et al., 2011). Our paper combines both domestic and global indicators, as well as their interactions, in an early warning framework. To our knowledge, only Frankel and Rose (1996) include global variables in addition to domestic variables in their probit model. However, they include only GDP growth and interest rates in advanced economies. Compared to Frankel and Rose, our paper includes a larger set of global

4 Kaminsky (1998) proposes a leading composite indicator of financial crises by calculating an average of a set of indicators weighted by their noise to signal ratio.

5 In their paper, Berg and Pattillo refer to Kaminsky et al. (1998).

6 In the out-of-sample exercise, Berg and Pattillo arbitrary set the threshold for a crisis signal at 50% and 25% of the crisis probability estimated with the probit model.

7 Recently, Schularick and Taylor (2011) and Jordá et al. (2011) proposed alternative evaluation methods for discrete choice models.

8 In particular, if baking crises are rare events and the cost of missing a crisis is high relative to the cost of issuing a false alarm, then minimising the noise to signal ratio could lead to too many missed crisis. As a consequence, the selected threshold could be not optimal from the point of view of the preferences of policy makers.

9 Recently, other papers embedded policy maker's preferences in the design of early warning models. Bussière and Fratzscher (2008) show that the design of an "optimal" early warning model depends on policy maker's aversion to fail to anticipate the events, the forecast horizon of the model, and the probability threshold for extracting warning signals. In particular, they show that for a given degree of risk aversion, there is a unique combination of the forecast horizon and of the probability threshold that maximizes the policymaker's preferences, yielding the best possible model from a policy perspective.

10 They correct the post crisis bias by using a multinomial logit model with three regimes: crisis, recovery and normal period. However, Bussiere and Fratzscher do not adopt a structured approach for the evaluation of the discrete choice models, as for example, the method proposed by Demirgüc-Kunt and Detragiache (2000). They simply set an arbitrary threshold for a crisis signal at 20% of the crisis probability estimated with the logit model.

11 A working definition of crisis similar to the one of Laeven and Valencia (2008) is adopted in several other studies, including Kaminsky and Reinhart (1999), Demirgüc-Kunt and Detragiache (2000), Berg and Pattillo (2000), Borio and Lowe (2002, 2004), Edison (2003), Berg et al. (2005), Bussière and Fratzscher (2006), Reinhart and Rogoff (2008, 2009), Schularick and Taylor (2011) and Jordá et al. (2011).

12 We discuss the construction of the financial stress index and the selection of the threshold in the next section.

13 The index of Eichengreen, Rose and Wyplosz is calculated as equal variance weighted average of exchange rate changes, interest rate changes, and reserve changes. Crises are defined as periods when the pressure index is at least two standard deviations above the mean.

indicators and studies the impact of the interaction between domestic and global variables on financial stability.[14]

Our empirical analysis covers a set of 28 emerging market and advanced economies with quarterly data between 1990 Q1 and 2009 Q4. In predicting the systemic events identified with the Financial Stress Index, we evaluate the early warning properties of a number alternative discrete choice models (including and excluding global variables), and compare them to stand alone macro-prudential indicators of vulnerabilities (as in Kaminsky and Reinhart, 1999). The evaluation of all the indicators is done with a methodology that takes into account policy maker's preferences between Type I and Type II errors, and is implemented as in Alessi and Detken (2011).

Our results highlight the importance of considering jointly various indicators in a multivariate framework, as we find that discrete choice models outperform stand alone indicators in predicting systemic events. We find that combining indicators of domestic and global macro-financial vulnerabilities substantially improves the ability to forecast systemic events. In addition, considering interactions between domestic and global macro-financial vulnerabilities further enhances the performance of the models.

Our framework displays a good out-of-sample performance in predicting the ongoing Global Financial Crisis. In fact, our model would have issued an early warning signal for the United States in 2006 Q2, five quarters before the emergence of the tensions in money markets that started the crisis in August 2007.

The remainder of the paper is organised as follows. Section 2 introduces the measure for financial stress and defines systemic events. Section 3 describes the data and the methodology, and presents the empirical analysis, while Section 4 concludes.

## 2. Measuring financial stress and identifying systemic events

### 2.1. Construction of the Financial Stress Index

To identify systemic events, we construct a *Financial Stress Index (FSI)* for each of the countries in our sample, and we analyse the joint occurrence of elevated levels of financial stress and economic downturns.

Typically, when negative shocks, such as bursts of asset price bubbles, or banking, financial or currency crises hit the economy, tensions in several financial market segments emerge. Normally, tensions coincide with bankruptcies of financial institutions or precede them if the market anticipates the occurrence of bank distress. The larger and broader the distress is (i.e. the more systemic the shock is), the higher the co-movement among variables reflecting tensions. Therefore, by aggregating variables measuring stress across markets segments, we calculate a Financial Stress Index that captures the beginning and the evolution of a crisis.[15]

Our FSI is a country-specific composite index, covering the main segments of the domestic financial market, and it consists of the following five components: (1) the spread of the 3-month interbank rate over the 3-month Government bill rate ($\text{Ind}_1$)[16]; (2) negative quarterly equity returns (multiplied by minus one, so that negative equity returns increase financial stress; positive returns are disregarded and set to 0) ($\text{Ind}_2$); (3) the realised volatility of the main equity index ($\text{Ind}_3$); (4) the realised volatility of the nominal effective exchange rate ($\text{Ind}_4$); and (5) the realised volatility of the yield on the 3-month Government bill ($\text{Ind}_5$).[17]

Each component $j$ of the index for country $i$ at quarter $t$ is transformed into an integer that ranges from 0 to 3 according to the country-specific quartile of the distribution the observation at quarter $t$ belongs to ($q_{j,i,t}$). For example, a value for component $j$ falling into the fourth quartile of the distribution would be transformed into 3.[18] Note that each variable is measured in a way that higher values indicate higher stress levels, therefore higher values of the transformed variables indicate higher stress.

The Financial Stress Index is computed for country $i$ at time $t$ as a simple average of the transformed variables as follows:

$$\text{FSI}_{i,t} = \frac{\sum_{j=1}^{5} q_{j,i,t}(\text{Ind}_{j,i,t})}{5} \tag{1}$$

Hollo et al. (2012) show that this standardization method based on quartiles is more robust than a standardization based on mean and variance, especially when the number of components of the index is small. More specifically, with the quartile standardization method, adding new observations to the sample produces only small revisions to the historical levels of the index (*ex post* stability). Large revisions of the historical levels of the index would complicate the analysis of the Financial Stress Index and its use in econometric models.[19]

In calculating the Financial Stress Index, we face a trade off between the degree of precision of the index at the country level and the degree of homogeneity of the index across countries and time. For some countries with more developed financial systems, it would be possible to calculate a more detailed Financial Stress Index aggregating the information from several financial instruments and several markets segments. The set of instruments and segments is, however, limited for the emerging economies that are included in our sample. Since in our study the cross-country dimension prevails, we give more importance to the homogeneity of the Financial Stress Index across countries. We believe this does not affect our results mainly for two reasons. First, once some crucial segments of the financial system are included in the index, adding components to it does not substantially change the shape of financial stress indices (Hollo et al., 2012). Second, our focus is on the detection of systemic events, i.e. we look only at extreme values of financial stress. Identified extreme values are robust to the composition of the FSI and also to the standardization method.[20]

[14] In addition, compared to Frankel and Rose (1996), our analysis is not limited to currency crashes and adopts a system to evaluate the early warning properties of discrete choice models, as described in the next paragraphs.

[15] Several indices for the measurement of financial tensions have been developed in recent years, examples are Illing and Liu (2006), Misina and Tkacz (2009), and Hakkio and Keeton (2009). The IMF (2008, 2009) presented the work by Cardarelli et al. (2011) and Balakrishnan et al. (2009), who constructed financial stress indices for a broad set of advanced and emerging economies. The ECB (2009a) presented a Financial Stress Index for the global economy. It should be highlighted that research on the measurement of financial stress and the construction of indices capturing systemic events is currently very active (see e.g. Hollo et al., 2012).

[16] Or the spread among the interbank and the T-bill rates of the closest maturity to 3months when the latter maturity is not available.

[17] In the calculation of realised volatilities for equity, nominal effective exchange rate and Government bill rate, i.e. components ($\text{Ind}_3$) to ($\text{Ind}_5$), average daily absolute changes over a quarter were used.

[18] The only exception to this standardisation method is the indicator for negative stock market returns. As this indicator is most of the time equal to zero, the standardisation through quartiles leads often to a very volatile variable jumping from 0 to 3 directly. Therefore to standardise this variable we just divide this indicator by its maximum value over the sample. We then rescale the transformed indicator so that it ranges from 0 to 3.

[19] Hollo et al. (2012) discuss advantages and disadvantages of approaches for the calculation of financial stress indices.

[20] The financial stress indices for countries in the sample are plotted in Figs. A1 and A2 in the Appendix A. As it can be seen from the Figures, the FSIs capture well past episodes of high financial stress or crises, such as the Asian financial crisis in 1997, the Russian crisis in 98, the burst of the IT bubble in 2000–2001 and the last Global Financial Crisis. For many advanced economies, the Global Financial Crisis led to the highest level of financial stress since the start of the sample in 1990, while in many emerging economies the level of financial stress was higher during the Asian financial crisis, or during some country-specific crisis, such as the Russian crisis of 1998 or the crisis in Argentina of 2001.

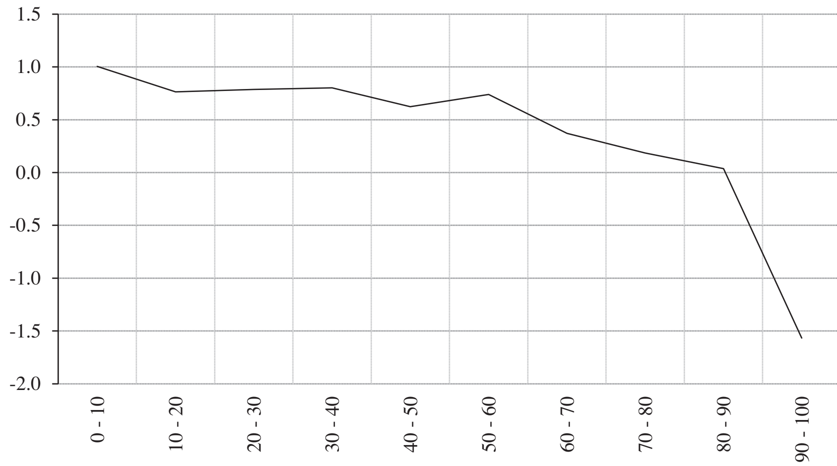


**Fig. 1.** Level of the Financial Stress Index at quarter *t* and the median deviation of real GDP from trend at quarter *t* + 2. Note: The *X*-axis represents the percentile of the country distribution of the Financial Stress Index at time *t*, while the *Y*-axis represents the median real GDP deviation from its trend at time *t* + 2, measured in percent of the trend.

### *2.2. Identification of systemic events*

Policy maker's main concern regarding financial stress is that financial instability could become so widespread that it impairs the functioning of the financial system to the extent that economic growth and welfare suffer materially. Therefore, in this study, we focus on episodes of extreme financial stress that have often (i.e. on median cases) been followed by negative economic developments. We define these episodes as systemic events.[21]

To identify systemic events, we analyse the relationship between the Financial Stress Index and measures of real economic activity. Fig. 1 reports the median deviation (in percent) of the real GDP from its trend[22] (i.e. output gap) for different percentiles of the distribution of the Financial Stress Index (two quarters ahead). As it can be seen from the figure, the levels of the Financial Stress Index above the 90th percentile of the country distribution of the index anticipate median negative deviations of the real GDP from its trend (i.e. economic slowdowns or recessions).[23]

Based on the graphical analysis, in our benchmark case, we identify systemic events when the Financial Stress Index exceeds the 90th percentile of the country distribution. Following this approach, we identify 94 systemic events in our sample. We find the following starting dates for well-known crisis episodes in the 1990s and 2000s: 1994 Q1 for Brazil, 1994 Q4 for the Mexican crisis; 1997 Q2 for the Asian crisis in Thailand, 1997 Q3 for Hong Kong and other main Asian countries, 1998 Q3 for the Russian crisis, 1999 Q1 for the Brazilian crisis; 2001 Q3 for the Argentinean crisis; 2007 Q3 for the last financial crisis in the United States. In many cases, these crisis episodes spread to several other economies. For example, after starting in 2007 Q3 with severe problems in money markets and volatility in other market segments in the United States and in the euro area, the latest crisis spread in successive waves across countries in 2008 Q1, 2008 Q3, and finally it reached emerging markets in 2008 Q4. Several of the episodes that we identify are also identified by earlier studies and are also in the list of crises compiled by Laeven and Valencia (2008).

## 3. Empirical analysis

### *3.1. Definition of the dependent variable*

The objective of the study is to predict the occurrence of systemic events within a given time horizon that in our benchmark specification is set to 6 quarters.[24] To do this, we proceed by defining the dependent variable using the following three steps:

First, we transform the Financial Stress Index into a binary variable that we call systemic event. The variable takes value 1 in the quarter, when the FSI moves above the predefined threshold of the 90th percentile of the country distribution.

Second, we set the dependent variable to 1 in the 6 quarters preceding the systemic event and to 0 in all the other periods. The dependent variable mimics the ideal leading indicator that perfectly signals systemic events in the 6 quarters before the event.

Finally, we drop from the sample all observations that are not informative about the transition from tranquil times to systemic events. It means that we delete the subsequent time periods, when financial stress remains above the predefined threshold after the identified systemic events. We also drop tranquil periods that are shorter than 6 quarters, as the short period between the extreme stress episodes suggests that the subsequent episodes might simply be the continuation of the first one.[25]

[21] By defining systemic events only those episodes when financial stress led to negative real economic consequences, we would incur in a selection bias. This bias emerges if a policy action undertaken after the occurrence of extreme financial stress prevented the negative economic outturn. To avoid this bias, we define systemic events those episodes of extreme financial stress that in the past had high probability of being followed by real consequences.

[22] The trend is calculated using Hodrick–Prescott filter with the smoothing parameter set to 1600.

[23] A level of stress above the 90th percentile anticipates a significant slowdown in economic activity that lasts up to 5 quarters in the median case. During the period while GDP remains below the trend, the median cumulated costs range between 4% and 5% of the GDP.

[24] The time horizon of 6 quarters is chosen because within this time interval policy makers can adopt measures to prevent the materialisation of systemic events. Shorter time horizons are less relevant for policy making because the potential for effective pre-emptive actions is lower. For robustness, we use time horizons of 2, 4 and 8 quarters. The results are discussed in the section on the robustness tests.

[25] Bussière and Fratzscher (2006) point out that including in the estimation of early warning models the period of economic recovery after a crisis produces the so called "post crisis bias". In recovery periods, economic variables go through an adjustment process before reaching again the path they have during tranquil periods. The recovery period therefore should be excluded from the analysis as it is not informative of the path leading from the pre-crisis regime to the crisis. Bussière and Fratzscher address this issue by using a multinomial logit model with "three regimes" for the dependent variable (calm period, crisis and recovery). In our paper, as we drop periods in which stress is high, we already disregard recovery periods, at least partially. However, we check the robustness of our results by dropping observations up to two quarters after the end of the stress periods to ensure that the post crisis bias is addressed. Only marginal gains in the performance of the model are obtained when dropping the additional two quarters.

**Table 1**
List of variables.

| Description | Domestic variables | | | | | | Global average |
|---|---|---|---|---|---|---|---|
| | 1 Level | 2 Deviation from moving average (short) | 3 Deviation from moving average (long) | 4 Annual change | 5 Deviation from trend (short) | 6 Deviation from trend (long) | 7 |
| Ratio money to GDP | | x | x | | x | x | x |
| Ratio m2 to GDP | | x | x | | x | x | x |
| Real effective exchange rate | | x | x | x | x | x | |
| Nominal effective exchange rate | | x | x | x | x | x | |
| Real GDP | | x | x | x | x | | x |
| Consumer price index | | x | x | x | x | | x |
| Ratio credit to the private sector to GDP | | x | x | | x | x | x |
| Real money | | | | x | | | x |
| Real M2 | | | | x | | | x |
| Real house prices | | | | x | | | x |
| Real equity prices (MSCI based | | | | x | x | x | x |
| Real credit to the private sector to GDP | | | | x | | | x |
| General government debt (% of GDP) | x | | | | | | |
| General government deficit (% of GDP) | x | | | | | | |
| Current account deficit (% of GDP) | x | | | | | | |
| Price-earning ratios | x | | | | x | x | x |
| Stock market capitalisation over GDP | | | | | x | x | x |

*Notes*: The table lists the core set of variables used in the empirical analysis; other transformations of the variables (i.e. interactions among them) that are used in the analysis are described in the main text. Real variables have been calculated by deflating the original nominal variable by the consumer price index (CPI). The first column reports the description of the variable; column (1) indicates whether the level of the original variable is used in the analysis; columns (2 and 3) indicate whether percentage deviations from short (8 quarters) and long (20 quarters) moving averages of the variable are used; column (4) indicates whether the annual percentage change of the original variable is used in the analysis; columns (5 and 6) indicate whether percentage deviations from Hodrick–Prescott trends of the variable are used; the "short" ("long") Hodrick–Prescott trend is computed with the smoothing parameter $\lambda$ set to 1600 (400,000) following Borio and Lowe (2004); The last column (7) indicates whether global averages have been computed for all the transformations of the variables listed in columns (1–6).

### *3.2. Explanatory variables: indicators of domestic and global macro-financial vulnerabilities*

To assess the level of systemic risks and to predict systemic events, we construct indicators commonly used in the literature to predict crises (Kaminsky and Reinhart, 1999; Borio and Lowe, 2002, 2004, and Alessi and Detken, 2011). These indicators capture the building up of vulnerabilities and imbalances both in the domestic and global economy. In this regard, we focus on asset price and credit developments and valuation levels, as well as proxies for leverage in the economy. In addition, we also calculate indicators for macroeconomic conditions (real GDP growth, inflation, current account balance, government balance and debt).

We build a comprehensive dataset of quarterly macro and financial data for the period 1990 Q1–2009 Q4 for 28 countries, of which 10 advanced countries and 18 emerging economies.[26] The data is obtained either from Haver Analytics, Bloomberg and Datastream.

Table 1 summarises the core variables included in the study.

Following the literature (e.g. Alessi and Detken, 2011), we test several transformations of the indicators, such as annual changes and deviations from moving averages or trends.[27] To proxy for global macro-financial imbalances and vulnerabilities, we calculate a set of global indicators by averaging the transformed variables for the following four countries or regions: the United States, euro area, Japan and the United Kingdom.[28]

Starting from the core set of variables listed in Table 1, we calculate interactions among domestic variables, among international variables and between domestic and international variables. The interactions are aimed to capture the joint dynamic of two indicators.

Our analysis is conducted as much as possible in a real-time analysis fashion.[29] At each point in time, only information available to the policy makers up to that point in time is used. This implies that we take into account that certain variables, as for example GDP, are not available in real time because of publication lags. To take into account publications lags, we used lagged variables. For GDP, money and credit related indicators the lag ranges from 1 to 2 quarters depending on the country and on the indicator.

The real time analysis also implies that de-trended variables are computed using only real time information. Therefore, we recursively calculate trends at each time $t$, using only the information available up to that moment.

[26] The advanced countries are the following: Australia, Denmark, Euro area, Japan, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The emerging economies are the following: Argentina, Brazil, China, Czech Republic, Hong Kong, Hungary, India, Indonesia, Malaysia, Mexico, the Philippines, Poland, Russia, Singapore, South Africa, Taiwan, Thailand and Turkey.

[27] We estimate the trend with the Hodrick–Prescott filter. Following Borio and Lowe (2004), we use two different values of the smoothing parameter, namely 1600 and 400,000.

[28] We also calculated global averages by using weighted GDP averages of all countries in our sample. We only report the results for the global variables calculated using the United States, euro area, Japan and the United Kingdom as the results are substantially the same.

[29] The literature on early warning models deals with large datasets of macro data for several countries of which several are emerging markets. "Real time datasets" that contain information on the revisions of data after the first publication do not exist yet for several countries in our sample. Our analysis is therefore a real-time analysis in the sense that we take into account publication lags, as in other early warning models (Alessi and Detken, 2011).

### 3.3. Evaluation of the predictive power of the indicators and calculation of the optimal early warning threshold

To evaluate the performance of the indicators in predicting systemic events, we use a method that takes into account policy maker's preferences between issuing false alarms and missing systemic events[30] to rank the indicators and find optimal early warning thresholds. The approach is used to evaluate the performance of both, individual stand alone indicators and the discrete choice models, the latter being evaluated, in the current literature, on the basis of arbitrary thresholds (with the exception of Demirgüc-Kunt and Detragiache, 2000).

First, following the method proposed by Alessi and Detken (2011), under a given threshold for issuing an early warning, we define a loss function that depends on the preferences of the policy maker with respect to Type I and Type II errors:

$$L(\mu) = \mu(\text{Type I}) + (1 - \mu)(\text{Type II}) \tag{2}$$

More specifically, Type I is the ratio of missing signals (i.e. when no early warning signal was issued despite a crisis occurred) to the number of periods when a signal should have been issued, while Type II is the ratio of wrong signals (i.e. when a signal was issued while no crisis occurred) to the number of periods when no signal should have been issued. Finally, the parameter $\mu$ describes the relative preference of the policy maker between Type I and Type II errors.

Second, we define usefulness $U$ the gain that the policy maker obtains by using the indicator as compared to ignoring it:

$$U = Min[\mu, 1 - \mu] - L(\mu) \tag{3}$$

where $Min\ [\mu,\ 1 - \mu]$ is the loss that the policy maker faces when disregarding the indicator (i.e. either she assumes that a signal is never issued or that the signal is always issued).

Optimal early warning thresholds are calculated by maximising $U$ and indicators are ranked according to the maximum $U$ they achieve.

Regarding policy maker's preferences with respect to Type I and II errors, in our benchmark analysis we take the point of view of a policy maker, who is equally concerned of issuing false alarms and missing systemic events, i.e. we assume that $\mu = 0.5$. This could be considered the point of view of a neutral external observer who does not want to commit any mistake and is only concerned of correctly anticipating a systemic event.[31]

### 3.4. Logit models

In our analysis, the country specific probability of a systemic event, i.e. systemic risk, is a function of financial vulnerabilities that are shown to perform well as stand alone indicators for predicting crises. In our benchmark model, the explanatory variables are grouped into three main sets, namely the domestic, the global and the interactions between domestic and global factors.

The first set consists of variables that measure domestic conditions and vulnerabilities. It includes growth in domestic asset prices (equity) and bank credit, asset price valuation levels, and the level of leverage in the economy. In our benchmark specification, growth in equity prices and bank credit are measured by the real (net of inflation) annual growth of the local MSCI equity index and of the amount outstanding of credit granted to the private sector. Asset price valuations are measured by the deviation of the ratio equity market capitalisation to GDP from its trend, while leverage is measured as the deviation of the ratio private credit to GDP from its trend.[32] The domestic set of variables also includes the interaction between asset price developments and valuation levels, as well as the interaction between credit growth and leverage. The interactions are computed by the product of the two relevant variables. Finally, the domestic macroeconomic environment is controlled for with the following variables: annual real GDP growth, annual CPI inflation, current account deficit in percentage of GDP, and government deficit in percentage of GDP.

The second set of explanatory variables aims at capturing the global macro-financial environment. These variables are included because the recent literature (Borio and Drehmann, 2009; Alessi and Detken, 2011) and our empirical analysis of stand alone indicators of vulnerabilities show that global factors have a significant influence on domestic financial stability. Similarly to the domestic set of variables, we include growth in global asset prices and bank credit, global asset price valuation levels, and the global level of leverage to the model. In addition, the set of explanatory variables also includes the interaction between global asset price developments and valuation levels, as well as the interaction between global credit growth and leverage. Finally, global macroeconomic conditions are captured by real GDP growth and inflation.

The third set of explanatory variables includes the interplay between domestic and global indicators of vulnerabilities, computed as the product between the relevant domestic and international variables. The introduction of this group of variables captures additional fragilities that emerge when the overheating of the domestic economy coincides with the vulnerabilities in the global conditions.

In the robustness section, we evaluate our results by changing the specification of the benchmark model and the variables used to measure the different vulnerabilities.

Regarding the estimation strategy, due data limitations, we pool the information of our unbalanced panel, and assume that the constant $c$ and the slope coefficients $\beta$ of the logit model do not change across time and countries. The appropriateness of a pooled approach is discussed by Demirgüc-Kunt and Detragiache (1998), Fuertes and Kalotychou (2006) and Davis and Karim (2008).[33]

To take into account potential cross country differences in the scale of the regressors, as well as to avoid that our results are affected by large outliers, we follow the method by Berg et al. (2005), and measure variables in country specific percentile terms.

Table 2A reports the estimated coefficients for the benchmark model (column 5), that includes the set of explanatory variables

[30] Normally, the threshold for an indicator is chosen based on some kind of information criteria, e.g. noise-to-signal ratio. Demirgüc-Kunt and Detragiache (2000), Bussière and Fratzscher (2008) and Alessi and Detken (2011) highlight that this approach has several drawbacks.

[31] The point of view of local policy makers or international institutions in charge of giving policy recommendations could be different, as the costs of missing systemic events and issuing false alarms are different (e.g. through reputational costs or real costs). It is likely that the last financial crisis increased the concerns of policy makers of missing systemic events. However, it is difficult to assess whether policy makers could be assumed to be relatively more concerned of missing crises versus issuing false alarms. For a more comprehensive discussion of the issue, see Bussière and Fratzscher (2008) and Alessi and Detken (2011).

[32] Trends are computed with the Hodrick-Prescott filter setting the smoothing parameter $\lambda$ to 400,000. Regarding, equity valuations it would be optimal to use price earning ratios, however time series for these data are not available since 1990 for a large portion of our set of countries. Therefore, we opted to use the ratio equity market capitalisation to GDP as a proxy for valuations after de-trending the ratio to correct for the non-stationarity due the progress in developing local stock markets. Regarding leverage, the deviation from the trend of the ratio private credit to GDP is a commonly used measure of leverage (Borio and Lowe, 2002), against the background of the lack of uniform coverage across countries and time of data on leverage of financial intermediaries, households and corporations.

[33] As Davis and Karim (2008) state, a fixed effects model would mean that the estimated country-specific effect and the systemic event indicator would be perfectly correlated also for countries that never experienced a systemic event, while excluding these countries would generate a biased sample and biased coefficients. Thus a pooled logit model is used in this study.

**Table 2A**
Estimation results of the logit models.

| | | (1) Currency crisis model | (2) Macro-prudential model | (3) Domestic model | (4) Domestic and international (no interactions) | (5) Benchmark | (6) Benchmark (marginal effects |
|---|---|---|---|---|---|---|---|
| Domestic variables | Real GDP growth | $0.0152^{***}$ | $0.0069^{**}$ | 0.0047 | 0.0057 | $0.0066^{*}$ | $0.0008^{*}$ |
| | Inflation | 0.0027 | $0.0082^{***}$ | $0.0070^{**}$ | 0.0050 | 0.0061 | $0.0008^{*}$ |
| | Current account deficit | 0.0012 | $0.0060^{**}$ | $0.0070^{***}$ | $0.0079^{**}$ | $0.0075^{*}$ | $0.0009^{**}$ |
| | General government deficit | | | $-0.0073^{**}$ | −0.0032 | −0.0011 | −0.0001 |
| | Real effective exchange rate overvaluation | 0.0014 | | | | | |
| | Real equity growth | | $0.0049^{*}$ | 0.0015 | $0.0089^{***}$ | $0.0188^{***}$ | $0.0023^{***}$ |
| | Equity valuation | | $0.0291^{***}$ | $0.0266^{***}$ | $0.0142^{***}$ | $0.0119^{***}$ | $0.0015^{***}$ |
| | (A1) equity interaction (growth & valuation) | | | $0.0063^{**}$ | | 0.0045 | 0.0006 |
| | Real credit growth | $0.0107^{***}$ | 0.0011 | −0.0011 | 0.0013 | −0.0068 | −0.0008 |
| | Leverage | $0.0195^{***}$ | $0.0222^{***}$ | $0.0160^{***}$ | $0.0180^{***}$ | $0.0105^{*}$ | $0.0013^{*}$ |
| | (B1) credit interaction (growth & leverage) | | | 0.0032 | | $0.0135^{***}$ | $0.0017^{**}$ |
| | Interaction leverage & valuation | | | $0.0067^{**}$ | | | |
| | Interaction equity & credit growth | | | 0.0019 | | | |
| Global variables | Real GDP growth | | | | 0.0015 | 0.0004 | 0.0000 |
| | Inflation | | | | $0.0170^{***}$ | $0.0303^{***}$ | $0.0038^{***}$ |
| | Real equity growth | | | | −0.0014 | $-0.0123^{***}$ | $-0.0015^{**}$ |
| | Equity valuation | | | | $0.0396^{***}$ | $0.0347^{***}$ | $0.0043^{***}$ |
| | (A2) equity interaction (growth & valuation) | | | | | $0.0133^{***}$ | $0.0017^{***}$ |
| | Real credit growth | | | | $0.0146^{***}$ | 0.0093 | 0.0012 |
| | Leverage | | | | 0.0032 | $0.0415^{***}$ | $0.0052^{***}$ |
| | (B2) credit interaction (growth & leverage) | | | | | $-0.0430^{***}$ | $-0.0054^{**}$ |
| Interaction between domestic and global variables | Interaction domestic & international leverage | | | | | $0.0078^{**}$ | $0.0010^{*}$ |
| | Interaction domestic & international valuations | | | | | −0.0029 | −0.0004 |
| | Interaction domestic & international credit growth | | | | | $-0.0074^{**}$ | $-0.0009^{**}$ |
| | Interaction domestic & international equity growth | | | | | $0.0168^{***}$ | $0.0021^{***}$ |
| | A1 × A2 | | | | | $-0.0124^{**}$ | $-0.0015^{**}$ |
| | B1 × B2 | | | | | −0.0008 | −0.0001 |
| | Constant | $-3.7562^{***}$ | $-5.1805^{***}$ | $-4.8874^{***}$ | $-8.4201^{***}$ | $-9.2450^{***}$ | |
| | Number of countries | 28 | 28 | 28 | 28 | 28 | |
| | Number of observations | 902 | 1275 | 1275 | 1275 | 1275 | |
| | Pseudo $R$-squared | 0.1278 | 0.1903 | 0.2036 | 0.3398 | 0.3894 | |

*Notes*: $\mu$ = 0.5 and forecasting horizon 6 quarters. Robust standard errors have been used in the estimation.
* Statistical significance at 10% level.
** Statistical significance at 5% level.
*** Statistical significance at 1% level.

described above, as well as statistics for alternative models that are used for comparison (columns 1–4).[34] The alternative models are:

1. "*Currency crisis*" model: it includes explanatory variables often used in the currency crises literature, as for example the real exchange rate, macro-conditions and credit growth.
2. "*Macro-prudential*" model: it adds equity price growth and valuation to the set of explanatory variables of the "Currency crisis" model.[35]
3. "*Domestic model*": it includes the explanatory variables of the "Macro-prudential" model with the addition of (i) the general government deficit, (ii) the interaction between equity growth and equity valuation, and (iii) the interaction between credit growth and leverage.
4. "*Domestic and international (no interactions)*" model: it includes the explanatory variables of the "Domestic" model with the exclusion of the interactions terms. It also includes global growth and inflation, as well as global credit growth and leverage, and global equity growth and valuation.

Table 2A also includes the estimated marginal effects of the independent variables in the benchmark model (column 6). Finally Table 2B presents the estimated coefficients for the "Domestic", "Domestic and international (no interactions)" and the "Benchmark" models estimated alternatively on emerging markets only or advanced economies only.

As the main objective of the paper is to evaluate the performance of the models in predicting systemic events according to the framework that takes into account policy maker's preferences,

[34] Note that the correlation among the explanatory variables included in the benchmark specification is on average low (0.16 is the average absolute correlation between the explanatory variables). Only a small number of bilateral absolute correlations exceed 0.5 (8 in total) of which only 3 exceed 0.6.

[35] However, it excludes the real exchange rate due to data limitations.

**Table 2B**
Estimation results of the logit models.

| | | EMEs only | | | Advanced Economies only | | |
|---|---|---|---|---|---|---|---|
| | | (7) Domestic model | (8) Domestic and international (no interactions) | (9) Benchmark | (10) Domestic model | (11) Domestic and international (no interactions) | (12) Benchmark |
| Domestic variables | Real GDP growth | −0.0017 | 0.0015 | −0.0002 | 0.0032 | 0.0007 | −0.0027 |
| | Inflation | 0.0118$^{***}$ | 0.0164$^{***}$ | 0.0266$^{***}$ | 0.0108$^{*}$ | 0.0020 | −0.0036 |
| | Current account deficit | 0.0059$^{*}$ | 0.0118$^{***}$ | 0.0107$^{**}$ | 0.0131$^{***}$ | 0.0105$^{**}$ | 0.0154$^{***}$ |
| | General government deficit | −0.0106$^{**}$ | −0.0040 | −0.0018 | −0.0013 | −0.0004 | −0.0006 |
| | Real effective exchange rate overvaluation | | | | | | |
| | Real equity growth | −0.0018 | 0.0146$^{***}$ | 0.0313$^{***}$ | 0.0049 | 0.0124 | 0.0104 |
| | Equity valuation | 0.0185$^{***}$ | 0.0011 | 0.0036 | 0.0482$^{***}$ | 0.0335$^{***}$ | 0.0408$^{***}$ |
| | (A1) equity interaction (growth & valuation) | 0.0033 | | 0.0012 | 0.0131$^{***}$ | | 0.0076 |
| | Real credit growth | −0.0045 | 0.0043 | −0.0198$^{**}$ | −0.0012 | −0.0015 | −0.0163 |
| | Leverage | 0.0069 | 0.0201$^{***}$ | 0.0050 | 0.0588$^{***}$ | 0.0313$^{***}$ | 0.0486$^{***}$ |
| | (B1) credit interaction (growth & leverage) | 0.0153$^{***}$ | | 0.0362$^{***}$ | −0.0190$^{**}$ | | −0.0323$^{***}$ |
| | Interaction leverage & valuation | −0.0011 | | | 0.0107$^{*}$ | | |
| | Interaction equity & credit growth | 0.0060 | | | 0.0005 | | |
| Global variables | Real GDP growth | | −0.0003 | 0.0003 | | 0.0090 | 0.0169$^{**}$ |
| | Inflation | | 0.0182$^{***}$ | 0.0685$^{***}$ | | 0.0112 | 0.0141 |
| | Real equity growth | | −0.0062 | −0.0177$^{***}$ | | −0.0009 | −0.0096 |
| | Equity valuation | | 0.0403$^{***}$ | 0.0434$^{***}$ | | 0.0293$^{***}$ | 0.0227$^{***}$ |
| | (A2) equity interaction (growth & valuation) | | | −0.0021 | | | 0.0231$^{***}$ |
| | Real credit growth | | 0.0266$^{***}$ | 0.0004 | | 0.0000 | −0.0168 |
| | Leverage | | 0.0017 | 0.1195$^{***}$ | | 0.0113 | 0.0169 |
| | (B2) credit interaction (growth & leverage) | | | −0.1292$^{***}$ | | | 0.0045 |
| Interaction between domestic and global variables | Interaction domestic & international leverage | | | 0.0103$^{*}$ | | | 0.0059 |
| | Interaction domestic & international valuations | | | −0.0065 | | | −0.0077 |
| | Interaction domestic & international credit growth | | | −0.0020 | | | −0.0076 |
| | Interaction domestic & international equity growth | | | 0.0355$^{***}$ | | | 0.0322$^{***}$ |
| | A1 × A2 | | | −0.0181$^{**}$ | | | 0.0011 |
| | B1 × B2 | | | 0.0020 | | | 0.0016 |
| | Constant | −3.5639$^{***}$ | −9.2059$^{***}$ | −12.0821$^{***}$ | −8.8773$^{***}$ | −9.3149$^{***}$ | −10.3401$^{***}$ |
| | Number of countries | 17 | 17 | 17 | 11 | 11 | 11 |
| | Number of observations | 745 | 745 | 745 | 525 | 525 | 525 |
| | Pseudo $R$-squared | 0.1485 | 0.3191 | 0.4585 | 0.3700 | 0.3990 | 0.4490 |

Notes: $\mu = 0.5$ and forecasting horizon 6 quarters. Robust standard errors have been used in the estimation.
* Statistical significance at 10% level.
** Statistical significance at 5% level.
*** Statistical significance at 1% level.

we draw the attention of the reader only on a few features that emerge from Tables 2A and 2B.[36]

First, the fit of the Benchmark model is better than the fit of any alternative model (Table 2A). The Benchmark model is outperforming other models also when the estimation includes emerging markets or advanced economies only (Table 2B). Second, the information criteria support the choice of the Benchmark model and the inclusion of interaction terms.[37] Third, in the Benchmark model, domestic factors, as well as global factors and the interaction between domestic and global ones are statistically significant and have, in most cases, the expected signs (both Tables 2A and 2B). Fourth, the estimated coefficients capturing the impact of global variables as well as those capturing the interaction between domestic and global variables are, in most cases, larger and have higher statistical significance level in the model for emerging markets only (Table 2B) compared to the model for advanced economies only (and also compared to the benchmark in Table 2A). This suggests that the determinants of systemic risks are the same in emerging and advanced economies. The main difference between emerging markets and advanced economies is the relative importance of the different factors. In particular, emerging economies seem to be more exposed to global factors.[38] This is confirmed by the better fit of the

[36] In addition, due to non-linearities and data transformations, the interpretation of the estimated coefficients is not straightforward.

[37] The AIC criterion for the models is the following: "Macroprudential" AIC = 1235.8; "Domestic" AIC = 1228.2; "Domestic and international (no interactions)" AIC = 1041.2; "Benchmark" AIC = 986.454. Models for emerging markets: "Domestic and international (no interactions)" AIC = 623.9; "Benchmark" AIC = 522.0.

[38] This conclusion is supported by Dungey et al. (2010).

**Table 3**
In-sample performance of logit models.

| Model | Threshold (percentile) | $U$ | NtSr | % Predicted | Cond Prob (%) | Prob Diff (%) |
|---|---|---|---|---|---|---|
| Benchmark (AEs only) | 66 | 0.34 | 0.21 | 85.14 | 64.62 | 36.95 |
| Benchmark (EMEs only) | 65 | 0.33 | 0.22 | 84.73 | 63.24 | 35.99 |
| Benchmark | 68 | 0.32 | 0.20 | 80.95 | 65.83 | 37.83 |
| Benchmark (no interactions – AEs only) | 66 | 0.32 | 0.25 | 85.13 | 60.87 | 33.20 |
| Domestic model (AEs only) | 64 | 0.32 | 0.25 | 85.14 | 60.87 | 33.21 |
| Benchmark (no interactions) | 58 | 0.31 | 0.31 | 88.80 | 55.52 | 27.52 |
| Benchmark (no interactions – EMEs only) | 68 | 0.30 | 0.24 | 78.33 | 61.39 | 34.14 |
| Domestic model | 67 | 0.26 | 0.28 | 71.71 | 57.79 | 29.79 |
| Macro-prudential | 63 | 0.24 | 0.34 | 74.23 | 53.32 | 25.32 |
| Best stand alone indicator* | 55 | 0.21 | 0.45 | 76.91 | 48.44 | 18.72 |
| Domestic model (EMEs only) | 58 | 0.19 | 0.47 | 71.43 | 44.62 | 17.33 |
| Currency crisis | 69 | 0.19 | 0.38 | 60.33 | 49.16 | 22.33 |

*Notes*: *The best stand alone indicator is the percentage deviation from Hodrick–Prescott trend (with $\lambda$ set to 400,000) of the ratio equity market capitalisation to GDP in G4 countries. The optimal threshold is expressed as percentile on the basis of the country distribution of the indicator. The optimal threshold indicates the level at which the indicator issues an early warning signal. Thresholds are calculated for $\mu = 0.5$ ("neutral" observer) and forecasting horizon 6 quarters. The columns of the table report the following measures to assess the efficiency of the models: usefulness "$U$" (see formula 3); the noise to signal ratio (NtSr) i.e. the ratio between false signals as a proportion of periods in which false signals could have been issued and good signals as a proportion of periods in which good signals could have been issued; the percentage of crisis predicted by the indicator (% predicted) i.e. the ratio between good signals and the number of periods in which good signals could have been issued; the probability of a crisis conditional on a signal (Cond Prob) i.e. the ratio between good signals and the total number of signals issued; the difference between the conditional and the unconditional probability of a crisis (Prob Diff). See Kaminsky et al. (1998) for details on the calculations of the measures of efficiency.

**Table 4**
Out-of-sample performance of logit models.

| Model | $U$ | NtSr | % Predicted | Cond Prob (%) | Prob Diff (%) |
|---|---|---|---|---|---|
| Benchmark | 0.18 | 0.57 | 83.91 | 48.34 | 13.54 |
| Benchmark (no interactions) | 0.15 | 0.64 | 85.06 | 45.40 | 10.60 |
| Domestic model | 0.12 | 0.67 | 72.41 | 44.37 | 9.57 |
| Macro-prudential | 0.10 | 0.75 | 77.01 | 41.61 | 6.81 |
| Currency crisis | 0.06 | 0.84 | 77.01 | 38.95 | 4.15 |

*Notes*: See notes to Table 3. Due to limitations in the number of observations we only estimate in real time the model for all countries without differentiating between EMEs and advanced economies.

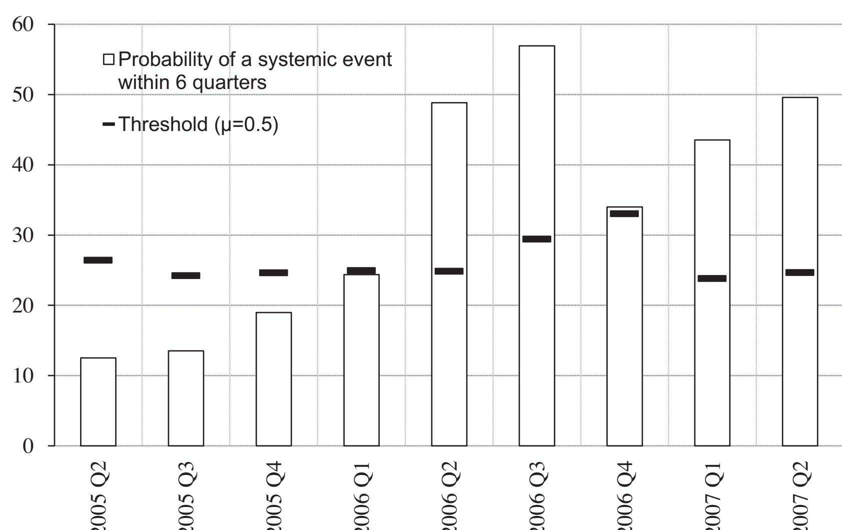


**Fig. 2.** Predicting the Global Financial Crisis in the United States. Out-of-sample performance of the Benchmark logit model in 2005 Q2–2007 Q2. Note: The *X*-axis represents time (in quarters), while the *Y*-axis represents the probability of a systemic event within the next 6 quarters (threshold optimised for $\mu = 0.5$). The probability is the output of the benchmark logit model.

models that include global factors compared to the "Domestic" model for emerging markets only (Table 2B) (while the increase in fit is only marginal for advanced economies).

We now turn to the evaluation of the performance of the models in predicting systemic events. The selection of the best model is done in the following way: once the probability of financial stress is estimated, we use the approach by Alessi and Detken (2011) to evaluate whether the policy maker can extract useful signals from it. Thus, we find the thresholds for the estimated probability that maximises the $U$ statistic for each model for the given preference parameter $\mu = 0.5$. The best model is the one that that achieves the highest usefulness $U$ score for the given preference parameter.

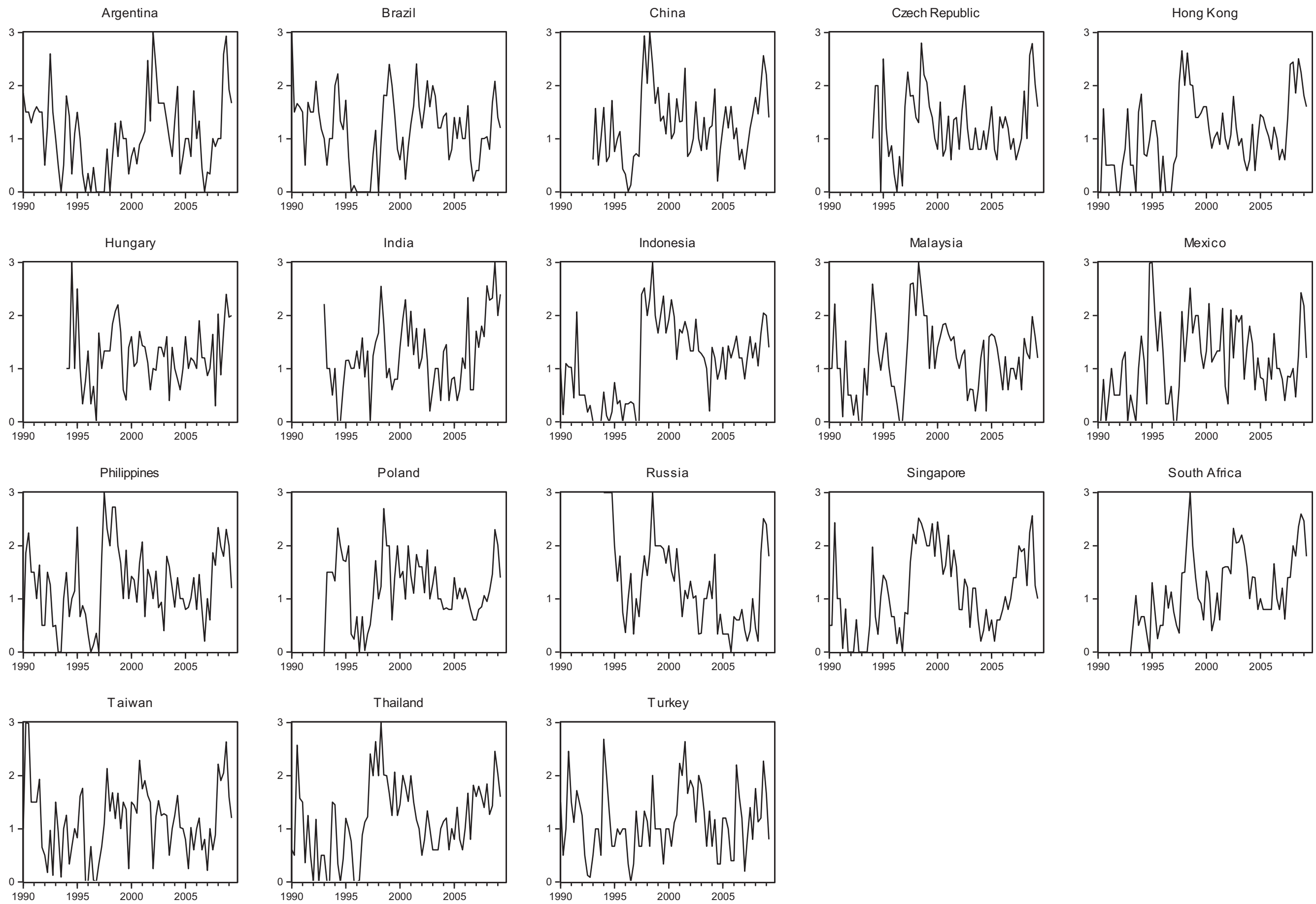


**Fig. A1.** Financial Stress Index for emerging economies in the sample.

Table 3 reports usefulness *U*, the noise-to-signal ratio (NtSr), the percentage of systemic events predicted by the indicator (%predicted), the probability of a systemic event conditional to a signal (Cond Prob) and the difference between the conditional and the unconditional probability of a systemic event (Prob Diff).[39] The main results are the following:

First, all models achieve a positive *U*, meaning that the models provide statistical gains for policy makers, who are equally concerned of issuing false alarms and missing systemic events. Second, all the models except the "Currency crisis" and the "Domestic (only EMEs)" models outperform the best stand alone indicator. The low performance of the "Domestic" model for emerging markets highlights the importance of including global factors in the model, especially for emerging economies. Third, the inclusion of global factors improves the performance of the models. Benchmark models with no interactions terms achieve usefulness gains, with the exception of the model for advanced economies. The latter performs as good as the "Domestic" model suggesting again that the role of global factors is more important for emerging economies. Finally, the inclusion of interaction terms further increases the performance of the models. As a matter of fact, Benchmark models (which include both global factors and interaction terms) are the best performing models. They achieve the highest *U*, successfully predict more than 80% of the systemic events, have the lowest noise to signal ratios and the highest conditional probability gains (i.e. difference between the unconditional probability and probability of a systemic event conditional to observing a signal from the model).

To sum up, our results highlight that analysing multiple signals from various sources of vulnerabilities in a multivariate framework, such as the discrete choice models, are more comprehensive tools than stand alone indicators to assist policy makers in evaluating systemic risks and predicting systemic events. Furthermore, it is important to take into consideration both domestic and international sources of vulnerabilities as well as their interactions.

We turn now to the evaluation of the out-of-sample performance of the logit models.

### 3.5. Out-of-sample performance of the models

We evaluate the out-of-sample performance of the logit models over the evaluation period 2005 Q2 to 2007 Q2 (8 quarters) in the following way[40]:

(1) We recursively estimate the model at each quarter *t* in the evaluation period using the information that would have been available in real time from the beginning of the sample (1990 Q1) to quarter *t*.

[39] See Kaminsky et al. (1998) for details on the calculations of these efficiency statistics.

[40] We choose 2005Q2 as starting point for our real time evaluation of the model in order to have long enough time series for all the countries for expressing regressors in country specific percentiles. We stop the evaluation period in 2007Q2 as in 2007Q3 the last financial crisis started in the US. Due to limitations in the number of observations we only estimate in real time the model for all countries without differentiating between EMEs and advanced economies.

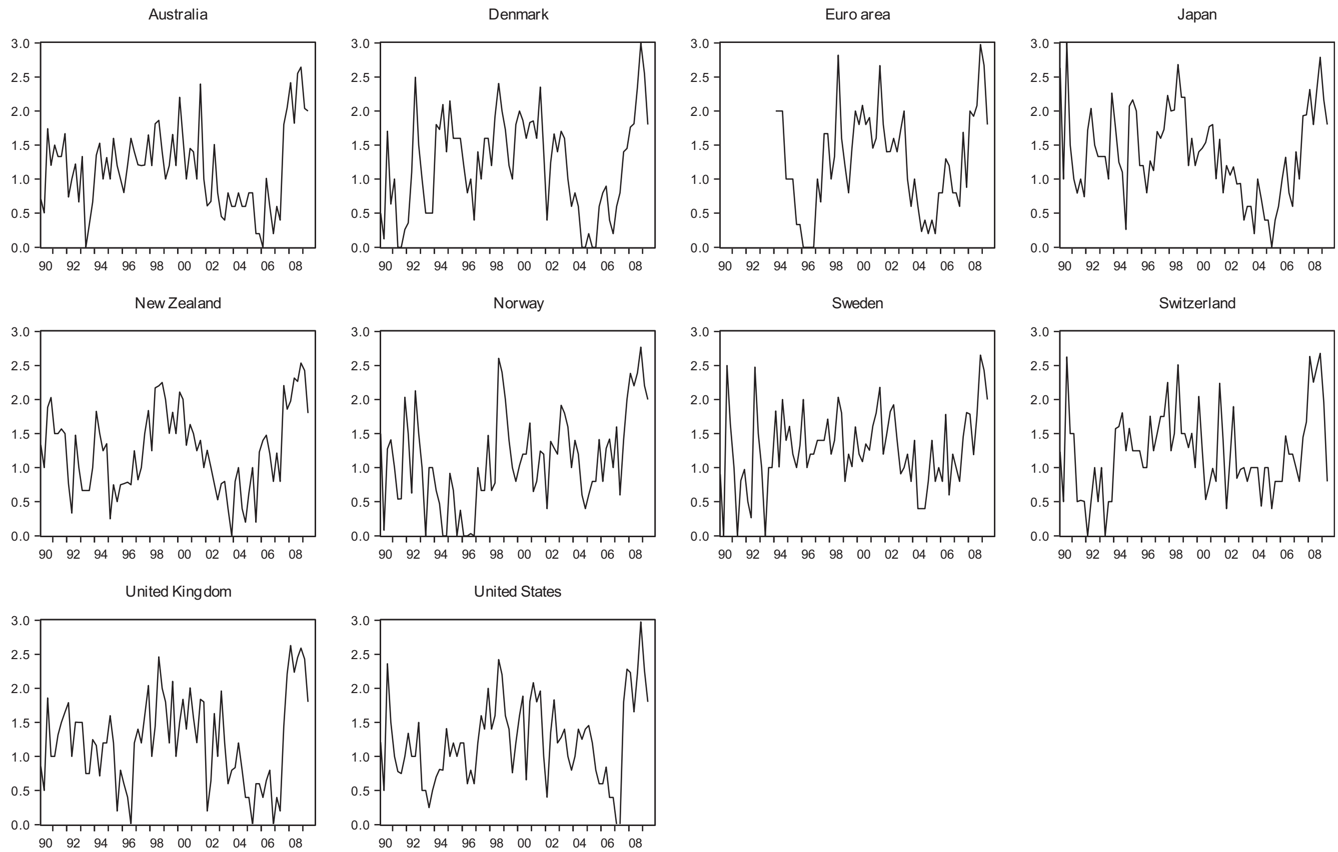


**Fig. A2.** Financial Stress Index for advanced economies in the sample.

**Table A1**
Performance of the benchmark logit model over different forecasting horizons with $\mu = 0.5$.

| Model | Forecasting horizon | Threshold (percentile) | $U$ | NtSr | % Predicted | Cond Prob (%) | Prob Diff (%) |
|---|---|---|---|---|---|---|---|
| Benchmark | 8 Quarters | 62 | 0.34 | 0.19 | 84.43 | 74.76 | 38.99 |
| Benchmark | 6 Quarters | 68 | 0.32 | 0.20 | 80.95 | 65.83 | 37.83 |
| Benchmark | 4 Quarters | 67 | 0.30 | 0.29 | 83.54 | 45.21 | 26.15 |
| Benchmark | 2 Quarters | 74 | 0.29 | 0.28 | 80.80 | 28.21 | 18.41 |

*Notes*: See notes to Table 3.

**Table A2**
Performance of the benchmark logit model using different values for the parameter $\mu$ with forecasting horizon 6 quarters.

| Model | $\mu$ | Threshold (percentile) | $U$ | NtSr | % Predicted | Cond Prob (%) | Prob Diff (%) |
|---|---|---|---|---|---|---|---|
| Benchmark | 0.5 | 68 | 0.32 | 0.20 | 80.95 | 65.83 | 37.83 |
| Benchmark | 0.4 | 69 | 0.23 | 0.19 | 79.83 | 67.06 | 39.06 |
| Benchmark | 0.6 | 65 | 0.22 | 0.23 | 83.47 | 62.74 | 34.74 |
| Benchmark | 0.7 | 53 | 0.14 | 0.37 | 91.88 | 51.57 | 23.57 |
| Benchmark | 0.3 | 72 | 0.13 | 0.17 | 74.51 | 69.27 | 41.27 |
| Benchmark | 0.8 | 53 | 0.07 | 0.37 | 91.88 | 51.57 | 23.57 |
| Benchmark | 0.2 | 81 | 0.06 | 0.11 | 57.14 | 77.27 | 49.27 |

*Notes*: See notes to Table 3.

(2) We collect the real time signals from the model over the evaluation period (assuming the benchmark scenario of a forecast horizon of 6 quarters and policy preference parameter of $\mu = 0.5$).

(3) We compute *ex post* the number of missed signals and false alarms issued by the model over the evaluation period and compute the measure of usefulness $U$ described in the previous chapter.

(4) We rank the models according to the usefulness parameter ($U$).

This approach provides a new, structured way to assess the out-of-sample performance of the models.

Table 4 summarises the results of the out-of-sample evaluation, which indicate that the Benchmark model and the three alternative models would have been useful tools for policy makers in predicting the ongoing Global Financial Crisis. As it was the case with in-sample predictions, the Benchmark model that incorporates both domestic and global variables as well as their interactions outperforms by far the other models.

Fig. 2 shows the out-of-sample performance of the Benchmark model for the United States for the ongoing Global Financial Crisis. It shows that the probability of a systemic event within 6 quarters in the United States was close to the early warning threshold already in 2006 Q1 and exceeded it in 2006 Q2. According to our Financial Stress Index, the switch from the tranquil period to the extreme financial stress period occurred in 2007 Q3, when the tensions in the money markets emerged and spread to other segments of the financial system. Thus, our Benchmark model is able to correctly anticipate the systemic event with a lead of 5 quarters. Furthermore, in the 5 quarters preceding the crisis, it keeps flagging that the systemic risks are elevated and a systemic event could be imminent.

### 3.6. Robustness analysis

In order to ensure the robustness of the results, we conducted the following robustness tests on discrete choice models:

#### 3.6.1. Definitions of vulnerabilities

We test various definitions and transformations of the vulnerability indicators (see Table 1). Overall, the results from the alternative models are qualitatively the same as in the Benchmark model. Moreover, they always have relatively high positive values of the usefulness parameter $U$, compared to the stand alone indicators of vulnerabilities. For instance, regarding asset valuations, we use also price/earnings ratios as it is common in the literature and obtain similar results to the Benchmark model. However, in this case, due to availability of data, our sample size is reduced and the analysis is not possible for all the countries in the sample.

#### 3.6.2. Contagion effects

To capture contagion effects we add the current average level of financial stress at the global level or, alternatively, in the region, to the set of explanatory variables. Our results indicate that adding contagion variables does not improve the performance of the Benchmark model in predicting systemic events.

#### 3.6.3. Role of capital inflows

Adding net capital inflows to the set of explanatory does not increase the performance of the model. The impact of capital inflows is indirectly captured by domestic asset price and credit dynamics.

#### 3.6.4. Forecasting horizon

We test the following forecast horizons for predicting systemic events: 2 quarters, 4 quarters, 6 quarters (benchmark) and 8 quarters. Overall, the performance of the model is relatively robust across forecasting horizons (see Table A1 in the Appendix A). The best performance is achieved, on average, over the 8 quarter period, followed by the 6 quarter period. Normally, over the 4 and 2 quarter periods, the model performance slightly decreases.[41]

[41] The performance also decreases for time horizons longer than 8 quarters (not reported in the table).

#### 3.6.5. Policy maker's preferences

In our benchmark analysis, we assume that the policy maker has the preferences of a neutral observer (she is equally concerned of Type I and Type II errors). By changing this assumption (see Table A2 in the Appendix A), we find that, overall, policy makers would benefit from the signals of the models as their usefulness score is positive. This is particularly the case, when the policy maker's preferences are close to the balanced preferences (i.e. either $\mu = 0.4$ or $\mu = 0.6$).

#### 3.6.6. Post crisis bias

We test whether our results are affected by a post crisis bias. Bussière and Fratzscher (2006) point out that including in the estimation of early warning models the economic recovery period after a crisis produces so called post crisis bias. In recovery periods, economic variables go through an adjustment process before reaching again the path they have during tranquil periods. The recovery period, therefore, should be excluded from the analysis as it is not informative of the path leading from the pre-crisis regime to the crisis. Bussière and Fratzscher address this issue by using a multinomial logit model with three regimes for the dependent variable (calm period, crisis and recovery). In this paper, as we exclude from the estimation sample the periods in which financial stress is high following the transition from tranquil regime to an extreme financial stress regime, we at least partially disregard some periods of economic recovery. However, we analyse the robustness of our results by excluding observations up to two quarters after the end of the stress periods to ensure that the post crisis bias is addressed. Only marginal gains in the performance of the model are obtained when dropping the additional two quarters from the sample.

## 4. Conclusions

This paper contributes to the financial crisis literature by developing a unified framework for assessing systemic risks, stemming from domestic and global macro-financial vulnerabilities, and for predicting (out-of-sample) systemic events i.e. periods of extreme financial instability with potential real costs.

We extend the existing literature on predicting financial crises in several ways. First, we identify past systemic events by using a composite index measuring the level of systemic tensions in the financial system of one country. Second, in predicting the identified systemic events, we evaluate the joint role of domestic and global vulnerabilities. In addition, we also analyse the role of the interactions between domestic factors and the interplay of global developments with the domestic conditions. Third, we evaluate both stand alone macro-prudential indicators of vulnerabilities and multivariate indicators estimated using discrete choice models. The evaluation of the indicators is done with a methodology that takes into account policy maker's preferences (Demirgüc-Kunt and Detragiache, 2000; Bussière and Fratzscher, 2008; Alessi and Detken, 2011).

The empirical analysis covers a set of 28 emerging market and advanced economies with quarterly data since 1990. Our results highlight the importance of considering jointly various indicators in a multivariate framework, as we find that discrete choice models outperform the stand alone vulnerability indicators in predicting systemic events. We find that combining indicators of domestic and global macro-financial vulnerabilities substantially improves the ability to forecast systemic events. In addition, considering interactions between domestic and global macro-financial vulnerabilities further improves the performance of the models.

Our framework displays a good out-of-sample performance in predicting the ongoing Global Financial Crisis. Our model would have issued an early warning signal for the United States in 2006

Q2, five quarters before the emergence of the tensions in money markets that started the crisis in August 2007.

The framework can also be used to identify potential vulnerabilities on the basis of a scenario analysis of the evolution of the domestic and global macro-financial environment.

## Acknowledgments

The authors want to thank for useful comments and discussions the editor, Ike Mathur, an anonymous referee, Lucia Alessi, Daniela Bragoli, Arjana Brezigar-Masten, Carmen Broto, Alexander Chudik, Mardi Dungey, Carsten Detken, Michael Fidora, Marcel Fratzscher, Philipp Hartmann, Jean Imbs, Manfred Kremer, Gilles Noblet, Peter Sarlin, Livio Stracca, Jouko Vilmunen and the participants to the following workshops and conferences: European Economic Association Annual Meeting 2010, Infinity 2010 conference, the 2010 Workshop of the Eurosystem and Latin American Central Banks, the 2010 ECB workshop ''A Global Dimension on Early Warning Models and Macro-prudential Analysis'', the 8th ESCB Workshop on Emerging Markets, the 2011 Bank of Korea and BIS conference on ''Macro-prudential Regulation and Policy'', the VI Annual Seminar on Banking, Financial Stability and Risk of the Central Bank of Brazil.

## Appendix A

See Figs. A1 and A2.
See Tables A1 and A2.

## References


Alessi, L., Detken, C., 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 (3), 520–533.

Balakrishnan, R., Danninger, S., Elekdag, S., Tytell, I., 2009. The Transmission of Financial Stress from Advanced to Emerging Economies. IMF Working Paper WP/09/133.

Berg, A., Pattillo, C. 2000. Predicting currency crises: the indicators approach and an alternative. Journal of International Money and Finance 18(4), 561–586.

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

Borio, C., Lowe, P., 2002. Asset Prices, Financial and Monetary Stability: Exploring the Nexus. BIS Working Papers, No. 114.

Borio, C., Lowe, P., 2004. Securing Sustainable Price Stability: Should Credit Come Back from the Wilderness? BIS Working Papers, No. 157.

Borio, C., Drehmann, M., 2009. Assessing the risk of banking crises – revisited. BIS Quarterly Review (March), 29–46.

Bussière, M., Fratzscher, M., 2006. Towards a new early warning system of financial crises. Journal of International Money and Finance 25 (6), 953–973.

Bussière, M., Fratzscher, M., 2008. Low probability, high impact: policy making and extreme events. Journal of Policy Modeling 30 (2008), 111–121.

Calvo, G., 1996. Capital flows and macroeconomic management: tequila lessons. International Journal of Finance and Economics 1 (3), 207–223.

Cardarelli, R., Elekdag, S., Lall, S., 2011. Financial stress and economic contractions. Journal of Financial Stability 7 (2), 78–97.

Davis, E.P., Karim, D., 2008. Comparing early warning systems for banking crisis. Journal of Financial Stability 4 (2), 89–120.

De Bandt, O., Hartmann, P., 2000. Systemic Risk: A Survey. ECB Working Paper No. 35.

Demirgüc-Kunt, A., Detragiache, E., 1998. The determinants of banking crises in developed and developing countries. IMF Staff Paper 45 (1), 81–109.

Demirgüc-Kunt, A., Detragiache, E., 2000. Monitoring banking sector fragility. A multivariate logit. World Bank Economic Review 14(2), 287–307.

Dungey, M., Fry, R., Martin, V., Tang, C., Gonzalez-Hermosillo, B., 2010. Are Financial Crises Alike? IMF Working Paper, WP/10/14.

ECB, 2009a. Global index for financial turbulence. Financial Stability Review (December), 21–23 (Box 1).

ECB, 2009b. The concept of systemic risk. Financial Stability Review (December), 134–142 (Special Feature B).

Edison, H.J., 2003. Do indicators of financial crises work? An evaluation of an early warning system. International Journal of Finance and Economics 8 (1), 11–53.

Eichengreen, B., Rose, A., Wyplosz, C., 1995. Exchange market mayhem: the antecedents and the aftermath of speculative attacks. Economic Policy 10 (21), 249–312.

Eichengreen, B., Rose, A., Wyplosz, C., 1996. Contagious currency crises: first tests. Scandinavian Journal of Economics 98 (4), 463–484.

Frankel, J., Rose, A., 1996. Currency crashes in emerging markets: an empirical treatment. Journal of International Economics 41(3–4), 351–366.

Fuertes, A.-M., Kalotychou, E., 2006. Early warning system for sovereign debt crisis: the role of heterogeneity. Computational Statistics and Data Analysis 5, 1420–1441.

Gavin, M., Hausmann, R., 1996. The roots of banking crises: the macroeconomic context. In: Hausmann, R., Rojas-Suarez, L. (Eds.), Volatile Capital Flows: Taming their Impact on Latin America. Inter-American Development Bank, Washington, DC.

Hakkio, C.S., Keeton, W.R., 2009. Financial stress: what is it, how can it be measured, and why does it matter? Federal Reserve Bank of Kansas City Economic Review, Second Quarter 2009, 5–50.

Hollo, D., Kremer, M., Lo Duca, M., 2012. CISS – A Composite Indicator of Systemic Stress in the Financial System. ECB Working Paper Series, No. 1426.

Honohan, P., 2000. Banking system failures in developing and transition countries: diagnosis and prediction. Economic Notes 29(1), 83–109.

Illing, M., Liu, Y., 2006. Measuring financial stress in a developed country: an application to Canada. Journal of Financial Stability 2 (3), 243–265.

IMF, 2008. World Economic Outlook, October 2008: Financial Stress and Economic Downturns. World Economic and Financial Surveys.

IMF, 2009. World Economic Outlook, April 2009: How Linkages Fuel the Fire: The Transmission of Financial Stress from Advanced to Emerging Economies. World Economic and Financial Surveys.

Jordá, O., Schularick, M., Taylor, A.M., 2011. Financial crises, credit booms, and external imbalances: 140 years of lessons. IMF Economic Review 59 (2), 340–378.

Kaminsky, G., 1998. Currency and Banking Crises: The Early Warnings of Distress. International Finance Discussion Paper 629. Board of Governors of the Federal Reserve System.

Kaminsky, G., Lizondo, S., Reinhart, C.M., 1998. Leading indicators of currency crises. IMF Staff Papers 45 (1), 1–48.

Kaminsky, G., Reinhart, C.M., 1999. The twin crises: the causes of banking and balance of payments problems. American Economic Review 89 (3), 473–500.

Laeven, L., Valencia, F., 2008. Systemic Banking Crises: A New Database. Working Paper WP/08/224, International Monetary Fund.

Misina, M., Tkacz, G., 2009. Credit, asset prices, and financial stress. International Journal of Central Banking 5 (4), 95–122.

Mishkin, F., 1996. Understanding Financial Crises: A Developing Country Perspective. NBER Working Paper No. 5600.

Reinhart, C.M., Rogoff, K.S., 2008. Is the 2007 US sub-prime financial crisis so different? An international historical comparison. American Economic Review 98 (2), 339–344.

Reinhart, C.M., Rogoff, K.S., 2009. The aftermath of financial crises. American Economic Review 99 (2), 466–472.

Sachs, J., Tornell, A., Velasco, A., 1996. Financial Crises in Emerging Markets: The Lessons from 1995. NBER Working Paper No. 5576.

Schularick, M., Taylor, A.M., 2011. Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870–2008. American Economic Review 102 (2), 1029–1061.