Comparing different early warning systems: Results from a horse race competition among members of the Macro-prudential Research Network

Alessi, Lucia and Antunes, Antonio and Babecky, Jan and Baltussen, Simon and Behn, Markus and Bonfim, Diana and Bush, Oliver and Detken, Carsten and Frost, Jon and Guimaraes, Rodrigo and Havranek, Tomas and Joy, Mark and Kauko, Karlo and Mateju, Jakub and Monteiro, Nuno and Neudorfer, Benjamin and Peltonen, Tuomas and Rodrigues, Paulo and Rusnak, Marek and Schudel, Willem and Sigmund, Michael and Stremmel, Hanno and Smidkova, Katerina and van Tilburg, Ruben and Vasicek, Borek and Zigraiova, Diana

16 February 2015

Comparing different early warning systems: Results from a horse race competition among members of the Macro-prudential Research Network

Lucia Alessi[1], António Antunes[2], Jan Babecký[3], Simon Baltussen[4], Markus Behn[5], Diana Bonfim[2], Oliver Bush[6], Carsten Detken[1], Jon Frost[4], Rodrigo Guimarães[6], Tomáš Havránek[3], Mark Joy[6], Karlo Kauko[7], Jakub Matějů[3], Nuno Monteiro[2], Benjamin Neudorfer[8], Tuomas Peltonen[1], Paulo M. M. Rodrigues[2], Marek Rusnák[3], Willem Schudel[1], Michael Sigmund[8], Hanno Stremmel[9], Kateřina Šmídková[3†], Ruben van Tilburg[4], Bořek Vašíček[3], Diana Žigraiová[3]

[1]European Central Bank, [2]Banco de Portugal, [3]Czech National Bank, [4]De Nederlandsche Bank, [5]Deutsche Bundesbank, [6]Bank of England, [7]Bank of Finland, [8]Oesterreichische Nationalbank, [9]WHU – Otto Beisheim School of Management

Preliminary: February 2015

Abstract

Over the recent decades researchers in academia and central banks have developed early warning systems (EWS) designed to warn policy makers of potential future economic and financial crises. These EWS are based on diverse approaches and empirical models. In this paper we compare the performance of nine distinct models for predicting banking crises resulting from the work of the Macroprudential Research Network (MaRs) initiated by the European System of Central Banks. In order to ensure comparability, all models use the same database of crises created by MaRs and comparable sets of potential early warning indicators. We evaluate the models' relative usefulness by comparing the ratios of false alarms and missed crises and discuss implications for pratical use and future research. We find that multivariate models, in their many appearances, have great potential added value over simple signalling models. One of the main policy recommendations coming from this exercise is that policy makers can benefit from taking a broad methodological approach when they develop models to set macro-prudential instruments.

I. Introduction

The global financial crisis has led researchers and policy makers around the world to put considerable effort into understanding and predicting systemic banking crises. In doing so, the empirical literature concerned with predicting banking crises has been focusing on developing early warning systems (EWS) which seek to predict future crises.

The aim of this paper is to elaborate on these advances in the field of EWS in a rather unique exercise. Specifically, we compare the performance of nine distinct models resulting from the work of the Macroprudential Research Network (MaRs).[1] All models use the same database of crises created by

[1] In the spring of 2010 the General Council of the European Central Bank (ECB) approved the establishment of the Macroprudential Research Network (MaRs) with the objective of developing core conceptual frameworks, models and tools that would provide research

MaRs and comparable sets of potential early warning indicators, so that we get a high degree of comparability. We evaluate the models' relative usefulness by comparing the ratios of false alarms and missed crises and discuss implications for pratical use and future research. Most models depart from traditional discrete choice (logit/probit) models, introducing dynamics and accounting for nonlinearities (binary regression trees) in these models as well as addressing uncertainty in the selection of individual early warning indicators (by means of Bayesian econometric techniques) Moreover, a potential of cross country heterogeneity is taken into accout (random coefficient models). Besides comparing the relative performance of these models, we discuss also relative advantages of different models in terms of economic intuition.

We find that despite the great diversity of approaches, one common result of these multivariate approaches is that they possess great potential to provide very useful early warning results, while offering considerable improvements over univariate signalling variables in terms of crisis prediction performance. Among the many results, we find evidence supporting the importance of accounting for private sector credit growth. The paper is organised as follows: Section II presents a concise overview of the related literature. Section III discusses the framework of the exercise and presents a set of common rules established ex-ante for all the approaches to follow to ensure meaningful comparability of model performance. Section IV elaborates on individual methodologies. Section V discusses the common left-hand side variable, while Section VI offers a structured comparison of the right-hand side variables, along with a summary of country coverage, time periods used for estimations, leads of early warning indicators (EWI) and model forecast horizons for each approach. Section VII compares the presented methodologies while it also highlights the specific features of each approach that contribute to their estimation success. Section VIII pinpoints in a concise manner the strengths and weaknesses of each approach while section IX lays ground for the future research to be yet done in the field.

II. Related literature

The early warning literature dates back to the late 1970s, when several currency crises generated interest in leading indicators (Bilson, 1979) and theoretical models (Krugman, 1979) explaining such crises. Nevertheless, only in the 1990s a wide-ranging methodological debate started, including studies on banking and balance-of-payments problems (Kaminsky and Reinhart, 1996) and currency crashes (Frankel and Rose, 1996).

Starting with the identification of single indicators, variable selection in most papers has usually followed the early work on signalling models from, inter alia, Kaminsky et al. (1998). The univariate signalling approach essentially maps the historical time series of a single indicator on past crises and extracts a threshold value above which crises are likely to happen. This univariate approach is transparent and straightforward to apply, which makes it attractive for policy makers. Yet, it contains a risk of underestimating the probability of a crisis if several other, potentially important, factors are close to (but below) their individual threshold values (Borio and Lowe 2002). More recent multi-variable early warning models have reduced this risk by estimating the probability of a future event

support in order to improve macro-prudential supervision in the European Union (EU). MaRs was set up for an initial period of two years, during which time it has followed three work streams. First, work stream 1 incorporates macro-financial models linking financial stability and the performance of the economy. Second, work stream 2 focuses on early warning systems and systemic risk indicators while work stream 3 assesses contagion risks. The review of early warning methods conducted in this paper falls into the second work stream.

(financial instability or a crisis) from a set of several potential early warning indicators (Frankel and Saravelos 2010 and Rose and Spiegel 2009). For both the univariate and the multivariate approaches it holds that in the case of a discrete left hand side variable, each early warning indicator or model is evaluated by minimising either the signal-to-noise ratio (Kaminsky 1999) or a policymaker's loss function (Demirgüc-Kunt and Detragiache 1999 and Bussière and Fratzscher 2006). In the case of the discrete choice approach (multinomial logit) and the continuous left-hand side variable approach, models have been estimated retaining significant indicators. Various papers also assessed the importance of banking sector characteristics for financial stability (Jeitschko and Jeung 2005, Cihak et al. 2009, Fahlenbrach et al. 2011).

More recent research has strived to improve early warning models by developing new techniques and employing more extensive datasets. Specifically, it has offered policy makers the explicit choice to pre-select their preferences regarding missed crises and false alarms and has subsequently evaluated indicators according to their usefulness given these preferences (Alessi and Detken 2011). In addition to traditional binary logit or probit models (Berg and Pattillo, 1998), the literature has been extended with multinomial models (Bussiere and Fratzscher, 2006), which generalise the discrete choice from two (yes/no) to more states, such as crisis, post-crisis, and tranquil periods. Departing from discrete choice models, continuous crisis indicators have been proposed (Rose and Spiegel, 2009, Frankel and Saravelos, 2010) that allow the EWM to explain the scale of real costs without the need to decide if it is sufficiently high to produce a 'yes' value. Markov switching models (Peria, 2002, Abiad, 2003) have also found their use in this area. Moreover, the literature improved the toolkits with methods such as non-parametric clustering methods like the binary recursive tree method (Barrel et al. 2009). A recent extensive literature review of both old and new publications is presented for example by Kauko (2014) and Mayes and Stremmel (2014).

Concerning computational methodology, Gaytán and Johnson (2002) distinguish qualitative approaches, signal extraction methods, dependent regression analysis, and other EWM approaches. As for qualitative models, they predict financial crises by exploring logical dependencies between risk factors and crises. In many cases, these links are simulated for different scenarios. For instance, the Bank of England uses risk transmission maps and feedback techniques to analyze financial crises (Bank of England, 2008; Aikman et al., 2009). Another such approach applies network theory to links in financial markets. For example, Elsinger, Lehar, and Summer (2006) model domino effects between large UK banks, and run simulations on the basis of a network of linkages between banks.

Moreover, statistical approaches are primarily data-focused and concentrate on regression models or on leading indicator and signal-extraction models. Kaminsky, Lizondo, and Reinhart (1998), and Kaminsky and Reinhart (1999) first apply this technique to financial crises by signalling a potential crisis when indicators exceed a previously defined threshold. Demirgüç-Kunt and Detragiache (2005) compare leading-indicator and regression models, concluding that the logit regression model is more suitable. However, Davis and Karim (2008) find that both models are useful, the signal model being better at predicting country-specific crises and the regression model more suitable for detecting global stress. Other methodological improvements offer policy makers an explicit choice to preselect their preferences regarding missed crises and false alarms (Alessi and Detken 2011).

As for less traditional approaches, tools based on artificial intelligence may provide valuable assistance in the complex and dynamic environment of today's financial systems. Lin et al. (2006) use a neuro-fuzzy approach to identify the drivers of currency crises and find that it improves the crisis prediction. A neural network is especially useful in detecting the main drivers of risk and their relative importance. On the other hand, input data must be chosen by the user, and the way risk patterns in the model are detected remains opaque. These drawbacks thus impede the use of neuro-fuzzy models for those wishing to get a more comprehensive picture of the crisis mechanism. Manasse and Roubini (2009) pioneered the use of classification and regression trees for predicting financial crises. Alessi and Detken (2014) apply the "Random Forest", a popular machine learning method based on decision trees, to the issue of identifying excessive credit and the associated build-up of systemic risk in the financial system. The "Random Forest" proves to be an extremely accurate predictor and a solid basis for the selection of the relevant indicators.

The latest works propose EWMs that consist of several components. First, they specify which costly events they intend to warn against. They explicitly distinguish (i) between diverse types of crises, (ii) between crisis onsets and occurrence, (iii) between crisis onset/current and crisis incidence in terms of real costs. Second, they select which countries should be incorporated into the EWMs and then identify which indicators could potentially provide a useful early warning about these costly events. By shifting the focus from emerging to developed countries, researchers contributed to working with arguably more homogeneous samples of countries. Third, they define time lags for these leading indicators in order to give policy makers some time to respond, allowing also for the fact that different indicators can provide useful information on different horizons. Fourth, they apply an empirical methodology that allows deciding which potential leading indicators exhibit sufficient predictive power and which shall be discarded.

III. Comparing EWS in a Horse race

Although the early warning literature has been subject to substantial methodological developments over the last decade, a robust comparison of different empirical frameworks has not been carried out so far. Indeed, different EW studies employ different frameworks and report results that are arguably conditioned on the selected framework. The purpose of this paper is to shed some light on relative merits of different empirical frameworks by means of a joint exercise (a so-called horse race). To make the performance of differing approaches to building early warning models (in our case for systemic banking crises) comparable, a set of ground rules needed to be established for the approaches to adhere to. This section lays out the specific conditions for all early warning models in this paper to follow.

First, the same dataset for all the authors was provided to draw on, even though everyone was free to rescale or generate different computations based on these data as long as these are reported. The constraint was put on trend and filter calculations as these were allowed only if they would exploit (quasi) real-time data (i.e. data that was available up to a particular moment in time) to ensure usability to policy makers. In line with this limitation, only one-sided (backward-looking) Hodrick-Prescott filters were allowed, or when a two-sided filter is used, recursive model forecasts should be employed.

Second, the left-hand-side variable, which defines banking crises and the related pre-banking crises periods, in all the EWMs was based on the banking crisis database that had been previously collected by the Czech National Bank and cross-checked by the heads of research of the ESCB within the MaRs network (Babecký et al. 2012) (see details in section IV).

Third, the contributors were asked to evaluate their models' performance for a homogeneous time window, specifically of 20 to 4 quarters before a banking crisis as well as for two sub-periods: the "early period" (20-12 quarters before the crisis), and the "late period" (12-4 quarters before the crisis). In other words, the models had to predict a crisis well in advance so as to ensure their robustness over time, which is key to make the models useful for policymakers who 1) require enough time to identify potential vulnerabilities which might need policy action and 2) need to give sufficient time to banks in case such actions require banks to build up capital (which applies to, for example, the counter-cyclical capital buffer).

Fourth, the right-hand-side variables used in the models needed to be either extracted from the provided dataset or from other publicly available sources. To increase viability of the forecast exercise, the variables selected into models needed to account for the data availability, such as publication lags for national account data. Financial market prices could be used contemporaneously.

Fifth, given the EU-wide focus of the exercise, the authors were asked to incorporate as many EU countries as possible and to provide out-of-sample predictions to accompany the reporting of in-sample model performance. In particular, models should demonstrate the quality of their out-of-sample performance for crisis episodes that precede the current global financial crisis (2007-), e.g. the Danish/Finnish/Swedish crises in the late 1980s/early 1990s. The respective crisis episode used for this purpose should thus have been excluded from the estimation.

Finally, all participants were asked to evaluate the in-sample fit of each horizon applying the so-called 'signalling approach' which has originally been developed by Kaminsky et al. (1998) and extended by Demirgüç-Kunt and Detragiache (2000), Alessi and Detken (2011) and Lo Duca and Peltonen (2013). In this framework, an indicator issues a warning signal whenever its value in a particular period (in this case one of the three time periods discussed above) exceeds a threshold which is defined as a percentile of the country-specific distribution of predicted probabilities. For example, a multivariate probability model (such as the multivariate logit) issues a warning whenever the predicted probability from this model exceeds a threshold. The predictive abilities of the different models can then be evaluated by comparing the signals issued by them in relation to the actual crisis observations. Each observation can be allocated to one of the quadrants in the contingency matrix depicted below in Figure 1: A signal by a specific indicator can either be followed by a banking crisis in a given horizon or not, *i.e.*, it can be a true positive (TP) or false positive (FP). Similarly, a period where no signal was issued can either be followed by a banking crisis or not, *i.e.*, a false negative (FN) or a true negative (TN), respectively.

Figure 1: Contingency matrix

		Actual class *Cj*	
		1	0
Predicted class *Pj*	1	*True positive (TP)*	*False positive (FP)*
	0	*False negative (FN)*	*True negative (TN)*

Note: This figure shows the relationship between model prediction and actual outcomes. Observations are classified into those where the indicator issues a warning that is indeed followed by a banking crisis (TP), those where the indicator issues a warning that is not followed by a crisis (FP), those where the indicator issues no warning and there is no crises (TN), and those where the indicator issues no warning although there is a crisis coming (FN).

In order to obtain the optimal threshold, it is key that the policy maker's preference vis-à-vis so-called type-I (T1) errors (missing a crisis by not issuing a warning although a crisis is approaching within the pre-set time horizon) and type-II (T2) errors (i.e. issuing a false alarm) are taken into account. This can be done by using a loss function which is minimised depending on the frequency of the two error types and on the policy maker's preference for either type. The loss of a policy maker consists of T1 and T2, weighted according to her relative preferences (P) between missing crises (μ) and giving false alarms ($1 - \mu$), which leads to the following loss function: $L(\mu) = \mu T1P1 + (1 - \mu)T2P2$. Using the loss function $L(\mu)$, the absolute usefulness (U) of a model can be defined in the following way: $U = \min(\mu P1, (1 - \mu)P2) - L(\mu)$, which computes the extent to which a model performs better than a coin toss (having no model). For the horse race, participants were asked to compute both Type I and Type II errors of their models.

In addition to assessing the absolute usefulness of a model, participants were also asked to report the Area under the Receiver Operating Characteristics curve (AUROC) as this is also a viable measure for comparing early warning model performance. The AUROC measures the probability that a randomly chosen distress event is ranked higher than a tranquil period. A perfect ranking has an AUROC equal to 1, whereas a coin toss has an expected AUROC of 0.5.

IV. Methodologies

This section presents the individual approaches to early warning model construction and highlights their unique features. The contributions are listed by their authors in order of their increasing dissimilarity with traditional approaches.

Baltussen et al. build a discrete probability (probit) model for crisis prediction. The unique feature of their approach is the inclusion of an interdependency index. Whereas other approaches proxy country interdependency only indirectly via international variables, this exogenous spatial lag captures country interdependency explicitly by simulating financial conditions in partner countries. Accounting for this is roughly analogous to the idea behind international application of the counter-cyclical capital buffer (CCB) (BCBS, 2010). Cross-border financial exposures are approximated by the bilateral weights in the BIS NEER index. The model draws inspiration from approaches used to research the spatial dependency of business cycles (Kakamu and Wago, 2010) and contagion of currency crises (Novo, 2003; Eichengreen, Rose and Wyplosz, 1996). Although probit and logit models lead to fairly similar results, a probit is chosen to coincide with more sophisticated spatial econometric research. In the latter the logistic distribution results in intractable multivariate specifications (Anselin, 2002). The two-step approach can be formalized as:

$$P(y_{1,i,t}) = \Phi(a + \beta_1' x_{1.i,t} + u_{1,i} + \varepsilon_{i,t}) \tag{1}$$

$$P(y_{2,i,t}) = \Phi(a + \gamma' W_{i,j}' {}_i \bar{y}_{1,j,t} + \beta_2' x_{2.i,t} + u_{2,i} + \varepsilon_{2,i,t}) \tag{2}$$

where $x_{i,t}$ is a vector of the exogenous variables, β is a vector of the coefficients and u_i is the country-specific random element. The reasoning behind the random effects model is that the (endogenous) spatial lags would amplify selection bias of a fixed effects model. Equation (1) includes important non-EU partners used to construct the exogenous interdepency index. The fitted values are used to construct the interdependency index, including a mechanical predictor (which takes on a value of 1 a fixed number of quarters after a crisis has materialized) to adjust for the post-crisis bias. These variables are used in equation (2), where $\bar{y}_{1,i,t}$ is a vector of the resulting (adjusted) fitted values, W_i the vector of financial linkage weights and γ' the vector of coefficients.

Bush et al. present another discrete choice model, a logit regression, to explain crisis periods. In comparison to the other approaches, their variable choice is rather data driven. They employ a principal components analysis (PCA) in a data rich environment with international data availability to guide their choice of right-hand side variables. The outcome of the PCA analysis suggests the usage of the following variables: credit-to-GDP gap, accounting leverage ratio, liquid asset ratio and a proxy for the price of risk (VIX). The simple logit panel regression model can be presented as follows:

$$Prob(C_{it} = 1) == \frac{1}{1+e^{-(\alpha_i + \mathrm{X}_{it}\beta'_{it})}} \tag{3}$$

where $Prob(C_{it} = 1)$ is the probability that country i at time t is in a crisis state. As independent variables, the vector X_{it} contains the four different explanatory variables. In addition, the parameter α_i represents country fixed effects. The approach also employs robust standard errors. The data frequency is different from the other approaches due to data availability issues. Since the calculation of accounting leverage relies on bank balance sheet data, this model has to employ annual data. The underlying aim of their specification is to find a combination of indicators, which anticipates crises with sufficiently long lead for policymakers to act.

Antunes et al. estimate dynamic probit models, comparing them with simple multivariate probits. The latter models make the restrictive assumption that all observations are independent, hence no dynamics is considered. When observations over time (time series) are available, time dependence may exist and it may be of interest to account for these dynamics. One possibility is to add lags of the dependent variable to the models, thereby estimating dynamic probits. Antunes et al. consider variants of the general dynamic probit model representation

$$y_{it}^{*} = \alpha + \sum_{k=1}^{p}\sum_{j=1}^{d}\beta_{kj}x_{ij,t-k} + \sum_{k=1}^{p}\gamma_k y_{i,t-k} + \sum_{k=1}^{p}\sum_{j=1}^{d}\delta_{kj}D_{ij,t-k}^{0.75} + u_{it} \tag{4}$$

where y_{it} is a binary crisis indicator, y_{it}^{*} is a latent variable such that $y_{it} = 1$ if $y_{it}^{*} > 0$ and 0 otherwise; $x_{ij,t}, j = 1,..., p$ corresponds to a set of p exogenous covariates, $y_{i,t-k}$, $k=1,..,p$ corresponds to the k^{th} lag of the crisis indicator and $u_{i,t}$ denotes a normally distributed random error term.
Hence, based on (1), for empirical purposes two distinct models are considered: i) a marginal model which results from setting $\gamma_1 = ... = \gamma_p = 0$, *i.e.*, only considers the effects of covariates on the probability outcomes and treats serial dependence as a nuisance which is captured through association parameters; ii) a transitional model which explicitly incorporates the history of the response in the regression for y_{it}^{*} (complete model (1)). Hence, in this way, each unit specific history can be used to generate forecasts for that unit, as opposed to the marginal model which makes forecasts solely on the basis of the values of the exogenous variables. In addition, Antunes et al. also explore dynamic probit models with extreme behaviour indicators $D_{ij,t}^{0.75}$, $j = 1,..., p$. The objective is to better capture periods in which some explanatory variables reach extreme values. In practice, when the 75th percentile of a variable is exceeded, a dummy variable is included.

Behn et al. investigate how early warning performance can be improved in a multivariate (logit) framework. Starting from the notion that the domestic credit-to-GDP gap (as used in BCBS 2010) performs well in univariate signalling models predicting financial crises, the authors build a multivariate model which includes credit variables, other macro-financial variables and banking sector variables. Moreover, they attempt to capture some of the potential spillover effects by incorporating global[2] variables into their models. Formally, the logistic regression model can be presented as follows:

$$Prob(y_{it} = 1) = \left[\frac{e^{\alpha_i + X'_{it}\beta}}{1+e^{\alpha_i + X'_{it}\beta}}\right] \qquad (5)$$

where $Prob(y_{it} = 1)$ depicts whether a country is in a pre-crisis vulnerable state (i.e. the common left hand side variable in the exercise. As independent variables, the vector X_{it} includes credit and macro-financial variables on the domestic and on the global level as well as domestic banking sector variables. The estimations also include a set of country dummy variables α_i in order to account for unobserved heterogeneity at the country level (country fixed effects).[3] Finally, the approach uses robust standard errors clustered at the quarterly level in order to account for potential correlation in the error terms that might arise from the fact that global variables are identical across a given quarter. The analysis is conducted as much as possible in a real-time fashion, meaning that only information that is available at a particular point in time is used. As such, all de-trended variables have been calculated using backward trends, thereby only using information that was available up to that point. Furthermore, the explanatory variables have been lagged by one quarter, also to account for endogeneity bias through simultaneity.

Neudorfer and Sigmund employ a Bayesian Random Coefficient Logit Panel model. This model is a special case of a generalized linear mixture model that contains fixed and random effects. In a way the term "fixed" and "random" are a bit misleading: arguably it is more appropriate to call them "general" and "country-specific" effects. If these effects are translated into an estimation framework they can be interpreted as follows: for each explanatory variable two coefficients are estimated: a "general" coefficient which is common for all countries and a "country-specific" coefficient which is different for each country. As a consequence, a random coefficient model covers the middle ground between the implausible assumption of homogeneity across countries (as assumed in a pooled model) and the infeasibility of treating all countries differently in the sense of being estimated separately.
To account for the dichotomous property of the pre-crisis variable, we apply a logistic transformation of the pre-crisis variable which does not change the ideas behind the generalized linear mixture models. Formally, the model can be presented as follows:

$$y_{i,t} = \beta'_i x_{i,t} + \varepsilon_{i,t}\beta_i = \beta + \gamma_i \qquad (6)$$

where $u_{i,t}$ denotes the random error term, $x_{i,t}$ is a K × 1 vector of exogenous variables and $\beta_{i,t}$ is the K × 1 vector of coefficients. Further β denotes the fixed effects that are the same for all individuals i. $\gamma_{i,t}$ is a Kx1 vector of a stationary random variable with a zero mean and a constant variance. The vector represents the country specific slope heterogeneities that are thus called random effects. Neudorfer and Sigmund then model response probabilities using the logit function. The aim of the exercise is to maximise the contingency matrix's diagonal, i.e. to maximise true positives and true negatives.

[2] Global variables are depicted by the (GDP-weighted) averages of the domestic variables across all EU countries plus Canada, Japan and the United States.
[3] There is an argument for omitting these dummies from the estimations as they automatically exclude all countries without a crisis from the estimation, hence introducing selection bias. However, not including them also induces bias, namely omitted variable bias caused by unit effects. As it is unlikely that financial crises are homogeneously caused by identical factors and as a Hausman test indicates unit heterogeneity, we have decided to include unit dummy variables in our estimations. See also Behn et al. (2013).

Kauko applies a univariate signalling approach which is based on the assumption that long lasting excessive credit growth gradually becomes more problematic. In previous research, it has been commonplace to measure the growth of the loan stock in percentages, i.e. by dividing the growth of the loan stock by the size of the loan stock itself, consistently with the odd assumption that a credit expansion of given size is more dangerous if the existing stock of loans is small. However, it may be more realistic to assume that long lasting excessive credit growth gradually becomes more and more problematic. Hence, the 12-month difference in the loan stock is scaled by dividing it by GDP, not by the amount of debt. More specifically, the denominator is the five-year moving average of the nominal GDP. A serious depression is no sign of excessive credit growth, and the moving average of the GDP is largely immune to the short-term cyclical variation of output.
The raw indicator is complemented by assuming that additional risk factors are needed to trigger a crisis. A crisis is almost impossible unless there is either contagion from abroad or a current account deficit. Previous research has demonstrated that mere credit growth is in most cases harmless unless either of these additional risk factors is present (Jordà et al 2011). Therefore, the difference indicator is divided by 10 unless either of these additional risk factors prevails. Formally, the measure is defined as follows:

$$K_1 = 500\,000 * \left[\frac{0.3333 \sum_{t=-3}^{-1} L_t}{\left(\frac{1}{19}\right) \sum_{t=-1}^{-19} Y_t} - \frac{0.25 \sum_{t=-4}^{-7} L_t}{0.05 \sum_{t=-4}^{-23} Y_t}\right] \qquad (7)$$

where L equals loan stock and Y equals GDP. The K_1 indicator is computed as the difference of credit stock divided by a five-year moving average of nominal GDP and is not very sensitive to business cycles. Moreover, its value is dependent on additional conditions. The indicator takes value K_1 from the formula if either the current account of the previous year is smaller than zero or the contagion risk is detected or both. In contrast, it takes the value of $\frac{K_1}{10}$ if there is neither a current account deficit nor contagion risk. The underlying idea is derived from the findings of Jordà et al.(2011) who based on a handful of advanced countries over 140 years observed that a banking crisis in the making could be identified based on credit growth and some additional factor, which could either be current account deficit or contagion from abroad. Moreover, the idea that a current account surplus is an almost perfect deterrent of banking crises is corroborated by e.g. the tables presented by Laeven and Valencia (2008) and various publications reviewed by Kauko (2014).

Babecký et al. aim to deal with model uncertainty inherent to EWS by means of Bayesian model averaging (Madigan and Raftery, 1994; Raftery, 1995, 1996). Bayesian model averaging (BMA) considers model combinations based on weights determined according to the models' fit. The following linear regression model is considered:

$$y = \alpha_\gamma + X_\gamma \beta_\gamma + \varepsilon \quad \varepsilon \sim (0, \sigma^2 I) \qquad (8)$$

where y is the dummy variable for crisis onset, α_γ is a constant, β_γ is a vector of coefficients, and ε is a white noise error term. X_γ denotes some subset of all available relevant explanatory variables, i.e., potential early warning indicators X. The number K of potential explanatory variables yields 2^K potential models. Subscript γ is used to refer to one specific model out of these 2^K models. The information from the models is then averaged using the posterior model probabilities that are implied by Bayes' theorem:

$$p(M_\gamma \mid y, X) \propto p(y \mid M_\gamma, X) p(M_\gamma) \qquad (9)$$

where $p(M_\gamma|y,X)$ is the posterior model probability, which is proportional to the marginal likelihood of the model $p(y|M_\gamma,X)$ times the prior probability of the model $p(M_\gamma)$.

The robustness of a variable in explaining the dependent variable can be expressed by the probability that a given variable is included in the regression. It is referred to as the posterior inclusion probability (PIP) and is computed as follows:

$$PIP = p(\beta_\gamma \neq 0 \mid y) = \sum_{\beta_\gamma \neq 0} p(M_\gamma \mid y) \qquad (10)$$

The PIP captures the extent to which we can assess how robustly a potential explanatory variable is associated with the dependent variable. Variables with a high PIP can be considered robust determinants of the dependent variable, while variables with a low PIP are deemed not robustly related to the dependent variable.

Joy et al.[4] aim to identify a set of economic "rules of thumb" that characterise economic conditions preceding the onset of banking crises. They employ a novel nonparametric approach - Classification and Regression Tree methodology (CART), specifically a Binary Classification Tree (BCT). This methodology permits the detection of key variables driving binary crisis outcomes, allows for interactions among key variables and determines critical tipping points. The multivariante CART tries to provide a more organized selection of the crises triggers from a relatively rich set of variables. The data are partitioned recursively and within each partition a simple prediction model is fitted. As a result, the partitioning can be represented graphically as a decision tree. CART thus searches through different possible splits for all explanatory variables and selects those splits that best separate crisis episodes from no-crisis episodes. The splitting criterion is the minimization of a loss function based on a cost that rises when the actual split deviates from the perfect split, i.e. where the perfect split partitions all crisis episodes into one node and all no-crisis episodes into another.

Let $p(i|t)$ be the fraction of occurrences belonging to class i at node t. In a two-class problem, such as here, and omitting the reference to node t, the class distribution at any node can be written as (p_0, p_1), where p_0 is the posterior probability of a *no-crisis* observation falling into node t, and p_1 is the posterior probability of a *crisis* observation falling into node t. Measures for selecting the best split are based on the degree of *impurity* in the child nodes. The more skewed the distribution, the smaller the degree of impurity. A node with class distribution (0, 1), for instance, has zero impurity, while a node with class distribution (0.5, 0.5) has maximum impurity.

The *Gini* criterion is employed as a primary splitting rule, which corresponds to the following impurity (or loss) function $i(t)$, which is minimised:

$$i_{gini}(t) = \sum p_0(t) p_1(t) \qquad (11)$$

Alessi and Detken aspire to provide policy makers with a set of early warning indicators helpful in guiding decisions on when to activate counter-cyclical capital buffers (CCB). Similarly to Joy et al., they use decision tree learning to build a predictive model. However, they apply a more advanced learning method than the Binary Classification Tree, to overcome its lack of robustness. Indeed, there could be multiple similarly important splitting variables at the same node, but only one would be shown in the tree while the information content of the others would remain hidden. Moreover, an

[4] Joy et al. (2014) aim both at banking and currency crises using different prediction horizons that defined above.

indicator with a generally poor predictive power could happen to feature in the tree due to the specific sample selection, but it would not survive a robustness check. For this reason, Alessi and Detken propose the application of the "Random Forest" technique (see Breiman 2001), which improves the stability and accuracy of the predictions via bootstrapping and aggregating a multitude of trees. Based on the results of the "Random Forest", they then select the most relevant early warning indicators and construct a benchmark classification tree on them.

The multivariate methodology proposed by Alessi and Detken to adopt decisions on the macroprudential instruments described above is decision tree learning, a greatly underutilized technology in economics. Indeed, while Classification and Regression Trees (CARTs) are extensively used in other disciplines, their economic applications are rare. Classification trees are a transparent tool which would also enable the public at large to understand and possibly anticipate macroprudential decisions. However, the main drawback of the tree technology is that, while it can be very good in-sample, it is known for being not particularly robust to the inclusion of additional predictors or observations.

Alessi and Detken overcome this problem by using the Random Forest method proposed by Breiman (2001). This framework is a state-of-the-art machine learning technique which consists in bagging, i.e. bootstrapping and aggregating, a multitude of trees. Each of the trees in the forest is grown on a randomly selected set of indicators and country-quarters. Analogously to the tree, the forest allows for interaction across the various indicators, is able to handle large datasets, is not influenced by outliers and does not require distributional or parametric assumptions. Together with being an extremely powerful predictor, the Random Forest allows to measure the importance of each of the input variables by evaluating the extent to which it contributes to improve the prediction. Notwithstanding the remarkably good performance of the Random Forest, we acknowledge that this is a black-box model and its predictions would be hard to defend, in particular if they would support the activation of a macroprudential instrument. Therefore, they rely on the Random Forest in order to identify the key indicators, on which we construct our benchmark early warning tree. By doing so, Alessi and Detken ensure that the variables selected to grow the tree are truly the most important ones in the pool and we rule out the possibility that the tree selects a relatively weak indicator which just happens to seem useful but would not survive a robustness check.

V. Definition of Crises

The left hand side variable is based on the so-called ESCB Heads of Research Database which has been collected by a team at the Czech National Bank in collaboration with the ESCB. This MaRs project (Babecký et al. 2012) has constructed discrete indices of the occurrence of banking, debt, and currency crises for EU-27 (and other OECD countries) by aggregating the available data sources, which, besides academic studies, included a survey of the ESCB Heads of Research who were provided with information about crisis occurrence in their respective countries alongside the corresponding crisis definitions. The influential papers that have been included are (in alphabetical order): Caprio and Klingebiel (2003); Detragiache and Spilimbergo (2001); Kaminsky (2006); Kaminsky and Reinhart (1999); Laeven and Valencia (2008, 2010, 2012); Levy-Yeyati and Panizza (2011); and Reinhart and Rogoff (2008, 2011).

The database covers financial crises in EU and OECD countries over a period from 1970:Q1 to 2010:Q4 and has been constructed as follows. For each country, three binary variables capture the timing of banking, debt, and currency crises. The corresponding crisis occurrence index takes a value of 1 when a crisis occurred (and a value of 0 when no crisis occurred).

With regard to the identification of banking crises used in this exercise, the experts were asked to use the following definition: a banking crisis is defined by significant signs of financial distress in the banking system as evidenced by (i) bank run(s) in relevant institutions or losses in the banking system (nonperforming loans above 20% or bank closures of at least 20% of banking system assets) or (ii) significant public intervention in response to or to avoid the realisation of losses in the banking system. In terms of the latter, intervention is considered significant if at least one of the following applies: (a) extensive liquidity support (ratio of central bank claims on the financial sector to deposits and foreign liabilities exceeds 5% and more than doubles relative to its pre-crisis level), (b) bank restructuring costs (the component of gross fiscal outlays directed to restructuring of the financial sector, excluding asset purchases and direct liquidity assistance from the treasury, is at least 3% of GDP in at least one fiscal year), (c) significant bank nationalisations (takeovers by the government of systemically important financial institutions, including cases where the government takes a majority stake in the capital of such financial institutions), (d) significant guarantees (either a full protection of liabilities or guarantees extended to non-deposit liabilities of banks; actions that only raise the level of deposit insurance coverage are not deemed significant), (e) significant asset purchases (including those implemented through the treasury or the central bank of at least 5% of GDP, cumulated) and (f) deposit freezes and bank holidays.

For the exercise, the dependent variable is set to 1 between (and including) 20 to 4 quarters prior to a banking crisis as identified by the ESCB HoR database and to 0 for all other quarters in the data. In order to overcome crisis bias, we omit all country quarters which either fall within the period from three quarters before the onset of a banking crisis up until the end of a banking crisis. Moreover, we also exclude all quarters as of 2006q1 (i.e. 20 quarters before the end of the data series). Indeed, the dataset ends in 2010q4 and strictly speaking ignores a crisis happened in any country as of 2011. Therefore, the period immediately preceding the end of the sample cannot be classified as pre-crisis or non pre-crisis. Moreover, some exeptions are detailed below.

Baltussen et al. use the ESCBs' Heads of Research (HoR) definition and database for identifying the pre-crisis period for all EU banking crises. Since estimating the probability of crises in major trading partners is necessary to construct the interdependency index, Baltussen et al. also use the Laeven and Valencia (2013) database to construct the same LHS-variable of pre-crisis periods for non-EU countries (only) in equation (1).

Bush et al. use two different crisis indicators as left-hand side variables to verify their results. They employ the ESCBs' Heads of Research (HoR) banking crises database. Further, they use the Laeven and Valencia (2012) database. Although the two databases use deviating methodology to collect and identify banking crisis event, they do share considerable overlaps. Both LHS variables are adjusted for post crisis bias identified by Bussière and Fratzscher (2006).

Antunes et al. use the the ESCBs' Heads of Research (HoR) banking crises database as a basis for their dependent variable, as described in Antunes et al. (2014). To ensure the comparability of the results in this exercise, they transform this variable into a vulnerability indicator, following the approach of Behn et al. This variable takes the value 1 in the (4 to 20) quarters before the emergence of a banking crisis.

In the same spirit, Sigmund and Neudorfer make use of a crises dummy dependent extracted from HoR database, where they assign 1 to crisis periods and 0 to others. Moreover, they add 0 to each quarter from 2011Q1 to 2012Q4 for all countries in their sample.

In contrast, Kauko does not require in his approach a crisis dependent as he focuses on building an indicator of crisis incidence that signals crisis periods based on meeting complementary conditions. No econometric model is presented in this approach; instead, the idea is to base the indicator on findings presented in previous literature.

VI. Potential Early Warning Indicators

VI.1. Overview and Sources

Subsection VI.1. offers an overview of the right-hand side variables used in each early warning model along with their sources. The employed variables are also presented for each method in table A in the appendix. Unlike in case of the left-hand side variables certain discretion was allowed to construct the right hand side variables.

Baltussen et al. select as leading indicators domestic credit and debt service ratio (BIS), GDP (IMF IFS, OECD), gross capital flows (IMF IFS) and REER appreciation (BIS). The raw variables are transformed in the following way. Credit gap is the deviation of private credit to GDP from a one-sided HP trend (λ =400000) where the first 5 years are excluded because of the shortage of lags. REER growth, GDP growth and the ratio of credit to GDP growth are defined as the change from 4 quarters earlier, and capital flows as a 4-quarter rolling sum of all gross inflows to GDP. The rationale for this set of variables is to attempt to capture both domestic imbalances and global risk and contagion through financial flows. Furthermore, an asymmetric definition of both capital flows and credit growth is used to allow for a dispersion of the financial cycle as a result of excessive credit versus credit crunches and different types of extreme capital flow episodes (see Forbes and Warnock, 2012). Similar methodologies are used in the research on the effects of shocks on real economy cycles (Jiménez-Rodríguez and Sánchez, 2004, Lilien, 1982).

Bush et al. use the BIS long series on total credit to the private non-financial sector as their credit data source and obtain the GDP data from IMF/OECD. They construct the credit-to-GDP gap series using the standard methodology by employing a one-sided recursive HP filter with a lambda of 400.000. In addition, they incorporate banking sector variables. The accounting leverage ratio (Common Equity/Total Assets) is obtained by aggregating the individual banking institutions' leverage ratios on the country level for each year. The underlying data source of the accounting leverage data is Worldscope/Datastream.[5] The liquid assets of the banking sector are obtained from the IMF IFS dataset. Finally, the price of risk measures (VIX proxy) is constructed by calculating the deviations of the general country stock market index from Datastream. The rationale behind their methodology is two-fold. Bush et al.'s purpose is to find a robust combination of indicators that is able to anticipate crises with sufficiently long lead for policymakers to act. In addition, they also aim to demonstrate the theoretical relevance of bank balance sheets and risk pricing in an empirical model.

Antunes et al. used the dataset provided by the ECB for the horse race exercise. In order to maximize the information set available, in some cases the authors combined the series from a given source with data from other sources available in the dataset. The authors implemented a few transformations of the variables provided. First, they computed several ratios, such as credit-to-GDP and total assets of the banking system as a percentage of GDP. Second, they computed year-on-year growth rates for most of the variables. Finally, they estimated deviations from long-term trends, using one-sided Hodrick-Prescott filters with different smoothing parameters. Ultimately they select the following variables

[5] Bush et al. employ certain coverage thresholds on the country level to ensure a solid representation of the banking sector in each country.

with better performance based on a univariate analysis: equity prices (index), debt service ratio (year-on-year), credit-to-GDP gap (one-sided HP filter with lambda = 400 000) and house prices (year-on-year) for their analysis.

Neudorfer and Sigmund choose credit-to-nominal-GDP gap computed as one sided HP filter (lambda of 400 000, BIS data), real GDP growth (EU data), house price growth (for each country the longest available time series), total asset growth, equity growth (longest available time series), capital and reserve growth and the debt service ratio.

Kauko selects for his crisis incidence a measure of contagion data (BIS), which is used as a binary variable, that is equals 1 if the number of countries in crisis state is larger than it was 4 quarters before. Moreover, the, current account (IMF) is also treated as binary (surplus, deficit). These two variables are combined; it is assumed that crises are very unlikely unless there is either a current account deficit, contagion from abroad, or both.

Behn et al. select as their credit variables the BIS long series on total credit to the private non-financial sector, and the credit-to-GDP gap as a deviation of the credit-to-GDP ratio from its long-term trend (one-sided (recursive) HP filter, λ=400 000). As for macro-financial indicators GDP growth, inflation, stock price growth and house price growth (from various sources, sourced through Haver Analytics) are chosen. Moreover, they incorporate banking sector variables such as aggregate capitalisation and profitability (OECD). The latter variables are interesting especially in the context of the counter-cyclical capital buffer, as it enables the authors to assess the effect of higher banking sector capitalisation on the probability of future financial crises. The variables are lagged by one quarter.

Babecký et al. use around 30 variable described in Babecký et al. (2012) and add for the purpose of this early warning exercise some additional variables. All HP filters they use are one sided, i.e. real time. They employ country fixed effects and transform the variables into percentiles for robustness check as in Behn et al. Moreover, they proxy global variables by using US data. Consequently, their dataset includes credit to GDP gap, credit growth, global credit to GDP gap, global credit growth, housing prices growth, equity prices growth, debt service ratio, real GDP growth, nominal GDP growth, inflation, money market rate, yield curve, M3, current account, government balance, global housing prices growth, global equity prices growth, BAA and AAA spread and interaction terms: credit to GDP gap and credit growth; global credit to GDP growth and global credit growth; global credit to GDP gap and credit to GDP gap; and global credit growth and credit growth.

Joy et al. extend the set of potential leading indicators of Babecký et al. (2012) from macro-financial variables by incorporating domestic structural factors and international factors. One sided HP filter, i.e. real time, is applied while US data is used as a proxy for the world variable. Specifically, they select credit to GDP gap, credit growth, global credit to GDP gap, global credit growth, housing prices growth, equity prices growth, debt service ratio, real GDP growth, nominal GDP growth, inflation, money market rate, yield curve, M3, current account, government balance, global housing price growth, global equity price growth and BAA and AAA spread.

Alessi and Detken include around 30 potential variables as triggers. With respect to credit-related indicators, they consider the private sector total credit and bank credit aggregates, in the form of ratios to GDP, growth rates and ‘gaps’, i.e. trend deviations; sectoral credit aggregates, namely loans extended to households and non-financial corporations, as ratios to GDP and rates of growth; and the debt service ratio, for the whole economy as well as for households and non-financial corporations. The level of public debt is also considered. The macroeconomic variables include real GDP growth, the unemployment rate, the inflation rate, M3 growth and ‘gap’, the change in the real effective exchange rate, and the current account. The market-based indicators considered are the real short and long term interest rates, as well as real equity price growth. The real estate indicators are based on residential and commercial property prices and include property price growth and ‘gaps’, as well as standard valuation measures such as the house price to income and house price to rent ratios, taken in both forms: levels and ‘gaps’.

VI.2. Time Horizons and Frequencies

Subsection VI.2. presents a comparison of country coverage, time periods and data frequencies for all studied methodologies. Table 5.1 concisely summarises the information.

Baltussen et al. use variables over the time period from 1970 to 2010 for all EU countries. To calculate the interdependency index, they add important trading partners of the EU, including major G20 economies, Norway and Switzerland. Their forecast horizons are 4-12Q, 12-20Q and 4-20Q before a banking crisis, the variables are lagged by 1 period and quarterly data is used.

Bush et al. employ an unbalanced panel with annual data from 1980 to 2010 for 15 EU countries (AT, BE, DE, DK, ES, FI, FR, GR, IR, IT, NL, PL, PT, SE, and UK). The EU country sample is derived from their original data set of 22 OECD countries. The authors are bound to the annual data frequencies, since the employed variables (e.g. accounting leverage ratio) constrain their choice. Hence, they are compelled to use annual horizons instead of quarterly horizons. The quarterly horizons are transformed to annual horizons to ensure comparability. Bush et al. estimate their model using the different forecast time horizons (late (1-2Y), early (3-5Y) and total (1-5Y) horizons) which are comparable and in line with the spirit of quarterly horizons.

Antunes et al. use quarterly data for all EU countries from 1970 to 2010 to estimate their model. The authors attempt to use the longest time series available, which they in some cases aggregate from different sources using simple retropolation procedures. For some countries, there is no information for some of the variables used, thereby implying that these countries are not included in the multivariate analysis (Belgium, Bulgaria, Cyprus, Estonia, Croatia, Hungary, Lithuania, Luxembourg, Latvia, Malta, Poland, Romania, Slovenia and Slovakia). The final sample thus consists of 14 European countries.

Neudofer and Sigmund estimate their model for late (4-12Q ahead), early (12-20Q ahead) and total horizon (4-20Q ahead). The data are on quarterly basis over all available periods. They use overall the panel of 27 countries in their estimation (AT, BE, BG, CY, CZ, DK, EE, FI, FR, DE, GR, HU, IE, IT, LV, LT, LU, MT, NL, PL, PT, RO, SK, SI, ES, SE and UK).

Kauko constructs K1 indicators for the whole EU 27 sample, excluding cases with insufficient data, and separately for two Scandinavian countries (Finland and Sweden) over the period of 1981-2010. The data he uses for calculating the indicators are on a quarterly basis.

Behn et al. use a sample of 23 EU countries over the period of 1982Q2-2012Q3, using data with a quarterly frequency. They estimate their model over the late (4-12Q), early (12-20Q) and total (4-20Q) horizon.

Babecký et al. employ the same variables as Joy et al., plus they incorporate also credit interaction terms. They cover 17 EU countries over the period of 1970-2010 on quarterly basis with prediction horizons of 4-20Q, 12-20Q, and 4-12Q. All entered right-hand side variables are of lag 1 to account for publication delay.

Joy et al. focus on EU-27 countries and use quarterly data over the period from 1970 to 2010. The model estimation is executed over late (4-12Q), early (12-20Q) and total (4-20Q) horizon. The right-hand side variables are lagged by 1 quarter due to publication delay.
The sample of quarterly data considered by Alessi and Detken includes each of the EU countries, with data at best as of 1970Q1 until 2006Q1. All variables except market data are entered in the model with a 1 quarter lag to proxy for publication lags.

In general, there appears to be a limitation in out-of-sample forecasting as 40% of crises episodes occur between 2008-2012. Due to the high concentration the 2008 crisis should be excluded from the sample.

Author	***Countries Covered***	***Time Period***	***Data Frequency***
Baltussen et al.	EU 25	1970-2010	Quarterly
Bush et al.	22 OECD countries	1980-2010	Annual
Antunes et al.	all EU countries	1970-2010	Quarterly
Neudorfer, Sigmund	EU 27	1970-2010	Quarterly
Kauko	EU 27	1981-2010	Quarterly
Behn et al.	EU 23	1982Q2-2012Q3	Quarterly
Babecký et al.	EU 17	1970-2010	Quarterly
Joy et al.	EU 17	1970-2010	Quarterly
Alessi, Detken	all EU countries	1970Q1-2006Q1	Quarterly

Table 1: Country and Dataset Overview

VII. Results of Different Methodologies

This section focuses on presenting the main results and contributions of individual approaches to Early Warning.

Baltussen et al. find that the interdependency variables emerge as robust predictors over all prediction horizons. Moreover, both the (positive) credit gap and negative credit to GDP growth emerge as significant counterparts over all three horizons, signaling that both credit excesses and acceleration are likely to precede full-blown financial crises. Positive capital flows show good predictive power over the shorter horizons, reflecting that capital flow surges may directly precede sudden stop episodes.

Bush et al. demonstrate that their four variables (credit-to-GDP gap, accounting leverage ratio, liquid asset ratios, and price of risk) perform well among the different prediction horizons. The overall accuracy in terms of the AUROC remains stable throughout the horizons. In addition, Bush et al.

confirm their choice of variables by conducting further robustness checks like adding further variables (current accounts, spreads, GDP growth, inflation, etc.). The roles of credit gap, leverage, liquidity ratios and risk proxy are all consistent with their hypotheses and with the theory on bank balance sheets and risk pricing.

Antunes et al. observe the performance of all three probit models (simple, dynamic and dynamic with an extreme behaviour indicator) for the late (4-12Q ahead), early (12-20Q ahead) and total (4-20Q ahead) horizons. They use two different dependent variables: the ESCB Heads of Research crisis dummy and the vulnerability indicator proposed by Behn et al. In the first case, all explanatory variables are lagged by 4-12, 12-20 or 4-20 quarters, depending on the forecasting horizon. The same number of lags is included for the dependent variable in the dynamic specification. In the second case, the explanatory variables are lagged by 1-12, 12-20 or 1-20 quarters. To improve the comparability of results across methodologies, only the latter results are presented in Section VII (the former are presented in Antunes et al. 2014). As mentioned before, Antunes et al. also explored the role of exuberance in the explanatory variables. Since the specifications using the vulnerability indicator do not have enough degrees of freedom to ensure an adequate quality of the results, these are therefore not reported in this paper. Antunes et al. also run out-of-sample and out-of-period exercises, which also maintain good fitting properties (in the first case, the Nordic countries were excluded from the estimation, while in the second case the global financial crisis was excluded). As mentioned above, the explanatory variables used are the equity price index, the year-on-year growth rate of the debt-to-service ratio, the credit-to-GDP gap, and the year-on-year-growth rate of the house price index. However, in the model specification exercise, only the lags of variables which showed statistically significance at a 10% level were considered, thereby leading to the estimation of a more parsimonious model. Though in this model all the indicators provide statistically significant signs in several lags, the growth of debt-to-service ratios seems to provide particularly useful guidance for policymakers significantly ahead of crises. The credit-to-GDP gap also provides strong signals in all horizons.

Behn et al. evaluate models based on the usefulness (U) measure capturing a trade-off between the two error types (missing crises, false alarms) depending on policy-makers' preferences. They find that the inclusion of global variables adds value, in particular for shorter prediction horizons. They ultimately observe that the domestic credit-to-GDP gap is the most stable single explanatory variable across different models, while the inclusion of other macro-financial variables improves in particular earlier warning model performance. Global variables and banking capitalisation are important especially as one is closer to a crisis.

Neudorfer and Sigmund observe that the Bayesian approach is a natural way to model probabilities or in this case, crisis probabilities, and is more helpful with variable selection than the simple logit model. Moreover, the authors note that the country specific random coefficients (the dispersion from average slope) are very important. The potential added value of this approach becomes evident when looking at the main empirical results. The cross-country dispersion is relatively high for total asset growth, debt service and house price growth. Random coefficients for equity price index growth and real GDP growth were not included as their dispersion was low. Overall, debt service ratio shows the best results among all variables. If debt service ratio is included in the model then HP-filtered credit to GDP does not add further explanatory power. The authors also find that equity price growth is a good early crisis indicator in the build-up face of the observed crisis periods while real GDP growth serves mainly as a control variable.

Kauko finds that the AUROC over the total period (1-5 years) is larger than the one for a horizon of 3-5 years. Both Kauko and Behn et al. find that international factors play an important role in EWMs, e.g. global credit variables or contagion.

Babecký et al. in their original paper (Babecký et al., 2012) found domestic private credit seems to be a good early warning indicator across crises and horizons. The usefulness of private credit is obtained by minimizing policy makers' loss function with respect to Type I errors (missed crises) and Type II errors (false alarms). In terms of the horserace, over the horizon of 12-20 quarters BMA finds global variables significant while over the total horizon of 4 to 20 quarters a mixture of variables is selected that are also found significant over both late and early horizons. Babecký et al. execute out-of sample performance for UK and French crises in 1990s and find that prediction is more accurate closer to a crisis than over earlier horizons. Moreover, inclusion of global indicators proves to be useful as global credit, BAA and AAA spread or global housing price growth as well as debt service ratio emerge to be important early warning indicators over the observed horizons.

The binary tree methodology by Joy et al. allows explicitly for the fact that not all crises are alike and accommodates non-linearities by including conditional thresholds. Moreover, the approach accounts for variable interactions. Joy et al. extend the list of common leading indicators by incorporating domestic structural factors and international factors as they contribute to crises indirectly through their interaction with domestic variables. They find that a shallow yield curve coupled with high money market rates and low bank profitability are the most reliable indicators of banking crises. For currency crises, however, high domestic money market rates coupled with overvalued exchange rates are common predictors. Domestic structural characteristics, such as trade openness, do not seem to affect substantially the sensitivity to either type of crisis while international variables, such as world GDP and world credit, interact significantly with country variables. All in all, the CART method is a useful alternative to traditional methods.

Alessi and Detken find that, irrespective of the prediction horizon, among the most important variables there are house price valuation measures. Total and bank credit and public debt also feature among the most relevant indicators, as well as the long term yield and key macroeconomic indicators, namely unemployment, inflation and the current account. With respect to bank credit, the conditional relationship between gaps, ratios to GDP and rates of growth should be considered. The benchmark early warning trees grown on the selected best indicators (one for each prediction horizon, see section II) identify the respective early warning thresholds. The trees take into account policymakers' preferences between Type I and Type II errors; for comparison purposes it is assumed that the weight attached to missed crises equals that attached to false alarms. The main message from the classification trees is that the conditional relationships among different early warning indicators change depending on the prediction horizon considered.

VIII. Comparison of outcomes

Section VIII presents strengths and weaknesses of each early warning model analysed in this paper. The objective of this section is to provide a brief assessment of each method by stating explicitly their advantages and disadvantages.

Overall, Baltussen et al. find a good in-sample fit of their model both over the sample and for earlier banking crises, while maintaining a solid theoretical foundation. The multivariate model with an interdependency index outperforms a similar model without interdependency, as demonstrated by the statistically significant higher AUROC ($p < 0.01$). The approach opens avenues to further research not only as a specific model, but also as a method that has the potential to increase the explanatory power of other models. On the other hand, the proposed model has a more moderate out-of-sample performance. Another important downside is that the simplifications resulting from a controlled (EU-only) comparison and the implementation in traditional models, still suffers from some suspected biases. For example, performing an out-of-sample test by deleting selected crisis countries leads to the exclusion of important cross-section-dependent episodes and is thus counter-intuitive in this approach.

The model by Bush et al. does feature a solid model fit. The inclusion of country effects in the logit model yields higher and more significant AUROC than logit models without these effects. In addition, fixed effects and pooled models successfully pick up the recent crisis for Denmark and Sweden. Contrary to these strengths, the use of annual data may be too raw as the use of less-frequent data reduces the frequency of model updates. Less-frequent updates, in turn, reduce a model's power to forecast a crisis significantly before it occurs (Gramlich et al., 2010). However, this potential drawback has to be balanced against the novelty of their approach. The incorporation of the additional variables in comparison to the sole credit gap, especially the accounting leverage, seems to add additional explanatory power. Another drawback of the approach is the data requirements for the explanatory variables. Detailed bank-individual data is needed to construct the accounting leverage ratio. The scope of the countries is heavily constrained by the availability and coverage of banks balance sheets. Clearly, for more recent years the data availability is much better. In consequence, their sample and their findings do rely to a large extent on the last 15 years. In this spirit, it is not really surprising that the model does not pick up the Swedish 1991 crisis. Nonetheless, their model works quite well for the more recent times, but the generalizability of the results may be arguable due to the high dependence on the global financial crisis. Moreover, Bush et al. provide evidence on the importance of crisis dating. They are able to improve the regression results considerably by changing the dependent variable (ESCBs HoR crisis vs. Laeven and Valencia (2012) database). This implies that a regression may depend to a large extent on the crisis definition. For policy purpose, a clear definition of the forecast objective seems inevitable to obtain appropriate results.

As for Antunes et al., the encouraging results obtained show that it is beneficial to include dynamics via lagged dependent into early warning models. A large battery of metrics confirms that adding a dynamic component to early warning crises models substantially improves the quality of the results, most notably in reducing the percentage of missed crises and in increasing the percentage of those that are correctly predicted. Antunes et al. (2014) find that the best performance is always obtained for the total period estimation. In contrast, the early period estimations provide the weakest results. However, when using the vulnerability crisis indicator used in this exercise, it was not possible to achieve convergence in the early period. Going forward, the dynamic nature of the model may be improved through the inclusion of lags of the latent variable, which is still work in progress. Though the model seems to work well both in and out of sample, it is necessary to bear in mind that all crises are different. Therefore, it is always possible that a crisis driven by factors not taken into account in the model is not captured with this and other similar methodologies.

The greatest strength of the Bayesian random coefficient model used by Neudorfer and Sigmund is that cross country heterogeneity can be accounted for in a common model framework. In the early warning exercise almost each country has experienced pre-crisis periods that might have been caused by slightly different combination of adverse developments in various explanatory variables. Under these circumstances a common model framework that also allows for country specific effects makes perfect sense. A "fixed" coefficient for each variable measures the "average" impact on the pre-crisis probability whereas a "random" coefficient estimates the country-specific "random" impact. The combination of "fixed" and "random" coefficients gives the overall impact for each country and each variable. Moreover, for each variable the dispersion of the "random" effects gives a good indication of the cross-country heterogeneity. Finally, the Bayesian approach estimates the posterior mean of all coefficients in the same way and does not rely on a rather complex two-step general least square estimator as in classical statistical inference.

The methodology of Neudorfer and Sigmund manages to avoid several shortcomings of limited dependent variable panel models by accounting for country differences with respect to crisis indicators such as heterogeneity in signs for different countries. As for the pitfalls, the estimation results rely on data quality. The authors used longer time series for GDP data for all countries instead of BIS GDP data provided in the horse race database which as a result complicates method comparison. Moreover, it is well known in classical statistical interference random coefficient models are difficult to estimate. One needs to apply a two-step general least squares. Usually there is no closed-form solution to estimate the random coefficients. In a Bayesian estimation framework this weakness is less restrictive as one can use a hierarchical prior structure. Nevertheless there are convergence issues related to the posterior distribution of the coefficients. Depending on the data there might be not enough information in the data, especially for highly unbalanced panels, such that all posterior distributions of the coefficients converge to a stationary distribution as required in a Markov Chain Monte Carlo framework. From a practical point of view, the model provides a very good in-sample fit that might be problematic for out-of-sample forecasts.

Kauko's approach is not based on any kind of in-sample optimization; variables to be included or parameter values are not selected to suit the data. The main ideas have been derived from previous econometric research with data that only partly overlaps with the sample used in the analyses of this paper. Hence, the predictive power of the indicator is not likely to be limited to one specific part of the world during one specific era. Kauko proposes a simple statistic that is easy to compute and is well rooted in existing literature. The approach also correctly identifies the 1990s crises in both Sweden and Finland. However, the developed measure has less than 50% ability to correctly predict crises while the missing crises rate is at least 65.15%, making it quite unreliable.

Behn et al. achieve by their model a promising out-of-sample performance that signals, e.g. the Scandinavian crisis several years before its outbreak. As for weaknesses, the authors use fixed effects while some other approaches (Neudorfer and Sigmund) suggest that the inter-country differences are quite large. On a more general note, estimation over too large windows might not yield any meaningful results, e.g. Babecký et al. identify a mixture of early and late warning indicators as useful over the total horizon.

Babecký et al. use the Bayesian model averaging technique to effectively resolve model uncertainty. BMA is able to predict crisis onset quite well and also serves as an insurance against using just one model that might be misspecified. In addition, the method does not suffer from sensitivity to left-hand side variable specification. As for the methodology pitfalls, there appear to be fewer early warning indicators selected by BMA over longer horizons and for the total horizon of 4-20 quarters AUROC value is the smallest from among all horizons. Moreover, not all coefficient signs are intuitive and results tend to be sensitive to prediction horizon selection. In order to account for nonlinearities and variable interactions they need to be input explicitly, i.e. by hand.

Joy et al. identify a different set of variables as potential leading indicators for varying horizons. Over the late horizon of 1 to 3 years global credit growth and global growth of housing prices act as banking crisis triggers. Over the early horizon of 3 to 5 years global credit to GDP gap and money market rate emerge as the most important crisis triggers while over the total horizon BAA and AAA spread and global credit to GDP gap are of interest. Despite these results being quite decisive, there are several pitfalls to keep in mind for this method. First, AUROC values appear similar to those reported by other approaches but equally suffer from many false alarms. In line with previous observations, the smallest AUROC is for the total horizon. Second, usually only the variables in the first three levels of a binary tree are informative as crises triggers while variables in lower levels of a tree are not very intuitive. Third, trees are very sensitive to prediction horizon selection. Similarly, given a limited number of crises, time period and country coverage matter in the final estimation.

There are several strong aspects of the methodology by Alessi and Detken. First, the model is able to attain a very high AUROC despite excluding from the analysis the quarters immediately preceding the crisis and the crisis periods themselves, in which arguably the relationship between the explanatory variables and the crisis indicator is strongest. Second, the benchmark classification trees yield positive double digit usefulness, although their predictive performance has been penalized by "pruning" those nodes which make no economic sense but nevertheless help lowering the error rate. Third, as there is no time dimension in the decision tree framework, the out-of-sample prediction can be executed for every crisis, and in fact the "Random Forest" algorithm consists in carrying out *n* out-of-sample exercises. To mention weak points, the benchmark regression trees sacrifice robustness to some extent in favour of clearly identified early warning thresholds.

To conclude the outcome comparison section, the performance of individual EWMs is presented in accordance to the rules of the game, i.e. by reporting the size of an area under receiver operating characteristics curve (AUROC) which measures the forecast quality, the percentage of type I errors (missing crises), percentage of type II errors (false alarms) and by the measure of usefulness, which weights both error types with respect to policymakers´ preferences assuming balanced preferences. These four evaluation criteria are reported for each methodology over the three horizons, late of 4-12 quarters, early of 12-20 quarters and total of 4-20 quarters in tables 2, 3 and 4, respectively.

4-12 quarters Horizon	*AUROC*	*Type I error (%)*	*Type II error (%)*	*Absolute usefulness*
Baltussen et al.	0.875	12.0	31.0	0.287
Bush et al.	0.730	38.0	36.0	0.130
Antunes et al.	0.912	40.0	4.65	0.277
Neudorfer, Sigmund	0.989	8.9	2.3	0.210
Kauko	0.870	79.3	1.44	0.096
Behn et al.	0.920	5.6	24.7	0.349
Babecký et al.	0.892	5.6	34.8	0.298
Joy et al.	0.952	3.2	12.8	0.42
Alessi, Detken	0.925	38.0	4.0	0.29

Table 2: In-sample performance statistics over the late horizon

12-20 quarters Horizon	*AUROC*	*Type I error (%)*	*Type II error (%)*	*Absolute usefulness*
Baltussen et al.	0.885	8.0	30.0	0.308
Bush et al.	0.700	48.0	26.0	0.130
Antunes et al.	N/A	N/A	N/A	N/A
Neudorfer, Sigmund	0.9929	10.3	2.45	0.220
Kauko	0.720	65.15	25.04	0.05

Behn et al.	0.947	9.5	11.1	0.397
Babecký et al.	0.882	7.8	30.9	0.307
Joy et al.	0.945	0,9	11.8	0.436
Alessi, Detken	0.93	30.5	6.0	0.32

Table 3: In-sample performance statistics over the early horizon

4-20 quarters Horizon	***AUROC***	***Type I error (%)***	***Type II error (%)***	***Absolute usefulness***
Baltussen et al.	0.889	6.2	31.1	0.314
Bush et al.	0.720	50.0	23.0	0.135
Antunes et al.	0.974	16.0	2.1	0.4097
Neudorfer, Sigmund	0.9928	5.5	2.93	0.395
Kauko	0.893	88.75	1.74	0.047
Behn et al.	0.931	7.3	22.0	0.354
Babecký et al.	0.856	7.9	59.9	0.161
Joy et al.	0.8416	0.0	42.5	0.288
Alessi, Detken	0.928	48.0	3.0	0.245

Table 4: In-sample performance statistics over the total horizon

IX. Concluding Remarks

The Early Warning literature, in particular, has so far almost uniquely relied on two approaches, namely the signalling approach and the categorical dependent variable regression. The signalling approach has the advantage of being extremely straightforward. Indeed, the early warning signal is issued when the considered indicator breaches a pre-specified threshold, set by optimizing the past predictive performance. The downside of this approach is that it considers early warning indicators separately. Logit/probit regression, contrary to the signalling approach, offers a multivariate framework within which one can assess the relative importance of several factors. However, the model offers only an estimate of the contribution of each factor to the increase in the overall probability of a crisis, rather than a threshold value for each regressor. The early warning threshold is eventually set in a second step, referring to the estimated probability of a crisis. Moreover, this framework is unable to handle unbalanced panels and missing data, which is a serious issue in particular with credit data, with the result that the regression can ultimately be estimated only on a relatively short sample. Decision trees, and classification trees in particular, retain the advantages of both approaches as they are on the one hand very easy to explain and use, and on the other hand able to provide an early warning system where the relevant indicators are considered in a unitary framework.

If anything, the main conclusion to draw from the horse race exercise is that multivariate approaches, in their many variations, generate potentially very useful early warning results and offer considerable improvements over univariate signalling variables in terms of crisis prediction performance. Having said this, each multivariate approach has its strengths and weaknesses. For example, multivariate logit models tend to reduce both type I and type II errors (although admittedly less so than decision trees in the case of type II errors) and enable researchers to easily gauge the marginal contribution of each individual variable. Yet, the reliability of their results tends to be sensitive to exact model specification issues. In turn, decision trees tend to reduce type II errors as they allow for conditioning the effect of one variable on particular values of other variables, thereby giving very specific circumstances in which a signal is released. Also, they are less restrictive to the inclusion of, for example, level information as non-stationarity is not an issue for this type of approach. Still, not much is known yet regarding the out-of-sample performance of decision trees.

In the context of applying these results to macro-prudential policy and taking the strengths and potential weaknesses of each approach into consideration, there is a reason to argue that the use of a suite of multivariate models could be a superior choice when developing empirical macro-prudential policy instruments. This argument holds especially under the assumption that policy makers do not have strong ex ante preferences towards minimising type I versus type II errors. Moreover, as it seems logical that such preferences vary between policymakers in different jurisdictions and evolve over time, a broad empirical approach based on several early warning methods seems warranted, in particular as policy makers are still to discover the potential effects of macro-prudential instruments on financial stability and the real economy.

The previous analysis of various early warning models is used to highlight that some issues are worthy of further research with the aim of improving future contributions to this stream of literature.

First, a large number of lags leads to a reduction in the number of crises in the sample, i.e. over more distant horizons their crises coverage ratio declines. Second, the differences among countries are significant, thus estimating an EU-wide model, that disregards country effects, has its costs. Third, the contingency matrix over the late 4-12 quarter horizon shows more errors (false positives and false negatives) than over the whole horizon while the results are even worse for the early 12-20 quarter horizon. Fourth, some variables change signs for different estimation horizons which could either mean a change of regime or a correction of situation, indicating that nothing needs to be done. The two, however, are difficult to distinguish.

Bibliography

Abiad, A. (2003). "Early-Warning Systems: A Survey and a Regime-Switching Approach." IMF Working Paper No. 03/32.

Aikman, David; Alessandri, Piergiorgio; Eklund, Bruno; Gai, Prasanna; Kapadia, Sujit; Martin, Elizabeth; Mora, Nada; Sterne, Gabriel; Willison, Matthew (2009). "Funding Liquidity Risk in a Quantitative Model of Systemic Stability". Bank of England Working Paper, No. 372, London, June 2009.

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, Vol. 27, Issue: 3, Pages: 520-533, 2011.

Alessi, L. and C. Detken (2014). "Identifying excessive credit growth and leverage". ECB Working Paper, forthcoming.

Antunes, A., D. Bonfim, N. Monteiro, and P. Rodrigues (2014). "Early Warning Indicators of Banking Crises: Exploring New Data and Tools". Economic Bulletin Banco de Portugal April 2014.

Babecký, Jan; Havránek, Tomáš; Matějů, Jakub; Rusnák, Marek; Šmídková, Kateřina; Vašíček, Bořek (2012). "Banking, Debt and Currency Crises: Early Warning Indicators for Developed Countries." Working Paper Series 1485, European Central Bank.

Anselin, Luc (2002). "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." Agricultural economics, 27(3), pp.247-267.

Bank of England (2008, April). Financial Stability Report, London, April 2008.

Basel Committee on Banking Supervision (2010). Guidance for National Authorities Operating the Countercyclical Capital Buffer. December 2010,

Berg, A. and C. Patillo (1998). "Are Currency Crises Predictable? A Test." IMF Working Paper No. 98/154.

Bilson, J. F. O. (1979). "Leading Indicators of Currency Devaluations." Columbia Journal of World Business 14, pp. 62–76.

*Breiman, Leo (2001). "*Random Forests." Machine Learning 45 (1): 5–32.

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

Davis, Philip E.; Karim, Dilruba (2008). "Comparing Early Warning Systems for Banking Crises, Journal of Financial Stability." Vol. 4 (2008), Iss. 2, pp. 89-120.

Demirgüç-Kunt, Asli; Detragiache, Enrica (2005). "Cross-Country Empirical Studies of Systemic Bank Distress: A Survey". IMF Working Paper, No. 96/05, Washington 2005.

Eichengreen, Barry; Rose, Andrew K; Wyplosz, Charles (1996). "Contagious Currency Crisis". NBER Working Paper No. 5681.

Elsinger, Helmut; Lehar, Alfred; Summer, Martin (2006). "Using Market Information for Banking System Risk Assessment". International Journal of Central Banking, Vol. 2 (2006), No. 1, pp. 137-165.

Fabian Valencia & Luc Laeven (2012). "Systemic Banking Crises Database: An Update." IMF Working Papers 12/163, International Monetary Fund.

Forbes, K.J., Warnock, F.E. (2012). "Capital Flow Waves: Surges, Stops, Flight, and Retrenchment." Journal of International Economics 88, pp.235-251.

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

Frankel, J. A. and G. Saravelos (2010). "Are Leading Indicators of Financial Crisis Useful for Assessing Country Vulnerability? Evidence from the 2008–09 Global Crisis." NBER Working Paper No. 16047.

Gaytán, Alejandro; Johnson, Christian A. (2002). "A Review of the Literature on Early Warning Systems for Banking Crises." Central Bank of Chile Working Paper, No. 183, Santiago, October 2002.

Jiménez-Rodríguez, R. and Sánchez, M. (2004). "Oil Price Shocks and Real GDP Growth; Empirical Evidence for OECD Countries." Working Paper Series 362, European Central Bank.

Jordà, Òscar; Schularick, Moritz and Taylor Alan M. (2011). "When Credit Bites Back: Leverage, Business Cycles, and Crises." Working Paper Series 2011-27, Federal Reserve Bank of San Francisco.

Kakamu, Kazuhiko, and Hajime Wago (2010). "Estimation of Regional Business Cycle in Japan Using Bayesian Panel Spatial Autoregressive Probit Model." Working paper. Available at: http://www. mssanz.org. au/modsim05/proceedings/papers/kakamu_2. pdf, 2007.

Kaminsky, G. L. and C. M. Reinhart (1996). "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems." Board of Governors of the Federal Reserve System, International Finance Discussion Papers 544.

Kaminsky, Graciela L.; Lizondo, Saul; Reinhart, Carmen M. (1998). "Leading Indicators of Currency Crises". IMF Staff Paper, Vol. 45 (1998), No. 1, pp. 1-48.

Kaminsky, Graciela L.; Reinhart, Carmen M. (1999). "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems". American Economic Review, Vol. 89 (1999), No. 3, pp. 473-500.

Kauko, Karlo (2014). "How to Foresee Banking Crises, a Survey of Empirical Literature." Economic Systems (forthcoming).

Krugman, P. (1979). "A Model of Balance-of-Payments Crises." Journal of Money, Credit and Banking 11(3), pp. 311–325.

Laeven, L. and Valencia, F. (2008). "Systemic Banking Crises: A New Database." International Monetary Fund, WP/08/224.

Luc Laeven & Fabian Valencia (2013). "Systemic Banking Crises Database." IMF Economic Review, Palgrave Macmillan, vol. 61(2), pages 225-270, June.

Lilien, D.M. (1982). "Sectoral Shifts and Cyclical Unemployment." Journal of Political Economy 90 (4), pp. 777-793.

Lin, Chin-Shien; Khan, Haider A.; Wang, Ying-Chieh; Chang, Ruei-Yuan (2006). "A New Approach to Modelling Early Warning Systems for Currency Crises: Can a Machine-Learning Fuzzy Expert System Predict the Currency Crises Effectively?" Center for International Research on the Japanese Economy Discussion Paper, CIRJE-F-411, Tokyo, April 2006.

Mayes, David G. & Hanno Stremmel, (2014)."The Effectiveness of Capital Adequacy Measures in Predicting Bank Distres." SUERF Studies: 2014/1.

Novo, Álvaro (2003). "Contagious currency crises: a spatial probit approach," Working Paper 5-03, Banco de Portugal.

Peria, M. S. M. (2002). "A Regime-Switching Approach to the Study of Speculative Attacks: A Focus on EMS Crises." In James D. Hamilton and Baldev Raj, editors, Advances in Markov-Switching Models. Heidelberg: Physica-Verlag.

Rose, A. K. and M. M. Spiegel (2009). "Cross-Country Causes and Consequences of the 2008 Crisis: Early Warning." NBER Working Paper No. 15357.

Appendix A

Table A

Right-hand side variables											Left-hand side variable	Author
Other variables	**Global variables, interactions**	**Banking sector variables**	**Debt, government balance**	**Current account**	**Debt service ratio**	**Rates**	**GDP**	**Equity prices**	**House prices**	**Credit data**		
capital flows					Debt service ratio	REER	GDP growth			private credit to GDP gap, private credit to GDP growth	ESCBs' HoR, Laeven&Valencia for non-EU	*Baltussen et al.*
VIX proxy		leverage ratio, liquid asset ratio								Credit –to-GDP gap	ESCB HoR, Laeven&Valencia	*Bush et al.*
					Debt service ratio growth			equity price index	house price growth	credit to GDP gap	ESCBs' HoR	*Antunes et al.*
equity growth, capital and reserve growth		total asset growth			Debt service ratio, debt service ratio gap		real GDP growth		house price growth	Credit to nominal GDP gap	ESCBs' HoR but extended	*Sigmund, Neudorfer*
	contagion data			Current account						Credit data	ESCBs' HoR	*Kauko*
	global variables (average of USA, CAN, JAP, EU27)	aggregate capitalisation and profitability				inflation	GDP	stock prices	house prices	Credit to GDP gap	ESCBs' HoR	*Behn et al.*
yield curve, M3, BAA and AAA spread	USA proxies global variables		government balance	Current account	Debt service ratio	inflation, money market rate	real and nominal GDP growth	equity price growth	house price growth	credit to GDP gap, credit growth	ESCBs' HoR	*Joy et al.*
unemployment rate, M3 growth			public debt	Current account	Debt service ratio	inflation, REER, short, long term interest rates	real GDP growth	equity prices	residential and commercial property prices	aggregate private sector total credit/GDP gap, total credit and bank credit, credit to households and non-financial corporations	ESCBs' HoR	*Alessi, Detken*
yield curve, M3, BAA and AAA spread, variable interactions	USA proxies global variables		government balance	Current account	Debt service ratio	inflation, money market rate	real and nominal GDP growth	equity price growth	house price growth	credit to GDP gap, credit growth	ESCBs' HoR	*Babecký et al.*

Appendix B

Horserace contributions (Workshop on Early Warning Tools and Tools for Supporting Macroprudential Policies (ECB, Frankfurt, 29 November 2013):

Lucia Alessi and Carsten Detken (2014). "Identifying excessive credit growth and leverage". ECB Working Paper, forthcoming.

Antonio Antunes, Diana Bonfim, Nuno Monteiro and Paulo M. M. Rodrigues (2013). "Forecasting banking crisis: A Dynamic Panel Probit Approach".

Simon Baltussen, Jon Frost, Ruben van Tilburg (2013). "One for all and all for one: A multivariate probit approach with interdependency".

Markus Behn, Carsten Detken, Tuomas A. Peltonen and Willem Schudel (2013). "Setting Countercyclical Capital Buffers Based on Early Warning Models".

Oliver Bush, Rodrigo Guimarães and Hanno Stremmel (2013). "Beyond the credit gap: bank balance sheets, credit and price of risk".

Benjamin Neudorfer and Michael Sigmund (2013). "A Bayesian Random Coefficient Logit Approach to Modeling the Crisis Dummy".

Karlo Kauko (2013). "Credit Growth Is Not Always Dangerous".

Mark Joy, Marek Rusnák, Kateřina Šmídková and Bořek Vašíček (2013). "Rules of Thumb for Systemic Banking Crises using Classification and Regression Trees".

Jan Babecký, Tomáš Havránek, Jakub Matějů, Marek Rusnák, Kateřina Šmídková and Bořek Vašíček (2013). "Predicting Systemic Banking Crises using Bayesian Model Averaging".