

Comparing different early warning systems: Results from a horse race competition among members of the MaRs

L. Alessi, A. Antunes, J. Babecký, S. Baltussen, M. Behn,
D. Bonfim, O. Bush, C. Detken, J. Frost, R. Guimarães, T.
Havránek, M. Joy, K. Kauko, J. Matějů, N. Monteiro, B. Neudorfer, T.
Peltonen, P. Rodrigues, M. Rusnák, W. Schudel, M. Sigmund, H.
Stremmel, K. Šmídková†, R. van Tilburg, **B. Vašíček**, D. Žigraiová

Final ESCB MaRs Conference

Frankfurt

23 June 2014

1. Introduction

- Over the recent decades researchers in academia and central banks have developed EWS designed to warn policy makers of potential future economic and financial crises.
- The aim of this paper is to elaborate on these advances in the field of EWS in a rather unique exercise:
 - we compare the performance of nine distinct empirical models resulting from the work of the MaRs
 - all models use the same database of crises created by MaRs and comparable sets of potential EWI
 - we evaluate the models' relative usefulness by comparing the ratios of false alarms and missed crises

1. Introduction

- Most models depart from traditional discrete choice (logit/probit) models (some using less standard indicators)...
- ...introducing dynamics (dynamic probit) and accounting for nonlinearities (binary regression trees, random forest)...
- ... 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 account (random coefficient models).

2. Related literature

- The literature on EWI of crises is extensive:
 - Univariate (signalling) (Kaminsky and Reinhart, 1999, Kaminsky, et al. 1998) vs. multivariate (regression) approaches (Frankel and Rose, 1996, Berg and Patillo, 1999).
 - Minimization of the signal-to-noise ratio (Kaminsky 1999) or a policymaker's loss function (Demirgüç-Kunt and Detragiache 1999), explicit choice on missed crises and false alarms (Alessi and Detken 2011)
 - multinomial models (Bussiere and Fratzscher, 2006), continuous crisis indicators (Frankel and Saravelos, 2010)
 - Markov switching models (Peria, 2002), recursive tree method (Manasse and Roubini, 2009).

3. Comparing EWS in a Horse race - rules

- 1. the same LHS variable: banking crises and the related pre-banking crises periods based on Babecký et al. 2012
- 2. the same dataset for all the authors was provided, only one-sided (backward-looking) filters were allowed
- 3. a time window of 20 to 4 Q before a banking crisis and two subperiods 20-12 Q and 12-4 Q before the crisis.
- 4. the RHS variables used either extracted from the provided dataset or from other publicly available sources.
- 5. incorporate as many EU countries as possible and to provide out-of-sample predictions (before 2007)
- 6. evaluate the in-sample fit of each horizon applying the so-called 'signalling approach'

Figure 1. Contingency matrix

		Actual class C_j	
		1	0
Predicted class P_j	1	<i>True positive (TP)</i>	<i>False positive (FP)</i>
	0	<i>False negative (FN)</i>	<i>True negative (TN)</i>

Different approaches deliver different combination of missed crises (Type I error) and false alarms (Type II error) for given preferences.

4. Comparing EWS in a Horse race - methodologies

- 1. Baltussen et al. (DNB) - a probit model with an interdependency index (exogenous spatial lag captures country interdependency explicitly by simulating financial conditions in partner countries).
- 2. Bush et al. (BoE) - a logit model with a principal components analysis (PCA) in a data rich environment with international data availability to guide their choice of RHS variables.
- 3. Behn et al. (DB/ECB) – a logit model with specific transformations of credit variables and global variables.
- 4. Antunes et al. (BdP) - general dynamic probit model to account for time dependence with lags of the dependent variable.

4. Comparing EWS in a Horse race - methodologies

- 5. Kauko (BoF) - a univariate signalling approach based on the assumption that long lasting excessive credit growth gradually becomes more problematic, uses ratios of loan stocks on GDP.
- 6. Neudorfer and Sigmund (OeNB) - a Bayesian random coefficient logit panel model, a “general” coefficient which is common for all countries and a “country-specific” coefficient
- 7. Babecký et al. (CNB) - a Bayesian model averaging that considers model combinations based on weights determined according to the models’ fit.

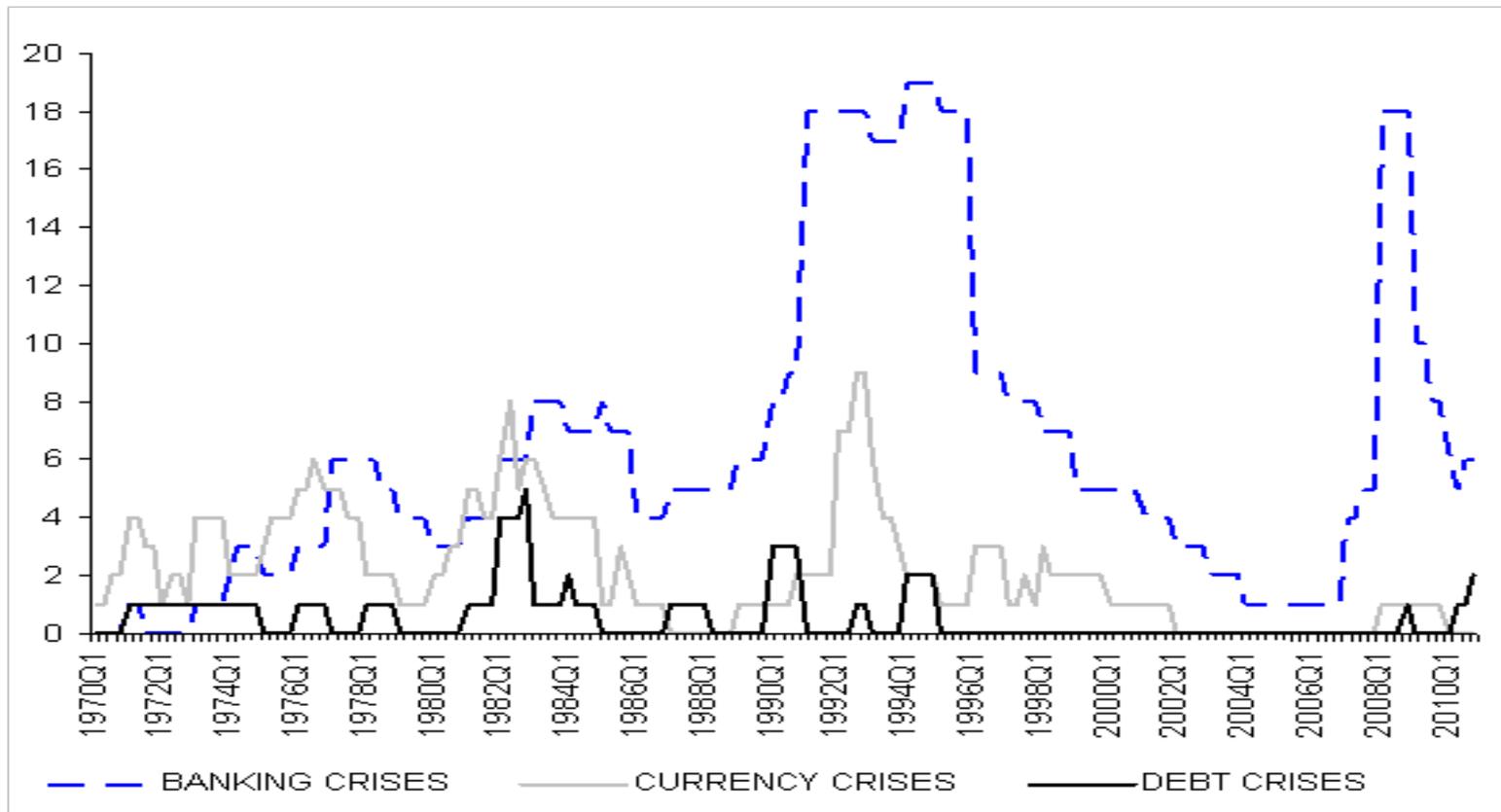
4. Comparing EWS in a Horse race - methodologies

- 8. Joy et al. (BoE/CNB) - a Binary Classification Tree (BCT), which allows for interactions among key variables and determines critical tipping points. CART thus searches through different possible splits for all explanatory variables and selects those splits that best separate crisis episodes from no-crisis episodes.
- 9. Alessi and Detken (ECB) - the Random Forest method, 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, unlike BCT is not influenced by outliers.

5. Definition of crises

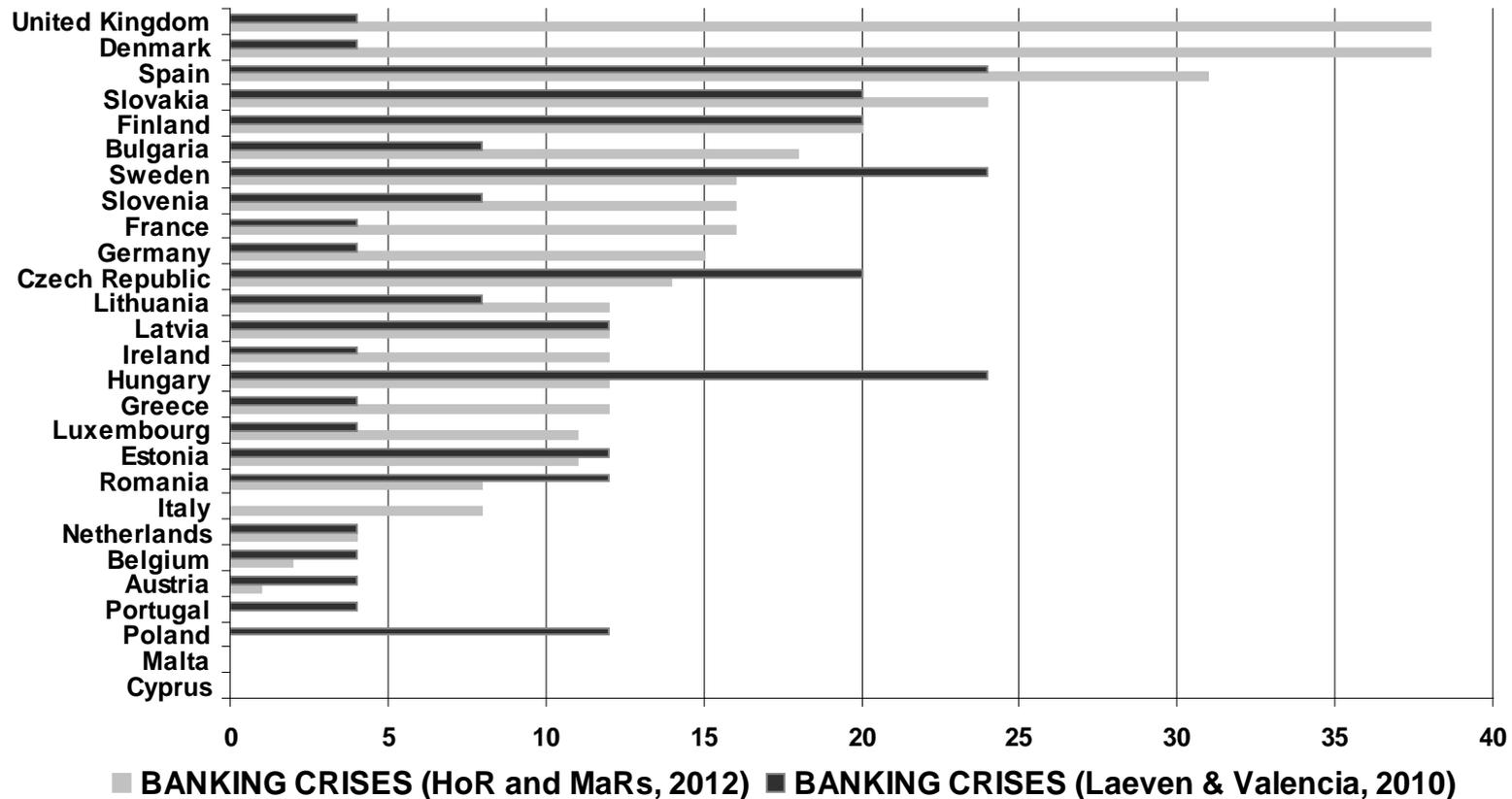
- Babecký et al. (2012, 2014) - a quarterly database of various types of crisis occurrence in EU-27 + OECD for 40 years (1970-2010): **banking crises**, debt crises, currency crises.
- **Systematic synthesis of the literature:** Caprio and Klingebiel (2003); Detragiache and Spilimbergo (2001); Kaminsky (2006); Kaminsky and Reinhart (1999); Laeven and Valencia (2008, 2010, 2012); Levy-Yeyati and Panizza (2011); and Reinhart and Rogoff (2008, 2011).
- and country experts asked for country-specific inputs (**for ESCB countries the HoR** provided with information about crisis occurrence in their respective countries alongside the corresponding crisis definitions).

Figure 2. Number of developed countries in crisis: 1970:Q1–2010:Q4



The database shows that developed countries have had their share of crises.

Figure 3. Example of disagreement: Number of quarters spent in banking crises



For the EU-27, experts add interesting views compared to the literature.

5. Definition of crises

- The LHS variable is set to 1 between (and including) 20 to 4 Q to a banking crisis onset and to 0 for all other quarters in the data.
- To overcome pre-crisis and crisis bias, all country quarters which either fall within the period from 3 Q before the onset of a banking crisis up until the end of a banking crisis excluded. Moreover, all quarters as of 2006q1 are excluded (i.e. 20 quarters before the end of the data series).
- We still try to further homogenize the exercise (especially the RHS). Indeed, we confirm that EWI exercises are conditioned by rather subjective preferences and interpretations by the researchers.

6. Potential EWIs

Author	Right-hand side variables										
	Other variables	Global variables, interactions	Banking sector variables	Debt, government balance	Current account	Debt service ratio	Rates	GDP	Equity prices	House prices	Credit data
<i>Baltussen et al.</i>	capital flows					Debt service ratio	REER	GDP growth			private credit to GDP gap, private credit to GDP growth
<i>Bush et al.</i>	VIX proxy		leverage ratio, liquid asset ratio								Credit –to-GDP gap
<i>Antunes et al.</i>						Debt service ratio growth			equity price index	house price growth	credit to GDP gap
<i>Sigmund, Neudorfer</i>	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
<i>Kauko</i>		contagion data			Current account						Credit data
<i>Behn et al.</i>		global variables (average of USA, CAN, JAP, EU27)	aggregate capitalisation and profitability				inflation	GDP	stock prices	house prices	Credit to GDP gap
<i>Joy et al.</i>	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
<i>Alessi, Detken</i>	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
<i>Babečský et al.</i>	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

7. Results of different EWI models

- 1. Baltussen et al. (DNB) - the interdependency variable, credit-to-GDP gap and negative credit to GDP growth perform well over all three horizons, positive capital flows over the shorter horizon.
- 2. Bush et al. (BoE) - credit-to-GDP gap, accounting leverage ratio, liquid asset ratios, and price of risk perform well over all three horizons.
- 3. Behn et al. (DB/ECB) – credit-to-GDP gap performs well over all three horizons, macro-financial variables over the longer horizon, global variables and banking capitalisation over the shorter horizon.
- 4. Antunes et al. (BdP) - credit-to-GDP gap performs well over all three horizons, debt-to-service ratios over the longer horizon.¹⁵

7. Results of different EWI models

- 5. Kauko (BoF) - the growth of the loan stock in percentages (12-month difference in the loan stock to 5-year moving average of nominal GDP), global credit variables perform well for all three horizons .
- 6. Neudorfer and Sigmund (OeNB) - debt service ratio performs well for all three horizons (and HP-filtered credit to GDP does not add further explanatory power), equity price growth over the shorter horizon. The cross-country dispersion is relatively high for total asset growth, debt service and house price growth.
- 7. Babecký et al. (CNB) - global indicators (global credit, BAA spread, global housing price growth) and debt service ratio perform well for all three horizons .

7. Results of different EWI models

- 8. Joy et al. (BoE/CNB) - shallow yield curve, coupled with high money market rates and low bank profitability performs well for all three horizons.
- 9. Alessi and Detken (ECB) - house price valuation measures, total and bank credit, public debt perform well for all three horizons. With respect to bank credit, the conditional relationship between gaps, ratios to GDP and rates of growth should be considered.

Credit-to-GDP gap and debt service ratio emerge as the most consistent indicators across models.

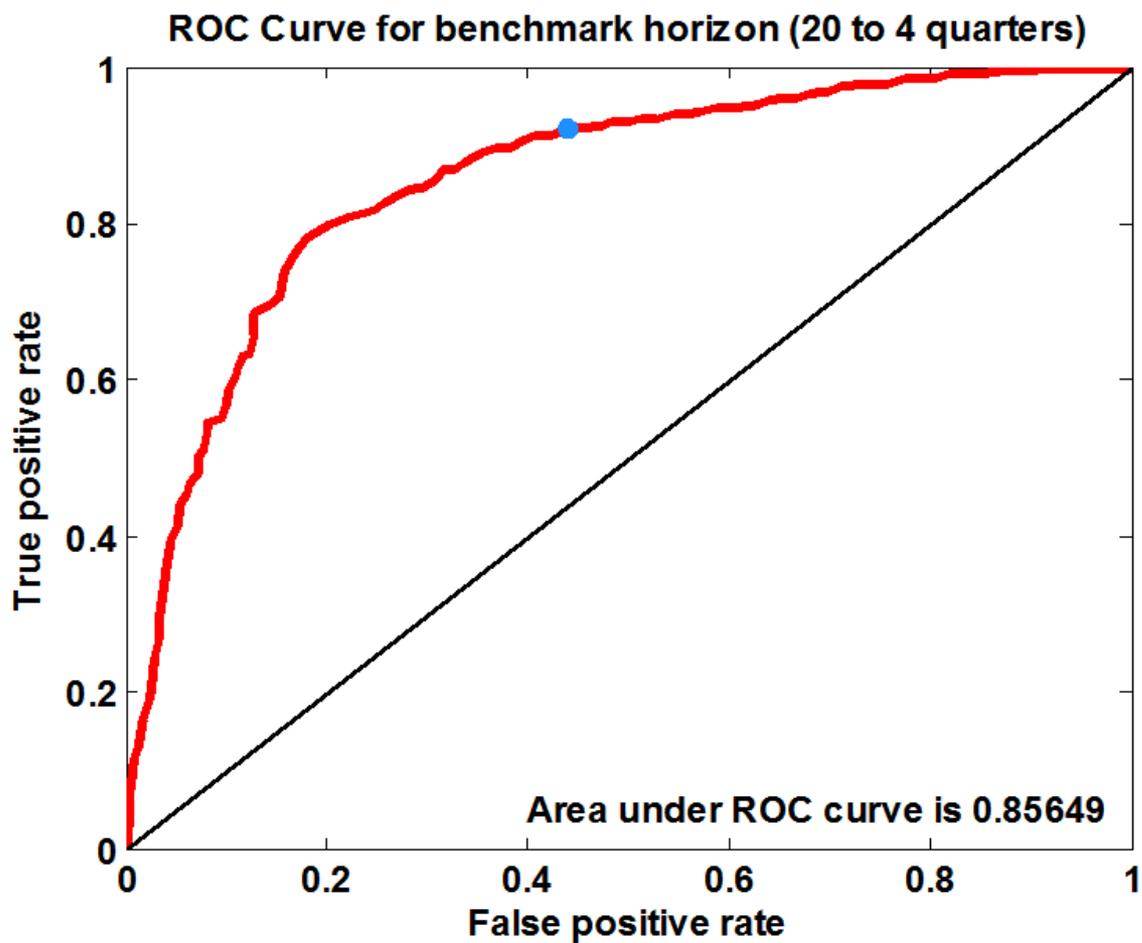
8. Comparison of outcomes

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

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

The performance of individual EWMs according to: a) the size of an area under receiver operating characteristics curve (AUROC) (measures the forecast quality) – summary measure for any preferences, b) the percentage of type I errors (missing crises), c) percentage of type II errors (false alarms) and d) the measure of usefulness (weights both error types) – derived under assumption of balanced preferences ($\theta = 0.5$).

Figure 4. AUROC



9. Concluding remarks

- The Early Warning literature has so far mostly relied on two empirical approaches, namely the signalling approach and the categorical dependent variable regression.
- There are now some new methods that overcome limitations of the former two approaches, e.g. classification and regression trees that allow interactions of different variables (like multivariate regressions) and identify the individual thresholds (like univariate signalling approach)
- Finding of empirical exercises varies due to a) different empirical methodology, b) different empirical choices and c) different underlying datasets of crises and potential EWI.

9. Concluding remarks

- Our joint exercise aims to provide a comparative analysis using different methodologies on a homogeneous dataset.
- The general conclusion is that multivariate approaches generate potentially very useful early warning results and offer considerable improvements over univariate signalling variables in terms of crisis prediction performance but each multivariate approach has its strengths and weaknesses.
- The use of a suite of multivariate models could be a superior choice when developing empirical macro-prudential policy instruments, especially as policy makers do not have strong ex ante preferences towards minimizing type I versus type II errors.

9. Concluding remarks

- The specific findings of the horse race can be summarized:
- The differences among countries are significant, thus estimating an EU-wide model has its costs.
 - The contingency matrix over the late 4-12 Q 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 Q horizon.
 - Some variables change signs for different estimation horizons which could be due to regime change or a correction of situation.
 - Among the many results, we find evidence supporting the importance of accounting for private sector credit growth and debt service ratio.



Thank you for your attention

Bořek Vašíček, borek.vasicek@cnb.cz