



Bank of Russia



# Incorporating financial development indicators into early warning systems

WORKING PAPER SERIES

No. 58 / July 2020

Alexey Ponomarenko  
Stas Tatarintsev

**Alexey Ponomarenko**

Research and Forecasting Department, Bank of Russia. Email: [ponomarenkoa@cbr.ru](mailto:ponomarenkoa@cbr.ru)

**Stas Tatarintsev**

Research and Forecasting Department, Bank of Russia. Email: [tatarintsevsa@cbr.ru](mailto:tatarintsevsa@cbr.ru)

The authors are grateful to Ramis Khabibullin for his valuable comments and suggestions.

Bank of Russia Working Paper Series is anonymously refereed by members of the Bank of Russia Research Advisory Board and external reviewers.

Cover image: Shutterstock.com

© Central Bank of the Russian Federation, 2020

**Address:** 12 Neglinnaya street, Moscow, 107016  
**Tel.:** +7 495 771-91-00, +7 495 621-64-65 (fax)  
**Website:** [www.cbr.ru](http://www.cbr.ru)

All rights reserved. The views expressed in this paper are solely those of the authors and do not necessarily reflect the official position of the Bank of Russia. The Bank of Russia assumes no responsibility for the contents of the paper. Any reproduction of these materials is permitted only with the express consent of the authors.

---

## Abstract

We set up an early warning system for financial crises based on the Random Forrest approach. We use a novel set of predictors that comprises financial development indicators (e.g. levels of credit to GDP ratio) in addition to conventional imbalances measures (e.g. credit gaps). The evaluation of the model is conducted using a three-step procedure (i.e. training, validation and testing sub-samples). The results indicate that combining financial imbalances and financial development indicators helps to improve the out-of-sample accuracy of the early warning system.

**JEL-classification:** C40, C52, G01, E44

**Keywords:** Early warning indicators, financial crisis, financial development, credit gap, random forest

## 1. Introduction

The last financial crisis stimulated a new wave of research on the early warning indicators models. The main problem is a forehanded reaction on the rise risk in the financial system and smoothing consequences of an oncoming financial crisis.<sup>1</sup> Therefore, there is a need to find an appropriate early warning system (EWS) to define a right time of macro prudential actions.

The existing EWSs are usually based on a set of indicators of financial imbalances: most notably the credit gap measure (Drehmann et al. 2011) as well as other subsidiary cyclical financial risk measures (Giese et al. 2014, Detken et al. 2014, Tölö et al. 2018, Gersl and Jasova 2018). If such EWSs signalises rising financial risks, the introduction of macro prudential actions should be considered.

This paper makes several contributions to this strand of literature. Firstly, the existing EWSs rarely make use of financial development indicators (FDIs) although the evidence of their relevance for credit cycle analysis is significant (Cottarelli et al. 2005, Égert et al. 2006, Buncic and Melecky 2014, Bahadir and Valev 2015, Naceur et al. 2019). In this paper we combine the traditional set of cyclical financial imbalances measures with the set of FDI indicators.

Another novelty is that for this purpose we use the machine learning algorithms in the spirit of Holopainen and Sarlin 2017, Alessi and Detken 2018, Beutel et al. 2019. Namely, we employ the Random Forest modelling approach, which is arguably well suited for capturing the potential non-linear relationship between financial development, financial imbalances and the probability of a financial crisis.

The rest of the paper is structured as follows. Section 2 presents the set of indicators used in the early warning systems and defines the financial crises events. Section 3 outlines the modelling set-up and the model evaluation algorithm. Section 4 discusses the design of experiments and presents the main empirical results. Section 5 provides more detail on the role of the interaction between financial imbalances and financial development measures. Section 6 concludes.

---

<sup>1</sup> See Jerman and Quadrini (2012), Mumtaz and Zanetti (2016), and Zanetti (2019) who show that financial factors are also critical for business cycle fluctuations.

## 2. The data

The set of the early warning indicators by category is presented in Table 1. The financial imbalances categories naturally include the standard credit gap measure, which is recommended by the Basel III regulation, the real credit growth indicators (Schularick and Taylor 2012, Gersl and Jasova 2018) and changes in the credit-to-GDP ratio (Jordà et al. 2013). We also augment the dataset with the GDP growth rate (Borio and Lowe 2002).

Financial development is represented by the level of credit-to-GDP ratio and the financial development indicators reported by the IMF (see Svirydzenka 2016 for a detailed description). The data sources OECD, BIS and IMF databases (Table 4 in the Appendix for details).

All series apart from FIA, FIE and FMA indicators have the quarterly frequency of initial data. Consequently, we apply linear interpolation for these financial development indicators. Also in rare cases, when the data on some individual indicators are unbalanced, we fill the missing observation with the extremely large big values of  $10^{18}$  (meaning that a missing observation never prevents issuing a signal). The description of these indicators is presented in the Appendix.

**Table 1.** List of indicators

<b>Category of indicators</b>	<b>Indicators</b>
Credit gap	Credit-to-GDP gap
Financial imbalances	GDP one year growth rate GDP two year growth rate Total credit one year real growth rate Total credit two year real growth rate Credit-to-GDP one year growth rate Credit-to-GDP five year growth rate
Financial development	Credit-to-GDP ratio, per cent Financial institution access (FIA) Financial institution efficiency (FIE) Financial market access (FMA)

We use the database for banking crisis compiled by Babecky et al. (2014) and Laeven and Valencia (2018). The former database covers the banking crisis episodes that take place over 1970–2010, whereas in the latter source is used to extend the data until 2017. In the cross-section we have 30 countries with observations from 1970 Q1 to 2016 Q4 and 6 more emerging market economies with observations in period from 2006 Q1 to 2016 Q4 (the latter are only used in the test sub-sample as described in Section 4). See Table 5 in the Appendix for details. However, we do not include post-2008 observations into the main dataset (although the results for the extended dataset are reported in Figures 6–8 in the Appendix). No new financial crises were detected in that period and the results obtained using these observations mostly reflect whether the signalling indicators were able to dissipate fast enough to stop producing false warnings.

Finally, in constructing the dependent variable (i.e. binary series of the banking crises) we take into account policy lag and want to prevent a forthcoming banking crisis. Therefore, initial binary series of banking crisis is modified in the same way as in Alessi and Detken 2018. Namely, 12 quarters before the crises events are set to 1, the crisis events and 3 quarter before the start of crisis events were removed from the dataset, all other observation are set to 0. The sample is unbalanced (i.e. only about 13% of the observations equals 1). As a result, we define as correct any warning signals issued during the period of 2–4 years preceding the start of a crisis; signals issued up to three quarters before the crisis or during the crisis are discounted and all other signals are assumed false.

### 3. The model

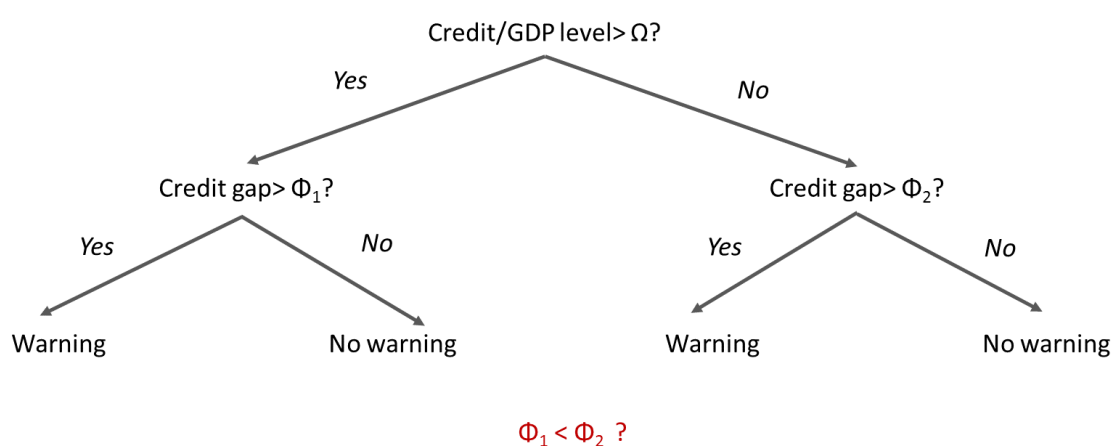
We use the modelling approach that is based on a decision tree technique. Arguably, this approach is suitable for modelling the interplay between the cyclical fluctuations of financial imbalances, the long-run trends in financial development and the risk of a financial crisis. A stylised example is presented in Figure 1. Presumably, an economy becomes more vulnerable to financial crises at higher financial development stages.<sup>2</sup> A decision tree may reflect this relationship by, for example, allocating the

---

<sup>2</sup> This presumption is in line with the empirical findings that show that the size of the financial sector exceeding a certain threshold negatively affects economic growth (see e.g. Cecchetti and Kharroubi 2012, Arcand et al. 2015, Naceur et al. 2018, Yang and Su 2018). This presumption may also reflect the concept of rapid development of the financial sector during the catch-up phase in emerging markets that is no longer regarded as benign once the convergence is achieved (see e.g. Égert et al. 2006, Buncic and Melecky 2014,

economies into different classes based on the level of financial development at the first stage and setting different critical thresholds for the financial imbalances variables at the second stage. Notably, this means that our dataset may include both cyclical and non-stationary variables, which is problematic when alternative multivariate modelling techniques (e.g. probit regressions) are used.

**Figure 1.** An example of a hypothetical decision tree



### 3.1 The modelling set-up

A Random Forest (Breiman 2001) is a popular machine learning technique which is widely used in classification problems. It is comprised of tens or hundreds of different trees. Presumably, this helps to prevent overfitting in comparison with a single binary classification tree. Even though the signal issued by each individual tree may be noisy, averaging across signals may help to obtain less biased and more precise prediction of classification. The mean prediction of the averaging of classification results is a probability of belonging to the positive or negative class.

The algorithm for the splitting subsample in each trees and selecting the single indicator with corresponding indicator's threshold is carried out with impurity measure. We use a standard impurity measure which is the Gini index:

$$\text{GINI}(f) = \sum_{i=1}^n f_i(1 - f_i)$$

---

Bahadir and Valev 2015). However, the assumption is used here only for illustrative purposes. The actual model is flexible enough to accommodate other types of non-linear interplay between financial development and financial imbalances.

where  $f_i$  is the fraction of values belonging to  $i$ 's class.

We use the Random Forest algorithm from the standard Python library “scikit-learn”. We are making a hyperparameter optimisation for five hyperparameters: four defined in the *RandomForestClassifier* function and one more is described from the *predict\_proba* function. Hence, we tested model for different hyperparameters with corresponding values from Table 2.

**Table 2.** List of hyperparameters

Hyperparameters	Python function	Tested values
n_estimators	<i>RandomForestClassifier</i>	{20, 30, 40, 50, 60, 70, 80, 90, 100, 300, 800}
max_depth		{2, 3, 5, 7, 10, 12, 15, 18, 25}
max_feature		{None, sqrt, log2}
class_weight		{1, 2, 5, 7, 10, 50, 80, 100, 150, 200, 1500}
threshold	<i>predict_proba</i>	{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8}

The hyperparameter *n\_estimators* determine the number of trees in the forest. Higher number of trees provide the better performance. However, more trees slow down the training process considerably. Hence, we search for an optimal number of trees for balance of training time and performance.

*max\_depth* is another important hyperparameter which represents the maximum depth of the individual tree in the forest. If the tree is very deep, it will have more splits and may capture more information about the data. On the other hand, overfitting may occur if trees are too deep (i.e. the trees perfectly predicts all of the train data, but it fails to generalise the findings for test data).

*max\_feature* determines a maximum number of indicators in the individual tree (reported in Table 2 as the functions of the total number of indicators in the dataset).

*class\_weight* sets the weights associated with classes for the purpose of the impurity measure calculation. This parameter may be important when the sample is unbalanced which is our case.



*Threshold* is the hyperparameter which shows the probability of the pre-crisis event from the realisation of Random Forest function, after overcoming the current value we highlight the event as the pre-crisis event: 1–pre-crisis event, 0–tranquil event. Other hyperparameters which is not specified in the model are used as the default. Usefulness (see Section 3.2) is the metric of optimisation for choosing best model and consequently, the best hyperparameters.

### 3.2 The model evaluation

The predictive ability of EWSs is conventionally tested using receiver operating characteristic (ROC) analysis, but this approach has several shortcomings. The ROC summarises an EWS's performance over all regions of the ROC space including those in which one would rarely operate (see Baker and Pinsky 2001). For example, extreme right and left sides of the ROC space are generally useless, as they correspond to high false-positive and high false-negative rates, respectively. A related criticism is that the ROC approach weights all types of errors equally, while in applications to predicting financial crises the type I and type II errors may not have the same importance. When misclassification costs are unequal, summarising over all possible threshold values is flawed (Adams and Hand 1999).

A more comprehensive evaluation approach is the analysis based on the “usefulness” measure, which is calculated conditionally on the policymaker's relative aversion to missed crises as opposed to false alarms. We believe that it is important to test the indicator's performance under different preferences. Therefore (as in e.g. Alessi and Detken 2011) the model evaluation is based on the usefulness indicator:

$$usefulness = \min(\theta, 1 - \theta) - loss \quad (1)$$

$$loss = \frac{\theta * FN}{TP + FN} + \frac{(1 - \theta) FP}{TN + FP}, \quad (2)$$

where FN are false negative outcomes, FP are false positive outcomes, TN true negative outcomes, TP true positive signal comes.

The usefulness represents the gain using an EWS in comparison with the naïve approach in a decision making process. Note that a policy maker can always realise a loss of  $\min[\theta ; 1 - \theta]$  by ignoring the indicator (i.e. by issuing the signal either always or never). If  $\theta$  is smaller than 0.5, it follows that we never having issued signals and  $TP = FP$

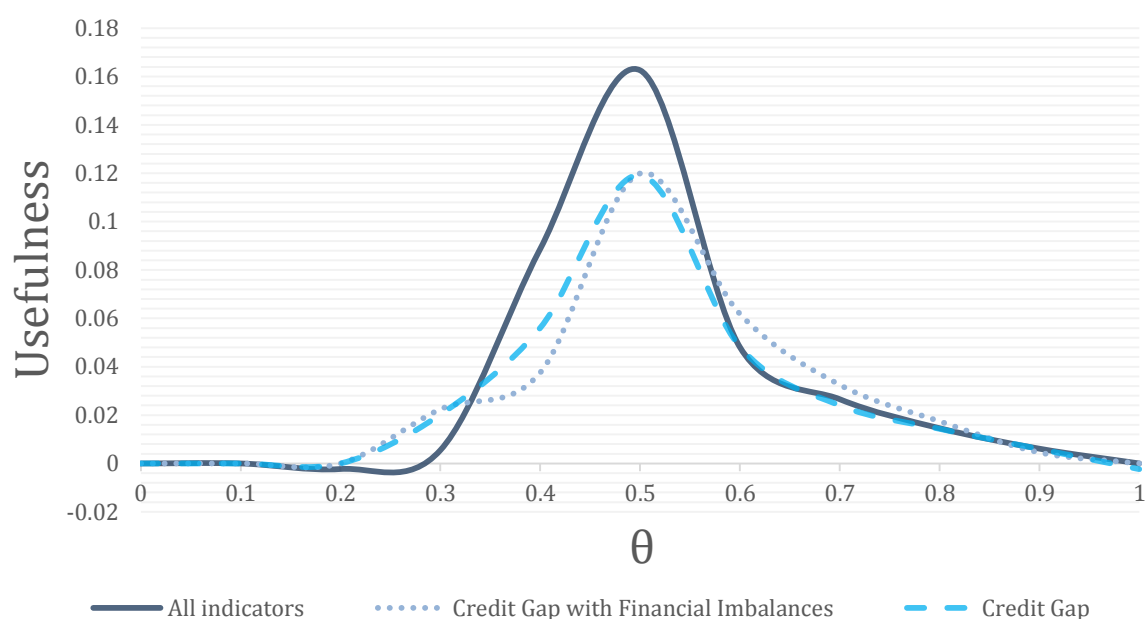
= 0: benchmark is obtain by ignoring the indicator. As a result, the value of loss is  $\theta$ . On the contrary, if  $\theta$  exceeds 0.5 it is assuming that the signal is always issued and  $FN = TN = 0$ . In this case the value of loss is  $1-\theta$ . Therefore, an indicator reduce a loss that the loss is becoming lower than  $\min[\theta ; 1 - \theta]$  for a given  $\theta$ —that is, relying on the indicator reduces the loss compared to a situation in which the indicator is ignored.

The parameter  $\theta$  represents the policy maker's preferences, larger  $\theta$  reflecting the preference to avoid missing a crisis rather than avoid issuing a false alarm. We test our EWSs for the full range of the values of  $\theta$ .

#### 4. The results

As noted in Beutel et al. (2019) the conventional two-step validation approach that implies splitting the observations into two parts for estimation and forecasting may not be suitable for the models with substantial specification flexibility. Notably, although the forecast (evaluation) sample is not used for model estimation it still may be used for model selection (i.e. if different values of hyperparameters' are searched through). Accordingly, the usefulness of the strategy on the whole may be overestimated. Therefore, a more demanding three-step validation exercise is conducted. Namely, we divide the dataset into three parts: the "train" sub-sample for model estimation, the "validation" sub-sample for model selection and the "test" subsample for model evaluation

Firstly, we report the results of a conventional cross validation procedure: the K-fold approach (see e.g. Murphy 2012). Namely, we split the countries in the cross-section into 5 folds. We estimate the models using all the cross-sections except for those in  $i^{th}$  and  $j^{th}$  folds. We validate the various models (i.e. models with the different values of hyperparameters) on the  $i^{th}$  fold. The best performing model is selected and tested on the  $j^{th}$  fold. We repeat the process using all 5 folds as test samples and compute the overall usefulness indicator. We conduct this exercise for various preferences  $\theta$ . The models are estimated using either the full set explanatory variables or the subset that does not include FDIs. We also measure the performance of a stand-alone credit gap indicator. The results for models are reported in Figure 2.

**Figure 2.** Usefulness obtained with different set of indicators (“K-fold” validation approach)

As expected the results show that the maximum value of the usefulness function is achieved when  $\theta = 0.5$ . The preference parameter of 0.5 represents a policymaker who is equally concerned about missing crises than issuing false alarms. The decrease in usefulness when  $\theta$  deviates from 0.5 represents the increased difficulty of outperforming the static strategy in cases when the preferences are clear. In other words, the competitiveness of always (or never) issuing the signal strategy increases in case a policymaker clearly wants to avoid missing the crisis (or issuing a false signal). Notably, the EWS that uses the full set of variables outperforms the competitors for most of  $\theta$  below 0.6.

We also conduct an alternative validation exercise that represents the historical case of forecasting the Great Financial Crisis. The “test” sub-sample includes 2004–2007 observations. We also add the emerging markets data into the “test” sub-sample (see Table 4 in the Appendix).<sup>3</sup>

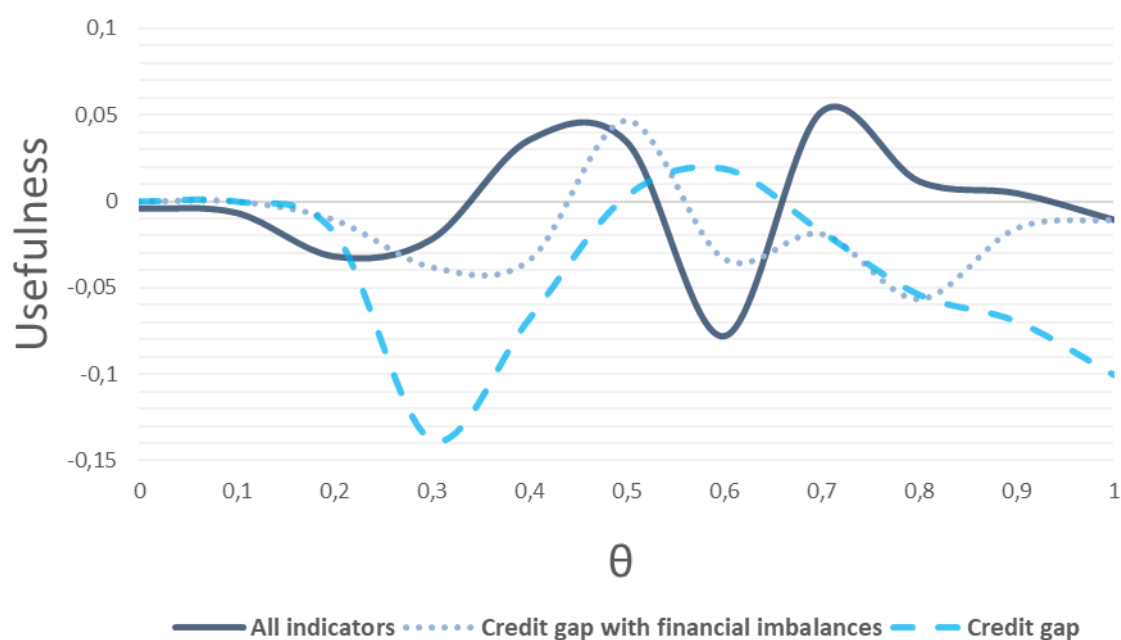
The rest of the sample is divided into five folds. We do this by creating five random samples with balanced number of crisis and non-crisis observations. Similarly to the “K-fold” approach we estimate the model using on the sub-sample consisting of four of these folds (the “train” sub-sample). The model is validated on the remaining fifth fold (the “validation sample”). This procedure is repeated five times using different folds as for

<sup>3</sup> In practice, due to limited data availability, applying the EWSs estimated for advanced economies to emerging markets is far from unprecedented (see e.g. Ponomarenko 2013). Excluding these observations from the dataset does not dramatically change the outcome.

validation, the results (i.e. FP, FN, TP and TN outcomes) are pooled and the usefulness measure is calculated. This procedure is repeated for different sets of hyperparameters and the best performing specification is selected for each  $\theta$ . Finally, the selected models are evaluated using the “test” sub-sample. The results are presented in Figure 3. The results for the test sample that include observations in 2004–2016 are reported in Figure 7 in the Appendix. The detailed results (i.e. case by case FP, FN, TP and TN outcomes) are reported in Figure 8 and Figure 9 in the Appendix.

The usefulness of all EWSs that was obtained in this experiment is significantly lower compared to the results of the “K-fold” exercise. In fact, the usefulness of the EWSs is negative in many cases. The deterioration of the EWSs’ performance may be attributed to a more demanding evaluation algorithm but also to the peculiarity of the test time sample. Nevertheless, the results of this exercise clearly indicate Random Forest approach outperforms the stand-alone credit gap indicator under the majority of policymaker’s preferences. Also, adding the FDIs to EWSs clearly improve their performance in forecasting the Great Financial Crisis at least in cases the policy makers had strong preferences against missing the crisis (i.e. large  $\theta$ ).

**Figure 3.** Usefulness obtained with different set of indicators (“historical” validation approach)



## 5. The role of financial development indicators

In this section we provide further insight on how the developments of FDIs, financial imbalances and the financial crisis risk interact. In order to illustrate relationship we conduct the following exercise. We estimate the EWS (for  $\theta=0.5$ ) using the full dataset using the best specification according to the “historical; validation approach. In Table 3 we report the Gini importance (see Menze et al. 2009 for details) of individual indicators in the EWS.

**Table 3.** Gini importance

Indicators	Gini importance
<i>Financial imbalances</i>	
Credit to GDP gap	18.58%
GDP one year growth rate	3.27%
GDP two year growth rate	4.44%
Credit to GDP one year growth rate	12.40%
Credit to GDP five growth rate	14.91%
Total credit one year growth rate	7.04%
Total credit two year growth rate	10.16%
<i>Financial development</i>	
Credit to GDP ratio	11.31%
Financial institution access	5.19%
Financial institution efficiency	9.40%
Financial market access	3.29%

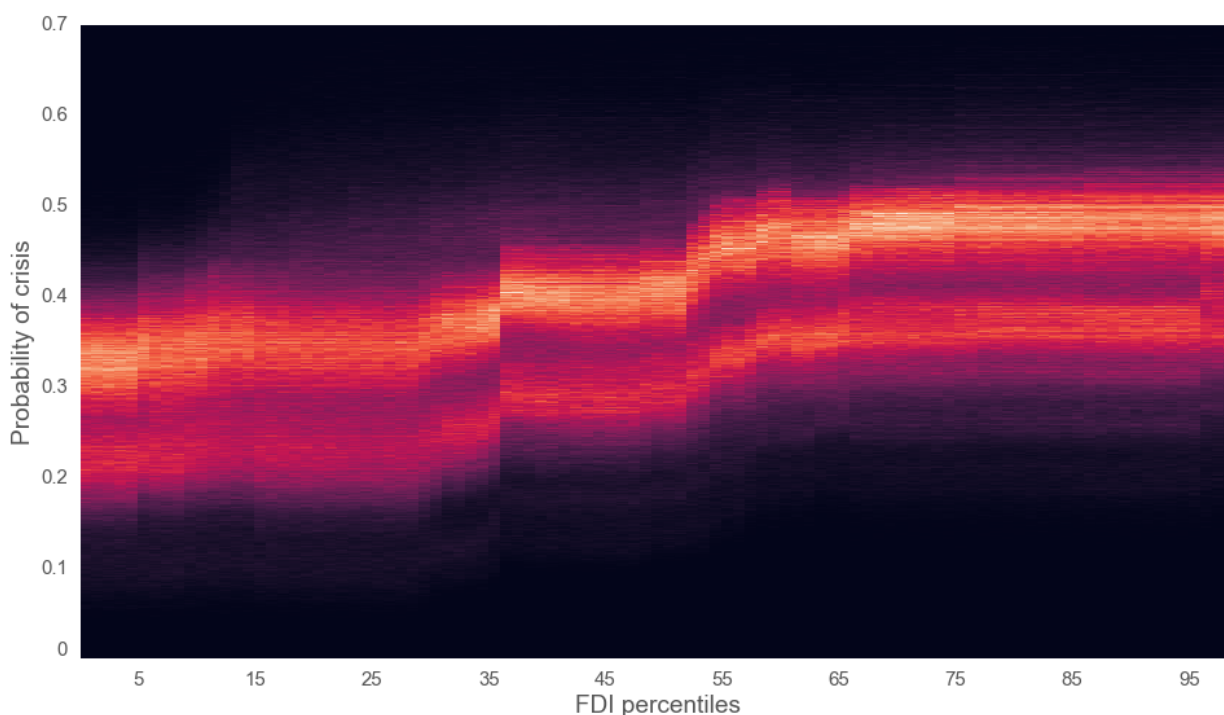
The Gini importance indicates that all of the indicators are used in the EWS. As expected, the credit to GDP ratio (in gaps, growth rates and levels) plays a prominent role, although other subsidiary indicators are also employed.

Illustrating the interplay between different values using a Random Forest model is notoriously difficult. Nevertheless, we proceed by characterising the interaction between credit gaps and credit-to-GDP with other FDI indicators via a simple exercise.

We calculate the crisis probability for 99 hypothetical datasets. The FDIs' values in each of these datasets are fixed at  $i = \{1 \dots 99\}$  percentile of the distribution<sup>4</sup> and the values of the financial imbalance are randomly drawn from the empirical distribution (100000 draws are made for each  $i$ ). The obtained distribution is presented on the heatmap (Figure 4) and represents the density of crisis probability conditional on the level of financial development.

The results indicate that the probability of a crisis is generally higher at later stages of financial development. Interestingly, the density of financial crisis probability becomes less dispersed and its two peaks<sup>5</sup> are more pronounced at higher FDI values (kernel density of crisis probability for FDIs at 25<sup>th</sup> and 75<sup>th</sup> percentile levels are presented in Figure 5). Arguably, this observation shows that the interaction between financial imbalances and FDIs is actually captured by the model (i.e. not only the higher levels of financial development per se increase the crisis probability but rather when FDIs are high the financial imbalances indicators may be used to estimate the crisis probability more precisely).

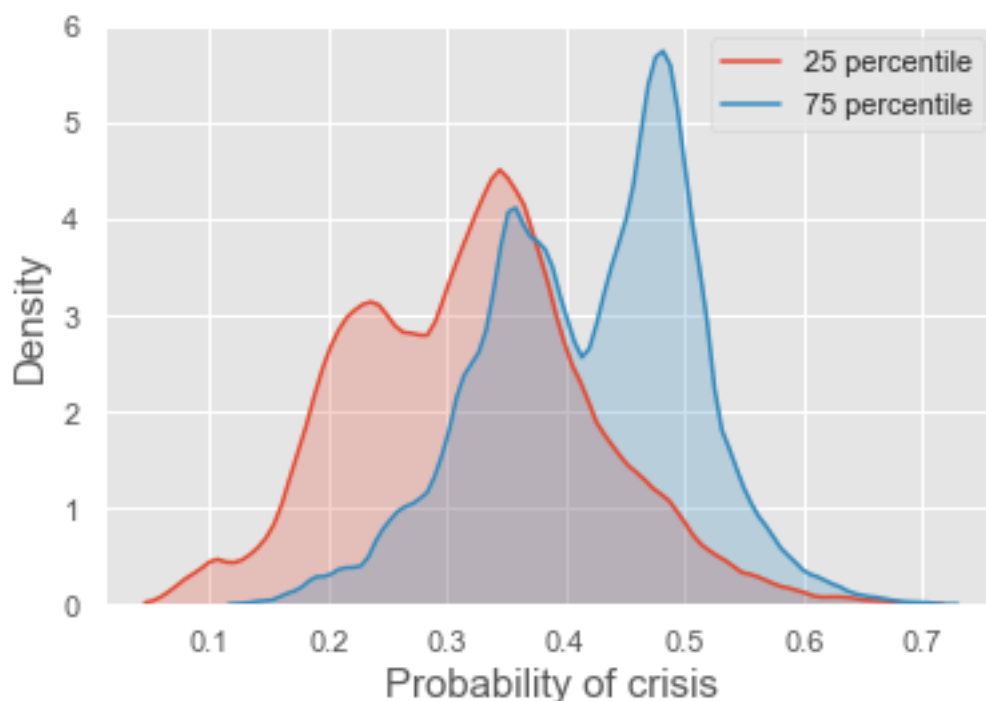
**Figure 4.** Probability of financial crises at different levels of financial development



<sup>4</sup> Note that in all cases, larger values of the FDIs correspond to higher financial development.

<sup>5</sup> Presumably these two peaks represent a threshold in an important financial imbalance indicator

**Figure 5.** Kernell density (%) of financial crisis probability at different levels of financial development (at 25<sup>th</sup> and 75<sup>th</sup> percentiles of FDI)



## 6. Conclusions

We set up an early warning system for financial crises based on the Random Forrest approach. We use a novel set of predictors that comprises financial development indicators (e.g. the levels of credit to GDP ratio) in addition to conventional imbalances measures (e.g. the credit gaps). The evaluation of the model is conducted using a three-step procedure (i.e. over train, validation and testing sub-samples). The results indicate that combining financial imbalances and financial development indicators helps to improve the out-of-sample accuracy of the early warning system. The probability of a crisis is generally higher at later stages of financial development. The interaction between financial imbalances and FDIs is also captured by the model: as FDIs increase the crisis probability, estimates become more sensitive more to the fluctuations of financial imbalances indicators.

## References

- 1) Adams, N.M., Hand, D.J. (1999). "Comparing classifiers when the misallocation costs are uncertain." *Pattern Recognition*, 32: 1139–1147.
- 2) Alessi, L., Detken, C. (2011). "Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity." *European Journal of Political Economy*, 27: 520–533.
- 3) Alessi, L., Detken, C. (2018). "Identifying excessive credit growth and leverage." *Journal of Financial Stability*, 35: 215–225.
- 4) Arcand, J.L., Berkes, E., Panizza, H. (2015). "Too much finance?" *Journal of Economic Growth*, 20 (2):105–148.
- 5) Babecky, J., Havranek, T., Mateju, J., Rusnak, M., Smidkova, K., Vasicek, B. (2014). "Banking, bedt and currency crisis: stylized facts and early warning indicators." *Journal of Financial Stability*, 15: 1–17.
- 6) Bahadir, B., Valev, N. (2015). "Financial development convergence." *Journal of Banking and Finance*, 56: 61–71.
- 7) Baker, S., Pinsky, P. (2001). "A proposed design and analysis for comparing digital and analog mammography: special receiver operating characteristic methods for cancer screening." *Journal of the American Statistical Association*, 96: 421–428.
- 8) Beutel, J., List, S., Schweinitz, G. (2019). "Does machine learning help us predict banking crises?" *Journal of Financial Stability*, 45, 100693.
- 9) Borio, C., Lowe, P. (2002). "Asset Prices, Financial and Monetary Stability: Exploring the Nexus." BIS Working Paper No. 114.
- 10) Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1): 5–32.
- 11) Buncic, D., Melecky, M. (2014). "Equilibrium credit: The reference point for macroprudential supervisors." *Journal of Banking & Finance*, 41: 135–154.
- 12) Cecchetti S.G., Kharroubi E. (2012). "Reassessing the impact of finance on growth." BIS Working Papers No. 381.
- 13) Cottarelli, C., Dell’Ariccia, G., Vladkova-Hollar, I. (2005). "Early birds, late risers, and sleeping beauties: Bank credit growth to the private sector in Central and Eastern Europe and in the Balkans." *Journal of Banking and Finance*, 29 (1): 83–104.
- 14) Detken, C., Weeken, O., Alessi, L., Bonfm, D., Boucinha, M., Castro, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N., Kakes, J., Klaus, B., Lang, J., Puzanova, N., Welz, P. (2014). "Operationalising the Countercyclical Capital Buffer: Indicator



- Selection, Threshold Identification and Calibration Options.” ERSB Occasional Paper Series No. 5.
- 15) Drehmann, M., Borio, C., Tsatsaronis, K. (2011). “Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates.” BIS Working Papers No. 355.
  - 16) Égert, B., Backé, P., Zumer, T. (2006). “Credit growth in Central and Eastern Europe—new (over)shooting stars?” ECB Working Paper No. 687, European Central Bank.
  - 17) Geršl, A. and Jašová, M. (2018). “Credit-based Early Warning Indicators of Banking Crises in Emerging Markets.” *Economic Systems*, 42: 18–31.
  - 18) Giese, J., Andersen, H., Busha, J., Castro, C., Farag, M., and Kapadia, S. (2014). “The credit-to-gdp gap and complementary indicators for macroprudential policy: evidence from the UK.” *International Journal of Finance and Economics*, 19: 25–47.
  - 19) Holopainen, M., Sarlin, P. (2017) “Toward robust early-warning models: a horse race, ensembles and model uncertainty.” *Quantitative Finance*, 17(12): 1933–1963.
  - 20) Jermann, U., Quadrini, V. (2012) “Macroeconomic Effects of Financial Shocks.” *American Economic Review* 102(1), 238-71.
  - 21) Jordà, Ò., Schularick, M., Taylor, A.M. (2013) “When Credit Bites Back.” *Journal of Money, Credit, and Banking*, 45(2): 3–28
  - 22) Laeven, L., Valencia, F. (2018). “Systemic Banking Crisis Revised.” IMF Working Paper, WP/18/206.
  - 23) Menze, B., Kelm, M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F. (2009). “A comparison of Random Forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data.” *BMC Bioinformatics* 10:213.
  - 24) Mumtaz, H., Zanetti, F. (2016). “The Effect Of Labor And Financial Frictions On Aggregate Fluctuations,” *Macroeconomic Dynamics*, 20(1), 313-341.
  - 25) Murphy, K.P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, Massachusetts.
  - 26) Naceur, S.B., Candelon, B., Lajaunie, Q. (2019). “Taming financial development to reduce crises.” *Emerging Markets Review*, 40, 100618.
  - 27) Ponomarenko, A. (2013). “Early warning indicators of asset price boom/bust cycles in emerging markets.” *Emerging Markets Review*, 15(C): 92–106.

- 
- 28)Schularick, M., Taylor, A.M. (2012). "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review*, 102 (2): 1029–61.
- 29)Sviryndzenka, K. (2016) "Introducing a New Broad-based Index of Financial Development." IMF Working Paper 16–5.
- 30)Tölö, E., Laakkonen, H., Kalatie, S. (2018). "Evaluating Indicators for Use in Setting the Countercyclical Capital Buffer." *International Journal of Central Banking*, 14(2): 51–112.
- 31)Yang, L., Su, J.-J. (2018). "Debt and growth: Is there a constant tipping point?" *Journal of International Money and Finance*, 87: 133–143.
- 32)Zanetti, F. (2019). "Financial Shocks, Job Destruction Shocks, And Labor Market Fluctuations." *Macroeconomic Dynamics*, 23(3), 1137-1165.

## Appendix

**Table 4.** Data sources

Series	Period	Transformation	Source
Crisis events	1970 Q1:2010 Q4	-	Babecky et al. (2014)
Crisis events	2010 Q1:2016 Q4	-	Laeven and Valencia (2018)
Real GDP	1970 Q1:2016 Q4	-	OECD data
Credit-to-GDP gap	1970 Q1:2016 Q1	-	BIS Statistics
Total credit, local currency	1970 Q1:2016 Q1	Log, deflated with the GDP deflator	BIS Statistics
Credit-to-GDP, per cent	1970 Q1:2016 Q1	-	BIS Statistics
Financial Institutions Access	1980:2016	-	IMF data
Financial Institutions Efficiency	1980:2016	-	IMF data
Financial Markets Access	1980:2016	-	IMF data

The definitions of the financial developments indicators are obtained from the IMF website:

*Financial Institutions Access index (FIA)* – this indicator consist of compilation data on bank branches per 100 000 adults and ATMs per 100 000 adults;

*Financial Institutions Efficiency index (FIE)*–this indicator consist of compilation data on banking sector net interest margin, lending-deposits spread, non-interest income to total income, overhead costs to total assets, return on assets and return on equity;

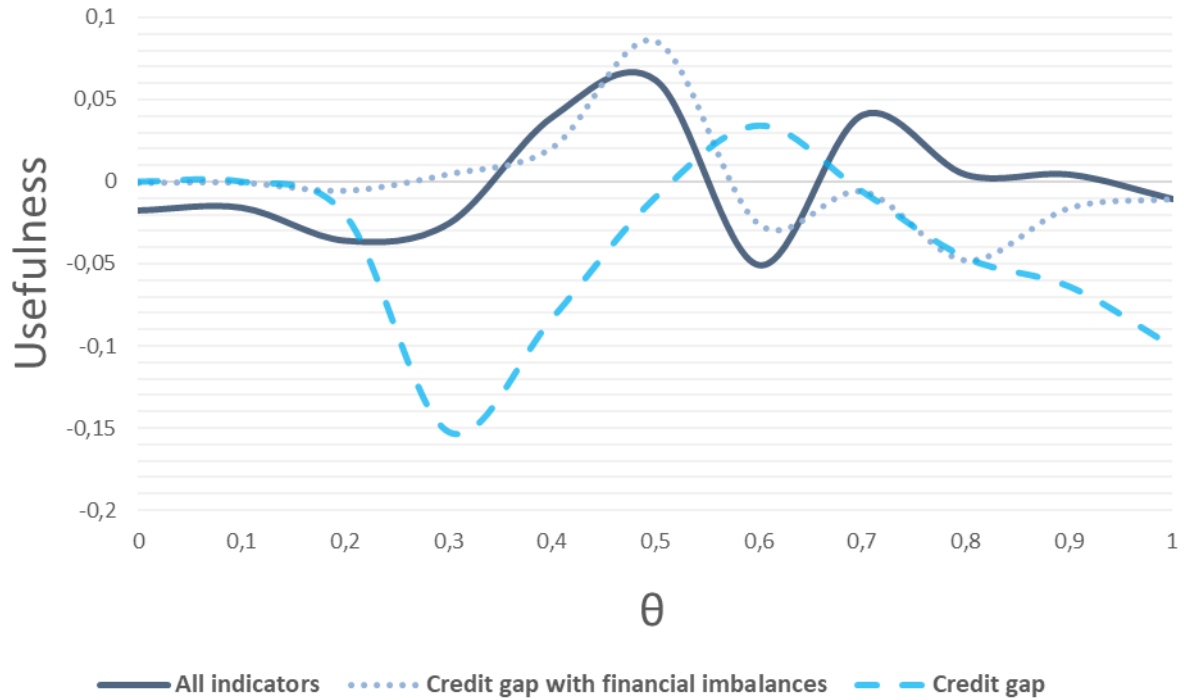
*Financial Markets Access index (FMA)*–this indicator consist of compilation data on per cent of market capitalisation outside of top 10 largest companies and total number of issuers of debt (domestic and internal, nonfinancial and financial corporations) per 100 000 adults.

**Table 5.** Cross-section of countries

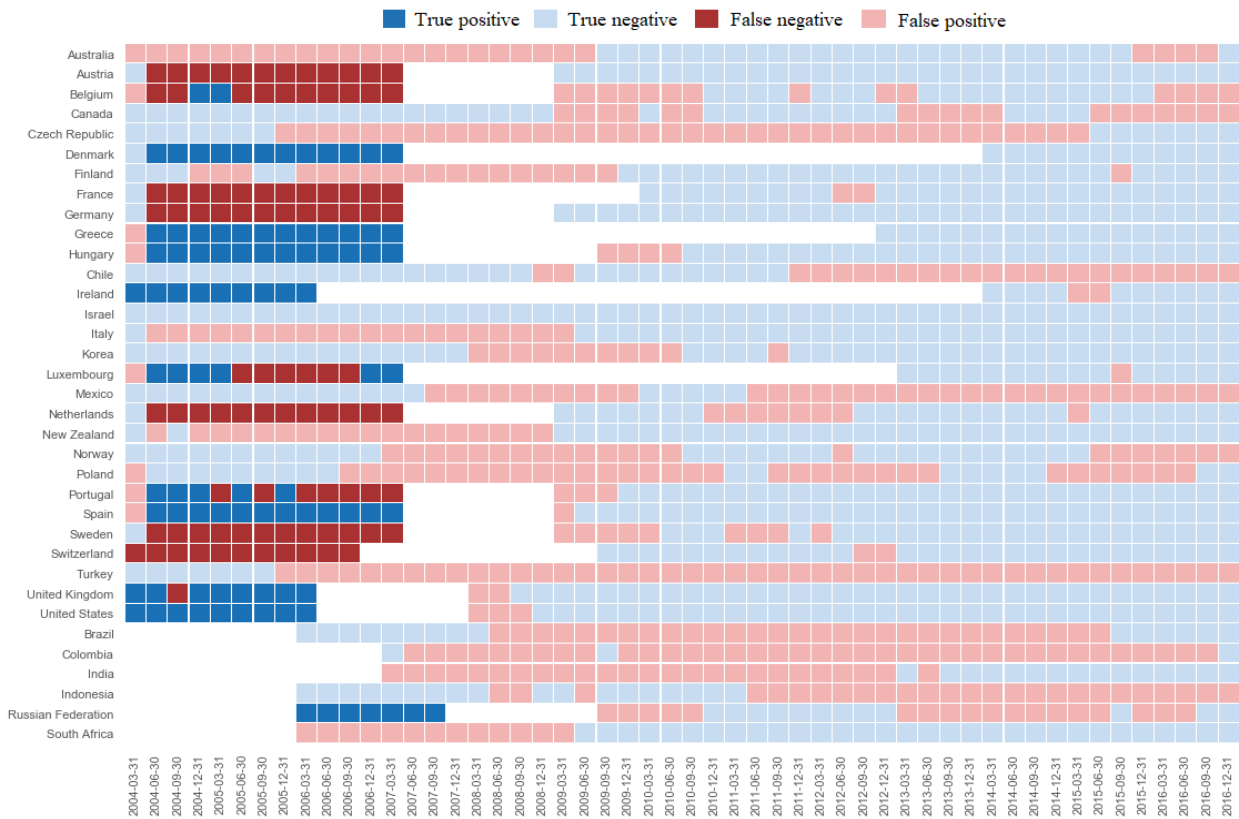
Countries	Observations	Countries	Observations	Countries	Observations
Australia	1971–2016	Ireland	1991–2016	Spain	1971–2016
Austria	1971–2016	Israel	1996–2016	Sweden	1971–2016
Belgium	1971–2016	Italy	1971–2016	Switzerland	1971–2016
Canada	1971–2016	Japan	1971–2016	Turkey	1999–2016
Chile	1987–2016	South Korea	1971–2016	United Kingdom	1971–2016
Czech Republic	1993–2016	Luxemburg	2000–2016	United States	1971–2016
Denmark	1995–2016	Mexico	1981–2016	<b>Brazil</b>	2006–2016
Finland	1971–2016	Netherlands	1971–2016	<b>Colombia</b>	2006–2016
France	1971–2016	New Zealand	1971–2016	<b>India</b>	2006–2016
Germany	1992–2016	Norway	1971–2016	<b>Indonesia</b>	2006–2016
Greece	1996–2016	Poland	1996–2016	<b>Russian Federation</b>	2006–2016
Hungary	1992–2016	Portugal	1971–2016	<b>South Africa</b>	2006–2016

Cross-sections in **bold** are only used in the “historical” test sub-sample as described in Section 4. The main dataset only includes the observations prior to 2008.

**Figure 6.** Usefulness obtained with different set of indicators (“historical” validation approach, the test sample is 2004–2016)



**Figure 7.** Detailed results of forecasting financial crises in 2004–2016.



**Figure 8.** Detailed results of forecasting financial crises in 2004–2016 (country-specific outcome shares).

