A machine learning approach to rank the determinants of banking crises over time and across countries

A machine learning approach to rank the determinants of banking crises over time and across countries

Elizabeth Jane Casabianca[a,b,1], Michele Catalano[a,c], Lorenzo Forni[a,d,*], Elena Giarda[a,e], Simone Passeri[a,2]

[a] Prometeia, Piazza Trento e Trieste 3, 40137 Bologna, Italy

[b] Polytechnic University of Marche, Department of Economics and Social Sciences (DISES), Piazzale Martelli, Via Villarey 4, 60121 Ancona, Italy

[c] IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

[d] University of Padua, Department of Economics and Management (DSEA), Via del Santo 33, 35123 Padova, Italy

[e] University of Modena and Reggio Emilia, Centre for Studies in Banking and Finance (CEFIN), Via Jacopo Berengario 51, 41121 Modena, Italy

Abstract

We use a machine learning approach, namely AdaBoost, to rank the determinants of banking crises over time and across countries. We cover a total of 100 countries, advanced and emerging, over the years from 1970 to 2017. The paper first shows that AdaBoost has a better predictive performance than the logit model, both in-sample and out-of-sample; then, it employs AdaBoost to classify the major macroeconomic factors leading to banking crises. The baseline analysis reveals that the US 10yr Treasury interest rate and world growth play a key role in anticipating a crisis, and that these two variables explain a growing share of the results over time, for both country groups. Other variables, which have been highlighted as important in the literature on crises - such as inflation, current account, public and external debt and credit - are relevant in the lead up to banking crises, but their role has been decreasing over time compared to the aforementioned variables. We present also extensions of the model, which confirm and add to the main results of the baseline model.

Keywords: banking crises; predictive models; machine learning.

JEL classification: C53; G01; C25; E44.

Corresponding author. E-mail address: lorenzo.forni@unipd.it (L. Forni).

---

[1] Present address: Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra (VA), Italy.
[2] Present address: Sogei, Via Mario Carucci 99, 00143 Rome, Italy.

1. Introduction

The 2007-2008 financial crisis that hit advanced economies triggered a worldwide economic downturn with severe and widespread losses across the real and financial sectors. It unfolded as a systemic banking crisis and reinforced the attention of national and supranational institutions on the links between money and credit fluctuations and the insurgence of a crisis, with an eye towards mitigating the propagation of similar crises.

A better understanding of countries' financial vulnerabilities is crucial to contain the contagion effects in case a new crisis should occur. In particular, recognising the economic factors that carry valuable information to identify vulnerabilities is key to countries' resilience to economic shocks. The ultimate goal is to inform macroprudential policies addressing such vulnerabilities and limit them from building up further and spreading across the economic system.

In this paper we study the macroeconomic determinants of banking crises for advanced and emerging economies by means of a boosting algorithm (AdaBoost), belonging to the field of machine learning (ML) models. We select a set of leading indicators of banking crises suggested by the literature and implement a framework that allows us to assess their contribution in anticipating a crisis. First, we validate the AdaBoost in-sample. Second, we assess its performance against a logit model both in-sample and out-of-sample. Third, based on the outcomes of the AdaBoost, we expand on the existing literature by providing an evaluation of the contribution of each macroeconomic factor to anticipate a crisis over time and across country groups by means of two measures, specifically "relative importance" and Shapley values.

This paper contributes to the existing literature by applying ML algorithms to the study of banking crises on a sample of 100 countries over a period of almost 50 years. There is a growing literature which applies ML algorithms to study financial crises, but either the set of countries analysed or the sample period are substantially smaller. The sample size allows us to have a larger number of crisis episodes to work with. Therefore, we can compare results for different groups of countries and over time. In particular, we provide evidence on the evolution of the main determinants of financial crises over a period when the global financial system has undergone substantial changes.

The literature aimed at detecting the risks that a systemic banking crisis may arise has evolved following various approaches, from the signals approach to discrete choice models and machine learning techniques. Kaminsky and Reinhart (1999) and Kaminsky (1999) represent two of the early contributions using the signals approach. Further work along this line is provided by Borio and Lowe (2002) and Davis and Karim (2008), among others. Demirgüc-Kunt and Detragiache (1998) make use of logit models, followed by contributions analysing different subsets of countries and periods (e.g. Arteta and Eichengreen, 2000; Demirgüc-Kunt and Detragiache, 2005; Barrell et al., 2010; Schularick and Taylor, 2012). More recently, machine learning methods have been employed by economists to improve the performance of predictive models. Examples in this direction include Duttagupta and Cashin (2011), Holopainen and Sarlin (2017), Alessi and Detken (2018), Beutel et al. (2019), Bluwstein et al. (2020).

Our work is also related to the literature on credit booms and financial crises. Credit booms during expansions can be particularly damaging if the ensuing recession is characterized by the presence of a financial sector crisis. Indeed, research shows that a rapid accumulation of private debt is a leading indicator of banking crises, which usually are followed by extremely acute downturns (Reinhart and Rogoff, 2009; Jordà et al., 2011; Gourinchas and Obstfeld, 2012). In addition, in the presence of financial sector disruptions, the empirical association between debt service and borrowing constraints (Johnson and Li, 2010) is likely to be amplified. Furthermore, the recent history of banking crises shows that financial sector intervention policies, such as bank recapitalizations and asset purchases, often involve significant fiscal resources, therefore requiring countries to have fiscal space to address them and in turn contributing to the deterioration of the public balance sheet (Laeven and Valencia, 2013; Bernardini and Forni, 2020).

For the purpose of our work, we gather information on banking crisis episodes from various sources to maximise coverage across both time and countries (Laeven and Valencia, 2008, 2013 and 2018; Reinhart and Rogoff, 2009; Jordà et al., 2017). Our dataset includes 100 countries, 33 advanced and 67 emerging ones, and covers the period 1970-2017. The total number of banking crises amounts to 142, 55 in advanced economies and 87 in emerging ones. Since we are interested in the factors leading to a crisis, we look at pre-crisis years, as in Arteta and Eichengreen (2000), Alessi and Detken (2018), Beutel et al. (2019) and Filippopoulou et al. (2020), among others. As for the macroeconomic factors that may lead to a crisis, in our baseline model we focus on the most important variables identified in the literature, such as current account balance to GDP, public debt to GDP, credit to GDP, external debt to GNI, and inflation. We also include the US 10yr Treasury rate and world GDP growth in real terms.

We start the analysis by splitting the samples of advanced and emerging countries into a training set and a testing set, on which we validate the models and test them, respectively. The choice of separating parameter selection (on the training set) from model assessment (on the testing set) is motivated by the fact that this should limit the risk of overfitting, especially common in empirical applications of machine learning algorithms.

First, we validate both models in-sample. Once identified, the optimised parameters are employed in the out-of-sample exercise. Second, we employ a recursive procedure to evaluate the performance of our models out-of-sample. Specifically, for each year of our testing set, we retrieve pre-crisis predicted probabilities, confusion matrices and performance indicators, including AUROC, sensitivity, specificity and relative usefulness. Following Beutel et al. (2019), we also implement a bootstrap procedure and obtain interquartile ranges for each of these indicators. Finally, we evaluate the "relative importance" (Carmona et al., 2019, and Antulov-Fantulin et al., 2021) and Shapley values (Bluwstein et al., 2020) of the macroeconomic factors that may anticipate a crisis based on the outcomes of the AdaBoost, since the model is shown to deliver a better performance than the logit both in-sample and out-of-sample.

Our baseline results show that, among the macroeconomic variables used, the strongest contribution in anticipating a crisis is given by the US 10yr Treasury rate, for both advanced and emerging economies. Our results reveal that the importance of the US 10yr Treasury rate grows over time, which is consistent with the rising role of gross financial flows and capital mobility across countries as a determinant of financial stability. The second most relevant variable is world growth for both groups of countries. Country-specific variables that have been highlighted as important in the literature on crises, including inflation, current account, public and external debt, and credit are relevant in the lead up to banking crises, but their role has been decreasing over time compared to the two aforementioned variables. Finally, when we extend the model to include additional variables, we find that in advanced countries the real effective exchange rate has a relevant role as leading indicator of banking crises. In the case of emerging ones, the real effective exchange rate cannot be included due to lack of sufficient data. In this case, the key role of the US 10yr Treasury rate is confirmed.

The remainder of the paper is organised as follows. Section 2 provides a review of the literature on systemic banking crises. Section 3 provides descriptive statistics on banking crises and the macroeconomic context. In Section 4, we build our models for advanced and emerging countries by means of a logit model and AdaBoost; we then validate them in-sample and assess their out-of-sample predictive performance. In Section 5, based on the AdaBoost results, we perform a ranking exercise of the determinants of banking crises over time. Section 6 provides a robustness analysis. Finally, Section 7 concludes with a summary of the main findings.

2. Literature review

The literature on how to identify, explain and predict crises has a long-lasting tradition. Work on the topic covers both the definition of banking crises and the construction of predictive models to help governments and international financial institutions act promptly to prevent risks of possible future bank runs and bank failures from turning into systemic banking crises.

As for the definition of systemic banking crises, Baron et al. (2019 and 2021) suggest the distinction between the policy-based approach and the narrative-based one. With regards to the former, Laeven and Valencia (2008, 2013 and 2018) − improving on Demirgüc-Kunt and Detragiache (1998) − identify a banking crisis by looking at factors such as a high level of nonperforming loans to total loans or relevant fiscal direct interventions and restructuring costs. However, banks’ losses can be mitigated by policy interventions and hence may be difficult to quantify. Therefore, the authors look at whether a certain number of measures were implemented (see **Table A.1** in Appendix A for details).[1] Instead, the narrative approach looks at the narrative sources of events, such as bank runs or policy intervention. This allows the inclusion of a number of crises backed by a strong historical narrative, which are “forgotten” in the policy-based framework. Examples of the narrative approach are Bordo et al. (2001), Reinhart and Rogoff (2009 and 2011), Schularick and Taylor (2012) and Jordà et al. (2017). Beside these two approaches, Schularick and Taylor (2012) view financial crises as credit booms gone bust. Their dataset is the result of a critical scrutiny and merge of previously compiled datasets (e.g. Bordo et al., 2001; Laeven and Valencia, 2008; Reinhart and Rogoff, 2011), partially updated by Jordà et al. (2017). Finally, Baron et al. (2019 and 2021) adopt an alternative method in which they develop a crisis indicator based on the decline in each country’s bank equity index to refine the chronology of banking crises.

In all the approaches used to define a banking crisis, the resulting variable is discrete, binary or multinomial, depending on the classification of crises, non-crisis spells and normal times. The variable is generally employed as the outcome variable in most of the predictive models in the literature. The earliest methodology to tackle the prediction of crises is the signals

[1] See also Caprio and Klingebiel (1997 and 2002) and Demirgüc-Kunt and Detragiache (1998).

approach, a non-parametric method that studies the ex-post behaviour of macroeconomic variables and verifies whether the indicators follow a pattern in the pre-crisis periods that differs from that in tranquil or normal times. In its initial version, a variable is considered to signal a crisis if it exceeds a pre-defined threshold (Kaminsky and Reinhart, 1999). The method evolved to look at composite indicators (Borio and Lowe, 2002; Davis and Karim, 2008; Alessi and Detken, 2011).

Another widely accepted line of research employs discrete outcome models (i.e. logit or multinomial), which, despite being reduced-form models, allow an economic interpretation of the links between the outcome variable and the explanatory variables. In the simplest case, it is a binary variable (0/1), where the 1s define systemic banking crisis episodes (Demirgüc-Kunt and Detragiache, 1998) and the 0s all the other periods. Arteta and Eichengreen (2000), Demirgüc-Kunt and Detragiache (2005), Barrell et al. (2010) and Schularick and Taylor (2012) identify a binary variable as well, but they differ in how they classify the non-crises years. Other authors, such as Alessi and Detken (2018) label the pre-crisis years 1s. Filippopoulou et al. (2020) build on Demirgüc-Kunt and Detragiache (1998, 2000 and 2005) and estimate a binary logit model for systemic banking crises with the aim of evaluating the predictive validity of the risk indicators included in the macroprudential database of the European Central Bank. Fielding and Rewilak (2015) estimate a dynamic probit model with the aim of quantifying the persistence of the crisis. Finally, in works such as those of Hardy and Pazarbasioglu (1999) and Caggiano et al. (2014), the outcome variable takes on more than two values (i.e. 0/1/2) since it distinguishes among pre-crisis, crisis and post-crisis spells.

In recent years, economists have been increasingly employing "supervised" ML methods where the main objective is to perform predictions (for a review, see Athey, 2019). ML is a data mining tool able to analyse complex datasets, fit multifaceted and flexible functional forms to the data and find functions that perform well out-of-sample (Mullainathan and Spiess, 2017). Most of the empirical literature employing ML methods usually provides a comparison with the outcomes of traditional discrete choice models. However, evidence of which of the approaches perform better and provide the best forecasting ability is mixed. Manasse and Roubini (2009) employ a Classification and Regression Tree (CART) to study sovereign debt crises in 47 emerging economies for the period 1970-2002. They implement a logit model to assess whether the non-linear interactions and threshold effects found in the tree are statistically significant. In Duttagupta and Cashin (2011) banking crises are analysed by means of a Binary Classification Tree (BCT) for 50 developing and emerging countries, with data from 1990 to 2005. They conclude that the BCT is superior to the logit model since it is able to identify non-linear relationships underlying banking crises. Alessi and Detken (2018) implement the Random Forest (RF) algorithm and apply it to banking crises in the European Union (EU) and the UK, by using quarterly data from 1970 to 2013. They find evidence of a better performance of the RF compared to the logit, although it may depend on the sample of countries covered. Beutel et al. (2019) perform a horserace among models by comparing a benchmark logit approach to several ML algorithms, including a BCT and a RF, covering European countries, Japan and the US and using quarterly data from 1970 to 2016. They find that the former outperforms the latter suggesting that further enhancements to ML models are needed before they can bring value added for predicting banking crises. Bluwstein et al. (2020) also develop a range of models to predict financial crises by applying various ML algorithms on 17 countries over the period 1970-2016. Their results show that ML algorithms perform better than the logistic regression. Also, Holopainen and Sarlin (2017), working on quarterly data on 15 European countries over 1976Q1-2014Q3, find that the conventional statistical approaches are outperformed by machine learning methods, particularly by model aggregation approaches. Moreover, Fouliard et al. (2020) employ a general framework of sequential predictions called online machine learning to forecast crises out-of-sample for France, UK, Germany and Italy. Finally, also Suss and Treitel (2019) find that the RF outperforms the logit and other ML algorithms in predicting financial distress in the UK using firms' data from supervisory assessments of bank riskiness from 2006 to 2012. More recently, Hellwig (2021) uses ML algorithms to assess fiscal crises and show that they deliver significant improvements in accuracy compared to econometric approaches widely used in the literature.

3. Dataset and descriptive statistics

3.1 Definition of banking crises and of the target variable

To build our predictive models we employ the banking crisis dataset of Laeven and Valencia (2008, 2013 and 2018) supplemented with the ones of Reinhart and Rogoff (2009) and of Jordà et al. (2017). We obtain a total of 142 crisis episodes, 55 in the advanced countries and 87 in the emerging ones, between 1970 and 2017, for a total of 1584 and 3216 annual observations for advanced and emerging countries respectively (see **Table A.2** in Appendix A for the full list of the countries covered and **Table A.3** for a full list of crisis episodes).

The literature adopts two different approaches to build the target variable. One that aims at predicting the occurrence of a crisis (e.g. Demirgüc-Kunt and Detragiache, 1998 and 2005;

Schularick and Taylor, 2012; Richter et al., 2021) and in which the target variable is a dummy equal to 1 in crisis years. The other that aims at signalling the building up of macroeconomic imbalances that may lead to a crisis (e.g. Arteta and Eichengreen, 2000; Alessi and Detken, 2018; Beutel et al., 2019; Filippopoulou et al., 2020); in this second case, the target variable identifies the pre-crisis periods. In this approach, crisis years are commonly dropped from the dataset (Beutel et al., 2019).

Since our interest lies in building a model that may help anticipate the occurrence of a crisis, we follow the second approach. To this end we need to define the pre-crisis years. Following Arteta and Eichengreen (2000), the "pre-crisis" spells are the three years preceding each banking crisis. The "post-crisis" ones are the three years following each banking crisis and the time spells that are at least three years past a crisis and at least three years prior to a crisis are defined "normal times". Our dataset is therefore partitioned into four groups: crisis, pre-crisis, post-crisis and normality. When, due to the frequent occurrence of a crisis, a post-crisis period overlaps with the pre-crisis period of a subsequent crisis, we give priority to the post-crisis episode.[2] Therefore, it may happen that we do not observe a pre-crisis spell before a crisis (for instance, for the USA between 1984 and 1988), or that a period of pre-crisis is shorter than the predefined three-year spell (i.e. Japan between 1992 and 1997).[3]

Figure 1 and **Figure 2** display the occurrence of crises, pre- and post-crisis episodes and normal times, for a selection of advanced and emerging economies, respectively. Among the advanced economies, all countries recorded at least a crisis episode, with the exception of Hong Kong. The UK is the one with the highest number of crises (five), followed by the USA, Iceland and Korea with three. Nineteen crises (35% of the total) occurred between 2007 and 2008. Among the emerging economies, sixteen countries did not experience any crisis based on the definition adopted in this paper.[4] The distribution of the crises is much more disperse over time than among developed countries and shows the highest concentration of episodes in

[2] Alessi and Detken (2011) address this issue by classifying two asset price boom episodes that closely follow each other as one boom.

[3] We decide to follow this approach as it allows us to focus on original crisis episodes rather than on the crises that closely follow previous ones.

[4] Barbados, Belize, Suriname, Trinidad and Tobago, Syria, Brunei, Pakistan, Botswana, Gabon, Libya, Mauritius, Seychelles, Namibia, Fiji, Turkmenistan and finally Serbia and Montenegro.

the 1990s.[5] Only four emerging economies (Kazakhstan, Russia, Ukraine and Hungary) experienced a banking crisis in 2008.

FIGURES 1 AND 2 ABOUT HERE

Targeting either the crisis or the pre-crisis years entails three definitions of the outcome variable, which is binary in all three cases (**Table 1**). In definition 1, the value 1 identifies the crisis, while the value 0 all other periods. In definition 2, the target variable equal to 1 identifies the pre-crisis spells. However, we have two options in defining the 0s. We either include (definition 2a) or exclude (definition 2b) the post-crisis episodes. In line with Demirgük-Kunt and Detragiache (2005) and Holopainen and Sarlin (2017) and to avoid the post-crisis bias (Bussière and Fratzscher, 2006), we adopt definition (2b) and drop the post-crisis periods from the definition of the 0s. The sample sizes based on these different definitions of the target variable are displayed in **Table 2**.

TABLES 1 AND 2 ABOUT HERE

3.2 Leading indicators of banking crises

The literature has identified a set of macroeconomic variables that may act as factors leading to a banking crisis. We distinguish between country-specific and global variables. The country-specific variables that we consider in our baseline model are: current account balance to GDP, external debt to GNI, public debt to GDP, credit to GDP and inflation. The global variables are: US 10yr Treasury rate and world GDP growth in real terms. **Table A.4** in Appendix A provides their exact definition and source.[6]

Richter et al. (2021) stress how a larger current account deficit indicates increased financial flows from abroad, which might increase financial fragility because of possible capital flow reversals. External and public debt are used as a proxy for countries' vulnerability to solvency and liquidity shocks (Manasse and Roubini, 2009). Countries with lower levels of public debt are expected to be less fragile and thus better able to counteract the emergence of a banking crisis by activating fiscal space. Moreover, higher levels of external debt may

[5] For this reason, banking crises for emerging countries are concentrated in the training sample.

[6] In Section 6 and Appendix B, we extend this baseline model to include country-specific GDP growth rates and energy prices expressed in percentage changes - for both sets of countries. Due to data limitations, we were able to add house prices and real effective exchange rates, both expressed as percentage changes, for advanced economies only.

indicate a country's greater reliance on foreign investors making it more vulnerable to external shocks. Credit to GDP is a leading indicator of banking crises based on the literature on credit booms. According to this line of research, excessive credit growth is a sign of an overheated economy that, if hit by an adverse shock, could trigger a banking crisis (Schularick and Taylor, 2012; Richter et al., 2021). As suggested by Demirgüc-Kunt and Detragiache (1998), inflation may provide indications of macroeconomic mismanagement, which adversely affects the economy.

Our second set of indicators (US 10yr Treasury rate and the world GDP growth in real terms) provides information on a global scale. The relevance of the US interest rates for the global economy has been recently stressed by Iacoviello and Navarro (2019). Their results show that the foreign spillovers of higher US interest rates are large, and depend on foreign countries exchange rate arrangements with the US dollar, intensity of trade flows with the US and financial channels. Regarding financial channels, Rey (2015) and Miranda-Agrippino and Rey (2019) show that changes in interest rates in ‘‘core” countries can trigger a global financial cycle that, regardless of the exchange rate regime, may generate global spillovers. Bruno and Shin (2015) find evidence of monetary policy spillovers on cross-border capital flows. The conditions of the global economy as a whole are captured by the real growth of world GDP.

Before building our model, where needed, we perform some data manipulation, i.e. detrending and standardisation. Detrending allows us to make some variables stationary by removing their time trend and capturing the cyclical component, while the standardisation smooths heterogeneities among countries. Variables are detrended by using a 1-sided Hodrick-Prescott (HP) filter with a smoothing parameter equal to 6.25 (Ravn and Uhlig, 2002).[7] An exception is the current account balance to GDP that is only standardised. Finally, no manipulation is performed on US 10yr Treasury rate and world GDP growth in real terms, while inflation is winsorised.[8]

Table 3 presents summary statistics of the dependent variables for advanced and emerging economies separately. Additionally, it presents the means of the raw data used in the analysis, discriminating between means in the full sample ("overall"), and the means for the

[7] We opt for a 1-sided HP filter with the idea that a policy maker would feed the model with information available up to time *t* to make predictions on time *t+1*.

[8] We replaced values of inflation below the 1st percentile and above the 99th percentile of the distribution with the 1st and 99th percentile, respectively.

four sub-periods (pre-crisis, crisis, post-crisis and normal times) defined in the previous section. Two main observations arise.

TABLE 3 ABOUT HERE

First, there is a significant difference in a number of country-specific variable means between advanced and emerging economies, as shown by the t-test reported in **Table A.5** in Appendix A. This further justifies our choice to analyse the two groups separately. Current account deficits are larger in emerging economies, which puts them in a more vulnerable position compared to advanced countries. External debt to GDP is relatively high in advanced economies. This result is not surprising as advanced countries are usually more integrated in the global economy compared to the developing world. Heterogeneities are also present with regards to credit to GDP. It suggests that in emerging economies the banking sector is not as developed as in the other countries and that banking crises are less likely to be triggered by an overheated credit market. Moreover, inflation is much higher in emerging economies. Although this result is partly due to countries experiencing significant inflationary pressures, prices are generally much more volatile in emerging countries than in developed ones.

Second, as expected, most of the selected macroeconomic variables indicate a worsening of the macroeconomic situation in the run up to the crisis. For example, current account deficits deteriorate as we approach the crisis. External debt to GNI is relatively low in normal times and it increases as we move towards the outbreak of the crisis. For emerging economies, inflation increases substantially in the wake of the crisis and more so in the midst of it. Credit to GDP increases as we approach the crisis but more for developed countries. Overall, this suggests that the four sub-periods classification is appropriate. Yet, only a few of the selected macroeconomic variables improve in the post-crisis period, consistently with findings that banking crises tend to have a lasting negative effect on activity. Moreover, a number of indicators present a similar behaviour to that exhibited in the pre-crisis period. Thus, to avoid any confounding factors that could affect our results, we exclude the post-crisis years from our empirical application.

4. Predictive models of banking crises

We employ our dataset on pre-crisis spells to build two sets of predictive models, one for 33 advanced economies and one for 67 emerging countries, over the period 1970-2017: a traditional logit model (Section 4.1) and a machine learning tool, i.e. the AdaBoost (Section

4.2).[9] For our baseline models, in both logit and AdaBoost and for the two sets of countries, the dataset is split into a training set covering the period 1970-2005 and a testing set covering the period 2006-2017.[10] Before proceeding to the estimation of models, we drop observations with missing information for our macroeconomic variables. Estimation sample sizes are displayed in **Table 4**.

TABLE 4 ABOUT HERE

On the one hand, the training set is used to estimate the models' "hyperparameters" (Section 4.3). The logit and the AdaBoost share one hyperparameter, the cut-off, used to classify the estimated probabilities of incurring a pre-crisis as 1 or 0. Other parameters are specific to the AdaBoost, i.e. tree depth, minimum split and number of trees. On the other hand, the testing set is used to evaluate the predictive performance of the model (Section 4.4). Separating hyperparameter selection from model evaluation allows us to limit overestimation of the model performance, in particular overfitting, as remarked by Beutel et al. (2019).

In both the logit and the AdaBoost, the set of explanatory variables for both sets of countries, advanced and emerging, includes current account to GDP, external debt to GNI, public debt to GDP, credit to GDP and inflation. We also control for the US 10yr Treasury rate and world growth in real terms. Finally, we interact external debt to GNI and the US 10yr Treasury rate. We do not include country fixed effects, since the number of 1s (i.e. pre-crises) is scarce at country level. However, because of the existing heterogeneity among emerging countries we control for country aggregates.[11] We also provide an "extended" model in which we include additional variables, i.e. country-specific GDP growth and global energy prices expressed as percentage changes, for both advanced and emerging countries. House prices and real effective exchange rates (REER) both expressed in percentage changes are added in the

[9] See Freund and Schapire (1996) on Adaptive Boosting.

[10] In the robustness analysis of Section 6 and Appendix B, we show results taking 1999 as the cut-off year and splitting the sample into a training set covering the period 1970-1999 and into a testing set covering the period 2000-2017.

[11] The country aggregates are the following: "Balcans and Eastern Europe", "Central Asia", "East Asia", "South Asia", "Russia and Belarus", "Middle East", "Latin America", "Caribbean", "Central America", "Pacific", "North Africa" and "Africa".

"extended model" for advanced countries only (see Section 6 and Appendix B).[12] The goal of adding these variables is to check the validity of our baseline results.

4.1 Logit model

The first step of our analysis consists of applying standard econometric techniques to identify the macroeconomic indicators that significantly affect the likelihood of the occurrence of a pre-crisis spell in the period 1970-2005. In particular, we estimate two logit models, one for advanced economies and one for emerging ones:

$$Prob(y_{it} = 1|X_{it}) = \frac{exp\,(X'_{it}\beta)}{1 + exp\,(X'_{it}\beta)} \tag{1}$$

where $Prob(y_{it} = 1|X_{it})$ denotes the probability that country *i* in year *t* is in a pre-crisis state and X_i is the set of regressors outlined above. Given that we are looking at pre-crisis spells, the information set included in X_i is taken at time *t*.

The estimation results, expressed as marginal effects, are displayed in **Table 5**. As for advanced economies, public debt to GDP and credit to GDP display statistically significant and positive marginal effects. These results are reasonable, since countries with higher levels of public debt are expected to be more fragile and less able to counteract the emergence of a banking crisis. The same holds if a country has recorded a significant expansion of credit to GDP that may not reflect economic fundamentals. Turning to emerging economies, the likelihood of experiencing a pre-crisis falls as public debt increases. A possible interpretation is that a pre-crisis period could be characterized by a decrease in public debt to GDP thanks to GDP growth and pro-cyclical improvement of the primary balance, while when the crisis breakes out, fiscal measures implemented by government could burden public debt. External debt to GNI has a negative marginal effect, possibly related to external official support, while the US 10yr Treasury rate is positively related to the likelihood of incurring a pre-crisis; also, their interaction positively affects the likelihood of observing a pre-crisis period.

TABLE 5 ABOUT HERE

4.2 AdaBoost algorithm

[12] In the model for emerging economies, we do not include house prices and the real effective exchange rate because they are available only for a limited number of countries and years resulting in a relevant drop in the number of observations.

Turning to the machine learning approach, a possible tool to solve classification problems consists of decision tree classifiers, which select features (variables) and their critical values to classify the outcome variable. They offer the advantage of being able to capture non-linear relationships between features and the target variable. A decision tree classifier is a partitioning algorithm that recursively chooses the predictors and the thresholds that are able to best split the sample into the relevant classes (in our case, pre-crisis and normal times) according to a measure of dispersion or "impurity". A typical example of impurity is the Gini index, which takes a value of zero when all observations belong to the specified class - pure classification - and a value of 1 when observations are randomly distributed across classes. Technically, the tree starts from a root node, which collects all the training set observations. The initial sample is split into two child nodes, according to the impurity criterion. Each of these child nodes can be further divided into two more child nodes based on the variable that best splits the corresponding subsamples. This recursive procedure stops when there is no further gain in splitting a subset (i.e. the impurity measure does not improve) or a binding rule applies (i.e. the pre-set maximum number of splits has been reached in order to avoid overfitting and poor forecast performance). The nodes that cannot be optimally split further are called terminal nodes. **Figure 3** depicts a simple tree.

FIGURE 3 ABOUT HERE

The AdaBoost belongs to the family of so-called "ensemble models", which estimate a multitude of trees (to grow a forest) instead of a single tree as in the CART (Classification and Regression Tree).[13] Specifically, the algorithm builds N base models sequentially: in each replication, observations incorrectly classified at the preceding step are attributed a higher probability to be selected in the new training set ("boosting"). By doing so, the model gives more weight to the observations that are more difficult to predict with the aim to reduce the model bias and capture crises of different nature. The model implies selection of the hyperparameters, i.e. tree depth (maximum number of nodes along the longest path from the root note down to the farthest leaf node), minimum split (minimum number of observations in a node to allow for a split), number of trees and, like in the logit model, cut-off. The AdaBoost is run on the same dataset of the logit model, separately for advanced and emerging economies,

[13] An empirical application of AdaBoost can be found in Alfaro et al. (2008) who apply it to predict corporate failures in Europe.

and employing the same explanatory variables. Setting the hyperparameters is the objective of the next sub-section.

4.3 In-sample cross-validation

The rationale for cross-validating the model in-sample lies in the fact that, as remarked at the beginning of the section, separating hyperparameter selection from model evaluation allows us to limit overfitting and improve out-of-sample predictions. The hyperparameters are chosen such that they optimize a performance criterion, the so-called "relative usefulness" function. To this end, we follow the evolution in the recent literature (Holopainen and Sarlin, 2017, and Beutel et al., 2019) and implement a five-year panel-block cross-validation on the training set (1970-2005) for the selection of the hyperparameters that will be used in the out-of-sample exercise on the testing set (2006-2017).[14] Selecting the optimal hyperparameters generally depends on the loss function of the forecaster. The current standard is to choose the classification hyperparameters such that they maximize a relative usefulness function (Alessi and Detken, 2011; Holopainen and Sarlin, 2017; Beutel et al., 2019), which weighs type I and type II errors (as a share of the respective actual class) by a parameter (μ) representing the forecaster's preferences.[15] [16] For each set of hyperparameters, the relative usefulness $U_r(\mu)$ is a function of μ and is defined as:

$$U_r(\mu) = 1 - L(\mu)/min\,(\mu, 1-\mu) \tag{2}$$

where $L(\mu)$ is the loss function:

$$L(\mu) = \mu[FN/(FN + TP)] + (1-\mu)[FP/(FP + TN)] \tag{3}$$

[14] Our five-year panel-block cross validation consists in a rolling estimation of the model that recursively adds a panel block in the training sample and makes the prediction on the next panel block. We settle for a block of five years as our pre-crisis window is of three years, thus making us sure to maintain the auto-correlation and cross-correlation features of our dataset. For each combination of hyperparameters, the performance is computed by averaging the results for each panel-block used as a testing sample. In the logit model, where the only hyperparameter is the cut-off, we build a grid of cut-off values and the corresponding relative usefulness. We choose the cut-off value that maximizes this measure of model performance. In the AdaBoost, the parameter-grid also includes minimum split, number of trees and tree depth. We follow the same approach in all the robustness analyses of Section 6 and Appendix B.

[15] The relative usefulness function of Holopainen and Sarlin (2017) is an evolution of Sarlin's (2013).

[16] Very recently, Sarlin and Schweinitz (2021) have suggested two alternative methods: one in which they include preferences in the estimation itself and one in which they set thresholds ex-ante.

and where FN stands for false negatives, FP for false positives, TP for true positives and TN for true negatives. The relative usefulness $U_r(\mu)$ reaches its maximum of one if $L(\mu) = 0$, and a value of zero if $L(\mu) = (\mu, 1 - \mu)$. When $U_r(\mu) > 0$ the model is more informative than the naïve decision rule (coin toss). As standard in the literature, we set $\mu = 0.5$, thus weighting type I and type II of errors equally. **Table 6** displays the selected hyperparameters obtained from the panel-block cross-validation. In the logit model the cut-off values that maximize the loss function are 0.112 and 0.08 in advanced and emerging countries, respectively. In the AdaBoost, the corresponding value is 0.45 in both sets of countries.[17] The other optimizing hyperparameters are as follows: the minimum split is equal to 70 and 30, the number of trees is 30 in both sets of countries and finally, the tree depth is 4 in both advanced and emerging economies.

TABLE 6 ABOUT HERE

Now that we have selected the hyperparameters, we use them to evaluate which model between the logit and the AdaBoost performs better in-sample in terms of performance indicators, such as AUROC, sensitivity, specificity and relative usefulness.[18] To this end, we run a panel-block bootstrap as in Beutel et al. (2019), where we sample 100 different bootstrap datasets with replacement and where the length of each block is 5 consecutive years.[19] **Table 7** displays the median values of the in-sample performance indicators and their interquartile

[17] Beutel et al. (2019) find similar values in all specifications of the logit model and values ranging between 0.097 and 0.373 in the machine learning algorithms. Regarding the AdaBoost cut-off, it is useful to clarify that it should not be considered as a probability in the traditional sense. In fact, in order to increase the performance of other machine learning algorithms, AdaBoost combines weak classifiers by focusing on misclassified observations at each iteration in order to gradually reduce the model bias. The probability assigned to each observation thus results as a weighted average of the predictions made at each step, where the weights of the samples depend on their classification error.

[18] The AUROC is calculated from the ROC curve, which plots the combinations of true positive (i.e. 1) and false positive (i.e. 0) rates attained by the model. It corresponds to the probability that a classifier ranks a positive instance higher than a negative one. The AUROC ranges from 0 to 1, where 0.5 corresponds to the AUROC of a random classifier, while 1 that of a perfect classifier. The closer the AUC is to one, the better the model predicts. Sensitivity and specificity rates derive from the confusion matrix, a two-way table that compares predicted values with observed ones. Sensitivity is the ratio between the correctly predicted positives and total observed positives, while specificity measures the proportion of negatives that are correctly identified.

[19] This allows us to maintain the autocorrelation structure of the data and at the same time to capture cross-sectional correlation. The number of observations of each bootstrap dataset is constrained to be equal to that of the training set.

range (IQR, i.e. the interval between the 25th and the 75th percentile of the distribution) for advanced and emerging countries, for the logit model and AdaBoost. The IQRs are computed by estimating the models 100 times and by taking the first and third quartile of the distribution of each indicator (AUROC, sensitivity, specificity and relative usefulness). Statistically significant differences between indicators are assessed by looking at whether IQRs overlap or not: when they do we say differences are not statistically different, while when they do not overlap we say differences are statistically different. For advanced economies, the AdaBoost displays a better performance than the logit in terms of all indicators and it attains a higher relative usefulness. For the sample of emerging economies, both the AdaBoost and logit perform worse than in the case of advanced economies. Nevertheless, almost all the indicators confirm the superiority of the in-sample performance of the AdaBoost when we compare the median values, although they are not statistically different between models with the exception of specificity.

TABLE 7 ABOUT HERE

4.4 Out-of-sample performance

We rely on a recursive, or rolling, out-of-sample exercise to evaluate the performance of our models to retrieve estimated pre-crises and performance measures (AUROC, sensitivity, specificity and relative usefulness) for both sets of countries.[20] **Table 8** shows the out-of-sample performance measures and the related IQRs of both models calculated from the

[20] For each year from 2006 to 2017, we estimate both the logit and the AdaBoost recursively, using the information available up to that point. For example, to predict 2007, we use the predictions obtained from estimating the model from 1970 to 2006. In the next step, we use the information up to 2007 to estimate 2008. We proceed in the same way until the last year of our sample. In each step, we predict pre-crisis probabilities based on which we retrieve the number of estimated pre-crises. Then, we build confusion matrices and calculate out-of-sample performance measures, i.e. AUROC, sensitivity, specificity and relative usefulness. In order to obtain confidence intervals, we implement the panel-block bootstrap procedure described above. This exercise is performed on the sample of advanced and emerging economies separately. We obtain predictions for 256 observations, including 36 pre-crisis periods, for the sample of advanced economies, and 779 observations, including 10 pre-crisis periods, for the sample of emerging economies. As in most of the papers working with crisis episodes (see for example Bernardini and Forni, 2020), we have to rely on a limited number of crisis episodes. Noteworthy, is that the number of pre-crisis episodes for emerging countries is significantly lower than the number of pre-crisis episodes for advanced economies because the latter are mainly concentrated in the training sample. In fact, the number of pre-crisis episodes for emerging countries increases to 24 when we apply 1999 as the cut-off year between the training and testing set.

distributions obtained by estimating each model 100 times for each year from 2006 to 2017. Again, we look at whether IQRs overlap to gauge statistical significance of the differences between indicators across models. The AdaBoost always outperforms the logit model, which attains an out-of-sample performance that is barely better than a coin toss for the sample of both advanced and emerging economies.[21] Furthermore, the AdaBoost delivers a comparable performance across samples in terms of AUROC and specificity, while sensitivity and relative usefulness are lower in the emerging economies than in the advanced ones. Overall, these results suggest that the AdaBoost model is better performing for our recursive out-of-sample forecasting exercise. Although we are aware that it is not easy to compare our results with similar work in the literature because of the differences in the data employed and the countries covered, the results of a better performance of ML algorithms are in line with those of Holopainen and Sarlin (2017) and Bluwstein et al. (2020), while are in contrast with those of Beutel et al. (2019).[22] Since the AdaBoost has a better performance, in what follows we present the predicted probabilities, the predicted pre-crises and the confusion matrices we obtained from it.[23]

TABLE 8 ABOUT HERE

[21] Both the in-sample performance and out-of-sample one of the AdaBoost are also superior to those of a Random Forest, as detailed in Section 6 and Appendix B. Furthermore, it emerges that for the set of advanced countries the RF outperforms the logit model both in-sample and out-of-sample. Meanwhile for emerging economies, the RF performs better than the logit in the out-of-sample exercise. Overall, these results suggest that the substantial improvement of the performance of the AdaBoost model compared to the logit model comes from the boosting part of the algorithm. This finding indicates that boosting algorithms can substantially improve the performance of ML models applied to predicting the build-up of systemic banking crises. Moreover, it reinforces the contribution of our paper to the existing literature where the use of boosting algorithms in a macro-financial context is still scarce.

[22] We identify many possible reasons why we obtain different results from Beutel et al. (2019). First, we exclude post-crisis spells from the sample, while they do not. Second, we use yearly data, while they employ quarterly ones. Third, we employ the AdaBoost algorithm, which is not included in the set of ML techniques they estimate. Fourth, we also account for global factors by means of world GDP growth and US 10yr Treasury rate. Fifth, we estimate our models on all advanced countries, while they restrict theirs to thirteen European countries, the US and Japan. We run a robustness check by estimating our models on the same sample of countries of Beutel et al. (2019) and find that the AdaBoost still performs better than the logit (see Section 6 and Appendix B).

[23] The corresponding results derived from the logit model are available upon request. They show that the logit exhibits a lower performance than the AdaBoost, especially in terms of the large number of false positives, which may also arise from the low value of the optimized cut-off.

Figure 4 shows the estimated predicted probabilities of pre-crises, for both groups of countries. As for the advanced countries, it indicates that the distribution of predicted probabilities by year performs well in replicating the path of the number of financial crises. It is noteworthy the upward shift of the distributions in 2017, below the levels reached in the years preceding the 2007-2008 crisis, but closer to the levels reached during the 2011-2012 sovereign debt crisis in Europe. This increase may signal a mounting fragility for the advanced economies. As for the emerging economies, predicted probabilities fit poorly and remain low.

FIGURE 4 ABOUT HERE

In **Figure 5**, the upper panels display the subsample of observed pre-crises, split into those that are correctly predicted (true positives, grey bars) and those that are incorrectly classified (i.e. false negatives, black bars). The lower panels display the subsample of predicted pre-crises, split into the ones that were observed (true positives, striped grey bars) and the ones that are incorrectly classified (i.e. false positives, striped black bars). The black and the striped black bars inform on how often the model fails to predict an observed crisis and on how often the model predicts a false crisis, respectively. The absolute values of the grey and the striped grey bars provide the same information, i.e. the observed crises that are correctly predicted.

FIGURE 5 ABOUT HERE

For advanced countries (**Figure 5a**), a large number of incoming crises are correctly predicted in 2006 and 2007, because the model exploits the information on 2005 that is now available due to the rolling method employed in the exercise. This is not true for emerging economies (**Figure 5b**), for which only few pre-crisis episodes are signalled by the model over the prediction horizon. On the other side, in both groups of countries the number of false alarms is limited. Moreover, these predicted false alarms do not necessarily signal a false crisis because it is possible that policy makers/monetary institutions enforced policies aimed at preventing the outburst of a potential banking crisis. Obviously, no machine learning method can account for this.

5. Ranking the determinants of banking crises

One goal of this paper is to identify the main leading indicators of financial crises among those we have considered. To this end, for each variable we calculate two measures: (i) the so-called "relative importance" (e.g. Carmona et al., 2019, and Antulov-Fantulin et al., 2021) and (ii) Shapley values (Bluwstein et al., 2020). "Importance" indicates how useful or valuable each variable is in the construction of the decision trees within the model. Technically, the more a variable contributes to distinguish among the possible outcomes, reducing the impurity measure, the higher its relative importance. Shapley values provide the contribution, positive or negative, of each variable to the probability of a pre-crisis. Loosely speaking, they provide the marginal effect of each variable in anticipating a crisis and are specific to each country and year.

5.1 Relative importance

Relative importance is calculated for a single decision tree by the amount that each indicator split point reduces the impurity measure – i.e. the Gini index – weighted by the number of observations the node is responsible for. The relative importance is then averaged over all trees by giving different weights to each tree based on its classification error. For ease of interpretation, we measure the relative importance of each variable in percentage terms so that the assigned value provides immediate and intuitive information on the weight of that indicator compared to others in detecting pre-crisis periods. It follows that the sum of the percentages across the selected indicators is equal to 100. Moreover, we calculate the relative importance using the sample from 2000, instead of just focusing on our post 2006 sample. Looking at a longer timeframe allows us to observe changes over time in each variable's role in anticipating banking crises. For this purpose, we implement the same recursive procedure used for the out-of-sample exercise of Section 4.4, but we apply it for each year from 2000 to 2017.

Figure 6 and **Figure 7** show the time evolution of the relative importance of each macroeconomic indicator for the two sub-samples. For the advanced countries, at the beginning of our sample the current account to GDP, inflation and public debt to GDP were the most relevant indicators for the classification task with a weight of 18.6%, 17.7% and 16.8% in 2000, respectively (**Figure 6**). Their role declined overtime, especially the current account. In the last years of the sample the relative importance of the US 10yr Treasury rate and world growth increased over the sample period and by 2017 they represented the most important indicators with a weight of 20.6% and 14.5%, possibly indicating the strong inter-connection among countries that facilitates the transmission of risk across countries.

In the "extended" model introduced in Section 4 (see also Section 6 and Appendix B), we add country-specific GDP and energy prices growth, for both advanced and emerging countries, and – only for advanced ones, due the lack of sufficient data for emerging countries

– also house price changes and real effective exchange rate changes. We find that for advanced economies a key role is played by the real effective exchange rate. Inflation, global energy index, credit to GDP and the US 10yr Treasury rate also explain a significant share of the variability in outcomes, although with fluctuations throughout time (**Figure B.1** in Appendix B). A possible explanation of the relevance of the real effective exchange rate is that the countries that experienced a banking crisis around the global financial crisis and the succeeding euro area sovereign debt one lost competitiveness before 2008. In pre-crisis periods, capital flew from wealthier countries to peripheral ones leading to an appreciation of the REER. During crises, instead, uncertainty and market panic triggered a sudden stop and reversal of capital flows, thus causing an adjustment in the REER.

FIGURE 6 ABOUT HERE

For the emerging countries, at the beginning of our sample (2000), inflation, world growth and public debt to GDP were the most relevant indicators for the classification task with a weight of 16.9%, 17.4% and 19.1%, respectively (**Figure 7**). World growth and public debt maintain a relevant weight throughout our sample. The US 10yr Treasury rate, instead, was not among the most relevant macroeconomic indicators until the mid-2000s, while its importance started to increase from 2004 and, by the end of our sample, its weight was equal to 42% and thus, represented the most relevant variable along with world growth (with a weight of 13.3%). In the “extended” model, the role of the US 10yr Treasury rate as the variable with the strongest contribution to anticipating a crisis is confirmed. It is followed by variables such as current account to GDP, inflation and public debt to GDP (**Figure B.1** in Appendix B).

FIGURE 7 ABOUT HERE

It is important to stress that the relative importance of the variables should not be interpreted as their marginal contribution to a pre-crisis situation. In the AdaBoost set up, variables with a higher relative importance are those that contribute the most in minimizing the impurity measure in each node and therefore contribute the most in the classification task. With this caveat in mind, it is remarkable the increasing role of global factors over time in explaining pre-crisis situations in both advanced and emerging economies revealed by our baseline model. This is in line with the recent literature on global financial cycles (Rey 2018; Miranda-Agrippino and Rey 2019), documenting the growing importance of monetary policy in “core” countries in determining the financial condition at the global level. Indeed, the changes that we uncover starting from the early 2000s seem to reflect some structural changes in global financial conditions. This is likely related to the growing role of the US dollar as international currency and reserve asset starting from the 2000s (Farhi and Maggiori, 2018; Maggiori et al. 2019), which went along a large and persistent US current account deficit. An alternative interpretation, is that as we move closer to 2008, global factors start do dominate because the 2008 crisis has been global in nature and triggered by strong financial and trade interconnections among countries more than country-specific characteristics.

5.2 Shapley values

Since relative importance does not provide the marginal contribution of each predictor to the probability of pre-crises, we follow Bluwstein et al. (2020) who borrow from the cooperative game theory literature (e.g., Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017) to compute Shapley values (Shapley, 1953; Young, 1985). The predicted pre-crisis probability for each observation is decomposed into a sum of contributions from each variable, i.e. its Shapley values.[24] Individual values are then averaged to obtain the Shapley value of each predictor by year. This allows us to identify which variables are driving our prediction and, more importantly, to quantify their contribution and direction.

FIGURE 8 AND FIGURE 9 ABOUT HERE

For advanced economies, the variable with the largest predictive share in absolute terms across the whole period is the US 10yr Treasury rate, while world growth contributed especially between 2006 and 2015 with a positive contribution until 2010 and then a negative one (**Figure 8**). The role of world growth, however, narrowed substantially in the last two years of our sample. It is noteworthy that the US 10yr Treasury rate has acted against the build-up of a crisis (its Shapley values are negative, with the exception of the years 2005-2007), mitigating the positive contribution of other factors such as external debt, credit and current account. For emerging economies, we detect the same behaviour of the US 10yr Treasury rate with respect to relevance and sign (although the negative-positive-negative sign pattern is a couple of years ahead of that of the advanced countries), while world growth contributes with a positive sign to the build-up of crises (**Figure 9**). Furthermore, current account provides an increasing and positive contribution to fragility from 2009, partially compensated by the interaction between external debt and the US Treasury 10yr rate. Finally, the role of credit reduces in importance throughout the period of analysis.

[24] For details on how Shapley values are computed, see Bluwstein et al. (2020).

Accordingly, **Figure B.2** of Appendix B displays Shapley values of the extended model. For both advanced and emerging economies, the results confirm that the most relevant effect is that of the US 10yr Treasury rate, with a positive contribution up to 2008 and a negative one afterwards, when the US monetary policy has been very accommodative.

6 Robustness analysis

In this Section we discuss the results of a series of different model specifications designed to test the robustness of our "baseline" logit and AdaBoost models of Section 4 to (i) different variable sets and to (ii) the cut-off year used to split the sample into the training and testing sets. Other robustness checks are presented in **Appendix B**.

To estimate the models and test them in-sample and out-of-sample we follow the same methodology presented and discussed in Section 4. We recall that in the baseline model the cut-off year is 2005 and the variables employed are current account balance to GDP, external debt to GNI, public debt to GDP, credit to GDP, US 10yr Treasury rate and world GDP growth in real terms (see Section 3.2 for a detailed description). Table B.1 in Appendix B displays the full list of variables of all models. The new specifications are the following:

(1) "Cut-off 1999": cut-off year at 1999 and same variable set as in the baseline.

(2) "Extended": cut-off year at 2005 (as in the baseline) and variable set extended to include country-specific GDP and energy prices growth, for both advanced and emerging countries, and – only for advanced ones – also house price changes and real effective exchange rate changes.

(3) "à la Beutel": cut-off year at 2005 and countries as in Beutel et al. (2019), i.e. thirteen European countries, US and Japan.

Finally, we also train a Random Forest model on both the "baseline 2005" setting, i.e. cut-off year at 2005 and same variables as in the baseline model, and on the three new specifications.

Figure 10 displays the median values of relative usefulness and the interquartile range (IQR) for the above three model specifications for advanced and emerging economies.[25] It also shows those of the baseline model for ease of reading. A comparison between models is performed by looking at their relative usefulness and their IQRs. In-sample and out-of-sample

[25] The specification "à la Beutel" is only available for a selection of advanced economies by construction.

IQRs are computed as in Sections 4.3 and 4.4. Three main takeaways emerge. First, for the advanced economies in-sample, the AdaBoost displays statistically higher relative usefulness values than the logit model in model specifications (1) and (2). In the out-of-sample case, this evidence is even more marked. In the latter case, AdaBoost also shows smaller IQRs than the logit, indicating its greater stability. For the emerging economies, similar results hold. When using 1999 as the cut-off year, the performance of the AdaBoost is in line with that reported for the baseline specification, thus confirming that our results are robust regardless of the cut-off year used. Second, the AdaBoost displays higher relative usefulness than the Random Forest in most circumstances, although differences are not always statistically significant. However, both the Random Forest and AdaBoost display a higher relative usefulness compared to the logit model. Third, the AdaBoost performance of the baseline model is very similar to that of the extended model for advanced economies both in sample and out of sample and for emerging economies out of sample, while it is lower in the case of emerging economies in sample.

7. Conclusions

The dramatic worldwide losses triggered by the 2007-2008 financial crisis urged policy makers to understand the macroeconomic vulnerabilities that led to its build-up. In particular, as the crisis spread as a systemic banking crisis, the connections between money and credit fluctuations and financial crises took centre stage. The ultimate goal was to layout macroprudential policies that could warrant a timely response to countries' weaknesses, thereby significantly limiting the burdensome costs entailed by similar crises.

In this paper, we contribute to the literature by developing predictive models of banking crises for both advanced and emerging economies with the aim of performing a ranking of a set of recognized determinants of banking crises and to quantify their contribution. To this end, we employ an integrated dataset of banking crises and macroeconomic indicators that includes a total of 100 countries (33 advanced and 67 emerging) over the period 1970-2017. We build our models by using both traditional econometrics – the logit model – and a supervised machine learning algorithm – the AdaBoost – and relate pre-crisis spells to a set of possible country-specific determinants: current account balance to GDP, public debt to GDP, credit to GDP, external debt to GNI and inflation. We also include two global variables, i.e. the US 10yr Treasury rate and world GDP growth in real terms. An extended version of the model includes additional variables i.e. country-specific GDP growth and global energy prices, for both advanced and emerging countries, and also – but only for advanced ones – house prices and real effective exchange rate changes.

The two models, estimated separately for advanced and emerging countries, are validated in-sample by splitting the sample into a training set and into a test set to estimate the models' hyperparameters and to evaluate the predictive performance of the models, respectively. This procedure has the advantage of limiting overfitting. In the second part of our empirical exercise, we test the out-of-sample predictive ability of our models and the AdaBoost shows a higher performance in terms of AUROC, sensitivity, specificity and relative usefulness. These results are robust to different model specifications.

We then use the AdaBoost to rank the major determinants among those selected leading to a banking crisis by means of two indicators – relative importance and Shapley values – that provide an indication of each variable's relevance in solving the classification problem and quantify their contribution to the predictions, respectively. Our results reveal that the US 10yr Treasury rate plays a key role and growing-over-time as a leading indicator of banking crisis for both advanced and emerging economies. As for the sign of its contribution, it acts against the build-up of banking crisis especially until 2003-2004 and from 2007-2008. World GDP growth is the second most relevant variable in both sets of countries, although it shows a different pattern: for advanced economies, it increases until 2010 and stabilises thereafter, while for emerging ones, it is rather stable, although with slight fluctuations. Its sign becomes negative at the start of the sovereign debt crisis for advanced economies, while it remains positive for emerging ones. Among the country specific variables, inflation is the one with the highest relative importance, followed by public debt, current account and credit, for advanced countries. For emerging countries, other relevant variables are inflation, current account and public debt, which, however, display a decreasing pattern over time compared to the global variables. When extending the model to account for additional leading factors, it emerges that – for advanced countries – changes in the real effective exchange rate are an important leading indicator. At the same time, the key role of the US 10yr Treasury rate is confirmed for both advanced and emerging economies.

These results suggest that the determinants of banking crises change over time with the evolving domestic and international context, and that variables that are relevant in explaining events located in a distant past might be less relevant today. The recent developments in global financial conditions, and specifically the increased dimension of dollar denominated capital flows in recent years and the growing role of the "global financial cycle", are affecting many dimensions of financial conditions, including by exposing banking sectors' fragilities, and might explain the leading role of the US interest rate. Interestingly enough, inflation, which can be interpreted as a proxy for poorly managed macroeconomic policies, is still relevant for emerging economies, although less so than in the past. This decreasing role of inflation likely reflects the improvements that many emerging economies have achieved in their management of monetary and fiscal policies. Further research will be necessary to establish the causation channels of the evidence presented in this paper.

Acknowledgments and disclaimer

We thank the editor and two anonymous referees for their valuable comments. We are also grateful to seminar participants at Prometeia (2018 and 2019) and at the University of Bologna (2019) as well as participants at BigNOMICS Workshop on Big Data and Economic Forecasting (2019), 16th EUROFRAME Conference (2019), 34th Annual EEA Congress (2019) and Virtual ISF Conference (2020). All remaining errors are our responsibility. The opinions expressed herein are our own and do not necessarily reflect those of the affiliated institutions.

Bibliographic references

Alessi, S., and 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*, Vol. 27, pp. 5203-533.

Alessi, S., and Detken, C., 2018, "Identifying excessive credit growth and leverage", *Journal of Financial Stability*, Vol. 35, pp. 215-225.

Alfaro, E., García, N., Gámez, M. and Elizondo D., 2008, "Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks", *Decision Support Systems*, Vol. 45, pp. 110-122.

Antulov-Fantulin, N., Lagravinese, R. and Resce, G., 2021, "Predicting bankruptcy of local government: A machine learning approach", *Journal of Economic Behavior and Organization*, Vol. 183, pp. 681-699.

Arteta, C., and Eichengreen, B., 2000, "Banking crises in emerging markets: presumptions and evidence", Center for International and Development Economics Research (CIDER) Working Papers, C00-115, August.

Athey, S., 2019, "The impact of machine learning on economics", In: Agrawal, A., Gans, J., and Goldfarb, A., (Eds.) *The economics of artificial intelligence: An agenda*, University of Chicago Press. pp. 507-547.

Babecký, J., Havránek, T., Jakub Matějů, J., Rusnáka, M., Smídkováa, K., and Vasícek, B., 2014, "Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators", *Journal of Financial Stability*, Vol. 15, pp. 1-17.

Baron, M., Verner, E., and Xiong, W., 2019, "Salient Crises, Quiet Crises", Available at SSRN: https://ssrn.com/abstract=3116148.

Baron, M., Verner, E., and Xiong, W., 2021, "Banking Crises without Panics", *Quarterly Journal of Economics*, Vol. 136(1), pp. 51-113.

Barrell, R., Davis, E. P., Karim, D., and Liadze, I., 2010, "Bank regulation, property prices and early warning systems for banking crises in OECD countries", *Journal of Banking & Finance*, Vol. 34(9), pp. 2255-2264.

Bernardini, M. and Forni, L., 2020, "Private and Public Debt Interlinkages in Bad Times", *Journal of International Money and Finance*, Vol. 109.

Beutel J., List, S., and von Schweinitz, G., 2019, "Does Machine Learning Help Predicting Banking Crises?", *Journal of Financial Stability,* Vol. 45.

Bluwstein, K., Buckmann, M., Andreas, J., Miao, K., Sujit, K., and Özgür, S., 2020, "Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach", Bank of England, Staff Working Paper No. 848.

Bordo, M., Eichengreen, B., Klingebiel, D., Martinez-Peria, M.S., and Rose, A.K., 2001, "Is the Crisis Problem Growing More Severe?", *Economic Policy,* Vol. 16(32), pp. 51-82.

Borio, C., and Lowe, P., 2002, "Assessing the risk of banking crises*", BIS Quarterly Review,* December, pp. 43-54.

Bruno, V., Shin, H.S., 2015. "Capital flows and the risk-taking channel of monetary policy", *Journal of Monetary Economics,* Vol. 71, pp. 119-132.

Bussiere, M., and Fratzscher, M., 2006, "Towards a new early warning system of financial crises", *Journal of International Money and Finance,* Vol. 25(6), pp. 953-973.

Caggiano, G., Calice, P., and Leonida, L., 2014, "Early warning systems and systemic banking crises in low income countries: A multinomial logit approach", *Journal of Banking & Finance,* Vol. 47, pp. 258-269.

Caprio, G., and Klingebiel, D., 1997, "Bank Insolvency: Bad Luck, Bad Policy, or Bad Banking?", Annual World Bank Conference on Development Economics 1996, The International Bank for Reconstruction and Development/The World Bank.

Caprio, G., and Klingebiel, D., 2002, "Episodes of Systemic and Borderline Financial Crises", In: Klingebiel, D. and Laeven, L. (Eds.), *Managing the Real and Fiscal Effects of Banking Crises,* The World Bank Discussion Paper n. 428, pp. 31-48.

Carmona, P., Climent, F., and Momparler, A., 2019, "Predicting failure in the U.S. banking sector: An extreme gradient boosting approach", *International Review of Economics & Finance,* Vol. 61, pp. 304-323.

Cesa-Bianchi, A., 2013, "Housing cycles and macroeconomic fluctuations: A global perspective", *Journal of International Money and Finance,* Vol. 37, pp. 215-238.

Davis, E.P., and Karim, D., 2008, "Comparing early warning systems for banking crises", *Journal of Financial Stability,* Vol. 4, pp. 89-120.

Demirgüc-Kunt, A., and Detragiache, E., 1998, "The Determinants of Banking Crises in Developing and Developed Countries", *IMF Staff Papers,* Vol 45(1), pp. 81-109.

Demirgüc-Kunt, A., and Detragiache, E., 2000, "Monitoring banking sector fragility: a multivariate logit approach", *World Bank Economic Review,* Vol. 14, pp. 287-307.

Demirgüc-Kunt, A., and Detragiache, E., 2005, "Cross-country empirical studies of systemic bank distress: a survey", *National Institute Economic Review,* Vol. 192(1), pp. 68-83.

Duttagupta, R., and Cashin, P., 2011, "Anatomy of banking crises in developing and emerging market countries", *Journal of International Money and Finance,* Vol. 30(2), pp. 354-376.

Farhi, E., and Maggiori, M., 2018, "A Model of the International Monetary System", *Quarterly Journal of Economics,* Vol. 133(1), pp. 295-355.

Fielding, D., and Rewilak, J., 2015, "Credit booms, financial fragility and banking crises", *Economics Letters,* Vol. 136, pp. 233-236.

Filippopoulou, C., Galariotis, E., and Spyrou, S., 2020, "An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach", *Journal of Economic Behavior and Organization,* Vol. 172, pp. 344-363.

Fouliard, J., Howell, M., and Rey, H., 2020, "Answering the Queen: Machine learning and financial crises", NBER working paper no. 28302.

Freund, Y., and Schapire, R. E., 1996, "Experiments with a New Boosting Algorithm", In: Saitta, L. (Ed.) *Machine Learning: Proceedings of the Thirteenth International Conference (ICML '96),* Bari, Italy, 3-6 July, Morgan Kaufmann Publishers Inc., San Francisco, CA, pp. 148-156.

Gourinchas, P.-O., Obstfeld, M., 2012, "Stories of the twentieth century for the twenty-first", *American. Economic. Journal: Macroeconomics,* Vol. 4(1), pp. 226-265.

Hardy, D.C., and Pazarbasioglu, C., 1999, "Determinants and Leading Indicators of Banking Crises: Further Evidence", *IMF Staff Papers*, Vol. 46(3), pp. 247-258.

Hellwig, K.-P. (2021), "Predicting Fiscal Crises: A Machine Learning Approach", IMF WP 150.

Holopainen, M., and Sarlin, P., 2017, "Toward robust early-warning models: A horse race, ensembles and model uncertainty", *Quantitative Finance,* Vol. 17(12), pp. 1933-1963.

Iacoviello, M. and Navarro, G., 2019, "Foreign effects of higher U.S. interest rates", *Journal of International Money and Finance*, Vol. 95, pp. 232-250.

Johnson, Kathleen W., Li, Geng, 2010, "The debt-payment-to-income ratio as an indicator of borrowing constraints: evidence from two household surveys", *Journal of Money Credit Banking*, Vol. 42(7), pp. 1373-1390.

Jordà, Ò., Schularick, M., Taylor, A.M., 2011, "Financial crises, credit booms, and external imbalances: 140 years of lessons", *IMF Economic Review*, Vol. 59(2), pp. 340-378.

Jordà Ò., Richter, B., Schularick, M., and Taylor, A.M., 2015, "Leveraged Bubbles", *Journal of Monetary Economics*, Vol. 76, pp. 1-20.

Jordà Ò., Richter, B., Schularick, M., and Taylor, A.M., 2017, "Macrofinancial History and the New Business Cycle Facts", *NBER Macroeconomics Annual 2016*, Vol. 31, pp. 213-263.

Kaminsky, G.L., 1999, "Currency and banking crises: The early warnings of distress", IMF working paper, No. 99/178.

Kaminsky, G.L. and Reinhart, C.M., 1999, "The twin crises: The causes of banking and balance-of-payments problems", *The American Economic Review*, Vol. 89(3), pp. 473-500.

Laeven, L., and Valencia, F., 2008, "Systemic Banking Crises: A New Database", IMF working paper, No. 08/224.

Laeven, L., and Valencia, F., 2013, "Systemic Banking Crises Database", *IMF Economic Review*, Vol. 61(2), pp. 225-270.

Laeven, L., and Valencia, F., 2018, "Systemic Banking Crises Revisited", IMF working paper, No. 18/206.

Lundberg, S., M. and Lee, S.-L., 2017, "A unified approach to interpreting model predictions", *Advances in Neural Information Processing Systems*, pp. 4765-4774.

Maggiori, M., Neiman, B., and Schreger, J., 2019, "The Rise of the Dollar and Fall of the Euro as International Currencies", *AEA Papers and Proceedings*, Vol. 109.

Manasse, P., and Roubini, N., 2009, "Rules of Thumb for sovereign debt crises", *Journal of International Economics*, Vol. 78, pp. 192-205.

Mian, A., and Sufi, A., 2011, "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis", *American Economic Review*, Vol. 101(5), pp. 2132-56.

Miranda-Agrippino, S., Rey, H., 2019, "US Monetary Policy and the Global Financial Cycle", NBER working paper 21722.

Mullainathan, S., and Spiess, J., 2017, "Machine learning: An applied econometric approach", *Journal of Economic Perspectives,* Vol. 31(2), pp. 87-106.

Ravn, M. O., and Uhlig, H., 2002, "On adjusting the Hodrick-Prescott filter for the frequency of observations", *The Review of Economics and Statistics,* Vol. 84(2), pp. 371-380.

Reinhart, C.M., and Rogoff, K., 2009*, This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press.

Reinhart, C.M., and Rogoff, K., 2011, "From Financial Crash to Debt Crisis", *American Economic Review*, Vol. 101, pp. 1676-1706.

Richter, B., Moritz Schularick M., and Wachtel, P., 2021, "When to Lean against the Wind", Journal of Money, Credit and Banking, Vol. 53(1), pp. 5-30.

Rey, H., 2018, "Dilemma not Trilemma: The Global Financial Cycle and Monetary Policy Independence", NBER working paper 21162.

Sarlin, P., 2013, "On policymakers' loss functions and the evaluation of early warning systems", *Economics Letters*, Vol. 119(1), pp. 1-7.

Sarlin, P., and von Schweinitz, G., 2021, "Optimizing policymakers' loss functions in crisis prediction: before, within or after?", *Macroeconomic Dynamics,* Vol. 25, pp. 100-123.

Schularick, M., and Taylor, A.M., 2012, "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008", *American Economic Review*, Vol. 102(2), pp. 1029-1061.

Shapley, L. S., 1953, "A value for n-person games", *Contributions to the Theory of Games*, Vol. 2(28), pp. 307-317.

Strumbelj, E., and Kononenko, I., 2010, "An efficient explanation of individual classifications using game theory", *Journal of Machine Learning Research*, Vol. 11, pp. 1-18.

Suss, J., and Treitel, H., 2019, "Predicting bank distress in the UK with machine learning", Bank of England, Staff Working Paper 831.

Young, P., 1985, "Monotonic solutions of cooperative games", *International Journal of Game Theory*, Vol. 14, pp. 65-72.

TABLES

Table 1 Definitions of the target variable

	Target variable = 1	Target variable = 0
Approach 1	crisis	Normal times + pre-crisis + post-crisis
Approach 2		
(a)	pre-crisis	Normal times + post-crisis
(b)	pre-crisis	Normal times

Note: In definition 2a, crisis episodes are dropped from the dataset. In definition 2b, crisis and post-crisis episodes are dropped from the dataset.

Table 2 Number of crises/pre-crises and percentage composition

	Advanced economies			Emerging economies		
	1	0	Total	1	0	Total
Approach 1	55	1529	1584	87	3129	3216
	3.5%	*96.5%*	*100.0%*	*2.7%*	*97.3%*	*100.0%*
Approach 2(a)	150	1379	1529	224	2905	3129
	9.8%	*90.2%*	*100.0%*	*7.2%*	*92.8%*	*100.0%*
Approach 2(b)	150	1216	1366	224	2653	2877
	11.0%	*89.0%*	*100.0%*	*7.8%*	*92.2%*	*100.0%*

Source: Authors' own elaborations.

Table 3 Summary statistics (mean values)

	Overall		**Normal times**		**Pre-crisis**		**Crisis**		**Post-crisis**		**No. obs.**
	Advanced	Emerging	Advanced	Emerging	Advanced	Emerging	Advanced	Emerging	Advanced	Emerging	
Country specific variables											
Current account to GDP	-0.02	-1.53	0.41	-1.44	-1.69	-2.18	-2.08	-2.44	-0.58	-1.42	3953
Credit to GDP	124.41	39.42	120.92	39.47	128.58	38.50	136.64	42.83	140.24	38.49	3955
External debt to GNI	164.29	45.55	149.21	43.20	202.94	48.56	205.20	53.31	213.81	60.46	3971
Public debt to GDP	53.51	48.41	53.71	47.44	47.46	48.07	49.56	51.25	58.73	56.35	4052
Inflation	4.91	34.04	5.04	20.22	5.64	88.67	5.25	121.85	3.35	82.13	3971
Real effective exchange rate*	0.35	0.64	0.45	0.79	0.57	1.13	1.47	-0.77	-0.78	-0.36	2285
House price*	2.77	2.09	3.34	2.84	5.71	6.18	-1.78	-0.96	-2.13	-4.89	1823
GDP growth rate	3.15	4.05	3.40	4.50	3.72	1.81	1.94	1.36	1.33	2.75	4159
Global variables											
Energy price index*	8.35	8.35	9.01	9.10	10.58	7.33	9.22	2.31	1.21	3.69	4700
World real GDP growth	3.14	3.14	3.20	3.19	3.39	3.09	2.65	2.66	2.61	2.84	4800
US 10yr Treasury rate	6.48	6.48	6.47	6.26	7.15	7.82	6.61	7.81	5.91	7.20	4800

Source: Authors' own elaborations.
Note: statistics are computed on the original variables, i.e. before data manipulation.
** Expressed as yoy percentage changes.*

Table 4 Estimation sample sizes

	In-sample			Out-of-sample			Overall total
	Advanced	Emerging	Total	Advanced	Emerging	Total	
Normality	568	1120	1688	220	658	878	2566
Pre-crisis	83	159	242	36	10	46	288
Total	651	1279	1930	256	668	924	2854

Source: Authors' own elaborations.

Table 5 In-sample logit estimation results: 1970-2005

	Advanced economies		Emerging economies	
	coefficients	margins	coefficients	margins
Current account to GDP	-0.159	-0.017	-0.180	-0.018
	(0.166)	(0.017)	(0.172)	(0.017)
Credit to GDP	0.427**	0.046**	-2.946	-0.290
	(0.209)	(0.023)	(2.070)	(0.209)
External debt to GNI	-1.139**	-0.122**	-1.141**	-0.112**
	(0.535)	(0.059)	(0.548)	(0.054)
Public debt to GDP	0.416**	0.044**	-0.447***	-0.044***
	(0.186)	(0.021)	(0.157)	(0.015)
World real GDP growth	0.106	0.011	-0.075	-0.007
	(0.085)	(0.009)	(0.067)	(0.007)
US 10yr Treasury rate	0.031	0.003	0.123**	0.012**
	(0.076)	(0.008)	(0.058)	(0.006)
External debt to GNI * US 10yr Treasury rate	0.124*	0.013	0.125*	0.012*
	(0.074)	(0.008)	(0.067)	(0.007)
Inflation	-0.010	-0.001	0.001	0.000
	(0.037)	(0.004)	(0.001)	(0.000)
constant	-2.614***	-	-3.310***	-
	(0.606)		(0.813)	
Country fixed effexts	No		No	
Macroregion fixed effects	No		Yes	
Log-likelihood	-236.6	-236.7	-423.5	-423.6
Pseudo-R2	0.0476	0.0476	0.118	0.118
AUROC	0.6451	0.6451	0.7521	0.7521
Observations	651	651	1279	1279

Source: Authors' own elaborations.
*Note: Standard errors are clustered at country level. Significance levels: * p<0.05; ** p<0.01; ***p<0.001.*

Table 6 Cross validation: estimated hyperparameters

	logit		Adaboost	
	advanced	emerging	advanced	emerging
cutoff	0.112	0.080	0.45	0.45
minimum split	-	-	70	30
number of trees	-	-	30	30
tree depth	-	-	4	4

Source: Authors' own elaborations.

Table 7 In-sample performance indicators

	Advanced economies			Emerging economies		
	25th percentile	median	75th percentile	25th percentile	median	75th percentile
Logit model						
AUROC	0.657	0.726	0.766	0.755	0.777	0.791
Sensitivity	68.354	76.110	84.466	83.871	86.736	90.370
Specificity	39.427	56.514	68.014	40.000	48.796	57.381
Relative usefulness	-0.242	0.062	0.197	0.184	0.238	0.335
AdaBoost						
AUROC	0.851	0.913	0.954	0.711	0.799	0.868
Sensitivity	90.109	94.895	97.321	86.576	88.644	90.909
Specificity	84.632	91.172	92.678	77.145	80.187	81.632
Relative usefulness	0.302	0.543	0.776	0.020	0.234	0.390

Source: Authors' own elaborations.
Note: The interquartile values, i.e. 25th and 75th percentile, are computed by estimating each model 100 times and by taking the values corresponding to the 25% and 75% of the distribution of each indicator.

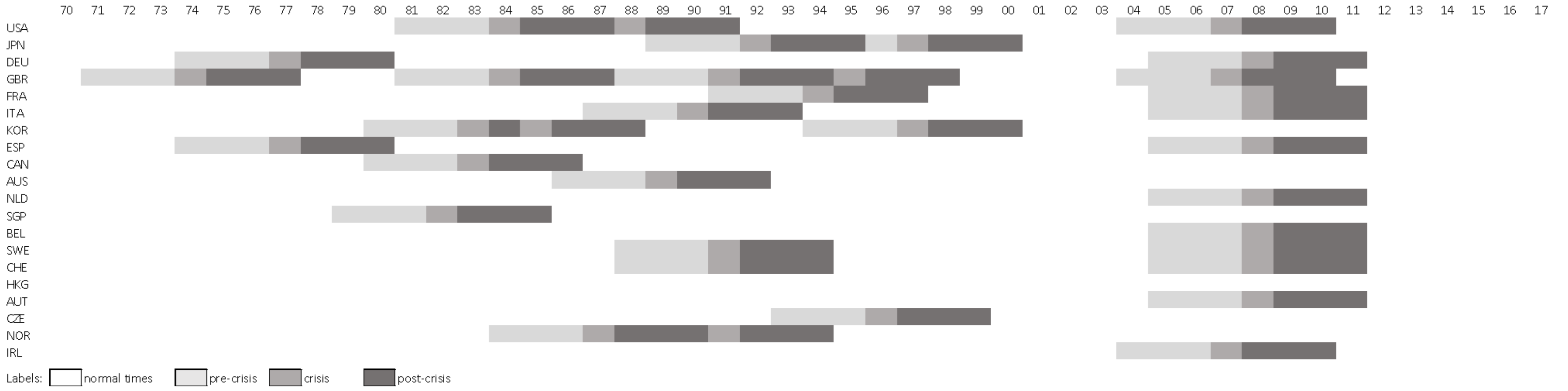
Table 8 Out-of-sample performance indicators

	Advanced economies			Emerging economies		
	25th percentile	median	75th percentile	25th percentile	median	75th percentile
Logit model						
AUROC	0.331	0.481	0.568	0.383	0.636	0.714
Sensitivity	20.329	54.258	66.900	53.198	71.167	77.216
Specificity	38.066	48.741	62.861	24.086	27.202	47.715
Relative usefulness	-1.888	-0.744	-0.303	-0.499	-0.190	-0.034
AdaBoost						
AUROC	0.821	0.874	0.915	0.767	0.822	0.857
Sensitivity	81.633	90.195	94.678	63.125	66.385	73.118
Specificity	86.109	93.674	96.765	84.693	91.099	95.446
Relative usefulness	0.208	0.510	0.765	-0.250	0.212	0.456

Source: Authors' own elaborations.
Note: The interquartile values, i.e. 25th and 75th percentile, are computed by estimating each model 100 times for each year from 2006 to 2017 and by taking the values corresponding to the 25% and 75% of the distribution of each indicator.

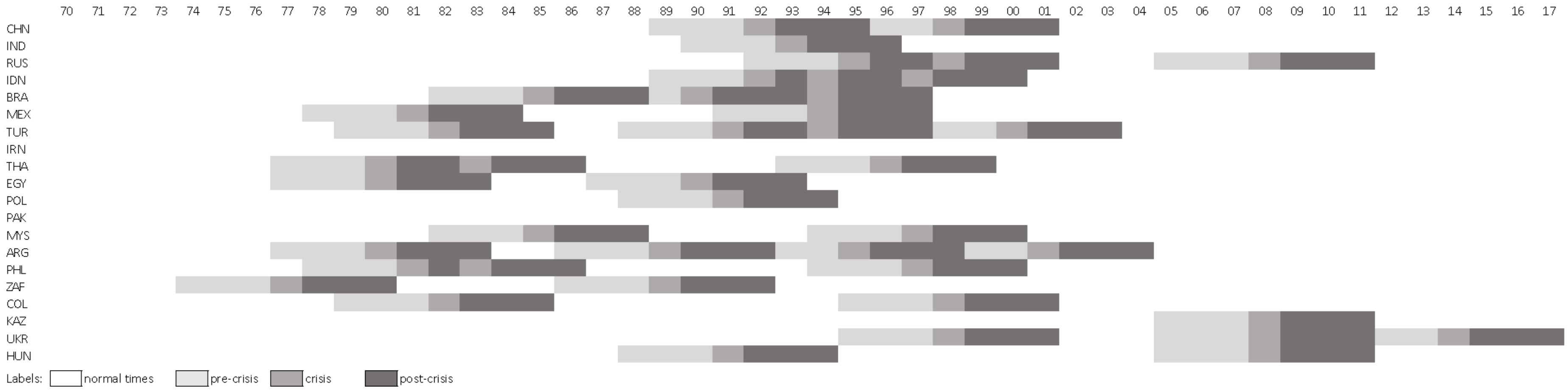
FIGURES

Figure 1 Systemic banking crises in the advanced economies



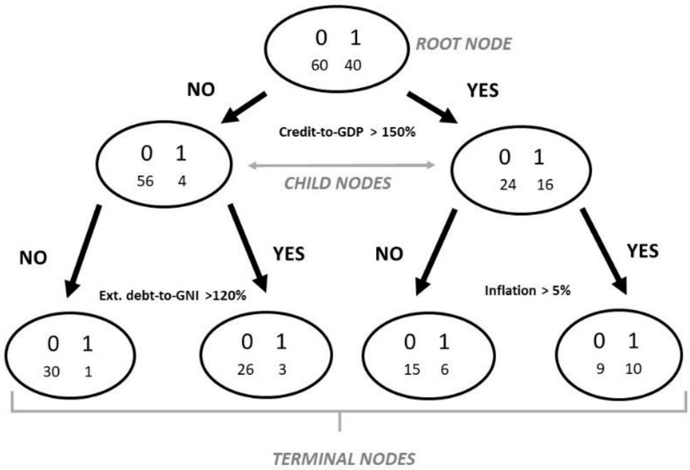
Note: for the sake of brevity and readability of the table, we selected twenty advanced economies based on their contribution to world GDP.

Figure 2 Systemic banking crises in the emerging economies



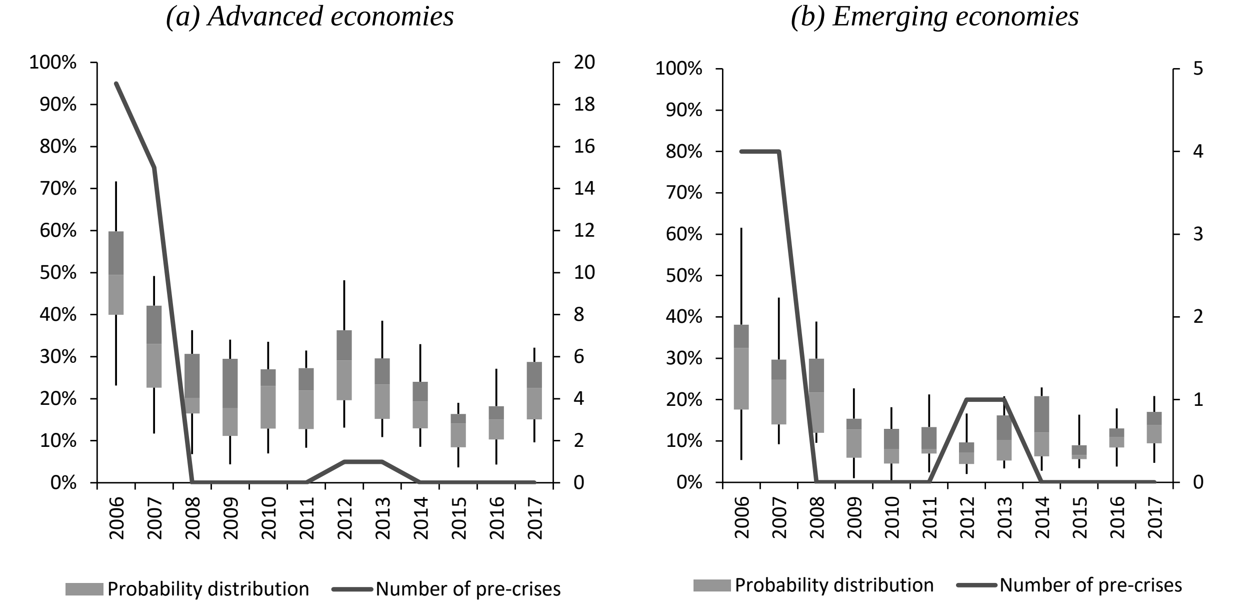
Note: for the sake of brevity and readability of the table, we selected twenty emerging economies based on their contribution to world GDP.

Figure 3 Binary classification tree: an example



Source: Author's own elaborations.

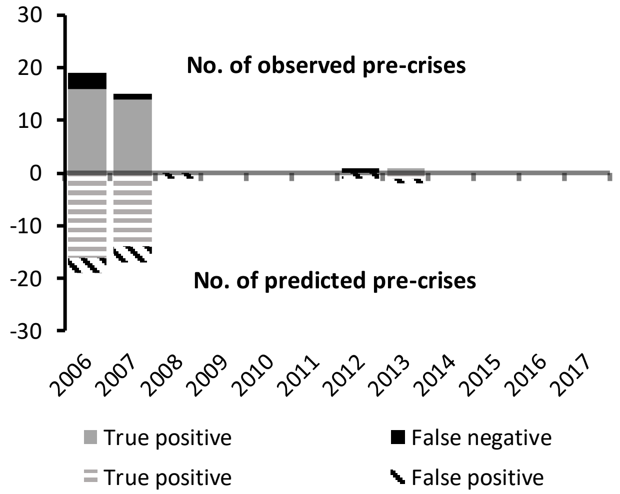
Figure 4 AdaBoost: Box plot of estimated pre-crisis probabilities and true number of pre-crises



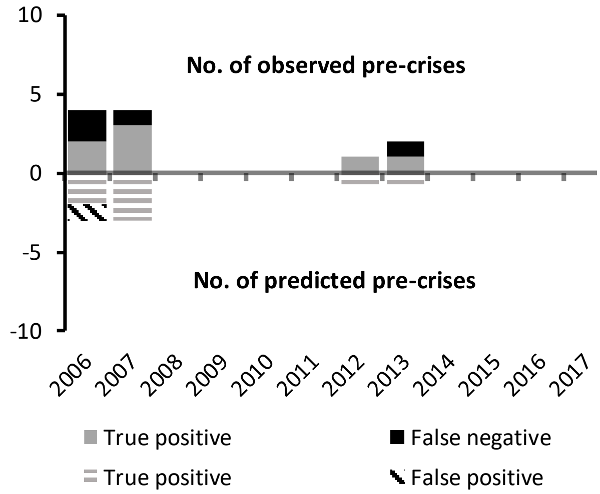
Source: Author's own elaborations.
Note: The lower lines of the box plot represent the first quartile, the upper lines are the fourth quartile while the two central bars correspond to the second and third quartiles, respectively.

Figure 5 AdaBoost: plot of the confusion matrix by year

(a) Advanced economies

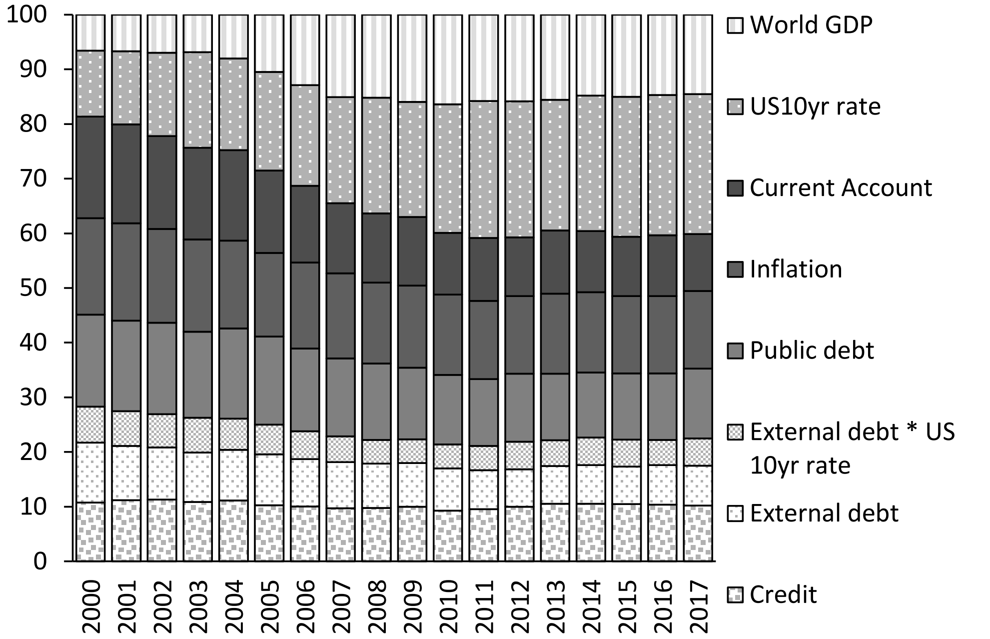


(b) Emerging economies



Source: Author's own elaborations.
Note: The absolute values of the grey and the striped grey bars (i.e. true positives) provide the same information, namely the observed crises that are correctly predicted

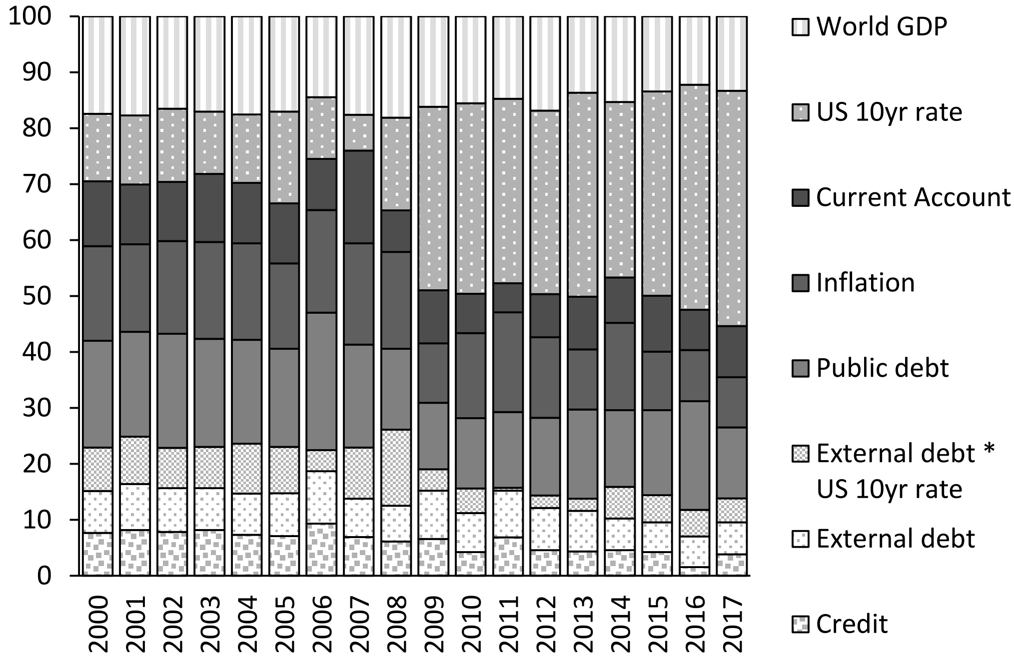
Figure 6 Relative importance by year: Advanced economies



Source: Author's own elaborations.

Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI.

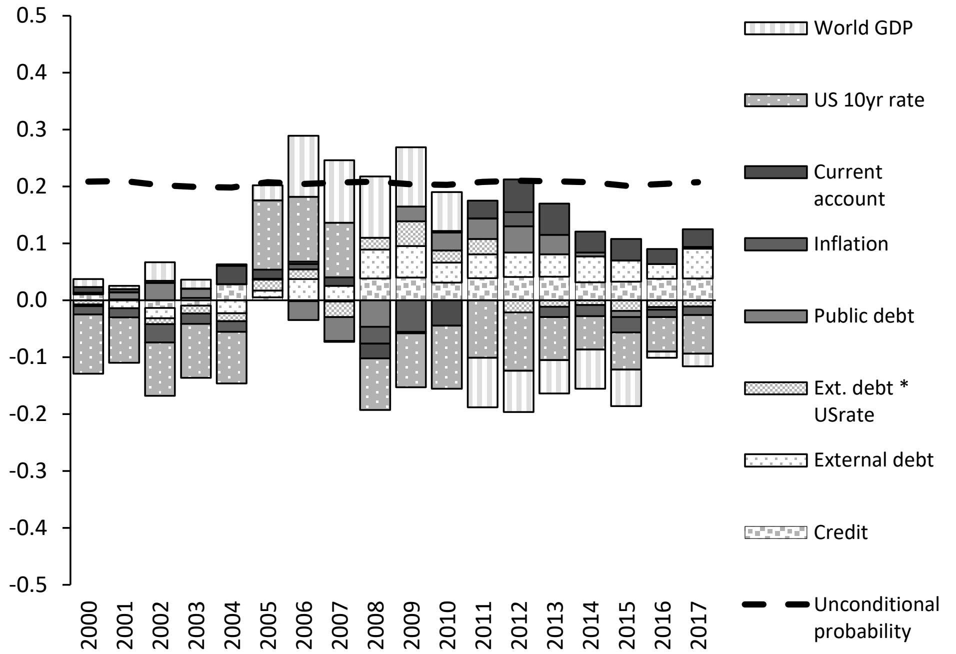
Figure 7 Relative importance by year: Emerging economies



Source: Author's own elaborations.

Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI.

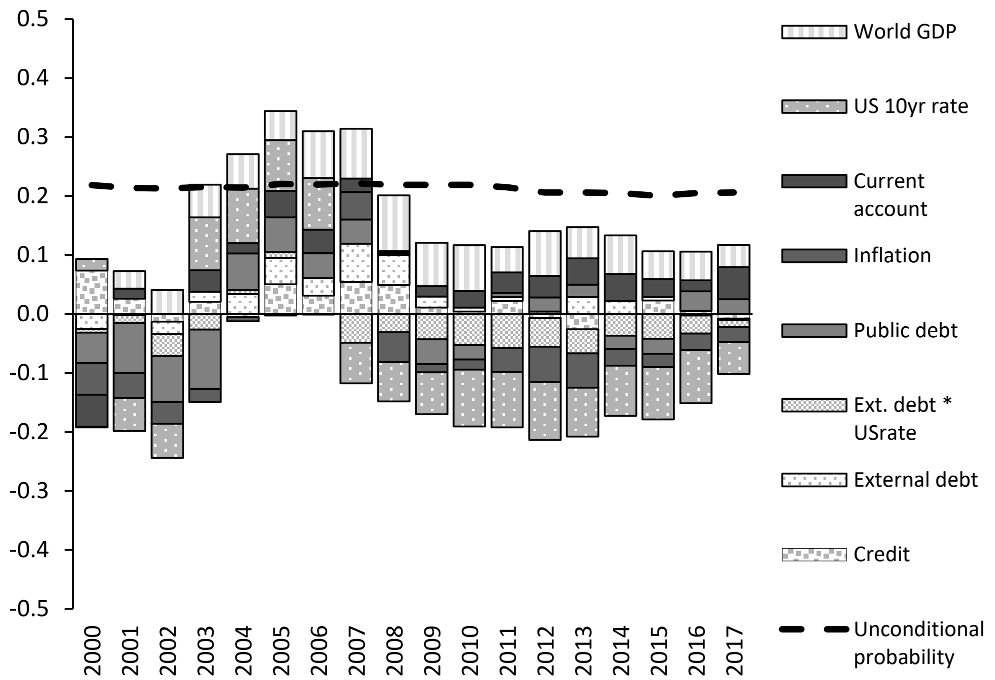
Figure 8 Shapley values: Advanced economies



Source: Author's own elaborations.

Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI. The unconditional probability represents the average probability of a pre-crisis if no variable had been included.

Figure 9 Shapley values: Emerging economies

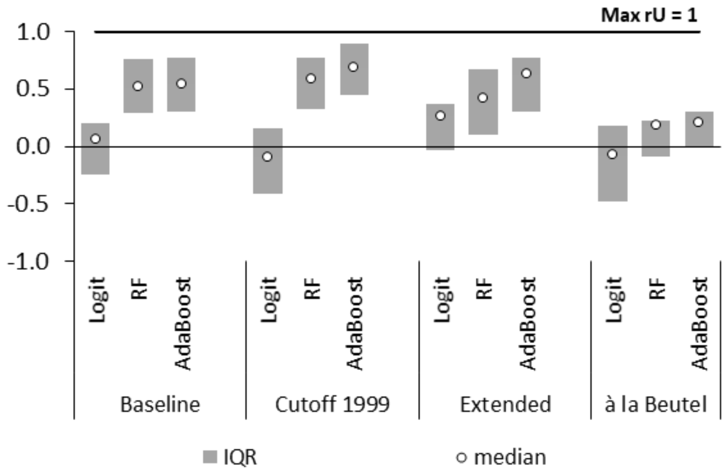


Source: Author's own elaborations.

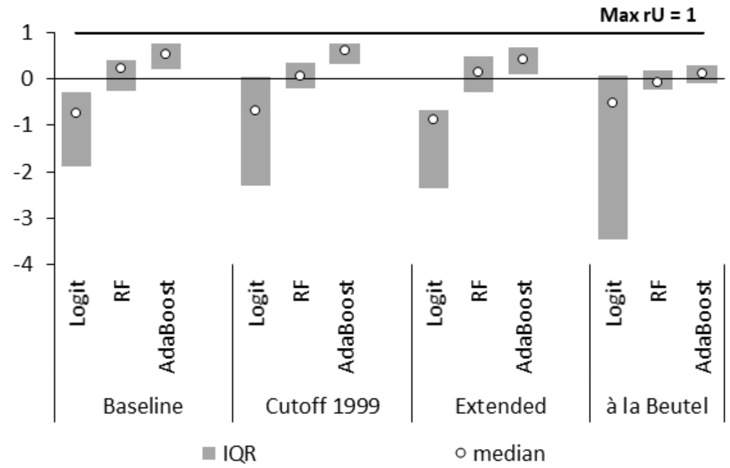
Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI. The unconditional probability represents the average probability of a pre-crisis if no variable had been included

Figure 10 In-sample and out-of-sample relative usefulness: median values and interquartile ranges

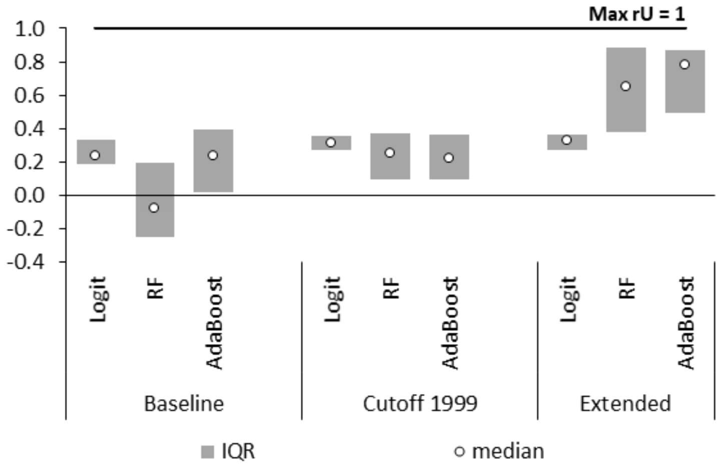
(a) Advanced economies: in-sample



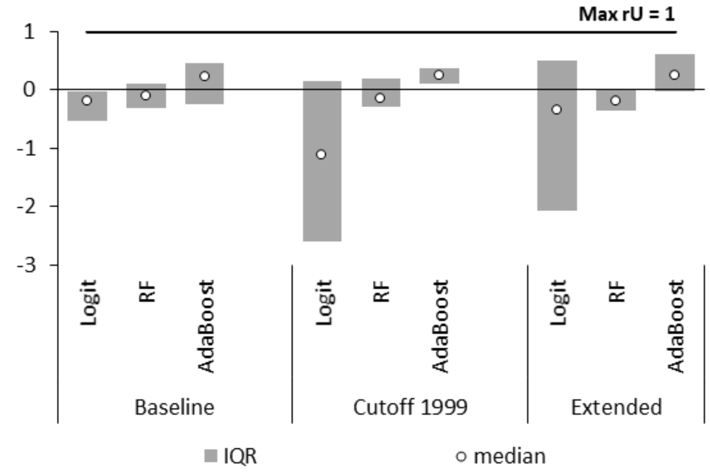
(b) Advanced economies: out-of-sample



(c) Emerging economies: in-sample



(d) Emerging economies: out-of-sample



Source: Authors' own elaborations.
Note: For the in-sample estimates, the interquartile ranges are computed by estimating each model 100 times and by taking the values corresponding to the 25% and 75% of the distribution of each indicator. For the out-of-sample estimates, the interquartile ranges are computed by estimating each model 100 times for each year from 2006 to 2017 and by taking the values corresponding to the 25% and 75% of the distribution of each indicator.

APPENDIX A - Data and definitions

Table A.1 Definitions of banking crises

Author	Definition
Caprio and Klingebiel (1997)	An episode of bank distress is systemic if much or all of the bank capital has been exhausted. Otherwise, it is classified as borderline. To distinguish between systemic banking crises and borderline cases, they also provide detailed information about NPLs, uncollectible loans, bank liquidations, revoked licences, takeover by the public sector and some other relevant variables.
Demirgüc-Kunt and Detragiache (1998, 2005)	An episode of distress is defined as a full-fledged crisis if at least one of the following four conditions holds: (1) the ratio of NPAs to total assets is higher than 10%; (2) the cost of the rescue operation is at least 2% of GDP; (3) a large-scale nationalization of banks has occurred; (4) bank runs take place or government measures (deposit freeze, deposit guarantees) are enacted.
Laeven and Valencia (2008, 2013, 2018)	A banking crisis is systemic if two conditions are met: "(1) Significant signs of financial distress in the banking system (significant bank runs, losses, bank liquidations); (2) Significant banking policy interventions in response to significant losses".
	When losses are severe, the first criterion is sufficient to date a systemic banking crisis. They consider that losses are severe when either (1) the share of NPLs is above 20 percent of total loans or bank closures of at least 20 percent of banking system assets or (2) fiscal restructuring costs of the banking sector are sufficiently high (> 5% of GDP).
	When quantifying the degree of financial distress is problematic or losses are mitigated by policy response, policy interventions are to be significant to date a crisis episode. A policy intervention is significant if at least three out of the following six measures were used: "(1) extensive liquidity support; (2) bank restructuring costs; (3) significant bank nationalizations; (4) significant guarantees; (5) significant asset purchases; (6) deposit freezes and bank holidays".
Reinhart and Rogoff (2009, 2011)	They mark a banking crisis by two types of events: (1) bank runs that lead to the closure, merging, or takeover by the public sector of one or more financial institutions; (2) if there are no runs, the closure, merging, takeovers, or large-scale government assistance of an important financial institution (or group of institutions) that marks the start of a string of similar outcomes for other financial institutions.
Schularick and Taylor (2012), Jordà et al. (2017)	The focus is on the documentary descriptions contained in Bordo et al. (2001) and Reinhart and Rogoff (2009), two widely-used historical data sets that they compare and merge for a consistent definition of event windows. In line with the previous studies, they define a financial crisis when a country's banking sector experiences bank runs or sharp increases in default rates, accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions.
Baron et al. (2019; 2021)	They take the union of all crisis dates as the Joint Crisis List from many sources and uncover new banking crises that are not in existing databases but for which two criteria are satisfied: "(1) there is a decline in the bank equity index of at least 30%, and (2) there is an abundance of narrative evidence consistent with a banking crisis". Then, they remove spurious crises when both of the following criteria are met: (1) bank stock prices do not display a crash of at least 30%, and (2) we cannot find evidence in the historical record that there were either widespread bank failures or bank runs. By adding new crises and removing spurious crises, they create a revised chronology.

Table A.2 List of countries included in the dataset, 1970-2017

Advanced economies	Emerging economies
Australia Austria Belgium Canada Czech Republic Denmark Estonia Finland France Germany Greece Hong Kong SAR Iceland Ireland Israel Italy Japan Korea Latvia Lithuania Luxembourg Netherlands New Zealand Norway Portugal Singapore Slovak Republic Slovenia Spain Sweden Switzerland United Kingdom United States	Albania Algeria Angola Argentina Armenia Azerbaijan Barbados Belarus Belize Bosnia and Herzegovina Botswana Brazil Brunei Bulgaria Chile China Colombia Costa Rica Croatia Dominican Republic Ecuador Egypt El Salvador Equatorial Guinea Fiji Gabon Georgia Guatemala Hungary India Indonesia Iran Jamaica Jordan Kazakhstan Kuwait Lebanon Libya Macedonia Malaysia Mauritius Mexico Morocco Namibia Pakistan Panama Paraguay Peru Philippines Poland Romania Russia Serbia Seychelles South Africa Sri Lanka Suriname Swaziland Syria Thailand Trinidad and Tobago Tunisia Turkey Turkmenistan Ukraine Uruguay Venezuela

Table A.3 Crisis episodes

ADVANCED ECONOMIES								
Country	Year	Source	Country	Year	Source	Country	Year	Source
Australia	1989	RR - JST	Israel	1977	LV	Singapore	1982	RR
Austria	2008	LV - RR	Italy	1990	RR - JST	Slovak Republic	1998	LV
Belgium	2008	LV - RR - JST	Italy	2008	LV - JST	Slovenia	1992	LV
Canada	1983	RR	Japan	1992	RR	Slovenia	2008	LV
Czech Republic	1996	LV	Japan	1997	LV - JST	Spain[3]	1977	LV - RR
Denmark	1987	RR - JST	Korea	1983	RR	Spain	2008	LV - RR - JST
Denmark	2008	LV - RR - JST	Korea	1985	RR	Sweden	1991	LV - RR - JST
Estonia	1992	LV	Korea	1997	LV - RR	Sweden	2008	LV - JST
Finland	1991	LV - RR - JST	Latvia	1995	LV	Switzerland	1991	JST
France	1994	RR	Latvia	2008	LV	Switzerland	2008	LV - RR - JST
France	2008	LV - RR - JST	Lithuania	1995	LV	United Kingdom	1974	RR - JST
Germany	1977	RR	Luxembourg	2008	LV	United Kingdom	1984	RR - JST
Germany	2008	LV - RR - JST	Netherlands	2008	LV - RR - JST	United Kingdom	1991	RR - JST
Greece	1991	RR	New Zealand	1987	RR	United Kingdom	1995	RR
Greece	2008	LV - RR	Norway[2]	1987	RR	United Kingdom	2007	LV - RR - JST
Iceland	1985	RR	Norway	1991	LV	United States	1984	RR - JST
Iceland	1993	RR	Portugal	2008	LV - RR - JST	United States	1988	LV
Iceland[1]	2007	RR	Portugal	2014		United States	2007	LV - RR - JST
Ireland[1]	2007	RR						

Notes: (1) 2008 in LV; (2) 1988 in JST; (3) 1978 in JST.
Legend: LV = Laeven and Valencia (2008, 2013 and 2018); RR = Reinhart and Rogoff (2009); JST = Jordà, Schularick, and Taylor (2017)

Table A.3 Crisis episodes (cont.)

EMERGING ECONOMIES

Country	Year	Source	Country	Year	Source	Country	Year	Source
Albania	1994	LV	Equatorial Guinea	1983	LV	Philippines	1981	RR
Algeria	1990	LV - RR	Georgia	1991	LV	Philippines	1983	LV
Angola	1992	RR	Guatemala	1990	RR	Philippines	1997	LV - RR
Argentina	1980	LV - RR	Guatemala	2001	RR	Poland[4]	1991	RR
Argentina	1989	LV - RR	Guatemala	2006	RR	Romania	1990	LV - RR
Argentina	1995	LV - RR	Hungary	1991	LV - RR	Russia	1995	RR
Argentina	2001	LV - RR	Hungary	2008	LV - RR	Russia	1998	LV - RR
Armenia	1994	LV	India	1993	LV - RR	Russia	2008	LV - RR
Azerbaijan	1995	LV	Indonesia	1992	RR	South Africa	1977	RR
Belarus	1995	LV	Indonesia	1994	RR	South Africa	1989	RR
Bosnia and Herzegovina	1992	LV	Indonesia	1997	LV - RR	Sri Lanka	1989	LV - RR
Brazil	1985	RR	Jamaica	1996	LV	Swaziland	1995	LV
Brazil	1990	LV - RR	Jordan	1989	LV	Thailand	1980	RR
Brazil	1994	LV - RR	Kazakhstan	2008	LV	Thailand	1983	LV
Bulgaria	1996	LV	Kuwait	1982	LV	Thailand[5]	1996	RR
Chile	1976	LV - RR	Lebanon	1990	LV	Tunisia	1991	LV - RR
Chile[1]	1981	LV	Macedonia	1993	LV	Turkey	1982	LV - RR
China	1992	RR	Malaysia	1985	RR	Turkey	1991	RR
China	1998	LV	Malaysia	1997	LV - RR	Turkey	1994	RR
Colombia	1982	LV - RR	Mexico	1981	LV - RR	Turkey	2000	LV - RR
Colombia	1998	LV - RR	Mexico	1994	LV - RR	Ukraine	1998	LV
Costa Rica	1987	LV - RR	Morocco	1980	LV	Ukraine	2008	LV
Costa Rica	1994	LV - RR	Morocco	1983	RR	Ukraine	2014	LV
Croatia	1998	LV	Panama	1988	LV - RR	Uruguay	1971	RR
Dominican Republic	1996	RR	Paraguay	1995	LV - RR	Uruguay	1981	LV - RR
Dominican Republic	2003	LV - RR	Paraguay	2002	RR	Uruguay	2002	LV - RR
Ecuador[2]	1981	RR	Peru	1983	LV - RR	Venezuela	1978	RR
Ecuador	1998	LV - RR	Peru	1999	RR	Venezuela[6]	1993	RR
Egypt[3]	1980	LV						
Egypt	1990	RR						
El Salvador	1989	LV - RR						

Notes: (1) 1982 in RR; (2) 1982 in LV; (3) 1981 in RR, (4) 1992 in LV; (5) 1997 in LV; (6) 1994 in LV.
Legend: LV = Laeven and Valencia (2008, 2013 and 2018); RR = Reinhart and Rogoff (2009); JST = Jordà, Schularick, and Taylor (2017).

Table A.4 Variable description and sources

Variable	Description	Source
Country-specific		
Ratios		
Current Account-to-GDP	Current account balance, % of GDP	WB[(a)]
External Debt-to-GNI	External debt stocks, % of GNI	WB, International Debt Statistics
Public Debt-to-GDP	Gross general governement debt, % of GDP	IMF, Global Debt Database
Credit-to-GDP	Credit to the private sector, % of GDP	BIS (WB when BIS data not available)
yoy % changes		
Inflation	GDP deflator, ratio of GDP in current local currency to GDP in constant local currency	WB[(b)]
Real effective Exchange Rate[(a)]	Real effective Exchange Rate based on CPI Index	WB[(c)]
Real GDP growth	Annual percentage growth rate of GDP at constant prices	WB[(b)]
House price	Real house price index (yoy %)	BIS, FRED, OECD, Cesa-Bianchi (2013)
Global		
10y US Treasury Rate	10-Year Treasury Constant Maturity Rate	Federal Reserve Dallas
yoy % changes		
Energy Price Index	Average weighted prices of energy raw materials (weight = 4.7), crude oil (weight = 84.6) and natural gas (weight = 10.8)	World Bank Commodity Price Data
Real World GDP growth	Annual percentage growth rate of world GDP at constant prices	WB[(b)]

Source: Authors' own elaborations based on BIS, CBOE, IMF and WB.
Notes: (a) Only included in extended model for the sample of advanced countries; (a) WB cites as source "International Monetary Fund, Balance of Payments Statistics Yearbook and data files, and World Bank and OECD GDP estimates"; (b) WB cites as source "National accounts data and OECD National Accounts"; (c) WB cites "International Monetary Fund, International Financial Statistics".

Table A.5 T-test on the differences in the means: advanced vs emerging countries

	Difference	t-stat
Current account to GDP	1.50 ***	4.34
Credit to GDP	84.99 ***	30.39
External debt to GNI	118.74 ***	11.76
Public debt to GDP	5.10 ***	4.23
Inflation	-29.13 ***	-6.43
GDP growth rate	-0.89 ***	-4.27

Source: Authors' own elaborations based on BIS, CBOE, IMF and WB.
* $p<0.05$; ** $p<0.01$; *** $p<0.001$.

APPENDIX B – Additional robustness checks

Table B.1 displays the full list of variables in all the models detailed in the robustness analysis of Section 6. **Table B.2** displays the marginal effects of the in-sample logit model applied to the specifications detailed in Section 6 and those of the baseline model of **Table 5**. The Table shows the very poor statistical performance of the "cut-off 1999" model, especially for advanced economies. Compared to the "baseline" model, in the model "à la Beutel" credit to GDP loses its role and the only statistically significant variable is external debt to GNI (alone and in interaction with the US 10yr Treasury rate). The results of the "extended" model for advanced economies show that, on the one hand, the credit to GDP ratio and external debt to GNI lose their statistical significance, and that, on the other hand, the newly introduced variables play a significant role as leading indicators of banking crises. In the model for emerging economies, previously significant variables maintain their role and country-specific GDP growth and energy prices show a statistically significant weight.

Figure B.1 displays variables' relative importance of the "extended" model. For advanced economies, we observe a change in the ranking of the variables contributing to the build-up of crises, i.e. the most relevant variables become the real effective exchange rate, inflation, global energy index, credit to GDP and – with fluctuations throughout time – the US 10yr Treasury rate. For emerging economies, the most relevant role of the US 10yr Treasury rate is confirmed, followed by current account to GDP, inflation and public debt to GDP, which exhibit a similar relative importance.

Finally, **Figure B.2** displays Shapley values. For advanced economies, the most relevant effect is that of the US 10yr Treasury rate. As for the other variables, it is worth noting the role of house prices in the second half of the 2000s. For emerging economies, again the US 10yr Treasury rate contributes the most, with a negative sign to the probability of a pre-crisis after the 2008 crisis. As for the other variables, we identify the positive sign of current account to GDP and credit to GDP from 2011.

Table B.1 Variables used in each model specification

Sample	**Model**			
	Main text	**Robustness analysis**		
	Baseline 2005	*Cutoff 1999*	*Extended*	*à-la-Beutel*
Advanced Economies	Curent account-to-GDP	Curent account-to-GDP	Curent account-to-GDP	Curent account-to-GDP
	Credit-to-GDP	Credit-to-GDP	Credit-to-GDP	Credit-to-GDP
	External debt-to-GNI	External debt-to-GNI	External debt-to-GNI	External debt-to-GNI
	Public debt-to-GDP	Public debt-to-GDP	Public debt-to-GDP	Public debt-to-GDP
	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate
	World GDP growth rate	World GDP growth rate	World GDP growth rate	World GDP growth rate
	US 10y Treasury rate	US 10y Treasury rate	US 10y Treasury rate	US 10y Treasury rate
			House prices	
			Real effective exchange rage	
			GDP growth rate	
			Energy prices	
Emerging Economies	Curent account-to-GDP	Curent account-to-GDP	Curent account-to-GDP	Curent account-to-GDP
	Credit-to-GDP	Credit-to-GDP	Credit-to-GDP	Credit-to-GDP
	External debt-to-GNI	External debt-to-GNI	External debt-to-GNI	External debt-to-GNI
	Public debt-to-GDP	Public debt-to-GDP	Public debt-to-GDP	Public debt-to-GDP
	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate	External debt -to-GNI X US 10y Treasury rate
	World GDP growth rate	World GDP growth rate	World GDP growth rate	World GDP growth rate
	US 10y Treasury rate	US 10y Treasury rate	US 10y Treasury rate	US 10y Treasury rate
			GDP growth rate	
			Energy prices	

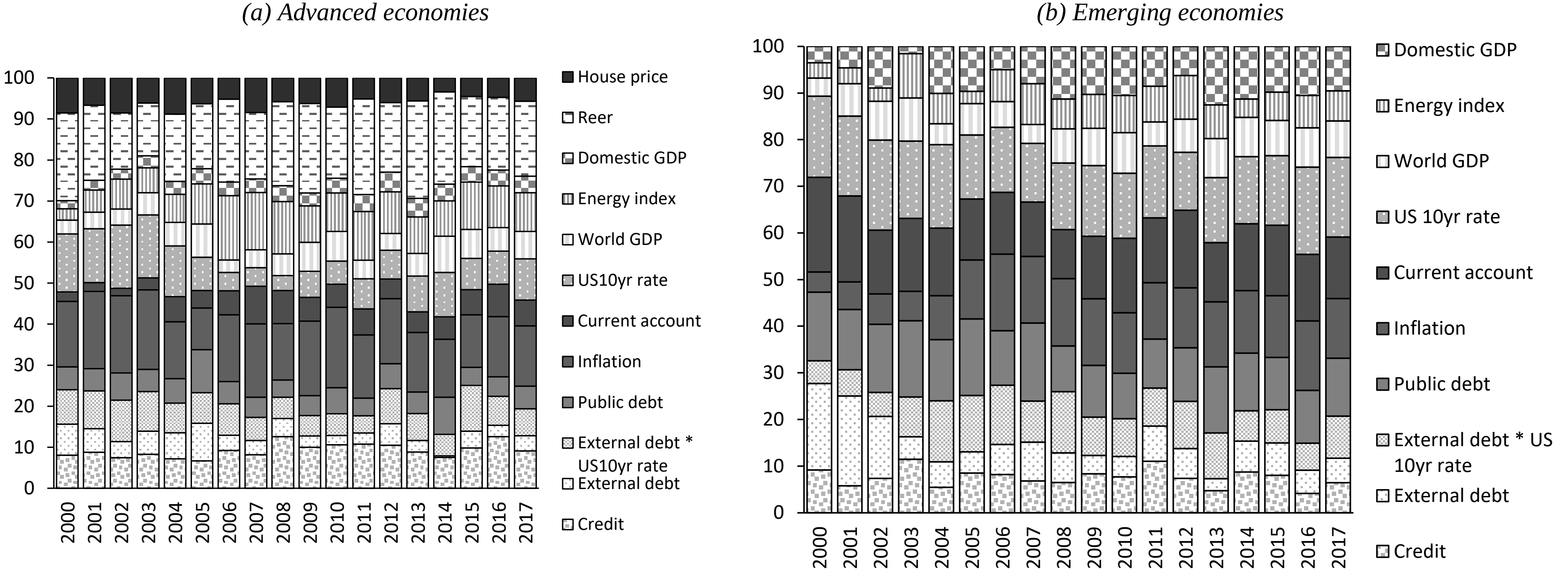
Table B.2 Logit model: in-sample estimated marginal effects

	Baseline 2005		"Cutoff 1999"		"Extended"		"à la Beutel"
	Advanced	Emerging	Advanced	Emerging	Advanced	Emerging	
Current account to GDP	-0.017	-0.018	-0.00202	-0.0118	-0.0109	-0.0180	-0.025
	(0.017)	(0.017)	(0.0216)	(0.0210)	(0.0176)	(0.0176)	(0.023)
Credit to GDP	0.046**	-0.290	0.0458	-0.388	0.0144	-0.334	0.021
	(0.023)	(0.209)	(0.0298)	(0.253)	(0.0229)	(0.207)	(0.027)
External debt to GNI	-0.122**	-0.112**	-0.0691	-0.00981	-0.0884	-0.122**	-0.159**
	(0.059)	(0.054)	(0.0919)	(0.0716)	(0.0639)	(0.0564)	(0.068)
Public debt to GDP	0.044**	-0.044***	0.0362	-0.0456**	0.0365*	-0.0494***	0.030
	(0.021)	(0.015)	(0.0276)	(0.0191)	(0.0210)	(0.0150)	(0.026)
World real GDP growth	0.011	-0.007	0.00138	-0.0140*	0.0178	-0.00631	0.013
	(0.009)	(0.007)	(0.0110)	(0.00823)	(0.0132)	(0.00747)	(0.013)
US 10yr Treasury rate	0.003	0.012**	0.0102	0.00216	0.00889	0.0145**	-0.000
	(0.008)	(0.006)	(0.0109)	(0.00703)	(0.00781)	(0.00585)	(0.011)
External debt to GNI * US 10yr Treasury rate	0.013	0.012*	0.00788	0.000679	0.00975	0.0134**	0.021**
	(0.008)	(0.007)	(0.0108)	(0.00908)	(0.00924)	(0.00666)	(0.009)
Inflation	-0.001	0.000	-0.000809	0.0000475	-0.00713	-0.0000196	-0.006
	(0.004)	(0.000)	(0.00409)	(0.0000657)	(0.00518)	(0.0000720)	(0.007)
Real GDP growth	-	-	-	-	0.0129**	-0.00695***	-
					(0.00620)	(0.00264)	
Energy price index	-	-	-	-	0.00271**	-0.00123*	-
					(0.00115)	(0.000707)	
Real effective exchange rate	-	-	-	-	0.00447***	-	-
					(0.00118)		
Real house price index	-	-	-	-	0.00312**	-	-
					(0.00148)		
Country fixed effexts	No	No	No	No	No	No	No
Macroregion fixed effects	No	Yes	No	Yes	No	Yes	No
Log-likelihood	-236.7	-423.6	-174.9	-366.2	-170.8	-414.2	-155.1
Pseudo-R2	0.0476	0.118	0.0393	0.1019	0.1645	0.1374	0.0541
AUROC	0.6451	0.7521	0.6254	0.7297	0.7740	0.771	0.6642
Observations	651	1279	489	961	526	1279	441

Source: Authors' own elaborations

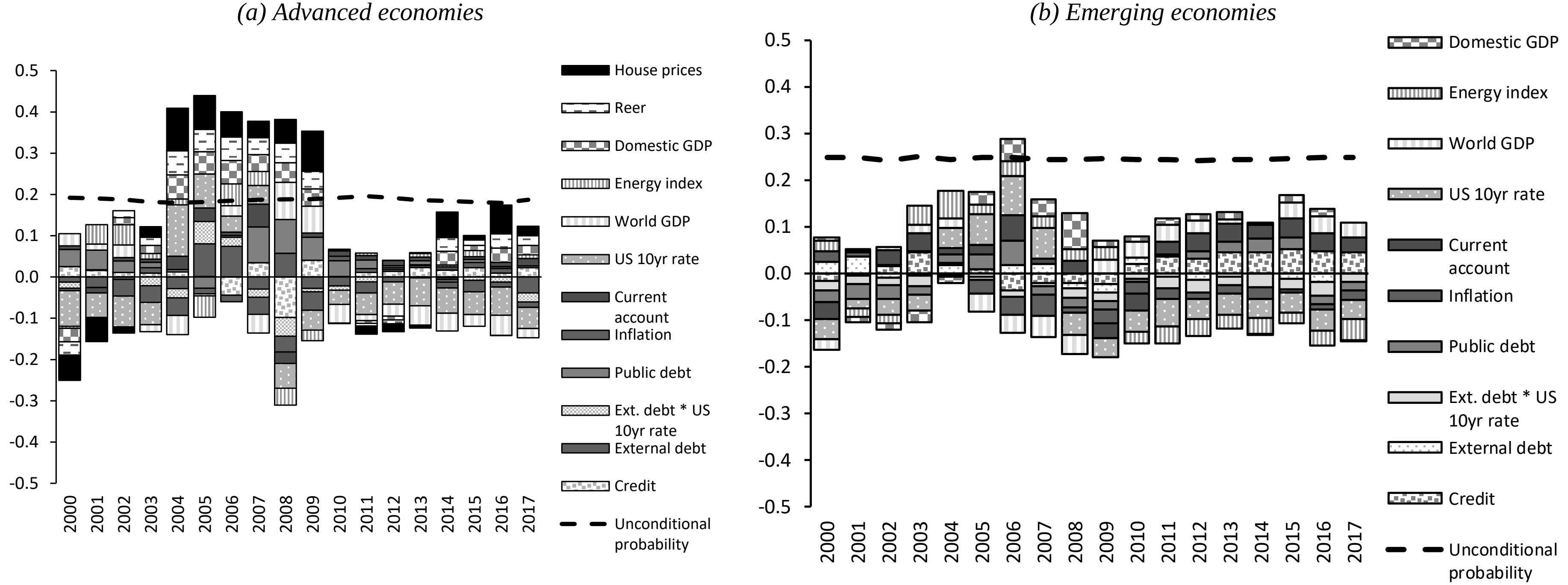
*Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Figure B.1 Extended model: Relative importance by year



Source: Author's own elaborations.
Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI.

Figure B.2 Extended model: Shapley values



Source: Author's own elaborations.

Note: credit, current account and public debt are expressed as ratios to GDP, while external debt as ratio to GNI. The unconditional probability represents the average probability of a pre-crisis if no variable had been included