Elizabeth Jane Casabianca, Prometeia Associazione per le Previsioni Econometriche, DiSeS (Polytechnic University of Marche)

Michele Catalano, Prometeia Associazione per le Previsioni Econometriche

Lorenzo Forni, Prometeia Associazione per le Previsioni Econometriche, DSEA (University of Padua)

Elena Giarda, Prometeia Associazione per le Previsioni Econometriche, Cefin (University of Modena and Reggio Emilia)

Simone Passeri, Prometeia Associazione per le Previsioni Econometriche

AN EARLY WARNING SYSTEM FOR BANKING CRISES: FROM REGRESSION-BASED ANALYSIS TO MACHINE LEARNING TECHNIQUES

August 2019

Marco Fanno Working Papers - 235

An Early Warning System for banking crises:
From regression-based analysis to machine learning techniques

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

[a] Prometeia Associazione per le Previsioni Econometriche;
[b] DiSeS (Polytechnic University of Marche);
[c] DSEA (University of Padua); [d] Cefin (University of Modena and Reggio Emilia)

This version: July 2019

Abstract

Ten years after the outbreak of the 2007-2008 crisis, renewed attention is directed to money and credit fluctuations, financial crises and policy responses. By using an integrated dataset that includes 100 countries (advanced and emerging) spanning from 1970 to 2017, we propose an Early Warning System (EWS) to predict the build-up of systemic banking crises. The paper aims at (i) identifying the macroeconomic drivers of banking crises, (ii) going beyond the use of traditional discrete choice models by applying supervised machine learning (ML) and (iii) assessing the degree of countries' exposure to systemic risks by means of predicted probabilities. Our results show that ML algorithms can have a better predictive performance than the logit models. All models deliver increasing predicted probabilities in the last years of the sample for the advanced countries, warning against the possible build-up of pre-crisis macroeconomic imbalances.

Keywords: banking crises; EWS; machine learning; decision trees; AdaBoost.

JEL classification: C40; G01; C25; E44; G21.

Note: We are grateful to seminar participants at Prometeia (Bologna, 7 November 2018 and 28 February 2019) and University of Bologna (Rimini, 9 April 2019) as well as participants at "BigNOMICS Workshop on Big Data and Economic Forecasting" (Ispra, 16-17 May 2019) and "16th EUROFRAME Conference" (Dublin, 7 June 2019) for valuable comments. All remaining errors are our responsibility. The opinions expressed in this article are our own and do not necessarily reflect those of Prometeia Associazione.

* Corresponding author: Lorenzo Forni, University of Padua, Department of Economics and Management, Via del Santo 33, 35123 Padova (lorenzo.forni@unipd.it) and Prometeia Associazione, Via G. Marconi 43, 40122 Bologna (lorenzo.forni@prometeia.com).

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 developing countries' resilience to economic shocks. The ultimate goal is to design macroprudential policies addressing such vulnerabilities and limit them from building up further and spreading across the economic system.

Against this background, economists have developed Early Warning Systems (henceforth, EWSs) aimed at detecting the risks that a systemic banking crisis may arise. This literature 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 one of the first 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 and were 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, and Schularick and Taylor, 2012). More recently, machine learning methods have been employed by economists to improve the predictive performance of EWSs. Duttagupta and Cashin (2011) and Alessi and Detken (2018) implement these modelling techniques to analyse banking crises. Another example in this direction is Manasse and Roubini (2009) on sovereign debt crises.

This empirical literature stems from – and partially overlaps with – a wide field of research aimed at identifying banking crisis episodes, according to a variety of criteria. Demirgüc-Kunt and Detragiache (1998) build on the early attempts in the literature (e.g. Lindgren et al., 1996; Caprio and Klingebiel, 1997) and identify banking crisis episodes based on the occurrence of a number of disruptive events related to the banking sector. More recently, Laeven and Valencia (2018), evolving from their previous work, put forth a more sophisticated definition of systemic banking crises.

With this paper, we contribute to the literature by developing an EWS for advanced and emerging economies. Our goal is threefold. First, to identify macroeconomic indicators that could contain valuable information to uncover vulnerabilities leading to a banking crisis should an economic shock occurs. Second, to propose a EWS by using both a modelling technique taken from traditional econometrics, namely the logit model, and from machine learning, namely Adaptive Boosting (AdaBoost) and compare their performance. Third, to assess the degree of countries' exposure to systemic risks by means of predicted probabilities.

For these purposes, we collect information on banking crisis episodes from various sources to maximise coverage across both time and countries. The banking crisis dataset is then merged with information on selected macroeconomic indicators. In particular, we select variables that have been suggested to serve as leading indicators of banking crises by similar research. We end up with an integrated dataset that includes 100 countries – 33 advanced and 67 emerging – over the period 1970-2017.

Our work brings a number of novelties to existing research. First, we combine banking crises and macroeconomic information from different sources and we update them according to the latest available data. Second, we put together different lines of research and attempt to shed light on the most meaningful leading indicators of banking crises. Third, we adopt the AdaBoost modelling technique to develop an EWS, which to the best of our knowledge, has never been done so far. By doing so, we overcome some of the limitations of traditional regression analysis, especially its predictive performance, while still retaining some of its advantages, namely ease of use and interpretation.

Our results are promising. Using a baseline set of 8 macroeconomic indicators, we show that the AdaBoost performs better than the logit model. Nevertheless, both models deliver increasing predicted probabilities in the last years of the sample, warning against the possible build-up of pre-crisis macroeconomic imbalances. Having established that the AdaBoost is a better classifier, we further test its predictive performance on an enlarged version of our baseline variable set.

The remainder of the article is organised as follows. Section 2 reviews the literature on the definition of systemic banking crises and on the empirical analyses that aim at predicting them by means of EWSs. Section 3 provides a discussion on how to build an appropriate binary variable employable as target variable in the empirical applications. In Section 4, we show some stylized facts on banking crises and macroeconomic contexts. In Section 5, we estimate an EWS by means of logit models. Section 6 introduces the main features of supervised machine learning methods. In Section 7, we implement an EWS by applying a supervised ML algorithm, i.e. Adaptive Boosting (AdaBoost) and compare its predictive performance with that of the logit model. We further develop the AdaBoost to include additional macroeconomic indicators to the set of explanatory variables used for its estimation. Section 8 concludes with a summary of the main findings.

2. Review of the literature

The widespread losses of the 2007-2008 global financial crisis in the advanced economies brought to the forefront the need for an effective Early Warning System (EWS) to help governments and international financial institutions act promptly to prevent risks of possible future bank runs and bank failures from turning into a systemic banking crisis.

The literature on how to identify, explain and predict "crises" has a long-lasting tradition. In the last two decades – and with renewed attention in the last one – the focus has shifted from balance of payments and currency crises to systemic banking crises. The definition of banking crises is not straightforward and economists provide different criteria to identify their occurrence (Section 2.1). The literature also provides ways to link macroeconomic imbalances with crisis episodes, to explore their role as leading indicators and to assess the ability of econometric models to predict banking crises or the risks that a crisis may occur (Section 2.2).

2.1 The definition of systemic banking crises

There is a wealth of definitions of banking crises. Baron et al. (2018) suggest a classification of the approaches to identify them: (i) the "policy-based" approach (Caprio and Klingebiel, 1997 and 2003; Demirgüc-Kunt and Detragiache, 1998 and 2005; Laeven and Valencia, 2008, 2013 and 2018) and (ii) the "narrative-based" approach (Bordo et al., 2001; Reinhart and Rogoff, 2009 and 2011; Schularick and Taylor, 2012; Jordà et al., 2017a).

Demirgüc-Kunt and Detragiache (1998) observe that the most dated literature (among others, see Lindgren et al., 1996; Caprio and Klingebiel, 1997) provides an overview of banking sector fragility, but it does not always distinguish either financial distress from banking crises or local crises from systemic crises.[1] They take from these studies to build a new framework to classify an episode of distress as a systemic banking crisis. This is based on four conditions: excessive level of nonperforming assets (NPAs) to total assets, substantial rescue operations, large-scale nationalization of banks, and finally bank runs and deposit freezes (see **Table A1** for details). If at least one of these events occurs, they define the episode of distress as a systemic banking crisis. They apply their definition to 29 countries for the years 1980-1994, identifying 31 crisis episodes.

More recently, Laeven and Valencia (2008) and subsequent updates (Laeven and Valencia, 2013 and 2018) put forth a more articulated definition of systemic banking crises, adding on what proposed by Demirgüc-Kunt and Detragiache (1998). They identify a banking crisis when losses are severe, i.e. a high level of nonperforming loans (NPLs) to total loans or relevant fiscal restructuring costs. However, if these losses are mitigated by policy intervention or it is difficult to quantify them, they look at whether three out of six measures were implemented (four of them are partially retrieved from Demirgüc-Kunt and Detragiache, 1998) (see **Table A1** for details). If this happens, the episode of distress is defined as a systemic banking crisis. Their most recent database covers 165 countries over the period 1970-2017 and identifies 151 crisis episodes.[2]

The policy-based approach requires richness of data and economic-related information to identify banking crises, causing limited time and country coverage. This prompted a new strand of the literature, the narrative-based approach, which refers to narrative sources of events such as bank runs or policy intervention, to identify banking crisis and fill-in the gaps and extend coverage of the policy-based approach. This approach gives the opportunity to include a number of banking crises backed by a strong historical narrative that are “forgotten” in the policy-based framework (Baron et al., 2018).

Reinhart and Rogoff (2011) provide the first systematic contribution in this direction. They extend preliminary analysis of economic historians such as Bordo et al. (2001) for the pre-World War II period, while take from Caprio and Klingebiel, (1997 and 2003) for the post-1970 period. In their study, a banking crisis is identified when bank runs lead to the closure, merging, or takeover by the public sector of one or more financial institutions. If there are no bank runs, crises are marked by events such as the closure, merging, or takeover by the public sector of an important financial institution that later spread to other financial institutions (see **Table A1** for details). Their database spans from 1800 to 2009, covers 70 advanced and emerging countries and identifies 290 banking crises.[3]

In line with these studies, Schularick and Taylor (2012, p. 1038) define financial crises as “events during which a country’s banking sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions”. In their view, banking crises are credit booms gone bust. Their final dataset is the result of a critical scrutiny and merge of previously compiled datasets (i.e. Bordo et al., 2001; Laeven and Valencia, 2008; Reinhart and Rogoff, 2011) and covers 70 countries for the period 1870-2008 (**Table A1**). Jordà et al. (2017a) update this dataset, extending the analysis to 17 countries up to 2013 (**Table A1**).

According to Baron et al. (2018), both approaches suffer from some shortcomings. The narrative-based one may be biased because it takes account only of the most relevant events, while the policy-based one because the policy intervention response may be endogenous, subjective and not always timely. To overcome what they think is a subjectivity bias they adopt an alternative approach by using

[1] For details on how Caprio and Klingebiel (1997) define banking crises, see **Table A1**.
[2] Their dataset is complemented with 236 currency crises and 74 sovereign debt crises.
[3] They also identify 209 sovereign default episodes.

a “hard” measure such as countries’ bank equity index and by developing a crisis indicator based on the decline in the index to refine the chronology of banking crises (see **Table A1** for details). Their database consists of 113 crisis episodes, for 46 countries over the period 1870-2016.

The resulting variable of all the approaches is a discrete variable. In the simplest case, it is a binary variable (0/1), where the 1s define systemic banking crisis episodes and the 0s all the other periods. In other cases it may take on three values (i.e. 0/1/2), which distinguishes among pre-crisis, crisis and post-crisis. This variable is used as the outcome or target variable in most of the models of the EWS literature.

2.2 The prediction of banking crises by EWSs

Besides monitoring the occurrence of banking crises, building a variable that identifies these episodes is functional to the “estimation” of empirical models – EWSs – aimed at detecting the risks that a systemic banking crisis may arise. The literature on EWSs has evolved along different lines, from the signals approach to discrete choice models and to machine learning techniques.

The signals approach is a non-parametric method, which 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. A variable is considered to signal a crisis if it exceeds a pre-defined threshold.[4] Kaminsky and Reinhart (1999) are the first to apply this approach to balance of payments and banking crises for a number of industrial and developing countries for the period 1970-1995, covering 76 currency crises and 26 banking ones.

However, with the signals approach each indicator is used in isolation and the model does not allow the aggregation of the individual warnings. The simplest solution consists of counting the number of leading indicators signalling distress. Nonetheless, this statistic “may not be the best choice because the economy may be vulnerable, but still many of the indicators may not signal jointly that something is wrong” (Kaminsky, 1999, p. 23). Kaminsky (1999) develops, among others, a composite index that weights the signals of each variable by the inverse of their noise-to-signal-ratio to account for the forecasting accuracy of each variable.[5] Borio and Lowe (2002) and Davis and Karim (2008) take from here and apply this methodology to banking crises, for the time spells 1960-1999 and 1979-2003, respectively.

An alternative methodology that allows the simultaneous study of macroeconomic variables as determinants of banking crises is the logit model, a tool widely used in microeconometrics to estimate the probability of an event. The outcome variable is binary (crisis/non-crisis) and the probability that the event (crisis) occurs is estimated as a function of macroeconomic factors. From the estimated coefficients of the model, it is possible to retrieve the estimated probabilities of the crisis. Demirgüc-Kunt and Detragiache (1998) apply this method to a large sample of developed and developing countries in 1980-1994 and find that the main determinants of a banking crisis are low growth, high inflation and high real interest rates. The literature evolved along these lines, with contributions from, among others, Arteta and Eichengreen (2000), Demirgüc-Kunt and Detragiache (2005), Barrell et al. (2010) and Schularick and Taylor (2012) for a variety of countries and time spells.

---

[4] Thresholds are discretional and based on the distribution of the variable of interest.

[5] In a binary classification problem, the noise-to-signal ratio is defined as the ratio between (i) the ratio of the number of crises incorrectly predicted to all non-crisis episodes (false positive rate) and (ii) the ratio of the number of crises correctly predicted to all crisis episodes (sensitivity or true positive rate). The lower the noise-to-signal ratio associated with a variable, the better the ability of the variable to predict a crisis.

Although these papers employ the same econometric approach and discriminate between the crisis episodes (1) and all other periods (0), they may differ in what they classify as 0s. The non-crisis years are a non-homogeneous informative set since they encompass a mix of time spells with different characteristics – pre-crisis, post-crisis and normal (or tranquil) times.

A widely used method consists of dropping some of the non-crisis years. Demirgüc-Kunt and Detragiache (1998) follow two approaches, one in which they drop all the observations following the first crisis episode experienced by a country and one in which they exclude the post-crisis years. Demirgüc-Kunt and Detragiache (2005) apply the latter criterion. Arteta and Eichengreen (2000) drop the three years before and after the crisis and therefore the 0s denote tranquil times only. Conversely, Barrell et al. (2010) and Schularick and Taylor (2012) make no distinction among pre-, post-crisis and normal times and thus they use the 0s as indicators of all the non-crises years. More recently, Fielding and Rewilak (2015) estimate a dynamic probit model in which the post-crisis years are classified as 1s. The explanatory variables set includes not only the lags of the macroeconomic factors, but also the lagged outcome variable, with the aim of quantifying the persistence of the crisis.

Hardy and Pazarbasioglu (1999) and Caggiano et al. (2014) explicitly address the issue that post-crisis years may differ significantly from times of normality, i.e. they tackle what Bussiere and Fratzsche (2006) – in studying currency crises – label "post-crisis bias". Therefore, their target variable is not binary, but takes on three values. In both papers, the value 0 identifies tranquil times. They differ in how they classify the other two values. In the former, 1 identifies the pre-crisis years and 2 the crisis years, while in the latter, 1 labels the crisis year and 2 the crisis years other than the first. Given the nature of the dependent variable, in both cases, the methodology adopted is a multinomial logit.

In most of these studies, the explanatory variables are taken in lags, since the objectives of the analysis are (i) to build an EWS to link pre-crisis macroeconomic imbalances to the crisis episodes, and (ii) to perform forecasts to predict the risks that a crisis may arise in the future, should these imbalances occur again.

The use of discrete outcome models is widely accepted and employed in the literature. Despite they are not structural macroeconomic models – but reduced-form models – they allow an economic interpretation of the links between the outcome variable and the explanatory variables through their estimated signs and coefficients. Like any standard econometric technique, logit models heavily depend on data availability, particularly in cases where the analysis covers a wide range of countries. They are also best kept relatively simple for ease of interpretation. Most importantly, they are not optimised to solve prediction problems (Kleinberg et al., 2015), which instead is the focus of EWS. To overcome these shortcomings, economists are increasingly employing machine learning (ML) methods in empirical works where the main objective is to perform predictions (for a review, see Athey, 2018). ML (or rather "supervised" ML) is a data mining tool able to (i) analyse complex datasets, (ii) fit multifaceted and flexible functional forms to the data and (iii) find functions that perform well out-of-sample (Mullainathan and Spiess, 2017).

As regards the prediction of crises, 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. In Duttagupta and Cashin (2011) banking crises are analysed by means of a Binary Classification Tree (BCT). The paper covers 50 developing and emerging countries, with data from 1990 to 2005. Alessi and Detken (2018) implement the Random Forest (RF) algorithm and apply it to banking crises in the European Union (EU), UK, Denmark and Sweden by using quarterly data from 1970 to 2013.

3. The banking crisis dataset and the target variable

To identify a banking crisis, we start from Laeven and Valencia (2008) and subsequent updates (Laeven and Valencia, 2013 and 2018). This allows us to detect 97 crisis episodes for 100 countries, over the period 1970-2017. To identify additional crisis episodes, we merge information from further sources that apply different criteria to detect a banking crisis. We retrieve 43 additional crisis episodes from Reinhart and Rogoff (2009) and 1 additional crisis episode from Jordà et al. (2017a). Our final dataset contains 141 crisis episodes, covering 100 countries between 1970 and 2017 (33 advanced economies and 67 emerging ones, see **Table A2** for the complete list). This dataset is the basis of our empirical analyses of Section 5 and Section 7.

3.1 Target variable: crisis vs pre-crisis

After having defined what a banking crisis is and identified the crisis episodes, we turn our attention to the construction of the target variable employed in our empirical analyses. The literature adopts two different approaches: one that aims at predicting the occurrence of a crisis (see, among others, Demirgüc-Kunt and Detragiache, 1998 and 2005; Schularick and Taylor, 2012; Richter et al., 2017), and one that aims at signalling the building up of macroeconomic imbalances that may lead to a crisis (e.g. Alessi and Detken, 2018). In the former, the target variable is the crisis itself, while in the latter the target variable consists of the pre-crisis periods.

Since our interest lies in building an early warning system that may help anticipate the occurrence of a crisis, we follow the second approach both in the econometric analysis (Section 5) and in machine learning (Section 7).[6] For this reason, we need a definition of the pre-crisis years. Following Arteta and Eichengreen (2000), we label the three years preceding each banking crisis "pre-crisis". Moreover, we label the three years following each banking crisis "post-crisis". Finally, the time spells that are at least three years past a crisis and at least three years prior to a crisis are classified "normal times". Therefore, in addition to the crisis episodes, our sample is partitioned into three intervals: pre-crisis, post-crisis and normal (or tranquil) times (**Table 1**).

TABLE 1 ABOUT HERE

Table 2 and **Table 3** display the occurrence of crises, pre- and post-crisis episodes and normal times, for a selection of advanced and emerging economies, respectively.[7] 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. 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 normality is shorter than the predefined three-year spell (i.e. Japan between 1992 and 1997).

TABLE 2 and 3 ABOUT HERE

Over the whole period, we observe 55 banking crises for the 33 advanced economies and 87 for the 67 emerging economies. 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

[6] In the econometric analysis, we also show the results of logit models in which the target variable is the crisis.

[7] For the sake of brevity and readability of the tables, we selected twenty advanced economies and twenty emerging ones based on their contribution to world GDP, with the exception of Kazakhstan, Russia, Ukraine and Hungary, which are listed because they are the only emerging countries in which we record a crisis in 2007 or after.

between 2007 and 2008. Among the emerging economies, sixteen countries did not experience any crisis consistent with the definition adopted in this paper.[8] The distribution of the crises is much more disperse than among developed countries and shows the highest concentration of episodes in the 1990s. Only four economies (Kazakhstan, Russia, Ukraine and Hungary) experienced a banking crisis in 2008.

Targeting either the crisis or the pre-crisis years entails three definitions of the outcome variable, which in all three cases, is a binary (0/1) variable (**Table 4**). In Approach 1, the value 1 identifies the crisis, while the value 0 all other periods. In Approach 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 to avoid the post-crisis bias, we adopt definition (2b) and drop the post-crisis periods from the definition of the 0s. We use this characterisation of the target variable in our main specifications of the empirical analyses of Section 5 and Section 7. For comparison purposes, Section 5 also presents a model in which the outcome variable is defined according to Approach 1. The two approaches imply a different set of explanatory variables. In Approach 1, the occurrence of the crisis is explained by previous-period macroeconomic factors, while in Approach 2 all factors are contemporaneous with the pre-crisis period.

TABLE 4 ABOUT HERE

4. Descriptive statistics

The banking crisis dataset of Section 3 is merged with the macroeconomic indicators listed in **Table A3** in the Appendix. We select these variables since previous literature suggests that they could contain valuable information to identify vulnerabilities that may lead to a banking crisis. For this reason, they are used as explanatory variables in the empirical applications. As our pre-crisis indicator is at the yearly level, we gather macroeconomic data on a yearly basis. To have the widest possible coverage across both time and countries for a single indicator, we combine consistent data from different sources when needed. For instance, we complement data on credit-to-GDP from the Bank of International Settlement (BIS) with information taken from the World Bank (WB).[9]

The selected variables can be grouped in two sets according to their level of detail: (i) country-specific and (ii) global. As to the former, we identify a few macroeconomic fundamentals, in particular the current account balance as a share of GDP, external debt-to-GNI and public debt as a ratio of GDP. Richter et al. (2017) for example stress how a larger current account deficit indicates increased financial flows from abroad, which might increase financial fragility because of possible capital flow reversals. We also control for external and public debt as a proxy for countries' solvency and liquidity (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. Moreover, higher levels of external debt may indicate a country's greater reliance on foreign investors making it more vulnerable to external shocks.

As our focus is on banking crises, we choose two banking variables taken from the literature on credit booms. The first one is credit-to-GDP, while the second is bank credit-to-bank deposits. According to

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

[9] In some cases, we are still left with some missing values. If gaps are sparse, we recover the missing values by interpolation.

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. However, Schularick and Taylor (2012) and Richter et al. (2017), find weak evidence that excessive credit growth poses a threat to financial stability. Furthermore, we employ bank credit-to-bank deposits as a measure of aggregate liquidity of the banking sector. Jordà et al. (2017b) find that this indicator increases prior to banking crises and enhances the risk of credit booms ending badly.

We consider an additional set of indicators, namely inflation and an openness index. As suggested by Demirgüc-Kunt and Detragiache (1998), inflation may provide indications of macroeconomic mismanagement, which adversely affects the economy. We account for the degree of openness of a country as economies that are more open may be more exposed to financial fragilities coming from abroad.

The last country-specific variable refers to asset prices, more specifically the yearly growth of the real house price index. Recent literature stresses that house price booms are a key vulnerability of modern economies, especially in times of "credit bubbles" (Jordà et al., 2015). In this framework, we follow Alessi and Detken (2018).

Our second set of indicators provides information on a global scale. These include the 10yr US treasury rate, a composite energy price index expressed as year-on-year percentage changes and the real world GDP growth. The 10yr US Treasury rate is meant to highlight vulnerabilities affecting emerging economies especially. In particular, tight monetary conditions in the US may cause a reverse in capital flows to emerging economies and thus contribute to their debt servicing difficulties (Manasse and Roubini, 2009). By deteriorating the balance of payments of highly import dependent countries, the change in energy prices may provide information on the resilience of an economy to adverse exogenous shocks. Finally, the conditions of the global economy as a whole are captured by the real growth of world GDP. This variable may have an ambiguous explanatory effect on financial instability. For instance, high real GDP growth rates may signal either overheating or a buoyant economic environment.

4.1 Data manipulation

Before proceeding, it is important to describe how we prepare the data for the next steps of our analysis. As for the country-specific variables, we treat the data in the following way. The indicators expressed as year-on-year percentage changes are used as they are, namely inflation and the house price index. Each of the other time series is detrended using a two-sided Hodrick-Prescott (HP) filter and standardised (i.e. we subtract the country specific mean and divide by the standard deviation), with the exception of the current account balance as a share of GDP which is only standardised.[10] Detrending allows us to remove the time trend and capture the cyclical component, while the standardisation smooths heterogeneities among countries and across time. Overall, the detrending and standardisation allow us to compare the behaviour of diverse variables across different countries.

Regarding the set of variables at the global level, we plug them in our models without manipulation as they are expressed as year-on-year percentage changes and they do not vary across countries.

[10] In the HP filter, we use a smoothing parameter equal to 100 as the periodicity of our data is yearly. In Appendix C we also apply a one-sided HP filter as a robustness check.

4.2 *Crisis, pre-crisis and the macroeconomy*

Table 5 presents summary statistics of the control variables for the group of advanced and emerging economies separately. Additionally, it presents the means of the raw and transformed version of the indicators used in the analysis, discriminating between means in the full sample ("Overall"), and the means for the four sub-periods described in Section 3.1. In the following, we comment the raw variables as the transformed ones have mean equal to 0 by construction. Two main observations follow.

TABLE 5 ABOUT HERE

First, there is a significant difference in a number of country-specific variable means between advanced and emerging economies in the full sample, as shown by the t-test reported in **Table 6**. 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 ones. External debt-to-GDP and the openness index are 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 the means associated to the banking variables, especially credit-to-GDP. It suggests that in emerging economies the banking sector is not as developed as in the other group of 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.[11]

TABLE 6 ABOUT HERE

Second, the path of average values from tranquil times to the outbreak of the crisis is consistent with expectations. 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-GDP 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 the crisis. Banking variables increase as we approach the crisis but more for developed countries. Overall, this suggests that we correctly classified the four sub-periods. Yet, only a few of the selected macroeconomic variables improve in the post-crisis period. 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 the analysis of Sections 5 and 7.

Next, we look at the distribution of normal times, pre-crisis and crisis plus post-crisis years according to the quartile of each of our explanatory variables. This type of analysis allows us to identify which of the macroeconomic variables are more useful to detect vulnerabilities that may precede a crisis. Also, we acknowledge that the set of relevant variables may differ between advanced and emerging economies. For this reason, we continue to keep the two subsamples separate.

Figure 1 draws attention to a selection of macroeconomic indicators that, more than others, carry information on the existence of macroeconomic imbalances in the years prior to a crisis. Worthy of note is that the frequencies associated with pre-crisis years are relatively low, while the tranquil times represent the majority of the observations. **Panel (a)** shows that in developed countries current

[11] In our dataset there are 123 observations with inflation higher than 100% and, with the exception of 6 observations (Israel between 1980 and 1985), they belong to the group of emerging economies.

account deficits are associated with a higher occurrence of pre-crisis periods. As the current account balance improves (we move from the first to the fourth quartile), pre-crisis years are less common. For emerging economies, instead, no clear pattern emerges, unlike what we expected.

FIGURE 1 ABOUT HERE

A comparison of **panel (c)** and **panel (d)** shows that high levels of inflation is particularly informative for emerging economies. The two upper quartiles are characterized by a higher frequency of pre-crisis periods compared to the two lower quartiles. In developed countries, however, no clear pattern emerges. As regards the credit-to-GDP and bank credit-to-bank deposit ratio, the occurrence of pre-crisis periods increases with the value of these two banking variables. This evidence is especially visible for developed economies (**panel e** and **panel g**).

A number of additional variables provide insights on countries' vulnerabilities.[12] In developed economies, the external debt-to-GDP presents a higher number of pre-crisis occurrences in the upper tail of its distribution. Meanwhile, pre-crisis periods in emerging economies are more common in the lower tail. Public debt-to-GDP presents a higher occurrence of pre-crisis years in the first quartile, especially for developed countries. Regarding the openness index, no relevant evidence emerges.

Summarizing, a number of interesting insights emerge from this descriptive analysis. Current account deficits are usually associated to increased vulnerabilities that may lead to a banking crisis. The same applies to hyperinflation, especially for emerging economies. In developed economies, macroeconomic imbalances are associated with overheated credit markets as well as high levels of external debt-to-GDP ratios. In the next sections, we aim at corroborating these results by employing econometric and machine learning techniques.

5. Logit models: Estimation results and predictive performance

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 banking crisis or a pre-crisis period, according to the specification. In particular, we estimate a pooled logit model as follows:

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

where $Prob(y_{it} = 1|X_{it})$ denotes the probability that country *i* in year *t* is in a crisis or pre-crisis state, X_i is a set of regressors and α_i are geographic dummies. We run three specifications of Equation (1) in line with the definition of the target variable provided in Section 3.1. According to the outcome variable, the information set included in X_i is taken either at time *t-1* or at time *t*. Moreover, we apply the model to the subsample of developed and emerging economies separately, as we acknowledge heterogeneities between the two groups of countries.[13] In particular, we recognize that in developed

[12] For this final set of macroeconomic indicators, we do not report the corresponding graphs for the sake of brevity. However, they are available upon request.

[13] See **Table A2** in the Appendix for the complete list of countries included in our dataset. The sample of advanced economies includes 33 countries observed over the period 1970-2017 for a total of 1,584 observations. We drop Estonia, Hong Kong, Israel, Lithuania and Singapore (240 observations in total) as a complete time series for the selected macroeconomic indicators is not available. We end up with 51 crisis episodes and 1,293 non-crisis episodes. The sample of emerging economies includes 67 countries observed over the period 1970-2017 for a total of 3,216 observations. We observe 87 crisis episodes and 3,129 non-crisis episodes.

and emerging economies vulnerabilities are related to different set of macroeconomic factors. Coherently, the set of explanatory variables for developed economies includes: current account-to-GDP, external debt-to-GNI, public debt-to-GDP, credit-to-GDP, while for emerging economies, we replace credit-to-GDP with inflation following the findings that arise from the descriptive analysis of Section 4. We call these two sets of variables "baseline".

Before presenting the results, some clarifications are in order. The selection of regressors to include in Equation (1) is heavily affected by data availability across both time and countries. For instance, house prices are available only for a subset of countries. If this indicator is plugged in the model it would greatly reduce the sample size thereby jeopardizing the validity of our results. Another reason why we can only include a limited number of indicators in our set of explanatory variables is that we need to attenuate potential correlation and endogeneity bias. In our framework, correlation and endogeneity stem from the fact that macroeconomic indicators react in unison to large scale events, such as banking crises, or show a similar behaviour in the run up to a crisis. We therefore opt to include only the most meaningful country specific macroeconomic indicators disclosed by the descriptive analysis above plus a number of controls at the global level.

The number of observations varies according to the specification of Equation (1). When the outcome variable corresponds to the pre-crisis period, we end up with 953 and 2,010 observations for the sample of advanced and emerging economies, respectively.

5.1 The logit models results

Table 7 and **Table 8** present the estimated marginal effects for the subsample of advanced and emerging economies, respectively.[14] The models for the sample of developed countries include country dummies, while those for the group of developing countries include region dummies.[15] Standard errors are clustered at the country level to account for any leftover serial correlation among observations belonging to the same cluster.

In both tables, in Column (1) the outcome variable identifies the year when the crisis occurs, in line with Approach 1 of **Table 4**. The explanatory variables are taken at time *t*-1, as it is reasonable to assume that banking crises at time *t* are generated by previous year macroeconomic imbalances.[16] The dependent variable in Column (2) identifies the pre-crisis periods and corresponds to the outcome variable of Approach 2(a). From Column (3) to (5) the outcome variable identifies the pre-crisis periods but post-crisis periods are excluded from the 0s, in line with Approach 2(b). This is our preferred outcome variable as it drops observations that may suffer from the post-crisis bias (Bussiere and

[14] In the logit model, as in any non-linear model, estimated coefficients are not directly interpretable. Therefore, we show the derived marginal effects, which allow us to quantify changes in probabilities when a regressor changes by one unit. In our framework, a positive (negative) coefficient means that higher levels of the associated macroeconomic indicator increases (decreases) the probability of observing a crisis or pre-crisis period, according to the specification.

[15] We are forced to include region dummies, instead of country dummies, for the subsample of emerging economies, as a significant number of these countries did not experience a banking crisis. In particular, Barbados, Belize, Botswana, Brunei, Fiji, Gabon, Iran, Libya, Mauritius, Namibia, Pakistan, Serbia, Seychelles, Suriname, Syria, Trinidad and Tobago and Turkmenistan. We clustered emerging economies in the following regions: Africa, Asia, the Balkans & East Europe, Caribbean, Central America, Central Asia, East Asia, Latin America, Middle East, North Africa, Pacific and South Asia.

[16] Additionally, one period lagged variables are used to mitigate potential endogeneity issues. Indeed, contemporaneous variables may not be exogenous if the effects of the banking crisis propagate quickly to the rest of the economy (Demirgüç-Kunt and Detragiache, 1998).

Fratzsher, 2006). For this last set of regressions, the explanatory variables are taken at time *t*. For each specification, we also report the Area Under the Receiver Operating Characteristic curve (AUROC), a standard measure used to evaluate the predictive performance of a logit or, more generally, any binary classification model.[17]

Overall, the logistic regression confirms the evidence emerging from the descriptive statistics. In advanced economies, the occurrence of a crisis is significantly associated to higher levels of external debt (**Column 1** of **Table 7**). Meanwhile, the likelihood of experiencing a crisis falls as public debt increases. A possible interpretation of this result 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 brakes out, fiscal measures implemented by government could burden public debt. We also find that higher levels of current account deficits and credit-to-GDP increase the probability of observing a crisis, although only the latter is statistically meaningful. As for the global variables, the probability of the occurrence of a crisis increases with the 10yr US Treasury rate and world GDP growth. This result provides evidence that, similarly to what we expect for emerging economies, a tight US monetary policy produces imbalances potentially leading to a crisis in advanced economies as well, while world GDP growth could increase crisis probability by fostering a worldwide easing of credit standards and a growing inter-dependence among countries. The same kind of information is conveyed when we look at the probability of being exposed to pre-crisis periods (**Column 2**).

TABLE 7 ABOUT HERE

These findings are robust to the exclusion of the post-crisis periods (**Column 3**). Noteworthy is the improvement in the predictive performance of the model, measured by the AUROC at the bottom of the table, as we move from the second to the third specification that is, when we drop observations corresponding to the post-crisis periods. This result confirms the presence of post-crisis bias in our data.

We enhance the model of Column 3 by adding the interaction between external debt and the 10y US Treasury rate (**Column 4**). Previous results are confirmed, with the exception of the coefficient associated with external debt, which loses statistical significance.

Turning to emerging economies (**Table 8**), the probability of the occurrence of a crisis is positively associated to higher levels of inflation, while it is negatively related to higher levels of public debt and, albeit mildly, current account deficits (**Column 1**). Yet, external debt does not meaningfully affect the probability of experiencing a crisis episode. As for the global variables, only the 10yr US Treasury rate is significantly related to the likelihood of observing a crisis. This suggests that vulnerabilities in emerging countries cannot be detected through changes in world GDP growth, as they are less open economies.

TABLE 8 ABOUT HERE

When we consider the specification where the outcome variable identifies pre-crisis periods (**Column 2**), inflation loses its statistical relevance, although the associated coefficient is still positive. These results are confirmed when we drop observations corresponding to the post-crisis periods and, as expected, the predictive power of the model increases (**Column 3**). When we include the interaction

[17] The AUROC is calculated from the ROC curve, which plots the combinations of true positive and false positive 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 0.5 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.

between the 10yr US Treasury rate and the external debt in Equation (1), this term positively affects the likelihood of observing a pre-crisis period and the external debt gains significance with a negative sign (**Column 4**).

All in all, our findings are in line with those of similar work.[18] Richter et al.'s (2017) analysis of banking crises for a sample of 17 developed economies find a positive, although insignificant, coefficient associated with credit-to-GDP. They also find that current account deficits increase the probability of observing a crisis. In Caggiano et al. (2014), banking crises events in Sub-Saharan African countries are not significantly related to inflation. As regards world GDP growth, we reconcile the observed positive coefficient with findings from Kaminsky and Reinhart (1999) that output tends to peak about 8 months before the onset of a crisis.[19]

5.2 The predictive performance of logit models

To evaluate the predictive performance of the logit model, we split each of our subsamples into a *training* and a *testing* set. The training set is used to estimate the model and the testing set to assess how well the model fits the data. More specifically, we build our training set by randomly picking 80% of the observations. We use the training set to estimate Equation (1): from this first step we compute the predicted probabilities of observing a crisis or a pre-crisis for country i at time t. The second step consists of using the estimated coefficients to compute probabilities on the testing set and test our model's predictions.[20] We replicate this procedure 1,000 times.

From each replication, we calculate various performance indicators typically used to assess the goodness of fit of any classification model, including machine learning algorithms: ROC and associated AUROC (see Section 5.1), accuracy, precision and sensitivity rates. Accuracy, precision and sensitivity rates derive from the so-called "confusion matrix", which compares predicted values with observed ones (**Table 9**). Accuracy is defined as the ratio between the observations correctly predicted and total observations, while precision is the ratio between the correctly predicted 1s and total predicted 1s. Sensitivity is the ratio between the correctly predicted 1s and total observed 1s.

Performance indicators can be computed both for the training set and for the testing set. In the former case, the evaluation is in-sample, while in the latter it is out-of-sample. In the following, we only comment on the out-of-sample performance of our preferred specification of Equation (1), i.e. the one corresponding to Column (3) of **Table 7** and **8**.[21] As we perform 1,000 replications, and thus have 1,000 possible realisations, we need to summarize our results in the most convenient fashion.

TABLE 9 ABOUT HERE

Starting from the predicted probabilities, we take the average of the predicted probabilities calculated in each replication by country and year. We then plot the yearly distribution of these averaged probabilities in panel (a) of **Figure 2** and **Figure 3** for the subsample of advanced and emerging economies, respectively. For ease of comparison, panel (b) of each figure shows the number of pre-

---

[18] Results hold when we include real GDP growth rates at the country level in the set of regressors. See **Table B1** and **B2** in the Appendix for the corresponding marginal effects.

[19] They refer to countries' GDP growth, which determines world GDP growth.

[20] For our preferred specification (Column 3 of **Table 7** and **Table 8**) in each draw, the training set comprises 762 and 1,608 observations for the sample of advanced and emerging economies, respectively. Consequently, the testing set includes 191 and 402 observations for the sample of advanced and emerging economies, respectively. Worthy of note is that the number of observations for the training and the testing will always be the same across each of the 1,000 replications although not necessarily identical.

[21] In-sample performance indicators are available from the authors upon request.

crises actually observed in our dataset. Some additional comments follow. For advanced economies (**Figure 2**), the estimated probabilities are a good predictor of banking crises. They increase in the run up to the most widespread and severe crises, notably those of the beginning of the 1990s and of 2008. Also, they perform relatively well for the patchier crises, such as those of the early 80s. Turning to emerging economies (**Figure 3**), the fitted probabilities perform well for the cluster of crises concentrated at the beginning of the 80s. Yet, their performance is rather poor with regards to the banking crises of the 90s. A possible reason is that the model, and more specifically the set of explanatory variables, chosen is not the most suitable to detect this group of banking crises.

FIGURE 2 and 3 ABOUT HERE

Turning to the performance indicators, **Table 10** provides summary statistics of the distribution of the AUROC, accuracy, precision and sensitivity rates for the subsamples of advanced and emerging economies. Worth mentioning is that these indicators have been calculated after classifying each observation as a positive ("1" or pre-crisis year) or negative ("0" or normal times) outcome according to the associated predicted probabilities. For classification, we have to choose a cutoff, i.e. a threshold above which observations are classified as 1 and 0 otherwise. In this exercise, and in line with the ML exercise below, we choose a cutoff equal to 0.5.[22]

For advanced economies (**Table 10**), the AUROC is, on average, lower than the one resulting from the estimation of Equation (1) on the full sample (0.74 versus 0.80, see bottom of **Table 7**). The same observation applies to the sample of emerging economies (**Table 10**). The accuracy rate is very high for both country groups. It tells us that, on average, the model correctly predicts 9 observations out of 10 total observations for both samples of countries. However, accuracy rates can be misleading especially when there is a large class imbalance problem, in our case a high number of observed 0s compared to 1s. When the sample is unbalanced, the model is correctly predicting the majority class and thus, achieving a high classification accuracy. For this reason, accuracy rates can be a poor measure of the model's performance and additional measures, such as precision and sensitivity are required to evaluate the classifier.

The precision rate for advanced economies (**Table 10**) suggests that, on average, the model correctly predicts almost 7 out of 10 predicted pre-crisis episodes. Yet, the sensitivity rate implies that the model correctly predicts only 2 out of 10 observed pre-crises. Turning to the emerging economies, the model performs rather weakly. According to the precision rate, the model on average correctly predicts 3 out of 10 predicted pre-crisis episodes. The sensitivity rate suggests that the model predicts less than 1 out of 10 observed pre-crisis episodes.

TABLE 10 ABOUT HERE

Finally, **Figure 4** plots the ROC curves for the subsample of advanced – panel (a) – and emerging economies – panel (b). In particular, from the 1,000 ROCs obtained, for each country group we choose the one with a value of the AUROC nearest to the mean shown in **Table 10**. The AUROC corresponding to the chosen ROCs is 0.74 and 0.75 for advanced and emerging economies, respectively**.**

FIGURE 4 ABOUT HERE

[22] We choose a cutoff of 0.5 to make the results from the logit model comparable to those obtained when employing the AdaBoost. Another approach we use is to employ a cutoff equal to the mean of the predicted probabilities conditional on the true outcome being 1. The main insights do not change and the correspondingresults are available from the authors upon request.

6. Supervised machine learning: Decision tree classifiers

Economists are increasingly employing supervised machine learning in empirical works where the main objective is to perform predictions and where it is necessary to extrapolate information from large datasets characterized by high heterogeneity (for an overview, see Athey, 2018). With reference to the crises literature, examples in this direction are Manasse and Roubini (2009) for sovereign debt crises, and Duttagupta and Cashin (2011) and Alessi and Detken (2018) for banking crises.

Before moving to the novel empirical application of our paper, we shortly introduce machine learning and decision tree classifiers. According to Athey (2018), the field of ML is concerned with the development of algorithms suitable to be applied to large and heterogeneous datasets, with the main objectives being prediction, classification and clustering. ML is of two types, supervised and unsupervised. Unsupervised ML infers patterns from a dataset without reference to known or labelled outcomes. It can be applied to clustering (i.e. splitting the dataset into groups according to similarity) or dimensionality reduction (i.e. reducing the number of features in a dataset). Instead, supervised machine learning is suitable for a wide range of applications where the aim of the analysis is to predict an outcome based on the behaviour of a set of predictors or "features" (the equivalent of covariates or explanatory variables in econometrics). In other words, it revolves around the problem of prediction (Kleinberg et al., 2015; Mullainathan and Spiess, 2017). Here, we focus on supervised machine learning.

In this paper, we have a dataset that includes a binary outcome variable (pre-crisis vs normal times, see Section 3) and a set of features (see Section 4). We wish to perform out-of-sample forecasts to predict the likelihood that a banking crisis may occur within a three-year spell. We are dealing with a prediction problem, which fits within the framework of supervised machine learning. The most straightforward way to address the issue is to apply a logistic regression (as in Section 5). However, the ML literature suggests the use of alternative nonlinear methods that are concerned primarily with prediction, unlike traditional econometric methods, which are not optimised to solve prediction problems (Kleinberg et al., 2015).[23] A way forward is to employ supervised ML methods, such as Classification and Regression Trees (CART), Random Forests (RF) and Adaptive Boosting (AdaBoost).[24]

In our empirical exercise, we address what in machine learning terminology is called a *classification problem*. A way to solve it is to use a *decision tree classifier*. The simplest classifier is CART, while more complex ones (the so-called *ensemble models*) are RF and AdaBoost.[25] Broadly speaking, these methods select features – and their critical values – to classify the outcome variable. They offer some advantages. First, they are particularly appropriate when datasets are large and characterised by high heterogeneity. Second, they have the ability to capture non-linear relationships and to identify relevant interactions among two or more variables. Third, they are not sensitive to missing values – they replace them with the most probable value – or to outliers. Finally, they allow a large explanatory set, since the statistical algorithm is able to select the most relevant variables in predicting the outcome. Before applying a decision tree classifier, the dataset is conventionally split into a *training* and a *testing set*: the training set is used to estimate ("train") the model (or "tree") and the testing set to evaluate the predictive performance of the model.

[23] The empirical economic literature (e.g. Manasse and Roubini, 2009, Duttagupta and Cashin, 2011, and Alessi and Detken, 2018) benchmarks ML results against those of logit models.
[24] Other supervised ML techniques include penalised regression (e.g. LASSO and elastic nets), support vector machines (SVM), neural nets and matrix factorisation (for further details, see Varian, 2014, and Athey, 2018).
[25] See Freund and Schapire (1996) on Adaptive Boosting.

Specifically, 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 so-called "impurity measure".[26] 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 one of the aforementioned impurity criteria. 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). The nodes that cannot be optimally split further are called *terminal nodes*. **Figure A1** in Appendix A depicts an example of classification tree.

These models may suffer from two drawbacks: instability and overfitting. Instability implies that small changes in the training set may cause large changes in classification rules. For instance, we could obtain two different trees from two similar training samples if the algorithm does not select the same variable in the first split. Overfitting refers to the tree's generalisation capability: an overfitted model gives a highly accurate prediction in sample, but a poorly accurate one out of sample. This could happen when too many splitting rules are applied compared to data availability.

Ensemble models, such as random forest and adaptive boosting, seek to overcome these limitations. As regards instability, these algorithms train many decision trees on different subsamples ("folds") of the initial dataset and then combine them in order to give a final prediction. As regards overfitting, it can be avoided by correctly setting some parameters ("regularizers", see below).

Both RF and AdaBoost estimate a multitude of trees to grow a forest, allowing us to obtain a strong and stable model from many weak and unstable ones. However, they differ in how they aggregate trees to get a final overall result. On the one hand, RF randomly resamples the training set and estimates N single models in parallel by using a certain number of features that are randomly selected in each replication. It subsequently averages across models in order to improve the performance of the estimator ("bagging"). On the other hand, AdaBoost 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 so doing, the model gives more weight to the observations that are more difficult to predict. Both algorithms perform better than CART, but the interpretation of their outcome is less intuitive since no final tree is represented: we can compare the importance of the variables but not the way they interact.

Their implementation requires pre-setting the regularizers: (i) *tree depth*, i.e. maximum number of nodes along the longest path from the root note down to the farthest leaf node; (ii) *minimum split*, i.e. minimum number of observations in a node to allow for a split; and (iii) *number of final trees*, i.e. number of trees (base models) in the forest. In addition, AdaBoost involves choosing the function that attributes increasing weight to the incorrectly classified observations at each round. Since AdaBoost corrects for misclassified observations and its predictive power is expected to be higher, we rely on it for the empirical analysis of Section 7. In Appendix C, we implement a comparative analysis of the two methods, which shows that the predictive performance of AdaBoost is superior to that of RF when applied to our framework of analysis.

[26] Given the distribution of a discrete variable contained in a node, we define impurity a measure of its dispersion. In each node, the choice of the predictor and of the cut-off point is made in order to maximize the reduction of impurity from the parent node to its child nodes. Examples of impurity measures are the Gini index and entropy.

7. Machine learning results

In this section, we preliminarily clarify some technical issues concerning the AdaBoost implementation (Section 7.1) and show the superiority of the AdaBoost compared to the logit (Section 7.2). An "enlarged" AdaBoost model is presented in Section 7.3 and finally we assign a probability to the event that a country is involved in financial troubles over a three-year horizon (Section 7.4). For this purpose, our outcome variable is the one that classifies the "pre-crisis" years as 1 and normal times as 0. As discussed in Section 3, pre-crisis spells correspond to the three years preceding a banking crisis. In order to avoid the post-crisis bias, we drop the observations corresponding to three years following a crisis (the post-crisis years), so that the 0s mark normal times only. Therefore, our outcome variable is what we labelled definition 2b in **Table 4** of Section 3.

The analysis is performed separately on the two sub-samples of advanced and emerging economies to take into account any differences in variable selection, model parameterization and performance evaluation. The descriptive statistics of the variables employed in the analyses are those of **Table A3**. We carry out the analysis on three sets of variables:

(1) the "baseline" set (corresponding to that employed in the logit model for comparison reasons) includes a common group of variables for both advanced and emerging economies: current account-to-GDP, external debt-to-GNI, public debt-to-GDP, world GDP growth and US 10yr Treasury rate. We add credit-to-GDP for the advanced economies and inflation for the emerging countries. All variables are simultaneously detrended and standardised, with the exception of the current account-to-GDP, which is standardised-only, and inflation, world GDP growth and US 10yr Treasury rate, which are percentage values.

(2) the "enlarged" set includes two sets of common variables for both advanced and emerging economies: (i) current account-to-GDP, external debt-to-GNI, public debt-to-GDP, credit-to-GDP, openness index and bank credit to bank deposits and (ii) inflation, world GDP growth and energy prices. Finally, house prices are for the advanced countries. With the exception of current account-to-GDP, inflation, world GDP growth, US 10yr Treasury rate and house prices, all variables are simultaneously detrended and standardised. We also add their standardised-only transformation.

(3) the "alternative" set enriches the "enlarged" one with a variable transformation that is intended to capture the build-up of the crisis: for all the variables expressed as ratios to GDP, we include their 3-year differences at each point in time. Moreover, we lower the cutoff threshold to classify the estimated probabilities as 0 or 1 from 0.5 to 0.45 in order to give more emphasis on correctly predicting pre-crisis episodes at the cost of incurring a higher risk of issuing false alarms.The results are reported in Appendix D.[27]

After excluding all observations for which one or more variables are not available, in the baseline model we are left with 953 observations for advanced economies and 2,010 for emerging ones. In the enlarged model, we have 918 observations for advanced countries and 1,887 for emerging ones. The analyses of Sections 7.1 and 7.2 rely on the "baseline" dataset, those of Sections 7.3 and 7.4 on the "enlarged" one, while those of Appendix D on the alternative dataset and model setup.

In both cases, the samples are divided randomly into a training and a testing set. The former consists of 80% of the sample and is used to train the model, the latter contains the residual 20% and is employed to assess the out-of-sample performance of the model. We train the models 1,000 times, compute the performance indicators (AUROC, precision, sensitivity and accuracy rates) for each

[27] At the current stage of the paper, the analysis is performed only for the advanced economies.

replication and then take their average values: these are used to assess the model ability to predict pre-crisis episodes.

Finally, robustness analyses of different specifications of the "baseline" model (pre and post crisis definition, different cut-off probabilities, oversampling of the crisis events, alternative detrending procedure, and comparison with RF) are presented in Appendix C.

7.1 Setting the model parameters

Before proceeding to the estimation of the model, we need to set the regularizers set that maximises the predictive power of our model and that minimises overfitting. To this end, we build an objective function that depends on in-sample and out-of-sample precision and sensitivity, which in turn both depend on the regularizers. Our proposed objective, or utility, function $U(\theta)$ is the following:

$$\begin{aligned} U(\theta) = \ & \beta_1 PREC_{out}(\theta) + \beta_2 SENS_{out}(\theta) - \beta_3[PREC_{in}(\theta) - PREC_{out}(\theta)] + \\ & - \beta_4[SENS_{in}(\theta) - SENS_{out}(\theta)] \end{aligned} \tag{2}$$

and we maximise it with respect to the regularizers vector $\theta = (\theta_1, \theta_2, \theta_3)$, where θ_1 is tree depth, θ_2 is number of final trees and θ_3 is minimum split, as defined in Section 6. *PREC* and *SENS* represent precision and sensitivity, respectively, and the subscripts *in* and *out* their in- and out-of-sample values. Both measures are a function of vector θ.

The objective function $U(\theta)$ depends on four terms: (i) out of sample precision, (ii) out of sample sensitivity, (iii) the difference between in-sample and out-of-sample precision and (iv) the difference between in-sample and out-of-sample sensitivity. Terms (i) and (ii) allow us to maximise the predictive ability of the model, while terms (iii) and (iv) to minimise overfitting by introducing a penalization for the difference between in- and out-of-sample performance indicators. At the same time, by including the first two terms we wish to maximise simultaneously sensitivity (i.e. the pre-crisis spells correctly identified) and precision (i.e. the ability of the model to avoid false alarms). Moreover, the presence of the second two terms allows us to minimise simultaneously the gap between in- and out-of-sample values of both precision and sensitivity, i.e. the gaps between correctly identified pre-crises and avoided false alarms. Finally, the parameters $(\beta_1, \beta_2, \beta_3, \beta_4)$ are the weights assigned to the four terms of the equation. They must sum to 1 and we chose the combination (0.4, 0.4, 0.1, 0.1). This is an arbitrary choice, but it reflects our preference for optimising the forecasting ability of the Adaboost model (β_1) and minimising false alarms (β_2). It also avoids overfitting by anchoring both out-of-sample and in-sample sensitivity and precision rates ($\beta_3;\ \beta_4$).

The maximization procedure of $U(\theta)$ with respect to θ is applied to the baseline set of variables listed in Section 7. It is applied separately to advanced and emerging economies by splitting randomly our samples into the training and testing sets defined above. Moreover, it is a numerical procedure, which consists of a grid search over the regularizers set. **Figure 5** shows the values of our objective function for each combination of the three regularizers: maximum depth and minimum split can be read along the x-axis and y-axis, respectively, while the bar height describes the values of the objective function. The objective function is at its maximum when the model is trained with $(\theta_1, \theta_2, \theta_3) = (3, 35, 70)$ for advanced economies and $(\theta_1, \theta_2, \theta_3) = (4, 30, 30)$ for emerging countries. This result holds when using different random training datasets and different weights β.

FIGURE 5 ABOUT HERE

7.2 Machine learning vs logit

Herein we compare the logit performance with that of AdaBoost to assess which is more accurate in acting as an early warning system. **Table 11** shows the mean, median and standard deviation of the distributions of out-of-sample sensitivity, precision, accuracy and AUROC for both advanced and emerging economies for the AdaBoost model. The corresponding values for the logit model are reported in **Table 10**. As in the logit model, the threshold above which observations are classified as 1 is 0.5.[28]

Overall, the AdaBoost outperforms the logit.[29] Specifically, the AdaBoost delivers a better out of sample performance than the logit model in terms of AUROC, especially for advanced economies (0.846 vs 0.74). The corresponding values for emerging countries are 0.815 and 0.780 for AdaBoost and logit, respectively. The accuracy rate is around 0.9 for both country groups, but as we already warned in Section 5, we must use caution in interpreting this measure since we are working with a strongly unbalanced outcome variable with respect to the relative weight of 1s.[30]

The AdaBoost performs better than the logit model even when looking at the precision and sensitivity rates. As for the precision rate, the AdaBoost correctly predicts almost 7 pre-crisis events out of 10 predicted pre-crises in advanced economies (0.675), but only 4 out of 10 for the sub-sample of emerging economies (0.438). The logit model delivers a similar precision rate for advanced countries (0.680), but an even lower one for emerging countries (0.334). In terms of sensitivity, the AdaBoost correctly predicts almost 4 pre-crises out of 10 observed pre-crisis events for advanced economies (0.361), but only 1 in emerging economies (0.131). The logit model instead delivers lower sensitivity rates for both country groups (0.226 and 0.044, respectively).

TABLE 11 ABOUT HERE

7.3 Enlarged AdaBoost: performance indicators

Having established that AdaBoost outperforms the logit model, we further test its predictive performance on the “enlarged” dataset. The AdaBoost improves its performance in terms of all indicators (**Table 12**). Regarding the advanced economies, the out-of-sample performance shows a precision of 0.743 (equivalent to less than 3 out of 10 false alarms) and a sensitivity of 0.427 (equivalent to more than 4 out of 10 correctly predicted pre-crises): in this specification the misclassified observations are mainly “missed pre-crisis”, while only a low number of “false alarms” is issued. As expected, the accuracy rate is high (0.909). Conversely, for the subsample of emerging economies, sensitivity and precision are lower than those for advanced economies and equal to 0.175 and 0.472, respectively, but better than the baseline specification. The accuracy rate is the same as that for the advanced economies.

TABLE 12 ABOUT HERE

Although sensitivity, precision and accuracy are valid and commonly used measures to evaluate the predictive capacity of a binary classification problem, they are not independent of the threshold used

[28] A robustness analysis with different cutoffs (i.e. 0.15 and 0.3) is performed in Appendix C.
[29] In Appendix C we show that the AdaBoost outperforms the RF as well.
[30] See Appendix C for a robustness analysis in which we oversample the 1s until the dataset is balanced.

to discriminate between 0s and 1s. For this purpose, we refer to the ROC Curve and to the area below it (AUROC).[31] Concerning advanced economies, the model performance appears satisfactory, as the AUROC equals 0.885. A lower AUROC is recorded if we consider the emerging economies (0.854), coherently with the results obtained for the other performance measures.

Finally, we discuss the relevance of the variables included in the analysis. As already argued in Section 6, the AdaBoost does not return a single final tree and, hence, nothing can be said on how the variables interact. Moreover, the algorithm does not provide information about how and by how much these variables affect the probability of a pre-crisis, so that we can only rank them according to their importance, defined as the contribution to the reduction of the impurity measure.

Figure 6 shows the ranking of the ten most important variables for our two sets of countries. For the advanced countries, the global variables US 10y rate and world growth are among the most important ones (1st and 4th place, respectively), possibly indicating the strong inter-connection among countries that favours the spread of risk factors. Banking variables display a higher importance as standardised credit-to-GDP and credit to deposits ratio are 2nd and 5th, while both detrended and standardised public debt are also important. Results about emerging economies confirm the importance of US monetary policy to the outbreak of the banking crises as well as the banking variables. As we expected, the importance of inflation and current account represents a novelty with respect to advanced countries.

FIGURE 6 ABOUT HERE

7.4 "Enlarged" AdaBoost: Out-of-sample exercise and prediction

Since the goal of this paper is to develop an early warning system for banking crises, we perform a forecasting exercise on both advanced and emerging economies from 1990 to 2017. Specifically, we start by training the models on the subsample 1970-1989 to make predictions in 1990 and compare the predicted values with the observed ones. By doing so, we are able to detect whether an observation is correctly predicted and, if not, we are able to distinguish between a "missed pre-crisis" and a "false alarm". We recursively repeat the same exercise from 1991 to 2017 by adding a new year to the training sample at each round and testing it on the first excluded year.

Results are depicted in **Figure 7**: for advanced economies the model shows a good ability in learning from the past as the closer the crisis year, the greater the number of pre-crisis identified. In 2005, the first pre-crisis year for many countries, the model has a modest predictive performance, possibly related to the fact that data are poorly informative because of their proximity to normal times. In 2006 and 2007, more incoming crises are correctly predicted because the model exploits the information on 2005 that is now available because of the rolling method employed in the exercise. This is not true for the emerging economies, for which only few pre-crisis episodes are signalled by the model over the prediction horizon. On the other side, in both models 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 (as no other model) can account for this.

[31] The AUROC is calculated from the ROC curve, which plots the combinations of true positive and false positive 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 0.5 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.

FIGURE 7 ABOUT HERE

A different way to look into our results is to focus on the probability distributions over time retrieved from the AdaBoost for the two groups of countries. In **Figure 8**, for both sub-samples, we represent the maximum and the minimum predicted probabilities as well as the median and the values corresponding to the first and the third quartiles by year.

FIGURE 8 ABOUT HERE

We infer two main results from the probability distributions. Predicted probabilities (i) tend to evolve quite coherently with the number of observed pre-crisis periods in both sub-samples, and (ii) show a higher variance in pre-crisis periods than in tranquil times. Both results indicate that our predicted probabilities perform well in replicating the path of financial vulnerability across countries and in capturing the fact that in pre-crisis periods countries could follow different paths.

Moreover, we note an upward shift of the distributions in 2017 for the advanced economies, 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. By contrast, for emerging economies probabilities remain low, notwithstanding the problems occurred to some countries in the last year.

Table 13 shows country-specific probabilities. In 2006, as the financial crisis was building-up, Latvia, Spain, Iceland, Portugal and the US were the most vulnerable countries, with probabilities ranging between 62% for Latvia and 50% for the US. For emerging economies, the probabilities are lower, consistent with the smaller number of observed crises. The most exposed countries were Croatia, Georgia, Romania, Hungary and Indonesia.

In 2017, probabilities are smaller than those of 2006. Among the advanced countries, the most vulnerable ones are the Netherlands, Greece, the US, Ireland and New Zealand. Their probabilities of incurring a banking crisis over a three-year horizon ranges from 36% to 28% Conversely, emerging economies record low probabilities: the most vulnerable one is Jamaica (19%), followed by Bosnia and Herzegovina (17%), Croatia (16%), Fiji (16%) and Botswana (16%).

TABLE 13 ABOUT HERE

Coming back to the advanced countries, Sweden, Latvia, Singapore and Israel turn out to be the less exposed to financial risks in 2017, with a probability between 11% and 13%. Among the emerging economies, Brazil, Bulgaria, Belize, Ecuador and Macedonia result the less vulnerable (less than 5%).

For both models, a graphical illustration of the comparison between 2006 and 2017 predicted probabilities by means of heat maps is presented in **Figure 9**, where darker colours correspond to higher probabilities to be in a pre-crisis year.

FIGURE 9 ABOUT HERE

8. Concluding remarks

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.

Economists have responded to this appeal by developing EWSs aimed at detecting the risks that a systemic banking crisis may arise. This literature has evolved along different lines, from the signals approach, to discrete choice models to machine learning. These contributions build on, and partially overlap with, a wide field of research directed at characterising banking crises episodes according to a range of criteria.

With this paper, we contribute to the existing literature by developing an EWS to predict the build-up of banking crises in both advanced and emerging economies. To this end, we use an integrated dataset of banking crises and macroeconomic indicators that includes 100 countries (33 advanced and 67 emerging) spanning from 1970 to 2017. We develop an EWS by using both traditional econometrics, namely the logit model and a supervised machine learning algorithm, namely Adaptive Boosting (AdaBoost). Both models entail pros and cons. Ease of use and interpretation are the main advantages of the logit model together with the possibility of assessing the statistical relationship between the single indicator and the probability of observing a pre-crisis event. An advantage of AdaBoost lies in its ability to capture non-parametric relationships. Overall, the AdaBoost shows a higher predictive performance than the logit model. Among the AdaBoost models, the best performing one delivers a 60% pre-crisis predictive ability with a false-alarm propensity of 30%. Logit and AdaBoost models agree to provide increasing predicted probabilities in the last years of the sample, warning against the possible build-up of pre-crisis macroeconomic imbalances.

Bibliographic references

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

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., 2018, "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 (*forthcoming*).

Baron, M., Verner, E., and Xiong, W., 2018, "Bank equity and banking crises", Available at SSRN: https://ssrn.com/abstract=3116148.

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.

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.

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., 2003, "Episodes of Systemic and Borderline Financial Crises", available at: http://siteresources.worldbank.org/INTRES/Resources/469232-1107449512766/648083-1108140788422/23456_Table_on_systemic_and_non-systemic_banking_crises_January_21_2003.pdf, The World Bank, Washington DC (accessed on 12 Dec. 2018).

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., 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.

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

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., and Obstfeld, M., 2012, "Stories of the Twentieth Century for the Twenty-First", *American Economic Journal: Macroeconomics*, Vol. 4(1), pp. 226-65.

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.

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., 2017a, "Macrofinancial History and the New Business Cycle Facts", *NBER Macroeconomics Annual 2016*, Vol. 31, pp. 213-263.

Jordà Ò., Richter, B., Schularick, M., and Taylor, A.M., 2017b, "Bank Capital Redux: Solvency, Liquidity and Crisis", NBER Working Paper 23287.

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.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z., 2015, "Prediction policy problems", *American Economic Review*, vol. 105(5), pp. 491-95.

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.

Lindgren, C.J., Garcia, G., and Saal, M.I., 1996, *Bank Soundness and Macroeconomic Policy*. Washington, D.C.: International Monetary Fund.

López, V., Fernández, A., García, S., Palade, V., and Herrera, F., 2013, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics", *Information Sciences*, Vol. 250, pp. 113-141.

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

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

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., 2017, "When to Lean against the Wind", CEPR Working Paper 12188.

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.

Varian, H. R., 2014, "Big Data: New Tricks for Econometrics", *Journal of Economic Perspectives*, Vol. 28(2), pp. 3-28.

TABLES

Table 1 Definition of sub-periods

Pre-crisis	up to 3 years prior to the crisis
Post-crisis	up to 3 years after the crisis
Normal times	no crisis in the preceding 3 years and no crisis in the subsequent 3 years

Table 2 Systemic banking crises: advanced economies

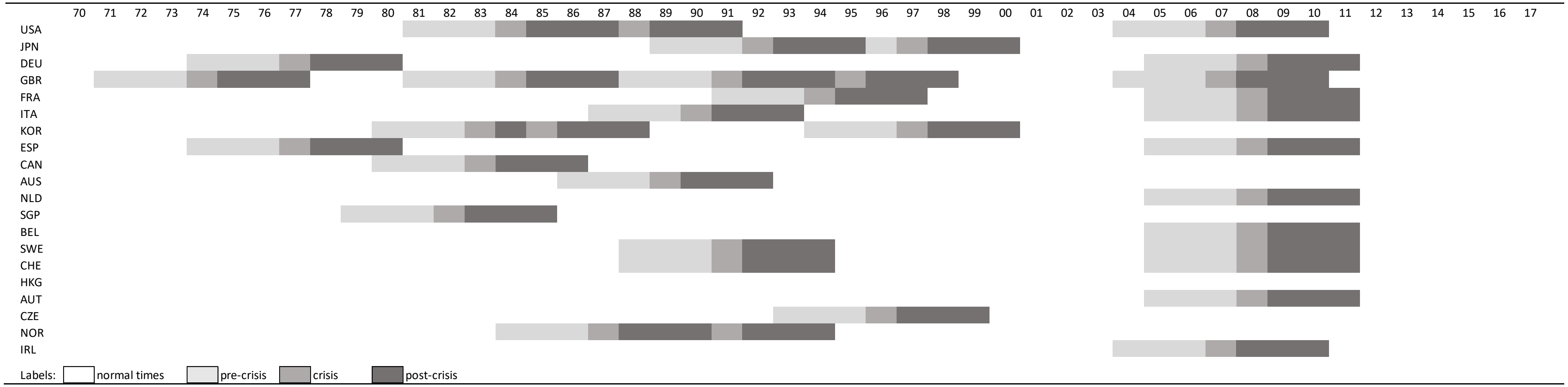


Table 3 Systemic banking crises: emerging economies

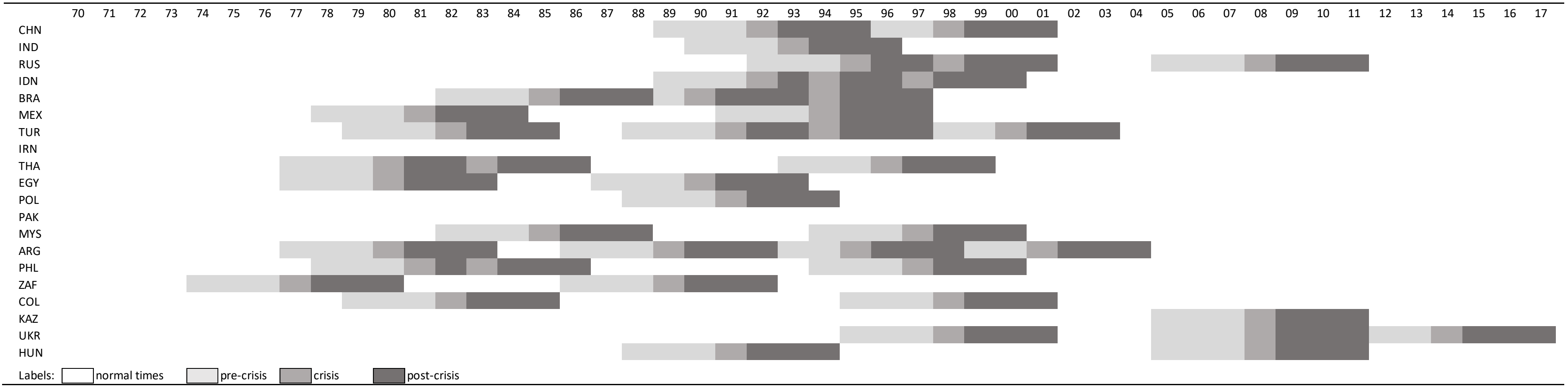


Table 4 Definitions of the target variable

	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 5 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											
Current Account-to-GDP	-0.02	-1.52	0.41	-1.43	-1.69	-2.18	-2.08	-2.44	-0.58	-1.4	3955
External Debt-to-GNI	164.29	45.70	149.21	43.40	202.94	48.56	205.20	53.31	213.81	60.5	3977
Public Debt-to-GDP	53.51	48.43	53.71	47.25	47.46	49.75	49.56	53.05	58.73	56.3	4078
Inflation	6.04	50.56	6.13	26.91	6.41	104.33	6.12	195.73	5.07	167.6	4159
Openness index	92.77	76.95	97.29	79.28	79.85	68.76	77.58	64.85	78.24	67.9	4150
Credit-to-GDP	124.41	39.36	120.92	39.45	128.58	38.50	136.64	42.83	140.24	38.1	3964
Bank Credit-to-Bank-Deposits	112.49	94.55	106.73	91.77	128.15	106.55	131.06	107.39	129.37	104.9	4088
House price	2.77	2.09	3.34	2.84	5.71	6.18	-1.78	-0.96	-2.13	-4.9	1823
Global											
10yr US Treasury rate	6.48	6.48	6.47	6.26	7.15	7.82	6.61	7.81	5.91	7.2	4800
Real world GDP growth	3.14	3.14	3.20	3.19	3.39	3.09	2.65	2.66	2.61	2.8	4800
Energy price index	8.35	8.35	9.01	9.10	10.58	7.33	9.22	2.31	1.21	3.7	4700
Detrended and standardised											
Current Account-to-GDP*	0.00	0.00	0.06	0.01	-0.25	-0.19	-0.41	-0.19	-0.02	0.13	3955
External Debt-to-GNI	0.00	0.00	-0.11	-0.02	0.28	-0.25	0.27	0.10	0.37	0.38	3977
Public Debt-to-GDP	0.00	0.00	0.08	0.00	-0.61	-0.39	-0.71	0.00	0.25	0.29	4078
Openness index	0.00	0.00	0.03	0.00	0.17	-0.02	0.04	-0.07	-0.34	0.07	4150
Credit-to-GDP	0.00	0.00	-0.15	-0.04	0.24	0.20	0.68	0.76	0.60	-0.03	3964
Bank Credit-to-Bank-Deposits	0.00	0.00	-0.11	-0.03	0.21	0.17	0.56	0.39	0.31	-0.03	4088

Source: Authors' own elaborations based on BIS, CBOE, IMF and WB.
Note: (*) Current account-to-GDP standardised only

Table 6 T-test on full sample averages

	Difference		t-stat
Current Account-to-GDP	1.50	***	4.34
External Debt-to-GNI	118.58	***	11.75
Public Debt-to-GDP	5.07	***	4.20
Inflation	-44.52	***	-4.09
Openness index	15.82	***	8.53
Credit-to-GDP	85.04	***	30.46
Bank Credit-to-Bank-Deposits	17.94	***	11.52
House price	0.69		1.21

Source: Authors' own elaborations based on BIS, CBOE, IMF and WB.
*Note: This table presents tests of differences in the means presented in **Table 5**. * p<0.05; ** p<0.01; ***p<0.001.*

Table 7 Marginal effects: Advanced economies, 1970-2017

	(1) crisis	(2) precrisis (2a)	(3) precrisis (2b)	(4) precrisis (2b)
Current account-to-GDP	0.003	-0.004	-0.000	-0.000
	(0.006)	(0.013)	(0.015)	(0.015)
External debt-to-GNI	0.027***	0.036***	0.048***	0.048
	(0.003)	(0.009)	(0.012)	(0.030)
Public debt-to-GDP	-0.034***	-0.090***	-0.095***	-0.095***
	(0.007)	(0.014)	(0.016)	(0.016)
Credit-to-GDP	0.010**	0.004	0.018	0.018
	(0.005)	(0.013)	(0.015)	(0.015)
10yr US Treasury rate	0.006***	0.011**	0.011*	0.011**
	(0.002)	(0.005)	(0.006)	(0.005)
Real world GDP growth	0.012***	0.024***	0.022***	0.022***
	(0.003)	(0.007)	(0.008)	(0.007)
*External debt-to-GNI*US 10yr Treasury*				0.000
				(0.005)
Country dummies	YES	YES	YES	YES
No. obs.	1,120	1,098	953	953
Pseudo R-squared	0.244	0.199	0.223	0.223
AUROC	0.845	0.795	0.802	0.801

Source: Authors' own elaborations

*Note: In Column (1) crisis = 1 identifies the year when the crisis occurs, 0 otherwise, and the set of explanatory variables are taken at t-1. In Column (2) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise. In Columns (3) and (4) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise except post-crisis years. From Columns (2) to (4) explanatory variables are taken at time t. All variables are detrended and standardised, with the exception of Current account-to-GDP (only standardised), 10yr US Treasury rate and Real world GDP growth. The AUROC is the area under receiving operating characteristic. Standard errors in brackets are clustered at the country level. * p<0.05; ** p<0.01; ***p<0.001.*

Table 8 Marginal effects: Emerging economies, 1970-2017

	(1) crisis	(2) precrisis (2a)	(3) precrisis (2b)	(4) precrisis (2b)
Current account-to-GDP	-0.007*	-0.010	-0.009	-0.010
	(0.004)	(0.009)	(0.011)	(0.010)
External debt-to-GNI	-0.000	-0.009	-0.003	-0.054***
	(0.007)	(0.013)	(0.015)	(0.020)
Public debt-to-GDP	-0.016**	-0.026*	-0.029**	-0.033**
	(0.007)	(0.013)	(0.015)	(0.015)
Inflation	0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
10yr US Treasury rate	0.006***	0.015***	0.020***	0.022***
	(0.001)	(0.002)	(0.003)	(0.003)
Real world GDP growth	0.001	0.006	0.004	0.003
	(0.003)	(0.004)	(0.004)	(0.004)
*External debt-to-GNI*US 10yr Treasury rate*				0.007***
				(0.002)
Region dummies	YES	YES	YES	YES
No. obs.	2,273	2,244	2,010	2,010
Pseudo R-squared	0.125	0.130	0.159	0.167
AUROC	0.794	0.785	0.803	0.808

Source: Authors' own elaborations

*Note: In Column (1) crisis = 1 identifies the year when the crisis occurs, 0 otherwise, and the set of explanatory variables are taken at t-1. In Column (2) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise. In Columns (3) and (4) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise except post-crisis years. From Columns (2) to (4) explanatory variables are taken at time t. All variables are detrended and standardised, with the exception of Current account-to-GDP (only standardised), Inflation, 10yr US Treasury rate and Real world GDP growth. The AUROC is the area under receiving operating characteristic. Standard errors in brackets are clustered at the region level. * p<0.05; ** p<0.01; ***p<0.001.*

Table 9 Confusion matrix

		Observed 0	Observed 1
Predicted	0	a_{00}	a_{01}
Predicted	1	a_{10}	a_{11}

- Accuracy = $\frac{a_{00}+a_{11}}{a_{00}+a_{01}+a_{10}+a_{11}}$

- Sensitivity (or True positive rate) = $\frac{a_{11}}{a_{01}+a_{11}}$

- Precision = $\frac{a_{11}}{a_{10}+a_{11}}$

Table 10 Logit model: Out of sample performance

	(a) advanced economies				(b) emerging economies		
	Average	Median	Std. Dev.		Average	Median	Std. Dev.
Sensitivity	0.226	0.222	0.077	Sensitivity	0.044	0.043	0.030
Precision	0.680	0.667	0.172	Precision	0.333	0.333	0.230
Accuracy	0.883	0.885	0.021	Accuracy	0.903	0.903	0.013
AUROC	0.740	0.740	0.050	AUROC	0.777	0.778	0.034

Source: Authors' own elaborations

*Note: Summary statistics of performance indicators on a total of 1,000 replications. Each AUROC is derived from estimating the logit model in Column (3) of **Table 7**. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5*

Source: Authors' own elaborations

*Note: Summary statistics of performance indicators on a total of 1,000 replications. Each AUROC is derived from estimating the logit model in Column (3) of **Table 8**. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5*

Table 11 Baseline AdaBoost: Out of sample performance

	(a) advanced economies				(b) emerging economies		
	Mean	Median	Std. Dev.		Mean	Median	Std. Dev.
Sensitivity	0.361	0.356	0.096	Sensitivity	0.131	0.128	0.056
Precision	0.675	0.667	0.127	Precision	0.438	0.429	0.159
Accuracy	0.889	0.890	0.021	Accuracy	0.904	0.905	0.013
AUROC	0.846	0.847	0.038	AUROC	0.815	0.817	0.029

Source: Authors' own elaborations
Note: Summary statistics of performance indicators on a total of 1,000 replications. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5.

Source: Authors' own elaborations
Summary statistics of performance indicators on a total of 1,000 replications. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5.

Table 12 Enlarged AdaBoost: Out of sample performance

	(a) advanced economies				(b) emerging economies		
	Mean	Median	Std. Dev.		Mean	Median	Std. Dev.
Sensitivity	0.427	0.426	0.101	Sensitivity	0.175	0.167	0.070
Precision	0.743	0.750	0.125	Precision	0.472	0.462	0.147
Accuracy	0.909	0.908	0.019	Accuracy	0.909	0.910	0.014
AUROC	0.885	0.890	0.039	AUROC	0.854	0.855	0.028

Source: Authors' own elaborations
Summary statistics of performance indicators on a total of 1,000 replications. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5.

Source: Authors' own elaborations
Summary statistics of performance indicators on a total of 1,000 replications. Precision, Sensitivity and Accuracy rates are calculated using a cutoff equal to 0.5.

Table 13 Enlarged AdaBoost: predicted probabilities

(a) Advanced economies				(b) Emerging economies				(b) Emerging economies (cont.)			
2006		2017		2006		2017		2006		2017	
Latvia	62%	Netherlands	36%	Croatia	58%	Jamaica	19%	Ecuador	24%	Iran	9%
Spain	55%	Greece	32%	Georgia	47%	Bosnia and Herzegovin	17%	Brazil	24%	Egypt	9%
Iceland	55%	United States	29%	Romania	44%	Croatia	16%	Armenia	23%	Mauritius	7%
Portugal	52%	Ireland	28%	Hungary	38%	Fiji	16%	China	23%	Poland	7%
United States	50%	New Zealand	28%	Indonesia	38%	Botswana	16%	Malaysia	23%	Indonesia	7%
United Kingdon	50%	Iceland	26%	Algeria	37%	Sri Lanka	16%	Paraguay	22%	Trinidad and Tobago	7%
Italy	49%	Lithuania	25%	Kazakhstan	36%	Panama	16%	Jamaica	22%	South Africa	7%
Denmark	48%	Austria	25%	Iran	35%	Syria	16%	Libya	22%	Belarus	7%
Lithuania	47%	Australia	24%	Belarus	34%	Malaysia	16%	Kuwait	22%	Azerbaijan	6%
Austria	46%	Germany	24%	Venezuela	32%	India	16%	Turkmenistan	21%	Chile	6%
Australia	46%	Norway	24%	Sri Lanka	32%	Paraguay	15%	Suriname	21%	Guatemala	6%
Germany	41%	Slovenia	24%	Argentina	31%	Venezuela	14%	Bulgaria	20%	Algeria	6%
Sweden	41%	Luxembourg	24%	Russia	31%	Seychelles	14%	Philippines	20%	Angola	6%
New Zealand	41%	Italy	23%	Swaziland	30%	Romania	13%	Barbados	19%	Lebanon	6%
Greece	39%	Spain	23%	South Africa	30%	Kazakhstan	13%	Equatorial Guinea	18%	Turkmenistan	6%
France	38%	Canada	22%	Turkey	30%	Colombia	12%	Morocco	17%	Georgia	6%
Israel	37%	France	22%	Azerbaijan	29%	Pakistan	12%	Peru	17%	Russia	6%
Luxembourg	37%	Finland	20%	Chile	29%	Jordan	11%	India	17%	Tunisia	5%
Canada	36%	Korea	19%	El Salvador	29%	Hungary	11%	Tunisia	17%	Gabon	5%
Slovenia	35%	Hong Kong SAR	19%	Pakistan	29%	Namibia	11%	Serbia	17%	Costa Rica	5%
Estonia	35%	Czech Republic	19%	Macedonia	29%	Swaziland	11%	Brunei	17%	El Salvador	5%
Netherlands	34%	Denmark	17%	Bosnia and Herzegovin	28%	Albania	10%	Fiji	16%	Armenia	5%
Slovak Republic	34%	Estonia	16%	Gabon	27%	Equatorial Guinea	10%	Uruguay	15%	Morocco	4%
Belgium	33%	Japan	16%	Trinidad and Tobago	27%	Libya	9%	Colombia	14%	Uruguay	4%
Ireland	31%	Slovak Republic	16%	Mexico	27%	Brunei	9%	Seychelles	14%	Philippines	4%
Finland	31%	Switzerland	16%	Angola	27%	Kuwait	9%	Namibia	14%	Thailand	4%
Norway	28%	Belgium	16%	Syria	27%	Barbados	9%	Mauritius	13%	China	4%
Hong Kong SAR	28%	United Kingdon	13%	Jordan	27%	Suriname	9%	Belize	12%	Serbia	4%
Singapore	26%	Sweden	13%	Albania	26%	Turkey	9%	Lebanon	12%	Macedonia	4%
Korea	26%	Latvia	12%	Costa Rica	25%	Argentina	9%	Poland	11%	Ecuador	3%
Switzerland	25%	Singapore	11%	Ukraine	25%	Dominican Republic	9%	Egypt	9%	Belize	3%
Japan	24%	Israel	11%	Botswana	24%	Mexico	9%	Panama	0%	Bulgaria	2%
Czech Republic	20%			Thailand	24%	Peru	9%			Brazil	1%

Source: Authors' own elaborations

FIGURES

Figure 1 Distribution of the sub-periods by quartile of selected macroeconomic indicators

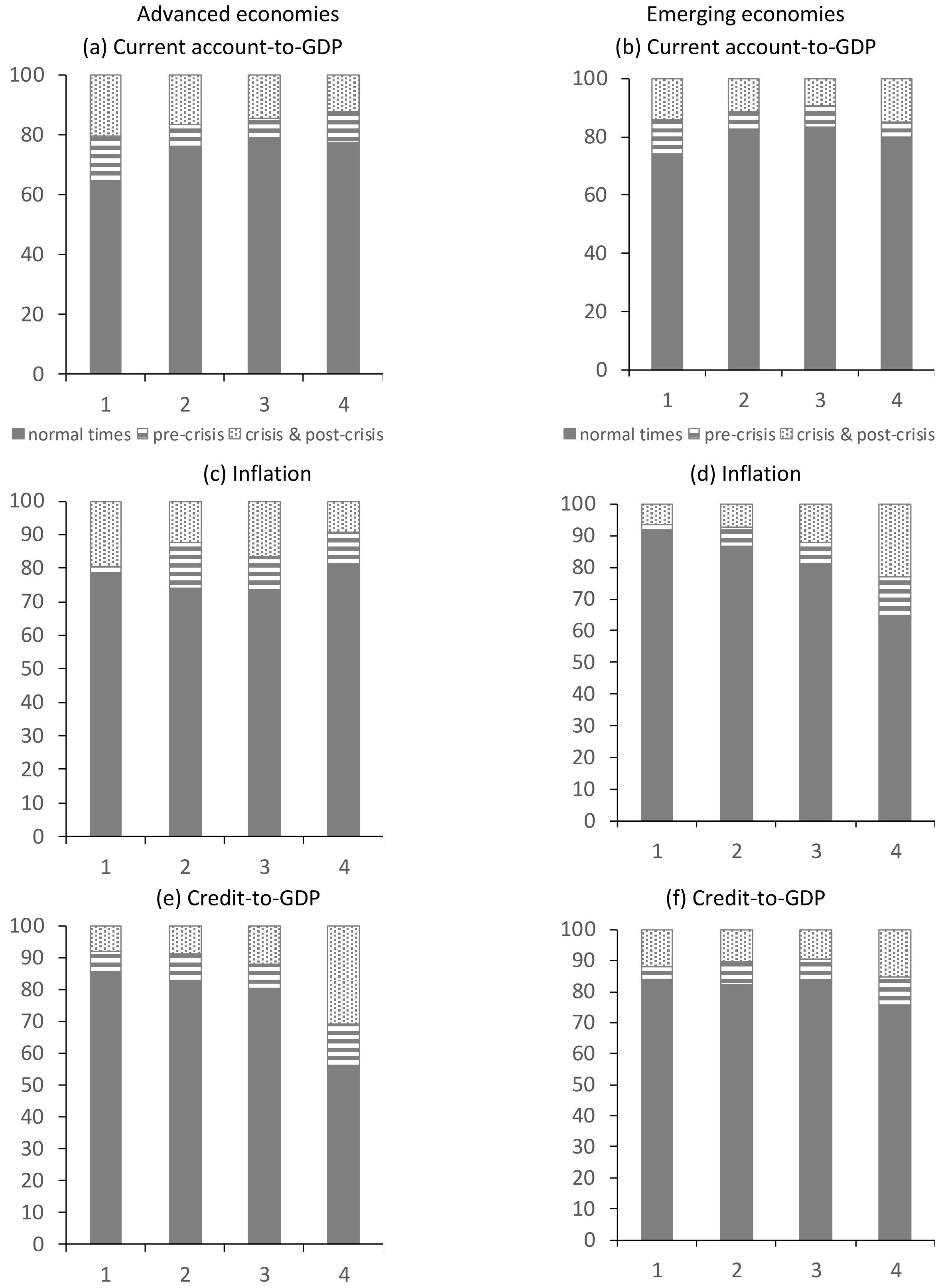
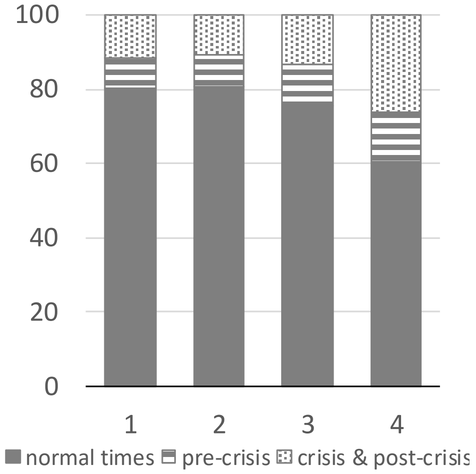


Figure 1 Distribution of the sub-periods by quartile of selected macroeconomic indicators (cont.)

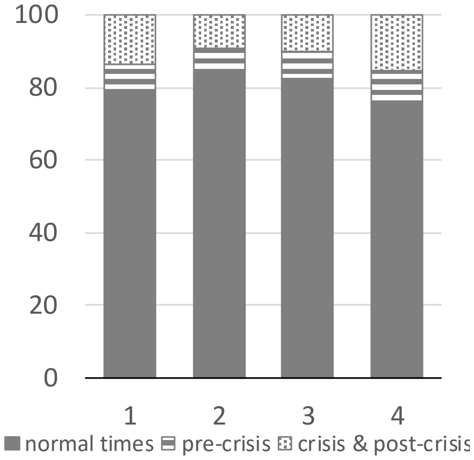
Advanced economies

(g) Bank credit-to-Bank deposits



Emerging economies

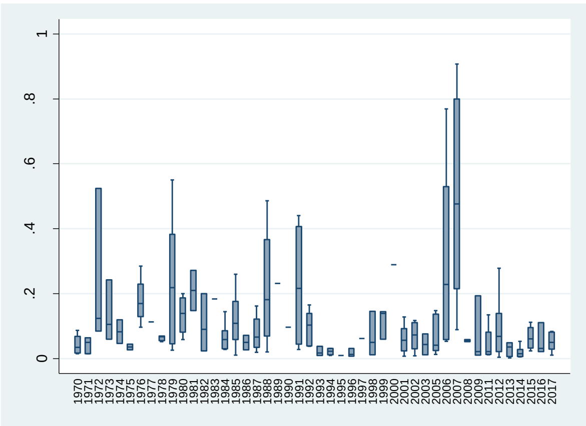
(h) Bank credit-to-Bank deposits



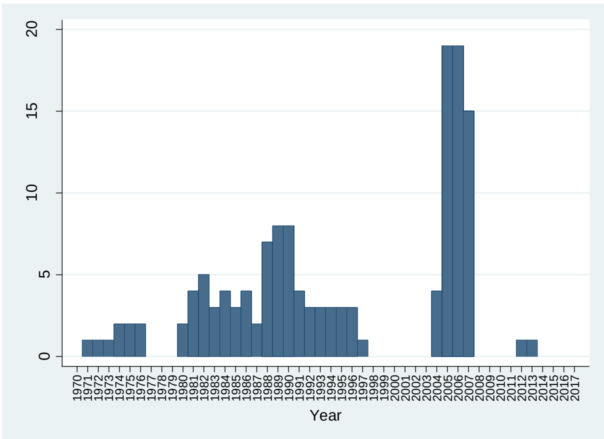
Source: Authors' own elaborations based on BIS, IMF and WB.
Note: This chart shows the distribution of the sub-periods by quartile of the selected variable.

Figure 2 Logit model: Predicted probabilities and number of pre-crisis spells, 1970-2017 (advanced economies)

(a) Distribution of predicted probabilities from logit model



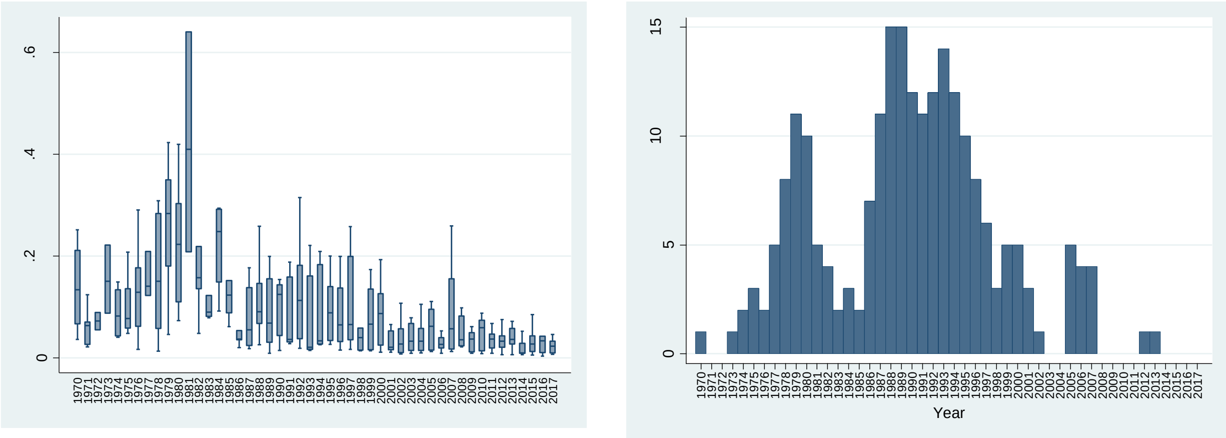
(b) Number of pre-crises



Source: Author's own elaborations
*Note: Predicted probabilities shown in panel (a) correspond to the out-of-sample predictions derived from estimating the logit model in Column (3) of **Table 7** 1,000 times. Probabilities obtained from each replication are then averaged by country and year and their distribution plotted in panel (a). They are based on a total of 191 out-of-sample observations.*

Figure 3 Logit model: Predicted probabilities and number of pre-crisis spells, 1970-2017 (emerging economies)

(a) Distribution of predicted probabilities from logit model (out-of-sample) *(b) Number of pre-crises*

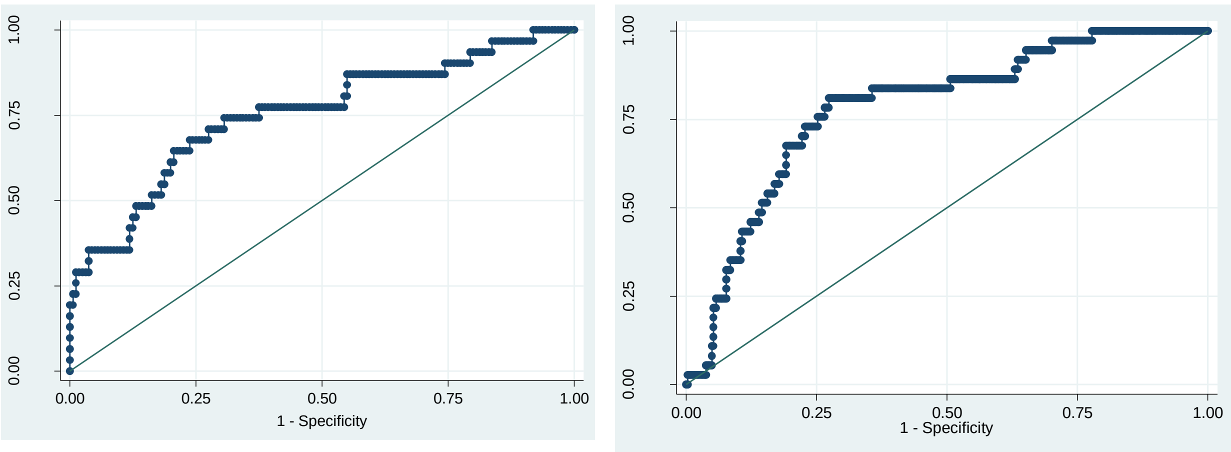


Source: Author's own elaborations

*Note: Predicted probabilities shown in panel (a) correspond to the out-of-sample predictions derived from estimating the logit model in Column (3) of **Table 8** 1,000 times. Probabilities obtained from each replication are then averaged by country and year and their distribution plotted in panel (a). They are based on a total of 402 out-of-sample observations.*

Figure 4 Logit model: ROC curves

(a) Advanced economies (b) Emerging economies

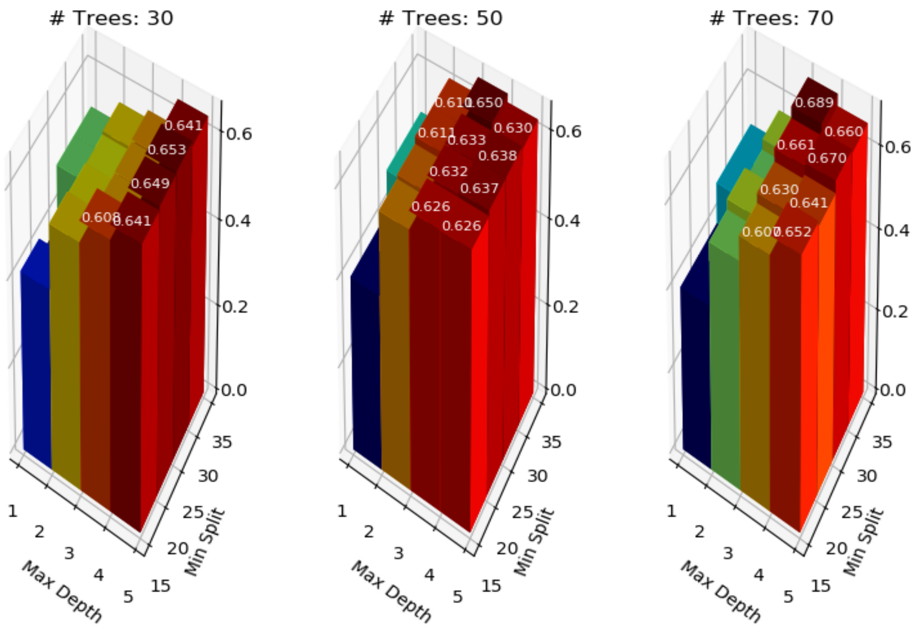


Source: Author's own elaborations.

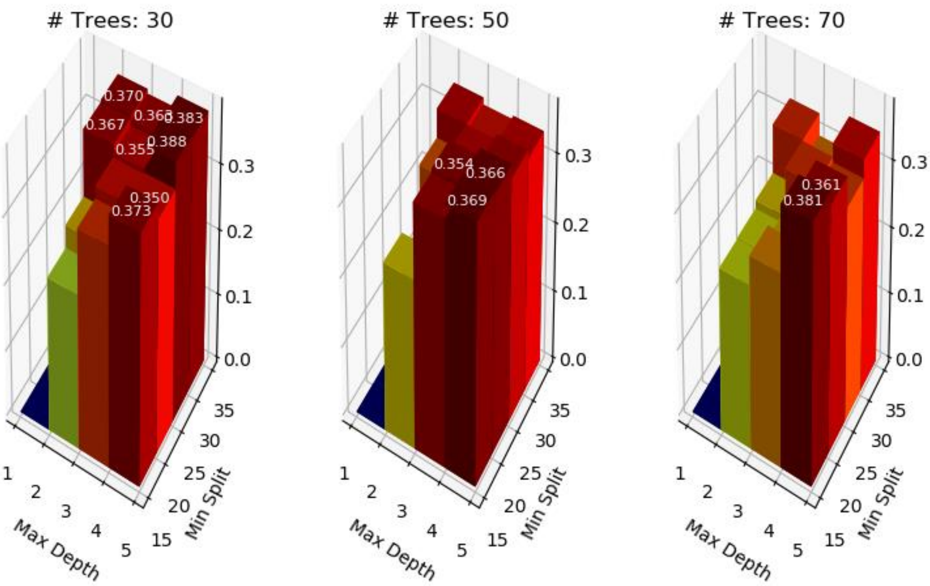
*Note: The ROC curve shows the combination of sensitivity (or true positive rate) – y-axis – and 1-specificity (or false positive rate) – x-axis - at various cutoff settings. ROC curves in panel (a) and (b) are derived from results of the logit model in Column (3) of **Table 7** and **Table 8**, respectively. In particular, from the 1,000 replications we choose the one which yields an AUROC nearest to the average values shown in **Table 10**. Panel (a) plots the ROC curve for the out-of-sample advanced economies. The corresponding AUROC is equal to 0.75. Panel (b) shows the ROC curve calculated for the out-of-sample emerging economies. The corresponding AUROC is 0.78.*

Figure 5 **Grid analysis for maximum depth, minimum split, number of trees and objective function**

(a) Advanced economies



(b) Emerging economies

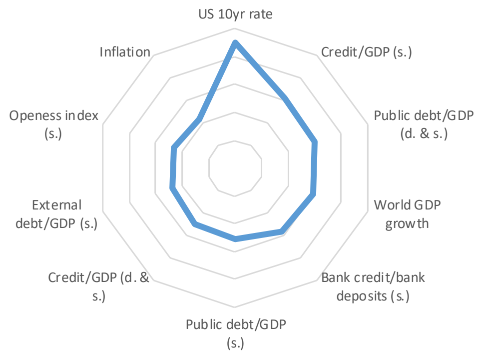


Source: Author's own elaborations.

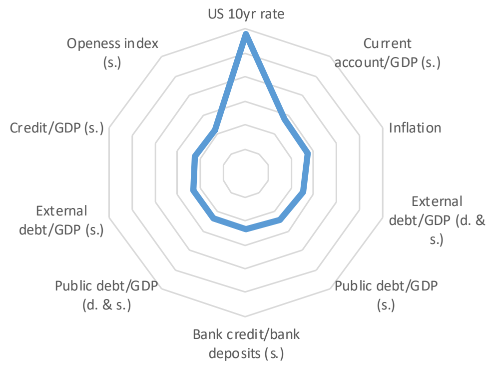
Note: Labels are as follows. Max Depth=maximum tree depth; Min Split=minimum split; MFin=number of trees. The bar height is the value of the objective, or utility, function.

Figure 6 Enlarged AdaBoost: Importance of variables

(a) Advanced economies



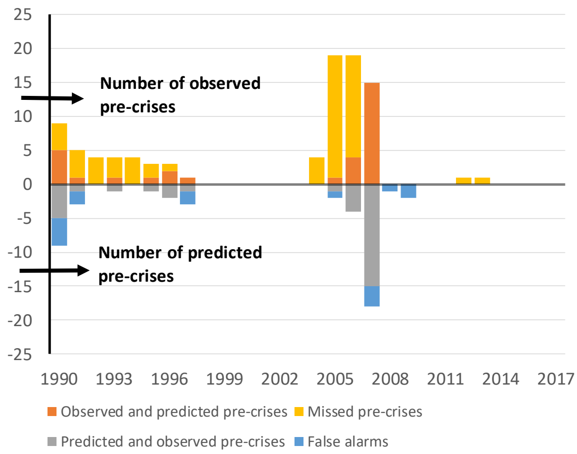
(b) Emerging economies



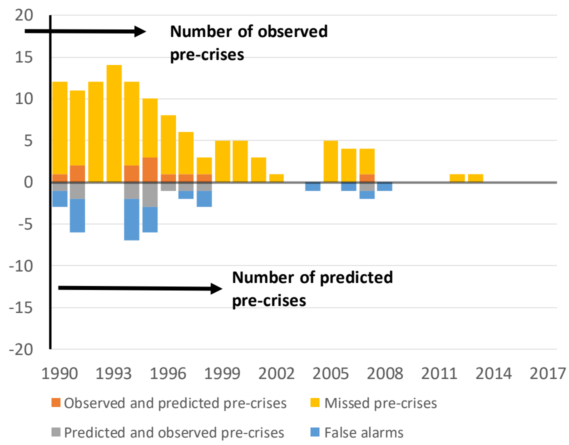
Source: Author's own elaborations.
Note: "s." stands for standardized, "d." stands for detrended and "s. & d." stands for detrended and standardized.

Figure 7 Enlarged AdaBoost: Forecasting ability

Advanced economies

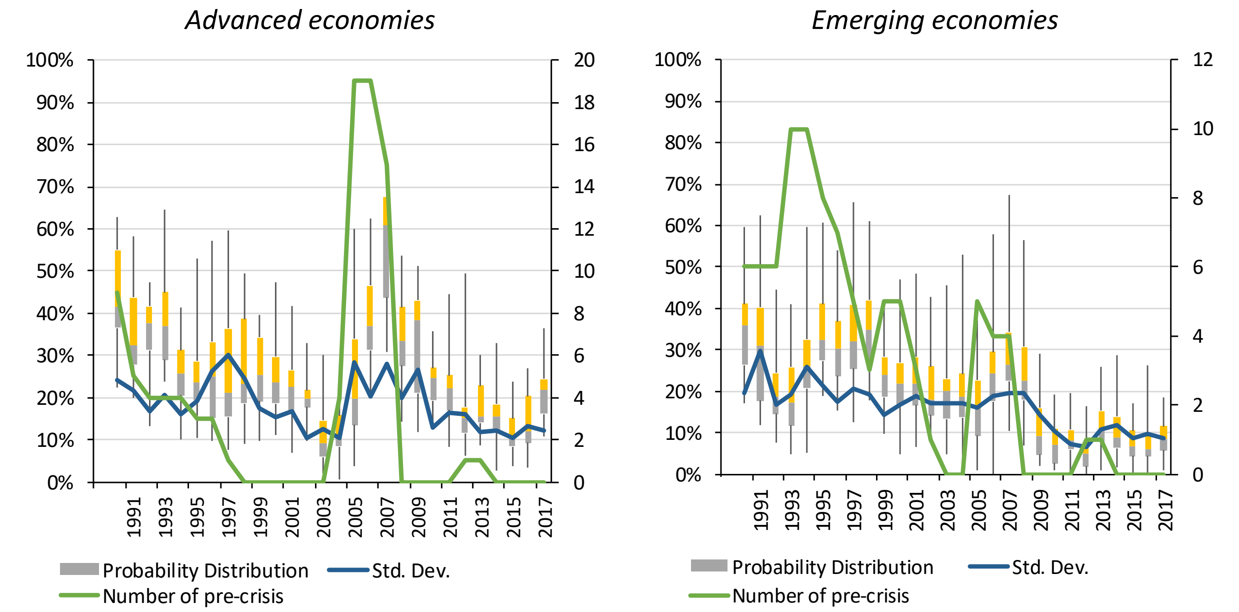


Emerging economies



Source: Author's own elaborations.
Note: the upper panels display the subsample of observed crises, split into those that are correctly predicted (orange bars) and those that are incorrectly classified (i.e. missed, yellow bars). The lower panels display the subsample of predicted crises, split into the ones that were observed (grey bars) and the ones that are incorrectly classified (i.e. false alarms, blue bars). The yellow and the blue bars inform on how often the models fail to predict an observed crisis and on how often the models predict a false crisis, respectively. The absolute values of the orange and the grey bars provide the same information, i.e. the observed crises that are correctly predicted.

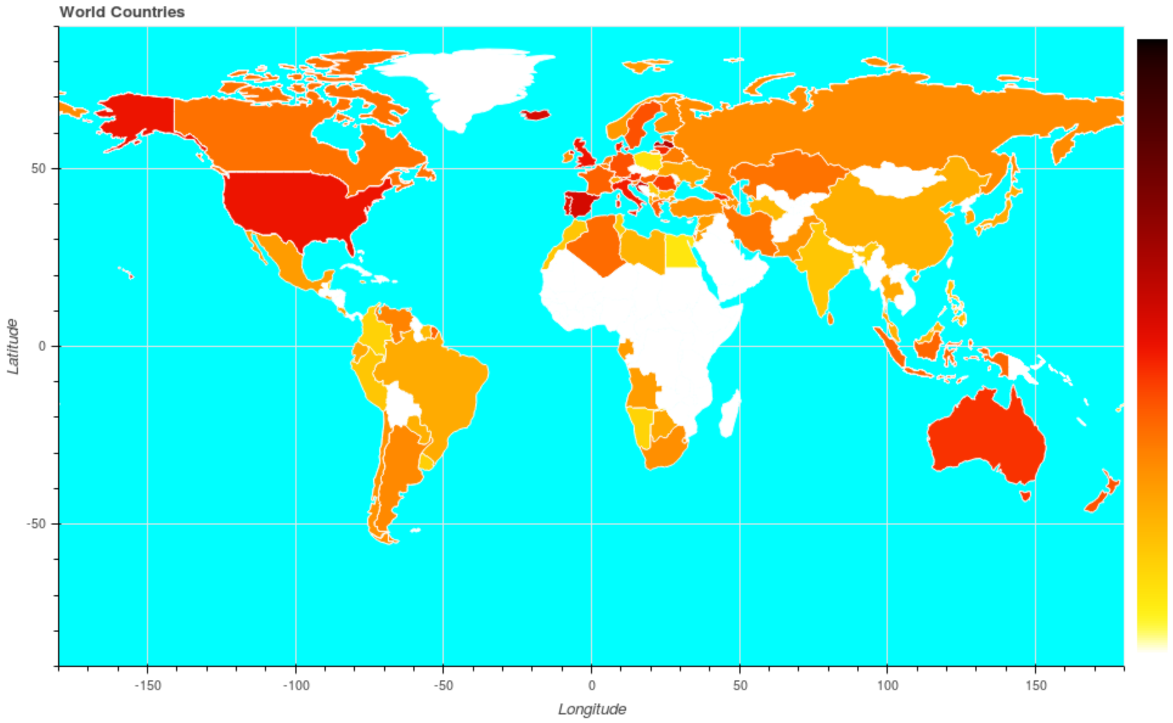
Figure 8 Enlarged AdaBoost: Pre-crisis probabilities



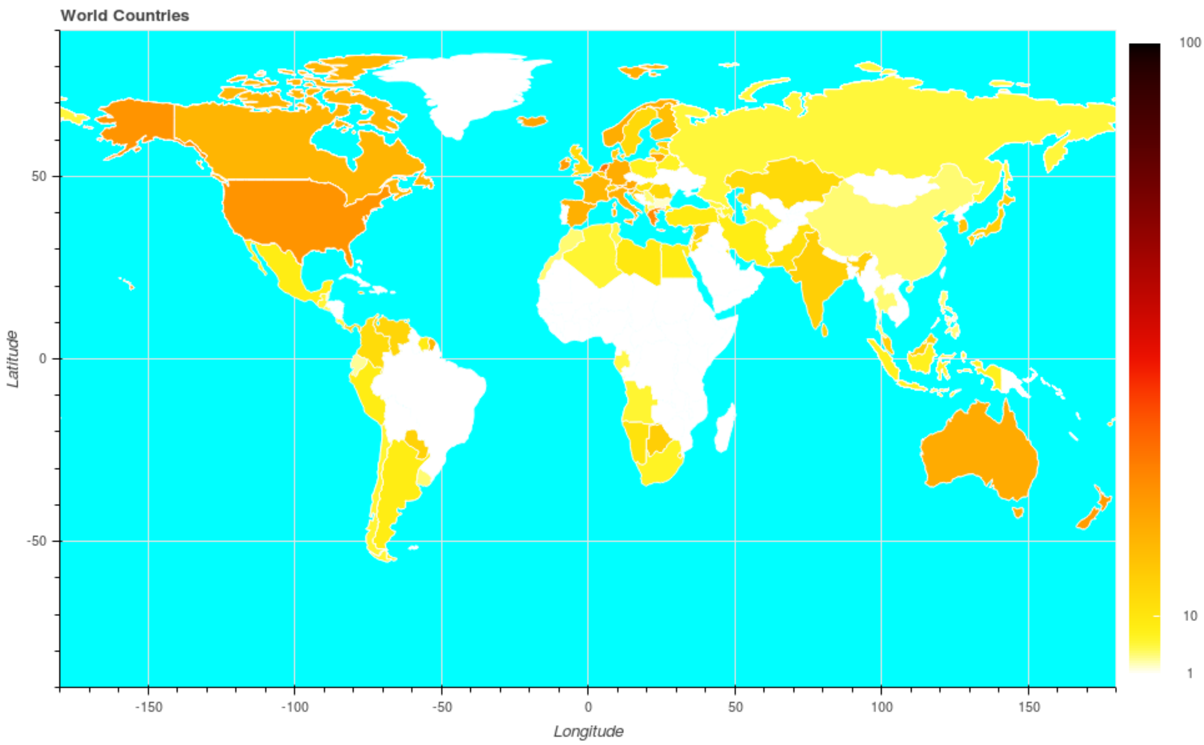
Source: Author's own elaborations.

Figure 9 Enlarged AdaBoost: Heat maps

2006



2017



Source: Author's own elaborations.

Note: Darker colours correspond to higher probabilities to be in a pre-crisis year.

Appendix A Data and definitions

Table A1 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. (2017a)	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. (2018)	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 A2 List of countries included in the dataset, 1970-2017

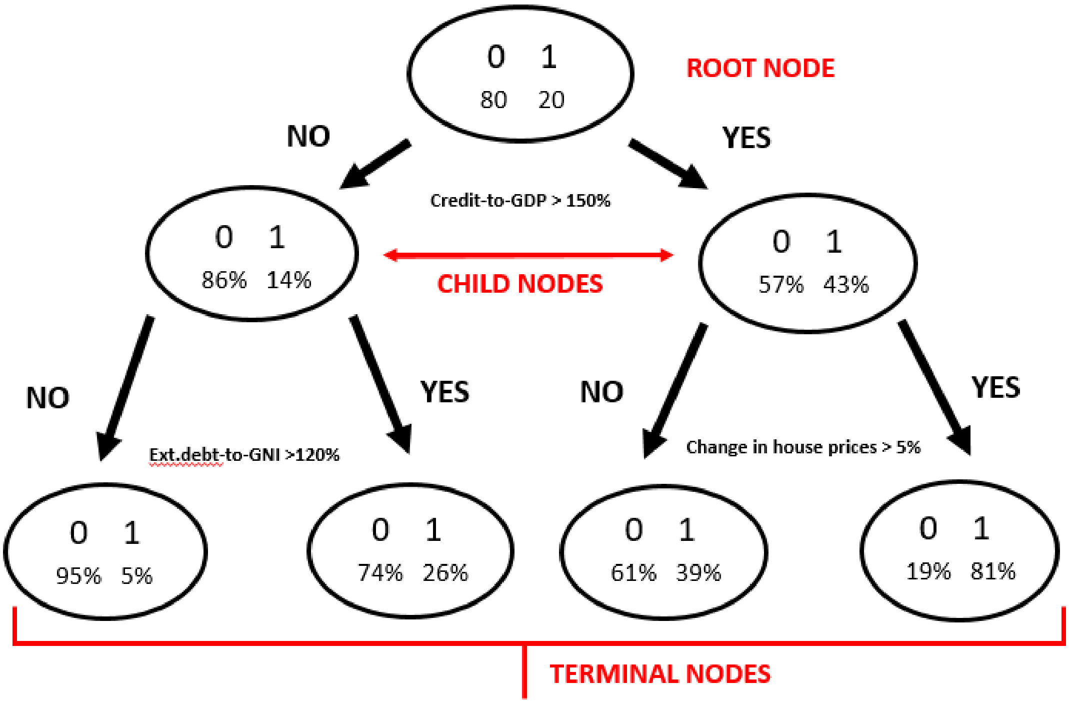
Advanced economies	Emerging economies
Australia Austria Belgium Canada Czech Republic Denmark Estonia Finland France Germany Greece Hong Kong SAR Iceland Ireland Israel Italy Japan Korea Latvia Lithuania Luxembourg Netherlands New Zealand Norway Portugal Singapore Slovak Republic Slovenia Spain Sweden 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 A3 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
Openness Index	Sum of exports and imports of goods and services, % of GDP	WB[(b)]
Credit-to-GDP	Credit to the private sector, % of GDP	BIS (WB when BIS data not available)
Bank Credit-to-Bank Deposits	Private credit by deposit money banks, % of demand, time and saving deposits	WB, Financial Development and Structure Dataset (updated July 2018) and IMF, International Financial Statistics
yoy % changes		
Inflation	GDP deflator, ratio of GDP in current local currency to GDP in constant local currency (yoy %)	WB[(b)]
House price	Real house price index (yoy %)	BIS, FRED, OECD, Cesa-Bianchi (2013)
Global		
Ratios		
10yr 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) (yoy %)	World Bank Commodity Price Data
Real World GDP growth	Annual percentage growth rate of world GDP at constant prices (yoy %)	WB[(b)]

Source: Authors' own elaborations based on BIS, CBOE, IMF and WB.
Notes: (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"

Figure A1 **Binary classification tree: an example**



Source: Author's own elaborations

Appendix B Logit model: Robustness analysis

Table B1 Marginal effects: Robustness for advanced economies, 1970-2017

	(1) crisis	(2) precrisis (2a)	(3) precrisis (2b)	(4) precrisis (2b)
Real GDP growth	0.005**	0.008	0.005	0.005
	(0.002)	(0.005)	(0.005)	(0.005)
Current account-to-GDP	0.001	-0.005	-0.002	-0.002
	(0.005)	(0.012)	(0.014)	(0.014)
External debt-to-GNI	0.028***	0.037***	0.047***	0.047*
	(0.004)	(0.009)	(0.012)	(0.028)
Public debt-to-GDP	-0.032***	-0.087***	-0.092***	-0.092***
	(0.007)	(0.014)	(0.016)	(0.016)
Credit-to-GDP	0.011**	0.005	0.017	0.017
	(0.005)	(0.013)	(0.015)	(0.015)
10yr US Treasury rate	0.005**	0.009*	0.010	0.010*
	(0.002)	(0.005)	(0.006)	(0.006)
Real world GDP growth	0.006*	0.020**	0.021**	0.021**
	(0.003)	(0.008)	(0.008)	(0.008)
*External debt-to-GNI*US 10yr Treasury rate*				0.000
				(0.005)
Country dummies	YES	YES	YES	YES
No. obs.	1,089	1,070	931	931
Pseudo R-squared	0.270	0.213	0.231	0.231
AUROC	0.862	0.801	0.804	0.804

Source: Authors' own elaborations

*Note: In Column (1) crisis = 1 identifies the year when the crisis occurs, 0 otherwise, and the set of explanatory variables are taken at t-1. In Column (2) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise. In Columns (3)-(5) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise except post-crisis years. From Columns (2) to (5) explanatory variables are taken at time t. All variables are detrended and standardised, with the exception of Current account-to-GDP (only standardised), Real GDP growth, 10yr US Treasury rate and Real world GDP growth. The AUROC is the area under receiving operating characteristic. Standard errors in brackets are clustered at the country level. * p<0.05; ** p<0.01; ***p<0.001.*

Table B2 Marginal effects: Robustness for emerging economies, 1970-2017

	(1) crisis	(2) precrisis (2a)	(3) precrisis (2b)	(4) precrisis (2b)
Real GDP growth	-0.001	-0.004**	-0.005***	-0.005***
	(0.001)	(0.001)	(0.002)	(0.002)
Current account-to-GDP	-0.007*	-0.011	-0.010	-0.011
	(0.004)	(0.009)	(0.011)	(0.011)
External debt-to-GNI	-0.001	-0.011	-0.005	-0.054***
	(0.006)	(0.013)	(0.014)	(0.02)
Public debt-to-GDP	-0.016***	-0.026**	-0.031**	-0.035**
	(0.006)	(0.012)	(0.014)	(0.014)
Inflation	0.000*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
10yr US Treasury rate	0.005***	0.014***	0.019***	0.021***
	(0.001)	(0.002)	(0.003)	(0.003)
Real world GDP growth	0.002	0.009**	0.008**	0.007*
	(0.003)	(0.004)	(0.004)	(0.004)
*External debt-to-GNI*US 10yr Treasury rate*				0.007***
				(0.002)
Region dummies	YES	YES	YES	YES
No. obs.	2,273	2,244	2,010	2,010
Pseudo R-squared	0.128	0.139	0.172	0.180
AUROC	0.803	0.795	0.812	0.817

Source: Authors' own elaborations

*Note: In Column (1) crisis = 1 identifies the year when the crisis occurs, 0 otherwise, and the set of explanatory variables are taken at t-1. In Column (2) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise. In Columns (3)-(5) pre-crisis = 1 identifies the 3 years preceding a crisis, 0 otherwise except post-crisis years. From Columns (2) to (5) explanatory variables are taken at time t. All variables are detrended and standardised, with the exception of Real GDP growth, Current account-to-GDP (only standardised), Inflation, 10yr US Treasury rat and Real world GDP growth. The AUROC is the area under receiving operating characteristic. Standard errors in brackets are clustered at the region level. * p<0.05; ** p<0.01; ***p<0.001.*

Appendix C AdaBoost: Robustness analysis

We perform a series of robustness checks to validate our AdaBoost results on the "enlarged" variable set: (i) pre/post crisis definition, (ii) cutoff probabilities, (iii) oversampling of the crisis events, (iv) detrending procedure, and (v) comparison with RF. For easiness of reading, **Table C1** replicates **Table 12** of the main text with the main results of our model specification (3 years of pre-crisis and 3 years of post-crisis).

(i) Pre/Post crisis definition. In the analysis presented in the main text, pre- and post-crisis spells are defined on a 3-year basis. In **Table C2,** we show the model performance by changing the spell length, which can now take on also the values -2 and -1 for the pre-crisis spells and +1 and +2 for the post-crisis ones. For the advanced economies, our model specification delivers the highest sensitivity rate (0.427). With reference to precision, there are combinations that deliver higher values than our model (0.743), ranging between 0.754 (-2/+2 years) and 0.759 (-3/+2 years). For the emerging economies, instead, sensitivity and precision are at their highest in our specification. Overall, our model performance is the best we could achieve.

(ii) Cutoff probabilities. AdaBoost classifies crisis events based on a cutoff probability. Our baseline scenario assumes a cutoff of 0.5. In **Table C3**, we test the model with two additional cutoff values, 0.3 and 0.15. We observe that the lower the cutoff probability the lower the precision and the higher the sensitivity. Despite an improvement in the sensitivity rate, the loss in precision is remarkable. We therefore prefer the 0.5 cutoff.

(iii) Oversampling of the crisis events. Our dataset is unbalanced since the 0s heavily outnumber the 1s. This could limit the ability of the AdaBoost model to learn from the data. The literature (e.g. Lopez et al., 2013) suggests oversampling to balance the 0s and the 1s. To this end, we randomly replicate the 1s until the dataset is balanced. **Table C4** shows that oversampling helps increase sensitivity at the cost of a strong decrease in precision: prediction of pre-crises improves, but at the same time we predict a larger number of false alarms.

(iv) Detrending procedure. Our dataset includes detrended variables. We perform a robustness check using the one-side HP-filter and compare the results with those obtained with two-side HP-filter employed in our main analysis. **Table C5** shows no significant differences for the advanced economies, while results worsen considerably for the emerging economies.

(v) *Random Forest vs AdaBoost.* Herein, we compare the performance of the RF with that of the AdaBoost by means of performance indicators. The aim is to justify our choice of using the AdaBoost as our preferred machine learning algorithm. **Table C6** shows performance indicators of a RF trained using 35 trees (the same number of trees used to train the AdaBoost). Comparison of **Table C1** with **Table C6** suggests that overall the AdaBoost outperforms the RF. Specifically, the AdaBoost delivers a better out of sample performance than the RF in terms of AUROC. Across all performance indicators, the AdaBoost performs better than the RF for both country groups. These results prove that the AdaBoost is a better classifier compared to the RF.

Table C1 Enlarged AdaBoost: Out of sample performance

(a) advanced economies				(b) emerging economies			
	Mean	Median	Std. Dev.		Mean	Median	Std. Dev.
Sensitivity	0.427	0.426	0.101	Sensitivity	0.175	0.167	0.070
Precision	0.743	0.750	0.125	Precision	0.472	0.462	0.147
Accuracy	0.909	0.908	0.019	Accuracy	0.909	0.910	0.014
AUROC	0.885	0.890	0.039	AUROC	0.854	0.855	0.028

Source: Authors' own elaborations

*This table replicates **Table 12** in the main text.*

Table C2 AdaBoost: Robustness to Several Pre/Post crisis timeframe definitions (Pre: +1,+2,+3; Post: +1,+2,+3): Average measures

(a) advanced economies				(b) emerging economies			
Sensitivity				Sensitivity			
Pre\Post	1	2	3	Pre\Post	1	2	3
1	0.244	0.267	0.297	1	0.039	0.040	0.042
2	0.379	0.409	0.415	2	0.104	0.098	0.107
3	0.419	0.424	0.427	3	0.151	0.145	0.175
Precision				Precision			
Pre\Post	1	2	3	Pre\Post	1	2	3
1	0.622	0.669	0.728	1	0.215	0.213	0.225
2	0.694	0.754	0.736	2	0.406	0.363	0.386
3	0.757	0.759	0.743	3	0.436	0.448	0.472
Accuracy				Accuracy			
Pre\Post	1	2	3	Pre\Post	1	2	3
1	0.940	0.935	0.942	1	0.963	0.963	0.964
2	0.913	0.911	0.913	2	0.932	0.930	0.929
3	0.891	0.885	0.909	3	0.907	0.907	0.909

Source: Authors' own elaborations

Table C3 AdaBoost: Robustness to different cut-off probability thresholds

(a) advanced economies			(b) emerging economies		
	Cutoff = 0.30	Cutoff = 0.15		Cutoff = 0.30	Cutoff = 0.15
Sensitivity	0.777	1	Sensitivity	0.385	0.576
Precision	0.323	0.190	Precision	0.170	0.168
Accuracy	0.738	0.366	Accuracy	0.767	0.705

Source: Authors' own elaborations

Table C4 AdaBoost: Robustness to oversampling methods

	(a) advanced economies				(b) emerging economies		
	Average	Median	Std. Dev.		Average	Median	Std. Dev.
Sensitivity	0.520	0.521	0.080	Sensitivity	0.271	0.256	0.078
Precision	0.600	0.605	0.073	Precision	0.179	0.176	0.045
Accuracy	0.891	0.893	0.038	Accuracy	0.905	0.906	0.018

Source: Authors' own elaborations

Table C5 AdaBoost: Robustness to detrending with one-side HP-filter

	(a) advanced economies				(b) emerging economies		
	Average	Median	Std. Dev.		Average	Median	Std. Dev.
Sensitivity	0.313	0.313	0.073	Sensitivity	0.038	0.011	0.063
Precision	0.650	0.651	0.103	Precision	0.113	0.101	0.021
Accuracy	0.778	0.777	0.038	Accuracy	0.905	0.906	0.011

Source: Authors' own elaborations

Table C6 Random Forest: Performance indicators

	(a) advanced economies				(b) emerging economies		
	Average	Median	Std. Dev.		Average	Median	Std. Dev.
Sensitivity	0.378	0.375	0.092	Sensitivity	0.135	0.133	0.052
Precision	0.727	0.727	0.121	Precision	0.478	0.467	0.159
Accuracy	0.895	0.895	0.020	Accuracy	0.906	0.908	0.013
AUROC	0.821	0.821	0.018	AUROC	0.773	0.774	0.016

Source: Authors' own elaborations

Appendix D Advanced Economies – Enlarged AdaBoost vs Alternative AdaBoost models

Table D1 Performance indicators by varying the set of variables and the cutoff

		Cutoff = 0.5	
	Enlarged	Build-up	Large build-up
Sensitivity	0.427	0.441	0.451
Precision	0.743	0.747	0.757
Accuracy	0.909	0.910	0.912
ROC	0.885	0.886	0.899
		Cutoff = 0.45	
	Enlarged	Build-up	Large build-up
Sensitivity	0.566	0.570	0.592
Precision	0.608	0.608	0.624
Accuracy	0.899	0.899	0.903
ROC	0.886	0.886	0.899

"Enlarged" = enlarged AdaBoost
"Build-up" = enlargded AdaBoost with build-up variables
"Large build-up" = "Build-up" + original country-specific variables

Having established that the AdaBoost outperforms the logit model, we further test its predictive performance on the "alternative" model introduced in Section 7. The AdaBoost slightly improves its performance in terms of all indicators compared to the Enlarged model (**Table D1**).

Including the "build up" variables allows us to increase the sensitivity rate by about 2 percentage points ("Build up" column), reaching a sensitivity rate equal to 0.441. When we also include the variables in their original form ("Large build-up"), we obtain an additional improvement of 1 percentage point (0.451). The precision slightly improves as well.

In addition, as the precision rate signals a low propensity to issue false alarms, we reduce the cutoff used to discriminate between 0s and 1s from 0.5 to 0.45. We obtain a significant increase in the sensitivity rate, reaching in the large build-up specification a sensitivity rate of almost 0.6. This comes at the cost of a reduction in the precision rate, which moves from 0.757 to 0.625. In other words, this model is able to predict about 2 out of 3 pre-crisis periods while, when the model predicts a pre-crisis, the probability to issue a false alarm is just above 1/3.

We view this setup as satisfactory since we are able to give more emphasis on correctly predicting pre-crisis episodes at the cost of incurring a higher risk of issuing false alarms. However, since the number of false alarms may be over-estimated because of the prompt activation of macro-prudential policies (and therefore pre-crises are not recorded), our choice of accepting a lower precision rate seems reasonable.