



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Early warning models for systemic banking crises: can political indicators improve prediction?

Tran Huynh | Silke Uebelmesser

JENA ECONOMIC RESEARCH PAPERS · # 2022-007

EARLY WARNING MODELS FOR SYSTEMIC BANKING CRISES: CAN POLITICAL INDICATORS IMPROVE PREDICTION? *

Tran Huynh[†]

Silke Uebelmesser[‡]

May 2022

Abstract

This study provides the first attempt to evaluate whether a logit early warning system (EWS) for systemic banking crises can produce better predictions when political indicators are used alongside traditional macro-financial indicators. Based on a dataset covering 32 advanced economies for the period 1975-2017, we show that the inclusion of political indicators helps improve the predictive performance of the model. While the improvement is small, it is statistically significant and consistent for several different performance measures and robustness tests. Among the newly employed political variables, variables indicating the political ideology of the ruling party and the time in office of the incumbent chief executive show significant correlations with the likelihood of systemic banking crises. The results suggest that a systemic banking crisis is less likely when the government is left-wing and when the chief executive officer has been in office longer.

JEL Classification: C35, C53, E60, F37, G01, G28.

Keywords: early warning systems, systemic banking crises, vulnerability, political indicators, macro-financial indicators.

* We would like to thank Volker Clausen, Oliver Holtemoeller, Michael Kvasnicka, Fabian Koenings, Xiang Li, Huyen Nguyen, Philipp Poyntner, Gregor von Schweinitz, Bernd Suessmuth, Roland Winkler, Christoph Wunder, and participants at the annual conference in economics of the Ruhr Graduate School, the annual workshop of the Central German Doctoral Program in Economics, and the economic research workshop at the Friedrich Schiller University of Jena for helpful comments and discussions. Tran Huynh gratefully acknowledges the funding by the Free State of Thuringia through the Graduate Programme 'The Economic Impact of Digital Transformation'.

[†] Friedrich Schiller University Jena, Bachstrasse 18k, 07743 Jena, Germany. Email: tran.huynh@uni-jena.de. Corresponding author.

[‡] Friedrich Schiller University Jena and CESifo, Carl-Zeiss-Str. 3, 07743 Jena, Germany. Email: silke.uebelmesser@uni-jena.de.

1. Introduction

History is a vast early warning system.

— Norman Cousins (1915-1990)

Throughout the course of financial history, the repeated occurrence of systemic banking crises, especially the 2007/2008 global credit crunch and its aftermath, has increased the need to understand the causes of such events and to devise mechanisms that can help prevent them. Early warning systems (EWSs) as a tool for monitoring the banking system's stability have therefore garnered significant attention from academia and policymakers (Kaminsky and Reinhart, 1999; Bussiere and Fratzscher, 2006; Alessi and Detken, 2018). Like many other predictive models with binary outcomes (warning vs. no warning), these early warning models often face a trade-off between missing crises and issuing false alarms. Given that misleading policy interventions following a false alarm could be as devastating to an economy as failing to anticipate an impending crisis, much research has been done to improve the prediction accuracy of EWSs.

This paper aims at investigating whether incorporating political indicators in a logit EWS model, alongside the traditional macro-financial indicators, can help improve its predictive performance. The related literature has hitherto developed in two directions for the same purpose. One strand of the literature mainly focuses on the methodological approach of EWSs, which ranges from the signal extraction method (Kaminsky and Reinhart, 1999; Drehmann et al., 2011) and discrete choice (logit/probit) models (Demirgüç-Kunt and Detragiache, 1998; Bussiere and Fratzscher, 2006) to machine learning techniques (Tanaka et al., 2016; Alessi and Detken, 2018). The fact that there is more than one approach to the early-warning task has further driven performance comparisons across methods. Notable is the work of Holopainen and Sarlin (2017), in which it is shown that EWSs based on novel machine learning algorithms can produce better predictive performance than those built on conventional methods (such as logit models and signal extractions). Nonetheless, the reverse result is found in the study of Beutel et al. (2019), in which the authors show that the new approach, while it often outmatches others in in-sample performance evaluations, performs poorly in out-of-sample exercises compared to logit models due to the so-called overfitting problem. Since each approach has its own strengths and drawbacks, and comparisons are further complicated by the different settings across studies,¹ it has not been decided which method is superior to others on every crisis-predicting task.

The other strand of the literature, motivated by the influential works of Allen and Gale (2000), Laeven and Valencia (2008), Reinhart and Rogoff (2009) and Schularick and Taylor (2012) gives more weights to the history of financial crises as well as the theories that can help understand the nature and the causes of these “rare events”. This research, therefore, contributes to the improvement of EWSs by identifying potentially good early-warning indicators. Until now, most of the focus has been put on the connection between the banking sector's vulnerability and macroeconomic imbalances, prominently the crucial roles of monetary policy and leverage cycles on the unfolding of systemic banking crises. Still, more often than not, economic policies are influenced by the political environment. The dynamic behavior of macroeconomic factors (such as inflation, GDP growth, money growth, taxes, government spending, etc.) during elections and changes of governments has motivated a

¹ Recent studies on EWSs have been developed based on various crisis databases, using different sets of indicators, and different evaluating procedures. The way in which post-crisis periods are treated also differs across studies.

growing literature on “political cycles” (see e.g., Nordhaus, 1975; Rogoff and Sibert, 1988 and Alesina and Roubini, 1992).

Despite this fact, the connection between the political environment and financial crises has often been neglected, and the literature on the link between the two is scarce. Manasse and Roubini (2009), for example, show that political uncertainty can increase the likelihood of a sovereign debt crisis. McCarty et al. (2013) and Calomiris and Haber (2014) shed light on the political origins of financial crises in case studies. More recently, Herrera et al. (2020) prove that political booms (i.e. rises in governments’ popularity) are good and important predictors of banking crises in emerging markets. Nguyen et al. (2020) stress the importance of capturing the political environment as a source of uncertainty in explaining financial crises in both advanced and developing economies. They examine the impact of several political factors (such as electoral cycles, political ideologies, governments’ majority and stability) on the likelihood of all types of financial crises (namely banking crises, currency crises, sovereign debt crises, as well as twin and triple crises). Nevertheless, the study of Nguyen et al. (2020) only suggests preliminary links between political factors and the occurrence of crisis events as the empirical model used is not that of an EWS. Furthermore, its set of independent variables is chosen such that only common indicators for all crisis types are included. Therefore, some relevant explanatory variables for specific types of crises have thus been omitted, potentially resulting in the confounding estimates of both political and economic variables.

Against the above background, this paper finds itself at the crossing of the two mentioned literature strands. On the one side, it advances the study of Nguyen et al. (2020) to investigate whether variables indicating the political ideologies of incumbent governments, the timing of political elections, and the time in office of chief executives are potential early-warning indicators for systemic banking crises, even though they have been omitted so far. On the other side, to achieve the stated goal, this paper follows the framework laid out in the methodologically related strand of the literature. This allows a thorough comparison of the predictive performances of two logit EWSs– one with political indicators and one without political indicators. The obtained comparative results can then be used to answer the question of whether the included political variables are potentially good crisis predictors.

Using an integrated dataset covering systemic banking crisis events as well as economic and political data for 32 advanced economies over the 1975-2017 period, this paper finds that incorporating political indicators to an EWS helps improve its predictive performance. The improvement, albeit small, is statistically significant and consistent for several different performance measures and robustness tests. Moreover, the paper shows that among the newly employed political variables, those indicating the political ideology of the ruling party and the time in office of the incumbent chief executive are significantly correlated with the likelihood of systemic banking crises. The results suggest that a systemic banking crisis is less likely when left-wing governments are in office, and when the time in office of the chief executives is longer.

This paper contributes to the existing literature in various ways. First and most importantly, it provides the first attempt to evaluate whether an EWS for systemic banking crises can perform better when political indicators are incorporated. The obtained results not only suggest one way to improve the performance of EWSs but also highlight the link between the political environment and systemic banking crises. Second, based on earlier empirical evidence, it proposes a relatively comprehensive set of macro-financial indicators with substantial predictive powers. Third, since the two EWSs involved only differ as far as political indicators are included, the evaluation strategy is designed in a suitable way. Specifically, after constructing the two logit EWSs, this paper proposes the usage of goodness-of-fit tests and

likelihood statistics to evaluate how good the two EWSs are in terms of classification models (i.e., models that classify inputs into different categories). This step is important to any model selection task where the models have different degrees of freedom. It then follows the literature and performs both in-sample and out-of-sample evaluations, with the latter using cross-validation techniques and quasi real time exercises. Furthermore, this paper also uses hypothesis tests and several variations of the original dataset as robustness checks.

The remainder of the paper proceeds as follows. Section 2 describes the dataset, which includes crisis events as well as macro-financial and political indicators. Section 3 presents the methodology, the predictive performance measures, and the evaluating strategies used to build and compare the two logit EWSs of interest. Section 4 analyzes the obtained empirical results on the predictive performances. Economic interpretation and policy implication of the main findings are discussed in Section 5. Section 6 concludes.

2. Data

The dataset used in this paper covers 32 advanced economies including 24 European countries and 8 non-European developed countries (see Table A.1 in the Appendix). The focus on advanced economies leads to a more homogeneous setting as there are considerable differences between advanced and emerging economies regarding their macroeconomic and political environments. The dataset is yearly and covers the period 1975-2017. It can be divided into three parts: (i) banking crisis events, (ii) macro-financial indicators, and (iii) political indicators.

2.1. The Banking Crises Dataset

There are two crisis databases often used in the EWSs literature. The first is the IMF's crisis database initially built by Laeven and Valencia (2008) and sequentially updated in Laeven and Valencia (2013, 2018, and 2020). The second has been established by the European System of Central Banks (ESCB) and the European Systemic Risk Board (ESRB), reported in Lo Duca et al. (2017). These crisis datasets are often complemented by compiling additional crisis events from other influential papers or by consulting country experts in order to better serve the studies' purposes (see e.g., Babecky et al., 2014; Detken et al., 2014). Since this paper studies a wide range of advanced economies both inside and outside of Europe, it thus uses the worldwide crisis dataset of Casabianca et al. (2019), which combines the crisis data of Laeven and Valencia (2018) with additional crisis episodes retrieved from Reinhart and Rogoff (2009) and Jordà et al. (2017).² Besides, the latest database of Laeven and Valencia (2020) is used as a robustness check. A full list of crisis events can be found in Table A.1 in the Appendix.

² The banking crisis dataset of Laeven and Valencia (2008, 2013, 2018, 2020) is constructed based on one of the most comprehensive definitions of a systemic banking crisis. Specifically, the authors classify a banking crisis as systemic if the two following criteria are met: (i) there are "significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations)"; (ii) there are "significant banking policy intervention measures in response to significant losses in the banking system". Though hitting the first criterion would be sufficient to deem a banking crisis as systemic if the losses in the banking sector and/or liquidations are severe, that is, when either "... significant losses result in a share of nonperforming loans above 20% of total loans, or bank closures of at least 20% of banking system assets, or ... fiscal restructuring costs of the banking sector are 4 sufficiently high, exceeding 5% of GDP". In addition, regarding the second criterion, policy interventions are considered significant if there are at least three out of six following measures in play: "... (1) deposit freezes and/or bank holidays; (2) significant bank nationalizations;

2.2. Traditional Sources of Vulnerabilities and the Macro-financial Indicators

In identifying potentially good early-warning indicators, the literature has put much focus on domestic macro-financial imbalances as well as global risks and international contagions, which are considered the primary sources of vulnerabilities. In the following, we will present a review of these sources and the corresponding macro-financial indicators that are employed in our EWSs. After that, we will introduce the data sources and the way in which these indicators are transformed.

Asset Price Bubble: A common driver of financial fragility and historical systemic banking crises is asset price bubbles, which refer to prolonged, inflated pricing of real or financial assets (Brunnermeier and Oehmke, 2013).³ These bubbles, which are the product of a low-interest-rate environment coupled with developments in mortgage-backed securities and excess supplies of global savings, burst in the presence of asymmetric information and deregulation in the financial sector (Allen and Gale, 2000; Babecky et al., 2014). To capture booms and bursts in asset prices, we use the *house price index* as an early-warning indicator.⁴

Credit Expansion: Credit plays a crucial role in the build-up phase of systemic banking crisis events since the burst of asset price bubbles causes more damage to the economy if they are credit-fueled (Mishkin, 2008). For instance, the US dot-com bubble in the early 2000s, which was an unleveraged stock market bubble, posed a limited risk to financial stability (Jordà et al., 2015). By contrast, credit-fueled asset price bubbles, such as Japan 1980s asset price bubbles, Scandinavian 1990s housing bubbles, and the US 2000s housing bubbles, had led to severe credit crunches (Brunnermeier and Oehmke, 2013; Jordà et al., 2015). Hence, we use data on *total credit to the private non-financial sector as a share of GDP* to capture credit developments.

Current Account Imbalance: As numerous past crises and the global financial crisis 2007/2008 were evidently preceded by large current-account deficits in several industrial countries, a current-account imbalance is a possible source of macro-financial vulnerability and thus a legitimate target of regulators (Obstfeld, 2012).⁵ We therefore employ the *current-account-to-GDP ratio* as an indicator in our EWSs.

Macroeconomic Environment: To reflect the macroeconomic environment and economic activities, we use two standard business-cycle indicators which are the *GDP growth rate* and the *inflation rate*. Whilst moderate GDP growth reflects economic stability, rapid or slow GDP growth may both be an alarming sign of vulnerability (Kauko, 2014). On the one hand, overheated economies are more vulnerable to external risks, which can be transmitted across borders through capital flows (Jordà et al., 2011). On the other hand, slow or negative economic developments signal real economic weaknesses, which potentially lead to crises. Similarly, high inflation may signal macro-financial instabilities while too low inflation or worse –

(3) bank restructuring fiscal costs; (4) extensive liquidity support; (5) significant guarantees put in place; and (6) significant asset purchases.”

³ Examples are the booms in real estate and stock markets in Japan in the late 1980s, and the housing bubbles in the US preceding the global financial crisis 2007/2008.

⁴ Another asset price indicator is equity price. However, Drehmann et al. (2011) and Schularick and Taylor (2012) show that equity-price-based indicators such as stock price have very little explanatory power in early warning models for banking crises. Given that data for stock price is largely missing for a subset of countries in our sample, we use the house price index to capture asset price bubbles.

⁵ Reinhart and Rogoff (2008) show that the current account starts to deteriorate several years prior to a banking crisis, and the average deficit in advanced economies is at its peak one year before the outbreak of the crisis.

deflation – is also a worrying sign as it restrains consumer spending and may even induce recessions.

Global Development: The global financial crisis 2007/2008 has raised concerns about the link between banking crises and economic globalization (Obstfeld, 2012). In an increasingly globalized environment, external shocks, e.g., changes in foreign interest rates, terms of trade, etc.) can affect domestic financial systems through various transmission channels (Kaminsky and Reinhart, 1999). We therefore add the *world GDP growth rate*, the *US 10-year treasury rate*, and the *trade-to-GDP ratio* to capture the potential conduits of contagion.

Sequencing Effects from Other Types of Financial Crises: Along with banking crises, two other types of severe financial failure are currency and sovereign debt crises. More often than not, these crises come in waves as each type of crisis may be preceded by, coincided with, or followed by another one. The fact that banking crises tend to precede or coincide with rather than follow currency and sovereign debt crises (Reinhart and Rogoff, 2011; Laeven and Valencia 2020) hints at a crucial point which needs to be taken into account when building EWSs. Specifically, it would be more useful to include the primary predictors of currency and debt crises rather than their actual timing, since these crises may have not yet materialized in the early-warning horizon. Still, the relating vulnerabilities might be already present. As an example, the public debt may have reached a critical level much before a debt crisis is realized, and by the time the debt crisis is perceived, a systemic banking crisis might have already occurred. The *total reserve-to-GDP ratio* as a leading indicator for currency crises (Kaminsky and Reinhart, 1999) and the *public-debt-to-GDP ratio* as the chief indicator for sovereign debt crises (Manasse and Roubini, 2009; Reinhart and Rogoff, 2011) are thus included in our EWSs.

Other Sources: Also mentioned as potential sources of vulnerabilities are the bank-level structural factors such as the degree of concentration in banking systems, the central bank's independence, banks' profitability, liquidity and solvency, banks' interest rates, as well as the deposit insurance scheme and the moral hazard problem it entails.⁶ However, no bank-level indicator will be included in our models in accordance with many other studies (e.g., Holopainen and Sarlin, 2017; Beutel et al. 2019), due to the following four reasons: (i) There is inconsistent empirical evidence as to whether bank-level indicators are good predictors for systemic banking crises. Instead, these indicators are more suitable for the purpose of studying individual bank failures (see Kauko 2014 and the references therein); (ii) Using a regression approach, there is naturally a trade-off between maximizing the time dimension and the cross-sectional dimension of the dataset. Given that bank-level data are not available for all countries in the sample or do not fully cover the time period 1975-2017, the inclusion of these variables, therefore, would considerably truncate the sample;⁷ (iii) With the purpose of comparing performances across models (as opposed to, say, finding a complete set of indicators for systemic banking crises), it is not necessary to have all possible early-warning indicators included; (iv) In addition, it is important to avoid data mining by including too many variables

⁶ For more details about deposit insurance scheme and the relating moral hazard problem, see e.g., Demirgüç-Kunt and Detragiache (2002).

⁷ For example, Klomp and de Haan (2009) suggest that central bank independence is positively related to financial stability. However, data for that indicator – which the authors borrowed from Arnone et al. (2007) – is available for only two points in time (late 1980s and 2003). In another instance, one potentially useful indicator is the CPIA financial sector rating provided by the World Bank, which assesses a country's financial sector structure and the policies and regulations that affect it. Yet, data for this indicator is only available from 2005 onwards, hence, including it to the model would significantly reduce the number of observations, given our studied time period is 1975-2017.

to an EWS, which might cause overfitting, and such overestimation of the model's predictive power (due to overfitting) might lead to wrong preventive actions (Beutel et al., 2019).

Overall, the chosen early-warning indicators are among the most important ones with substantial predictive powers as has been shown in the literature (e.g., Jordà et al., 2011; Schularick and Taylor, 2012; Holopainen and Sarlin, 2017 and Beutel et al., 2019). In most cases, the macro-financial indicators are transformed into, or already in form of, ratio to GDP (e.g., current-account-to-GDP, public-debt-to-GDP, reserves-to-GDP). These variables and those reflecting year-over-year percentage changes (such as the credit growth rates, inflation rates, domestic and world GDP growth rates, US 10-year treasury rate) will enter the models without further transformations. Exceptions are credit-to-GDP ratio, house price index, and trade-to-GDP ratio, whose deviations from trend will be used. Regarding the credit-to-GDP ratio, its trend deviation is more reliable and powerful as an early-warning indicator for banking crises than the ratio itself, according to the Basel III framework.⁸ Meanwhile, house prices and trade volumes naturally contain time trends that need to be removed ex-ante. For the detrending procedure, the standard two-sided Hodrick-Prescott (HP) filter is applied, and the smooth parameter is set to 100 ($\lambda = 100$) as the periodicity of data is annual. Since the one-sided (real-time) HP filter is often recommended in predicting tasks where no future information is yet available (see, e.g., Jokipii et al., 2021), it is used as a robustness check to compute trend deviations.⁹

Finally, to avoid losing observations due to lack of data, data of some variables are collected from multiple sources if appropriate. For instance, house price indexes are taken from the OECD i-Library for all countries except for Singapore, which is the only non-OECD country in the sample and whose information is sourced from its government's public database.¹⁰ Credit-related data is collected from BIS (Bank for International Settlements) and WDI (World Development Indicators), with the former serving as the primary source and the latter providing information for some countries that are not covered in the BIS dataset. Table A.2 in the Appendix presents the detailed descriptions, transformations, and sources of all macro-financial indicators, whereas their descriptive statistics are reported in Table A.4.

2.3. Political Environment as a New Source of Uncertainty

In the literature on EWSs, much focus has been put on the macro-financial imbalances of the economy at the expense of the political environment, which is also a potential source of vulnerability. The linkage between political environment and financial crises, however, is not entirely new as it has been highlighted in the works of McCarty et al. (2013), Fernandez-Villaverde et al. (2013), Calomiris and Haber (2014), Herrera et al. (2020), among others. To assess the role of the political environment in the likelihood of systemic banking crises, we investigate three factors of the political environment:

The timing of political elections: There is both theoretical and empirical evidence that political elections create socio-economic uncertainties. Theoretically, due to the temporary information asymmetry (e.g., a government knows about its own performance better, or observes it sooner, than the representative voters), an incumbent government has incentives to

⁸ See Basel Committee on Banking Supervision (2010) for more details.

⁹ One drawback of using one-sided HP filter is that it is prone to the end-of-sample bias and thus may generate unreliable gaps (Edge and Meisenzahl, 2011).

¹⁰ Even though the base year to compute the index of Singapore is different from that of OECD data (2009 and 2015, respectively), it is not problematic in terms of consistency since it is the trend deviation of the index that serves as an indicator rather than the absolute value of it.

send out signals about optimistic economic prospects before an election to enhance the chance of being re-elected (Rogoff and Sibert, 1988). Such opportunistic acts give rise to the “political business cycles” in many macroeconomic factors such as government spending, taxes, and money supplies (Nordhaus, 1975; Alesina and Roubini, 1992). For example, there is a tendency for governments to adopt expansionary fiscal policies in pre-election periods and contractionary policies after being re-elected (Frankel, 2010). The so-called “honeymoon hypothesis”, which implies major economic reforms right after political elections, is confirmed by Alesina et al. (1997).¹¹ The reforms are more likely and more extensive if the new government has opposite partisan interests to the previous one (Krueger, 1993). Empirically, several studies have proven the negative effects of policy uncertainty around elections on the banking system’s stability. Baum et al. (2010), for instance, show that Turkish banks’ efficiency reduces significantly around elections. Jackowicz et al. (2013) find that state-owned banks in Central European countries report significantly lower net interest income ratios during parliamentary election years. Eichler and Sobanski (2016) confirm the negative impact of national electoral cycles on bank default risk in the Eurozone. Even though these studies examine individual bank failures, we expect that the negative effects of policy uncertainty around elections on individual banks’ performance would increase the banking system’s vulnerability and thus raise the likelihood of a systemic banking crisis. We therefore use a dummy variable *election time*, which indicates the election year and the first year following the election year, to capture the uncertainty caused during and after political elections.¹²

The political ideology of the incumbent government: As economic policies are often shaped by the political ideology of the incumbent government, this political factor is expected to be a useful crisis predictor. More specifically, while left-leaning governments often impose more market regulations in pursuing higher levels of socio-economic equality, right-wing governments tend to value economic freedom, free trade, and certain social orders (classes/hierarchies). For instance, Potrafke (2010) provides empirical evidence from OECD countries showing that there are more (less) market regulations with left(right)-wing governments. Bechtel (2009) shows that right-leaning governments tend to offer a more favorable investment environment than do their political counterparts, which is appealing to investors who have a higher income and higher propensity to save. Castro and Martins (2021) confirm that right-wing governments tend to ease regulations and promote economic freedom. Yet, Nguyen et al. (2020) find that currency crises are more likely during right-wing governments’ terms, suggesting that economic freedom and market deregulations may foster financial fragility and international contagions. Likewise, Phan et al. (2021) find that a banking system with stricter capital regulations is less vulnerable to economic policy uncertainty, implying that deregulation might have a destabilizing effect on the financial system. Based on existing evidence, we conjecture that governments’ political ideology can influence the likelihood of a banking crisis. We hence employ the dummy variables *right-wing government*, *left-wing government*, and *central government* indicating the respective political ideologies of the incumbent governments as potential early-warning indicators.

The time in office of chief executives: Theoretically, rising uncertainty is one among four factors (besides deterioration in financial sector balance sheet, rises in interest rates, and deterioration of nonfinancial balance sheets) that lead to increases in asymmetric information and thus to financial instability (Mishkin, 1999). Since changes in incumbent chief executives create uncertainty regarding future policies, they may amplify a banking system’s vulnerability

¹¹ In political contexts, the “honeymoon period” indicates the first year in office of an incumbent.

¹² In a robustness check, we adjust the election time indicator to also capture the year before an election.

during distress events.¹³ Nguyen et al. (2020) find that the time in office of chief executives is negatively correlated with the likelihood of banking crises. The authors conjecture that a longer time in office allows policy continuity and thus facilitates economic stability. Phan et al. (2021) confirm the adverse effect of economic policy uncertainty on financial stability. We thus expect that longer time in office of chief executives, captured by the indicator *time in office of executive*, is associated with a lower likelihood of systemic banking crises.

Data of the employed political indicators are acquired from the Database of Political Institutions 2017 (DPI2017) as described in Cruz et al. (2018). Details of each indicator can be found in Table A.3 in the Appendix, while their descriptive statistics are reported in Table A.4.¹⁴

3. Methodology

In this section, we describe the setups of the two logit EWSs – one using both macro-financial and political indicators and one using macro-financial indicators only. The measures of the predictive performances and the evaluation strategies will also be introduced later in this section.

3.1. Early Warning Window

As EWSs have the purpose of helping predict future crises, it is of crucial importance that these systems resemble the underlying decision-making problem faced by policymakers, which is, in this case, the problem of dealing with low-chance-yet-high-impact events (Holopainen and Sarlin, 2017). Given that crises may be driven by unpredictable shocks, predicting the exact timing of these events is an extremely hard if not unrealistic task. Hence, rather than aiming at forecasting the occurrence of crises, recent EWSs are built in a way that they can signal the building up of macro-financial imbalances that may trigger crises. Accordingly, the target (dependent) variable in these models is not the crisis itself but the pre-crisis period. That entails the problem of defining pre-crisis episodes, which is described in the following.

Technically, an observation is classified as pre-crisis if it is followed by a crisis in the forecast window, which is a pre-defined future time interval; otherwise, it is considered as a tranquil or “normal” time. The length of the forecast window is arbitrarily determined by the policymaker, taking into account two criteria: (i) the starting point of the window should not be too close to the observation as, in case of an impending crisis, pre-emptive policies require some time to come into effect; (ii) the ending point of the window should not be too far in the

¹³ It should be noted that changes in the chief executive position and hence in the pursued economic policies can still happen without changes in the overall political ideology of the incumbent government, given the fact that the spectrum of political ideology has more positions than only left, right, and center (e.g., there are also far-left, center-left, radical center, center-right, or far-right positions).

¹⁴ Initially in this paper, the effect of majority governments on the likelihood of systemic banking crises has also been examined. It is based on the fact that policy changes under minority or divided governments require the approval of both incumbent and opposition parties – a process that often ends up in policy gridlocks (Bechtel, 2009). Conversely, the danger of gridlock is less likely under majority governments. That means, in response to negative economic shocks, minority governments might face more difficulties than majority governments in getting the necessary macro-prudential policies passed and implemented. Yet, Nguyen et al. (2020) shows that majority governments can only reduce the probabilities of twin or triple crises and that effect on the likelihood of a single crisis is statistically insignificant. In our model, this indicator is also highly insignificant and since its data are missing for many observations, it is thus dropped out of the core analyses. The result of the model including this indicator is nevertheless reported in a robustness check (see robustness check 7).

future as, in case the EWS sends out an alarm/signal and thus a crisis is anticipated, the closer the signal is to the crisis (if there is indeed one), the higher the credibility of the prediction (Hamdaoui, 2016; Beutel et al., 2019). With the same rationale, pre-crisis episodes can be determined by setting an early-warning window prior to each banking crisis. Observations that fall into the early-warning windows will then be classified as pre-crisis. Since EWSs are built based on historical data, i.e., past crisis observations, setting up an early-warning window before a crisis to define the pre-crisis episodes – the target/dependent variable – may be more intuitive.

Following the literature, this paper defines the three years prior to each crisis year as pre-crisis, implying a systemic banking crisis could be predicted in advance between one and three years. In addition, the crisis years and the three years after each crisis are excluded from the sample to address the so-called “post-crisis bias”.¹⁵ Proposed by Demirgüç-Kunt and Detragiache (1998) and elaborated by Bussiere and Fratzscher (2006), “post-crisis bias” implies that the behavior of explanatory variables after the onset of a banking crisis might bias the results of binomial logit early-warning models. For example, the credit-to-GDP ratio is likely to fall in the aftermath of a banking crisis and so is the GDP growth. Hence, these periods cannot be considered as tranquil times and should be excluded to avoid biased results. One drawback of dropping these periods is that the number of observations in the sample is substantially reduced.¹⁶

Once pre-crisis and tranquil periods are classified, they are transformed into binary values, taking the value of “1” and “0”, respectively. For some countries that experience two or more banking crises during a short period of time (e.g., the USA and South Korea in the 1980s and the UK and Japan in the 1990s, see Table A.1 in the Appendix), it happens that the post-crisis period of the first crisis may fully or partly overlap with the pre-crisis period of the next crisis. In these cases, the overlapped episodes will be excluded.

3.2. Logit-based Early Warning System

After building the complete dataset, the logistic regression is then used to compute for each observation the probability of being in a pre-crisis period, hereafter named “pre-crisis probability”. Logistic regression is the preferable approach in this paper compared to machine learning techniques for two reasons. First, logit models are less prone to the overfitting problem, which is a conundrum of machine learning techniques. Second, as the role of the new political indicators is of main interest, the regression approach helps to arrive at an intuitive economic interpretation while the superiority of machine learning methods in dealing with flexible patterns and heterogeneous data often comes at the cost of interpretation.

To examine whether political indicators can help improve the prediction of EWSs, two multivariate logit models are employed: one using both macro-financial and political indicators and one having macro-financial indicators only. For the sake of brevity, they will be mentioned

¹⁵ According to Nguyen et al. (2022), the average duration of a systemic banking crisis in advanced economies is 3.74 years, while that figure is 4.35 years in the study of Caggiano et al. (2016). Hence, excluding the crisis years and the three years after each crisis (4 years in total) would be adequate to avoid the post-crisis bias.

¹⁶ Proposed by Bussiere and Fratzscher (2006), another way to address the “post-crisis bias” problem without losing many observations is to use the discrete-dependent-variable approach, e.g., a multinomial logit model with three outcomes representing all three classifications (pre-crisis, post-crisis, and tranquil times). Given the purpose of this paper, which is to examine whether adding political indicators to a logit EWS can improve its predictive performance, it is not necessary to implement the multinomial logistic regression method.

as “the two EWSs of interest” hereafter. Under the former, the pre-crisis probability of country (*i*) in year (*t*) is given by:

$$Prob(Y_{it} = 1|MaFi_{it}, Pol_{it}) = F(MaFi'_{it}\alpha + Pol'_{it}\beta) = \frac{e^{MaFi'_{it}\alpha + Pol'_{it}\beta}}{1 + e^{MaFi'_{it}\alpha + Pol'_{it}\beta}} \quad (1)$$

where *Y* is the pre-crisis dummy; *MaFi* and *Pol* are the vectors of the macro-financial and political indicators, respectively; α and β are the corresponding vectors of coefficients; and $F(\cdot)$ is the cumulative logistic distribution function. Parameters α and β are obtained by a maximum likelihood function as follows:

$$\ln L = \sum_{i=1}^n \sum_{t=1}^T \{Y_{it} \ln F(MaFi'_{it}\alpha + Pol'_{it}\beta) + (1 - Y_{it}) \ln [1 - F(MaFi'_{it}\alpha + Pol'_{it}\beta)]\} \quad (2)$$

Similarly, the pre-crisis probability of country (*i*) in year (*t*) under the logit EWS using only macro-financial variables is given by:

$$Prob(Y_{it} = 1|MaFi_{it}) = F(MaFi'_{it}\gamma) = \frac{e^{MaFi'_{it}\gamma}}{1 + e^{MaFi'_{it}\gamma}} \quad (3)$$

where γ represents the vector of macro-financial coefficients under the new model setting and is determined by maximizing the log-likelihood function below (other specifications are analogous to those of the first model):

$$\ln L = \sum_{i=1}^n \sum_{t=1}^T \{Y_{it} \ln F(MaFi'_{it}\gamma) + (1 - Y_{it}) \ln [1 - F(MaFi'_{it}\gamma)]\} \quad (4)$$

Using panel data, empirical models often include cross-sectional fixed effects to capture potential heterogeneity in the explanatory variables across sections. In the EWSs literature, however, pooled data is the preferable approach (see e.g., Holopainen and Sarlin 2017; Beutel et al. 2019). The rationale behind this is that introducing country-specific effects requires leaving out from the panel countries that did not experience banking crises during the studied period, which in turn would significantly truncate the sample and may even induce sample-selection bias (Demirgüç-Kunt and Detragiache, 1998; Davis and Karim 2008). Furthermore, comparing pooled, fixed-effect, and random-effect EWSs, Bussiere and Fratzscher (2006) find that omitting fixed and random effects does not produce biased estimations in EWSs for banking crises. For those reasons, this paper estimates pooled logit models and uses robust standard errors as an attempt to address heteroskedasticity in the error terms. Also, the models use data of all variables contemporaneously, i.e., the explanatory variables are not lagged. Since the dependent variable in the EWSs measures the vulnerabilities prior to crises, it cannot be the case that the behavior of explanatory variables during a pre-crisis period is influenced by the crisis itself, which is yet to occur. Put differently, there is no risk of simultaneity in the target models.

3.3. Predictive Performance Measures

To transform predicted probabilities into correct signals, an EWS uses a specific threshold $\tau \in [0,1]$, based on which the system will send out alarms if predicted pre-crisis probabilities exceed the cut-off value τ . The compatibility between the signal produced by the EWSs and

the actual outcome can be summarized in the so-called contingency matrix for all cases as shown in Table 1.

Table 1: A contingency matrix

		Actual class	
		Pre-crisis period	Tranquil period
Predicted class	Signal	Correct call <i>True positive (TP)</i>	False alarm <i>False positive (FP)</i>
	No signal	Missed crisis <i>False negative (FN)</i>	Correct silence <i>True negative (TN)</i>

Source: Holopainen and Sarlin (2017)

Based on the contingency matrix, a wide range of measures on the predictive performance of an EWS can be derived. The most straightforward measures include but are not limited to: Accuracy rate = $(TP+TN)/(TP+TN+FP+FN)$; Precision rate = $TP/(TP+FP)$; Sensitivity rate = $TP/(TP+FN)$; Specificity rate = $TN/(FP+TN)$; False negative rate (type I error) = $FN/(TP+FN)$; False positive rate (type II error) = $FP/(FP+TN)$. In practice, policymakers face a trade-off when using EWSs between missing crises (type I errors) and issuing false alarms (type II errors). Intuitively, failing to perceive the build-up of vulnerabilities may be more costly to the policymakers than taking unnecessary preventive actions. To capture unbalanced preferences between the two types of errors, policymakers can construct their own loss function, in which the two types of errors are weighted by a pre-set parameter. In this paper, the loss function proposed by Sarlin (2013) is employed and described as follows:¹⁷

$$L(\mu) = \mu T_1 P_1 + (1 - \mu) T_2 P_2 \quad (5)$$

where $\mu \in [0,1]$ is the preference parameter; T_1 and T_2 refer to type I and type II errors, respectively; P_1 and P_2 denote the unconditional probability of pre-crisis periods and tranquil periods, respectively, with $P_1 = \frac{TP+FN}{TP+TN+FP+FN}$ and $P_2 = (1 - P_1)$. With $\mu = 0.5$, the two types of errors are weighted equally. Likewise, if policymakers would rather issue a false alarm than miss a crisis, a higher μ can be set (e.g., $\mu = 0.8$). Based on the loss function, two other performance measures – absolute Usefulness (U_a) and relative Usefulness (U_r) – can be set up as follows (Sarlin, 2013):

$$U_a(\mu) = \min(\mu P_1, (1 - \mu) P_2) - L(\mu) \quad (6)$$

$$U_r(\mu) = \frac{U_a(\mu)}{\min(\mu P_1, (1 - \mu) P_2)} \quad (7)$$

The absolute Usefulness (U_a) is the difference between the loss faced by policymakers in case of not using any model ($\min(\mu P_1, (1 - \mu) P_2)$) and the loss when an EWS is employed ($L(\mu)$). Specifically, ignoring EWSs, policymakers can achieve the loss of $\min(\mu P_1, (1 - \mu) P_2)$ by opting for one out of two simple decisions: always issuing alarms ($T_1 = 0, T_2 = 1, L(\mu) = (1 - \mu) P_2$) or never doing so ($T_1 = 1, T_2 = 0, L(\mu) = \mu P_1$), depending on which action incurs the smaller loss. This measure underlines the fact that building a useful EWS is a challenging task as a not-so-good model could easily be beaten by the decision to not use it. The measure of relative Usefulness (U_r) is conceptually close as it relates the absolute

¹⁷ Another loss function with slightly different specification can be found in Alessi and Detken (2011).

Usefulness to the maximum usefulness a model can achieve (e.g., when $L(\mu) = 0$). In general, the higher the usefulness, the better the model is.

It should be noted that all entries of the contingency matrix and thus all measures specified above depend on the threshold τ . The lower the threshold, the more signals the model sends out, and hence, the lower the possibility of missing a crisis, at the expense of more false alarms. Policymakers can use the default threshold $\tau = 0.5$ or another cut-off value that best suits their preferences regarding the two types of errors. For instance, Bussiere and Fratzscher (2006) choose a threshold of 20%, Hamdaoui (2016) sets a lower value at 10%, while the cut-off probability in Alessi and Detken (2018) is 30%. Yet, instead of an ad-hoc choice, policymakers can also determine the optimal threshold of their models by maximizing the model's predictive performance in terms of the above measures (e.g., U_a or U_r), given a specific preference parameter μ . Empirical results of Beutel et al. (2019) show that a threshold around 10% appears to be optimal for their logit EWSs while the optimal cut-off value for machine-learning-based models is often higher than that.

Nevertheless, Sarlin and von Schweinitz (2021) point out two undesirable properties of an optimal threshold. First, since the threshold is optimized based on the in-sample performance of EWSs, it is likely to change when new observations are added, i.e., the optimal threshold is time-varying. Second, as future crises may differ from past ones, good in-sample performance does not ensure good out-of-sample performances, thus, the optimal threshold may become sub-optimal in out-of-sample predictions. For these reasons, this paper uses the above measures for the in-sample evaluation of EWSs only, with preference parameter μ set at 0.9 and the threshold τ set at 0.1.¹⁸ Using available data, Figure A.1 in the Appendix illustrates that this much lower threshold (compared to the default value of 0.5) is more useful for logit EWSs as it helps not missing too many pre-crisis episodes while having a moderate number of false alarms when mapping pre-crisis probabilities into binary signals.

It follows from above that a measure, which is independent of the threshold τ and thus independent of the policymaker's preference, would be ideal for both in-sample and out-of-sample performance evaluations. Fortunately, such measure exists. It is the Area Under the ROC (Receiver Operating Characteristic) Curve, so-called AUROC or AUC (see Figure A.2 for an example). The ROC curve plots the True Positive Rate against the False Positive Rate of the outcome for all possible values of the threshold $\tau \in [0,1]$. Specifically, for each threshold, the ROC is calculated as follows:

$$ROC = \frac{\text{True Positive Rate}}{\text{False Positive Rate}} = \frac{\frac{TP}{TP + FN}}{\frac{FP}{FP + TN}} = \frac{\frac{TP}{TP + FN}}{1 - \frac{TN}{FP + TN}} = \frac{\text{Sensitivity Rate}}{1 - \text{Specificity Rate}} \quad (8)$$

Technically, the value of AUC (the integral of the area under the ROC curve) ranges from 0.5 to 1. While an AUC of 0.5 can be achieved by a random classifier (e.g., a coin toss), an AUC of 1 means that the model is a perfect classifier. Accordingly, a good EWS should have an AUC closer to 1 than to 0.5, and the higher the AUC, the better the model is. To provide objective comparisons between the two logit EWSs of interest, the AUC will be the chief measure in the in-sample evaluation and the only measure in the out-of-sample exercises.

¹⁸ Following Sarlin and von Schweinitz (2021), the choice of threshold τ is based on the optimal setting: $\tau = 1 - \mu$, where the loss function is minimized and the policymaker is indifferent between a signal and no signal.

3.4. Prediction Evaluations

Following the literature, predictions of the two EWSs of interest are assessed on both in-sample and out-of-sample performances.¹⁹ Yet, as future crises are unknown events, one way of testing a model's out-of-sample performance is splitting the dataset into two subsamples: one *training sample* for model estimation and one *testing sample* for a so-called “pseudo” out-of-sample forecasting experiment. Two evaluation designs operating on that idea, namely *k*-fold cross-validation and quasi real-time exercise, will be performed in this paper. Properties of these evaluation exercises and the estimation strategies involved are summarized in the following.

K-fold cross-validation: Cross-validation is a model validation method that is widely used to assess how good a model's estimation results can be generalized to new (independent) observations.²⁰ The idea of *K*-fold cross-validation (Geisser, 1975) is to randomly divide the data into *K* partitions (folds) of approximately equal size (n/K), from which fold K^{th} is kept out for the out-of-sample validation while the remaining ($K-1$) folds are used to fit a model in. The estimation process repeats with each fold sequentially playing the role of the validation sample. In principle, *K*-fold cross-validation is analogous to the exhaustive leave-*p*-out cross-validation (Shao, 1993) as the latter also takes out every possible subset of *p* observations from the sample to play the role of a validation sample. With $K = n$, there are *n* folds and the *K*-fold cross-validation turns into the classic leave-one-out cross-validation (Geisser, 1975).

In applications, one important question is how big *K* should be? Evidence from the related literature (see, e.g., Shao, 1993; Breiman and Spector, 1992; among others) shows that leave-one-out cross-validation (with *K* as big as *n*) often leads to high-variance estimates as the estimations from each fold are highly correlated.²¹ Nonetheless, as *K* gets larger, the bias of the technique gets smaller, since the difference in the size of the resampling training sets decreases (Kuhn and Johnson, 2013). Until now, the literature has preferred using $K = 5$ or $K = 10$, as these values are shown empirically to provide a balanced bias-variance trade-off. Following Holopainen and Sarlin (2017), this paper uses the value of $K = 10$ and thus performs a 10-fold cross-validation as an out-of-sample exercise.

Quasi real-time estimation: As suggested by Holopainen and Sarlin (2017), the application of cross-validation in model comparisons is not without problems. Specifically, the cross-validation process does not take into account the existence of serial and cross-sectional correlations in panel data when partitioning. However, while this may make cross-validation yield results biased in favor of flexible approaches like machine learning compared to logit models, it is not a problem in this paper where the two models are both based on logistic regressions. Nonetheless, another out-of-sample exercise under a real-time setting would be conducive to examining how different the two models perform in real time.

For this, the real-time out-of-sample exercise will be introduced to test how precisely the two logit models can predict the buildup of vulnerabilities prior to the global financial crisis 2007/2008. The estimation proceeds as follows. Data from 1975 to 2005 will be used as a training sample to estimate the two models, which in turn will provide forecasts for all observations in 2006, but not in 2007. To mimic real-time usage, predictions for 2007 will be obtained by repeating the process with data up to 2006 as the new training set. As forecasts for

¹⁹ The importance of out-of-sample experiments in comparing different EWSs has been advocated by Holopainen and Sarlin (2017).

²⁰ It should be noted that, as we use parametric models (logistic regression), cross-validation is not employed here to tackle models' estimation uncertainty but to create a pseudo out-of-sample experiment that is conducive to examining how well the two models of interest perform in different subsamples/time periods.

²¹ For an intensive review on various cross-validation techniques, see Arlot and Celisse (2010).

each year are produced using only information available up to that year, and the exercise is an ex-post assessment based on historical data, it is hence termed quasi real-time.²²

4. Empirical Analyses

4.1. EWS With versus Without Political Indicators

Before advancing to the predictive performance evaluations, we test how good the two logit EWSs are in terms of classification models in general. For that purpose, Table 2 reports the estimated coefficients, goodness-of-fit tests, and likelihood-based statistics of the two models.

Regarding Panel A, as to logit models or any non-linear models, the coefficients do not represent marginal effects and thus require careful interpretations. More details on that will be discussed in Section 5, after the better performer between the two models is determined through various comparisons. The focus at this point is on Panels B and C of Table 2. With respect to Panel B, the two goodness-of-fit tests – Pearson chi-squared (χ^2) test and Hosmer-Lemeshow test – examine how well the predicted values generated by each model match with the observed/actual outcomes. As a small p-value (usually under 5%) corresponds to a poor fit (Hosmer et al., 2013), the obtained large p-values ($\text{prob} > \chi^2$) mean that we cannot reject both models as potentially good classifiers.

Yet Panel C shows that the model with political indicators (Model 1) is the better classifier in terms of all performed likelihood-based statistics. The model is better if the former two statistics (Likelihood-ratio test and McFadden R-squared) are higher or the AIC (Akaike Information Criterion) is lower. Since the AIC punishes the inclusion of more variables through a penalty term that increases with the number of regressors, it provides a trade-off measure between the goodness-of-fit and the parsimony, or simplicity, of a model (Verbeek, 2012). It is thus conducive to any model selection task where the comparative models have different degrees of freedom, which is clearly the case in this paper. In sum, we see that both models are of good fit in terms of classification models but the model incorporating political indicators is preferable, given the better results of likelihood statistics.

As another necessary step, the presence of multicollinearity is also checked. The pairwise correlation matrix among independent variables is reported in Table A.5 in the Appendix. In general, no high correlation shows up in the results.

²² It should be noted that 2005 is also a pre-crisis year according to the definition specified in this paper. Yet, a training sample with data up to 2004 would not provide enough observations for a reliable estimation of models. Thus, following the literature, 2006 is chosen as the starting point for the real-time exercise.

Table 2: Logit-based EWS with vs. without political indicators

	EWS with political indicators (Model 1)	EWS without political indicators (Model 2)
Panel A: Estimated coefficients of logit models (Dependent variable: Pre-crisis)		
Constant	-3.3811***	-5.2970***
<i>Macro-financial variables:</i>		
House price gap	0.2388***	0.2469***
Credit growth rate	0.0346**	0.0322**
Credit-to-GDP gap	0.0624***	0.0608***
Current-account-to-GDP	0.1137***	0.1272***
Inflation rate	-0.1710***	-0.1542***
GDP growth rate	-0.0565	-0.0390
World GDP growth rate	0.5941***	0.5771***
US 10-year treasury rate	0.4037***	0.3872***
Trade-to-GDP gap	0.0610	0.0480
Public-debt-to-GDP	0.0011	0.0035
Reserves-to-GDP	-21.0344***	-17.8184***
<i>Political variables:</i>		
Election time	-0.4658	—
Time in office of executive	-0.1038**	—
Right-wing government	-0.6166	—
Left-wing government	-1.4967**	—
Center government	-1.0949	—
Panel B: Goodness-of-fit tests		
Pearson χ^2 test	428.49	446.12
Pearson prob > χ^2	1.0000	1.0000
Hosmer–Lemeshow test	5.41	9.85
Hosmer–Lemeshow prob > χ^2	0.7131	0.2759
Panel C: Likelihood statistics		
Likelihood ratio (LR) test	198.044***	184.049***
McFadden (or Pseudo) R-squared	0.3840	0.3569
AIC	351.684	355.679
Observations	708	708

Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively.

4.2. In-sample Performance Evaluation

Using the full sample, several measures on the predictive performance of the two logit EWSs are obtained based on the correspondence between the fitted (predicted) values and the actual outcomes. Table 3 reports the results.

As can be seen from Table 3, the model with political indicators outperforms the other one in all measures. To evaluate whether the difference in the performances of the two models is statistically significant, we test the hypothesis (H_0) of whether the AUCs produced by the two

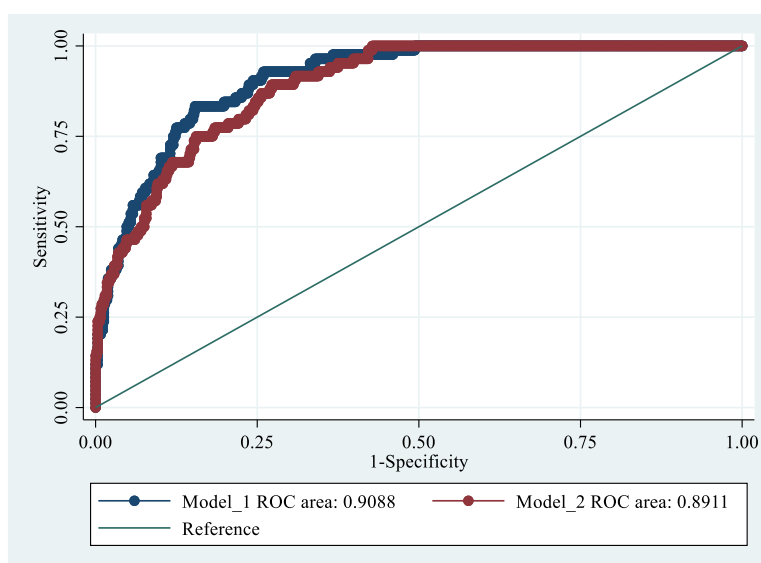
models are equal. Since this measure is independent of policymaker's preference between missing crises (type I errors) and issuing false alarms (type II errors), the obtained result of the test would be representative of the overall result. Table 3 presents the standard error and the 95% confidence interval of the AUC for each model, as well as the result of the Chi-squared test. The yielded significance probability (p-value) of 0.0079, which is smaller than 1%, means that we can reject the null hypothesis (H_0). In other words, the difference between the predictive performances of the two EWSs, albeit small, is statistically significant. This performance improvement implies that there are clear links between the newly employed political indicators and the banking crisis events. We will discuss this in more detail in Section 5. The ROC curves yielded by the two models are visualized in Figure 1.

Table 3: In-sample predictive performance

Predictive performance measurements	EWS with political indicators (Model 1)	EWS without political indicators (Model 2)
Accuracy Rate	0.7768	0.7585
Precision Rate	0.3318	0.3117
Sensitivity Rate	0.8690	0.8571
Specificity Rate	0.7644	0.7452
False Negative Rate (Type I Error)	0.1310	0.1429
False Positive Rate (Type II Error)	0.2356	0.2548
Absolute Usefulness (U_a)	0.0534	0.0504
Relative Usefulness (U_r)	0.6057	0.5721
Area Under the ROC Curve (AUC)	0.9088	0.8911
Standard Error	(0.0137)	(0.0152)
95% Confidence Interval	[0.8819; 0.9358]	[0.8613; 0.9209]
H_0 : AUC (Model 1) = AUC (Model 2)	Chi-squared χ^2 : 7.07	Prob > χ^2 : 0.0079

Source: Own elaborations. Notes: Better measures in bold. Of all measures included, the higher the value the better, except for type I and type II errors in which lower values are preferred.

Figure 1: ROC curves for in-sample estimations



Source: Own elaborations. Notes: Model 1 and Model 2 correspond to EWSs with and without political indicators, respectively. The reference line is the line of no-discrimination.

4.3. Out-of-sample Performance Evaluation

Table 4 reports the out-of-sample AUCs of the two EWSs. Regarding the cross-validation exercise, the model with political indicators, once again, performs better. It should be noted that, repeated cross-validation can be implemented to obtain the level of significance for the cross-validation exercise. However, Vanwinckelen and Blockeel (2012) empirically show that for the purpose of testing how well a model derived from the training set will perform in new cases, repeated cross-validation is not useful and could be a waste of computational efforts, as the training and testing samples are already exhaustively drawn from the original dataset. In addition, this method is only advocated when different modeling techniques are involved. Therefore, this paper does not perform this additional step, given that the two models of interest use the same estimation method.

The model with political indicators also performs better on average in the real-time exercise. Nevertheless, performing the same hypothesis test as in the in-sample evaluation, it is shown that the difference in predictions between the two models is no longer statistically significant. This is because the real-time exercise excludes the global financial crisis 2007/2008 from the training set, thus a large numbers of pre-crisis observations is left out in the model estimation phase, which in turn affects the prediction accuracy in new cases (as can be seen from Table A.1 in the Appendix, most systemic banking crisis events covered in the dataset occur during the 2007/2008 global financial crisis).

Table 4: Out-of-sample predictive performance

	EWS with political indicators (Model 1)	EWS without political indicators (Model 2)
<i>10-fold cross-validation exercise</i>		
AUC	0.8734	0.8634
Standard Error	(0.0886)	(0.0750)
Bootstrap bias corrected 95% CI	[0.8403; 0.9140]	[0.8320; 0.9070]
<i>Quasi real-time exercise</i>		
Prediction for pre-crisis year 2006:		
AUC	0.8611	0.8611
Standard Error	(0.0811)	(0.0752)
95% Confidence Interval	[0.7022; 1.0000]	[0.7137; 0.9945]
Ho: AUC (Model 1) = AUC (Model 2)	Chi-squared χ^2 : 0.00	Prob > χ^2 : 1.0000
Prediction for pre-crisis year 2007:		
AUC	0.7875	0.7500
Standard Error	(0.1008)	(0.1011)
95% Confidence Interval	[0.5900; 0.9850]	[0.5519; 0.9481]
Ho: AUC (Model 1) = AUC (Model 2)	Chi-squared χ^2 : 0.55	Prob > χ^2 : 0.4582

Source: Own elaborations. Notes: Higher AUC in bold.

In sum, the EWS with political indicators has better predictive performance compared to the other one in both out-of-sample exercises, despite the great uncertainty in the real-time experiment. These results, together with the previous in-sample comparison, suggest that the included political variables are potentially useful in predicting systemic banking crises and are able to improve the prediction. Figure B.1 and Figure B.2 in the Appendix report the ROC curves produced by the two models in the two out-of-sample exercises. Additionally, it can be seen from Tables 3 and 4 that the out-of-sample performances of both models are weaker than

their in-sample performances. This is understandable as no model can produce better, or at least the same, performance with new observations unless such model is able to capture the true data generation process (DGP).

4.4. Robustness Checks

To assess the robustness of the above results in favor of the EWS with political indicators, several robustness checks are performed. For each test, both in-sample and out-of-sample performance evaluations are conducted. Also, to ensure fair comparisons, the measure of area under the ROC curve (AUC) is solely employed. Table 5 reports the results.

In the first robustness test, the two logit models are estimated using the new sample in which only crisis years are excluded whilst post-crisis episodes are retained and considered as tranquil times. In the second test, the crisis end dates from Laeven and Valencia (2020) are employed instead of post-crisis windows whenever the information is available. The results hardly challenge by these variations of the original dataset. Specifically, the model with political indicators produces higher AUCs in all cases.

In the third test, we examine the extent to which the main results are conditional on the definition of systemic banking crises. For that purpose, the latest crisis database of Laeven and Valencia (2020) is employed. The outperformance of the model with political indicators continues to be observed throughout the test, except for the cross-validation estimation in which the model without political indicators prevails. An explanation for this result will be provided in the next section when we take a closer look at the underlying models.

For the fourth robustness check, the one-sided (real-time) HP filter is used to detrend some macro-financial variables instead of the standard two-sided HP filter. While the two-sided filter requires future data at each point in time, the one-sided filter is purely backward looking, as only information available up to each point is used to calculate the trend. The fourth check confirms that the model using political indicators has better predictions in all exercises on average, implying that the core results are robust to the choice of the detrending method.

The fifth check introduces an interaction term between the inflation rate and election time to the model with political indicators. The idea stems from the political-cycle study of Alesina and Roubini (1992), who find that inflation rates increase immediately after elections in several OECD countries, suggesting the result is due to pre-electoral opportunistic policies of the incumbent governments (e.g., expansionary monetary or fiscal policies) to enhance their chances of being re-elected. The results obtained are very close to those of the benchmark case, and so are the results of the sixth check, where the Federal Funds Rate (FFR) is used instead of the 10-year treasury rate to capture the US monetary policy.

We also perform five additional robustness checks (Robustness Test 7–11) concerning variations in the set of employed political indicators. In robustness check 7, for example, we include the indicator for majority governments as majority governments are expected to face less policy gridlocks than minority ones and thus might react more quickly in distress events (Bechtel, 2009; Eichler and Sobanski, 2016). Since policy uncertainty may already exist in pre-election periods, the election time indicator is defined to capture also the year before each political election in robustness check 8. As a piece of evidence, Liu and Ngo (2014) show that US banks' failures are about 45% less likely in the year leading up to an election. Lastly, statistically insignificant political indicators in the benchmark model such as those representing election time and center and right-wing governments are gradually excluded in the remaining

robustness checks. Overall, the comparative results in favor of the model using political indicators hold in all performed robustness checks.

Table 5: Robustness checks – Predictive performance measured in AUC

	EWS with political indicators (Model 1)	EWS without political indicators (Model 2)
<i>Robustness Test 1: including post-crisis periods</i>		
In-sample estimation (full sample)	0.8969 (0.0142)	0.8794 (0.0157)
10-fold cross-validation exercise	0.8791 (0.0407)	0.8650 (0.0372)
Quasi real-time exercise:		
Prediction for pre-crisis year 2006	0.8333 (0.0812)	0.7933 (0.0937)
Prediction for pre-crisis year 2007	0.7818 (0.0967)	0.7576 (0.0985)
<i>Robustness Test 2: using crisis end dates instead of post-crisis windows</i>		
In-sample estimation (full sample)	0.9061 (0.0140)	0.8876 (0.0157)
10-fold cross-validation exercise	0.8728 (0.0522)	0.8667 (0.0449)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8611 (0.0811)	0.8542 (0.0761)
Prediction for pre-crisis year 2007	0.7875 (0.1008)	0.7500 (0.1011)
<i>Robustness Test 3: using crisis dataset of Laeven and Valencia (2020)</i>		
In-sample estimation (full sample)	0.9320 (0.0135)	0.9265 (0.0151)
10-fold cross-validation exercise	0.9072 (0.0492)	0.9114 (0.0404)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8264 (0.0862)	0.7847 (0.0952)
Prediction for pre-crisis year 2007	0.7250 (0.1062)	0.7125 (0.1057)
<i>Robustness Test 4: detrending with one-sided HP filter</i>		
In-sample estimation (full sample)	0.8685 (0.0195)	0.8400 (0.0212)
10-fold cross-validation exercise	0.8410 (0.0501)	0.8188 (0.0651)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8672 (0.0846)	0.7500 (0.1135)
Prediction for pre-crisis year 2007	0.7540 (0.1085)	0.7143 (0.1122)
<i>Robustness Test 5: introducing an interaction term between inflation rate and election time</i>		
In-sample estimation (full sample)	0.9089 (0.0138)	0.8911 (0.0152)
10-fold cross-validation exercise	0.8728 (0.0888)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8642 (0.0819)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7875 (0.1008)	0.7500 (0.1011)
<i>Robustness Test 6: replacing the US 10-year treasury rate by the Federal Funds Rate (FFR)</i>		
In-sample estimation (full sample)	0.9104 (0.0143)	0.8931 (0.0157)
10-fold cross-validation exercise	0.8737 (0.0951)	0.8597 (0.0813)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8542 (0.0819)	0.8472 (0.0791)
Prediction for pre-crisis year 2007	0.7812 (0.1040)	0.7500 (0.1012)

Source: Own elaborations. Notes: Higher AUC in bold. Standard errors are given in parentheses.

Table 5 (cont.): Robustness checks – Predictive performance measured in AUC

	EWS with political indicators (Model 1)	EWS without political indicators (Model 2)
<i>Robustness Test 7: including indicator for majority government</i>		
In-sample estimation (full sample)	0.9028 (0.0147)	0.8822 (0.0164)
10-fold cross-validation exercise	0.8724 (0.0544)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8611 (0.0789)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7875 (0.1008)	0.7500 (0.1011)
<i>Robustness Test 8: the year before each election is also included in election time indicator</i>		
In-sample estimation (full sample)	0.9088 (0.0134)	0.8911 (0.0152)
10-fold cross-validation exercise	0.8753 (0.0781)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8611 (0.0811)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7875 (0.1008)	0.7500 (0.1011)
<i>Robustness Test 9: excluding indicator for election time</i>		
In-sample estimation (full sample)	0.9085 (0.0133)	0.8911 (0.0152)
10-fold cross-validation exercise	0.8762 (0.0795)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8611 (0.0811)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7812 (0.1013)	0.7500 (0.1011)
<i>Robustness Test 10: excluding indicators for election time and center government</i>		
In-sample estimation (full sample)	0.9071 (0.0132)	0.8911 (0.0152)
10-fold cross-validation exercise	0.8778 (0.0758)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8750 (0.0767)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7750 (0.0974)	0.7500 (0.1011)
<i>Robustness Test 11: excluding indicators for election time, center and right-wing governments</i>		
In-sample estimation (full sample)	0.9071 (0.0132)	0.8911 (0.0152)
10-fold cross-validation exercise	0.8797 (0.0766)	0.8634 (0.0750)
Quasi real-time exercise		
Prediction for pre-crisis year 2006	0.8681 (0.0726)	0.8611 (0.0752)
Prediction for pre-crisis year 2007	0.7562 (0.1002)	0.7500 (0.1011)

Source: Own elaborations. Notes: Higher AUC in bold. Standard errors are given in parentheses.

5. Economic Interpretation and Final Discussion

Besides performance evaluations, economic interpretation of early-warning indicators is of high relevance. Since the model without political indicators is nested in the model with political indicators, it is sufficient to provide the economic interpretation for the latter only. Tables B.1 and B.2 in the Appendix report the estimated coefficients and the average marginal effects, respectively, of the employed indicators in the benchmark case and in all robustness tests.

Similar results for the model without political indicators are provided in Tables B.3 and B.4 in the Appendix.

Most macro-financial indicators have their expected sign, and the results are robust across robustness tests. Increases in asset prices and credit growth raise pre-crisis probabilities, which is in line with historical observations of credit-fueled asset price bubbles preceding systemic banking crises (e.g., Drehmann et al., 2011; Jordà et al., 2011; Schularick and Taylor, 2012). Pre-crisis probabilities also increase in case of economic downturns, reflected by lower GDP growth rates and/or lower inflation rates (see, e.g., Demirgüç-Kunt and Detragiache, 1998; Davis and Karim, 2008). Meanwhile, increased total reserves consistently reduce the risk of systemic banking crises (Frankel and Saravelos, 2012).

Next, we are interested in the newly employed political indicators that are responsible for the difference in predictive performance between the two EWSs. As can be seen from Table B.1, the variable indicating the time around political elections does not play a significant role in the EWS. This is in contrast to what is expected when thinking about the uncertainty caused by elections as possible economic policy reforms during the “honeymoon period” of the newly elected government might increase the vulnerability of an economy and thus increase the possibility of a banking crisis. Nonetheless, the obtained results need not mean that the above hypothesis is rejected but only suggest that the “election time” variable may not be useful as an early-warning indicator. In fact, excluding this indicator from the model does not affect the comparative results on predictive performances. According to the empirical results of Nguyen et al. (2020), banking crises are more likely to materialize in the election years rather than the years after. This, together with the fact that all crisis years are excluded in the model to avoid bias, may explain the results.

On the other hand, the variable indicating the time in office of incumbent chief executives has the expected sign and is statistically significant in all cases. The negative coefficient of the variable means that a longer time in office is related to a lower pre-crisis probability. The underlying idea is that a longer time in office allows for a continuity of economic policies and thus promotes macro-financial stability. It should be noted that chief executives can lengthen their time staying in office by winning consecutive elections, where the maximum number of (presidential) terms one can hold as well as the length of each term (tenure) differ across countries and across political regimes. This result, however, need not be interpreted as an argument advocating for a longer presidential term or a particular political regime. Instead, it suggests that changes create uncertainties. Thus, a smooth transition in economic policies after an election would be beneficial for a country’s economic health.

Concerning political ideologies, there is a statistically significant negative correlation is detected between left-wing governments and pre-crisis probabilities, except in the second robustness test where the dataset with fewer systemic banking crisis events is used.²³ Nevertheless, regardless of the significance levels, the obtained coefficients of the three political-ideology indicators imply in all cases that the pre-crisis probability tends to be lowest when left-wing governments are in office, as the absolute value of the corresponding coefficient is the highest. Likewise, comparing center and right-wing governments, the pre-crisis probability is more likely to be lower if the political ideology of the ruling party is center rather than right leaning. These results support the argument that the higher level of economic

²³ The fact that none of the political-ideology indicators achieves statistical significance in the second robustness test explains for this test’s contradict comparative result in which the model without political indicator is the better performer in the 10-fold cross-validation estimation.

freedom often pursued by right-wing governments, may be favorable in tranquil times, but may also act as a catalyst which raises the vulnerability of an economy in pre-crisis periods.

For the average marginal effects, we look at Table B.2. For logit models, a variable's coefficient naturally does not represent its marginal effects, and the marginal effect itself is not constant but varies with the variable's value. Hence the presented marginal effect of each indicator is the average change of its marginal effects across all observations, with the marginal effect of each observation is computed using a one-standard deviation change in the respective indicator and the contemporary values of the other covariates. Notably, when interpreting marginal effects, it is also important to mind the scale of the indicators' values. For instance, the standard deviation of the reserves-to-GDP ratio for the full sample is about 0.18 (see Table A.4), and the sum of its mean and standard deviation is around 0.3, implying a country's total reserves rarely exceed 30% of GDP across observations. Therefore, it would be better to interpret the results at smaller scales of changes, e.g., the average marginal effect of a 0.1 standard deviation change in the reserves-to-GDP ratio on pre-crisis-probabilities is about -14 percentage point. In other words, a decrease in total reserves equivalent to 1.8% of GDP might raise the pre-crisis probability by nearly 14 percentage point.

Similarly, the average marginal effect of a one-standard deviation change of the years in office of the incumbent chief executive (around 3 years, see Table A.4) can reduce the pre-crisis probability by 0.7 percentage point on average. The corresponding measure for the left-wing government indicator is -10 percentage point. Given that the cut-off threshold τ in the logit model is 10%, these average marginal effects are considerable. Since the vulnerability level of the banking system and thus the crisis likelihood is affected by both macro-financial and political factors, a marginal change in the predicted crisis probability around the cut-off threshold might alter the EWS's signal outcome. For instance, while a 9.5% crisis probability results in a no-warning, a 10.2% crisis probability would produce an alarm. For an illustration, Figure B.3 in the Appendix provides the plots of predicted pre-crisis probabilities for the USA, the UK, Germany, and France during the 2003-2017 period.

Last but not least, it is important to highlight that the purpose of the EWSs is to signal the buildup of vulnerabilities. Therefore, the indicators employed should not be interpreted as the causes of systemic banking crises but as the contributing factors to the banking system's vulnerability. Furthermore, it is pivotal to evaluate whether the estimated crisis probability is significantly different from the cut-off threshold, based on which regulators can better tailor their policy responses using these models. As proposed by Beutel et al. (2019), policy recommendations based on EWSs' predictions can be categorized as follows: (i) clear advice of preventive actions if predicted probabilities are significantly above the threshold; (ii) no actions required if predicted probabilities are significantly below the threshold; (iii) further investigations needed if predicted probabilities do not differ significantly from the threshold. This interesting issue is out of the scope of this paper and is therefore left for future research.

6. Conclusion

Crises, of any type and under any circumstances, cause severe damages to the economy and to society as a whole. Since it has been unveiled that the nature of the 2007/2008 credit crunch has been a systemic banking crisis, this type of financial failure has drawn more attention from academia and policymakers. In identifying its causes, much focus has hitherto been put on the connection between the banking sector's vulnerability and macroeconomic imbalances. Nonetheless, Nguyen et al. (2020) empirically show that a country's political environment also

plays a role in the occurrence of these distress events. With the purpose of providing an examination of the proposed relationship within an EWS setting – which has not been done in the paper of Nguyen et al. (2020), this paper has shown that some political indicators are, indeed, potentially useful in predicting systemic banking crises.

Specifically, besides conventional macro-financial indicators that can signal the buildups of vulnerabilities prior to a banking crisis (such as house price gap, credit growth rate, etc.), indicators that reflect the political ideology of the incumbent government and the time in office of the incumbent chief executive also have significant crisis predictive power. An EWS incorporating these political indicators thus outperforms, on average, the one using only macro-financial variables in several crisis-predicting exercises. In line with the literature, the results are obtained based on both in-sample and out-of-sample performance evaluations, with the latter using *K*-fold cross-validation and (quasi) real-time exercises as out-of-sample forecasting experiments. Furthermore, this paper shows that the results in favor of the model using political indicators hold: (i) under different performance measures; (ii) independent of the policymaker's preference between Type I errors (missing crises) and Type II errors (issuing false alarms); (iii) regardless of how post-crisis periods are treated; and (iv) robust to the choice of the detrending method.

Beyond those predictive performance comparisons, economic interpretations and policy implications of statistically significant indicators are also provided. While the effects of macro-financial variables on the likelihood of systemic banking crises are consistent with the theoretical frameworks and previous empirical evidence, results of the newly employed political indicators point out some interesting implications. First, the probability of observing a banking crisis in the forecasting window is more likely to be lower if the incumbent government is controlled by a left-leaning party. The fact that economic policies pursued by left-wing governments tend to impose more market regulations, which in turn help reduce the banking systems' vulnerability, might explain the result. Second, the time in office of chief executives is found to be negatively correlated with the pre-crisis probability, implying a longer time in office of chief executives – which is conducive to the continuity of economic policies – may help reduce systemic risks. Overall, these two results are in line with earlier empirical findings of Nguyen et al. (2020). Nevertheless, different from the study of Nguyen et al. (2020), we find that the variable indicating the timing of political elections is not a useful early-warning indicator for systemic banking crises. As Nguyen et al. (2020) only include independent variables if they are indicators of all three crisis types, which they consider, some important factors regarding each type are thus omitted, potentially resulting in confounded estimates. For example, their set of independent variables does not include the indicators capturing credit developments and house prices – those that are of crucial importance in studying systemic banking crises. This and the fact that their empirical model is not that of an EWS have prompted us to further investigate whether the suggested political indicators are indeed useful in an EWS for systemic banking crises.

Finally, this paper also shows that the connections between the studied political indicators and the pre-crisis probabilities weaken when a different crisis dataset (with fewer systemic banking crisis events defined) is employed. Notwithstanding, given that there exists a substantial number of unknown factors that give rise to crises, the indicators of the political environment do hold promise in devising a better early warning system. Having said that, further investigations on the subject are needed, e.g., using more advanced data or different modeling techniques, to comprehend the role of political environment on the likelihood of systemic banking crises as well as other types of financial crises.

References

- Alessi, Lucia and Carsten Detken (2011). “Real time early warning indicators for costly asset price boom/bust cycles: a role for global liquidity”. *European Journal of Political Economy* 27(3), pp. 520-533.
- Alessi, Lucia and Carsten Detken (2018). “Identifying excessive credit growth and leverage”. *Journal of Financial Stability* Vol. 35, pp. 215-225.
- Alesina, Alberto and Nouriel Roubini (1992). “Political Cycles in OECD Economies”. *The Review of Economic Studies* Vol. 59, No. 4, pp. 663-688.
- Alesina, Alberto., Nouriel Roubini, and Gerald D. Cohen (1997). *Political cycles and the macroeconomy*. Cambridge, MA: MIT Press.
- Allen, Franklin and Douglas Gale (2000). “Financial contagion”. *Journal of Political Economy* 108 (1), pp. 1-33.
- Arlot, Sylvain and Alain Celisse (2010). “A survey of cross-validation procedures for model selection”. *Statistical Surveys* 4, pp. 40-79.
- Arnone Marco, Bernard J. Laurens, Jean-François Segalotto, and Martin Sommer (2007). “Central Bank Autonomy: Lessons from Global Trends”. *IMF Working Paper* No. 07/88.
- Babecky, Jan, Tomas Havranek, Jakub Mateju, Marek Rusnak, Katerina Smidkova, and Borek Vasicek (2014). “Banking, Debt, and Currency Crises in Developed Countries: Stylized Facts and Early Warning Indicators.” *Journal of Financial Stability* 15, pp. 1-17.
- Basel Committee on Banking Supervision (2010). *Guidance for national authorities operating the countercyclical capital buffer*. Bank for International Settlements. Basel, Switzerland.
- Baum, Christopher F., Mustafa Caglayan, and Oleksandr Talavera (2010). Parliamentary election cycles and the Turkish banking sector. *Journal of Banking and Finance* 34(11), pp. 2709–2719.
- Bechtel, Michael M. (2009). “The political sources of systematic investment risk: Lessons from a consensus democracy”. *The Journal of Politics* 71, pp. 661-677.
- Beutel, Johannes, Sophia List, and Gregor von Schweinitz (2019). “Does machine learning help us predict banking crises?” *Journal of Financial Stability* 45, pp. 1-28.
- Breiman, Leo and Philip Spector (1992). “Submodel selection and evaluation in regression. the x-random case”. *International Statistical Review* 60(3), pp. 291-319.
- Brunnermeier, Markus K. and Martin Oehmke (2013). “Bubbles, Financial Crises, and Systemic Risk”. *Handbook of the Economics of Finance*. Amsterdam: Elsevier Print.
- Bussiere, Matthieu and Marcel Fratzscher (2006). “Towards a New Early Warning System of Financial Crises”. *Journal of International Money and Finance* 25(6), pp. 953–973.
- Caggiano, Giovanni, Pietro Calice, Leone Leonida, and George Kapetanios (2016). “Comparing logit-based early warning systems: Does the duration of systemic banking crises matter?”. *Journal of Empirical Finance* 37, pp. 104-116.
- Calomiris, Charles W., and Stephen H. Haber (2014). *Fragile by Design: The Political Origins of Banking Crises and Scarce Credit*. Princeton, NJ: Princeton Univ. Press.
- Casabianca, Elizabeth J., Michelle Catalano, Lorenzo Forni, Elena Giarda, and Simone Passeri (2019). “An Early Warning System for banking crises: From regression-based analysis to machine learning techniques.” *Marco Fanno Working Papers* No.235.
- Castro, Vitor and Rodrigo Martins (2021). “Government ideology and economic freedom”. *Journal of Comparative Economics* 49(1), pp. 73-91.
- Cruz, Cesi, Philip Keefer, and Carlos Scartascini (2018). *Database of Political Institutions 2017 (DPI2017)*. Inter-American Development Bank. Numbers for Development.

- Davis, Philip E. and Dilruba Karim (2008). "Comparing early warning systems for banking crises". *Journal of Financial Stability* 4, pp. 89-120.
- Demirgüç-Kunt, Asli and Enrica Detragiache (1998). "The Determinants of Banking Crises in Developing and Developed Countries". *IMF Staff Papers* Vol 45(1), pp. 81-109.
- Demirgüç-Kunt, Asli and Enrica Detragiache (2002). "Does deposit insurance increase banking system stability? An empirical investigation". *DeJournal of Monetary Economics* 49, pp. 1373-1406.
- Detken, Carsten, Olaf Weeken, Lucia Alessi, Diana Bonfim, Miguel M. Boucinha, Christian Castro, Sebastian Frontczak, Gaston Giordana, Julia Giese, Nadya Jahn, Jan Kakes, Benjamin Klaus, Jan H. Lang, Natalia Puzanova, and Peter Welz (2014). "Operationalizing the Countercyclical Capital Buffer: Indicator Selection, Threshold Identification, and Calibration Options". *ESRB Occasional Paper* No. 5.
- Drehmann, Matthias, Claudio Borio, and Kostas Tsatsaronis (2011). "Anchoring countercyclical capital buffers: The role of credit aggregates". *International Journal of Central Banking* 7(4), pp. 89-240.
- Edge, Rochelle M. And Ralf R. Meisenzahl (2011). "The unreliability of credit-to-GDP ratio gaps in real time: implications for countercyclical capital buffers". *International Journal of Central Banking* 7(4), pp. 261-298.
- Eichler, Stefan, and Karol Sobanski (2016). "National politics and bank default risk in the eurozone". *Journal of Financial Stability* 26, pp. 247-256.
- Fernandez-Villaverde, Jesus, Luis Garicano, and Tano Santos (2013). Political Credit Cycles: The Case of the Eurozone. *Journal of Economic Perspectives* 27 (3), pp. 145–166.
- Frankel, Jeffrey (2010). "Monetary Policy in Emerging Markets". *Handbook of Monetary Economics* 3, pp. 1439-1520.
- Frankel, Jeffrey and George Saravelos (2012). "Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis". *Journal of International Economics* 87 (2), pp. 216-231.
- Geisser, Seymour (1975). "The predictive sample reuse method with applications". *Journal of the American Statistical Association* 70(350), pp. 320-328.
- Hamdaoui, Mekki (2016). "Are banking crises in developed and developing countries predictable?" *Journal of Multinational Financial Management* 37-38, pp. 114-138.
- Herrera, Helios, Guillermo Ordóñez, and Christoph Trebesch (2020). "Political Booms, Financial Crises". *Journal of Political Economy* 128(2), pp. 507-543.
- Holopainen, Markus and Peter Sarlin (2017). "Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty". *Quantitative Finance* 17 (12), pp. 1933–1963.
- Hosmer, David W. Jr., Stanley Lemeshow, and Rodney X. Sturdivant (2013). *Applied Logistic Regression*. (3rd ed). Hoboken, NJ: Wiley.
- Jackowicz, Krzysztof, Oskar Kowalewski, and Lukasz Kozłowski (2013). "The influence of political factors on commercial banks in Central European countries". *Journal of Financial Stability* 9(4), pp. 759–777.
- Jokipii, Terhi, Reto Nyffeler and Stéphane Riederer (2021). "Exploring BIS credit-to-GDP gap critiques: the Swiss case". *Swiss Journal of Economics and Statistics* 157:7.
- Jordà, Oscar, Moritz Schularick, and Alan M. Taylor (2011). "Financial crises, credit booms and external imbalances: 140 years of lessons". *IMF Economic Review* 59, pp. 340–378.
- Jordà, Oscar, Moritz Schularick, and Alan M. Taylor (2015). "Leverage Bubbles". *Journal of Monetary Economics* 76, pp. S1-S20.

- Jordà, Oscar, Moritz Schularick, and Alan M. Taylor (2017). “Macrofinancial History and the New Business Cycle Facts”. *Federal Reserve Bank of San Francisco Working Paper* No. 2016- 3.
- Kaminsky, Graciela L. and Carmen M. Reinhart (1999). “The Twin Crises: The Causes of Banking and Balance-of-Payments Problems.” *American Economic Review* 89 (3), pp. 473-500.
- Kauko, Karlo (2014). “How to foresee banking crises? A survey of the empirical literature”. *Economic System* 38, pp. 289-308.
- Klomp, Jeroen and Jakob de Haan (2009). “Central bank independence and financial instability”. *Journal of Financial Stability* 5(4), pp. 321-338.
- Krueger, Anne O. (1993). *Political economy of policy reform in developing countries*. Cambridge: MIT Press
- Kuhn, Max and Kjell Johnson (2013). *Applied Predictive Modeling*. Springer.
- Laeven, Luc and Fabian Valencia (2008). “Systemic Banking Crises: A New Database”. *IMF Working Paper* No. 08/224.
- Laeven, Luc and Fabian Valencia (2013). “Systemic Banking Crises Database”, *IMF Economic Review* Vol. 61(2), pp. 225-270.
- Laeven, Luc and Fabian Valencia (2018). “Systemic Banking Crises Revisited”. *IMF Working Paper* No. 18/206.
- Laeven, Luc and Fabian Valencia (2020). “Systemic Banking Crises Database II”. *IMF Economic Review* 68, pp. 307-361.
- Liu, Wai-Man and Phong T.H. Ngo (2014). “Elections, political competition and bank failure”. *Journal of Financial Economics* 112, pp. 251–268.
- grennLo Duca, Marco, Anne Koban, Marisa Basten, Elias Bengtsson, Benjamin Klaus, Piotr Kusmierczyk, Jan Hannes Lang, Carsten Detken, and Tuomas Peltonen (2017). “A New Database for Financial Crises in European Countries. ECB/ESRB EU crises database”. *ECB Occasional Paper Series* No. 194.
- Manasse, Paolo and Nouriel Roubini (2009). “Rules of thumb for sovereign debt crises”. *Journal of International Economics* 78, pp. 192-205.
- McCarty, Nolan, Keith T. Poole, and Howard Rosenthal (2013). *Political Bubbles: Financial Crises and the Failure of American Democracy*. Princeton, NJ: Princeton Univ. Press.
- Mishkin, Frederic S. (2008). “How should we respond to asset price bubbles?” *Financial Stability Review* 12, pp. 65-74.
- Nguyen, Thanh C., Vitor Castro, and Justine Wood (2020). “Political environment and financial crises”. *International Journal of Finance & Economics*, pp. 1-22.
- Nguyen, Thanh C., Vitor Castro, and Justine Wood (2022). “A new comprehensive database of financial crises: Identification, frequency, and duration”. *Economics Modelling* 108, pp. 105770.
- Nordhaus, William D. (1975). “The Political Business Cycle”. *The Review of Economic Studies* Vol. 42, No. 2, pp. 169-190.
- Obstfeld, Maurice (2012). “Does the current account still matter?” *American Economic Review* 102(3), pp. 1-23.
- Phan, Dinh H. B., Bernard N. Iyke, Susan S. Sharma, and Yoga Affandi (2021). “Economic Policy Uncertainty and Financial Stability – Is there a relation?” *Economic Modelling* 94, pp. 1018-1029.
- Potrafke, Niklas (2010). “Does government ideology influence deregulation of product markets? Empirical evidence from OECD countries”. *Public Choice* 143(1), pp. 135-155.

- Reinhart, Carmen M. and Kenneth S. Rogoff (2008). “Is the 2007 US Sub-Prime Financial Crisis So Different? An International Historical Comparison”. *The American Economic Review* 98 (2), pp. 339-344.
- Reinhart, Carmen M. and Kenneth S. Rogoff (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton and Woodstock.
- Reinhart, Carmen M. and Kenneth S. Rogoff (2011). “From Financial Crash to Debt Crisis”. *American Economic Review* 101, pp. 1676-1706.
- Rogoff, Kenneth S. and Anne Sibert (1988). “Elections and macroeconomic policy cycles”. *The Review of Economic Studies* 55(1), pp. 1-16.
- Sarlin, Peter (2013). “On policymakers’ loss functions and the evaluation of early warning systems”. *Economics Letters* 119(1), pp. 1-7.
- Sarlin, Peter and Gregor von Schweinitz (2021). “Optimizing policymakers' loss functions in crisis prediction: before, within or after?”. *Macroeconomic Dynamics* 25, pp. 100-123.
- Schularick, Moritz and Alan M. Taylor (2012). “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008”. *American Economic Review* 102(2), pp. 1029-1061.
- Shao, Jun (1993). “Linear model selection by cross-validation”. *Journal of the American Statistical Association* 88(422), pp. 486-494.
- Tanaka, Katsuyuki, Takuji Kinkyo, and Shigeyuki Hamori (2016). “Random Forests-Based Early Warning System for Bank Failures”. *Economics Letters* 148, pp. 118-121.
- Vanwinckelen, Gitte and Hendrik Blockeel (2012). *On estimating model accuracy with repeated cross-validation*. Proceedings of the 21st Belgian Dutch Conference on Machine Learning. Ghent.
- Verbeek, Marno (2012). *A Guide to Modern Econometrics* (4th ed.). John Wiley & Sons, Ltd.

Appendix A. Data

Table A.1: List of countries and systemic banking crisis events during the period 1975-2017.

Country	Systemic banking crises	
	Casabianca et al. (2019)	Laeven and Valencia (2020)
Australia	1989	—
Austria	2008	2008 – 2012
Belgium	2008	2008 – 2012
Canada	1983	—
Czech Republic	1996	1996 – 2000
Denmark	2008	2008 – 2009
Estonia	1992	1992 – 1994
Finland	1991	1991 – 1995
France	1994 // 2008	2008 – 2009
Germany	2008	2008 – 2009
Greece	2008	2008 – 2012
Iceland	2008	2008 – 2012
Ireland	2008	2008 – 2012
Israel	1983	1983 – 1986
Italy	1990 // 2008	2008 – 2009
Japan	1992 // 1997	1997 – 2001
Latvia	1995 // 2008	1995 – 1996 // 2008 – 2012
Lithuania	1995	1995 – 1996
Luxembourg	2008	2008 – 2012
New Zealand	—	—
Norway	1987 // 1991	1991 – 1993
Portugal	2008	2008 – 2012
Singapore	1982	—
Slovakia	1998	1998 – 2002
Slovenia	1992 // 2008	1992 – 1992 // 2008 – 2012
South Korea	1983 // 1985 // 1997	1997 – 1998
Spain	1977 // 2008	1977 – 1981 // 2008 – 2012
Sweden	1991 // 2008	1991 – 1995 // 2008 – 2009
Switzerland	1991 // 2008	2008 – 2009
The Netherlands	2008	2008 – 2009
The United Kingdom	1984 // 1991 // 1995 // 2007	2007 – 2011
The United States	1984 // 1988 // 2007	1988 – 1988 // 2007 – 2011

Notes: In the crisis dataset of Casabianca et al. (2019), 1977 is a crisis year in Germany. However, there was no banking crisis in Germany in 1977. Therefore, we change 1977 to a non-crisis year for Germany in our final dataset.

Table A.2: List of macro-financial indicators

Indicator name	Description	Transformation	Source(s)
House price gap	Deviation from trend of the nominal residential property price index	Absolute trend deviation using HP filter ($\lambda = 100$)	OECD, Singapore government's database
Credit growth rate	Growth rate of total credit to private non-financial sector	Annual percentage growth rate	BIS, WDI
Credit-to-GDP gap	Deviation from trend of the ratio between total credit to private non-financial sector and nominal GDP	Absolute trend deviation using HP filter ($\lambda = 100$)	BIS, WDI
Current-account-to-GDP	Nominal current account balance as a share of nominal GDP	Ratio	WDI
Inflation rate	GDP deflator	Annual percentage rate	WDI
GDP growth rate	Growth rate of gross domestic product	Annual percentage growth rate	WDI
World GDP growth rate	Growth rate of world GDP	Annual percentage growth rate	WDI
US 10-year treasury rate	US 10-year constant maturity rate	Annual average percentage rate	FRED
Trade-to-GDP gap	Deviation from trend of the ratio between total trade volume and nominal GDP, in which total trade volume is the sum of exports and imports of goods and services.	Absolute trend deviation using HP filter ($\lambda = 100$)	WDI
Public-debt-to-GDP	General government debt as a share of nominal GDP	Ratio	IMF
Reserves-to-GDP	Total reserves as a share of nominal GDP	Ratio	WDI

Table A.3: List of political indicators

Indicator name	Description	Source
Election time	Dummy variable that takes the value of 1 in the election years and the first year after each election, and 0 otherwise.	DPI2017 (source variable: LEGELEC, EXELEC)
Time in office of executive	The number of years that a chief executive holds office.	DPI2017 (source variable: YRSOFFC)
Right-wing government	Dummy variable that takes the value of 1 if the incumbent executive belongs to a right-wing party, and 0 if else.	DPI2017 (source variable: EXECRLC)
Left-wing government	Dummy variable that takes the value of 1 if the incumbent executive belongs to a left-wing party, and 0 if else.	DPI2017 (source variable: EXECRLC)
Center government	Dummy variable that takes the value of 1 if the incumbent executive belongs to a centrist party, and 0 if else.	DPI2017 (source variable: EXECRLC)
Majority government	Dummy variable that takes the value of 1 if the incumbent government has more than half of the total vote share and 0 if else.	DPI2017 (source variable: GOV1VOTE, GOV2VOTE, GOV3VOTE, GOVOTHVT)

Table A.4: Descriptive statistics

	Observations	Mean	Standard Deviation	Min	Max
<i>Macro-financial variables:</i>					
House price gap	708	-.1392	6.4147	-31.1961	38.9688
Credit growth rate	708	8.2494	13.2945	-20.9841	87.0078
Credit-to-GDP gap	708	-.8076	8.8362	-42.7116	123.104
Current-account-to-GDP	708	1.1233	5.7876	-23.0847	27.1433
Inflation rate	708	3.3661	3.588	-5.2139	20.8158
GDP growth rate	708	2.7806	2.5981	-14.4339	25.1625
World GDP growth rate	708	3.0621	1.1024	-1.674	5.271
US 10-year treasury rate	708	5.4157	3.0497	1.81	13.92
Trade-to-GDP gap	708	.2004	5.658	-30.1709	45.1299
Public-debt-to-GDP	708	61.2082	36.8963	3.7655	236.3346
Reserves-to-GDP	708	.1282	.1826	.0034	1.1928
<i>Political variables:</i>					
Election time	708	.5918	.4918	0	1
Time in office of executive	708	3.9958	3.031	1	24
Right-wing government	708	.3941	.489	0	1
Left-wing government	708	.3715	.4835	0	1
Center government	708	0.0777	0.2679	0	1

Source: Own elaborations.

Table A.5: Correlation matrix

	House price gap	Credit growth rate	Credit-to- GDP gap	Current- account-to- GDP	Inflation rate	GDP growth rate
House price gap	1.0000					
Credit growth rate	0.2535	1.0000				
Credit-to-GDP gap	0.1343	0.1248	1.0000			
Current-account-to-GDP	-0.1361	-0.1587	-0.1733	1.0000		
Inflation rate	0.1093	0.1870	0.0088	-0.2385	1.0000	
GDP growth rate	0.1528	0.3428	-0.1269	0.0642	0.1334	1.0000
World GDP growth rate	0.1447	0.2983	-0.0320	-0.0416	0.0458	0.4377
US 10-year treasury rate	0.0571	0.1439	-0.0060	-0.1570	0.6381	0.0792
Trade-to-GDP gap rate	-0.0799	-0.0579	-0.0457	0.0315	0.0667	0.2060
Public-debt-to-GDP	-0.1877	-0.2436	-0.0738	0.1316	-0.3327	-0.1730
Reserves-to-GDP	0.0188	-0.0024	0.0147	0.5968	-0.1445	0.1850

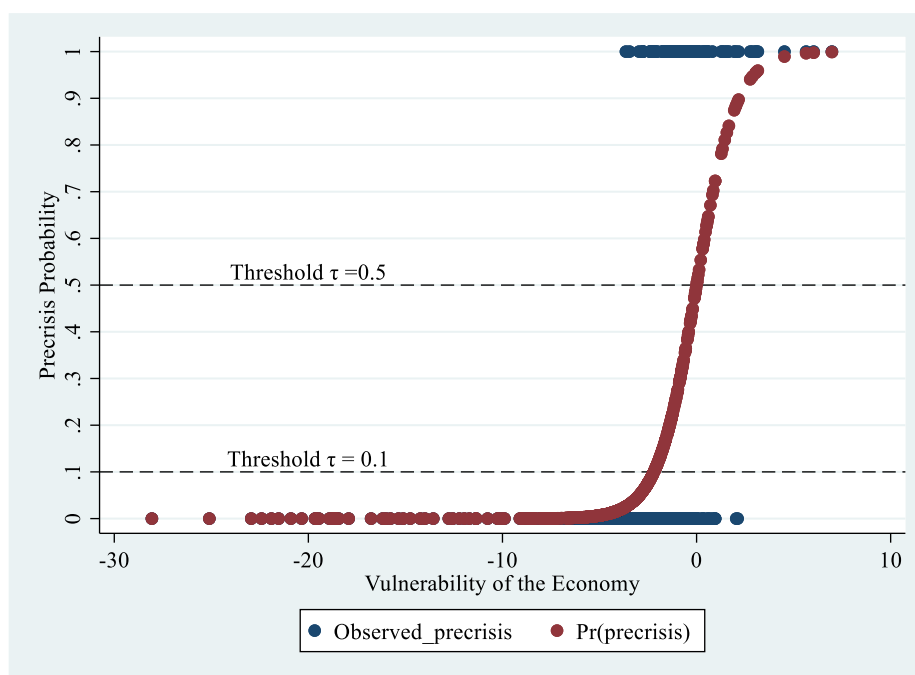
Source: Own elaborations.

Table A.5 (cont.): Correlation matrix

	World GDP growth rate	US 10-year treasury rate	Trade-to- GDP gap	Public- debt-to- GDP	Reserves- to-GDP
World GDP growth rate	1.0000				
US 10-year treasury rate	0.0308	1.0000			
Trade-to-GDP gap	0.1745	0.0511	1.0000		
Public-debt-to-GDP	-0.0490	-0.2841	0.0050	1.0000	
Reserves-to-GDP	-0.0652	-0.0939	-0.0135	0.1305	1.0000

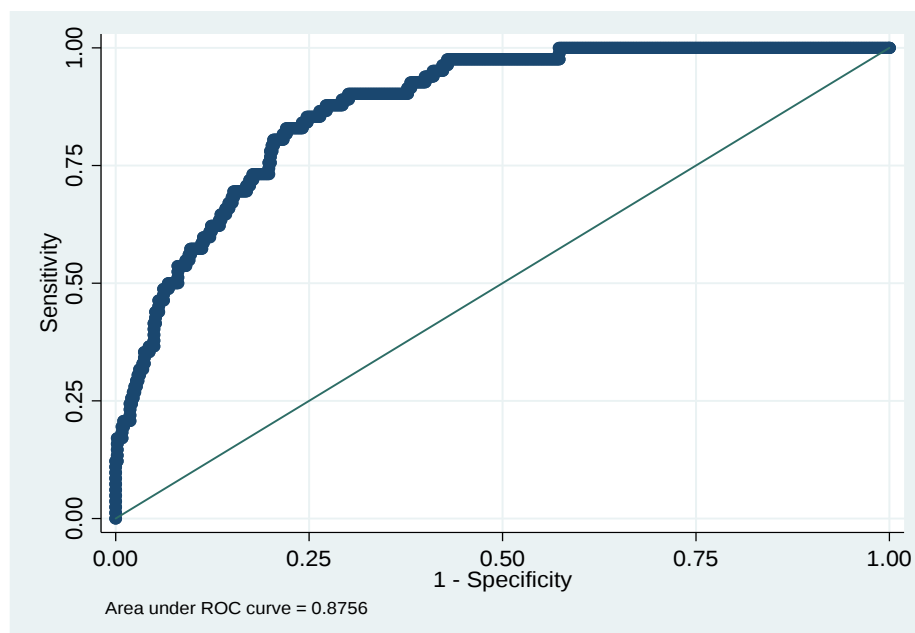
Source: Own elaborations.

Figure A.1: Illustration of Logit-based EWS



Source: Own elaborations. Notes: $Pr(precrisis)$ stands for predicted pre-crisis probabilities.

Figure A.2: Example of ROC curve



Source: Own elaborations. Notes: The 45-degree line corresponds to an AUC of 0.5, so-called line of no-discrimination.

Appendix B. Results

Table B.1: Estimated coefficients: Logit-based EWS with political indicators

	Benchmark model	Robustness Test (1)	Robustness Test (2)	Robustness Test (3)	Robustness Test (4)	Robustness Test (5)
<i>Macro-financial variables:</i>						
House price gap	0.2388*** (0.0350)	0.2206*** (0.0320)	0.2386*** (0.0354)	0.2467*** (0.0391)	0.1124*** (0.0426)	0.2384*** (0.0349)
Credit growth rate	0.0346** (0.0140)	0.0318** (0.0126)	0.0350** (0.0139)	0.0304** (0.0142)	0.0320** (0.0135)	0.0350** (0.0139)
Credit-to-GDP gap	0.0624*** (0.0165)	0.0516*** (0.0137)	0.0596*** (0.0161)	0.0677*** (0.0195)	0.0729*** (0.0246)	0.0622*** (0.0166)
Current-account-to-GDP	0.1137*** (0.0437)	0.0809** (0.0406)	0.1076** (0.0434)	0.0936** (0.0475)	0.0159 (0.0440)	0.1121** (0.0436)
Inflation rate	-0.1710*** (0.0554)	-0.1461*** (0.0478)	-0.1731*** (0.0559)	-0.2041** (0.0795)	-0.0801 (0.0488)	-0.1543** (0.0626)
GDP growth rate	-0.0565 (0.0656)	-0.0307 (0.0594)	-0.0556 (0.0682)	-0.1206 (0.1327)	-0.1018 (0.0981)	-0.0582 (0.0660)
World GDP growth rate	0.5941*** (0.1567)	0.5466*** (0.1351)	0.5863*** (0.1551)	1.5056*** (0.2806)	0.6867*** (0.1830)	0.5953*** (0.1569)
US 10-year treasury rate	0.4037*** (0.0716)	0.3889*** (0.0617)	0.4185*** (0.0723)	0.1078 (0.0767)	0.3125*** (0.0718)	0.4049*** (0.0718)
Trade-to-GDP gap	0.0610 (0.0374)	0.0554 (0.0345)	0.0617* (0.0367)	0.0753 (0.0480)	0.1486*** (0.0413)	0.0626* (0.0378)
Public-debt-to-GDP	0.0011 (0.0055)	-0.0064 (0.0055)	0.0009 (0.0055)	-0.0051 (0.0059)	-0.0044 (0.0047)	0.0011 (0.0055)
Reserves-to-GDP	-21.0344*** (4.3541)	-17.4111*** (3.3701)	-21.2011*** (4.4076)	-23.4028*** (4.5870)	-16.4746*** (4.2341)	-21.0753*** (4.3698)
<i>Political variables:</i>						
Election time	-0.4658 (0.3130)	-0.3485 (0.2921)	-0.4517 (0.3128)	-0.3654 (0.3865)	-0.2416 (0.2861)	-0.3267 (0.4549)
Time in office of executive	-0.1038** (0.0472)	-0.0814* (0.0429)	-0.0989** (0.0474)	-0.0946* (0.0553)	-0.0960** (0.0460)	-0.1046** (0.0474)
Right-wing government	-0.6166 (0.6870)	-0.4794 (0.6777)	-0.7661 (0.6966)	0.0537 (1.1307)	-0.8904 (0.5846)	-0.6456 (0.6833)
Left-wing government	-1.4967** (0.6968)	-1.4214** (0.6904)	-1.5824** (0.7067)	-0.4813 (1.1206)	-2.0429*** (0.6010)	-1.5229** (0.6927)
Center government	-1.0949 (0.8771)	-1.2153 (0.8652)	-1.1130 (0.8860)	-1.0155 (1.3643)	-1.9290** (0.7940)	-1.1427 (0.8934)
Election time*Inflation rate	—	—	—	—	—	-0.0344 (0.0796)
Pseudo R-squared	0.3840	0.3551	0.3782	0.4639	0.2939	0.3843
Observations	708	809	707	747	683	708

Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Table B.1 (cont.): Estimated coefficients: Logit-based EWS with political indicators

	Robustness Test (6)	Robustness Test (7)	Robustness Test (8)	Robustness Test (9)	Robustness Test (10)	Robustness Test (11)
<i>Macro-financial variables:</i>						
House price gap	0.2134*** (0.0338)	0.2366*** (0.0351)	0.2342*** (0.0343)	0.2352*** (0.0346)	0.2396*** (0.0344)	0.2396*** (0.0341)
Credit growth rate	0.0322** (0.0136)	0.0326** (0.0140)	0.0345** (0.0136)	0.0338** (0.0133)	0.0322** (0.0133)	0.0321** (0.0132)
Credit-to-GDP gap	0.0561*** (0.0151)	0.0635*** (0.0164)	0.0621*** (0.0161)	0.0624*** (0.0164)	0.0609*** (0.0160)	0.0609*** (0.0160)
Current-account-to-GDP	0.0993** (0.0409)	0.1135*** (0.0433)	0.1133*** (0.0437)	0.1140*** (0.0435)	0.1152*** (0.0438)	0.1152*** (0.0438)
Inflation rate	-0.1889*** (0.0571)	-0.1688*** (0.0567)	-0.1651*** (0.0546)	-0.1634*** (0.0542)	-0.1617*** (0.0518)	-0.1615*** (0.0500)
GDP growth rate	-0.0548 (0.0716)	-0.0557 (0.0658)	-0.0532 (0.0674)	-0.0575 (0.0678)	-0.0529 (0.0671)	-0.0530 (0.0663)
World GDP growth rate	0.5799*** (0.1748)	0.6033*** (0.1530)	0.5811*** (0.1614)	0.5866*** (0.1601)	0.5752*** (0.1604)	0.5752*** (0.1599)
US 10-year treasury rate	—	0.3936*** (0.0721)	0.3979*** (0.0709)	0.3968*** (0.0701)	0.3842*** (0.0675)	0.3839*** (0.0625)
Trade-to-GDP gap	0.0347 (0.0376)	0.0599 (0.0373)	0.0604 (0.0378)	0.0612 (0.0382)	0.0529 (0.0357)	0.0529 (0.0356)
Public-debt-to-GDP	0.0001 (0.0053)	0.0022 (0.0061)	0.0013 (0.0054)	0.0011 (0.0054)	0.0015 (0.0054)	0.0015 (0.0051)
Reserves-to-GDP	-20.2506*** (4.4492)	-20.8845*** (4.3246)	-20.5300*** (4.2020)	-20.3472*** (4.1140)	-18.6439*** (3.7754)	-18.6357*** (3.8491)
FFR	0.3365*** (0.0517)	—	—	—	—	—
<i>Political variables:</i>						
Election time	-0.5419* (0.3217)	-0.4517 (0.3131)	-0.2800 (0.3754)	—	—	—
Time in office of executive	-0.1157** (0.0479)	-0.1093** (0.0477)	-0.0978** (0.0470)	-0.0978** (0.0472)	-0.0979** (0.0473)	-0.0980** (0.0469)
Right-wing government	-0.4541 (0.6389)	-0.7891 (0.7179)	-0.6013 (0.6996)	-0.5912 (0.6998)	-0.0065 (0.5233)	—
Left-wing government	-1.4088** (0.6656)	-1.6855** (0.7270)	-1.4632** (0.7066)	-1.4421** (0.7070)	-0.8289 (0.5061)	-0.8242*** (0.3152)
Center government	-1.0652 (0.8462)	-1.2515 (0.8895)	-1.0135 (0.8689)	-0.9658 (0.8774)	—	—
Majority government	—	-0.1764 (0.3101)	—	—	—	—
Pseudo R-squared	0.3956	0.3722	0.3805	0.3794	0.3767	0.3767
Observations	708	653	708	708	708	708

Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Table B.2: Average marginal effects: Logit-based EWS with political indicators

	Benchmark model	Robustness Test (1)	Robustness Test (2)	Robustness Test (3)	Robustness Test (4)	Robustness Test (5)
<i>Macro-financial variables:</i>						
House price gap	0.0160*** (0.0021)	0.0148*** (0.0020)	0.0162*** (0.0022)	0.0115*** (0.0016)	0.0088*** (0.0031)	0.0159*** (0.0021)
Credit growth rate	0.0023** (0.0009)	0.0021** (0.0008)	0.0024** (0.0009)	0.0014** (0.0006)	0.0025** (0.0010)	0.0023** (0.0009)
Credit-to-GDP gap	0.0042*** (0.0011)	0.0035*** (0.0009)	0.0040*** (0.0011)	0.0031*** (0.0009)	0.0057*** (0.0019)	0.0042*** (0.0011)
Current-account-to-GDP	0.0076*** (0.0029)	0.0054** (0.0027)	0.0073** (0.0029)	0.0044* (0.0023)	0.0012 (0.0034)	0.0075*** (0.0029)
Inflation rate	-0.0114*** (0.0038)	-0.0098*** (0.0033)	-0.0117*** (0.0039)	-0.0095*** (0.0037)	-0.0063 (0.0039)	-0.0103** (0.0043)
GDP growth rate	-0.0038 (0.0043)	-0.0021 (0.0040)	-0.0038 (0.0046)	-0.0056 (0.0060)	-0.0080 (0.0076)	-0.0039 (0.0044)
World GDP growth rate	0.0397*** (0.0107)	0.0366*** (0.0093)	0.0397*** (0.0108)	0.0701*** (0.0123)	0.0539*** (0.0141)	0.0398*** (0.0108)
US 10-year treasury rate	0.0270*** (0.0050)	0.0261*** (0.0043)	0.0283*** (0.0051)	0.0050 (0.0036)	0.0245*** (0.0057)	0.0271*** (0.0050)
Trade-to-GDP gap	0.0041* (0.0024)	0.0037 (0.0023)	0.0042* (0.0024)	0.0035 (0.0021)	0.0117*** (0.0031)	0.0042* (0.0025)
Public-debt-to-GDP	0.0001 (0.0004)	-0.0004 (0.0004)	0.0001 (0.0004)	-0.0002 (0.0003)	-0.0003 (0.0004)	0.0001 (0.0004)
Reserves-to-GDP	-1.4063*** (0.2753)	-1.1668*** (0.2222)	-1.4361*** (0.2821)	-1.0891*** (0.2019)	-1.2929*** (0.3291)	-1.4085*** (0.2762)
<i>Political variables:</i>						
Election time	-0.0311 (0.0207)	-0.0234 (0.0195)	-0.0306 (0.0209)	-0.0170 (0.0181)	-0.0190 (0.0225)	-0.0218 (0.0302)
Time in office of executive	-0.0069** (0.0032)	-0.0055* (0.0029)	-0.0067** (0.0033)	-0.0044* (0.0026)	-0.0075** (0.0037)	-0.0070** (0.0032)
Right-wing government	-0.0412 (0.0459)	-0.0321 (0.0454)	-0.0519 (0.0471)	0.0025 (0.0526)	-0.0699 (0.0460)	-0.0432 (0.0456)
Left-wing government	-0.1001** (0.0461)	-0.0953** (0.0461)	-0.1072** (0.0474)	-0.0224 (0.0522)	-0.1603*** (0.0462)	-0.1018** (0.0459)
Center government	-0.0732 (0.0577)	-0.0814 (0.0573)	-0.0754 (0.0591)	-0.0473 (0.0631)	-0.1514** (0.0615)	-0.0764 (0.0587)
Election time*Inflation rate	—	—	—	—	—	-0.0344 (0.0796)
Pseudo R-squared	0.3840	0.3551	0.3782	0.4639	0.2939	0.3843
Observations	708	809	707	747	683	708

Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Table B.2 (cont.): Average marginal effects: Logit-based EWS with political indicators

	Robustness Test (6)	Robustness Test (7)	Robustness Test (8)	Robustness Test (9)	Robustness Test (10)	Robustness Test (11)
<i>Macro-financial variables:</i>						
House price gap	0.0139*** (0.0020)	0.0171*** (0.0023)	0.0157*** (0.0021)	0.0158*** (0.0022)	0.0161*** (0.0022)	0.0161*** (0.0022)
Credit growth rate	0.0021** (0.0009)	0.0024** (0.0010)	0.0023** (0.0009)	0.0023** (0.0009)	0.0022** (0.0009)	0.0022** (0.0009)
Credit-to-GDP gap	0.0037*** (0.0010)	0.0046*** (0.0011)	0.0042*** (0.0010)	0.0042*** (0.0011)	0.0041*** (0.0011)	0.0041*** (0.0011)
Current-account-to-GDP	0.0065** (0.0027)	0.0082*** (0.0031)	0.0076*** (0.0029)	0.0077*** (0.0029)	0.0078*** (0.0030)	0.0078*** (0.0030)
Inflation rate	-0.0123*** (0.0038)	-0.0122*** (0.0042)	-0.0111*** (0.0038)	-0.0110*** (0.0038)	-0.0109*** (0.0036)	-0.0109*** (0.0035)
GDP growth rate	-0.0036 (0.0046)	-0.0040 (0.0047)	-0.0036 (0.0045)	-0.0039 (0.0045)	-0.0036 (0.0045)	-0.0036 (0.0044)
World GDP growth rate	0.0379*** (0.0115)	0.0435*** (0.0114)	0.0390*** (0.0110)	0.0394*** (0.0109)	0.0388*** (0.0110)	0.0388*** (0.0110)
US 10-year treasury rate	—	0.0284*** (0.0054)	0.0267*** (0.0050)	0.0266*** (0.0050)	0.0259*** (0.0049)	0.0259*** (0.0044)
Trade-to-GDP gap	0.0023 (0.0024)	0.0043* (0.0026)	0.0041 (0.0025)	0.0041 (0.0025)	0.0036 (0.0024)	0.0036 (0.0024)
Public-debt-to-GDP	0.0000 (0.0003)	0.0002 (0.0004)	0.0001 (0.0004)	0.0001 (0.0004)	0.0001 (0.0004)	0.0001 (0.0003)
Reserves-to-GDP	-1.3225*** (0.2691)	-1.5075*** (0.2943)	-1.3769*** (0.2701)	-1.3659*** (0.2653)	-1.2563*** (0.2480)	-1.2557*** (0.2513)
FFR	0.0220*** (0.0034)	—	—	—	—	—
<i>Political variables:</i>						
Election time	-0.0354* (0.0206)	-0.0326 (0.0223)	-0.0188 (0.0252)	—	—	—
Time in office of executive	-0.0076** (0.0032)	-0.0079** (0.0035)	-0.0066** (0.0032)	-0.0066** (0.0032)	-0.0066** (0.0032)	-0.0066** (0.0032)
Right-wing government	-0.0297 (0.0415)	-0.0570 (0.0518)	-0.0403 (0.0469)	-0.0397 (0.0470)	-0.0004 (0.0353)	—
Left-wing government	-0.0920** (0.0426)	-0.1217** (0.0521)	-0.0981** (0.0471)	-0.0968** (0.0472)	-0.0559 (0.0345)	-0.0555*** (0.0212)
Center government	-0.0696 (0.0542)	-0.0903 (0.0631)	-0.0680 (0.0575)	-0.0648 (0.0582)	—	—
Majority government	—	-0.0127 (0.0224)	—	—	—	—
Pseudo R-squared	0.3956	0.3722	0.3805	0.3794	0.3767	0.3767
Observations	708	653	708	708	708	708

Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Table B.3: Estimated coefficients: Logit-based EWS without political indicators

	Benchmark model	Robustness Test (1)	Robustness Test (2)	Robustness Test (3)	Robustness Test (4)	Robustness Test (6)
House price gap	0.2469*** (0.0328)	0.2284*** (0.0305)	0.2463*** (0.0330)	0.2522*** (0.0389)	0.1267*** (0.0461)	0.2207*** (0.0314)
Credit growth rate	0.0322** (0.0132)	0.0321*** (0.0119)	0.0324** (0.0131)	0.0284** (0.0132)	0.0331*** (0.0124)	0.0299** (0.0127)
Credit-to-GDP gap	0.0608*** (0.0156)	0.0503*** (0.0126)	0.0583*** (0.0153)	0.0654*** (0.0190)	0.0569*** (0.0207)	0.0549*** (0.0144)
Current-account-to-GDP	0.1272*** (0.0432)	0.0963** (0.0394)	0.1212*** (0.0423)	0.0929** (0.0467)	0.0207 (0.0436)	0.1155*** (0.0406)
Inflation rate	-0.1542*** (0.0478)	-0.1520*** (0.0451)	-0.1524*** (0.0479)	-0.2099*** (0.0746)	-0.0838* (0.0440)	-0.1637*** (0.0489)
GDP growth rate	-0.0390 (0.0700)	-0.0061 (0.0598)	-0.0399 (0.0718)	-0.1187 (0.1375)	-0.1082 (0.1108)	-0.0406 (0.0782)
World GDP growth rate	0.5771*** (0.1582)	0.5141*** (0.1321)	0.5662*** (0.1558)	1.5095*** (0.2745)	0.6440*** (0.1838)	0.5590*** (0.1788)
US 10-year treasury rate	0.3872*** (0.0639)	0.3921*** (0.0595)	0.3925*** (0.0639)	0.1290* (0.0726)	0.2818*** (0.0690)	—
Trade-to-GDP gap	0.0480 (0.0388)	0.0431 (0.0323)	0.0500 (0.0387)	0.0679 (0.0429)	0.1362*** (0.0395)	0.0195 (0.0396)
Public-debt-to-GDP	0.0035 (0.0054)	-0.0036 (0.0052)	0.0031 (0.0054)	-0.0006 (0.0057)	-0.0026 (0.0046)	0.0030 (0.0053)
Reserves-to-GDP	-17.8184*** (3.6960)	-15.1166*** (2.8049)	-17.7832*** (3.7836)	-20.9603*** (4.5417)	-11.7672*** (4.1492)	-16.8878*** (3.9122)
FFR	—	—	—	—	—	0.3159*** (0.0460)
Pseudo R-squared	0.3569	0.3306	0.3522	0.4521	0.2508	0.3653
Observations	708	809	707	747	683	708

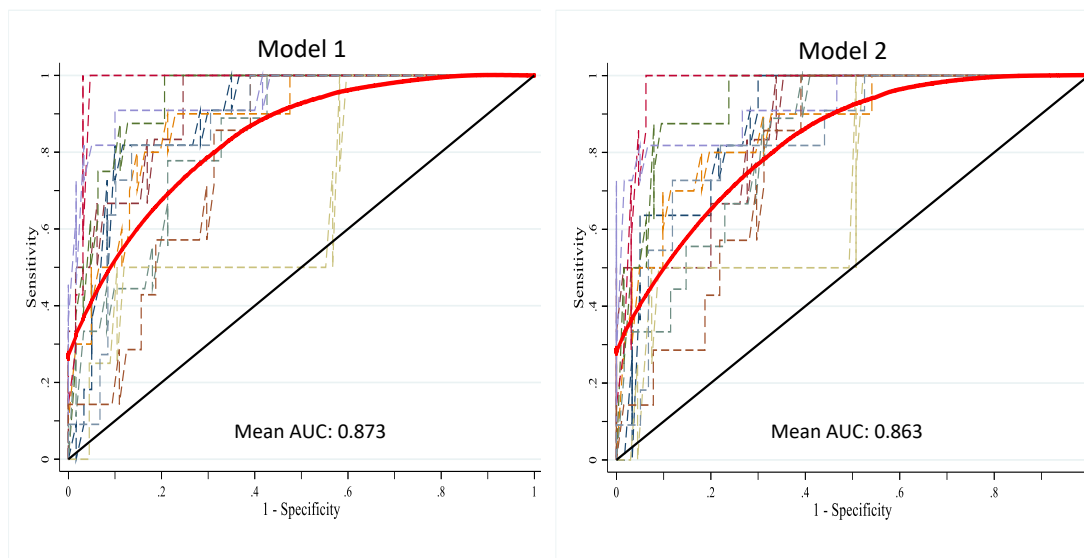
Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Table B.4: Average marginal effects: Logit-based EWS without political indicators

	Benchmark model	Robustness Test (1)	Robustness Test (2)	Robustness Test (3)	Robustness Test (4)	Robustness Test (6)
House price gap	0.0173*** (0.0021)	0.0159*** (0.0020)	0.0174*** (0.0021)	0.0120*** (0.0016)	0.0106*** (0.0036)	0.0152*** (0.0019)
Credit growth rate	0.0023** (0.0009)	0.0022*** (0.0008)	0.0023** (0.0009)	0.0014** (0.0006)	0.0028*** (0.0010)	0.0021** (0.0009)
Credit-to-GDP gap	0.0043*** (0.0011)	0.0035*** (0.0009)	0.0041*** (0.0011)	0.0031*** (0.0009)	0.0047*** (0.0017)	0.0038*** (0.0010)
Current-account-to-GDP	0.0089*** (0.0031)	0.0067** (0.0028)	0.0086*** (0.0030)	0.0044* (0.0023)	0.0017 (0.0036)	0.0079*** (0.0028)
Inflation rate	-0.0108*** (0.0034)	-0.0106*** (0.0032)	-0.0108*** (0.0035)	-0.0100*** (0.0034)	-0.0070* (0.0038)	-0.0113*** (0.0034)
GDP growth rate	-0.0027 (0.0049)	-0.0004 (0.0042)	-0.0028 (0.0050)	-0.0057 (0.0064)	-0.0090 (0.0091)	-0.0028 (0.0053)
World GDP growth rate	0.0404*** (0.0112)	0.0359*** (0.0093)	0.0400*** (0.0112)	0.0720*** (0.0118)	0.0537*** (0.0150)	0.0384*** (0.0122)
US 10-year treasury rate	0.0271*** (0.0048)	0.0273*** (0.0044)	0.0278*** (0.0048)	0.0062* (0.0034)	0.0235*** (0.0059)	—
Trade-to-GDP gap	0.0034 (0.0027)	0.0030 (0.0022)	0.0035 (0.0027)	0.0032* (0.0020)	0.0113*** (0.0032)	0.0013 (0.0027)
Public-debt-to-GDP	0.0002 (0.0004)	-0.0003 (0.0004)	0.0002 (0.0004)	-0.0000 (0.0003)	-0.0002 (0.0004)	0.0002 (0.0004)
Reserves-to-GDP	-1.2487*** (0.2456)	-1.0544*** (0.1907)	-1.2575*** (0.2534)	-0.9995*** (0.2085)	-0.9807*** (0.3422)	-1.1615*** (0.2538)
FFR	—	—	—	—	—	0.0217*** (0.0034)
Pseudo R-squared	0.3569	0.3306	0.3522	0.4521	0.2508	0.3653
Observations	708	809	707	747	683	708

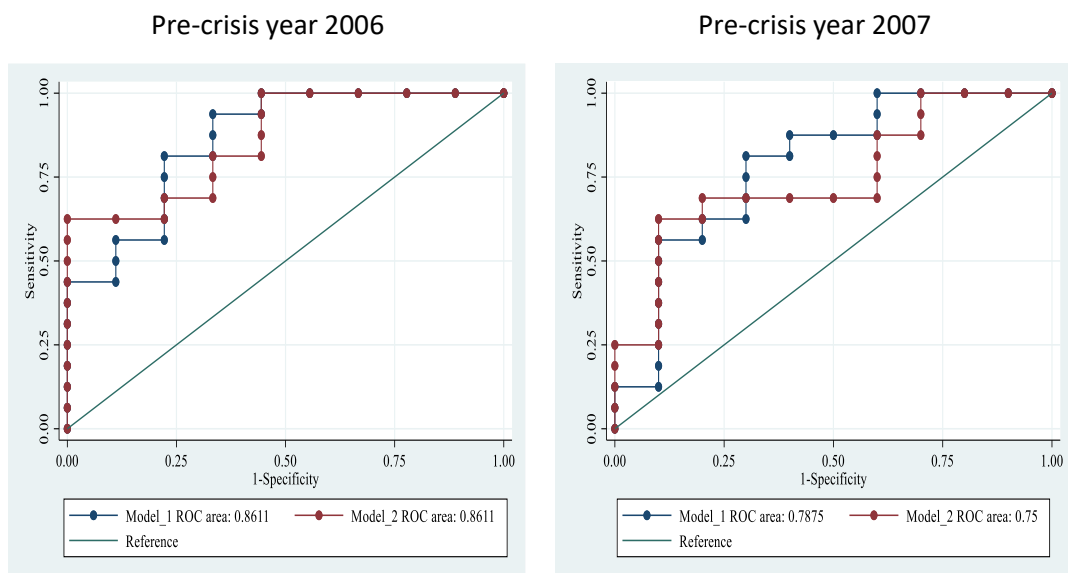
Source: Own elaborations. Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively. Robust standard errors are given in parentheses. See also Table 5 for the specification of robustness tests.

Figure B.1: ROC curves for the 10-fold cross-validation exercise



Source: Own elaborations. Notes: Model 1 and Model 2 correspond to EWSs with and without political indicators, respectively. For each model, the solid red curve represents the mean ROC curve whereas dashed curves represent the 10-fold ROC curves.

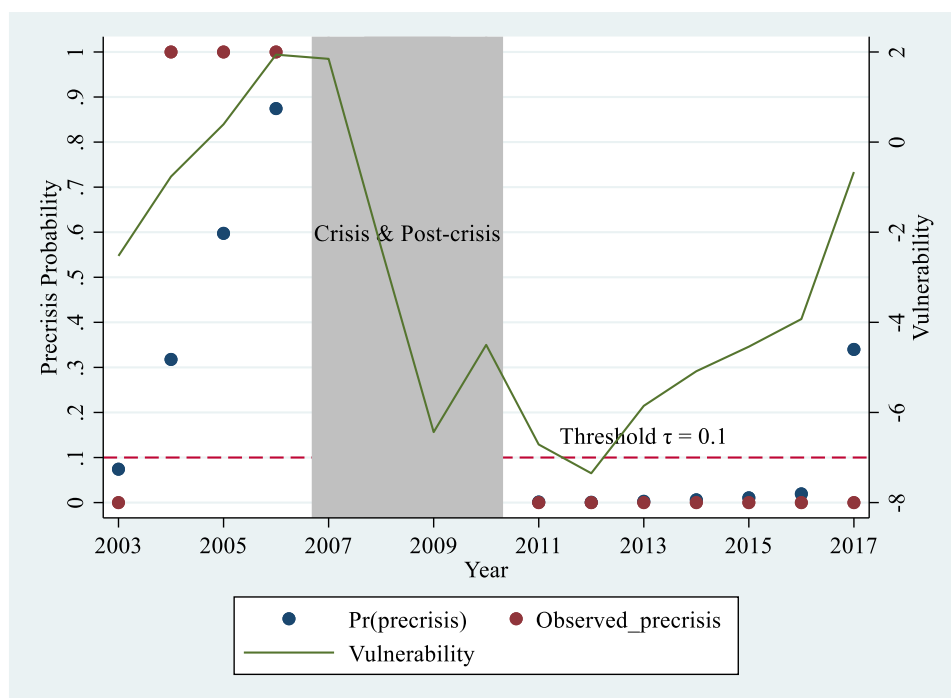
Figure B.2: ROC curves for the quasi real-time exercise



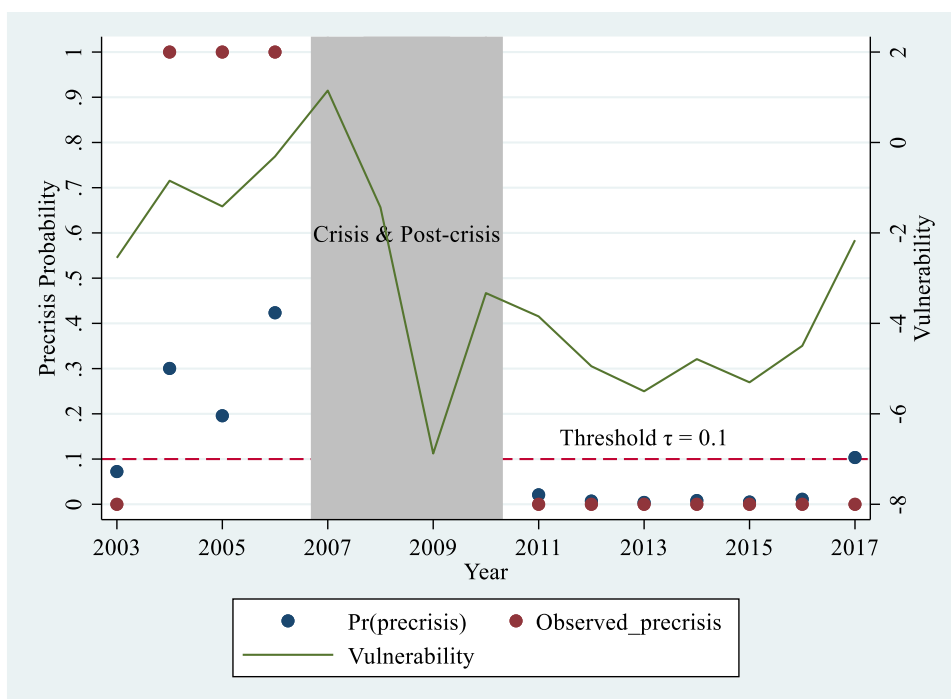
Source: Own elaborations. Notes: Model 1 and Model 2 correspond to EWSs with and without political indicators, respectively. Reference line is the line of no-discrimination.

Figure B.3: Plots of pre-crisis predictions of the benchmark EWS for the USA, the UK, Germany, and France in the 2003-2017 period.

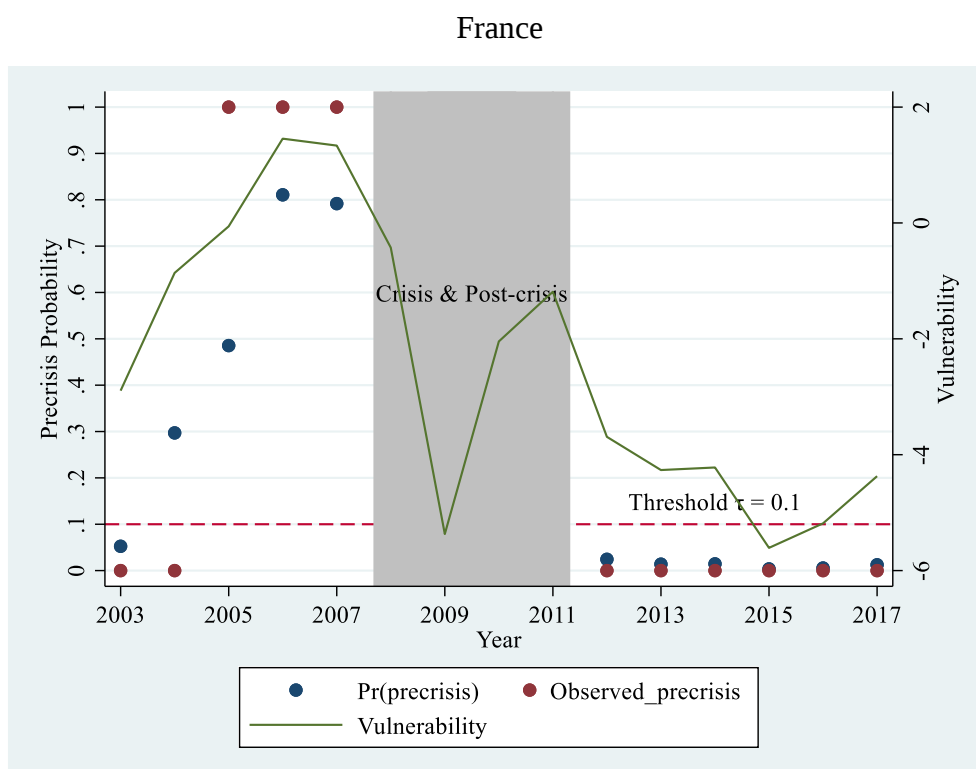
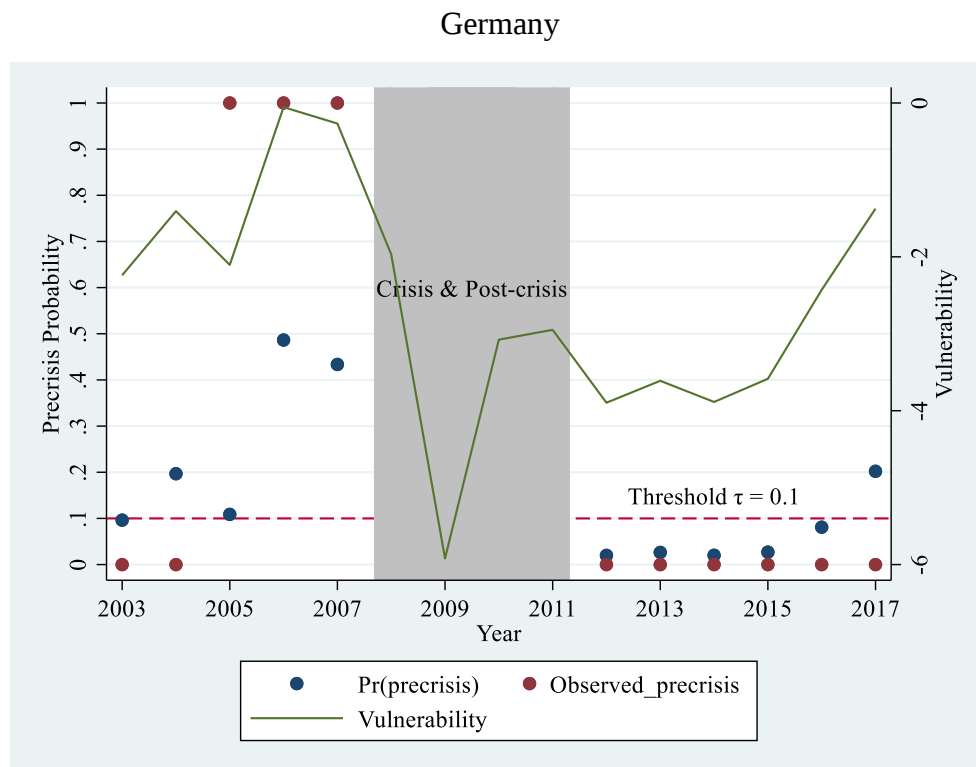
The USA



The UK



Source: Own elaborations. Notes: $Pr(\text{pre-crisis})$ stands for predicted pre-crisis probability.



Source: Own elaborations. Notes: Pr(precrisis) stands for predicted pre-crisis probability.

IMPRESSUM

Jena Economic Research Papers

ISSN 1864-7057

Friedrich Schiller University Jena

Faculty of Economics and Business Administration

Carl-Zeiss-Str. 3

D-07743 Jena, Germany

Email: office.jerp@uni-jena.de

Editor: Silke Übelmesser

Website: www.wiwi.uni-jena.de/en/jerp

© by the author