Economic Modelling 125 (2023) 106304

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

How to foresee crises? A new synthetic index of vulnerabilities for emerging economies[☆]

Irma Alonso-Alvarez [*], Luis Molina

Bank of Spain, Spain

ARTICLE INFO

JEL classification:
E44
F01
F34
F37
G01

Keywords:
Emerging economies
Crisis
Vulnerabilities
Early warning models
Risks
Index

ABSTRACT

We present a novel vulnerability index to monitor crises in emerging economies. To design the index, we identify the empirical regularities that precede sovereign, currency, and banking crises. Because we want to give policy makers the ability to react at an early stage, we focus on six-quarters before the onset of a crisis. We use data for 25 emerging economies and a new quarterly dataset of crisis events. The short-term interest rate is the unique variable that predicts all three types of crisis since it captures bank difficulties and sovereign and currency strains as monetary authorities use it to defend exchange rate pegs and to avoid capital flows. As the predictors of the three types of crisis generally differ, we define a different index for each type of crisis. The index, which is easy to update, outperforms the usual individual leading predictors of crises.

1. Introduction

Emerging Market Economies (EMEs) have become increasingly relevant in the world economy, being more and more interconnected with advanced economies. Given that these economies have traditionally been more prone to suffer crises, a key question for developed countries' policymakers should be to what extent their country, and in particular their banking sector, is exposed to turbulences in those EMEs. Could these stress episodes be anticipated sufficiently in advance to tame its local and global effects? And which EMEs' variables should be prominently monitored to detect the build-up of vulnerabilities in these economies?

In this paper we try to answer both questions based on the large and growing literature on Early Warning Systems (EWS). The contribution of this paper is twofold. First, we present a synthetic index of vulnerability for each emerging market, which is labelled SHERLOC (Signalling Heightened Emerging Risks that Lead to the Occurrence of Crises).[1] To design it, we identify key empirical regularities in the run-up to the sovereign, currency and banking crises, that is six quarters before the onset of the event. We find that the predictors of these three types of crisis tend to be different given their heterogeneous nature in terms of both origin and development. As a result, we present a different index for each type of crisis, and we prove that these synthetic indexes outperform an aggregate index for all types of crisis. A significant crisis predictor for the three different types of crisis is the short-term interest rate. This can be explained by the fact that it captures bank difficulties and sovereign and currency strains as monetary authorities have used it to defend exchange rate pegs and to avoid capital flows volatility. In the same vein, international reserves, fiscal deficits, increases in public debt, the leverage ratio of banks, or rating downgrades also play a key role in anticipating turbulences, and should be monitored in depth. Finally, the main advantages of SHERLOC are the following: (1) it anticipates accurately crises, (2) it outperforms the usual individual leading indicators of crises both in-sample and out-of-sample, that is, its forecast accuracy is greater than that of the individual indicators commonly used, such as the sovereign spread or the short-term interest

☆ We thank participants at Bank of Spain internal seminars, IRC meetings and the 17th Emerging Markets Workshop. We are very grateful to Sonia López Senra for her extraordinary and essential work in the earlier stages of the paper. We also thank Rodolfo Campos, Daniel Santabárbara, Juan Carlos Berganza, Pedro del Rio, Sonsoles Gallego, Enrique Alberola, José Manuel Marqués, Javier Pérez, Luis Orgaz and José Ramón Martínez for their useful comments. Finally, we thank comments from two anonymous referees and the editor of Economic Modelling, which improved the paper. The views expressed in the paper are those of the authors and do not necessarily reflect those of the Bank of Spain or of the ESCB.

* Corresponding author.
E-mail addresses: irma.alonso@bde.es (I. Alonso-Alvarez), lmolina@bde.es (L. Molina).

[1] The acronym SHERLOC also reminds us the main use of the index, which could provide clues to detect the skeleton in the closet.

https://doi.org/10.1016/j.econmod.2023.106304
Received 13 January 2022; Received in revised form 3 April 2023; Accepted 3 April 2023
Available online 3 May 2023

rate, and (3) it can be quickly and easily updated since it is based on few variables publicly available.

Second, we develop a new quarterly dataset of sovereign, currency and banking crises, for 25 emerging economies from 1993, partially based on the seminal works of Laeven and Valencia (2012, 2018, 2020) and Reinhart and Rogoff (2009). We define the beginning and the end of crises on a quarterly basis in order to develop a more timely early warning system.

SHERLOC is built sequentially using in a first step a signalling approach (AUROC) to assess, in a univariate setting, the predictive ability of each indicator before the onset of a crisis. Then, in a second step, these preselected variables are used to estimate vulnerable states in a standard logit model. This process mitigates the overfitting issues that usually plague the EWS estimations.[2]

Along with their greater presence in global markets both in trade and financial terms,[3] EMEs have experienced a large number of currency crashes, banking crises and sovereign defaults with high costs in terms of economic activity and employment. There are well-known examples of such situations: Mexico in 1995, Asia in 1997, Russia in 1998, or Argentina in 2001.[4] In recent years, crises have been less frequent, partly due to the measures implemented by policymakers to minimize risks – floating exchange rate regimes with inflation targeting and de-dollarization of public debt –. Nevertheless, EMEs have continued to face periods of heightened financial volatility, related either to their greater integration in international financial flows or to domestic imbalances. These stress periods resulted in increases in risk premia, declines in stock market indexes and significant currency depreciations, which in some cases also led to economic fall outs, as exemplified by the turbulences in Argentina and Turkey in 2018. At the time of writing, EMEs could face another wave of financial instability due to the consequences of the Russian invasion of Ukraine and the spike in inflation that have sparked an increase in interest rates throughout the world and an appreciation of the US dollar.[5]

Although most of these recent episodes reversed faster than in previous periods, it is essential to monitor the build-up of vulnerabilities in EMEs. First, this enables policy makers and domestic investors to assess the economic situation of these countries and the risks they might face, and to take informed and rational decisions on their portfolio composition avoiding herd behaviour and contagion. Secondly, in an increasingly globalized world, financial spillovers and spillbacks from EMEs to advanced economies are becoming increasingly relevant.[6] Finally, a crisis in one of these countries can affect the financial stability of other economies through the positions of financial institutions and firms in EMEs. That is why, for instance, the European Systemic Risk Board (ESRB) established the possibility of imposing countercyclical capital buffers to European banks that are exposed to third countries with unaddressed risks.[7]

Fig. 1 provides a stylized description of the potential imbalances that SHERLOC tries to capture. Financial markets can signal markets' perceptions of how the fundamentals of an economy are evolving. But they can also overreact and trigger a crisis, and could lead to contagion to other sectors within the same country (for example, a deterioration of sovereign risks could lead to a withdrawal of capital from foreign investors and to a currency crash) or to other countries. The real sector provides us with information about imbalances stemming from a recession or overheating of the economy. The fiscal sector focuses on fiscal policy uncertainty or sovereign solvency risk, and all items related to excessive public leverage and debt sustainability. The banking sector, which is monitored with a large number of variables to reflect the fact that banking crises have become more relevant in the last part of our sample, might capture a high leverage of the sector, lack of profitability, balance sheet mismatches, liquidity shortages, solvency or systemic risks. The external sector offers information on unsustainable current account balances, sudden stops or capital flight or excessive leverage on foreign currency of domestic agents. Finally, the institutional and political area covers risks such as the lack of willingness to pay back debt, policy uncertainty, the expropriation risk, geopolitical risk, violence and social strain risks, as well as high operational costs.

All these linkages could lead to an increase in EMEs vulnerabilities and, at the end of the road, to a crisis. For example, a growing public deficit could potentially lead to a sovereign default. A deterioration of global financial conditions could also make the level of public debt unsustainable through increasing its cost of financing. Capital outflows or a depreciation of the currency could put in dire straits the situation of the public sector depending on the proportion of debt denominated in foreign currency. In the case of currency crises, there is vast theoretical literature that links unsustainable public accounts, the deterioration of economic fundamentals or the currency mismatches both in banks or in banks' clients balance sheets to currency crashes. Excessive banking sector leverage, misvalued off balance assets and liabilities, or insolvency problems, are usually linked to banking crises. Finally, vulnerabilities in each area are closely interconnected and interact and reinforce one another. For instance, a recession can lead to social distress, and to an increase in non-performing loans that hit the banking sector. A worsening of the institutional framework can deter foreign investment and lead to sudden stops or capital flights, which could hinder the funding of public debt. Public sector solvency risks could affect banking sector profitability, and the solvency if banks own a high proportion of domestic public debt. Conversely, public sector implicit guarantees or even the nationalization of banks in the event of a banking crisis could also trigger a sovereign default. A strong depreciation of the currency could make currency mismatches in banks' balance sheets unsustainable, and the need for foreign liquidity in the case of a banking crisis could potentially lead to a currency crash.[8]

The aim of SHERLOC is then not to estimate the probability of the occurrence of a crisis but to identify underlying vulnerabilities and imminent tail risks that predispose a country to a crisis, using variables that proxy the risks on the left hand side of Fig. 1.

The rest of the paper is structured as follows. The next section provides a brief literature review and discusses the main pitfalls of EWS models. Section 3 presents the main features of the dataset used in the econometric analysis. Section 4 describes the empirical strategy used in the paper, the results of the non-parametric and parametric approaches to detect vulnerabilities, and the construction and validation of SHERLOC. In addition, it assesses the development of SHERLOC in the period of the COVID-19 pandemic and just before the invasion of Ukraine. Section 5 concludes and highlights future work.[9]

[2] Other methodological approaches are also proposed to build the index, such as a factor augmented logistic estimation, a simple average of the risk percentiles or an aggregation of preselected variables based on principal component analysis (PCA).

[3] For example, according to published earnings and profitability reports, some European banks obtained around half of their profits before taxes (EBITDA) in EMEs.

[4] For more information on these crises, see Appendix A.1.4.

[5] On the consequences of an unexpected increase in US official interests see for example (Banco de España, 2022a).

[6] See IMF (2016).

[7] See Decision of the European Systemic Risk Board of 11 December 2015 on the assessment of materiality of third countries for the European Union's banking system in relation to the recognition and setting of countercyclical buffer rates (ESRB/2015/3) (2016/C 97/11).

[8] For a detailed description of the theoretical models that supports the stylized briefing of the channels and triggers of crisis represented in Fig. 1, see Appendix A.1.4.

[9] This approach can be complemented by the use of vulnerability dashboards (heat maps), which have been useful to detect vulnerabilities in the past, as described in Alonso and Molina (2021).

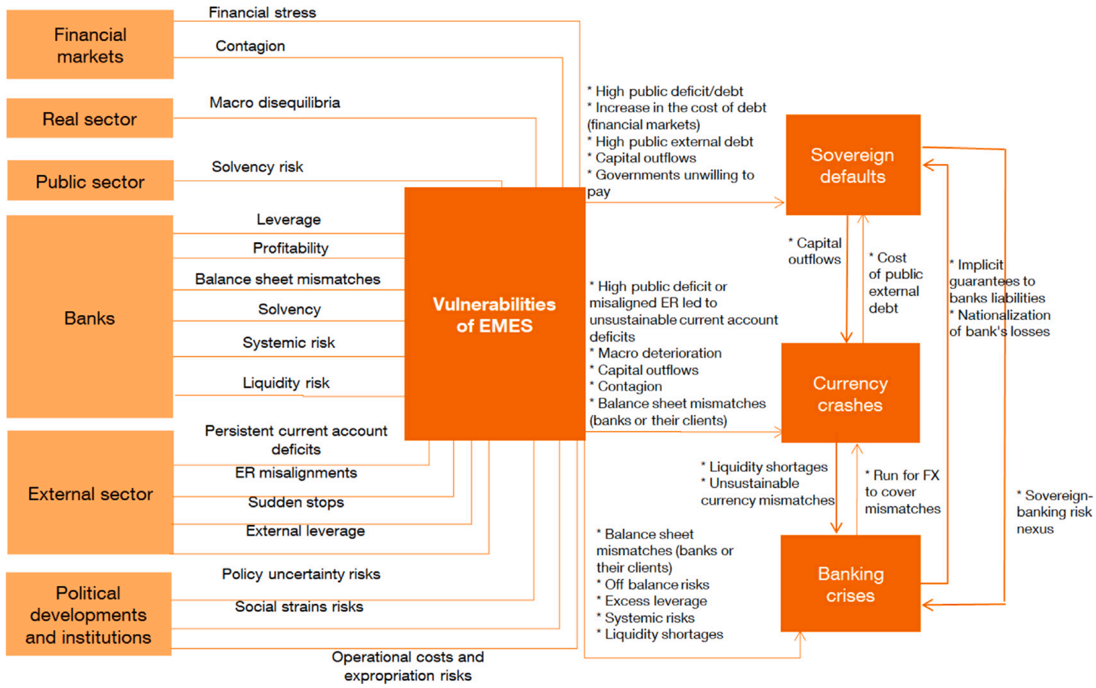

Fig. 1. The mechanics of vulnerabilities and crisis in EMEs.

2. Literature review and dealing with EWS' pitfalls

There is a widespread and growing literature on EWS, developed in the aftermath of the seminal paper of Kaminsky et al. (1998). In its early stage, most of the literature on EWS focused on analysing EMEs' risks, such as Kamin et al. (2007), Bussiere and Fratzscher (2006); and Lestano et al. (2004) for Asian countries. The increasing number of models and papers even prompted a horse race between them, as implemented in Berg et al. (2005). In the wake of the global financial crisis of 2008, many papers focused on developed countries (Rose and Spiegel, 2011; Frankel and Saravelos, 2012[10] or Catão and Milesi-Ferretti, 2014[11]), especially in the euro area (González-Minguez and Carrascal, 2019) or the G20 group (Qin and Luo, 2014), and also on assessing vulnerabilities that can affect countries' financial stability.[12] Recent papers on EWS focused on improving the methodology for the estimation of the probability of a crisis by introducing dynamics in the models (Candelon et al., 2014; Dabrovski et al., 2016) or by using new techniques suh as machine learning (Samitas et al., 2020; Beutel et al., 2018; Wang et al., 2021) or entropy (Billio et al., 2016), or by introducing new measures of uncertainty to address the political and social risks in emerging economies.

We rely on this large strand of the literature to select the potential leading indicators for our analysis, and consider all types of crises that occurred in emerging markets over the period 1993–2018.

Nevertheless, a large number of theoretical and empirical papers have also challenged the EWS methodology.[13] First, on the empirical front, one of the main caveats of EWS models is that there are too few crisis observations to obtain consistent estimators of probabilities. This problem is even more acute when dealing with out-of-sample calibration of EWS, as the number of crisis observations is reduced. In this paper we use six quarters previous to a crisis as the strain times, which partially mitigates the issue of lack of enough stress periods. As stated in the introduction, the aim of the SHERLOC is to identify underlying vulnerabilities that could lead to a crisis, and not to predict the crisis itself. Moreover, the use of random effects in the logistic estimation allows us to exploit the information of all the countries in the sample, even those with no crisis observations.[14]

Second, EWS critics remark the problem of "real time" EWS. This means that the authorities and economic modellers do not have the entire set of information used by a posteriori estimations.[15] The need to build real time indicators is a time consuming task in macro samples with heterogeneous countries and the use of a common lag of publication for all countries could bias the results.[16] While in the baseline specification we decide to keep the model as simple as possible discarding "the real-time issue", we carry out a validation exercise using a pseudo-real time approach (appendix A.7). Our results are mostly robust to this specification.

Third, EWS may present a post-crisis bias of discrete dependent variables as macro variables tend to be very persistent and therefore some indicators show an "erratic behaviour" in the recovery phase. Bussiere and Fratzscher (2006) predict financial crises in EMEs by relying on a multinominal logit model, which allows to distinguish between three

[10] Both papers propose different methodologies to determine the drivers of the global financial crisis using a cross-country approach.

[11] The paper highlight the role of foreign liabilities to explain external crises in advanced economies.

[12] Oet et al. (2013), Gramlich et al. (2010) and Castro et al. (2016) develop EWS for systemic risk to implement macroprudential tools.

[13] Some of the arguments discussed could be found in Boonman et al. (2019) and Bussière (2013).

[14] Another way to avoid this problem would be to build a kind of unique country in which crises and variables will be placed consecutively one after another regardless of when they occur, as explained in Gadea Rivas and Perez-Quiros (2015). However, this is appropriate when the degree of homogeneity when a shock occurs is higher than the homogeneity within countries, which is probably not the case in our sample.

[15] At time t they only have information on GDP in t−1, on Nominal Effective Exchange Rate on t, on Short term External Debt indicators at time t−2, and so on, but EWS are estimated at each point of time as if the entire set of information had been updated in t.

[16] Another additional problem is the "true real time" EWS, that is, to rigorously calibrate an EWS we should use the data that were actually published at the time of the calibration, which are usually revised later, especially those referred to activity and public sector balances. In our case, we do not have the vintages to carry out true real-time EWS.

states: a normal, pre-crisis, and post-crisis state. In our case, we show that considering the "post-crisis" bias, by eliminating the data of four quarters after a crisis, is relevant to slightly reduce type II errors (false alarms) out-of-sample, but it does not seem to affect the ratio of crisis predicted in-sample.

Fourth, these EWS models might also suffer from a problem of prediction failures outside the sample that could be partially attributed to an in-sample overfitting and variable selection bias. As many combinations of different variables could lead to the same prediction, the final selection of variables can be arbitrary, which leads to data mining or cherry picking issues. A way to circumvent those pitfalls would be the use of automated variable selection methods like the Lasso[17] or out-of-sample validation methodologies like the Random Forest[18] which basically rely on the statistical improvement of the results to decide whether a concrete indicator is included or not in the model. In this paper we deal with the overfitting issue in a simpler way than Lasso or Random Forest models. First, we use a signalling approach (AUROC) to pre-select the variables that issue the best signals ahead of a crisis,[19] and then use these variables to build the vulnerability indexes SHERLOC via a panel logit model.[20] As robustness checks we also estimate a factor augmented logit model – which exploits all the available information by using as explanatory variables the common factors extracted from the whole dataset – and also combine the preselected indicators using principal component analysis techniques.[21] These two methodologies have been widely used in the literature (Edison, 2003; Frankel and Rose, 1996; Duca and Peltonen, 2013), but as far as we know these papers do not provide a clear framework to preselect variables as we do.

As stated in the introductory section, the main contribution to the literature of this paper is twofold: (1) we develop a quarterly dataset of events, distinguishing by type of crisis, improving the usual dating of crises and, more importantly, (2), we propose a synthetic index of vulnerability for EMEs, the SHERLOC, that summarizes the state of vulnerability of each EME against three different types of crises (a sovereign default, a currency crash and a banking crisis).

Few papers have proposed a synthetic index of vulnerabilities for emerging economies. One of the most recent is the index of vulnerability for emerging economies proposed by Lepers and Serrano (2020). In contrast to our paper, they only focus on financial crises and do not evaluate the performance of their composite index. In this sense, our focus will be on all crises occurred in emerging markets over the period 1993–2018 in order to exploit the heterogeneity of different types of crises. Additionally, we incorporate a widespread evaluation method, developed by Alessi and Detken (2011), to assess the predictive ability of SHERLOC both in sample and out-of sample. Alessi and Detken (2011) propose a new measure of Usefulness, which compares the loss of the model with regards to the loss of disregarding the model, taking into account policy maker's preference. While the in-sample performance is quite adequate, in line with the results of Alessi and Detken (2011), the out-of-sample performance is poorer, as in other out of sample validations. Moreover, some of the criticisms to EWS still hold. First, noise (excessive issuance of signals) remains high. Second, we are not able to capture non-linearities as pointed out by Eichengreen (2002). Third, even if the use of different types of crises for each country partially mitigates the criticism of "this time is different" and the out-of-sample performance suggests that "not every time is different", a new type of crisis might not be detected by SHERLOC.

3. Data

3.1. Building an early warning system: Stress events

The first step of EWS models consists of identifying the relevant crises in the countries analysed. The definition of a crisis is crucial for an EWS and requires a thorough analysis to avoid misclassification issues and uncertainty about the results. We use stress events in order to proxy "vulnerable states" in the econometric analysis, and defined them as the period six quarters prior to the crisis.

There are two main ways to define stress events. First, one can rely on a binary indicator to define periods of crisis. This approach has a crucial advantage: it is possible to define "vulnerable states" as the pre-crisis periods, as we do in this paper. It is also useful to eliminate the post crisis bias and to distinguish between different types of crisis. But this approach faces some limitations because an exogenous definition of crisis requires expert judgment and can be subject to misclassification. Moreover, a sufficient number of stress periods are needed to obtain robust results. That is why some papers rely on the use of continuous variables, such as Financial Stability Indexes (FSI) or Exchange Market Pressure indexes (EMP) (Eichengreen et al., 1995, 1996 and Duca and Peltonen, 2013 and let the model endogenously choose the periods of crisis through the use of Markov Switching Models Martinez Peria, 2002). However, in our case it is difficult to follow this strategy due to the lack of continuous variables covering a long enough period for our sample of EMEs countries. For instance, we could build up FSI measures but as many of the required series start after 2000 for many EMEs, we would miss some well-known stress episodes of the 90 s. Moreover, it is debatable which indicator to use in order to capture all the stress events that we are interested in.

One of the contributions of this paper is precisely the construction of a quarterly dataset of stress events, distinguishing between sovereign, banking and currency crisis. To do so, we mainly identify sovereign and banking crises following (Laeven and Valencia, 2012), but defining the beginning and the end on a quarterly basis instead of an annual frequency, and updating the dataset until the end of 2018.[22] The beginning of a sovereign crisis is dated in the quarter in which the sovereign defaults or restructures its debt according to Standard and Poor's rating agency, while the end of a sovereign crisis is associated with the quarter in which an agreement with debt holders is reached,

[17] The Lasso (Least Absolute Shrinkage and Selection Operator) technique was introduced in order to improve the prediction accuracy and interpretability of least square regression models by altering the model fitting process to select only a subset of the provided covariates. It forces the sum of the absolute value of the regression coefficients to be less than a fixed value, which leads to certain coefficients to be set to zero, effectively leading to a simpler model. After repeating the estimation, variables that are retained most often are selected. See for example Li and Chen (2014).

[18] The random forest methodology operates by constructing a multitude of "decision" trees and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. In the case of EWS, it could consist of selecting the regressors based on their relative importance, that is, on the relative increase of the accuracy of the prediction once the concrete variable is included. See Breiman (2001).

[19] The signalling approach was initially developed by Kaminsky et al. (1998) to identify macroeconomic variables that can anticipate currency crises based on a "critical threshold" calculated by minimizing the noise-to-signal ratio for each indicator. But, it does not consider the interrelationships between variables.

[20] One of the earliest contributions was from Frankel and Rose (1996) who study the determinants of currency crashes in 100 emerging economies from 1971 to 1992. Duca and Peltonen (2013) also develop a framework for predicting systemic events, which incorporate both domestic and global indicators that improve the forecast performance of the model.

[21] Many variables are introduced in the regression in levels, although some of them (typically stocks variables like credit to GDP or those related to debt) could have a trend that could bias their accuracy as crises' predictors. In some cases we have introduced first time differences to avoid this problem.

[22] The update was made before the publication of Laeven and Valencia (2018, 2020). That is why there are slight differences. For instance, we consider a banking crisis in Russia in 2015 that is not included in Laeven and Valencia (2018). Nevertheless, our events database is easily updatable.

or alternatively, the date of the debt exchange, which implies a partial access of the sovereign to international markets -and a change of the sovereign rating. For banking crises, we date the quarter of the start of stress using national authorities' information or IMF's reports. Banking crises are dated when there are significant signs of financial stress in the banking system and/or banking interventions and banking takeovers by the authorities. The end date of the crisis is assigned to the quarter in which the eighth consecutive quarter of growth of both real GDP and real credit is reached, which is the criteria used by Laeven and Valencia (2012, 2018).[23] Finally, for currency crises we rely on a definition similar to that of Reinhart and Rogoff (2009) though with a more restrictive threshold. A crisis is assigned when the nominal exchange rate against the US dollar depreciates more than 30% quarter on quarter.[24] As a robustness check, we also define a currency crisis as a depreciation of at least 15% on a quarterly basis as long as this depreciation exceeds the average variation of the exchange rate plus a standard deviation (see table 8 for a detailed description of the dates of crises).

Once the crisis events are identified on a quarterly basis, we construct three dummy variables, one for each type of crisis, that are used in the econometric approach. Fig. 2 plots the number of crises identified per quarter.[25] As expected, the number of crisis diminished dramatically since 2004. Sovereign and currency crises are less frequent than in the nineties because many countries in the sample that used to rely on a fixed exchange rate to the US Dollar as the main tool to stabilize domestic inflation have now turned to inflation targeting with floating exchange rates. Finally the percentage of banking crises in terms of total crises has increased after 2008. Fig. 3 presents the exact number of crises in our sample, distinguishing between currency, sovereign and banking crises.[26] As shown in Fig. 3, in our dataset, there is a single twin sovereign-banking crisis (Argentina 2001 q4), one twin currency-banking crisis (Indonesia 1997 q4) and one triple crisis (Russia 1998 q3).[27]

As the objective of EWS is to anticipate the appearance of crises with enough time to enable authorities to react, indicators should send a correct signal in advance of the stress periods identified. Considering the type of crisis we are dealing with, we use an evaluation window up to 6 quarters prior to the event (consistent with the results of Kaminsky and Reinhart (1999) who stated that warnings of a crisis usually appear 10 to 18 months before the onset). This means that the dummy variable identifies the one-and-a-half year period before the crisis as the period in which indicators should send correct signals and the AUROC approach assesses the relevance and performance of each variable over this period.[28]

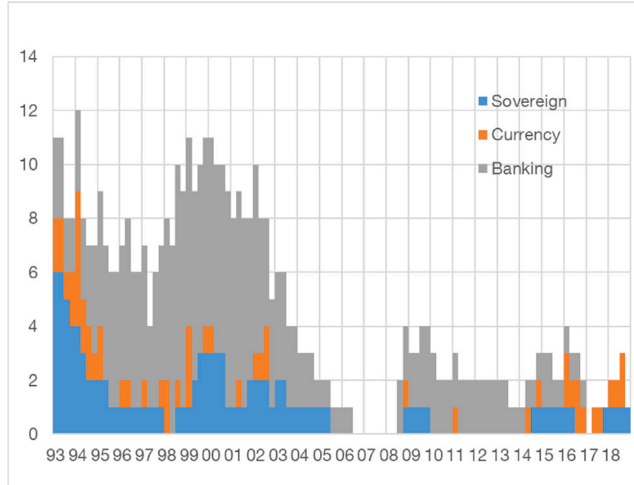

Fig. 2. Number of countries in crisis.

Fig. 3. Number of crises by type.

Crisis quarters are not taken into consideration as our objective is to identify leading indicators, and indicators tend to have an erratic behaviour during periods of stress. The rest of the quarters are considered normal or tranquil periods and therefore are equal to zero.

3.2. Macroeconomic determinants of crisis

Our dataset includes 25 countries, representing around 78% of the GDP of EMEs, and around 45% of world GDP. It comprises 9 Latin American countries (Argentina, Brazil, Mexico, Colombia, Chile, Peru, Venezuela, Ecuador and Uruguay), 5 Asian nations (China, South Korea, India, Indonesia and Thailand), 6 Eastern Europe countries (Czech Republic, Hungary, Poland, Romania, Russia and Turkey) and 5 countries from Africa and the Middle East (South Africa, Nigeria, Saudi Arabia, Egypt and Morocco). The selection of countries is determined by the availability of data and also by their economic and financial relevance.

Table 1 reports the 35 vulnerability indicators used in this paper. Their selection is mainly based on the most significant variables suggested by the literature on EWS. Notice, however, that we include a large number of variables related to the banking sector, in order to reflect the fact that banking crises have become more relevant for advanced economies in the period we consider. In addition, most indicators are easy to update and allow a frequent monitoring of the risks faced by the countries in the sample. Moreover, we use both

---

[23] However, in our dataset, we decide not to truncate the end of banking crises. Laeven and Valencia truncate banking crises 5 years after the blown up.

[24] As events are defined in a different way (banking and sovereign are proxied by events while currency by a quantitative threshold), there is a wide heterogeneity in terms of frequency and duration of crises. Indeed, currency crises tend to be more frequent but with a shorter duration. We have discarded the definition of a currency crisis as a change in the exchange rate regime because since the mid 00 s almost all countries in our sample have adopted freely floating regimes.

[25] Note that these numbers do not correspond exactly with the number of countries in crisis as one country could suffer from two or even three types of crisis at the same time.

[26] In Section 5 we update the number of crisis to the first quarter of 2022.

[27] This avoid us to conduct a concrete analysis of the determinants of twin crises. Therefore, the possible bias arising from the presence of twin crises does not seem to be too problematic. It is worth noting that the currency crises in Argentina in 2001 takes place a quarter after the sovereign and banking crises.

[28] In practical terms this implies that the dummies are defined with a 1 six quarters before the onset of the crisis, with missing values during the crisis, and zero in tranquil no pre-crisis times. Four and eight quarters are also considered as robustness checks (see Appendix A.5).

Table 1
Variables included in the AUROC exercise.

Financial markets	
Sovereign spread (bps level)	
Sovereign spread (change over 3 months)	
Stock Exchange index (change over 3 months)	
Exchange rate vis a vis the USD (change over 3 months)	
Macroeconomic fundamentals	
Real sector:	Banking sector:
GDP (change y-o-y)	Real credit to private sector (yoy)
Inflation rate	Real deposits on domestic banks (yoy)
Industrial production (12 month MA, y-o-y)	Loan to Deposit ratio
NEER overappreciation	Non performing loans (% total loans)
	Net foreign assets of domestic banks
Fiscal sector:	Bank Stock Exchange (3 month change)
Public sector balance (% GDP)	Spread of bank's external debt (3 month)
Public sector gross debt (% GDP)	Short term interbank rate (%)
	Intermediation margin (loan-deposit rate)
External sector:	
Current account balance (% GDP)	
Gross external debt (% GDP)	
FDI (% GDP)	
Short term external debt (% Reserves)	
Reserves (% GDP)	
External debt service (% exports)	
Portfolio gross inflows (% GDP)	
Wealth and institutional quality	Contagion risks
	VIX
Per capita GDP (USD PPP and % change)	10 year US Treasury bond yield
GPR index	Short term interest rate
Sovereign rating (average of the 3 agencies)	Trade links
	EMBI sovereign spread
	Oil prices

national and international sources in order to be able to extend the dataset back to the first quarter of 1993. This enables us to include some of the most relevant idiosyncratic crises (Mexico, Asia or Russia in the 90s).

The 35 variables are divided into four groups, mainly reflecting the frequency of update, their presumed capacity to react to an increase in risks, and the linkages described in Fig. 1. Financial markets variables – updated daily – are supposed to react more quickly to situations of high vulnerability, and even an overshooting of them could trigger a crisis. Fundamental variables – monthly or quarterly updated – tend to reflect the increase in risks more parsimoniously and constitute the core of the econometric analysis.[29] Finally, crises coming from a deterioration of the institutional quality – whose indicators are updated annually or show very little variability, which limits their usefulness for the empirical analysis – tend to have a longer ripening process, and their transmission channels differ from those of macroeconomic variables. For instance, a worsening of the institutional framework can deter foreign and internal investment and it can lead to sudden stops or capital flights even if fundamental variables are in good shape. The willingness of governments to pay back their external debt is also a relevant factor, which should be monitored. Finally, some variables have been included in the estimation strategy either to control for global factors that could bias the results or to take into account EME's interconnectedness and EME's dependence on global commodity prices.

For the econometric analysis all the indicators are transformed into a quarterly frequency, using either the last data of the corresponding quarter or the quarterly average (from daily or monthly to quarterly frequency) or a linear interpolation (from annual to quarterly frequency). Moreover, the lack of time series long enough leads us to discard many relevant qualitative variables.[30]

4. Empirical analysis

4.1. Signalling approach: AUROC results

The AUROC (Area Under the Receiver Operating Characteristics) approach is a univariate method that measure the performance of each variable to distinguish between two distributions (in this case, related to normal and stress periods). More specifically, it extracts signals from each indicator when they are above or below a certain threshold calculated with their historical data.[31] Based on these signals, the procedure estimates the performance of each variable depending on the number of correct signals (right guess of a crisis happening or not happening) and false signals (missed crises or predicted crises not happening) issued in the predetermined evaluation window. The AUROC evaluates each indicator's performance taking into account the true positive rates in terms of the false positive rates. A value of 1 implies that the specific indicator is only issuing correct signals, whereas a value of 0.5 provides no information at all since it would anticipate crises randomly (like flipping a coin).

More specifically, the AUROC methodology indicates how well the model is able to distinguish between the probabilities of being in a vulnerable state or in a normal period. Indeed, as the threshold imposed to differentiate between stress and normal time increases, the ratio of true positive rate and false positive rate declines. An AUROC equals to 0.5 implies that there is an overlap between the distributions of crisis and normal times (Fig. 4).

[29] Most of the data used is comparable since they are extracted from international or official national statistical sources. However, there remain differences, for instance with the definition of reserves which in some cases include reserves that cannot be mobilized by the monetary authority.

[30] Some of them are the BICRA indicators (Standard and Poor's), banking and political risks scores (EIU), or Doing Business and Absence of Violence (World Bank). All of them are introduced in the vulnerability dashboards described in Alonso and Molina (2021).

[31] See, for instance, Castro et al. (2016) for an application to the case of the financial sector in Spain.

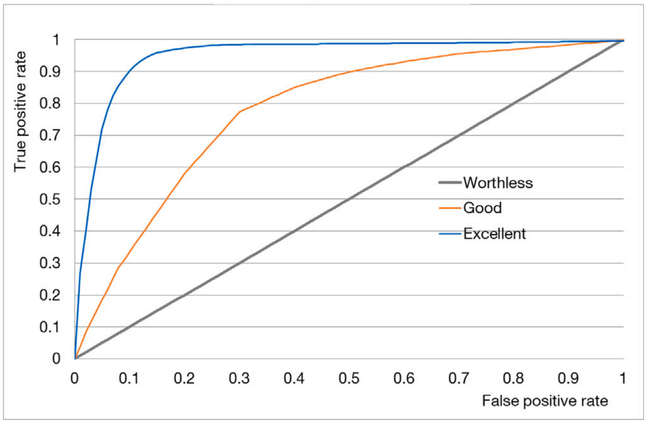

Fig. 4. Comparing ROC curves.

In technical words the AUROC estimates the area under the curves ROC (Receiver Operating Characteristics), which show the relation between the true positive rate and false positive rate for different thresholds:

$$AUROC = \int_0^1 ROC(FP)dFP \tag{1}$$

to evaluate the predictive ability of each variable as an early indicator of a crisis. FP stands for false positive rate. Implicitly the AUROC model used assumes that policymakers preferences are equally concerned about type I (missing crises) and type II (false positives) errors.[32]

The results for the signalling approach are presented in Table 2.[33] The variables that seem to send strong signals are marked in red and in orange, indicating a percentage of good signals above 70% and 65%, respectively. These are ad-hoc thresholds, which enable us to reduce the number of leading indicators to mitigate over-fitting issues.[34] In appendix A.2, we provide a table with all the results of the AUROC exercise.[35]

The main result is that, for any kind of crisis, only a few variables reach an adequate performance in anticipating them, in terms of the ratio of true positive rates and false positive rates (just 6 out of 48). Moreover, only half of the variables belong to the so-called "fundamental" indicators (the rate of inflation, the short-term interest rate, and the ratio of short term external debt to reserves), together with another variable reflecting market sentiment (the sovereign spread) and contagion from global turbulence (the US 10 year interest rate). As for the "qualitative" variables, just the sovereign rating seems to issue correct signals in an adequate proportion. These apparently disappointing results could be due to the double heterogeneity of our sample, on account of the diverse geographical composition and the different nature of crises. In addition, there seems to be a structural break around 2006 as the number of crises sunk to zero and then a new type of crisis emerged in 2007 (mainly contagion from advanced economies turbulence and banking crises), which may have affected the type of variables that issue anticipatory signals. For all these reasons, we carried out the exercise distinguishing between banking, currency and sovereign crises. Moreover, we estimate another AUROC splitting the sample into EME regions, and before and after the third quarter of 2006.

Results for banking crises are coherent with economic theory. The net foreign position of domestic banks (a proxy for an "excessive" reliance on external funding and for balance sheet currency mismatches) and the loan to deposit ratio (which again proxies for a higher reliance on market funding instead of traditional funding) have a ratio of true positive signals close to 70%, as well as the Nominal Effective Exchange Rate (NEER) deviation from trend, which could be a proxy for strains in banks' borrowers. Moreover, two variables related to the cost of external funding and external financial conditions – short and long term interest rates in the US –, tend to also anticipate EMEs banking crises, as well as short term domestic rate increases. Finally, non-performing loans have a noise to signal ratio of around 62%.[36]

In the case of sovereign crises there are a large number of variables issuing adequate signals (17 out of 48 show a percentage of correct signals above 65%, 12 of which have a percentage above 70%). Sovereign ratings and the sovereign spread deteriorate in anticipation to a public debt default. Moreover, activity growth variables deteriorate before a sovereign crisis, which can be capturing a tax base shrinkage. On the fiscal side, the relevant variables seem to be the level and increase of public debt rather than the public sector balance. As a large part of public debt of those countries was placed abroad, variables related to the level of external debt, such as maturity and debt service, are also leading indicators of sovereign defaults. In this sense, the short term domestic rate also exhibits a ratio of correct signals above 70%, probably as it is influenced by monetary authorities to limit capital outflows that could lead to a depreciation of the currency. As reserves constitute the last line of defence in case of strains derived from external debt, their level is also relevant to anticipate sovereign crises. Finally, the variable that measures the interconnectedness of EMEs – trade links – is also significant.

Currency crises seem to be anticipated by an increase in sovereign spreads and short term interbank rates, as well as by short term external debt and an acceleration of public debt growth. A decrease in reserves and a fall in activity growth are also relevant. Surprisingly the overvaluation of the NEER does not rank as one of the best leading indicators (58%), although the inflation rate could proxy for the loss of competitiveness in fixed exchange rate regimes like those of the 90 s. AUROC results also point to a high correlation between currency and sovereign crises, as they share some common determinants, though with a slightly higher AUROC ratio for public debt, sovereign rating and spreads and reserves in the latter case. As predicted by first generation currency crisis models, an increase in public deficit and public debt can lead to a sudden stop of capital inflows and a strong depreciation of the currency, which may ultimately lead to a sovereign default, as a large part of public debt is denominated in foreign currency and held by foreign investors.

[32] Nevertheless other AUROC specifications could give more weight to any type of error (pseudo AUROC).

[33] Each variable has been transformed so an increase implies greater risks.

[34] The threshold of 0.65 appears in several articles related to AUC estimates in medical screening. In economic studies, Castro et al. (2016) show the significance of the estimated AUC for each of the determinants at the theoretical level of the countercyclical capital buffer in Spain. A 99 percent level of significance corresponds in this case to a ratio of 0.65. On the contrary, Camacho and Palmieri (2021) used ROC techniques to evaluate the predictive content of the monthly OECD's main economic indicators for predicting both growth-cycle and business-cycle recessions at different horizons, sticking with those whose AUROC is greater than 0.5 in each of the 12 months prior to the onset of recessions. Therefore, the percentage of acceptance depends on the ultimate goal of the analysis.

[35] Another possibility would have been to choose all the variables with the lower bound of the confidence band of the AUROC above 51%. This would double the number of variables chosen for each crisis, possibly leading to overfitting and multicollinearity problems. However, including all these variables in the logit regression does not change our main results. Finally, in Appendix A.2, we show that the SHERLOC we propose outperform other indexes that include some of the variables left aside.

[36] For our two tail risk variables we have presented the result of an increase in real credit growth and huge portfolio inflows. Nevertheless, the results for the other tail risk (that could detect the appearance of domestic credit bubbles financed in part with foreign short term capital) are also non-significant (38% and 41% of good signals, respectively).

Table 2
AUROC results.

Variable	By type of crisis				By region				By time	
	All	Bank	Sov	Curr	LA	EE	Asia	Rest	Bef07	After
Sov.spread	0.73	0.59	0.90	0.80	0.77	0.64	0.18	0.83	0.68	0.77
GDP growth	0.60	0.53	0.83	0.67	0.67	0.58	0.35	0.57	0.60	0.63
Inflation rate	0.74	0.64	0.68	0.75	0.70	0.76	0.68	0.79	0.64	0.86
NEER dev from mean	0.59	0.68	0.59	0.58	0.58	0.49	0.72	0.62	0.63	0.55
Public sector balance	0.56	0.39	0.57	0.66	0.70	0.56	0.26	0.57	0.53	0.64
Public debt (first diff)	0.62	0.57	0.74	0.69	0.62	0.59	0.41	0.77	0.64	0.63
Net foreign assets dom.banks	0.59	0.68	0.56	0.56	0.53	0.51	0.65	0.57	0.66	0.51
Loan to Deposit ratio	0.54	0.71	0.62	0.46	0.45	0.48	0.93	0.40	0.71	0.31
Short term interbank rate	0.74	0.78	0.70	0.75	0.69	0.73	0.91	0.73	0.75	0.69
External debt	0.47	0.52	0.73	0.47	0.50	0.30	0.58	0.18	0.50	0.43
Short term ext.debt	0.67	0.66	0.87	0.71	0.72	0.56	0.82	0.25	0.70	0.62
Reserves over GDP	0.64	0.54	0.73	0.70	0.68	0.61	0.76	0.64	0.56	0.70
Ext.debt service	0.57	0.58	0.74	0.58	0.65	0.42	0.52	0.30	0.60	0.52
Portfolio inflows decline	0.48	0.38	0.67	0.51	0.47	0.45	0.43	0.62	0.39	0.57
GDP per capita (change)	0.60	0.50	0.84	0.68	0.68	0.55	0.41	0.58	0.59	0.63
Sov.rating	0.65	0.51	0.88	0.73	0.74	0.55	0.34	0.92	0.52	0.79
US 10Y treasury bond rate	0.65	0.78	0.56	0.58	0.61	0.76	0.92	0.30	0.73	0.56
US 3M interbank rate	0.62	0.75	0.60	0.56	0.57	0.73	0.85	0.41	0.64	0.53
Int. Trade interconnectedness	0.53	0.45	0.72	0.57	0.53	0.60	0.17	0.74	0.49	0.62

Turning to the results by region, Latin American crises seem to be well anticipated by a deterioration of public sector balance, an increase of short-term external debt and its service, a decrease of international reserves and a deterioration of the macro environment (higher inflation and lower activity growth). These factors fit well with first generation currency crisis models.[37] For Eastern European countries, banking sector variables (a fall in real credit and deposits growth and an increase of short term interest rate) tend to better anticipate the outbreak of crises, along with two global variables and total reserves.[38] Asian crises are correctly signalled by both banking sector variables (those related to currency mismatches as the net foreign position of domestic banks; or those referred to an excessive leverage in financial markets like the loan to deposit ratio) and external disequilibrium variables (current account balance, reserves).[39] For the five remaining emerging countries (Rest of EMEs), the picture is much less clear, global and financial market variables, an aggregate evaluation of the economies (sovereign rating) or some variables reflecting the macro environment (public sector accounts, reserves and inflation rate) issue good signals.[40] Finally, as mentioned above, from 2006q3 to 2008q2 none of the countries in our sample suffered any kind of crisis. Splitting the sample before and after 2007q1 implies that after 2007 the inflation rate, the sovereign rating and the level of international reserves gain predictive power. The short term interbank rate remains one of the best predictors for any kind of crisis.

Summing up, the heterogeneous nature of our sample leads us to separate the crisis indicator by type of crisis – instead of just using a dummy for all crises or to split the sample by region –, so that the results improve substantially.[41] According to the parametric exercise, the short term domestic rate tends to correctly anticipate vulnerable states, especially those related to the banking ones. Public sector disequilibria, both at the domestic and the external side, are relevant to determine the pre-crisis state that could lead to sovereign crises, which seem to be highly correlated with currency crises. Finally, variables measuring the deterioration of debt stocks, rather than flow variables, seem to be more useful to anticipate stress periods.

4.2. Building the SHERLOC

Once the variables have been preselected using a signalling approach (AUROC above 65%), we propose the best methodology, which we validate in Section 4.3, to aggregate the relevant variables and build the SHERLOC. This methodology consists in estimating the predicted probability of being in a pre-crisis (or vulnerable) state (six quarters previous to the crisis) using a logistic estimation. Then, we compare this methodology to three other ways of aggregating the relevant variables. First, the predicted probability is estimated from a factor-augmented logit approach. Second, we calculate an index using the two first principal components of the set of (preselected) standardized variables. Third, we just calculate the mean of the risk percentiles of the relevant variables.[42] In Section 4.3, we show that the logistic SHERLOC outperforms both in-sample and out-of sample the other methodologies to anticipate vulnerable countries.

4.2.1. A logistic estimation

The panel logit approach estimates the probability of being in a "vulnerable state" (i.e, six quarters prior to the crisis), which is assumed to follow a logistic distribution that depends on risk factors, that is:

$$Pr(Y_{it} = 1|X_{i,t}, \beta) = \frac{\exp^{(\alpha+X'_{i,t}\beta+\epsilon)}}{1 + exp^{(\alpha+X'_{i,t}\beta+\epsilon)}} \tag{2}$$

where Y_{it} is the period of vulnerability t (six quarters) prior to a crisis in country i. As already mentioned, in the baseline specification, we use an evaluation window up to six quarters before the crisis and distinguish between a sovereign, currency and a banking crisis. $X_{i,t}$ stands for the factors that signal a vulnerable state according to the AUROC approach. In order to deal with multicollinearity issues, we exclude some of the AUROC preselected variables that are highly correlated (i.e. GDP

[37] Latin America registered 29 crises from 1993q1 to 2018q4, of which 16 were currency crises (8 of them before 2006q4).

[38] Eastern European countries present by far the large number of banking crises on the sample, around 142 quarters, that is, 24 quarters by country on average.

[39] Almost all crises for this group of countries are banking or currency crises, and all are dated back before 2000q1.

[40] The five countries included in this cluster are probably the most heterogeneous of the sample, and for example the standard deviation of GDP growth and of GDP per capita is the highest of the four groups. Moreover the number of missing values is also the highest (almost 10% of the total possible data).

[41] For the sake of robustness, we have build SHERLOCs using the regional results, but out of sample validation is much poorer.

[42] Mean percentiles are calculated from the frequency distributions of the historical series of each variable form 1993 to 2018, as detailed in Alonso and Molina (2021).

Table 3
Logistic estimation using the variables preselected by the AUROC.

VARIABLES	Banking	Sovereign	Currency	Crises (any kind)
GDP growth		**−0.206*****	0.012	−0.014
		[0.056]	[0.024]	[0.023]
Inflation rate	−0.031***		−0.003	0.000
	[0.008]		[0.004]	[0.006]
NEER deviation from trend	**0.028*****			
	[0.006]			
Public debt (first difference)		0.057	**0.058*****	
		[0.053]	[0.086]	
Public sector balance			**−0.311*****	**−0.143*****
			[0.044]	[0.036]
Net Foreign assets dom banks	**−0.038*****			
	[0.007]			
Loan to Deposit ratio	**0.693*****			
	[0.243]			
Short-term interbank rate	**0.026*****	**0.065*****	**0.059*****	**0.059*****
	[0.010]	[0.022]	[0.009]	[0.010]
External debt		**0.038****		
		[0.018]		
Short-term external debt	−0.001	−0.035	**0.034*****	0.010
	[0.002]	[0.031]	[0.011]	[0.007]
Reserves (mm USD)		**−0.152*****	**−0.006*****	**−0.004*****
		[0.038]	[0.002]	[0.001]
Ext.debt service		−0.002		**0.011****
		[0.009]		[0.004]
Foreign direct investment				**−0.212*****
				[0.061]
Rating		−0.217	0.209***	0.194***
		[0.138]	[0.068]	[0.066]
US 10 Y interest rate	**0.689*****			
	[0.166]			
US 3 M interest rate	0.153			
	[0.109]			
Trade links		**0.237*****		
		[0.082]		
Observations	2197	2270	2407	2014
Number of id	25	25	25	25

Standard errors in brackets
***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.1.

growth and GDP per capita, or reserves in USD billion and reserves over GDP).

Individual (country) effects are incorporated into the model by using random effects, which assume that the country effects have a distribution.[43] The choice of random effects has several advantages. First, it represents a more efficient combination of "within" and "between information". Second, it enables us to exploit the information of all the countries in the sample, even those which have not suffered any type of crisis. On the contrary, the fixed effect model – which is a Conditional Logit model –, "eliminates" from the estimation the countries that have never faced a crisis. Understanding why some countries have never suffered a crisis is also relevant in our analysis. However, in order to use a random effect model, one needs to assume that the individual effects are not correlated with the independent variables.[44] Finally, as a robustness check, we also estimate the logit using pooled data (See Appendix A.11).

Results of the logistic estimation[45] are reported in Table 3.

In general the results of this multivariate analysis are coherent with those of the univariate (AUROC) estimations. For banking crises, the most relevant factors to explain a pre-crisis state are the proxy for global financial conditions (US long term interest rate) and the loan to deposit ratio, which points to an excessive leverage of domestic banks. Also the results for net foreign assets of domestic banks would indicate that the excessive leverage is carried out in foreign currency. Finally, the negative sign of the inflation rate can be explained by the fact that in high inflation environments banks have an advantage over their clients in the management of their assets and liabilities, for example lending at variable rates and taking deposits at fixed rates or longer maturities. For sovereign crises, a drop in activity and international reserves, an increase in short term domestic rates or external debt are highly relevant, as well as the trade links between EMEs, a proxy for one of the determinants of pure contagion. As in the AUROC approach, currency crises are closely related with sovereign crises, although in this case flows (public sector balance) also play a role.[46][47] The results for any kind of crisis summarize fundamentally those obtained for sovereign and currency crises, and add the most stable capital inflows and foreign direct investment, with the expected sign.

[43] While in the fixed effects, it is assumed to be fixed.

[44] The fixed effect model is also better at minimizing the "omitted variable" bias.

[45] All the logistic regressions in this section use a random effects model with standard errors derived from asymptotic theory (OIM) and include a constant term. Apart from the reasons already mentioned, the use of random effects is validated by a standard Hausmann test.

[46] The results shown in the table constitute the baseline for building the SHERLOC, although we have tested also the reliability of other specifications (see Appendix A.2 for details).

[47] We have excluded the sovereign spread as it tends to be a simultaneous indicator of crisis instead of being a leading variable. Indeed, the ratio of good signals increases monotonously as we get closer to the crisis date, which is not the case for the rest of variables. In addition, the sovereign differential has the highest pairwise correlation on average with the rest of the regressors, and also the highest among them (−0.62 with the rating). Nevertheless including the sovereign spread in the regression of currency and sovereign crises hardly vary the results.

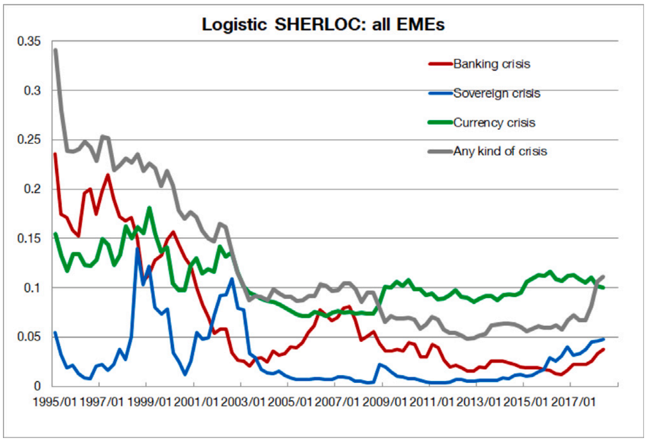

Fig. 5. The logistic SHERLOC.

SHERLOC is then calculated – on a quarterly basis – as the predicted probability of being in a pre-crisis state (6 quarters before a crisis) derived from the results of the logistic regressions reported in Table 3. Fig. 5 depicts the Logistic SHERLOC for the whole sample for each type of crisis, and Fig. 6 SHERLOCs by region.

Even in this aggregate form,[48] SHERLOC seems to capture the stress prior to the successive crises of the late 90 s, the tranquil times between 2006 and 2009, and the increase in vulnerability in some EMEs, especially in Latin America and Eastern Europe, at the end of 2018. Sovereign stresses are close to historical lows except in Latin America, as well as banking stress, with the exception of Eastern Europe.

4.2.2. A factor augmented logistic estimation

Another alternative to estimate the predicted probability of being in a vulnerability state consists of using a factor-augmented logit model. This methodology enables us to exploit all the available information in a dense model by using as explanatory variables the common factors extracted from the largest dataset that we have built. The underlying idea is to capture all the fundamental drivers of the economy that might lead to a vulnerable state. To do so, we first extract the common factors using a principal component analysis and incorporate all the factors with an eigenvalue higher than one in a logistic estimation.[49] The probability of being in a "vulnerable state" (i.e, six quarters prior to the crisis) is, therefore, assumed to follow a logistic distribution that depends on common factors, that is:

$$Pr(Y_{it} = 1|X_{i,t}, \beta) = \frac{\exp^{(\alpha + X'_{i,t}\beta + \epsilon)}}{1 + exp^{(\alpha + X'_{i,t}\beta + \epsilon)}} \tag{3}$$

where Y_{it} is the period of vulnerability t (six quarters) prior to a crisis in country i. $X_{i,t}$ stands for the six common factors extracted from the large dataset based on a principal component analysis.[50]

In the factor-augmented logit estimation, we follow the methodology proposed by Chen et al. (2011) due to two main reasons. First, our dummy variable is equal to one the six quarters previous to a crisis, and therefore we capture all the indicators that might affect crises from one lag to six lags prior to the onset of the crisis. In this sense, we do not consider necessary to include dynamics as (Bellégo and Ferrara, 2012) do.[51] Second, as pointed out by Chen et al. (2011), the extracted factors from the PCA "may already incorporate the lags of underlying dynamic factors".

Figs. 7 and 8 show the developments of vulnerabilities based on the FA-logistic approach. Results are qualitatively very similar to the logistic SHERLOC. Latin America is the most vulnerable region at the end of the sample due to an increase in sovereign stress and in currency stress, which also increased in Eastern Europe around 2018.

In Appendix A.3 we estimate the SHERLOC using two other alternative methodologies, a principal component analysis and just calculating the mean percentiles of the selected variables.

4.3. Validating SHERLOC

In order to assess the performance of our synthetic index (the logistic SHERLOC) in predicting "pre-crisis" events, we follow the methodology proposed by Alessi and Detken (2011), who develop a measure of Absolute Usefulness based on policy maker's preferences. This methodology has also been used by Duca and Peltonen (2013) to validate different models to predict systemic financial crisis and by Babecký et al. (2014) to assess the performance of early warning indicators of debt crises. An index is considered to be "useful" whenever there is a gain in using this index as compared to ignoring it. In other words, the loss of not using the index is larger than the loss of using the index.

Assume that $L(\theta)$ is the loss function when using the index proposed, which is dependent on a policy makers' preferences between type I and type II errors:

$$L(\theta) = \theta(TypeI) + (1 - \theta(TypeII)) \tag{4}$$

The parameter θ captures the policy maker's preferences between type I and type II errors. While Type I error calculates the share of missed vulnerable periods in terms of total number of stress periods, Type II errors represents the ratio of false alarms issued in terms of "tranquil periods". A value of θ larger than 0.5 reveals that a policy maker prefers receiving a false alarm rather than missing a crisis.

The measure of Absolute Usefulness (U) of Alessi and Detken (2011) is formally defined as:

$$U = min[\theta; 1 - \theta] - L(\theta) \tag{5}$$

where $min[\theta; 1-\theta]$ is the loss that the policy maker incurs when ignoring the index and therefore she either always assumes there is a signal and has to react if $\theta > 0.5$ or there is never a signal and does not react if $\theta < 0.5$; and $L(\theta)$ is the loss function when using the index proposed, previously defined. In our framework, we assume that policy makers are as concerned about missing a crisis as they are about issuing a false alarm, and therefore we set $\theta = 0.5$. Optimal thresholds are calculated by maximizing the measure of Absolute Usefulness (U). Additionally, we rank the indexes obtained using different methodological approaches according to its "Usefulness". If this measure is positive, it provides useful signals for policy makers compared to ignoring it. The larger the measures, the higher the benefit.

This methodology has several advantages. First, it takes into account policy makers' preferences over missed crises versus false alarms. Second, it enables us to calculate the optimal threshold beyond which an early warning signal is issued. Thresholds are optimized for each index for the given preference parameter $\theta = 0.5$ using all the information in the evaluation sample in order to get the percentile of the distribution that maximizes the Absolute Usefulness measure. Finally, it is quite intuitive and easily understood by policymakers. An index or model is "useful" whenever there is a benefit of using the index to detect vulnerable states or a pre-crisis period.

---

[48] The aggregate SHERLOC is the simple average of country's SHERLOCs. Using a weighted average (via GDP in PPP terms or the number of crises of each country, to capture those more prone to register turbulences) does not alter the picture.

[49] We end up with six common factors that account for 60% of the data variation among all the variables included in the principal component analysis.

[50] The logistic regressions use a random effect model with standard errors derived from asymptotic theory (OIM) and include a constant term.

[51] In their case, the dummy only takes value 1 when the economy is in recession and, hence, they need to include the dynamic effects in the explanatory variables.

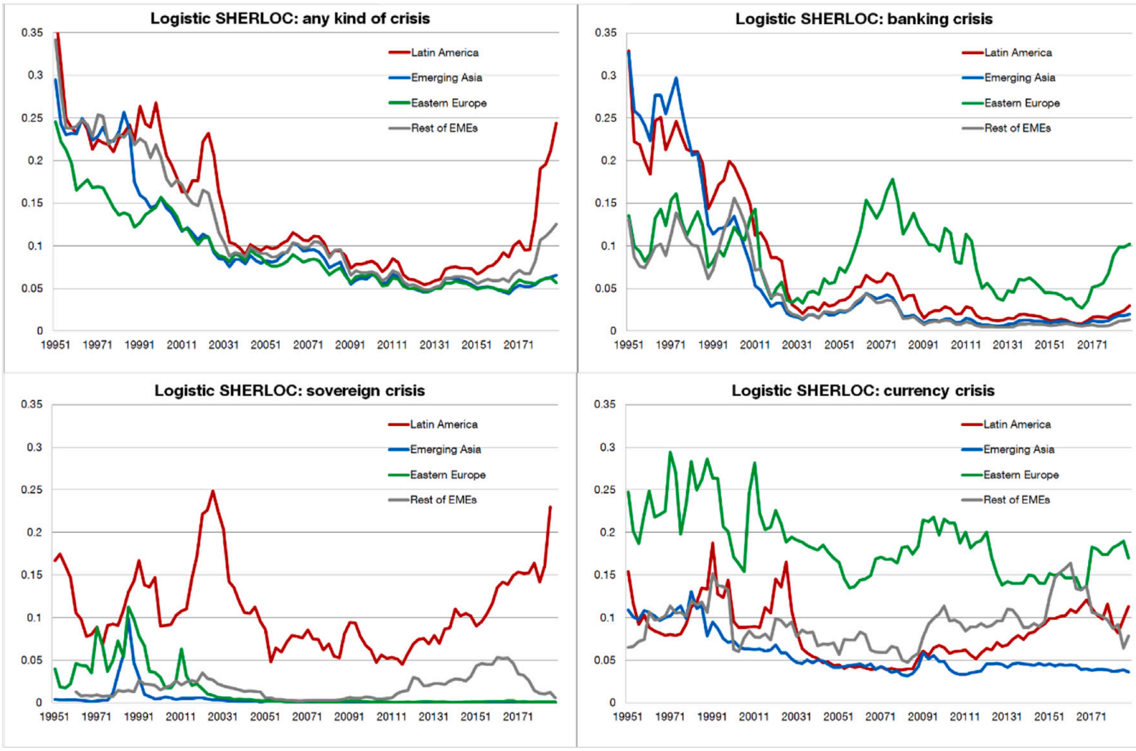

Fig. 6. The logistic SHERLOC.

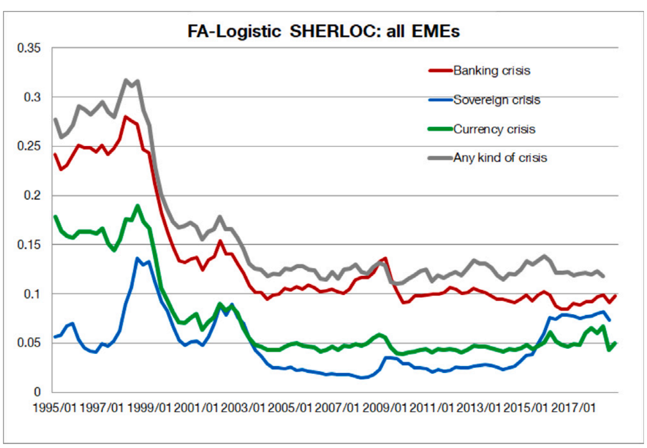

Fig. 7. The FA SHERLOC.

As a robustness check, we also use two other alternative evaluation methods: the Brier score and the Diebold and Mariano test. The Brier score measure the accuracy of probability forecasts using a quadratic score, where the square differences between actual vulnerable states, six quarters before the crisis, and the predictions of our models are calculated (see Brier et al. (1950)). The lower the score, the better. The Diebold and Mariano test measures the predictive accuracy between two competing forecasts by country (Diebold and Mariano, 2002).

4.3.1. *In-sample performance of SHERLOC*

We evaluate the performance of the different indexes in predicting a "vulnerable" or a "pre-crisis" state using the sample over the period 1993 to 2018. For each index proposed, we calculate the threshold for the estimated likelihood of a "vulnerable" state that maximizes the usefulness score.

Table 4 reports the in-sample evaluation for the banking, sovereign and currency indexes proposed and an aggregate index for all type of crises, based on the measure of Usefulness (U), the noise to signal ratio (NtSr), the percentage of predicted crisis (% Predicted) and the number of correct signals in terms of total signals issued (Cond Prob (%)). More specifically, this table presents the performance of the indexes for the methodology proposed-the predicted probability based on a logistic estimation (logistic SHERLOC) and the other three methodologies used to build alternative indexes: the predicted probability based on a factor-augmented logistic approach (FA logistic approach), a principal component analysis from the variables preselected using the AUROC (PCA index) and the simple average of the percentiles of the relevant variables according to the AUROC[52] (mean percentile index). For comparison, we also include the performance of the best single indicators derived from the AUROC: the short-term interbank rate in the case of the banking crises; and the sovereign spread for sovereign and currency indexes. The variation of reserves is also assessed for currency crises as it has been an early warning indicator broadly used in the literature, as well as the stock of banks and deposits in the case of the banking crises.[53] Finally, two additional evaluation measures are used: the Brier score and the Diebold–Mariano test. As expected, all models achieve a positive Usefulness measure, which suggests that the indexes proposed provide gains for policy makers. Indeed, the usefulness ratios in our exercise are similar to those reported by Alessi and Detken (2011) in a univariate setting and for the same preference parameter (θ) and by Babecký et al. (2014) relying on a composite early warning index. Both papers find similar usefulness values of around 0.15–0.25.[54] Duca and Peltonen (2013) obtain slightly larger Usefulness values than these two papers, with values around 0.19–0.34, in line with the results of the Logistic SHERLOC presented in this paper for each type of crisis (0.20–0.33).

52 We preselect the variables with an AUROC above 65 in the case of banking and currency crises, and above 70 for sovereign crises.

53 Results can be provided upon request.

54 Alessi and Detken (2011) report values around 0.2 and 0.25 while (Babecký et al., 2014) get usefulness values around 0.15 and 0.20.

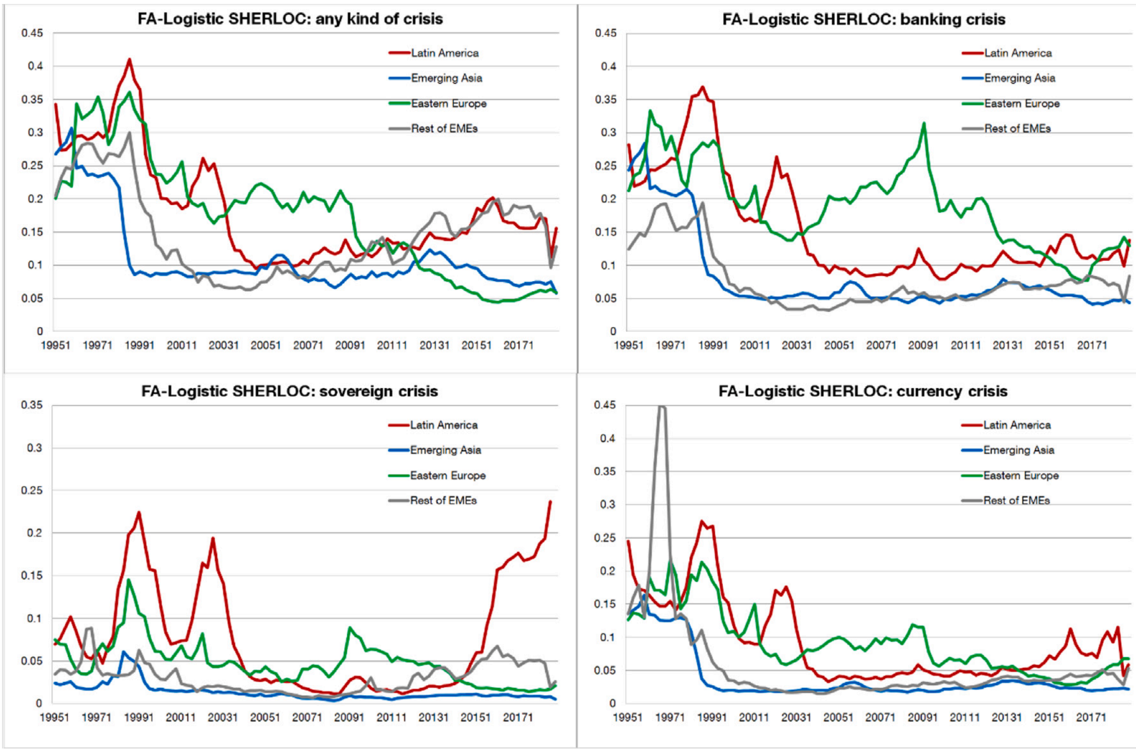

Fig. 8. The FA SHERLOC.

More significantly, all the models proposed outperform their respective best single indicators: the short-term interbank rate for banking crises and the sovereign spread for sovereign and currency crises. Therefore, the use of a composite early warning index (SHERLOC) seems to be more accurate to anticipate vulnerable states of countries than the use of a single indicator, which accords with the results of Babecký et al. (2014) and Duca and Peltonen (2013). However, in this paper we also show that the creation of an index for each type of crisis seems to outperform an aggregate index for all types of stress. Indeed, the usefulness values when using an aggregate index for all crises are much lower than the results of the SHERLOC, regardless of the methodology used and the type of crisis. The underlying reason is the wide heterogeneity of the nature of crises, which therefore should be treated in a different way.

Additionally, the usefulness measures enable us to rank the different indexes proposed. The higher the usefulness measure, the better. The Logistic SHERLOC outperforms the three other methodologies used (the FA logistic index; the PCA index and the mean percentile index), as it achieves higher usefulness scores for the banking and sovereign indexes.[55] In addition, the ratio of predicted crises are very high for the three types of crises considered (around 87% for banking crises, 79% for sovereign crises and 81% for currency crises). The Brier score and the Dielbold–Mariano test also suggest a relatively good accuracy of the Logistic SHERLOC.[56]

Finally, comparing the Logistic SHERLOC validation exercise with the outcome of previous literature, our index outperforms some of the indexes or variables proposed. Indeed, the usefulness measures of our analysis are similar or higher than the ones reported by Alessi and Detken (2011) and Babecký et al. (2014). They calculated values around 0.2 and 0.25, and 0.2 respectively. In the case of our Logistic SHERLOC, we obtained values of 0.33, 0.27 and 0.2 for banking, sovereign and currency, respectively. Therefore, our index seems to be better at predicting banking and sovereign crises at least. However, results are not strictly comparable since their analysis focused on developed economies, instead of emerging economies, over a different period of analysis. In addition, Alessi and Detken (2011) predict asset booms instead of crises. If we compare them with (Duca and Peltonen, 2013), they find similar values for a mixed sample of advanced and emerging economies. The percentage of predicted crises is also similar. However, they only focus on financial stress instead of taking a broader perspective and considering all types of stress. Finally, Lepers and Serrano (2020) do not provide any evaluation methodology, although they point out that their results are better than the usual credit to GDP gap. The creation of asset bubbles is also captured in our analysis using two tail risks for real credit growth, but the AUROC results led us to discard it.

Therefore, the SHERLOC proposed will be based on the logistic estimation. However, as the PCA index, whose interpretation is more straightforward and its update is simpler, also seems to perform adequately, it can be used as a robustness check. The FA logistic index seems quite appropriate for currency crises. The Mean percentile index will be discarded as its performance does not seem to be good enough.

4.3.2. Out-of-sample performance of SHERLOCs

In order to assess the predictive ability of each index out-of-sample, we split our sample into two subsamples. The period from 1993 to 2006 is used to estimate our models and calibrate the optimal threshold beyond which each index issues an early warning signal. The last years of the sample (2007–2018) are used to estimate the performance out-of-sample.[57]

Table 5 presents the results of the out-of-sample validation for the different indexes proposed for the banking, sovereign and currency

[55] The FA logistic index achieves a higher usefulness for currency crises.

[56] The FA logistic index also seems to perform well in the case of currency and all crises according to this two evaluation methods.

[57] As we noted before, there seems to be some structural changes around 2006 as the number of crises dwindles (reaching zero in some quarters). Using 2007 as the break point enables us to have more crises out-of-sample to validate SHERLOCs. Nevertheless we have also tested the performance of SHERLOCs from 2014 onward, and the results (provided upon request) do not

Table 4
In-sample performance of indexes.

Model	Threshold (percentile)	U	NtSr	Predicted %	Cond Prob (%)	Other validation tests Brier score	DM % logistic
Banking indexes							
Logistic SHERLOC	75	0.33	0.24	86%	17%	0.04	–
FA Logistic index	72	0.30	0.29	84%	15%	0.04	60%
PCA index	66	0.25	0.36	78%	12%	–	100%
Mean percentile index	76	0.20	0.37	63%	11%	0.18	96%
Short-term interbank rate	56	0.19	0.50	79%	9%	–	100%
Sovereign indexes							
Logistic SHERLOC	72	0.27	0.32	79%	6%	0.02	–
FA Logistic index	75	0.24	0.32	71%	6%	0.02	68%
PCA index	90	0.23	0.16	54%	12%	–	100%
Mean percentile index	85	0.19	0.27	52%	7%	0.22	100%
Sovereign spread	55	0.17	0.57	77%	3%	–	
Currency indexes							
Logistic SHERLOC	57	0.20	0.50	81%	10%	0.06	–
FA Logistic index	74	0.26	0.31	74%	16%	0.05	44%
PCA index	71	0.20	0.39	67%	13%	–	100%
Mean percentile index	79	0.15	0.39	50%	13%	0.20	96%
Sovereign spread	55	0.08	0.74	59%	8%	–	100%
All crises							
Logistic SHERLOC	74	0.14	0.43	50%	18%	0.08	–
FA Logistic index	80	0.22	0.26	60%	27%	0.07	40%
PCA index	64	0.16	0.47	60%	17%	–	100%
Mean percentile index	55	0.10	0.67	63%	12%	0.14	84%

This table presents the in-sample validation exercise for the banking, sovereign, currency indexes proposed and an aggregate index for all type of crisis (all crises) and the four different methodological approaches used to build the indexes: the predicted probability from a logistic estimation (Logistic SHERLOC) and from a factor augmented logistic estimation (FA Logistic index), a principal component analysis from the variables (PCA index), and the average of the risk percentiles (mean percentile index). For the PCA and the mean percentile index, see Appendix A.3. The validation exercise is based on the measure of Usefulness (U), the noise to signal ratio (NtSr), the percentage of predicted crisis (% Predicted) and the number of correct signals in terms of total signals issued (Cond Prob (%)). The threshold indicates the percentile beyond which the index issues a signal. Short-term interbank rate and sovereign spread represent the best single indicators according to the AUROC. This exercise is based on a neutral policy maker (θ = 0.5) and evaluation window six quarter previous to the crisis. Two additional tests are also presented. The Brier score measures the accuracy of probabilistic forecasts using a quadratic scoring rule. The lower the score, the better. The Diebold–Mariano test calculates a measure of predictive accuracy between two competing forecasts. In this table, we test the accuracy with regard to the logistic SHERLOC.

stress and an aggregate index for all type of crises. As in the in-sample validation exercise, we will assess the performance of our index based on the "usefulness" measure (U) developed by Alessi and Detken (2011). In addition, this table also presents the noise to signal ratio (NtSr), the percentage of predicted crisis (% Predicted) and the number of correct signals in terms of total signals issued (Cond Prob (%)), as well as two additional evaluation measures: the Brier score and the Dielbold–Mariano test.

As expected, out-of-sample performance is slightly poorer than in-sample performance although usefulness measures remain positive and higher than the best single predictors, with the exception of sovereign crises and the FA logistic index for currency and sovereign crises. Indeed, usefulness values hover around 0.06 and 0.15, which is somewhat in line with the out-of-sample results reported by Duca and Peltonen (2013). As was the case with the in in-sample predictions, the SHERLOC indexes perform much better than the best single indicators and the use of an index for each type of crisis outperforms an aggregate index of all crises in terms of usefulness.[58] The Brier score also provides encouraging results.

Finally, when comparing between the different methodologies employed to construct the SHERLOC, the Logistic SHERLOC seems to perform adequately. Indeed, it outperforms other methodologies for currency and banking crises, as suggested by the usefulness measures and the percentage of predicted crises, 49% and 58%, respectively. In the case of sovereign crises, the usefulness value (0.08) is lower than the scores obtained with the PCA and mean percentile (0.12). But, the percentage of predicted crises (50%) is higher or equal to the percentage predicted by the PCA and mean percentile. Moreover, one needs to be aware that there were only three episodes of sovereign crises over this period, including the "hold-out" crisis in Argentina, which is a particular case, difficult to anticipate with macro variables. In addition, the Brier score is also much smaller for the Logistic SHERLOC than for the mean percentile, very similar to the result of the FA logistic index. Finally, the Diebold–Mariano tests suggest that our index outperforms the other methodologies since most countries in our sample are better served by the Logistic SHERLOC than by the three other methodologies.[59] Indeed, while the PCA index also provides a positive usefulness measure out-of-sample, suggesting that it can also be an adequate measure to anticipate vulnerable states, the results are very poor in terms of the Diebold–Mariano test. The FA logistic index only provides substantial gains in the case of currency crises, while its out-of-sample forecast performance is very low in the case of banking and sovereign crises. Finally, the Mean Percentile index provides a low gain in the case of the currency index (0.06), and a poor performance according to the Brier score and the Dielbold–Mariano tests, and, therefore, it is discarded.

change significantly. Indeed, in some cases, the model improves in terms of performance, for instance for banking crises.

[58] Although the PCA index for all crises also seems to perform adequately and the performance of the FA logistic approach for sovereign and currency crises is quite poor (0.04 and 0.01 respectively).

[59] In the case of currency crises, the Diebold–Mariano test seems to slightly prefer the FA Logistic index but the results are not conclusive since the difference between both forecasts is not statistically significant.

Table 5
Out-of-sample performance of indexes.

Model	Threshold (percentile)	U	NtSr	Predicted %	Cond Prob (%)	Other validation tests: Brier score	DM % logistic
Banking indexes							
Logistic SHERLOC	70	0.14	0.53	58%	4%	0.05	–
FA Logistic index	67	0.04	0.81	42%	3%	0.04	88%
PCA index	51	0.13	0.65	75%	3%	–	100%
Mean percentile index	70	0.14	0.54	58%	4%	0.18	92%
Short-term interbank rate	53	0.07	0.77	63%	3%	–	100%
Sovereign indexes							
Logistic SHERLOC	69	0.08	0.68	50%	2%	0.04	–
FA Logistic index	70	0.01	0.93	33%	2%	0.03	88%
PCA index	80	0.12	0.45	44%	3%	–	100%
Mean percentile index	74	0.12	0.52	50%	3%	0.24	96%
Sovereign spread	51	0.11	0.69	72%	2%	–	100%
Currency indexes							
Logistic SHERLOC	76	0.11	0.55	49%	8%	0.09	–
FA Logistic index	70	0.15	0.49	60%	9%	0.04	44%
PCA index	61	0.08	0.70	54%	7%	–	100%
Mean percentile index	79	0.06	0.66	33%	7%	0.21	88%
Sovereign spread	51	0.00	1.01	49%	5%	–	100%
All crises							
Logistic SHERLOC	66	0.07	0.71	49%	10%	0.07	–
FA Logistic index	71	0.07	0.65	43%	11%	0.07	72%
PCA index	51	0.12	0.65	72%	11%		100%
Mean percentile index	64	0.04	0.81	46%	9%	0.13	80%

This table presents the out-of-sample validation exercise for the banking, sovereign, currency indexes proposed and an aggregate index for all type of crisis (all crises) and the four different methodological approaches used to build the indexes: the predicted probability from a logistic estimation (Logistic SHERLOC) and from a factor augmented logistic estimation (FA Logistic Sherloc), a principal component analysis from the variables (PCA SHERLOC), and the average of the risk percentiles(mean percentile SHERLOC). For the PCA and the mean percentile index, see Appendix A.3. The validation exercise is based on the measure of Usefulness (U), the noise to signal ratio (NtSr), the percentage of predicted crisis (% Predicted) and the number of correct signals in terms of total signals issued (Cond Prob (%)). The threshold indicates the percentile beyond which the index issues a signal. Short-term interbank rate and sovereign spread represent the best single indicators according to the AUROC. This exercise is based on a neutral policy maker (θ = 0.5) and evaluation window six quarter previous to the crisis. Two additional tests are also presented. The Brier score measures the accuracy of probabilistic forecasts using a quadratic scoring rule. The lower the score, the better. The Diebold–Mariano test calculates a measure of predictive accuracy between two competing forecasts. In this table, we test the accuracy with regard to the logistic SHERLOC.

4.3.3. *Summing up the validation exercise*

In light of the in-sample and out-of-sample exercises, the (preferred) vulnerability index proposed will be based on a logistic estimation as it outperforms other indexes in-sample and performs adequately out-of-sample. As a robustness check, the PCA index can also be used since its out-of-sample performance remains good and the FA logistic index can be useful but only for currency crises. We discard the mean percentile index as its predictions do not seem to be good enough, especially in-sample.[60]

The high type II errors are a source of concern as suggested by the high noise to signal ratios and the low percentage of conditional probability both in-sample and out-of-sample. Indeed, early signals tend to be noisier, but they are more valuable for policymakers. A possible way to reduce the elevated noise is to take into consideration the "post-crisis bias". Indeed, macroeconomic variables tend to show a strong persistence and therefore they have an "erratic" behaviour in the recovery phase. This is one of the issues of binary choice models since the model cannot distinguish between a period where the economy is back to normal times and a period in which the country is still adjusting. As a robustness check, we remove post-crisis years (4 quarters), as done with crisis years, at a cost of fewer observations. We show that considering the "post-crisis" bias is relevant to slightly reduce type II errors out-of-sample, but it does not seem to affect the ratio of crises predicted in-sample and it has only a marginal impact out-of-sample (See Appendix A.6).

Finally, our model is robust to different specifications and, broadly speaking, it tends to outperform alternative specifications. As robustness checks, we validate the performance of our SHERLOC using different logistic estimations: including and excluding some variables (Appendix A.2), using a different definition of currency crisis (Appendix A.5), taking into consideration the "pseudo-real time approach" (Appendix A.7), relying on different evaluation windows (Appendix A.8), assuming different preferences of policy makers (Appendix A.9), calculating the index by region (Appendix A.10), or using pooled data (Appendix A.11).

[60] A striking result of SHERLOC is the irrelevant role played by institutional indicators or variables related to social and political tensions. Some institutional indicators were included,- such as the World Bank's Doing Business or Stability and Absence of Violence- or political strains proxies -IHS Markit political risk index. The results of the AUROC for those variables are very poor (Appendix A.2), probably as we do not have series long enough. Moreover, some of these conflicts could be captured by the sovereign ratings. Nevertheless for the sake of completeness we introduced new text based or machine learning indicators of conflicts or uncertainty in SHERLOC. We have used the conflict indicators developed by Mueller and Rauh (2022). Results are interesting: indicators related to economic and politics topics issue adequate signals according to AUROC (0.74 for sovereign crises, 0.68 for banking crises in the case of economics and 0.69 for sovereign defaults in the case of politics). In the logistic regression, the coefficient for economics topics is significant with the expected (positive) sign for sovereign defaults, banking crises and currency crashes, without channging the sign and significance of the rest of the variables. Nevertheless when validating this conflict-augmented SHERLOC the results are disappointing. Across all methodologies their performance in-sample and out-of-sample is worst than the original SHERLOC (see Appendix A.2, indicator RC4). The rest of results can be submitted upon request.

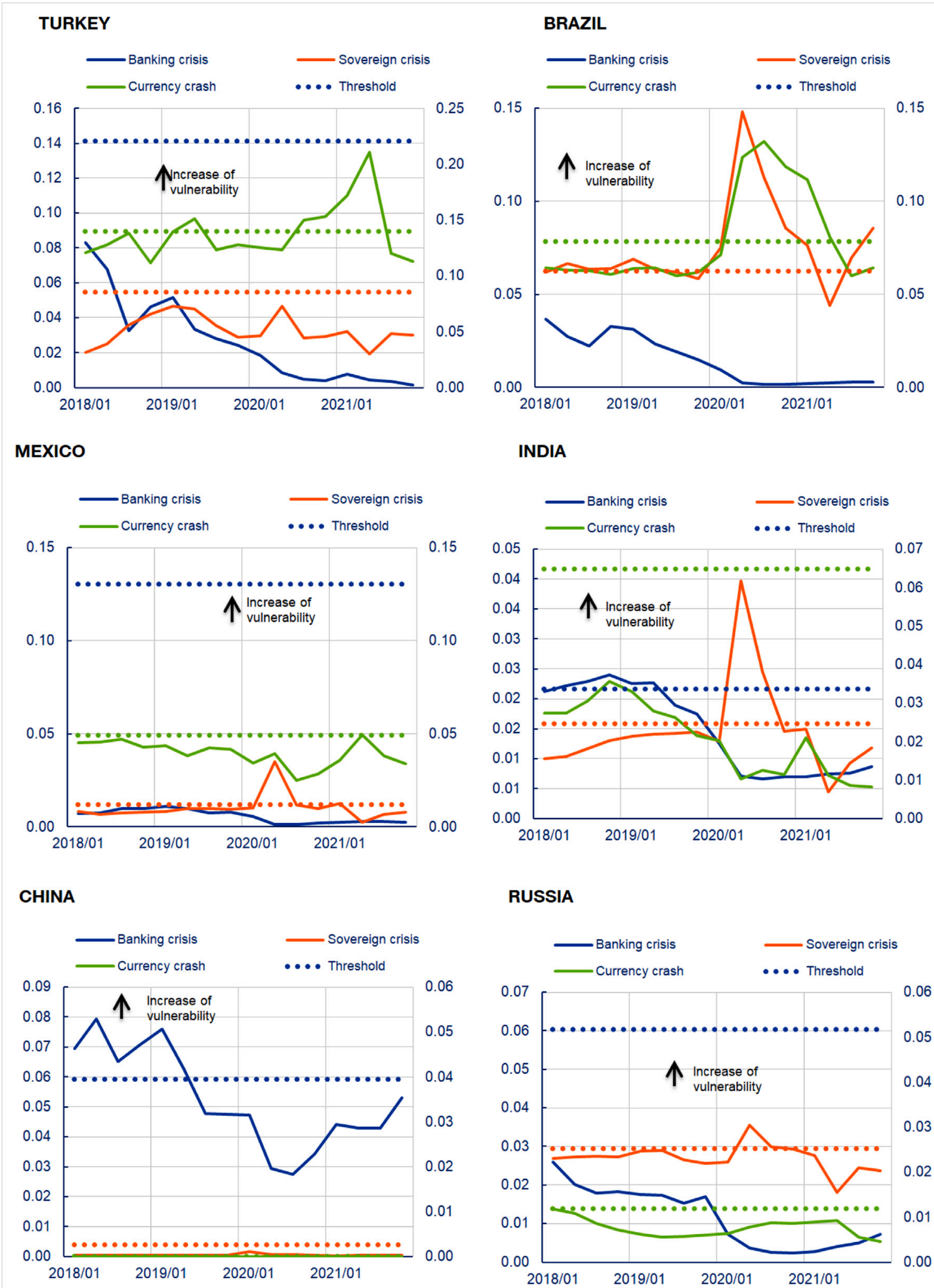

Fig. 9. Updating SHERLOC until the last quarter of 2021.

4.3.4. The usefulness of SHERLOC over the period of 2018 and 2021[61]

Throughout the paper we have shown that SHERLOC outperforms other methodologies to build synthetic indexes and the best single indicators both in-sample and out-of-sample. Moreover, one of the main advantages of this index is that it is easy and quick to update, and therefore, it can be used as a tool for assessing the likelihood of facing a crisis or a high degree of vulnerability in the regular analysis of the relevant economies.[62] In this section we briefly describe the developments of SHERLOC for some countries since the end of 2018, focusing on the pre-covid situation and in the global context at the end of 2021, characterized by a tightening of global financial conditions – propitiated by either an increase in official interest rates in developed markets or a strong appreciation of the US dollar –, a bout of uncertainty about the outlook for economic activity, and a surge in commodity prices.

In this context, this tool enables policy makers to detect the more vulnerable EMEs, and therefore those more prone to face any kind of crisis at some point in time, for instance at the end of 2019 and 2021. To do so, we first update the dependent variable, that is, the dummy variable for each type of crisis. According to Standard and Poor's, since 2018 there have been just two sovereign defaults (Argentina in the third quarter of 2019 – selective default – and in May 2020, and Ecuador in the second quarter of 2020[63]), one country with banking crises (China in the third quarter of 2020 – Baoshang bank – and the third quarter of 2021 -Huarong-), and a currency crisis (the Turkish lira at the end of 2021 depreciated by 30.9 percent against the USD).[64]

The development of SHERLOCs for the most relevant EMEs since the beginning of 2018 is shown in Fig. 9. The main takeaways from this figure would be the notable resilience of the banking sector, and, by contrast, the fact that the most vulnerable post-pandemic economies were those that suffered sharp increases in their external or public debt or in their deficit, such as Brazil, although none of them experienced a currency or a sovereign crisis. Finally, at the end of 2021, that is, just a quarter before the Fed started to rise official rates and Russia invaded Ukraine, the most vulnerable country was Turkey, which experienced a currency crisis as stated above. Note also that whereas the probability of being in a vulnerable state against currency crashes or sovereign defaults in China is zero, the probability of a banking crisis in this country has been hovering around the signal threshold since the beginning of 2018, and that the authorities in fact have had to intervene or liquidate two medium-sized banks.[65]

5. Conclusions and work ahead

Is it possible to detect vulnerabilities in emerging markets with enough time to implement measures to tame its effects from a financial stability point of view? In this paper we propose a user friendly tool to monitor vulnerabilities in 25 EMEs based on an synthetic index (called SHERLOC) for three types of crisis. These indexes capture the developments of the leading indicators to anticipate pre-crisis periods, selected using a signalling approach (AUROC) to mitigate the usual data mining and overfitting caveats. Statistical validation techniques suggest that the use of different SHERLOCs for each type of crisis implies a gain of utility with respect to the use of an aggregate index for all crises and with respect to the individual indicators with the best performance according to the AUROC. This can be explained by the heterogeneous nature of crises. Results are robust to considering different specifications of SHERLOC and different subsamples, which highlight the relevance of this tool to monitor risks in EMEs. A useful example is included in Section 4.3.4, with a brief analysis of the vulnerability situation of the main EMEs a quarter before the pandemic and the invasion of Ukraine.

Nevertheless one needs to be aware of the main pitfalls of the methodology used. First, noise remains elevated, partly due to the "post-crisis bias". A possible solution is to remove post-crisis episodes at the cost of fewer observations. Second, the criticism of "this time is different" also applies. However, as "not every time is different" is also a possibility, the use of variables that have anticipated crises in the past may be useful. Third, binary choice models require a sufficient number of stress events to get robust results. To overcome this shortcoming, it is possible to use continuous variables, such as Financial Stress Indicators (FSI) and apply Markov Switching models that endogenously determine the beginning and exit of the crisis. However, this choice comes at the cost of not evaluating the "vulnerable" state but the "crisis period". Additionally, the main drawback is that is difficult to construct continuous variables long enough to capture all the events and to cover a long period of time. Fourth, the results depend on policymakers' preferences for Type I and Type II errors, and also on the definition of crises. Fifth, these models are not able to capture non-linearities. And finally, the introduction of some variables that capture political and social strains or banking regulatory issues could also improve the performance of these models. These issues would be interesting topics for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econmod.2023.106304.

References

Alessi, L., Detken, C., 2011. Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. Eur. J. Political Econ. 27 (3), 520–533.
Alonso, I., Molina, L., 2021. A GPS Navigator to Monitor Risks in Emerging Economies: the Vulnerability Dashboard. Banco de EspaÑA Occasional Paper, No. 2111.
Babeckỳ, J., Havránek, T., Matěj, J., Rusnák, M., Šmídková, K., Vašíček, B., 2014. Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. J. Financ. Stab. 15, 1–17.
Banco de España, 2022a. Impact on the Emerging Economies Relevant to the Spanish Banking System of Tighter Global Financing Conditions and Rising Commodity Prices. Financial Stability Report, Box 1.2, No. Spring 2022.
Banco de España, 2022b. Report on the latin american economy, first half of 2022, Anal. Articles Banco de España.
Bellégo, C., Ferrara, L., 2012. Macro-financial linkages and business cycles: A factor-augmented probit approach. Econ. Model. 29 (5), 1793–1797.
Berg, M.A., Borenzstein, M.E., Pattillo, C.A., 2005. Assessing Early Warning Systems: How Have They Worked in Practice? IMF Working Papers, No. WP/04/52.
Beutel, J., List, S., von Schweinitz, G., 2018. An Evaluation of Early Warning Models for Systemic Banking Crises: Does Machine Learning Improve Predictions? Deutsche Bundesbank Discussion Papers, No. 48, Deutsche Bundesbank.
Billio, M., Casarin, R., Costola, M., Pasqualini, A., 2016. An entropy-based early warning indicator for systemic risk. J. Int. Financ. Mark. Inst. Money 45, 42–59.
Boonman, T.M., Jacobs, J.P.A.M., Kuper, G.H., Romero, A., 2019. Early warning systems for currency crises with real-time data. Open Econ. Rev. 30, 813–835.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brier, G.W., et al., 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78 (1), 1–3.
Bussière, M., 2013. In Defense of Early Warning Signals. Banque de France Working Paper Series, No. 420, Banque de France.
Bussiere, M., Fratzscher, M., 2006. Towards a new early warning system of financial crises. J. Int. Money Finance 25 (6), 953–973.
Camacho, M., Palmieri, G., 2021. Evaluating the OECD's main economic indicators at anticipating recessions. J. Forecast. 40 (1), 80–93, January.

[61] We provide data, STATA code and the definition of events.

[62] For example as in Banco de España (2022b).

[63] The default of Russia occurred in the second quarter of 2022, which are not included in the updated dataset.

[64] The Russian rouble depreciated by 20 percent in the first quarter of 2022, so technically it did not count as a currency crash.

[65] The results for the rest of EMEs may be provided upon request.

Candelon, B., Dumitrescu, E.I., Hurlin, C., 2014. Currency crisis early warning systems: Why they should be dynamic. Int. J. Forecast. 30 (4), 1016–1029.
Castro, C.E., Estrada, Á., Martínez-Pagés, J., 2016. The Countercyclical Capital Buffer in Spain: an Analysis of Key Guiding Indicators. Banco de EspaÑa Working Papers, No. 1601, Banco de España Working Paper.
Catão, L.A., Milesi-Ferretti, G.M., 2014. External liabilities and crises. J. Int. Econ. 94 (1), 18–32.
Chen, Z., Iqbal, A., Lai, H., 2011. Forecasting the probability of US recessions: a probit and dynamic factor modelling approach. Can. J. Econ./Rev. Can. d'Écon. 44 (2), 651–672.
Dabrovski, J.J., Beyers, C., de Villiers, J.P., 2016. Systemic banking crisis early warning systems using dynamic Bayesian networks. Experts Syst. Appl. 62, 225–242.
Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. J. Bus. Econom. Statist. 20 (1), 134–144.
Duca, M.L., Peltonen, T.A., 2013. Assessing systemic risks and predicting systemic events. J. Bank. Financ. 37 (7), 2183–2195.
Edison, H.J., 2003. Do indicators of financial crises work? An evaluation of an early warning system. Int. J. Finance Econ. 8 (1), 11–53.
Eichengreen, B., 2002. International financial crises: is the problem growing? Jahrb. Wirtschaftsgeschichte/Econ. Hist. Yearb. 43 (1), 89–104.
Eichengreen, B., Rose, A.K., Wyplosz, C., 1995. Exchange market mayhem: the antecedents and aftermath of speculative attacks. Econ. Policy 10 (21), 249–312.
Eichengreen, B., Rose, A.K., Wyplosz, C., 1996. Contagious Currency Crises. NBER Working Papers, No. 5681.
Frankel, J.A., Rose, A.K., 1996. Currency crashes in emerging markets: An empirical treatment. J. Int. Econ. 41 (3–4), 351–366.
Frankel, J., Saravelos, G., 2012. Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. J. Int. Econ. 87 (2), 216–231.
Gadea Rivas, M.D., Perez-Quiros, G., 2015. The failure to predict the great recession—a view through the role of credit. J. Eur. Econom. Assoc. 13 (3), 534–559.
González-Minguez, M.J.M., Carrascal, C., 2019. A crisis early warning model for euro area countries. Anal. Articles Econ. Bull. Banco España.
Gramlich, D., Miller, G., Oet, M.V., Ong, S.J., 2010. Early warning systems for systemic banking risk: critical review and modeling implications. Banks Bank Syst. 5, 199–211.
IMF, 2016. Global financial stability report, april 2016: Potent policies for a successful normalization.
Kamin, S.B., Schindler, J., Samuel, S., 2007. The contribution of domestic and external factors to emerging market currency crises: an early warning systems approach. Int. J. Finance Econ. 12 (3), 317–336.
Kaminsky, G., Lizondo, S., Reinhart, C.M., 1998. Leading indicators of currency crises. IMF Staff Pap. 45 (1), 1–48.
Kaminsky, G.L., Reinhart, C.M., 1999. The twin crises: the causes of banking and balance-of-payments problems. Amer. Econ. Rev. 89 (3), 473–500.
Laeven, M.L., Valencia, M.F., 2012. Systemic Banking Crises Database: An Update. International Monetary Fund.
Laeven, M.L., Valencia, M.F., 2018. Systemic Banking Crises Revisited. International Monetary Fund.
Laeven, L., Valencia, F., 2020. Systemic banking crises database II. IMF Econ. Rev. 68 (2), 307–361.
Lepers, E., Serrano, A.S., 2020. Decomposing financial (in) stability in emerging economies. Res. Int. Bus. Finance 51, 101068.
Lestano, L., Jacobs, J., Kuper, G., 2004. Indicators of Financial Crises Do Work! an Early-Warning System for Six Asian Countries. Technical Report, University Library of Munich, Germany.
Li, J., Chen, W., 2014. Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. Int. J. Forecast. 30 (4), 996–1015.
Martinez Peria, M.S., 2002. A regime-switching approach to the study of speculative attacks: A focus on EMS crises. In: Advances in Markov-Switching Models. Springer, pp. 159–194.
Mueller, H., Rauh, C., 2022. The hard problem of prediction for conflict prevention. J. Eur. Econom. Assoc..
Oet, M.V., Bianco, T., Gramlich, D., Ong, S.J., 2013. SAFE: An early warning system for systemic banking risk. J. Bank. Financ. 37 (11), 4510–4533.
Qin, X., Luo, C., 2014. Capital account openness and early warning system for banking crises in G20 countries. Econ. Model. 39, 190–194.
Reinhart, C.M., Rogoff, K.S., 2009. This Time Is Different: Eight Centuries of Financial Folly. Princeton University Press.
Rose, A.K., Spiegel, M.M., 2011. Cross-country causes and consequences of the crisis: An update. Eur. Econ. Rev. 55 (3), 309–324.
Samitas, A., Kampouris, E., Kenourgios, D., 2020. Machine learning as an early warning system to predict financial crisis. Int. Rev. Financ. Anal. 71.
Wang, T., Zhao, S., Zhu, G.Z., Zhengvon, H., 2021. A machine learning-based early warning system for systemic banking crises. Appl. Econ. 53 (26).