



WP/21/197

IMF Working Paper

Can Financial Soundness Indicators Help Predict Financial Sector Distress?

by Marcin Pietrzak

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

IMF Working Paper

Statistics Department

Can Financial Soundness Indicators Help Predict Financial Sector Distress?

Prepared by Marcin Pietrzak¹

Authorized for distribution by Marco Espinosa-Vega

July 2021

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

This paper shows how the role of Financial Soundness Indicators (FSIs) in financial surveillance can be usefully enhanced. Drawing from different statistical techniques, the paper illustrates that FSIs generate signals that can accurately detect, with 4 to 12 quarters lead, emerging financial distress—as measured by tight financial conditions.

JEL Classification Numbers: C40, E44, E60, G01

Keywords: financial stability, early warning signals, financial conditions, macroprudential

Author's E-Mail Address: marcin_pietrzak@brown.edu

¹Work on this paper began when I was a summer intern in the Financial Institutions Division of the IMF's Statistics Department. I would like to thank Marco Espinosa-Vega for his invaluable feedback and guidance and to Mindaugas Leika, Apostolos Panagiotopoulos, and Francisco Vazquez for excellent discussions. I am grateful to Meraj Allahrakha, Tadeusz Gałęza, Jesus Gonzalez-Garcia, Mahmut Kutlukaya, André Mialou, Samah Torchani, Giovanni Ugazio, and IMF seminar participants for their constructive comments and to Brian Bowling for the excellent editorial assistance.

CONTENTS

I. INTRODUCTION	4
II. LITERATURE REVIEW	5
III. DATA	6
A. Financial conditions data	6
B. Financial soundness indicators	9
IV. EMPIRICAL METHODOLOGY	10
A. Evaluation window	10
B. Accuracy measure	11
C. Generating signals	13
V. FINDINGS	16
VI. CONCLUSIONS	28
REFERENCES	29
APPENDIX	33
A. Data sources	33
B. Evolution of Financial Soundness Indicators	34
C. Single variable predictive accuracy in high-income economies (HIE) sample	42
D. Single variable predictive accuracy – LV crises dataset	43
E. Single variable predictive accuracy – LD crises dataset	48
F. Multivariate early warning models – additional crises datasets	50
BOX	
1 Financial Soundness Indicators	5
FIGURES	
1 Macro-financial linkages	4
2 Comparison of distress periods in the UK	8
3 Distribution of financial conditions indices in crisis and tranquil periods	9
4 Evaluation of signals	11
5 Receiver operating characteristic curve	13
6 Early warning signals for Chile based on tier 1 capital to assets ratio	18
7 ROC curve for signals based on tier 1 capital to assets ratio	19
8 Confusion matrices for tier 1 capital to assets ratio	21
9 Signals based on regulatory tier 1 capital ratio	23
10 ROC curves for multivariable model—FCIs, full sample	25
11 Confusion matrices for multivariate models	27
12 Regulatory capital to risk weighted assets	34
13 Regulatory tier 1 capital to risk weighted assets	35
14 Tier 1 capital to assets	36
15 Personnel expenses to non-interest expenses	37

16 Liquid assets to short term liabilities	38
17 Residential real estate loans to total loan	39
18 Return on equity	40
19 Interest margin to gross Income	41
20 ROC curve—LV crises dataset	50
21 ROC curve—LD crises dataset	50
22 Confusion matrices for LV crises, full sample	51
23 Confusion matrices for LD crises, full sample	52

TABLES

1 Prediction performance measured by AUC - FCIs, full sample	17
2 Prediction performance of FSIs measured by AUC - FCIs, full sample	26
3 Prediction performance measured by AUC - FCIs, HIE sample	42
4 Prediction performance measured by AUC – LV crises dataset, full sample	43
5 Prediction performance measured by AUC – LV crises dataset, HIE sample	44
6 Prediction performance measured by AUC – LV crises dataset, upper middle-income sample	45
7 Prediction performance measured by AUC – LV crises dataset, lower middle-income sample	46
8 Prediction performance measured by AUC – LV crises dataset, low-income sample	47
9 Prediction performance measured by AUC – LD crises dataset, full sample	48
10 Prediction performance measured by AUC – LD crises dataset, HIE sample	49
11 Prediction performance of FSIs measured by AUC - LV, full sample	53
12 Prediction performance of FSIs measured by AUC - LD, full sample	53

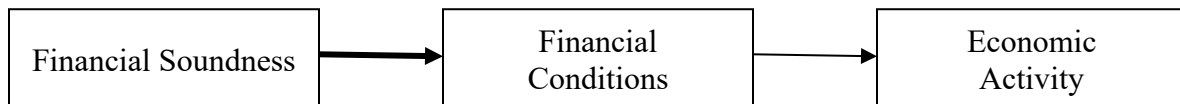
I. INTRODUCTION

Recessions associated with banking and financial crises are deeper and more protracted than those associated with other types of crises. In the short-term banking crises lead to output loss, decline in consumption and investment but there are also considerable long-term effects (Furceri and Mourougane, 2012). As documented by Laeven and Valencia (2018) median cumulative deviation of GDP from its trend in 3 years after banking crises is 35 percent for high-income economies (HIE) and 14 percent for low- and middle-income economies. On average banking crises tend to last relatively longer in advanced economies. Laeven and Valencia (2018) hypothesize that longer and more severe crises for HIE are due to the size of financial systems in these economies.

Financial or banking crises are usually associated with sharp tightening of financial conditions—higher-than-average borrowing costs, lower-than-average house price growth and lower-than-average credit growth (IMF, 2017). However, periods of tight financial conditions may or may not lead to crises. If they do, then recessions associated with financial crises are more pronounced than “normal” recessions (Jordà et al., 2011). Even when tight financial conditions do not result in a crisis, they still pose a challenge for policymakers by increasing downside risk of GDP growth (Adrian et al., 2019). Therefore, it is desirable to early detect tight financial conditions. Early enough, so that appropriate policy measures can be deployed well in advance of tightness.

Tight financial conditions can be accurately detected in 1 to 3 years ahead with the use of financial soundness indicators. We employ country-level data related to financial health and soundness and selected macroeconomic variables to generate early warning signals of tight financial conditions. Specifically, this paper uses financial indicators dataset: the IMF’s Financial Soundness Indicators (FSI) and macro variables, to investigate whether FSIs can offer reliable signals about worsening of financial conditions in the future. Box 1 provides a brief reference to the FSIs dataset. Left arrow in figure 1 illustrates the goal of this paper—can FSIs detect early enough tight financial conditions. Right arrow shows why this is worthwhile—tight financial conditions lead to lower economic activity whether they lead to a crisis or not.

Figure 1 Macro-financial linkages



To illustrate the paper’s strategy, consider one of the FSIs—the aggregate leverage (assets to Tier 1 capital) of banking sector. The goal is to find a threshold above which leverage issues a signal of tight financial conditions h -periods ahead. As illustrated in section V leverage exceeding 8.3 issues signals that 80 percent of times correctly detect tight financial conditions 4 to 12 quarters ahead.

Combining a few variables yields superior accuracy of early warning signals compared to single variable. In particular, signals based on FSIs are useful in early detection of

periods of tight financial conditions and in general in macro-financial surveillance. As a preview, it turns out that most accurate signals are issued when capital adequacy variables are low as well as earnings and profitability of deposit taker institutions are high. Capital adequacy variables dominate variables related to non-financial sector indebtedness in terms of signals accuracy. Authorities with financial stability mandates can benefit from deploying these FSIs in financial sector surveillance as a signal extraction tool for increased likelihood of tight financial conditions.

This paper contributes to the early warning literature in three ways: (i) it uses rarely used data with rich information about country-level financial soundness and health, (ii) it mostly uses recent data that is not subject to structural changes after Global Financial Crisis, and (iii) focuses on tight financial conditions rather than financial of banking crises.

The rest of the paper is structured as follows. Section II relates this paper to the early warning literature. Section III describes the data and transformations used in this paper. Section IV provides description of methods use to extract early warning signals from macroeconomic and financial variables. Section V presents the key results of the paper. Section VI concludes.

Box 1 Financial Soundness Indicators

IMF Financial Soundness Indicators give the detailed view of the current financial health and soundness of the financial institutions in a country, and of their corporate and household counterparts. FSIs include both individual institution data aggregated by country and indicators that are representative of the markets in which the financial institutions operate. FSI data cover almost up to 30 years and more than 100 countries. Apart from wide country coverage, FISs are compiled under internationally agreed methodology and they accurately measure internal and external financial sector risk exposure due to consolidation. FSIs are split into two sets of core and additional indicators. Core subset covers deposit taker institutions, while additional subset covers other financial institutions, non-financial private sector, and real estate market. See IMF (2019).

II. LITERATURE REVIEW

Pre-Global Financial Crisis (GFC), the early warning literature focused on macroeconomic variables—including real GDP growth, excessively high real interest rates, high inflation, real exchange rate—as the main leading indicators of crises. Demirgüç-Kunt and Detragiache (1998) show that banking crises are associated with low real GDP growth, excessively high real interest rates and high inflation. Note however that authors focus on contemporaneous relationship between set of predictors and banking crisis occurrence. Borio and Lowe (2002, 2004) conclude that the gap of credit to GDP from its long-run trend is the most useful early warning indicator of banking crises, followed by asset price index and real exchange rate.

The GFC called into question the usefulness of macroeconomic indicators for early warning purposes. In response, many authorities with financial stability mandates developed new analytical tools that help in early prediction of banking/financial crises (e.g.,

Juks and Melander, 2012; Antunes et al., 2014; Detken et al., 2014; Babecký et al., 2014; Lainà et al., 2015; Bańbuła and Pietrzak, 2017; Giordani et al., 2017; Holopainen and Sarlin, 2017; Joy et al., 2017; Deryugina and Ponomarenko, 2017; Aikman et al., 2018). Some of these papers focus entirely almost on macroeconomic and real estate variables, and a few considers banks' balance sheet information and the financial system health in general. Furthermore, these papers test the accuracy of signals against only one crises dataset which as discussed in section III.A can lead to wrong conclusions if the crises dataset is misspecified.

The GFC showed the need to consider variables like real estate prices and credit aggregates for early prediction of crises. The question remains which variables are predictors of tight financial conditions. Do structural changes post-GFC change the implications of early warning literature? Borio and Drehmann (2009) show that a combination of credit to GDP and property prices as well as equity market indices perform well in out-of-sample exercise in foreshadowing the GFC. Drehmann and Juselius (2012) construct the novel measure of aggregate debt service ratio (DSR) to proxy financial burden of non-financial private sector. DSR turns out to be an accurate predictor of banking sector distress at shorter horizon and can be a useful supplement for other early warning indicators. Detken et al. (2014) provide extensive analysis of early warning indicators underpinning European Systemic Risk Board policy recommendations. Holopainen and Sarlin (2017) document substantial difference in the predictive power across statistical methods.

III. DATA

This section describes the data used in the paper. Part III.A includes the description of dependent variable that serves as a proxy for tight financial conditions. Later, the reader can find a discussion on how this dependent variable compares to the approach used in the early warning literature. Part III.B describes FSIs and other macroeconomic and financial variables used in the estimation of early warning models—data sources and how these variables are transformed to extract additional information.

A. Financial conditions data

In this paper, financial conditions are proxied by the IMF (2017) financial conditions index (FCIs). Broadly speaking, FCIs measure the ease of financing in the economy based on variables such as short-term interest rates, credit growth, equity returns, asset price returns and spreads. FCIs time series are available for 43 economies from 1990 to 2016. Tight financial conditions are identified as higher-than-average corporate spreads, lower-than-average house price growth and lower-than-average credit growth and could lead to distress for the weakest, most exposed firms—with associated losses borne by banks, life insurers, mutual funds, pension funds, and overseas institutions (IMF, 2017). The other reason for choice of FCIs is that they are a predictor of future economic activity (Hatzius et al., 2010 and Koop and Korobilis, 2014) and are linked to the entire distribution of forecasted GDP growth, with the greatest impact on left tail of the distribution (Adrian et al., 2019). While there exist alternative financial conditions measures, they are: (i) not available for many countries and (ii) they are produced using differing methodologies which hinders comparability (example: ECB's Composite Indicator of Systemic Stress and Chicago Fed's

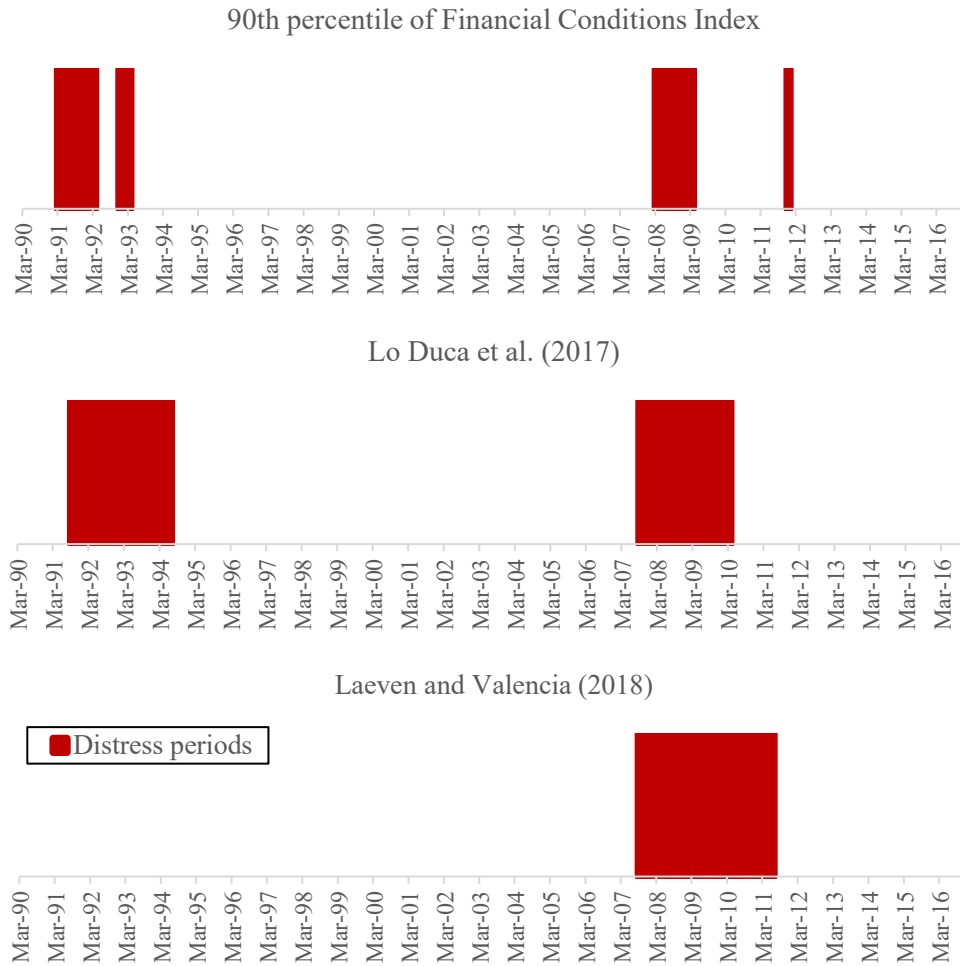
National Financial Conditions Index). To work with consistent measures across the countries, this paper used IMF's FCIs.

In line with the early warning literature, we construct a binary variable to track periods of financial distress. To this end, we transform a continuous FCI into a binary variable. Binary variable is set to 1 whenever FCI reaches historically high level, meaning that financial intermediation has deteriorated. The transformation threshold is set to 90th percentile of the country's FCI distribution. The rationale for this choice is a proportionality assumption that draws from the banking and financial crises literature: Laeven and Valencia (2018) and Lo Duca et al. (2017)²—find that 12.5 and 10.5 percent of sample are crises periods.³ Laeven and Valencia (2018) provide timing of systemic banking crises. On the other hand, Lo Duca et al. (2017) cover a broader measure, financial crises, though only for subset of European countries. The differences between the periods of tight financial conditions and two crisis databases are presented in Figure 2 for the UK as an example. There is overlap between FCI-based and Lo Duca et al. (2017) database for the periods of early 1990s and GFC, while Laeven and Valencia (2018) point out at the GFC as the only financial distress period in the UK.

² Hereafter, LV denotes database of Laeven and Valencia (2018), and LD database of Lo Duca et al. (2017).

³ There are other, well-known crises databases, that identify the timing of banking or financial crises (see Caprio et al., 2005; Reinhart and Rogoff, 2009) however the choice of datasets for this study is limited to these who are up to date.

Figure 2 Comparison of distress periods in the UK

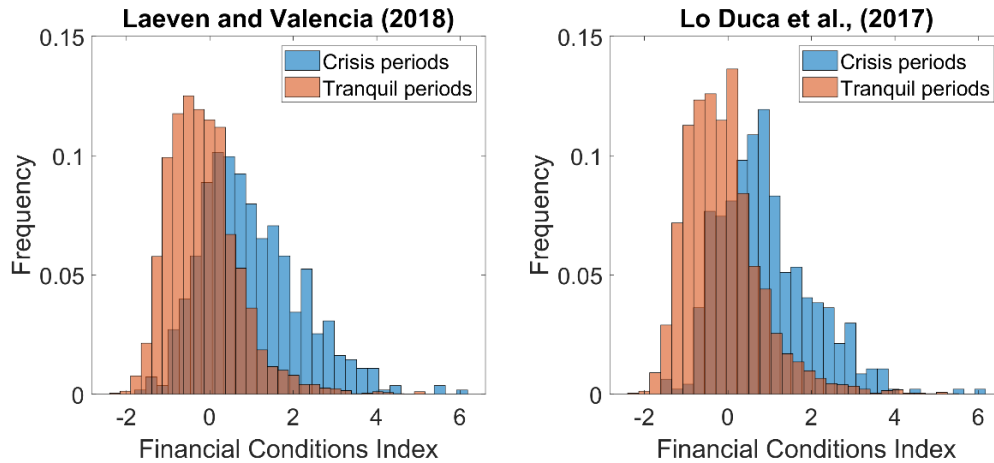


Note: Figure shows the discrepancy in the period of financial distress in the UK as an example. Top panel shows the periods of tight financial conditions, which are calculated based on whether the financial conditions index exceeds 90th percentile. Middle panel is based on Lo Duca et al. (2017) dataset, while bottom panel is based on Laeven and Valencia (2018) dataset.

The choice of two alternative crises datasets is meant to compare the usefulness of early warning signals based on FSIs with the literature that has been predominantly focused on crises. The other reason is the robustness of the results. Chaudron and de Haan (2014) reinvestigate the banking crises timing by complementing crises datasets with the data on the number and size of bank failures and conclude that Laeven and Valencia (2013) is the dataset that most closely aligns with their augmented crises timing. Therefore, we choose an updated version (Laeven and Valencia, 2018). Its main advantage is a wide coverage of countries which allows to test the predictive accuracy of early warning models not only for advanced economies and covers years 1970-2017. In this paper only banking crises are used from the set of three types of crises: currency, sovereign debt, and banking. Another dataset used is Lo Duca et al. (2017). It covers European countries from 1970 until 2016. The main advantage of this dataset is combining quantitative criteria (financial stress indices) for crisis classification with expert judgment of relevant European and national authorities of countries in the sample. It covers all periods of significantly heightened systemic stress in financial

system. Figure 3 shows the difference between all three notions of distress periods. As mentioned earlier crises are associated with tightening of financial conditions, though there are periods of elevated cost of borrowing which are not necessarily crises.

Figure 3 Distribution of Financial Conditions Indices in crisis and tranquil periods



Note: Left panel presents distributions of Financial Conditions Index for crisis and tranquil periods according to Laeven and Valencia (2018) crisis dataset, while right panel depicts analogous distributions based on Lo Duca et al. (2017) dataset.

B. Financial soundness indicators

The variables used to generate early warning signals use wide range of information about financial system health as proxied by FSI data. FSI data are complemented with a broad set of macroeconomic and financial variables⁴ covering years 1990 (when possible) to 2019. While some of these indicators have been studied in the literature (see section II) previous studies have not focused on FSIs and often used variables compiled according to different methodologies. In this study we draw on FSIs for a large set of reporting countries according to the same methodology. There are two sets of FSIs: core and additional. Core subset covers deposit taker institutions, while additional subset covers deposit taker institutions, other financial institutions, non-financial private sector, and real estate market. The evolution of selected FSIs is presented in appendix B.

To increase the information set available, all variables have been subject to several transformations. These transformations include one-quarter and one-year change of level (that is differences) and/or growth rate⁵ to capture the momentum. Another transformation is

⁴ Rate of inflation, real GDP, industrial production, debt service cost, current account, and financial account. Financial market prices and additional data from International Financial Statistics (IFS) were further included into the set of predictors.

⁵ For instance, change in level was applied to interest rate, since growth rate of an interest rate would not be meaningful. Same logic applies to GDP, where growth rate is calculated but not the change in level as this transformation would not be stationary.

detrending using one-sided Hodrick-Prescott filter⁶ as it is often done in early warning literature. Cyclical deviation, i.e. the difference between the level and the long-term trend, is computed for all variables except these which relate to growth rates, interest rates or spreads. The smoothing parameter is set to $\lambda = 1600$ since the data are at quarterly interval (Ravn and Uhlig, 2002). The last transformation is country-specific z-score which main benefit is that the thresholds that must be exceeded to issue signals are country-specific. The complete list of predictors can be found in appendix A.

IV. EMPIRICAL METHODOLOGY

This section describes how early warning signals are generated by different statistical techniques and how the accuracy of the signals is evaluated. First, we show how we extract the signals, second how we measure their accuracy, third we describe statistical methods that are used to extract the signals from the underlying variables.

A. Evaluation window

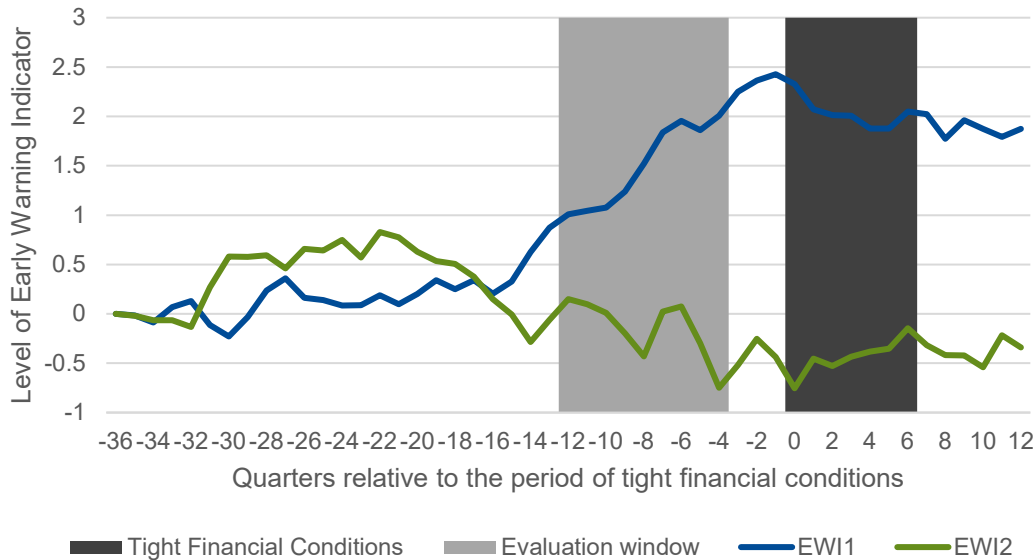
This paper provides early signals about the tightening of financial conditions from 4 to 12 quarters ahead. Optimal early warning signals are meant to be accurate, timely, stable, and interpretable (Drehmann and Juselius, 2014). Accuracy of signals is measured using ratio of true positive signals to false positive signals for different thresholds. Regarding timing, signals should give policy makers enough time to design measures that could mitigate systemic stress. Most often, evaluation horizon spans from 12-16 quarters before a distress starts until 4 quarters before its start. Lower bound of this timespan is dictated by the fact that some prudential measures need at least a year from a decision to operationalization and full implementation by regulated institutions. For instance, this applies to Countercyclical Capital Buffer. Therefore, in this study the evolution horizon starts 4 quarters before a deterioration in financial system occurs. The end of that time frame is set to 12 quarters. Stability-wise, increasing evaluation horizon from 12 to 16 quarters does not significantly impact the results, although it makes most of the signals less accurate as it is more challenging to issue correct signals 4 years in advance of tightening of financial conditions. In order to satisfy interpretability requirement, each method used in this paper yields the probability that the financial conditions worsen in horizon of 4 to 12 quarters.

Figure 4 illustrates the approach to evaluate the usefulness of the signals. The black area denotes the period of tight financial conditions, and the grey shaded area shows the preceding 4 to 12 quarters evaluation window. Suppose we use a binary choice model in which the dependent variable is 1 or 0. In such case dependent variable in the grey area has value of 1 meaning that in 4 to 12 quarters ahead there will be a distress in the financial system. Now suppose we have two hypothetical early warning indicators – say leverage (EWI1) and interbank rate spread (EWI2). Hypothetical early warning indicator EWI1 seems to be a

⁶ It is well-known that standard – two-sided – version of HP filter is sensitive to adding new data (Mise et al., 2005), therefore deviation from trend at time t estimated using data up to time t (real-time information) will be different from the deviation estimated using the data up to say $t + 10$. The standard remedy for that is the use of one-sided version of HP filter that for every period uses the data up to this period. Therefore, one-sided HP filter is based on the information that was available at a particular point in time.

better predictor of tight financial conditions as its value increases in the period preceding black area and is relatively low before the tightness period. In fact, for hypothetical threshold of 1.0 EW1 issues a signal exactly during the grey pre-distress period while the other variable, EW2, does not send signals. EW2 does not reach high levels hence it is not a useful predictor. When threshold is higher than both EW1 and EW2 issue less signals while for lower threshold they send more signals. Choice of thresholds is discussed in section IV.B.

Figure 4 Evaluation of signals



Note: This figure depicts evolution of two hypothetical Early Warning Indicators EW1 and EW2 before the period of tight financial conditions. Both variables are standardized by mean and standard deviation.

The last step in the preparation of the data is accounting for so-called postcrisis and crisis bias. Bussiere and Fratzscher (2006) suggest that crisis and postcrisis periods do not provide useful insights about the accuracy of signals since these periods are neither pre-crisis periods nor normal times. Given the structural changes that are ongoing during crisis and postcrisis periods this part of sample does not tell much about the transition from normal to pre-crisis state. Therefore, these periods should be discarded. In line with literature, we assume that two years after the crisis ends, system is back to normal. Observations for that periods are discarded as well. Crises database compiled by Lo Duca et al. (2017) does include “system back to normal” dates therefore in the case of this dataset, the actual length of structural changes is accounted for.

B. Accuracy measure

Signals generated by early warning indicators in both tranquil and pre-crisis periods need to be evaluated with regard to accuracy. One approach to the measurement of accuracy is based on noise to signal ratio. In this framework the best indicator minimizes noise to signal ratio subject to correct prediction of fixed fraction of crises. This fixed percentage is often set to 60-75 percent (see Borio and Lowe, 2002; Borio and Drehmann, 2009). Another approach is construction of loss function and usefulness of signals (see Alessi

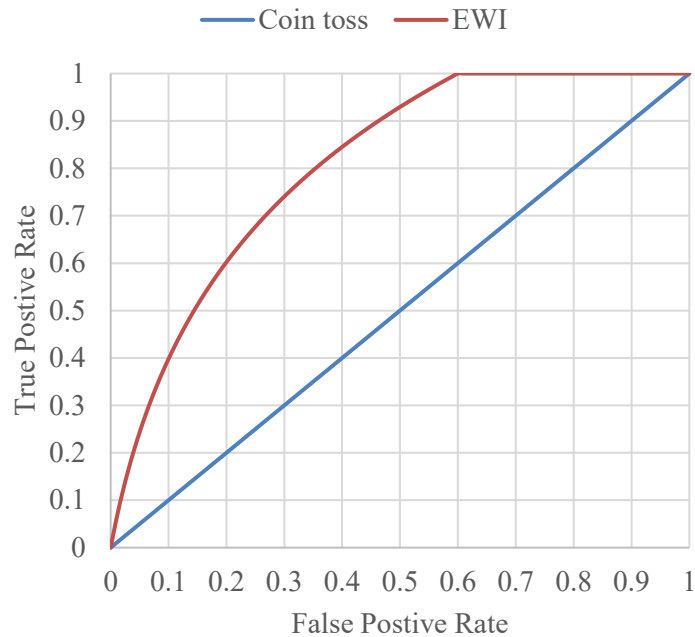
and Detken, 2011; Bussiere and Fratzscher, 2008). The main drawback of loss function is the need to know the relative preferences of policy makers regarding the cost of issuing false alarm and missing a crisis. Third way of evaluation of accuracy is by Receiver Operating Characteristic (ROC) curve that shows the tradeoff between True Positive Rate (TPR)—share of correctly predicted crises—and False Positive Rate (FPR)—share of false signals—for all possible values of thresholds.

Figure 5 illustrates relationship between FPR and TPR for hypothetical early warning indicator and—as a benchmark—random variable with uniform distribution at the interval from 0 to 1. This random variable would result in the signals that are random as well and would point out tight financial conditions in half of the cases. Setting threshold to the highest possible—one—would mean that these random signals do not signal any tightening of financial conditions. This means TPR and FPR would be both zero. For the lowest possible threshold—zero—TPR and FPR are both one. As the threshold goes down, more signals are issued and both TPR and FPR increase. For the random signals TPR and FPR increase linearly. The relationship between threshold and FPR and TPR looks different for a hypothetical early warning indicator—red line in Figure 5. To summarize, the lower the threshold that early warning indicator must exceed to issue a signal, the more crises are predicted (higher TPR). At the same time lower threshold means that the number of false alarms increases (higher FPR).

The trade-off of avoiding false alarms and missing a tightness periods can be summarized by the Area Under the ROC Curve (AUC). For the benchmark—random variable—AUC equals 0.5. For the hypothetical EWI it equals 0.8. The closer is AUC to 1, the more accurate are the signals issued by early warning indicator. The only exception regarding AUC interpretation occurs when signal extraction method is used (see next section). Then accurate predictor can have AUC level either close to 1 or close to 0. If it is close to 1 it means that a predictor is increasing before a crisis (for instance this applies to banks' return on equity), while when it is close to 0 it means that a predictor decreases before a crisis (for instance Tier 1 capital to assets ratio).

The AUC may lead overoptimistic picture of signals accuracy if the tightness periods are rare as is the case in these data. Binary dependent variable takes value of 1 only in 10% of cases. This means that the signals that perfectly predict tranquil periods and always miss the tightness periods will have high AUC. Therefore, high value of AUC is not informative if there is a high imbalance (a lot of tranquil periods and a few tightness periods). To correct for the imbalance, we apply Synthetic Minority Over-Sampling Technique (Chawla et al., 2002) and generate additional “tightness” observations. Number of synthetically generated observation is such that the number of tranquil and tightness periods is equal.

Figure 5 Receiver Operating Characteristic curve



Note: This figure depicts different combinations of True Positive Rate and False Positive Rate based on hypothetical signals from Early Warning Indicators. Blue line depicts a benchmark.

A remaining issue is the choice of a threshold above which signal is generated. As mentioned earlier, the true preferences regarding tolerance for type I vs. type II error are not known. However, thresholds can be decided for relative cost of missing a true positive (excessively tight financial conditions) to cost of issuing false signal. Let $\tau \in (0,1)$ be the cost of missing true positive and $(1 - \tau)$ be the cost of issuing a false alarm. Then relative cost of missing a crisis to cost of false signal is given by $\tau/(1 - \tau)$. The way to choose the optimal threshold is to choose a combination of TPR and FPR on ROC curve that is tangent to the inverse of this relative cost—that is $(1 - \tau)/\tau$. The higher the cost of missing a crisis τ , or more risk-averse is decision maker, the flatter the slope of the line and therefore the optimal point on the ROC curve will be more in the upper right region of FPR/TPR space.

C. Generating signals

This paper draws on eight methods that generate signals warning about impending tight financial conditions with a 4-12 quarters lead. The previous two sections describe how to evaluate the signals that are issued when a certain threshold is exceeded by a variable. There are multiple ways to generate the signals. The rationale behind using several methods is that due to different samples and different types of vulnerabilities there is no one method that generates signals that outperform others in every possible scenario. Therefore, rather than choosing one method it is better to compare different approaches without a priori assuming that some of them are inferior. It might be the case that one method underperforms in one sample (for instance for advanced economies) while outperforms other methods for other sample (for example developing economies). This paper uses eight methods, and their description follows. More formal description of these methods can be found in Hastie et al.,

(2005). Note that except signal extraction, all methods yield probability of tight financial conditions in a horizon of 4 to 12 quarters. Therefore, AUC for the best signals will be high. For signal extraction AUC level of a good signals will be either close to 1 (if high level of predictor issues signals about tightness) or close to 0 (if low level of predictor issues signals about tightness).

Signal extraction (SigExt)

This method issues a binary signal $s = 1$ whenever a level of analyzed predictor exceeds a threshold. Otherwise, signal is not issued. The optimal threshold is found using grid search. That is the list of thresholds is tested and then the one that achieves the highest performance is used. The performance can be defined as minimization of ratio of false alarms to correct alarms conditional on ability to predict at least prespecified share of tightness periods. The main disadvantage of signal extraction is the problem of analyzing multiple indicators at the same time as it increases the size of grid search exponentially. In this paper signal extraction is used for univariate early warning models.

Decision tree (Tree)

This method can be related to a sequential signal extraction. Root node, which is the first node, represents the entire data. Root node is then split into two, more homogenous, sub-nodes based on a rule which is the level of one of the analyzed variables. The process of partitioning using decision rules based on analyzed variables is continued until subsets become more homogenous. The size of a decision tree is controlled by the node size, which is set using grid search of exponentially spaced values from 1 through 100. Then I compute the cross-validated out-of-sample error. Optimal tree node size is chosen such that this tree is smallest possible within one standard error of the minimum cross-validated out-of-sample error on the grid of node sizes.

Discriminant analysis

Quadratic discriminant analysis (QDA) assumes that probability density function of predictors conditional on tranquil/crisis state are normally distributed with some means and covariance matrices. In linear discriminant analysis (LDA) covariance matrices are the same for both conditional probability density functions. The signal is issued when log ratio of “tightness” probability density function to “tranquil” probability density function exceeds a certain threshold. In this paper both linear and quadratic discrimination methods are used. For DA with more than one predictor model is regularized by grid search over γ and δ hyperparameters.

Support Vector Machine (SVM)

In linear version of this method a set of hyperplanes is constructed to find a good separation between classes. In case of this paper SVM constructs hyperplanes that separate two subsets of predictors – one is tranquil ($s = 0$) subset and the other is a “tightness” subset ($s = 1$). In practice it is not possible to linearly separate the data, therefore non-linear versions of SVM

are used. In this paper SVM with Gaussian kernel with a tuning parameter γ picked from the grid of values between 0.1 and 1 such that it minimizes cross-validated out-of-sample prediction error.

k-Nearest Neighbors (KNN)

In kNN regression the output is the probability of a tight financial conditions computed as the average values of dependent binary variables of k neighbors. The neighbors are the observations which have most similar features compared to the datapoint in question. This similarity is proxied by the Euclidean distance between the datapoint and all other points. k points with the smallest distance from the datapoint constitute k -neighborhood of that point. A free parameter of this method is the value of k . There are some useful rules of thumb for choosing k . In this paper k is rounded square root of sample size.

Naïve Bayes (NB)

This classification method consists of constructing probability density functions of predictors conditional on independent variable outcome that is tranquil ($s = 0$) or tightness period ($s = 1$). This conditional probability density functions are multiplied by prior probability of tranquil state and tightness state which are equal to fraction of these observations in the sample. Then the predicted value is chosen using maximum a posteriori decision rule which consists of choosing the prediction that is more probable through the lens of likelihood of joint probability density function of predictors.

Logistic Regression (LR)

The last method is the logistic regression without fixed effects. While fixed effects would increase the model fit, they would drive probability of excessive tightness to zero for countries that have never experienced it.⁷ Therefore, logistic regression model is estimated on a pooled data that was previously standardized with a country-specific z-score.

Variables generating signals that are significantly better than a coin toss are fed into the multivariate models to further enhance the accuracy of signals. Such variables generated signals whose AUC confidence interval is not overlapping with the value of 0.5. Another criterion for variable selection is maximization of overlap between these variables as missing observations are ignored.⁸ This procedure leads to choosing 19 variables.⁹ Many variables feature high degree of correlation. For instance, Regulatory Capital to Risk Weighted Assets

⁷ That pertains to datasets of Laeven and Valencia (2018) and Lo Duca et al. (2017).

⁸ Using all available variables would lead to only 556 periods in the sample with only 37 tightness periods.

⁹ Regulatory Capital to Risk Weighted Assets, Regulatory Tier 1 Capital to Risk Weighted Assets, Interest Margin to Gross Income, Non-interest Expenses to Gross Income, Return on Assets, Return on Equity, Liquid Assets to Short Term Liabilities, Liquid Assets to Total Assets Liquid Asset Ratio, Net Open Position in Foreign Exchange to Capital, Tier 1 Capital to Assets, Customer Deposits to Total Non-interbank Loans, Foreign Currency Denominated Liabilities to Total Liabilities, Foreign Currency Denominated Loans to Total Loans, Gross Asset Position in Financial Derivatives to Capital, Gross Liability Position in Financial Derivatives to Capital, Personnel Expenses to Non-interest Expenses, Spread Between Reference Lending and Deposit Rates, Trading Income to Total Income, Residential Real Estate Loans to Total Loans

and Regulatory Tier 1 Capital to Risk Weighted Assets have a correlation coefficient of 0.88, while Return on Assets and Return on Equity have a correlation coefficient of 0.85. To speed up estimation these 19 variables are subject to feature selection via Principal Component Analysis. The number of principal components that is fed into classification methods is chosen such that these components explain 95% of variance in these 19 variables. This gives 6 principal components that are used in the multivariate models. The only method that is not used for combining multiple variables is signal extraction as it requires extensive grid search that increases exponentially with number of predictors.¹⁰ To prevent overfitting, k-fold cross-validation with 10 folds is used in every repetition. The cross-validation is repeated 5 times. Results reported in tables are averages from all repetitions. Contrary to the standard out-of-sample exercises that are used in the literature, cross-validation allows for estimation and testing of models using full sample.

V. FINDINGS

Signals generated by FSIs have high accuracy relative to other macro-financial variables. The most accurate signals of tight financial conditions are based on capital adequacy variables. Accuracy of signals is measured against binary transformation of FCI, as discussed in section III.A. Table 1 presents AUC performance measures for eight different statistical methods based on sample of countries for which FCIs exist (see note to Table 1 for sample details). Each variable in this table generated signals that have AUC significantly different from a random choice (which would have AUC=0.5). To better understand AUC values in this table, we analyze the best-performing variable, Tier 1 Capital to Assets ratio,¹¹ for a Chile as an example due to a long history of reporting this variable. Top left panel of Figure 6 shows the raw data. How does it translate into early warning signals? Other panels depict probabilities of tightening based on that variable. Note that all methods yield probabilities that have negative correlation with Tier 1 Capital to Assets ratio, though probability series have different properties. For instance, Tree and KNN generate signals that feature spikes, while other methods seem to generate more smooth series. Horizontal dashed lines are the thresholds. Exceeding a threshold generates a signal. The lower the threshold the more tightness periods are correctly predicted, though at the cost of issuing more false positives—signals about the tightening that does not take a place.

¹⁰ This does not change the quality of signals as signal extraction performs rather poorly in multivariate case (see Holopainen and Sarlin, 2017).

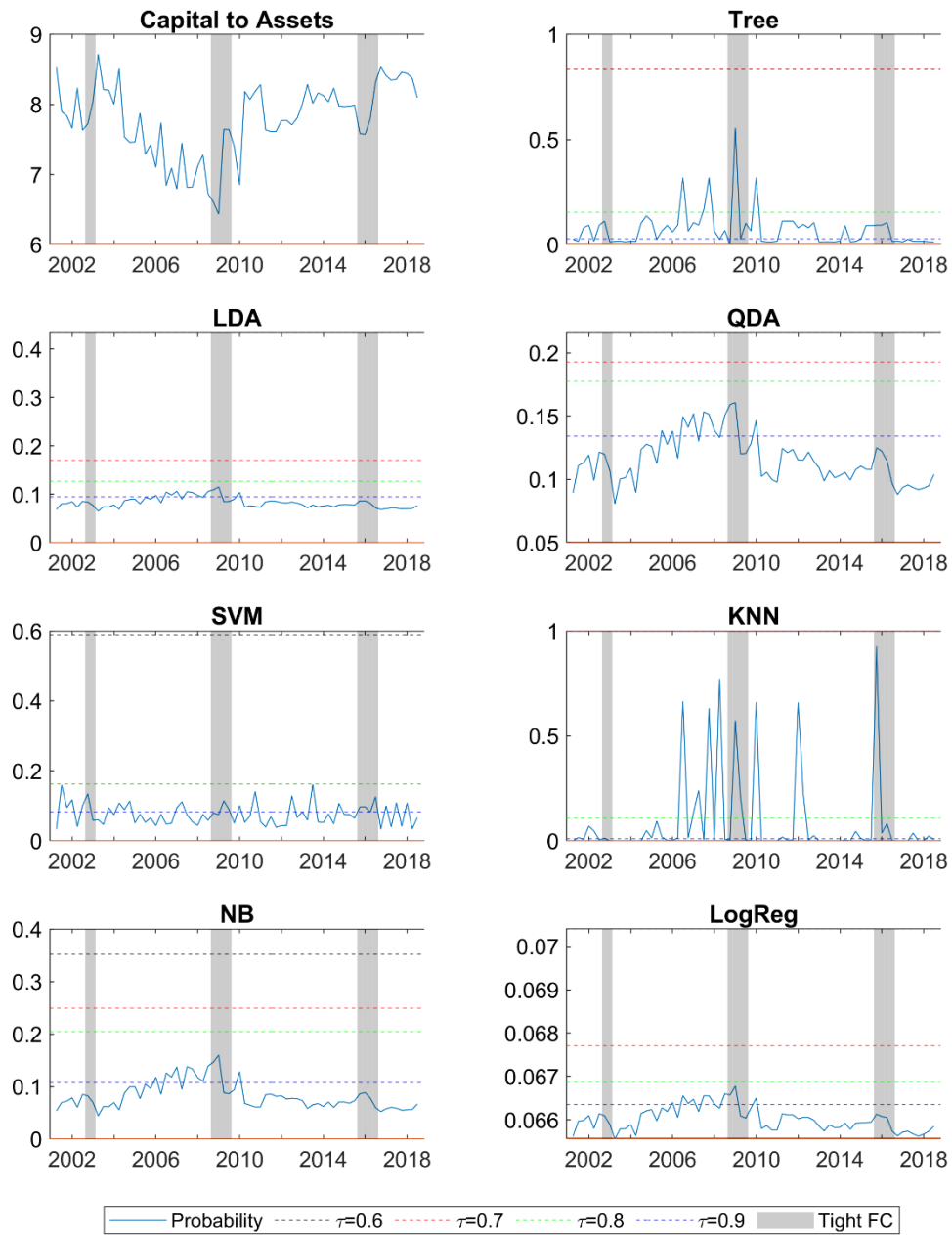
¹¹ For countries that have implemented Basel III, this variable is calculated using the new Basel III leverage ratio. The capital measure for the numerator is Tier 1 capital. The denominator comprises a new Basel III aggregate called “exposure” consists of all balance sheet assets and off-balance sheet exposures. For other countries, this FSI is calculated by using Tier 1 capital as numerator, and total balance sheet assets—without risk weighting—as denominator. See page 87 in IMF (2019).

Table 1 Prediction performance measured by AUC - FCIs, full sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Tier 1 Capital to Assets raw	0.227	0.836	0.772	0.771	0.768	0.830	0.790	0.772	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.222	0.852	0.778	0.813	0.778	0.866	0.847	0.778	Yes
Regulatory Capital to Risk Weighted Assets raw	0.212	0.854	0.788	0.787	0.787	0.854	0.803	0.788	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets z	0.249	0.820	0.750	0.750	0.769	0.830	0.792	0.750	Yes
Regulatory Capital to Risk Weighted Assets z	0.279	0.779	0.721	0.721	0.721	0.778	0.738	0.720	Yes
Nonbank Financial Institutions Claims on Priv Sect z	0.689	0.819	0.686	0.686	0.684	0.787	0.743	0.686	No
Banking Institutions Foreign Liabilities z	0.682	0.770	0.682	0.691	0.702	0.741	0.719	0.682	No
Residential Real Estate Loans to Total Loans raw	0.637	0.856	0.635	0.635	0.638	0.892	0.733	0.635	Yes
Banking Institutions Claims on Priv Sect z	0.671	0.797	0.670	0.682	0.680	0.756	0.731	0.670	No
Net Open Position in Foreign Exchange to Capital c4	0.590	0.783	0.589	0.751	0.775	0.806	0.776	0.588	Yes
Personnel Expenses to Non-interest Expenses raw	0.645	0.793	0.644	0.649	0.660	0.795	0.753	0.645	Yes
DSR hp	0.674	0.749	0.674	0.675	0.684	0.730	0.702	0.674	No
Return on Equity z	0.666	0.761	0.666	0.665	0.670	0.747	0.709	0.665	Yes
Foreign Currency Denominated Liabilities to Total Liabilities r	0.637	0.811	0.637	0.635	0.660	0.818	0.690	0.636	Yes
Foreign Currency Denominated Loans to Total Loans raw	0.618	0.785	0.617	0.680	0.681	0.789	0.713	0.615	Yes
Spread Between Reference Lending and Deposit Rates z	0.664	0.726	0.663	0.661	0.665	0.751	0.685	0.663	Yes
Residential Real Estate Loans to Total Loans z	0.338	0.735	0.661	0.659	0.661	0.755	0.686	0.661	Yes
Liquid Assets to Short Term Liabilities hp	0.355	0.767	0.645	0.573	0.647	0.781	0.702	0.642	Yes
Net Open Position in Foreign Exchange to Capital z	0.374	0.765	0.625	0.628	0.626	0.785	0.692	0.626	Yes
Tier 1 Capital to Assets z	0.341	0.697	0.658	0.657	0.659	0.729	0.655	0.658	Yes

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSIs. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: ARG, AUS, AUT, BEL, BGR, BRA, CAN, CHL, CHN, COL, CZE, DEU, DNK, ESP, FIN, FRA, GBR, GRC, HUN, IDN, IND, IRL, ISR, ITA, JPN, KOR, MEX, MYS, NLD, NOR, PER, PHL, POL, PRT, RUS, SWE, THA, TUR, USA, VNM, ZAF.

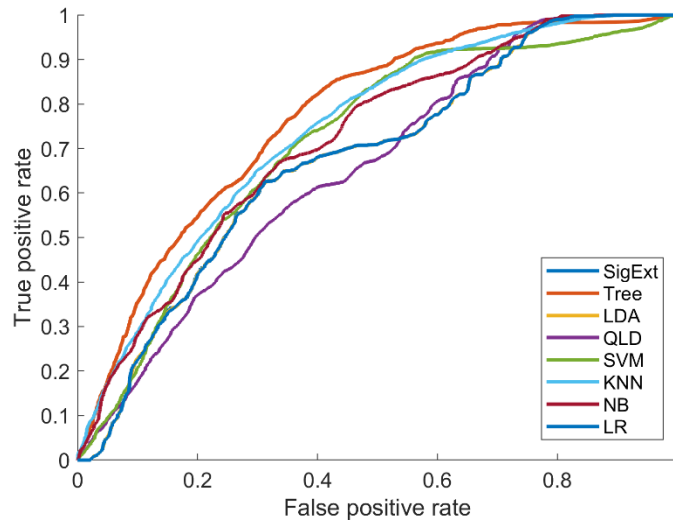
Figure 6 Early warning signals for Chile based on Tier 1 Capital to Assets ratio



Note: Top left panel shows the Tier 1 Capital to Assets time series, while other panels show probability of tight financial conditions in 1 to 3 years. Dashed lines show thresholds—for different τ —above which signal is issued. Grey areas show periods of tight financial conditions.

High AUC means that on average signals have high accuracy—correctly classify tranquil and tightness periods. Figure 7 shows ROC curves for signals generated based on Tier 1 Capital to Assets ratio—the best performing predictor of future tightness. Each curve shows the trade-off that is faced when choosing a threshold. High threshold implies low TPR and FPR, while low threshold the opposite. To better understand these trade-offs, we analyze confusion matrices in Figure 8.

Figure 7 ROC curve for signals based on Tier 1 Capital to Assets ratio



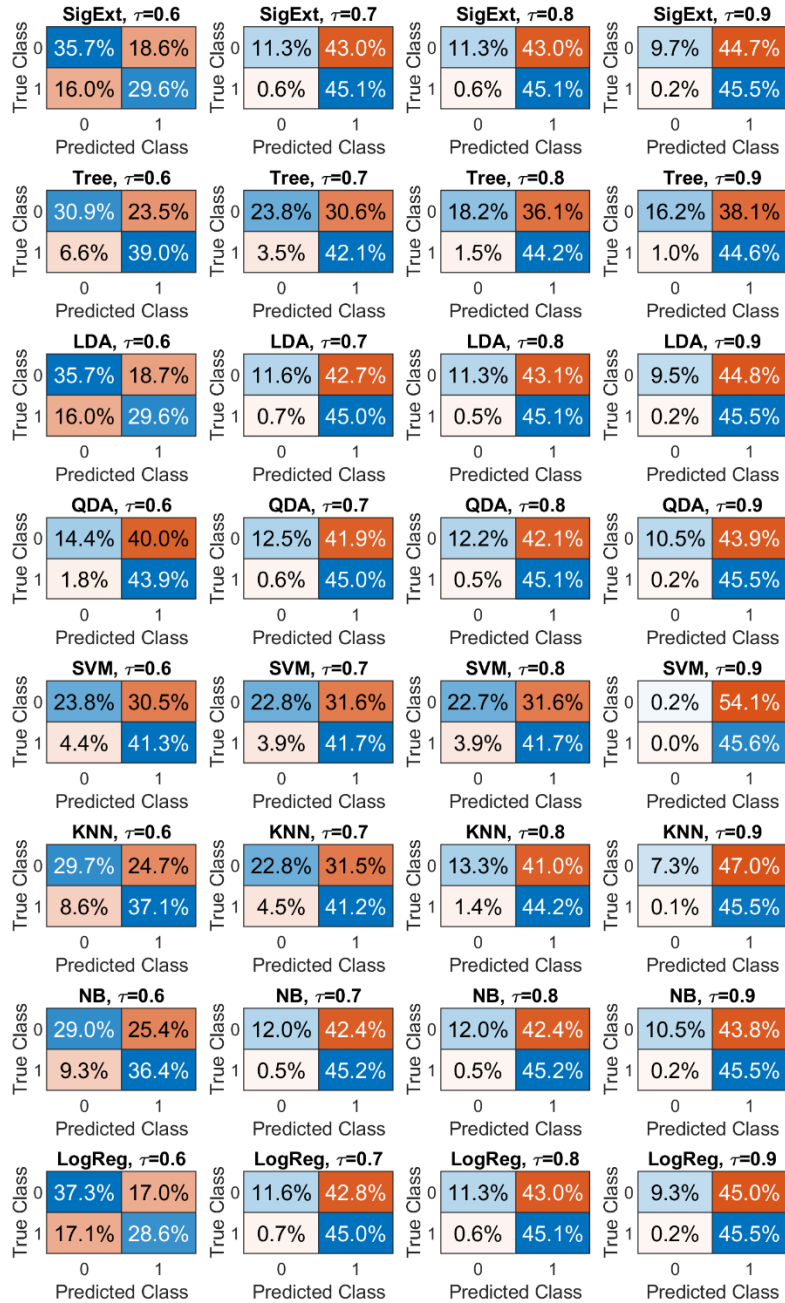
High AUC value is not informative of the trade-offs that a policymaker faces when choosing a threshold. Tight financial conditions do not emerge frequently, hence the data are imbalanced with mostly tranquil periods. It means that signals that predict perfectly tranquil periods but do not predict tightness periods can still have high AUC given that there are a few tightness periods. The imbalance of tranquil and tightness periods is addressed by synthetically generating tightness periods (as discussed in section IV.B)

Figure 8 shows confusion matrices for different thresholds. Each row shows matrices for the same method but across different τ ¹² (tightness-aversity), while each column shows matrices for the same τ but across different methods. For instance, panels in second row show that for lowest τ , decision tree generates signals that correctly classify 30.9 percent of true negatives and 39.0 percent of true positives. 30.1 percent of observations is misclassified. As we increase τ , policymaker becomes more “tightness-averse”, and the threshold is lower. The same signals more often exceed the lower threshold. Lower threshold (high τ) implies more false alarms. For instance, top right panel of Figure 8 shows that the same signals correctly classify 16.2 percent of true negatives and 44.6 percent of true positives. Now 39.1 percent of observations is misclassified. Note that lowering the threshold implies more false negatives. For $\tau=0.6$, 23.5 percent observations are classified as signals of future tightness while they

¹² For a specific value of τ , the relative cost of not detecting tightness is $\tau / (1 - \tau)$, meaning that missing a tightness is $\tau / (1 - \tau)$ more harmful than generating a false alarm.

are not. For $\tau=0.9$, this number totals to 38.1 percent. Applying a forward-looking policy actions when the threshold is low implies many of these actions are not necessary.

Figure 8 Confusion matrices for Tier 1 Capital to Assets ratio



As

summarized in Table 1, most accurate signals are generated by FSIs related to capital

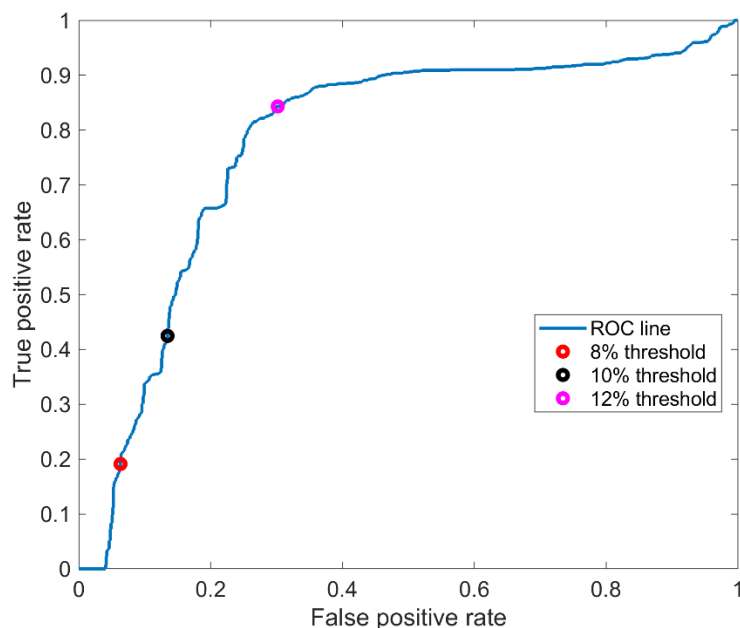
adequacy, followed by real estate markets and earnings and profitability¹³ variables. Among non-FSIs, the best predictors include banks and non-banks claims on private non-financial sector to GDP as well as Debt Service Ratio (DSR) which turned out to be most accurate predictors in the early warning literature. Surprisingly, these variables are inferior to many FSIs. Appendix C reports the results for a subsample of HIE based on World Bank countries classification by income level.¹⁴ The results in this subsample are consistent with the findings based on full sample. Note however that AUC values on the subsample are on average higher compared to full sample which resonates with suggestion of clustering the countries to achieve higher accuracy (Van den Berg et al., 2008).

Authorities in charge of financial sector surveillance can benefit by monitoring capital adequacy since it generates reliable signals of tight financial conditions. One of the interesting results of the univariate signaling exercise reported in Table 1 is the predictive power of regulatory capital ratios—contrary to the findings of Aikman et al. (2014) where regulatory capital ratios were found to be not useful predictors. Figure 9 charts TPR against FPR of signals extracted from Tier 1 capital. In line with that paper, we assess 8 and 10 percent thresholds to be non-informative whereas threshold set at 12 percent, correctly classifies impending tight financial conditions—TPR is 84 percent while FPR 30 percent. What does explain the discrepancy? This paper uses country level data, not a bank level data. Secondly, the goal of this paper is to early detect tightening of financial conditions not individual bank failure. Note also that cited paper focused on a sample of large global banks. As discussed in part III.A tightened financial conditions are not equivalent to bank failures. This result shows that, contrary to microprudential supervision of global banks, authorities in charge of financial sector surveillance can benefit by monitoring capital adequacy.

¹³ This FSI is subject to different methods of calculation since it is a ratio of a flow (income) to a stock (assets). The income annualization choice can be found in the FSI metadata. The denominator should be the average of the stock of total assets during the reporting period. At a minimum, the denominator can be calculated by using the average of the beginning and end-period positions. See page 90 in IMF (2019).

¹⁴ Low-income economies are defined as those with a GNI per capita, calculated using the World Bank Atlas method, of \$1,025 or less in 2018; lower middle-income economies are those with a GNI per capita between \$1,026 and \$3,995; upper middle-income economies are those with a GNI per capita between \$3,996 and \$12,375; high-income economies are those with a GNI per capita of \$12,376 or more.

Figure 9 Signals based on Regulatory Tier 1 Capital ratio



The main message remains unchanged when we switch attention from the full sample to the sample consisting of HIE. Table 3 in appendix C shows that the results for HIE sample are comparable with the results for the full sample. Again, variables related to capital adequacy are generating most accurate signals. This is not surprising given that the FCI data are available mostly for advanced economies.¹⁵ While the most accurate predictors remain the same, most of them feature even higher accuracy in this subsample, meaning that they are especially suitable for macro-financial surveillance in HIE. Higher values of AUCs suggest that it is optimal to estimate early warning models on a subset of homogenous countries. In this case, homogeneity is achieved by grouping countries according to per capita income level. Predictive accuracy of these indicators in other subsamples of countries cannot be established due to scarcity of the FCI data, which are rarely available for countries with GNI per capita less than \$12,376.

Apart from predicting tight financial conditions, FSIs generate signals that early detect both banking crises and financial crises. These types of crises are what the early warning literature (section II) focuses on. Appendix D shows predictive accuracy of single indicators for Laeven and Valencia (2018) crises dataset. Note that this crises dataset consists of banking crises, therefore signals provide early warning about imminent banking crises. Table 4 reports the results for full sample. Again, FSIs are useful in terms of generating accurate signals in advance. Categories of FSIs which are particularly useful are related to real estate markets, earnings and profitability, capital adequacy and sensitivity to market risk. Apart from FSIs, housing price to rent and price to income ratios provide useful signals. How does the usefulness of indicators look like in subsamples of countries? Table 5 shows the best

¹⁵ In fact, advanced economies compose 80 percent of the full sample in case when dependent variable is FCI.

indicators in sample of HIE. FSIs play prominent role in providing early warning signals about banking crises. Capital adequacy ratios as well as earnings and profitability indicators seem to be two categories of variables that generate most accurate signals. For upper middle-income countries (Table 6) set of best predictors includes real estate markets, interest rates on government bonds, financial institutions claims on private non-financial sector and real economy variables. For lower middle-income countries (Table 7) the most accurate signals are generated by indicators related to capital adequacy, foreign currency denominated debt, deposits to loan relation and earnings and profitability measures. For low-income economies (Table 8) best predictors include interest rates of government bonds, foreign liabilities of banking institutions and current account. The second alternative proxy for financial system distress is a financial crisis. Appendix E presents predictive accuracy of single indicators evaluated against Lo Duca et al. (2017) dataset of financial crises. In this case results are not tested for different groups of countries as this crisis dataset covers subset of European countries only, therefore there are not many countries other than HIE. Table 9 shows that FSIs are again top predictors of financial tightness. Among them, proxies for capital adequacy and earnings and profitability of deposit takers are most accurate. Other useful variables are related to real estate market and private non-financial sector indebtedness.

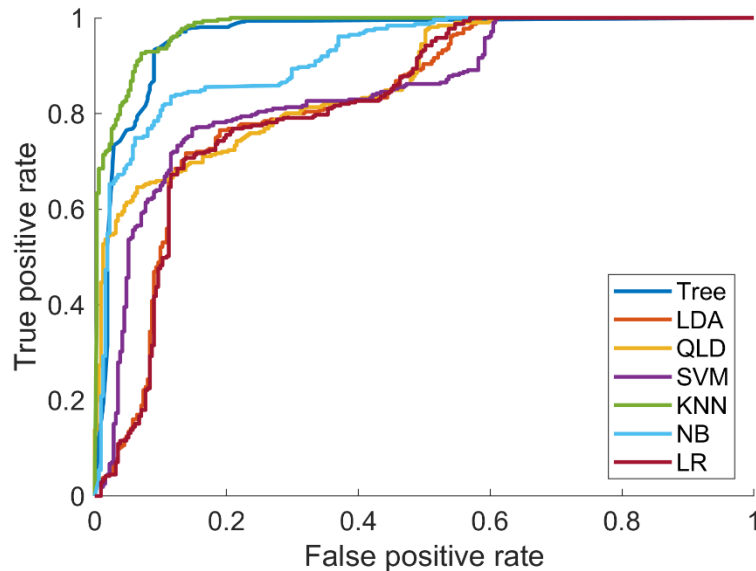
To sum up, the main finding is a strong empirical support for using FSIs as an additional data source for macro-financial surveillance. FSIs are overall helpful in early detection of tight financial conditions. The most accurate signals are generated by FSIs related to capital adequacy and earnings and profitability. Surprisingly, variables related to indebtedness of non-financial sector offer less accurate signals than the signals based on FSIs. In this section we focused on single variables as a source of signals. It seems natural to ask to what extent accuracy of signals can be enhanced by combining a few variables. Confusion matrices for capital Tier 1 Capital to Assets shows that even the best signals yield high number of false positive and false negatives. In the next subsection we show that using multivariate models to extract signals leads to significant improvements overall. This can be easily seen by comparing confusion matrices of univariate vs multivariate models.

Multivariate models

Extracting early warning signals based on models with a few variables yields superior accuracy. In the previous section single indicators were evaluated in terms of predictive accuracy of signals they issue—in this section multiple variables are combined in a model. Section IV.C explains which variables are chosen. To reiterate, only variables with AUC of signals which were higher than 0.5 are chosen. Another criterion for variable selection is minimization of missing observations. If an observation for a variable is missing it means that all other variables for this time period and country are ignored. This leads to choosing 19 variables (see footnote 9). Since many variables are highly correlated, we use Principal Component Analysis to reduce the data size and to speed up estimation. First six principal components explain 95% of variance in the data and these are the components used in the estimation. As in the previous section, due to model uncertainty, there are seven different methods used. The predictive accuracy is evaluated against benchmark—FCIs binary variable showing when financial conditions are excessively tight. The estimation and cross-

validation processes are the same as before. The final output of each method is a cross-validated probability of excessively tight financial conditions occurring in 4 to 12 quarters.

Figure 10 ROC curves for multivariable model—FCIs, full sample



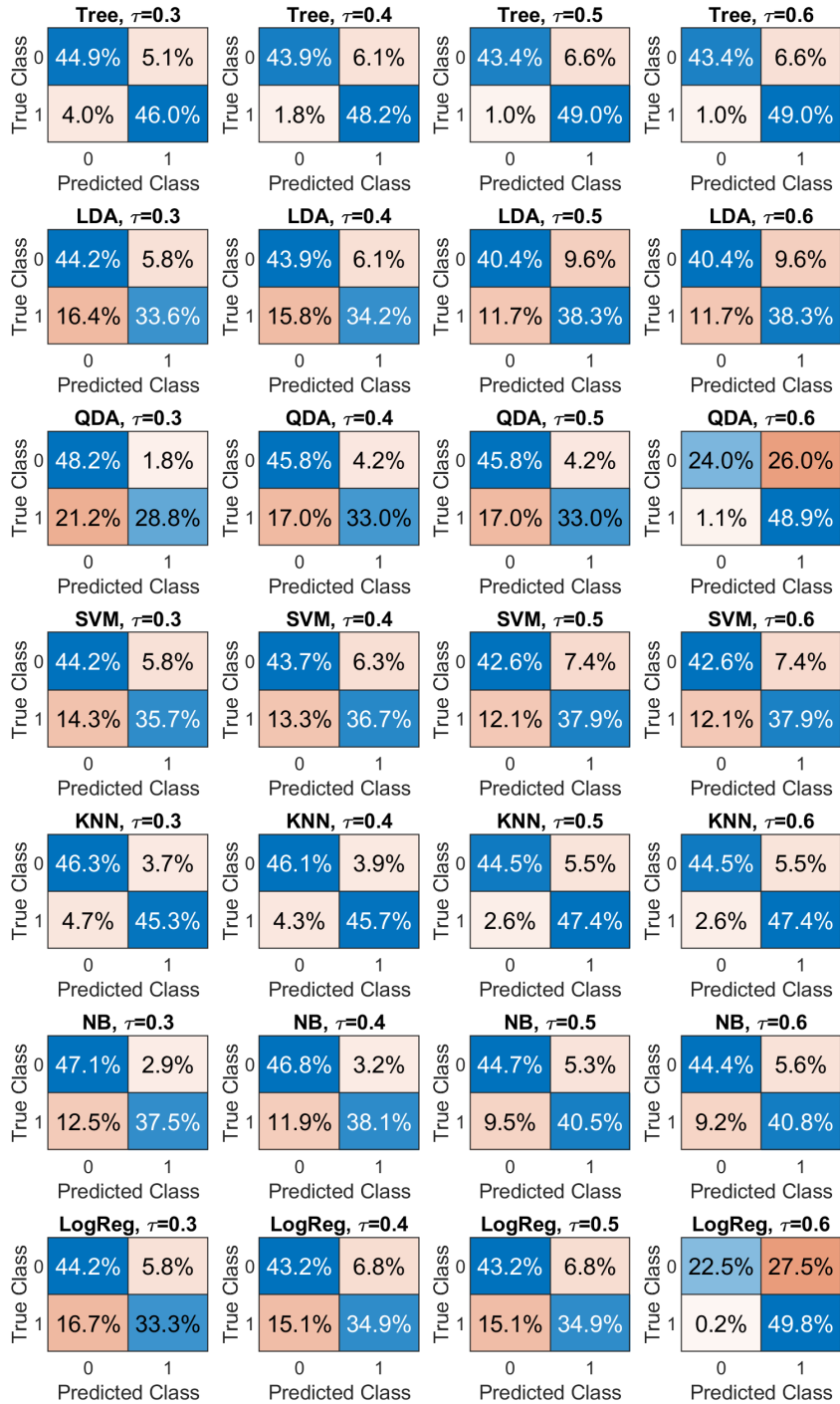
KNN and Tree dominate other methods in terms of signals accuracy as shown in Figure 10 which depicts feasible combinations of FPR and TPR based on the probabilities generated by seven methods. Note that at some point lowering the thresholds does not yield a benefit in terms of capturing more tightness periods. In fact, KNN and Tree can correctly classify most of the true positives with relatively low false positive rate of 20 percent. Figure 11 relates the ROC curves to the confusion matrices for specific τ . KNN and Tree methods feature high accuracy for any thresholds. The ratio of false positives and false negatives increase modestly as τ increases. Table 2 includes the details of the results presented in Figure 10. AUC values confirm the analysis of ROC curves. Interestingly, KNN is the best performing methods which coincides with the results of Holopainen and Sarlin (2017). Note that AUC values for both indicators are not significantly different however analysis of ROC curves shows that their performance differs in different regions of FPR. As shown in Figure 10 KNN clearly dominates Tree for low levels of FPR. Logistic regression and linear discriminant analysis generate least accurate signals. This is again in line with their performance for a different crisis dataset in Holopainen and Sarlin (2017). Surprisingly, most of the early warning literature uses logistic regression as a main tool to extract signals. It seems that this method is inferior to classification methods.

Table 2 Prediction performance of FSIs measured by AUC - FCIs, full sample

	τ	0.30			0.4			0.5			0.6		
	Relative cost	0.43			0.67			1			1.5		
Method	AUC	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR
Tree	0.96	0.75	0.10	0.94	0.29	0.12	0.97	0.12	0.14	0.98	0.12	0.14	0.98
LDA	0.83	0.67	0.12	0.68	0.65	0.12	0.69	0.54	0.19	0.77	0.54	0.19	0.77
QDA	0.87	0.93	0.04	0.58	0.92	0.08	0.66	0.92	0.08	0.66	0.60	0.52	0.98
SVM	0.85	0.63	0.12	0.72	0.58	0.13	0.74	0.54	0.15	0.76	0.54	0.15	0.76
KNN	0.97	0.67	0.07	0.91	0.67	0.08	0.92	0.54	0.11	0.95	0.54	0.11	0.95
NB	0.92	0.74	0.06	0.75	0.72	0.06	0.77	0.61	0.11	0.81	0.60	0.11	0.82
LR	0.83	0.63	0.12	0.67	0.59	0.14	0.70	0.59	0.14	0.70	0.31	0.55	1.00

Note: The table reports cross-validated out-of-sample performance measured by AUC for seven methods and evaluation horizon of 4-12 quarters. The second column shows Area Under the ROC curve. Next three columns report optimal threshold (Tr) in terms of probability from the model and False Positive Rate (FPR) and True Positive Rate (TPR) associated with the optimal threshold. The optimal threshold is chosen according to the cost of missing a crisis (τ). Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: ARG, AUS, AUT, BEL, BGR, BRA, CAN, CHL, CHN, COL, CZE, DEU, DNK, ESP, FIN, FRA, GBR, GRC, HUN, IDN, IND, IRL, ISR, ITA, JPN, KOR, MEX, MYS, NLD, NOR, PER, PHL, POL, PRT, RUS, SWE, THA, TUR, USA, VNM, ZAF.

Figure 11 Confusion matrices for multivariate models



VI. CONCLUSIONS

Early warning literature focuses on financial and banking crises. However, tight financial conditions—whether they lead to a crisis or not—still pose a risk to economic growth. Therefore, it is important to be able to early detect impending tightening financial conditions such that some policy measures can be implemented well in advance. So far, there exists no policy guidance whether tight financial conditions are detectable and if yes, then which macro-financial variables should be monitored to track the risk of tightening.

This paper showcased that Financial Soundness Indicators can be successfully used for macro-financial surveillance. Signals based on the FSIs offer accurate and early detection of financial sector distress as proxied by tightening financial conditions. Indicators that should be closely monitored are related to capital adequacy and earnings and profitability of deposit taker institutions. To high leverage and returns on assets and equity issue early warnings that vulnerabilities are building up in the financial sector.

Generating signals based on single variables is not enough as the accuracy of these signals is inferior compared to the multivariate models-based signals. Such signals offer superior accuracy of detecting tight financial conditions 1 to 3 years ahead. However, the accuracy depends on the statistical methods used to generate signals. For instance, widely used logistic regression seems to be a suboptimal choice for task of detecting tightening of financial conditions. These results show that using FSIs can be beneficial in conducting macro-financial surveillance as they can signal early enough tightness of financial conditions which ultimately may lead to lower and more volatile economic growth.

REFERENCES

- Adrian, T. (2017). Macroprudential policy and financial vulnerabilities. In Speech at the European Systemic Risk Board Annual Conference, September (Vol. 22).
- Adrian, T. and F. Duarte (2018): "Financial Vulnerability and Monetary Policy," CEPR Discussion Papers 12680, C.E.P.R. Discussion Papers.
- Adrian, T., & Liang, N. (2018). Monetary Policy, Financial Conditions, and Financial Stability. *International Journal of Central Banking*.
- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4), 1263-89.
- Aikman, D., Bridges, J., Burgess, S., Galletly, R., Levina, I., O'Neill, C., & Varadi, A. (2018). Measuring risks to UK financial stability (No. 738). Bank of England.
- Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., ... & Neumann, T. (2014). Taking uncertainty seriously: simplicity versus complexity in financial regulation. Bank of England financial stability paper, (28).
- Aikman, D., Kiley, M., Lee, S. J., Palumbo, M. G., & Warusawitharana, M. (2017). Mapping heat in the US financial system. *Journal of Banking & Finance*, 81, 36-64.
- Alessi, L., & Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3), 520-533.
- Antunes, A., Bonfim, D., Monteiro, N., & Rodrigues, P. (2014). Early warning indicators of banking crises: exploring new data and tools. *Economic Bulletin Banco de Portugal April*.
- Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., & Vašíček, B. (2014). Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, 15, 1-17.
- Bańbuła, P., & Pietrzak, M. (2017). Early warning models of banking crises applicable to non-crisis countries (No. 257). Narodowy Bank Polski, Economic Research Department.
- Bemanke, B., Gertler, M., & Gilchrist, S. (1996). The Financial Accelerator and the Flight to Quality. *The Review of Economics and Statistics*, 78(1), 1-15.
- Borio, C. E., & Drehmann, M. (2009). Assessing the risk of banking crises—revisited. *BIS Quarterly Review*, March.

Borio, C. E., & Lowe, P. (2004). Securing sustainable price stability: should credit come back from the wilderness? (No. 157). Bank for International Settlements.

Borio, C., & Lowe, P. (2002). Assessing the risk of banking crises. *BIS Quarterly Review*, 7(1), 43-54.

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

Bussiere, M., & Fratzscher, M. (2008). Low probability, high impact: Policy making and extreme events. *Journal of Policy Modeling*, 30(1), 111-121.

Caprio, G., Klingebiel, D., Laeven, L., & Noguera, G. (2005) Appendix: Banking crisis database”, published in. P. Honohan, L. Laeven (Eds.), *Systemic Financial Crises: Containment and Resolution*, Cambridge University Press, Cambridge, 307-340.

Chaudron, R., & de Haan, J. (2014). Dating banking crises using incidence and size of bank failures: Four crises reconsidered. *Journal of Financial Stability*, 15, 63-75.

Cihák, M. M., & Schaeck, K. (2007). How well do aggregate bank ratios identify banking problems? (No. 7-275). International Monetary Fund.

Costa Navajas, M., & Thegeya, A. (2013). Financial soundness indicators and banking crises (No. 13-263). International Monetary Fund.

Demirgüç-Kunt, A., & Detragiache, E. (1998). The Determinants of Banking Crises in Developing and Developed Countries. *Staff Papers-International Monetary Fund*, 81-109.

Deryugina, E., & Ponomarenko, A. (2019). Determination of the Current Phase of the Credit Cycle in Emerging Markets. *Russian Journal of Money and Finance*, 78(2), 28-42.

Detken, C., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M.M., Castro, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N. and Kakes, J., (2014). Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options (No. 5). ESRB Occasional Paper Series.

Drehmann, M., & Juselius, M. (2012). Do debt service costs affect macroeconomic and financial stability?. *BIS Quarterly Review* September.

Drehmann, M., & Juselius, M. (2014). Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3), 759-780.

- Furceri, D., & Mourougane, A. (2012). The effect of financial crises on potential output: New empirical evidence from OECD countries. *Journal of Macroeconomics*, 34(3), 822-832.
- Giordani, P., Spector, E., & Zhang, X. (2017). A new early warning indicator of financial fragility in Sweden. *Economic Commentaries*, (1), 1-17.
- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
- Hatzius, J., Hooper, P., Mishkin, F. S., Schoenholtz, K. L., & Watson, M. W. (2010). Financial conditions indexes: A fresh look after the financial crisis (No. w16150). National Bureau of Economic Research.
- Holopainen, M., & Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12), 1933-1963.
- International Monetary Fund. 2017. *Global Financial Stability Report: Getting the Policy Mix Right*. Washington, DC.
- International Monetary Fund. 2019. *Financial Soundness Indicators: Compilation Guide*. Washington, DC.
- Jordà, Ò., Schularick, M., & Taylor, A. M. (2011). Financial crises, credit booms, and external imbalances: 140 years of lessons. *IMF Economic Review*, 59(2), 340-378.
- Joy, M., Rusnák, M., Šmídková, K., & Vašíček, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics*, 22(1), 44-67.
- Kaminsky, G., Lizondo, S., & Reinhart, C. M. (1998). Leading indicators of currency crises. *IMF Staff Papers*, 45(1), 1-48.
- Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71, 101-116.
- Laeven, L., & Valencia, F. (2013). Systemic banking crises database. *IMF Economic Review*, 61(2), 225-270.
- Laeven, M. L., & Valencia, M. F. (2018). Systemic banking crises revisited. *International Monetary Fund*.
- Lainà, P., Nyholm, J., & Sarlin, P. (2015). Leading indicators of systemic banking crises: Finland in a panel of EU countries. *Review of Financial Economics*, 24, 18-35.

- Lo Duca, M., Koban, A., Basten, M., Bengtsson, E., Klaus, B., Kusmierczyk, P., Lang, J. H., Detken, C., & Peltonen, T. A. (2017). A new database for financial crises in European countries: ECB/ESRB EU crises database. No. 194. ECB Occasional Paper, 2017.
- Mise, E., Kim, T. H., & Newbold, P. (2005). On suboptimality of the Hodrick–Prescott filter at time series endpoints. *Journal of Macroeconomics*, 27(1), 53-67.
- Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2), 371-376.
- Reinhart, C. M., & Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly*. Princeton University Press.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Van den Berg, J., Candelon, B., & Urbain, J. P. (2008). A cautious note on the use of panel models to predict financial crises. *Economics Letters*, 101(1), 80-83.

APPENDIX

A. Data sources

International Monetary Fund

- Regulatory Capital to Risk Weighted Assets.
- Regulatory Tier 1 Capital to Risk Weighted Assets.
- Interest Margin to Gross Income.
- Non-interest Expenses to Gross Income.
- Return on Assets, Return on Equity.
- Liquid Assets to Short Term Liabilities.
- Liquid Assets to Total Assets Liquid Asset Ratio.
- Net Open Position in Foreign Exchange to Capital.
- Tier 1 Capital to Assets.
- Customer Deposits to Total Non-interbank Loans.
- Foreign Currency Denominated Liabilities to Total Liabilities.
- Gross Asset Position in Financial Derivatives to Capital.
- Gross Liability Position in Financial Derivatives to Capital.
- Large Exposures to Capital.
- Net Open Position in Equities to Capital.
- Personnel Expenses to Non-interest Expenses.
- Spread Between Highest and Lowest Interbank Rate.
- Spread Between Reference Lending and Deposit Rates.
- Trading Income to Total Income.
- Household Debt to Gross Domestic Product GDP.
- Assets to Gross Domestic Product GDP.
- Assets to Total Financial System Assets.
- Commercial Real Estate Loans to Total Loans.
- Residential Real Estate Loans to Total Loans.
- Residential Real Estate Prices.
- Current Account.
- Net Acquisition of Financial Assets.
- Index of Equities.
- Consumer Price Index.
- Portfolio Investment.
- Government Bonds Interest Rates.
- Treasury Bills Interest Rates.
- Real Gross Domestic Product.
- Real Industrial Production.

Bank for International Settlements

- Price To Rent.
- Price To Income.
- Housing Rent.
- Real House Prices.
- Debt Service Ratio.

B. Evolution of Financial Soundness Indicators

This appendix contains figures depicting the evolution of the variables that turned out to give accurate signals about the future financial distress. Each variable is depicted in four panels and for two income groups according to the World Bank classification. Top-left panel shows how a specific variable evolved over time (median). Top-right panel shows how many countries have reported the variable in a given time. Bottom panels show box plot with 25th, 50th and 75th percentiles of a variable in a specific time period for a group of countries. The whiskers extend to the most extreme observation that are not outliers while red dots show the outliers.

Figure 12 Regulatory capital to risk weighted assets

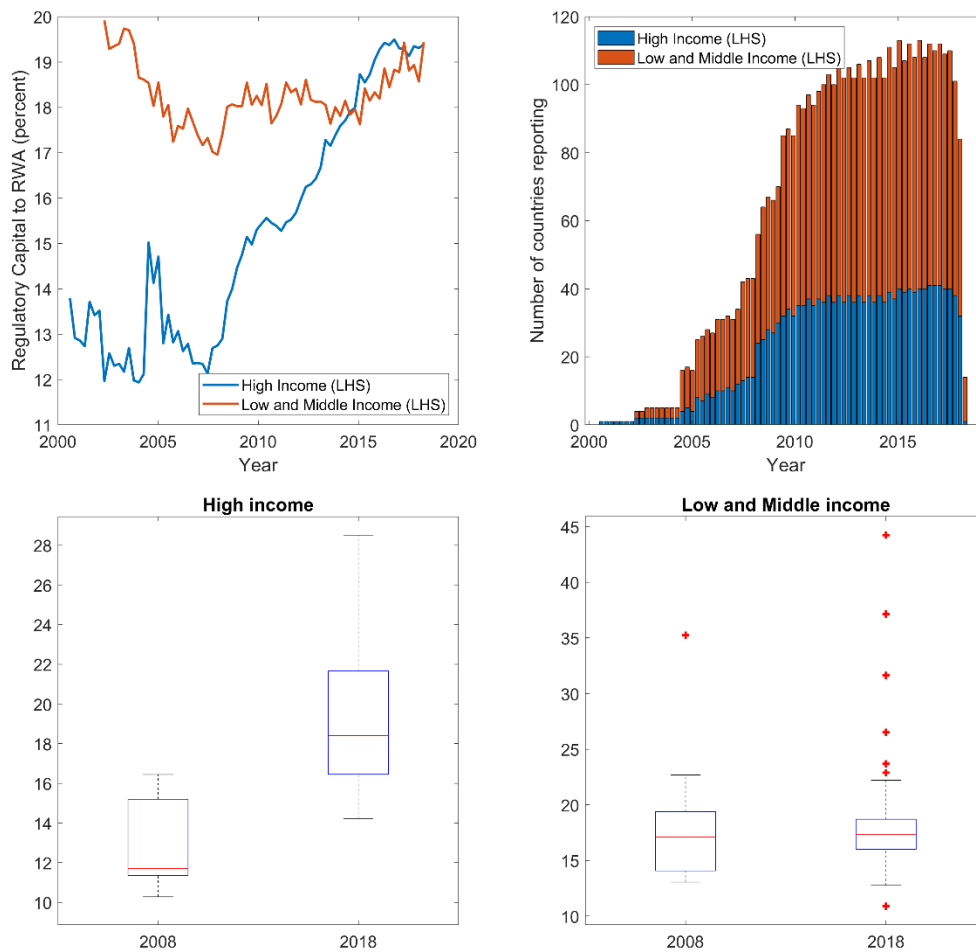


Figure 13 Regulatory Tier 1 Capital to risk weighted assets

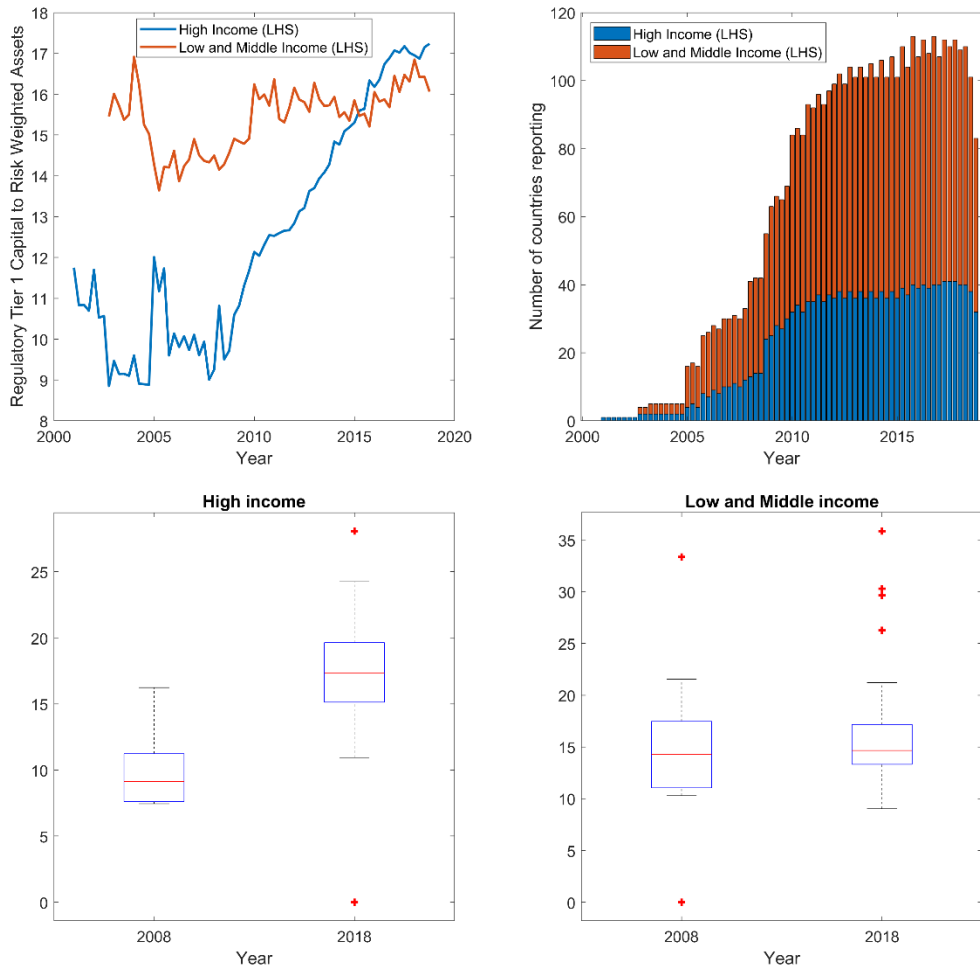


Figure 14 Tier 1 Capital to assets

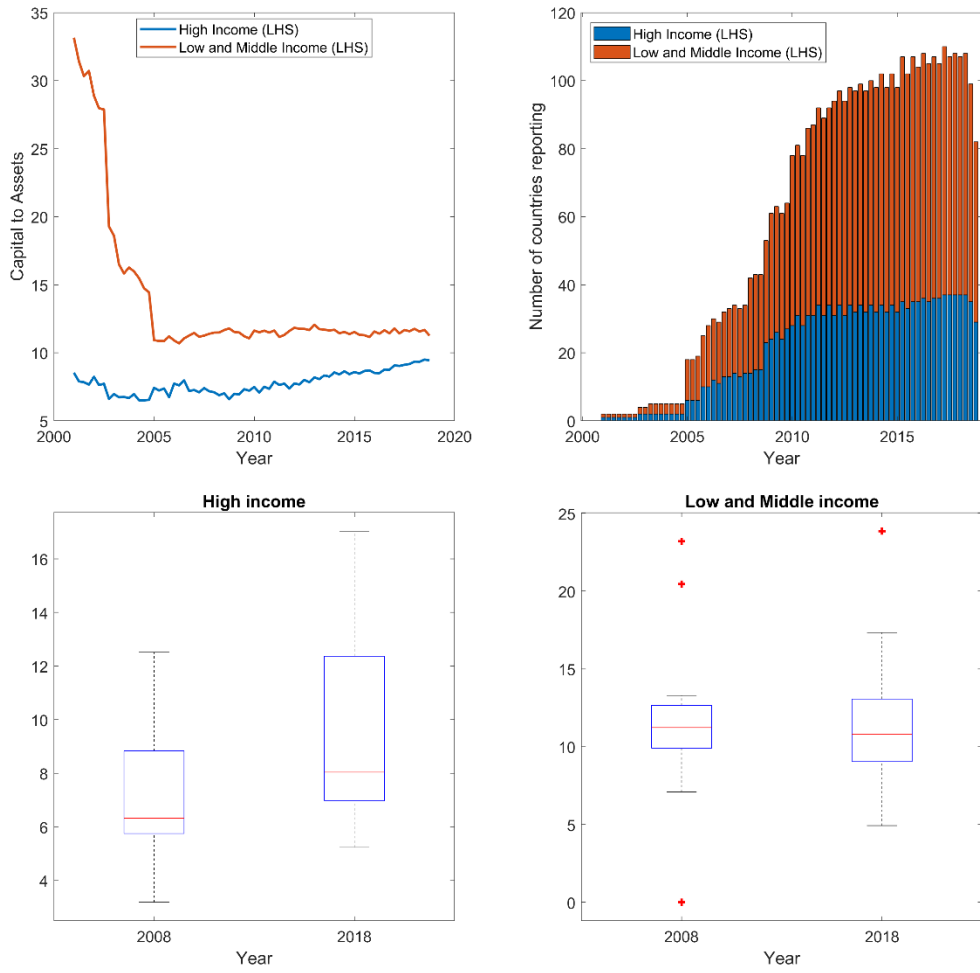


Figure 15 Personnel Expenses to Non-interest Expenses

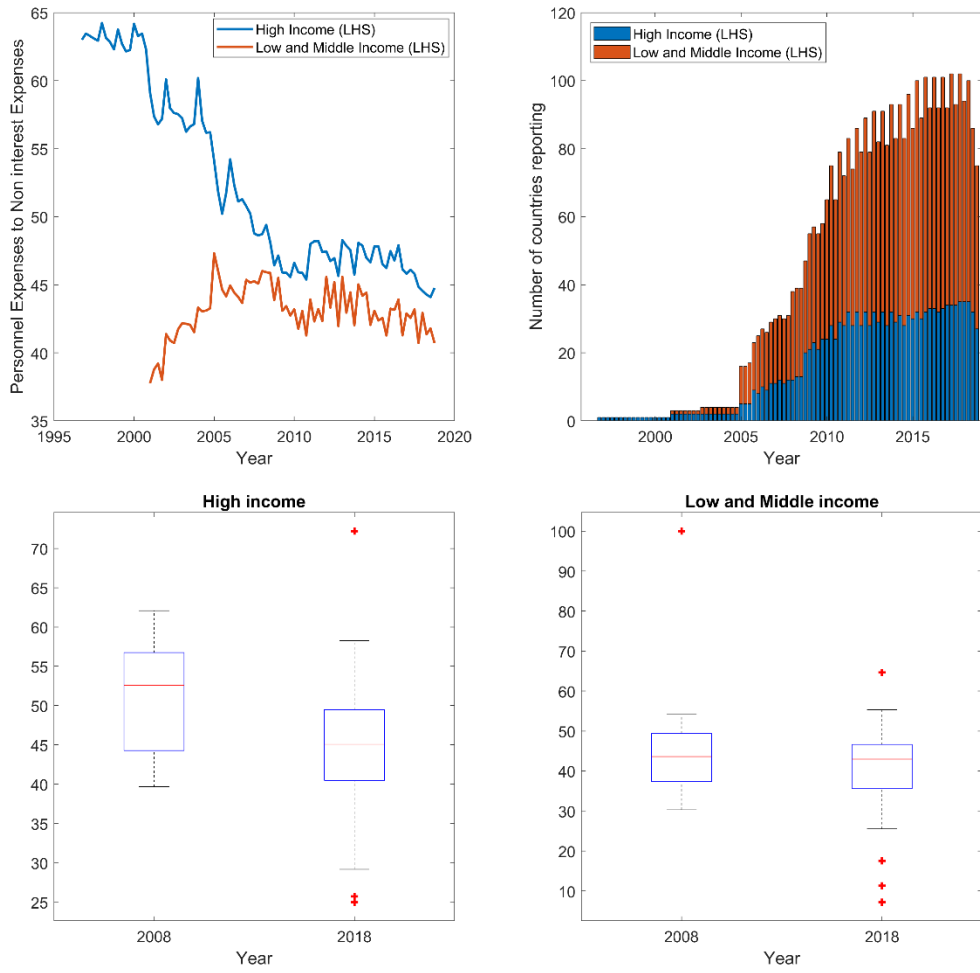


Figure 16 Liquid Assets to Short Term Liabilities

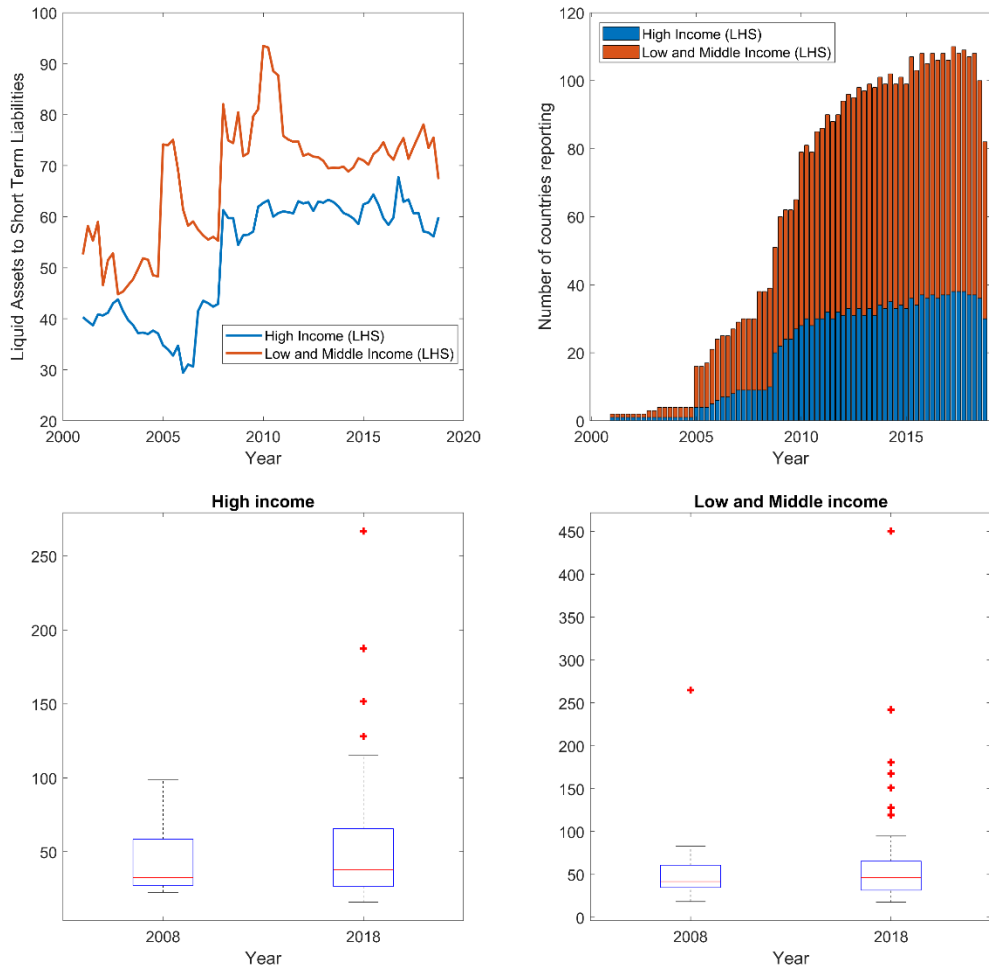


Figure 17 Residential Real Estate Loans to Total Loans

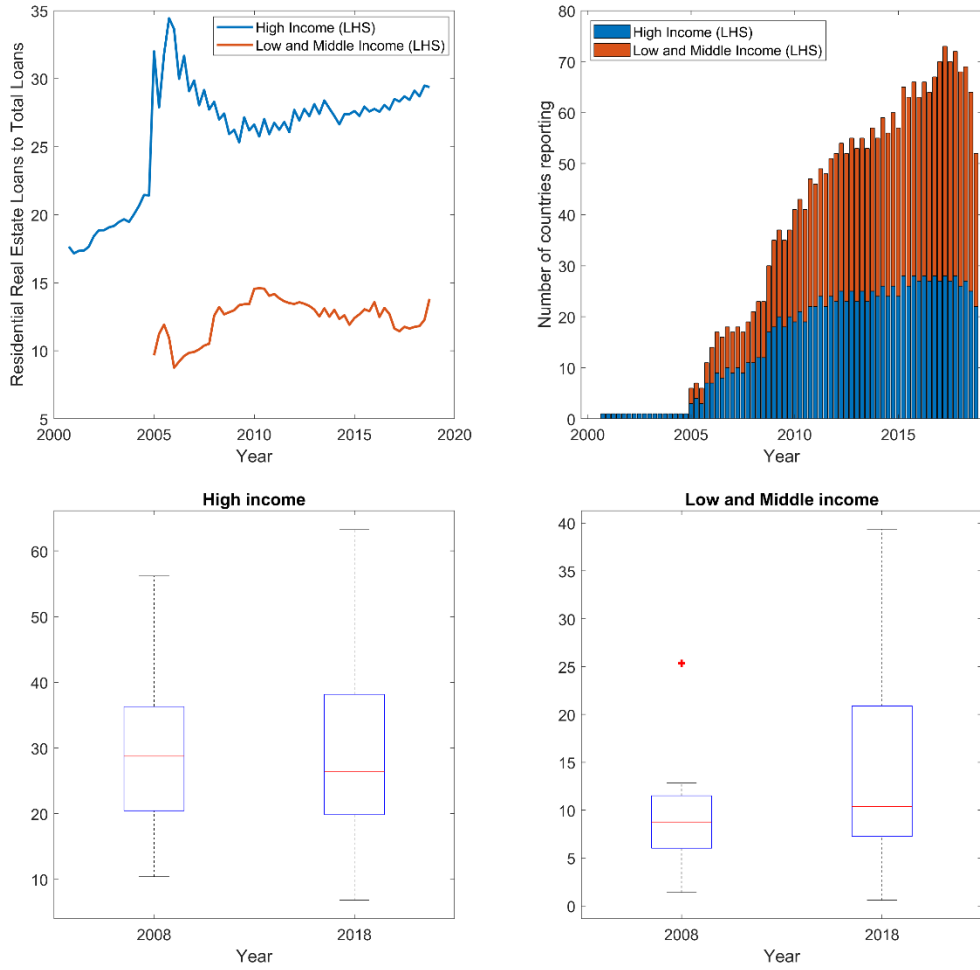


Figure 18 Return on Equity

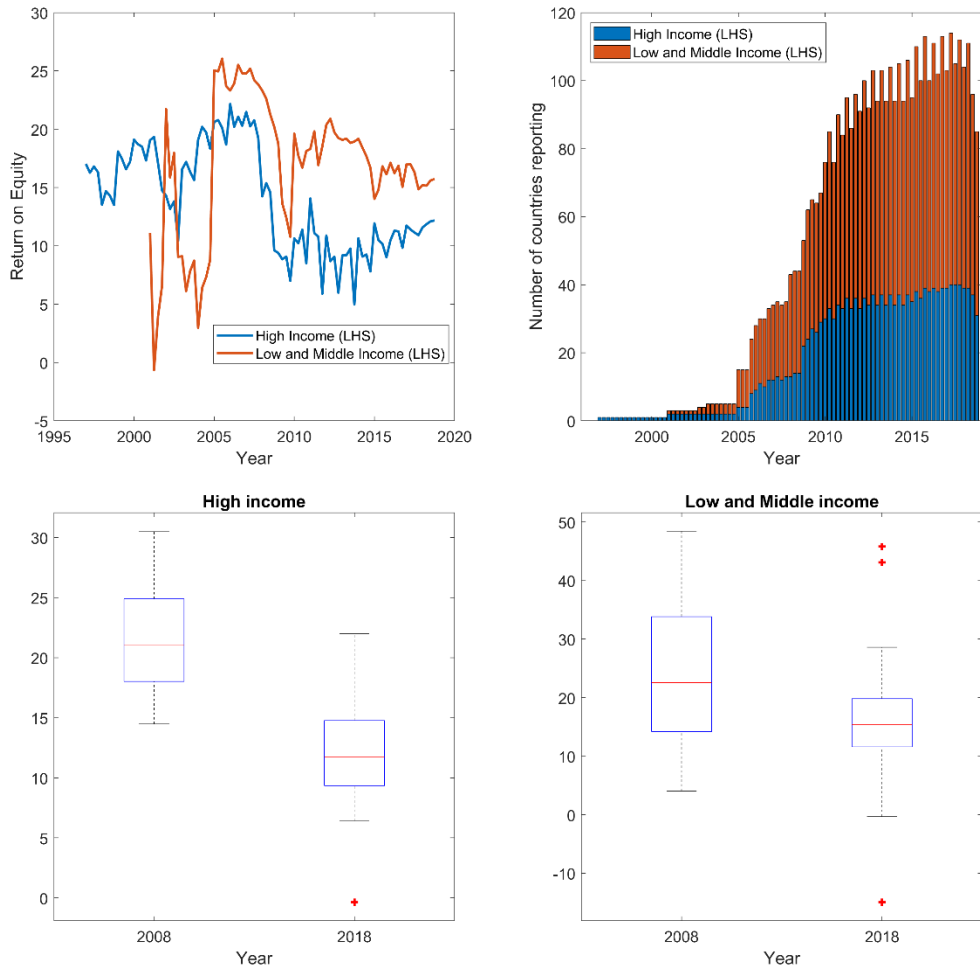
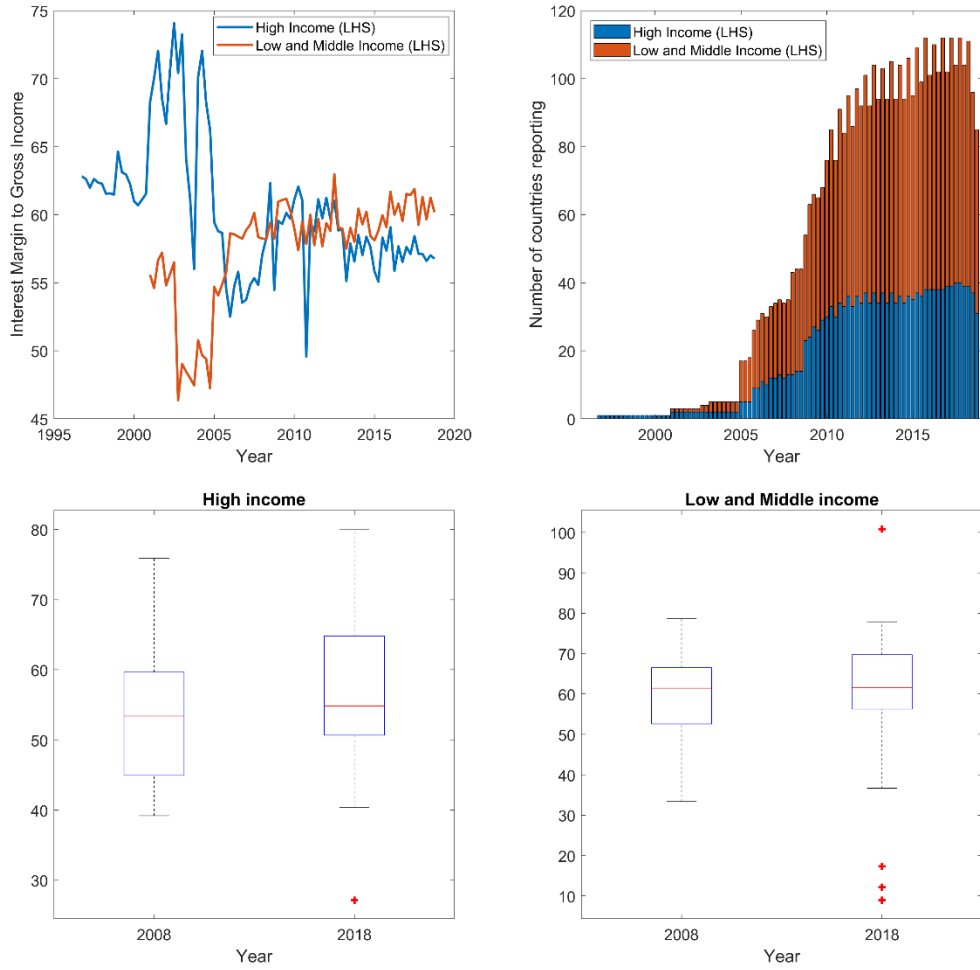


Figure 19 Interest Margin to Gross Income



C. Single variable predictive accuracy in high-income economies (HIE) sample

Table 3 Prediction performance measured by AUC - FCIs, HIE sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Regulatory Capital to Risk Weighted Assets raw	0.166	0.869	0.833	0.829	0.833	0.882	0.878	0.834	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.211	0.874	0.788	0.847	0.840	0.883	0.860	0.788	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets z	0.189	0.831	0.810	0.810	0.808	0.846	0.814	0.810	Yes
Net Open Position in Foreign Exchange to Capital raw	0.739	0.855	0.733	0.788	0.823	0.892	0.867	0.728	Yes
Regulatory Capital to Risk Weighted Assets z	0.229	0.766	0.770	0.770	0.770	0.806	0.767	0.771	Yes
Personnel Expenses to Non-interest Expenses raw	0.724	0.792	0.724	0.741	0.740	0.830	0.793	0.724	Yes
Return on Equity raw	0.744	0.780	0.743	0.743	0.743	0.812	0.745	0.743	Yes
Nonbank Financial Institutions Claims on Priv Sec z	0.721	0.806	0.717	0.718	0.720	0.820	0.757	0.719	No
Banking Institutions Claims on Priv Sec z	0.705	0.778	0.704	0.704	0.706	0.760	0.738	0.704	No
Tier 1 Capital to Assets z	0.283	0.735	0.716	0.716	0.717	0.749	0.714	0.716	Yes
Spread Between Reference Lending and Deposit Rates hp c4	0.333	0.782	0.667	0.720	0.728	0.810	0.717	0.665	Yes
Tier 1 Capital to Assets raw	0.297	0.770	0.702	0.668	0.700	0.770	0.714	0.703	Yes
DSR hp	0.695	0.770	0.695	0.693	0.695	0.763	0.721	0.695	No
Banking Institutions Foreign Liabilities z	0.689	0.768	0.688	0.688	0.701	0.778	0.709	0.688	No
Spread Between Reference Lending and Deposit Rates c1	0.389	0.809	0.611	0.743	0.763	0.803	0.747	0.610	Yes
Net Open Position in Foreign Exchange to Capital hp	0.346	0.769	0.653	0.637	0.613	0.810	0.781	0.655	Yes
Spread Between Reference Lending and Deposit Rates c4	0.340	0.789	0.657	0.603	0.661	0.807	0.728	0.656	Yes
Liquid Assets to Short Term Liabilities raw	0.372	0.791	0.627	0.664	0.627	0.839	0.748	0.627	Yes
Return on Equity z	0.681	0.723	0.679	0.679	0.679	0.743	0.686	0.680	Yes
Residential Real Estate Loans to Total Loans z	0.340	0.758	0.658	0.657	0.675	0.765	0.700	0.658	Yes

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSI. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AUS, AUT, BEL, CAN,

D. Single variable predictive accuracy – LV crises dataset

Table 4 Prediction performance measured by AUC – LV crises dataset, full sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Residential Real Estate Loans to Total Loans hp	0.174	0.524	0.802	0.801	0.721	0.672	0.758	0.814	Yes
Customer Deposits to Total Non-interbank Loans raw	0.223	0.519	0.776	0.756	0.668	0.707	0.779	0.777	Yes
Return on Assets z	0.745	0.525	0.742	0.733	0.696	0.689	0.746	0.742	Yes
Return on Equity z	0.748	0.476	0.745	0.740	0.683	0.738	0.728	0.746	Yes
Residential Real Estate Loans to Total Loans c4	0.230	0.478	0.761	0.763	0.661	0.677	0.726	0.765	Yes
Price To Rent raw	0.703	0.526	0.698	0.718	0.647	0.777	0.740	0.702	No
Regulatory Capital to Risk Weighted Assets d4	0.260	0.484	0.739	0.741	0.684	0.575	0.734	0.739	Yes
Price To Income z	0.719	0.501	0.718	0.715	0.594	0.714	0.724	0.717	No
Personnel Expenses to Non-interest Expenses z	0.719	0.527	0.717	0.706	0.661	0.628	0.683	0.713	Yes
Return on Equity c4	0.718	0.539	0.717	0.653	0.711	0.603	0.668	0.714	Yes
Regulatory Capital to Risk Weighted Assets c4	0.250	0.484	0.739	0.715	0.626	0.576	0.688	0.744	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.350	0.507	0.647	0.730	0.631	0.766	0.724	0.641	Yes
Portfolio Investment hp	0.713	0.548	0.686	0.654	0.585	0.720	0.669	0.713	No
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.367	0.689	0.631	0.656	0.604	0.739	0.679	0.630	No
Return on Assets c4	0.708	0.539	0.702	0.608	0.685	0.580	0.731	0.709	Yes
Net Open Position in Foreign Exchange to Capital z	0.301	0.475	0.698	0.669	0.636	0.679	0.680	0.692	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets c4	0.283	0.516	0.710	0.671	0.628	0.615	0.642	0.715	Yes
Price To Income raw	0.702	0.665	0.701	0.710	0.600	0.727	0.773	0.298	No
Price To Rent z	0.679	0.506	0.678	0.679	0.576	0.692	0.680	0.676	No
DSR raw	0.585	0.500	0.580	0.608	0.726	0.807	0.724	0.581	No

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSI. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AGO, ALB, ARG, ARM, AUS, AUT, BDI, BEL, BGD, BGR, BIH, BLR, BOL, BRA, BRN, BTN, BWA, CAF, CAN, CHL, CHN, CMR, COG, COL, COM, CRI, CYP, CZE, DEU, DJI, DNK, DOM, ECU, ESP, EST, FIN, FJI, FRA, GAB, GBR, GEO, GHA, GIN, GMB, GNQ, GRC, GRD, GTM, HKG, HND, HRV, HUN, IDN, IND, IRL, ISL, ISR, ITA, JPN, KAZ, KEN, KGZ, KHM, KNA, KOR, KWT, LBN, LKA, LSO, LTU, LUX, LVA, MDA, MDG, MDV, MEX, MKD, MMR, MUS, MWI, MYS, NAM, NGA, NIC, NLD, NOR, NPL, PAK, PAN, PER, PHL, PNG, POL, PRT, PRY, ROU, RUS, RWA, SGP, SLV, SVK, SVN, SWE, SWZ, SYC, TCD, THA, TJK, TTO, TUR, TZA, UGA, UKR, USA, UZB, VNM, ZAF, ZMB.

Table 5 Prediction performance measured by AUC – LV crises dataset, HIE sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.069	0.853	0.931	0.972	0.906	0.971	0.962	0.931	Yes
Regulatory Capital to Risk Weighted Assets raw	0.054	0.499	0.946	0.931	0.743	0.941	0.957	0.946	Yes
Personnel Expenses to Non-interest Expenses z	0.889	0.568	0.889	0.852	0.889	0.842	0.892	0.887	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets z	0.090	0.462	0.910	0.855	0.829	0.884	0.858	0.909	Yes
Capital to Assets raw	0.171	0.499	0.828	0.794	0.847	0.882	0.893	0.828	Yes
Personnel Expenses to Non-interest Expenses raw	0.830	0.504	0.827	0.858	0.777	0.864	0.860	0.829	Yes
Customer Deposits to Total Non-interbank Loans raw	0.211	0.474	0.787	0.707	0.774	0.852	0.779	0.787	Yes
Regulatory Capital to Risk Weighted Assets z	0.171	0.462	0.826	0.766	0.759	0.681	0.778	0.825	Yes
Capital to Assets z	0.199	0.551	0.796	0.790	0.682	0.702	0.776	0.796	Yes
Government Bonds Interest Rates raw	0.371	0.749	0.626	0.744	0.674	0.843	0.857	0.628	No
Return on Equity raw	0.727	0.500	0.721	0.753	0.749	0.653	0.791	0.722	Yes
Index of Equities hp	0.763	0.490	0.761	0.699	0.627	0.757	0.713	0.761	No
Return on Equity c1	0.679	0.482	0.669	0.732	0.738	0.790	0.777	0.680	Yes
Bank Credit to Pri Nonfin to GDP z	0.700	0.504	0.699	0.754	0.628	0.760	0.774	0.699	No
Liquid Assets to Total Assets Liquid Asset Ratio z	0.766	0.424	0.740	0.746	0.679	0.676	0.710	0.753	Yes
Government Bonds Interest Rates z	0.393	0.727	0.604	0.706	0.685	0.743	0.748	0.604	No
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.328	0.499	0.670	0.649	0.767	0.742	0.756	0.668	No
Price To Rent raw	0.695	0.499	0.694	0.709	0.609	0.772	0.734	0.693	No
Price To Income z	0.719	0.501	0.718	0.717	0.585	0.719	0.725	0.719	No
Interest Margin to Gross Income z	0.267	0.584	0.718	0.678	0.625	0.644	0.693	0.725	Yes

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSI. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AUS, AUT, BEL, BRN, CAN, CHL, CYP, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRC, HKG, HRV, HUN, IRL, ISL, ISR, ITA, JPN, KNA, KOR, KWT, LTU, LUX, LVA, NLD, NOR, PAN, POL, PRT, SGP, SVK, SVN, SWE, SYC, TTO, USA.

Table 6 Prediction performance measured by AUC – LV crises dataset, upper middle-income sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Housing Rent d4	0.887	0.452	0.873	0.862	0.814	0.851	0.876	0.881	No
Housing Rent d1	0.873	0.511	0.850	0.852	0.683	0.837	0.842	0.862	No
Treasury Bills Interest Rates raw	0.741	0.866	0.740	0.758	0.708	0.883	0.739	0.737	No
Personnel Expenses to Non-interest Expenses c4	0.747	0.765	0.745	0.741	0.631	0.812	0.742	0.746	Yes
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.654	0.742	0.644	0.680	0.747	0.848	0.813	0.644	No
Index of Equities raw	0.654	0.689	0.745	0.709	0.657	0.714	0.724	0.740	No
Regulatory Capital to Risk Weighted Assets raw	0.660	0.504	0.698	0.757	0.637	0.750	0.774	0.699	Yes
Real House Prices c4	0.745	0.547	0.731	0.691	0.656	0.606	0.689	0.743	No
Banking Institutions Foreign Liabilities z	0.719	0.501	0.719	0.714	0.573	0.723	0.734	0.719	No
Price To Rent raw	0.695	0.499	0.690	0.709	0.606	0.775	0.732	0.694	No
Banking Institutions Claims on Pri Sec to GDP z	0.356	0.746	0.644	0.665	0.630	0.744	0.670	0.644	No
Broad Money hp c1	0.688	0.438	0.676	0.438	0.762	0.767	0.871	0.734	No
Real House Prices d4	0.753	0.547	0.736	0.669	0.585	0.585	0.707	0.753	No
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.328	0.499	0.669	0.654	0.715	0.702	0.753	0.672	No
Housing Rent c1	0.336	0.454	0.665	0.732	0.724	0.711	0.714	0.663	No
Regulatory Capital to Risk Weighted Assets z	0.652	0.506	0.649	0.746	0.588	0.756	0.751	0.650	Yes
Index of Equities z	0.616	0.523	0.621	0.699	0.674	0.775	0.729	0.618	No
Tot Credit to Pri Nonfin to GDP raw	0.370	0.500	0.625	0.627	0.623	0.805	0.764	0.620	No
Price To Rent z	0.680	0.499	0.678	0.682	0.578	0.701	0.689	0.679	No
Government Bonds Interest Rates z	0.678	0.511	0.654	0.651	0.695	0.687	0.621	0.669	No

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSIs. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: ALB, ARG, ARM, BGR, BIH, BLR, BRA, BWA, CHN, COL, CRI, DOM, ECU, FJI, GAB, GEO, GNQ, GRD, GTM, KAZ, LBN, LKA, MDV, MEX, MKD, MUS, MYS, NAM, PER, PRY, ROU, RUS, THA, TUR, ZAF.

Table 7 Prediction performance measured by AUC – LV crises dataset, lower middle-income sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Foreign Currency Denominated Loans to Total Loans raw	0.773	0.894	0.772	0.967	0.919	0.927	0.966	0.772	Yes
Index of Equities raw	0.267	0.709	0.734	0.709	0.940	0.923	0.879	0.733	No
Foreign Currency Denominated Liabilities to Total Liabilities raw	0.633	0.817	0.614	0.787	0.936	0.972	0.958	0.611	Yes
Index of Equities z	0.258	0.777	0.741	0.699	0.841	0.886	0.826	0.742	No
Customer Deposits to Total Non-interbank Loans raw	0.209	0.609	0.786	0.786	0.800	0.799	0.806	0.789	Yes
Capital to Assets raw	0.811	0.497	0.810	0.790	0.783	0.760	0.807	0.811	Yes
Foreign Currency Denominated Loans to Total Loans hp	0.206	0.496	0.783	0.781	0.750	0.830	0.773	0.783	Yes
Regulatory Capital to Risk Weighted Assets d4	0.239	0.489	0.754	0.806	0.772	0.722	0.777	0.759	Yes
Interest Margin to Gross Income hp	0.228	0.477	0.766	0.709	0.755	0.705	0.761	0.765	Yes
Regulatory Capital to Risk Weighted Assets c4	0.217	0.489	0.777	0.750	0.675	0.716	0.710	0.777	Yes
Net Open Position in Foreign Exchange to Capital z	0.298	0.465	0.695	0.715	0.769	0.788	0.814	0.697	Yes
Return on Assets z	0.715	0.517	0.709	0.685	0.717	0.768	0.785	0.714	Yes
Liquid Assets to Total Assets Liquid Asset Ratio hp	0.750	0.516	0.741	0.737	0.704	0.671	0.679	0.743	Yes
Bank Credit to Pri Nonfin to GDP z	0.700	0.504	0.698	0.757	0.637	0.750	0.774	0.699	No
Banking Institutions Foreign Liabilities z	0.314	0.679	0.685	0.671	0.649	0.737	0.672	0.686	No
Return on Equity z	0.689	0.517	0.686	0.682	0.698	0.731	0.776	0.685	Yes
Portfolio Investment hp	0.752	0.510	0.700	0.489	0.784	0.759	0.697	0.750	No
Liquid Assets to Short Term Liabilities hp c4	0.426	0.510	0.563	0.806	0.802	0.737	0.860	0.574	Yes
Price To Income z	0.719	0.501	0.719	0.714	0.573	0.723	0.734	0.719	No
Price To Rent raw	0.695	0.499	0.690	0.709	0.606	0.775	0.732	0.694	No

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSI. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AGO, BGD, BOL, BTN, CMR, COG, COM, DJI, GHA, HND, IDN, IND, KEN, KGZ, KHM, LSO, MDA, MMR, NGA, NIC, PAK, PHL, PNG, SLV, SWZ, UKR, UZB, VNM, ZMB.

Table 8 Prediction performance measured by AUC – LV crises dataset, low-income sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Government Bonds Interest Rates z	0.851	0.982	0.854	0.858	0.979	0.973	0.925	0.852	No
Government Bonds Interest Rates raw	0.837	0.532	0.816	0.811	0.963	0.751	0.877	0.818	No
Treasury Bills Interest Rates z	0.685	0.830	0.658	0.491	0.885	0.942	0.896	0.668	No
Banking Institutions Foreign Liabilities z	0.618	0.499	0.782	0.771	0.740	0.832	0.779	0.781	No
Current Account to GDP hp	0.509	0.813	0.523	0.845	0.831	0.880	0.817	0.545	No
Government Bonds Interest Rates hp	0.537	0.453	0.731	0.730	0.713	0.772	0.819	0.760	No
Net Acquisition of Financial Assets hp	0.583	0.880	0.518	0.747	0.708	0.892	0.751	0.539	No
Banking Institutions Claims on Pri Sec to GDP z	0.696	0.500	0.703	0.711	0.680	0.807	0.759	0.703	No
Net Acquisition of Financial Assets raw	0.432	0.686	0.608	0.551	0.750	0.847	0.818	0.634	No
Bank Credit to Pri Nonfin z	0.700	0.504	0.698	0.757	0.637	0.750	0.774	0.699	No
Net Acquisition of Financial Assets hp c1	0.437	0.835	0.522	0.746	0.666	0.797	0.789	0.494	No
Price To Income z	0.719	0.501	0.719	0.714	0.573	0.723	0.734	0.719	No
Price To Rent raw	0.695	0.499	0.690	0.709	0.606	0.775	0.732	0.694	No
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.628	0.499	0.669	0.654	0.715	0.702	0.753	0.672	No
Tot Credit to Pri Nonfin z	0.652	0.506	0.649	0.746	0.588	0.756	0.751	0.650	No
Current Account to GDP c4	0.577	0.475	0.571	0.869	0.679	0.832	0.740	0.540	No
Treasury Bills Interest Rates raw	0.573	0.637	0.552	0.568	0.828	0.929	0.602	0.548	No
Price To Rent z	0.680	0.499	0.678	0.682	0.578	0.701	0.689	0.679	No
Price To Income raw	0.702	0.643	0.700	0.710	0.606	0.739	0.766	0.298	No
Banking Institutions Claims on Pri Sec to GDP c4	0.653	0.650	0.644	0.584	0.599	0.692	0.637	0.647	No

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSIs. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: BDI, CAF, GIN, GMB, MDG, MWI, NPL, RWA, TCD, TJK, TZA, UGA.

E. Single variable predictive accuracy – LD crises dataset

Table 9 Prediction performance measured by AUC – LD crises dataset, full sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Regulatory Capital to Risk Weighted Assets raw	0.031	0.927	0.968	0.972	0.969	0.970	0.971	0.968	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.119	0.933	0.881	0.989	0.979	0.985	0.985	0.881	Yes
Personnel Expenses to Non-interest Expenses raw	0.889	0.856	0.886	0.919	0.952	0.945	0.953	0.885	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets z	0.090	0.551	0.908	0.894	0.769	0.865	0.897	0.907	Yes
Capital to Assets raw	0.153	0.469	0.846	0.839	0.778	0.832	0.837	0.846	Yes
Capital to Assets z	0.155	0.538	0.843	0.832	0.687	0.793	0.822	0.840	Yes
Regulatory Capital to Risk Weighted Assets z	0.161	0.551	0.837	0.813	0.728	0.780	0.789	0.838	Yes
Return on Equity z	0.821	0.449	0.817	0.815	0.723	0.819	0.815	0.813	Yes
Personnel Expenses to Non-interest Expenses z	0.781	0.568	0.777	0.763	0.742	0.713	0.783	0.771	Yes
Residential Real Estate Loans to Total Loans c4	0.205	0.500	0.787	0.756	0.730	0.725	0.757	0.789	Yes
Return on Equity raw	0.745	0.537	0.742	0.762	0.681	0.772	0.816	0.738	Yes
Customer Deposits to Total Non-interbank Loans d4	0.255	0.451	0.742	0.769	0.792	0.733	0.795	0.739	Yes
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.547	0.781	0.488	0.522	0.970	0.969	0.953	0.487	No
Bank Credit to Pri Nonfin to GDP c4	0.734	0.611	0.734	0.728	0.652	0.752	0.771	0.734	No
DSR raw	0.664	0.720	0.660	0.672	0.821	0.812	0.675	0.661	No
Price To Income z	0.738	0.650	0.736	0.736	0.597	0.725	0.725	0.737	No
Return on Assets z	0.721	0.449	0.713	0.723	0.779	0.756	0.707	0.718	Yes
Customer Deposits to Total Non-interbank Loans raw	0.291	0.528	0.702	0.700	0.692	0.775	0.715	0.708	Yes
Bank Credit to Pri Nonfin to GDP z	0.700	0.504	0.698	0.757	0.637	0.750	0.774	0.699	No
Liquid Assets to Total Assets Liquid Asset Ratio z	0.748	0.552	0.739	0.723	0.606	0.655	0.682	0.735	Yes

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSIs. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AUT, BEL, BGR, CYP, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRC, HRV, HUN, IRL, ITA, LTU, LUX, LVA, NLD, NOR, POL, PRT, ROU, SVK, SVN, SWE.

Table 10 Prediction performance measured by AUC – LD crises dataset, HIE sample

Variable	SigExt	Tree	LDA	QDA	SVM	KNN	NB	LR	FSI
Regulatory Capital to Risk Weighted Assets raw	0.035	0.953	0.965	0.969	0.965	0.967	0.965	0.965	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets raw	0.133	0.929	0.866	0.986	0.976	0.985	0.983	0.867	Yes
Personnel Expenses to Non-interest Expenses raw	0.880	0.840	0.876	0.920	0.948	0.944	0.956	0.878	Yes
Regulatory Tier 1 Capital to Risk Weighted Assets z	0.086	0.513	0.912	0.898	0.799	0.861	0.902	0.911	Yes
Capital to Assets z	0.148	0.538	0.850	0.838	0.741	0.812	0.839	0.847	Yes
Regulatory Capital to Risk Weighted Assets z	0.150	0.513	0.849	0.834	0.766	0.800	0.816	0.849	Yes
Capital to Assets raw	0.170	0.497	0.827	0.825	0.755	0.805	0.818	0.825	Yes
Return on Equity z	0.822	0.497	0.816	0.822	0.733	0.826	0.821	0.818	Yes
Nonbank Financial Institutions Claims on Pri Sec to GDP z	0.547	0.897	0.497	0.561	0.969	0.966	0.955	0.486	No
Personnel Expenses to Non-interest Expenses z	0.778	0.552	0.776	0.756	0.708	0.731	0.793	0.769	Yes
Bank Credit to Pri Nonfin to GDP c4	0.734	0.617	0.733	0.729	0.661	0.763	0.771	0.733	No
DSR raw	0.664	0.776	0.659	0.675	0.810	0.815	0.666	0.663	No
Customer Deposits to Total Non-interbank Loans d4	0.277	0.450	0.723	0.737	0.795	0.709	0.797	0.713	Yes
Residential Real Estate Loans to Total Loans c4	0.217	0.502	0.773	0.754	0.620	0.723	0.712	0.779	Yes
Liquid Assets to Short Term Liabilities raw	0.716	0.492	0.703	0.767	0.700	0.758	0.792	0.710	Yes
Foreign Currency Denominated Liabilities to Total Liabilities raw	0.667	0.653	0.641	0.585	0.836	0.811	0.786	0.648	Yes
Return on Equity raw	0.728	0.490	0.718	0.748	0.680	0.741	0.795	0.723	Yes
Price To Income z	0.738	0.646	0.737	0.735	0.574	0.725	0.724	0.735	No
Return on Assets z	0.720	0.497	0.709	0.738	0.717	0.719	0.743	0.714	Yes
Residential Real Estate Loans to Total Loans c1	0.299	0.483	0.697	0.634	0.735	0.733	0.746	0.694	Yes

Note: The table reports cross-validated out-of-sample performance measured by AUC for eight methods and evaluation horizon of 4-12 quarters. The last column shows whether a variable is a FSIs. Note that AUC levels for signal extraction below 0 means that a variable generates a signal when it achieves low level. Unbalanced panel data covers periods from 2000 to 2016 subject to data availability. Countries included: AUT, BEL, CYP, CZE, DEU, DNK, ESP, EST, FIN, FRA, GBR, GRC, HRV, HUN, IRL, ITA, LTU, LUX, LVA, NLD, NOR, POL, PRT, SVK, SVN, SWE.

F. Multivariate early warning models – additional crises datasets

Figure 20 ROC curve—LV crises dataset

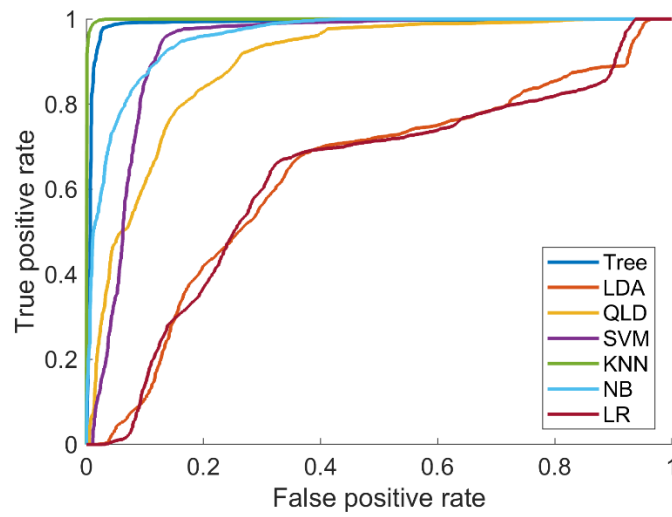


Figure 21 ROC curve—LD crises dataset

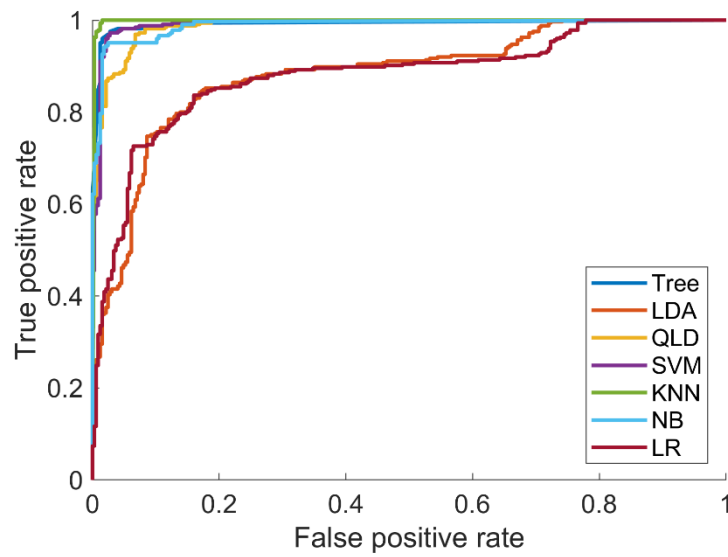


Figure 22 Confusion matrices for LV crises, full sample

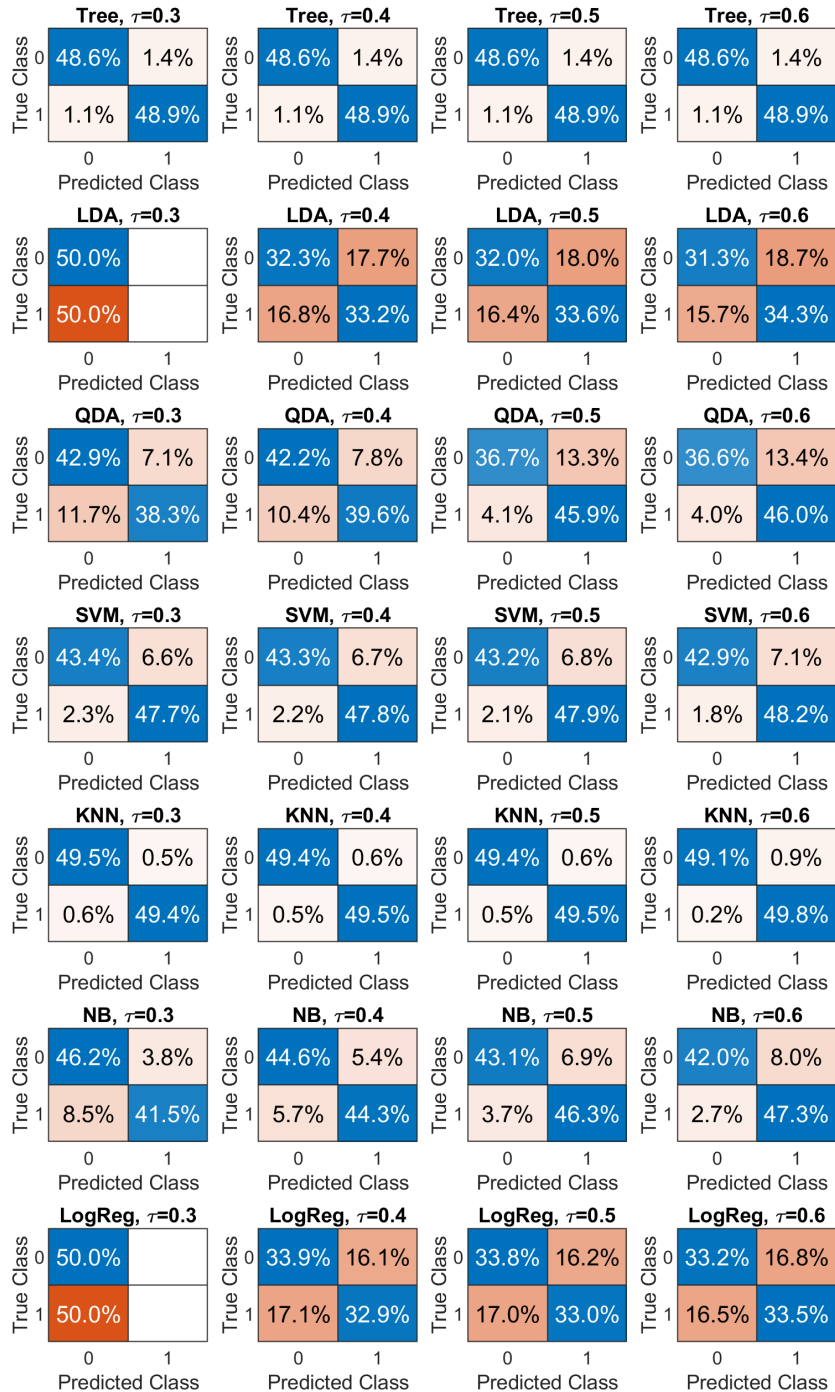


Figure 23 Confusion matrices for LD crises, full sample

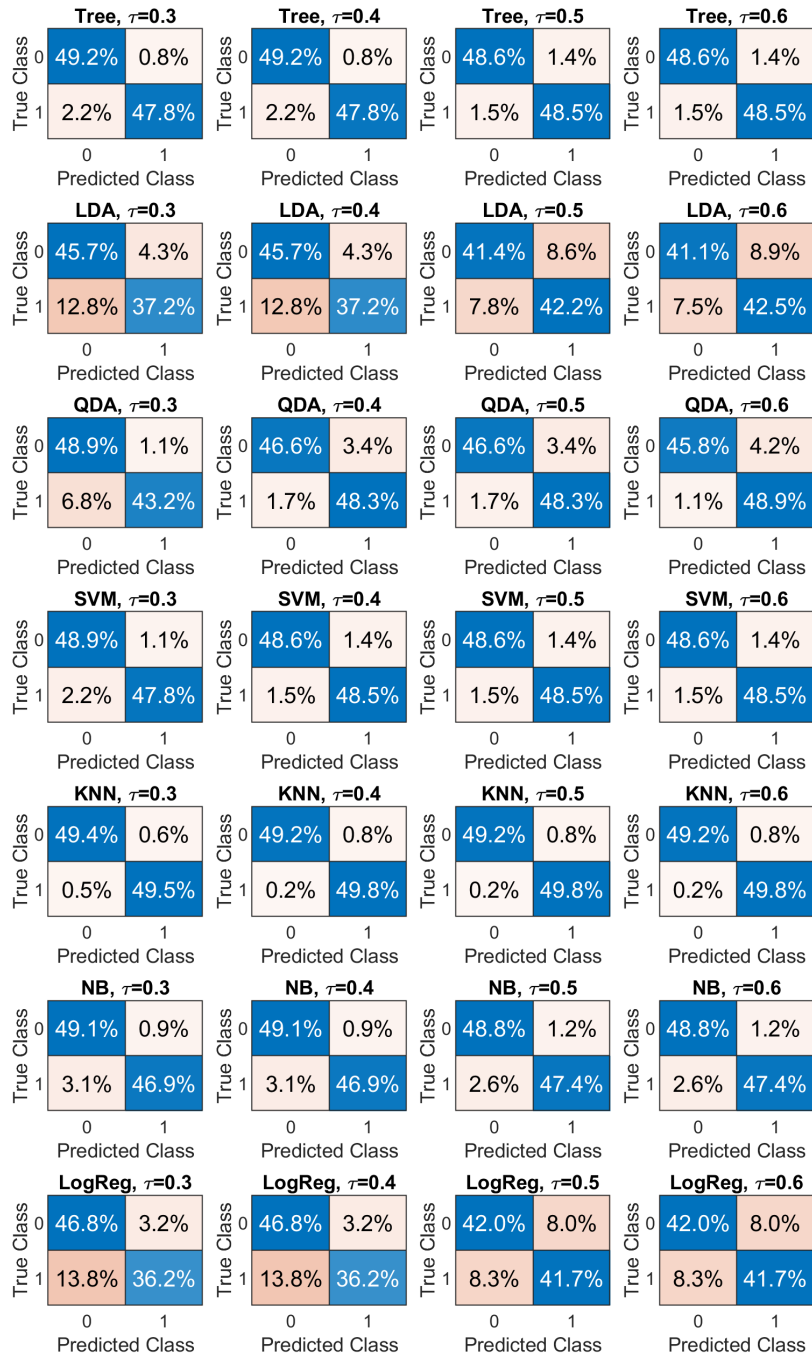


Table 11 Prediction performance of FSIs measured by AUC - LV, full sample

	τ	0.30			0.4			0.5			0.6		
	Relative cost	0.43			0.67			1			1.5		
Method	AUC	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR
Tree	0.99	0.56	0.03	0.98	0.56	0.03	0.98	0.56	0.03	0.98	0.56	0.03	0.98
LDA	0.63	0.84	0.00	0.00	0.51	0.35	0.66	0.51	0.36	0.67	0.51	0.37	0.69
QDA	0.89	0.75	0.14	0.77	0.73	0.16	0.79	0.52	0.27	0.92	0.51	0.27	0.92
SVM	0.93	0.63	0.13	0.95	0.63	0.13	0.96	0.62	0.14	0.96	0.61	0.14	0.96
KNN	1.00	0.93	0.01	0.99	0.92	0.01	0.99	0.92	0.01	0.99	0.87	0.02	1.00
NB	0.96	0.75	0.08	0.83	0.65	0.11	0.89	0.53	0.14	0.93	0.46	0.16	0.95
LR	0.63	0.71	0.00	0.00	0.51	0.32	0.66	0.51	0.32	0.66	0.51	0.34	0.67

Note: The table reports cross-validated out-of-sample performance measured by AUC for seven methods and evaluation horizon of 4-12 quarters. The second column shows Area Under the ROC curve. Next three columns report optimal threshold (Tr) in terms of probability from the model and False Positive Rate (FPR) and True Positive Rate (TPR) associated with the optimal threshold. The optimal threshold is chosen according to the cost of missing a crisis (τ). Unbalanced panel data covers periods from 2000 to 2016 subject to data availability.

Table 12 Prediction performance of FSIs measured by AUC - LD, full sample

	τ	0.30			0.4			0.5			0.6		
	Relative cost	0.43			0.67			1			1.5		
Method	AUC	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR	Tr	FPR	TPR
Tree	0.99	0.67	0.02	0.96	0.67	0.02	0.96	0.37	0.03	0.98	0.37	0.03	0.98
LDA	0.88	0.78	0.09	0.75	0.78	0.09	0.75	0.55	0.17	0.85	0.53	0.18	0.85
QDA	0.99	0.85	0.02	0.87	0.27	0.07	0.97	0.27	0.07	0.97	0.09	0.08	0.98
SVM	0.99	0.82	0.02	0.96	0.74	0.03	0.97	0.74	0.03	0.97	0.74	0.03	0.97
KNN	1.00	0.80	0.01	0.99	0.70	0.02	1.00	0.70	0.02	1.00	0.70	0.02	1.00
NB	0.99	0.58	0.02	0.94	0.58	0.02	0.94	0.50	0.02	0.95	0.50	0.02	0.95
LR	0.88	0.63	0.06	0.73	0.63	0.06	0.73	0.53	0.16	0.84	0.53	0.16	0.84

Note: The table reports cross-validated out-of-sample performance measured by AUC for seven methods and evaluation horizon of 4-12 quarters. The second column shows Area Under the ROC curve. Next three columns report optimal threshold (Tr) in terms of probability from the model and False Positive Rate (FPR) and True Positive Rate (TPR) associated with the optimal threshold. The optimal threshold is chosen according to the cost of missing a crisis (τ).