International Review of Financial Analysis 85 (2023) 102451

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Measuring financial soundness around the world: A machine learning approach

Alessandro Bitetto [a,*], Paola Cerchiello [a], Charilaos Mertzanis [b]

[a] University of Pavia, Italy
[b] Abu Dhabi University, United Arab Emirates

ARTICLE INFO

JEL classification:
E32
E42
E61
E02
F02

Keywords:
Financial soundness
Data-driven
Cross-country
Policy framework
Principal Component Analysis
Random Forest

ABSTRACT

We use a fully data-driven approach and information provided by the IMF's financial soundness indicators to measure the condition of a country's financial system around the world. Given the nature of the measurement problem, we apply different versions of principal component analysis (PCA) to deal with the presence of strong cross-sectional and time dependence in the data due to unobserved common factors. Using this comprehensive sample and various statistical methods, we produce an alternative data-driven measure of financial soundness that provides policy makers and financial institutions with a monitoring and policy tool that is easy to implement and update. We validate our index by using alternative macroeconomic factors, confirming its predictive power. Our index captures important aspects of financial intermediation around the world.

1. Introduction

The importance of a sound financial system for sustained economic growth is well documented (Allen & Gale, 2000). A sound financial system supports economic activity by pooling and mobilizing savings for productive use, providing information on existing and potential investment opportunities, improving corporate governance and facilitating trading, diversification, and risk management. The 2007–8 financial crisis has underscored the importance of financial system resiliency in providing finance to households and business throughout the business cycle. It has also emphasized the importance of limiting the types of financial and real imbalances that develop during times of prosperity. When such balances unwind, they can cause significant damage to the financial system and the economy.

Financial stability policies pursued by governments and central banks heavily depend on economic forecasts, especially during times of financial turmoil (Financial Stability Board, 2017). The accuracy of forecasting models is however still a challenging policy issue since modern economies are subject to various shocks that make the forecasting tasks very hard, in both the short and long terms. The global financial crisis has highlighted that credible information about the underlying financial risk dynamics requires the understanding of the complex and time-varying nature of the data (Basel Committee on Banking Supervision, 2018). In this respect, the application of modern data science technologies for improving forecasting has high potential. The rising amounts and frequency of data have opened up new opportunities to financial monitoring and analysis. Economic forecasting performed by modern machine-learning technologies can integrate and augment the information carried out by publicly available aggregated variables produced by national and international authorities. By lowering the level of granularity, data science technologies can uncover economic relationships that are often not evident when information is in an aggregated form over many variables, sections, or time periods. In this way, these technologies can improve the decision-making process.

One way to address the challenge of assessing the complexity and soundness of the financial system as a whole is to adopt a data-driven approach. This approach may provide a better evidence basis to support reasoning and decision. Given the complexity of the financial system, it may lead to better optimized assessments. Data-driven techniques are the result of an algorithmic modeling culture, in which input information, transformed through a partly unknowable system, produces predictions as outcomes (Breiman, 2001). The goal is to find a rule (algorithm) that transforms input information in order to predict outcomes more effectively without a priori assumptions about

* Correspondence to: Via San Felice al Monastero, 5 27100, Pavia, Italy.
E-mail addresses: alessandro.bitetto@unipv.it (A. Bitetto), paola.cerchiello@unipv.it (P. Cerchiello), charilaos.mertzanis@adu.ac.ae (C. Mertzanis).

https://doi.org/10.1016/j.irfa.2022.102451
Received 3 March 2022; Received in revised form 5 October 2022; Accepted 16 November 2022
Available online 28 November 2022

the relationships between input variables. The algorithms are designed to learn, in various ways, from the data with minimal human intervention. Standard empirical approaches cannot fully understand the risk dynamics, which are inherent in the financial systems and often cause financial crises, because structural forces interact in nonlinear and contingent ways (Arakelian et al., 2019). Thus, these complex risk dynamics cannot be easily reduced to common data models in which outcomes are generated from independent predictor variables and random noise. Under these circumstances, the outcomes reflect more the model structure and less the nature of data and information. While there are no data-driven algorithms that can lead to fully optimal assessments of the financial system's soundness, this approach has considerable advantages, such as avoiding post-hoc rationalization, as is typical for economic narrative-based methods.

Data-driven techniques can better analyze multi-dimensional datasets, such as financial data generated by many sources, better manage unbalanced datasets with missing information, do not require ex ante modeling assumptions, and better handle the policy versus forecasting trade-offs. These characteristics endow the data-driven models with higher power for mining information and using it for financial monitoring and prediction (L et al., 2019). Data-driven methods are widely used in many scientific disciplines, but there are scarcely used in financial stability analysis. Recently some studies use machine learning techniques for improving prediction performance, especially in financial institutions credit scoring models and in the construction of early warning models and the prediction of financial crises (Consoli et al., 2021). However, these data-driven approaches mostly focus on individual institutions or countries, use information that does not capture policy considerations and often study financial vulnerabilities associated with static rather dynamic (both cross-section and time-varying) conditions and cross-market/bank interdependencies.

In this paper, we contribute to the development of structural dynamic models of financial stability prediction by constructing a financial soundness index (FSIND) using a data-driven methodological approach. A further key novelty of our approach is the use of credible and objective information provided by the International Monetary Fund's (IMF) Financial Soundness Indicators (FSIs), which incorporate both institution-wide, market-wide and policy-level considerations. This is the first paper to capture these considerations simultaneously. Given the nature of the measurement problem across many dimensions, another novelty of our approach is the use of advanced principal component analysis (PCA) than can deal simultaneously with the presence of strong cross-section and time dependence in the data due to unobserved common factors. The PCA provides a more flexible but robust approach to capturing strong common factors on a dynamic basis. Indeed, the PCA method offers a variety of statistical tools not only to assess the quality of the results but also to interpret and replicate conditional insights. This choice is in line with the methodology used in related studies (Hakkio & Keeton, 2009; Illing & Liu, 2006). We build the FSIND index by using the IMF's Financial Soundness Indicators specifically tailored to measure the strengths and weaknesses of the financial system as a whole comprising information from markets, institutions and regulators. The FSIND index is available for 119 countries for the 2010–2017 period. We enrich the index by adding qualitative information capturing cultural and geographical specificities. We carry out extensive model validation and sensitivity analysis that makes the model robust and statistically sound. Finally, we make the index values available in both continuous and binary formats to accommodate alternative policy needs and research strategy specifications.

Our paper contributes to the financial stability literature in several ways. First, it creates a new financial soundness index for 119 countries that is fully data-driven, tested and validated, based on credible IMF information that captures both market-wide and policy-level considerations. We allow the data to "speak" by means of an unsupervised statistical learning technique, namely the Principal Component Analysis (hereafter PCA). The latter makes neither a-priori assumption on the relationship among the input variables nor a subjective decision on the variables to be possibly dropped. Further, the model does not need to define a target variable, thereby avoiding a further level of subjectivity. The only model assumption rests on the number of components built on the original variable space reflecting the desired level of captured variability and parsimony. Moreover, the new coordinates must, by construction, lie on a linear space and be mutually orthogonal (i.e. independent). Independence ensures that each new principal component is describing a specific and not known in advance latent phenomenon through the linear combination of the initial variables. Our approach is methodologically related to that of Brave and Butters (2011), Hakkio and Keeton (2009), Kliesen and Smith (2010) and Louzis and Vouldis (2011), but these authors use monthly market-based data and study single counties, whilst we use aggregated bank balance-sheet and structural macro data for many countries. Our analysis also complements recent risk assessments based on the use of machine learning methods (Lin et al., 2012). Indeed, the authors stress that, beside the efficiency of the machine learning algorithm (often ensemble models do the job), the dataset, the selection of leading variables and the pre-processing phase in general play a key role in producing accurate assessments. We have placed special emphasis on these aspects in our analysis.

Second, our paper contributes to the literature on stress-testing and early warning models aimed at assessing financial stability (Alessi & Detken, 2018b; Drehmann & Juselius, 2014; Lo Duca & Peltonen, 2011; Rose & Spiegel, 2012). These studies use different assessment methods but they do not explicitly consider data-driven methods. Other studies use data-driven methods but they focus on either individual institutions or single countries. For example, using random forests, Alessi and Detken (2018a) develop early warning models of systemic financial crises at the country level and Tanaka et al. (2016) predict failures at the level of individual banks. Our approach is also related to that of Samitas et al. (2020), who use machine learning and network modeling to detect financial contagion effects in 33 countries, thereby improving the prediction of financial crises. They apply network analysis to study dynamic conditional correlations, which they use to identify market interdependencies. It is further related to that of Liu et al. (2022), who find that machine learning algorithms predict default risk of online loan borrowers in China better than logistic models and therefore provide benefits to investors. Finally, our analysis is related to cross-country studies of financial market contagion (Baele, 2005; Corsetti et al., 2005; Dornbusch et al., 2000; Geraci & Gnabo, 2018; Masson, 1999; Pericoli & Sbracia, 2003), as well as to cross-country studies that analyze the network interconnectedness of banking systems (Babus, 2016; Hausenblas et al., 2015; Minoiu et al., 2015). However, in contrast to these studies, our study focuses on dynamic structural conditions and not the dissemination of cross-market disturbances and bank interdependencies themselves.

Third, our paper contributes to the empirical literature on credit scoring models aimed at assessing the financial vulnerabilities of financial institutions and the prediction of corporate default (Abdou et al., 2017, 2016; Alaka et al., 2018; Devi & Radhika, 2018; Kim et al., 2020). Different standard statistical techniques and data-driven approaches have been used for assessing the credibility of ratings and predictions of the different financial institutions in different institutional settings. However, while often innovative and powerful, the latter approaches are institution/unit focused and mostly structural in nature and do not look at the financial system as a whole. Our paper takes on system-wide approach and integrates the implications of financial stability policies in general and not those focusing on specific institutions.

The paper is organized as follows: in Section 2 we recall the background of our paper, in Section 3 we discuss the literature on financial soundness index and relative approaches, in Section 4 we present the methodology detailing the Principal Component Analysis, a more robust version employed and the validation index techniques. In Section 5 we extensively discuss the data and the preprocessing phase, in Section 6 results will be shown and discussed and conclusions are given in Section 7.

2. Background

Measuring financial stability is a formidable challenge (Gadanecz & Jayaram, 2008). The financial system is a complex one. It consists of many diverse actors, including banks, mutual funds, hedge funds, insurance companies, pension funds and shadow banks. All of them interact with each other and the real economy in complex ways. The 2007–8 financial crisis provided a clear example of this complexity and its consequences. Assessing financial stability therefore requires the consideration of the diverse macroeconomic, structural and institutional aspects of the financial system (European Central Bank, 2005). The large volume of international capital flows has made it increasingly important to strengthen the foundations of domestic financial systems in order to build up resilience to capital flow volatility and its effects on the economy. Maintaining financial resilience, strong macroeconomic performance and effective monetary policy at the national level requires sound financial institutions. Hence, central banks and governments monitor closely the health and efficiency of financial institutions and markets along with the macroeconomic and institutional developments that pose potential risks to financial stability.

The Basel Committee on Banking Supervision (2013) has stressed the need to focus on the ability of financial institutions to manage risk. The Committee identified the major shortcomings of financial institutions' ability to quickly and accurately aggregate risk exposures and identify risk concentration at the individual institution level, across business lines and between legal entities. These shortcomings affect the financial institutions' ability to quickly carry out stress testing in order to assess their exposure to risk associated with particular economic-financial scenarios. In order to achieve the efficiency, reliability and transparency required for performing large-scale risk analysis, decision-makers need a proper standardization of the data, rules for financial flow generating algorithms and sufficient computational power (Mertzanis, 2018).

Using absolute quantitative levels of risk may not adequately capture the extent of a financial institutions' or financial system's vulnerability. There is a need for implementing a "net risk" approach that combines both the quantitative and qualitative aspects of financial vulnerability based on proper information and measurement (Lo Duca & Peltonen, 2011). The net risk approach involves the quantitative evaluation of all risks faced by financial institutions and the qualitative adjustment for institutional factors. It helps to better assess whether risks are managed adequately through market discipline and internal governance in a financial institution, as well as through regulatory and supervisory frameworks in the financial system as a whole. The outcomes of these evaluations combine to produce an overall risk assessment for individual financial institutions and an overall stability assessment for the financial system as a whole.

The qualitative adjustments require the consideration of the institutional characteristics of the financial system. The nature of government intervention in the economy, the payments culture, the insolvency regime, the credit and deposit guarantees, the quality of supervision and regulation, moral hazards, corporate governance, and management quality all affect the general incentives structure of a financial system and must be taken into account in qualitative adjustments. Combining the qualitative and quantitative aspects of risk assessment is not an exact method and requires judgment. This introduces considerable complexity in the assessment process.

Advanced data-driven techniques have certain advantages relative to standard empirical models used to predict financial trends. First, they can better analyze multi-dimensional datasets, as is often the case with financial data. Central banks and financial market participants produce today rich information that comprises hundreds of indicators (Basel Committee on Banking Supervision, 2018; Financial Stability Board, 2017). Financial institutions are increasingly using advanced data technologies to ensure they are able to identify and address new financial risks. Financial supervisors are also increasing paying attention to the data-driven technologies and advanced data analytics (i.e., suptech) to increase the efficiency and effectiveness of managing large volumes of information. Second, data-driven algorithms can better manage unbalanced datasets with lower loss of important information. Large datasets contain large amounts of missing values, mixed frequencies, ragged edges, or irregular patterns, which standard empirical models are not able to handle. Third, data-driven techniques do not require making ex ante modeling choices and assumptions. For example, econometric techniques require selecting a restricted number of variables, and factor models, while able to handle large data, require the setting of the underlying driving assumptions. Moreover, econometric models require controlling for nonlinearities and interactions, whose existence must be hypothesized a priori, whilst data-driven methods can address the dynamics directly. These characteristics confer certain advantages of the data-driven models in predictive performance. Data-driven methods are relatively more robust in handling the fitting versus forecasting trade-offs (Clements & Hendry, 1998), which underlie the separation between models used for forecasting and those used for policymaking. When historical data is noisy (as is the case with financial crises episodes), the use of empirical models that overfit in-sample data runs the risk of producing spurious estimates that provide deficient forecasts and/or mis-identifying crisis events due to the noisy transmission of information. Data-driven techniques can overcome this problem. However, data-driven techniques have also disadvantages, the most important of which is that they are "black box" models. Indeed, while the data-driven models can control inputs and obtain broadly accurate outputs, they cannot explain the reasons behind the specific outcome produced by the algorithm. Therefore, it becomes very difficult to build an argument that would help policy-makers make sense of the outcomes, which is as important as the ability to make accurate predictions.

3. Relevant literature

3.1. Theoretical approaches

In carrying out financial stability assessment, it is important to evaluate how risk-taking financial institutions manage risk and how supervisors regulate the management of risk. Different financial institutions have different risk appetites. Further, the particular institutional and regulatory framework of the financial system strongly influences the level of risk-taking. This raises two challenges: the timely aggregation and measurement of financial risk and the usefulness of absolute risk levels.

The 1997 Asian financial crisis and the 2007–8 global financial crisis showed the importance of financial stability for economic activity. These crises demonstrated that certain structures of financial institutions' balance sheets could adversely affect the financial sector and lead to the origination and perpetuation of financial crises (Laeven & Valencia, 2012). These crises further demonstrated that the vulnerability of the financial institutions could exist along with robust macroeconomic conditions. As a result, the systematic and regular monitoring of both the financial institutions' balance sheets and the macroeconomic conditions emerged as key policy considerations throughout the world. These considerations have been associated with the emergence of the notion of "macro-prudential supervision". Hawkesby (2000) was the first to introduce the term referring to a set of indicators that collectively indicate the riskiness of financial institutions and its implication for financial system stability.

Financial institutions soundness is an important element of financial stability. The measurement and monitoring of the financial institution soundness can help the early detection of the potential buildup of systemic risk that may lead to a financial crisis. Researchers proposed several approaches to assess financial stability. The most important ones are the structured approach, the reduced-form approach, the network analysis approach and the indicators approach.

The structured approach measures financial system risk by calculating joint default risk or portfolio credit risk (Avesani et al., 2006; Huang et al., 2009; Segoviano & Goodhart, 2009). This approach uses bank balance sheet information and market price information, such as credit default swap spreads and financial security prices, and derives the marginal distribution of risk using specific copula structures to obtain the joint default probability or portfolio credit. Tunyi and Ntim (2016) stress the role of institutional factors in explaining heterogeneous corporate restructuring activity in line with the resource-curse and managerial inefficiency approaches.

The reduced-form approach measures financial system risk by using a quantile regression to calculate conditional values at risk (CoVaR) (Adrian & Brunnermeier, 2016). CoVaR is defined as the change in the value at risk (VaR) of the financial system as a whole conditional on a financial institution being under distress relative to the financial system median state. Acharya et al. (2012) use the concept of systemic expected shortfall (SES) to measure the contribution of a single financial institution to financial system risk. They argue that the undercapitalization of the financial system as a whole implies undercapitalized individual financial institutions.

The network analysis approach uses information on two-way financial exposures and transactions in the balance sheets of financial institutions to establish a network of inter-relatedness among them. It measures financial system risk by simulating the accumulation of exposures and interactions of individual institutions. Chan-Lau et al. (2009) constructed network models to analyze the network externalities of a single bank. Billio et al. (2012) use Granger causality tests of financial asset returns to define the edges of a network of financial institutions and show that Granger causality networks are highly dynamic and become densely interconnected prior to systemic shocks.

The financial indicator approach uses both quantitative and qualitative information from the implementation of the Financial Sector Assessment program (FSAP), administered jointly by the International Monetary Fund and the World Bank. This information arises from balance sheets and other financial sources (International Monetary Fund, 2019). Related studies have used variants of this approach. For example, Borio and Drehmann (2009) apply simultaneous extreme value theory on pairs of property prices, equity prices and credit spreads, and construct an indication of financial system risk. Alessi and Detken (2011a) use a broad range of real and financial trends in 18 OECD countries between 1970 and 2007 and constructed simple early-warning indicators. Lane (2019) followed a data-driven approach examining the implications of monetary union for macro-financial stabilization policies for all countries of the Euro area in the period 2003–2012, focusing on 2007–2012 for the global crisis and 2010–2012 for the area-specific crisis. Finally, Quinn et al. (2011) make a survey of main indicators of financial openness and integration used to capture the relationship of financial openness or integration with economic growth.

These theoretical approaches to financial stability assessment have been variously used to produce early warning signals, macro-stress tests and financial stability indicators (Alessi et al., 2011; Cavalcante et al., 2016; Demyanyk & Hasan, 2009; Lin et al., 2012; Quagliariello, 2009). Each approach has its advantages and limitations. Policy makers and researchers have focused on alternative statistical indicators and used various combinations to identify and describe the vulnerabilities of the financial system and achieve accurate financial stability assessments. These approaches and their metrics have evolved over time to address the transition from the micro-prudential to the macroprudential dimension of financial stability.

3.2. Empirical findings

One strand of the literature analyzes the financial stability problem using standard empirical methods. For example, Holmfeldt et al. (2009) constructs a financial stress index based on the deviation of actual trends from their historical average of variables in the credit and money market for Sweden and the US from 1997 to 2007. The European Central Bank (2009) used a variance-equal weighted method to compute the Global Index of Financial Turbulence (GIFT) for 29 European economies that comprises six market-based indicators, which capture stress in fixed income, equity and foreign exchange markets. Slingenberg and Haan (2011) assessed the financial stability of thirteen OECD countries using a simple unweighted sum of standardized IMF indicators, where positive sum values indicate financial stress. They tested the predictive power of their index using a GARCH model, which produced moderate results for most countries. Cevik et al. (2013) analyzed financial stability in four Eastern European countries and Russia. They used a PCA method to analyze aggregated indicators of riskiness of the banking sector, the securities market, the foreign exchange market, the external debt market, sovereign risk and trade finance markets. Creane et al. (2006) collected 35 indicators from a survey on 20 Middle-East and North-Africa countries from 2000 to 2003 on six main themes: non-bank financial sector development, monetary and banking sector development, regulation and supervision, financial openness and institutional environment. They constructed a Financial Development Index summarizing the indicators by the means of a weighted average, with a subjective set of weights. Then a PCA-based set of weights was used only as a sensitivity test to reduce the reliance on qualitative judgments in selecting the most relevant weights to be assigned. Islami and Kurz-Kim (2013) used daily data of financial market indicators, such as CDS spreads, EUR/USD exchange rate volatility, 3-month and 10-year interest rates, to build a financial stability index (FSI) for all European countries. After standardizing and rescaling the time series, they produced a simple average FSI index with a daily and monthly horizon. They also used a single-equation error correction model to assess the index's predictive power relative to other benchmarks. Finally, Vermeulen et al. (2015) applied a similar approach to assess the financial stability of twenty-eight OECD countries and seven stock market indices and interest rates. They further used logit models and their index to predict binary outcomes (crisis vs. no crisis) and multinomial outcomes (banking vs. currency vs. no crisis).

Another strand of the literature analyzes financial stability using data-driven methods. Among them, decision trees have been variously deployed by early warning models to predict financial instability. Duttagupta and Cashin (2011) use a binary classification tree to analyze banking crises in 50 emerging markets and developed economies. The tree identifies the economic and financial conditions and thresholds under which induce the likelihood of a banking crisis. Policymakers only need to monitor whether the warning thresholds are crossed. Manasse et al. (2016) use cross-validation aggregating algorithms to test 540 predictor variables and identify two "danger zones" associated with banking crises: the first occurs when high deposit interest rates interact with credit booms and capital flows; the second occurs when an investment boom is financed primarily with banks' net foreign exposure. Tanaka et al. (2016) apply random forest methods to assess vulnerabilities in the banking sector, including bank-level financial statements as predictor variables. Ward (2017) uses random forest techniques to measure the risk of banking crises in 17 countries. He finds that tree produces a significantly better predictive performance relative to the logit regression. Alessi and Detken (2011b) develop an early warning system for systemic banking crises, which focuses on the identification of unsustainable credit developments. They consider 30 predictor variables for all EU countries and apply the random forest approach, showing that it outperforms competing logit models out-of-sample. Casabianca et al. (2019) use AdaBoost to identify the buildup of systemic financial crises in 100 advanced and emerging economies, which has a better predictive out-of-sample performance than logit models. Gabriele (2019) uses the same cross-validation aggregating algorithms to identify vulnerabilities to systemic banking crises, based on a sample of 15 European Union countries. He finds that high credit aggregates and a low market risk perception are amongst the key predictors. Bluwstein et al. (2020) show that machine learning techniques outperform network- and regression-based techniques in out-of-sample predictions and forecasting of financial crises, providing also a theoretical narrative based on probability decomposition.

Decision trees have been used to address sovereign crises issues. Manasse et al. (2003) compare the logit and the classification and regression tree approaches for sovereign crisis prediction and conclude that the latter overperforms the logit one considerably regarding the prediction of banking crises, but it also produces more false alarms. Manasse and Roubini (2009) use classification and regression trees to investigate the roots of sovereign debt crises and find that they differ depending on whether the country faces public debt sustainability issues, illiquidity, or various macroeconomic risks. Savona and Vezzoli (2012) adopt a mixed multi-step cross-validation aggregating algorithm and a non-linear regression approach to detect the most important risk indicators with the corresponding threshold at different stages towards default. Savona and Vezzoli (2015) use cross-validation aggregating algorithms to predict sovereign crises in the southern EU countries, which outperforms competing models. Arakelian et al. (2019) use a recursive partitioning approach to detect specific European sovereign risk zones and associated thresholds.

Decision trees have finally been used to predict currency crises. Gosh and Gosh (2002) use binary classification decision trees to predict 52 currency crises in a sample of 42 countries, and identify two different groups of key predictors for advanced and emerging economies, respectively. Frankel and Wei (2004) use regression trees to analyze macroeconomic crises in emerging markets based on the estimation of a different regression effect in accordance with tree splits. Kaminsky (2006) uses regression tree analysis to classify 96 currency crises in 20 countries based on their different stylized characteristics. Finally, Joy et al. (2017) use classification and regression trees and the random forest to predict currency crises and banking crises in 36 economies in the short and in the long run, based on the identification of different causes.

Moreover, sparse models, centered around the least absolute shrinkage and selection operator (LASSO) and the Bayesian operator, have also been used to address financial stability issues. Eidenberger et al. (2014) use Bayesian techniques and 30 predictor variables to develop an early warning system designed to predict an index of financial stress and identify the most significant ones based on probability. Ray-Bing et al. (2017) use a Bayesian hierarchical model to analyze the determinants of the 2008 global financial crisis, allowing for a joint treatment of variable and group selection, and stress the role of the different priors. Holopainen and Sarlin (2017) compare standard and data-driven techniques, including a logit LASSO, decision trees and the Random Forests, embedded in a number of early-warning models used to predict crises, with different results based on the different parametric assumptions made. Lang et al. (2018) use a logistic LASSO approach with out-of-sample data in combination with cross-validation of penalty parameters and test its predictive effectiveness on bank-level and macro-financial data, with good in-sample and out-of-sample signaling properties. Guidolin et al. (2019) use Bayesian models to estimate the effects of the US subprime crisis, based on panel data and cross-asset contagion measures. L et al. (2019) use the LASSO to predict sovereign crises based on the cross-section of sovereign credit default swaps spreads in a recursive setting that takes into consideration the time-varying sensitivities of market to economic fundamentals. They identify distinct crisis regimes associated with different dynamics.

Finally, empirical network models have also been used to predict financial stability based on the notion that the financial system is a complex system with network characteristics (Battiston, Caldarelli et al., 2016; Battiston, Farmer et al., 2016; Glasserman & Young, 2016). Minoiu and Reyes (2011) used systemically important nodes in a financial network to identify too-big-to-fail financial institutions in 184 countries, linked through cross-border lending flows. Since then, the literature has burgeoned. Recently, Schuldenzucker et al. (2020) used network models to analyze the type of systemic risk associated with situations in which it is impossible to decide which banks are in default. Fioramanti (2008) used artificial neural networks and supervised learning to develop an early warning system used to predict financial crises. However, unsupervised learning methods are rarely used in the literature (Sarlin & Peltonen, 2013).

3.3. Research design

Overall, financial stability analysis has been mostly single-country based and uses standard statistical methods for prediction. While data-driven methods have been recently used to analyze financial instability, these do not adequately capture policy considerations, they do not account for qualitative factors, and mostly ignore the dynamic (both cross-section and time-varying) conditions and cross-market/bank interdependencies. We fill the void by constructing a financial soundness index (FSIND) for both developed and developing countries using a data-driven methodological approach and credible and objective information provided by the International Monetary Fund's (IMF) Financial Soundness Indicators (FSIs), which incorporate institution-wide, market-wide and policy-level considerations. Our data-driven approach uses advanced principal component analysis to deal simultaneously with the presence of strong cross-section and time dependence in the data due to unobserved common factors. Our financial soundness index is fully model-based and data-driven, properly tested and validated by using alternative machine learning techniques, and accounts for the role of qualitative factors in the different countries. In so doing, we analyze the relative efficiency of the data-driven algorithms, the credibility of the data source and the propriety of variable selection. We make the index values available in both continuous and binary formats to accommodate alternative policy needs and research strategy specifications. A similar approach has been successfully applied to a different data-driven problem regarding the quantification of the countries' preparedness in facing the risk of epidemiological harms (Bitetto et al., 2021).

4. Methodology

As already discussed, macroeconomic variables are often used to assess financial stability of countries. A common way to summarize information from these variables is to create synthetic indexes based on assumptions taken by financial institutions experts, typically resulting in the usage of a weighted average. However, these measures are subjective by nature and thus such indices can be questionable, leading to endless debate on which one should be used as a robust financial indicator.

In this paper, we want to present a data-driven statistical approach to building a financial index based on intrinsic information. We analyze a set of Financial Soundness Indicators (FSI) provided by the International Monetary Fund ranging from 2006 to 2017 for all available countries that span the globe, including both strong and developing economies. First, we assess the data quality and cope with issues related to the presence of missing data. Then we take advantage of a statistical methodology to build the index: Principal Component Analysis (PCA). By means of PCA, we create a low dimensional (1 to 2 way) continuous indicator, explaining the variance of the data at the highest possible level and considering each year separately. We subsequently set an appropriate threshold to the PCA-based indicator that allows us to assign a dichotomous label (stable vs. unstable) to each country. As a result, we produce an additional binary measure of our index, which offers an easy-to-use quantity for various purposes, like classification, clustering and comparison of countries. Our binary measure should be considered as a new, additional measure rather than a substitute of the continuous one. To ensure reliability, we construct both the continuous and binary measures of our index in a statistically robust manner that does not rely on subjectivity.

The aim of our analysis is to extract synthetic indicators that summarize at best the relationship among variables in a lower dimensional space. One of the most common methodology used for dimensionality reduction is Principal Component Analysis (*PCA*).

Briefly, *PCA* aims at creating one or more new components from a larger set of observed variables, where each component is a linear combination of the Y original variables. The model is represented by the following equation:

$$C_1 = w_1 Y_1 + \cdots + w_p Y_p \tag{1}$$

where C_1 is the new first principal component obtained as the linear combination of Y_i that are the original variables and w_i that are the weights of the combination. The following C_k components are built similarly.

We recall that our dataset has three dimensions, *Country*, *Variable* and *Time*, so we assume to apply the previous dimensionality reduction technique in the following way: we model country/variable interaction for each year. Thus, PCA has been evaluated on each year separately, resulting in T models. For sake of stability and robustness, we also decided to evaluate and compare three different PCA techniques: regular PCA, Robust PCA and Robust Sparse PCA.

According to the definition, PCA aims at finding new and linear-wise combinations of the original data, in a way that the amount of explained variance of the data is maximized. Those combinations are mathematically constrained to be mutually orthogonal (that is independent) and are called Principal Components (PC) or loadings. Given a $n \times p$ data matrix $\mathbf{X}$, where n is the number of observations and p is the number of variables, we want to find the $k \times p$ Principal Component matrix C, with usually $k \ll p$ such that the projected data matrix $W = XC^T$, also called scores matrix, will have dimension $n \times k$. The problem can be seen as:

$$\begin{aligned} &\underset{\mathbf{C}}{\text{minimize}} && \|\mathbf{X} - \mathbf{X}\mathbf{C}\mathbf{C}^T\|_F^2 \\ &\text{subject to} && \mathbf{C}^T\mathbf{C} = \mathbf{I} \end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius norm. We implement the model using R package `prcomp`.

If we want a robust estimation of the Principal Components, we can decompose the data matrix X into a low rank component L that represents the intrinsic low dimensional features and an outlier component S that captures anomalies in the data. The problem can be then solved by:

$$\begin{aligned} &\underset{\mathbf{L},\, \mathbf{S}}{\text{minimize}} && \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \\ &\text{subject to} && \mathbf{L} + \mathbf{S} = \mathbf{X} \end{aligned}$$

where $\|L\|_.$ is the nuclear norm and λ is a penalization term. Once fitted, $\mathbf{L}$ can be used as a proxy for $\mathbf{X}$ but cleaned up by extreme values. We implement the model as described in Candes et al. (2009).

As a further improvement, if we want a robust estimation and a sparse representation of the Principal Components, we can add a sparsity constraint on matrix C. The problem can be then solved by:

$$\begin{aligned} &\underset{\mathbf{C},\, \mathbf{W}}{\text{minimize}} && \|\mathbf{X} - \mathbf{W}\mathbf{C}^T - \mathbf{S}\|_F^2 + \psi(\mathbf{C}) + \phi(\mathbf{W}) + \lambda\|\mathbf{S}\|_1 \\ &\text{subject to} && \mathbf{C}^T\mathbf{C} = \mathbf{I} \end{aligned}$$

ψ and ϕ are regularizing functions (i.e. LASSO or Elastic Net) as described in Erichson et al. (2018).

The final index, hereinafter referred to as Financial Soundness Index (FSIND), will be based upon the scores matrix W that is k-dimensional. We aim to select the number of components k as a result of a trade-off between maximal explained data variance and smallest value of k (i.e. principal components).

By construction, PCA produces a continuous output vector of size n for each of the k selected principal components, also known as scores vectors. We assume that each principal component, according to its natural rank order (first component, second component, etc.), is a candidate variable for representing our Financial Soundness Index. The rationale behind is the following: the first component, being the most representative one by construction, is the building block of the index. Consequently, according to the amount of explained variability of the first principal component, we evaluate whether a second one is worth of being considered. Ideally the process iterates until no more advantage is foreseen according to the trade-off variability explained-minimal number of components.

We produce a k-dimensional continuous FSIND per country-year pair. At the same time, we are aware of the importance for policy makers and institutions to work with clear and intuitive measures. Thus, we pay particular attention on the choice of the thresholds necessary to produce binary indices. Since there is no a-priori information on the best thresholding value, we implement a process to get the most reliable value.

The procedure for obtaining the threshold that guarantees a minimal level of subjectivity is composed of two steps. We firstly set a candidate threshold and get the relative binary index, i.e. 0 or 1 labels. Then we perform a bunch of regressions using *Random Forest* and *Gradient Boosting Machine* algorithms where the target is an economic variable deemed as relevant for financial analysis (such as *GDP* or *Non Performing Loans*) and the regressors are the binarized principal components. We choose the aforementioned algorithms because their iterated binary splitting criteria are the most suitable for the dichotomous nature of the regressors. Finally, we evaluate prediction accuracy and outliers for different thresholds. The procedure is iterated with different values and the most appropriate one is selected according to prediction accuracy and outliers stability.

Finally, we validate the performance of our FSIND comparing its ranking with the IMF's Financial Institutions Efficiency (FIE) index and Financial Markets Efficiency Index (FME), which measure the ability of institutions to provide financial services at low cost with sustainable revenues, and the level of activity of capital markets, respectively.[1] In particular, the Financial Development Index has several nested versions: its highest level consists of a single dimension that is Financial Development, the middle level has two dimensions that are Financial Institutions (FI) and Financial Markets (FM), and the lowest level has three additional separations, i.e. Depth, Access and Efficiency, for both FI and FM. Given the scope of our index, we focused on the lowest level specification of efficiency for both FI and FM.

5. Dataset and preprocessing

The dataset consists of 17 Financial Soundness Indicators (FSI) provided by International Monetary Fund for 140 countries, spanning from 2006 to 2017. Tables 1 and 2 present the summary statistics of the index's constituent variables 1 to 17 and their pairwise correlations.

However, some countries present too many missing values, as well as, less than expected years. As a consequence, we decide to restrict our analysis on a subset of 119 countries from 2010 to 2017, selected with a NA incidence tolerance not exceeding 30%. The complete list of selected countries and relative percentage of NA is reported in AppendixA. Since the presence of many missing values can extremely impact the quality and the reliability of results, we set a protocol of missing values treatment and imputation.

As stated above, according to our protocol, missing values for the selected countries are still present, in fact 16 out of 119 countries show a percentage of NA between 20%–29%. A complete overview of selected countries and their missing values percentage is shown in Fig. 1. In order to deal with missing values and to apply further methodologies in a robust way, we test two different data imputation techniques: Matrix Completion with Low Rank SVD (MC-SVD) (Hastie et al., 2015) and Bayesian Tensor Factorization (BTF) (Khan & Ammad-ud-din, 2016)

[1] https://www.imf.org/external/pubs/ft/wp/2016/wp1605.pdf.

Table 1
List of used variables with sources, frequency, total number of observations, number of missing values and descriptive summary statistics.

Source	Variable	Frequency	Total observations	Missing values	Min	Max	Mean	Standard deviation	Variation coefficient
	1 - EMB Capital to assets			63	1.49	24.85	10.28	3.57	0.35
	2 - EMB Customer deposits to total non interbank loans			113	29.01	626.93	120.73	56.5	0.47
	3 - EMB Foreign currency liabilities to total liabilities			193	0	100	30.61	24.87	0.81
	4 - EMB Foreign currency loans to total loans			176	0	100.06	28.75	26.26	0.91
	5 - EMB Personnel expenses to non interest expenses			93	5.29	91.58	44.17	12.04	0.27
	6 - Interest margin to gross income			21	−294.33	142.77	59.01	18.4	0.31
	7 - Liquid assets to short term liabilities			79	10	690.37	69.13	61.11	0.88
	8 - Liquid assets to total assets			50	4.99	74.97	27.92	13.03	0.47
FSI	9 - Net open position of foreign exchange to capital	Yearly	952	221	−95.43	407.97	9.57	36.74	3.84
	10 - Non interest expenses to total income			21	−303.46	115.79	58.17	17.88	0.31
	11 - Non performing loans net of capital provisions			21	−51.61	413.56	18.78	38.28	2.04
	12 - Non performing loans to total gross loans			23	0	54.54	6.81	7.4	1.09
	13 - Regulatory capital to risk weighted assets			19	1.75	42.2	17.67	4.83	0.27
	14 - Regulatory tier 1 capital to risk weighted assets			24	2.18	40.3	15.43	4.86	0.31
	15 - Return on assets			21	−25.61	10.28	1.5	1.8	1.2
	16 - Return on equity			24	−505.64	65.4	13.22	21.93	1.66
	17 - Sectoral distribution of loans residents			127	20.67	100	87.85	16.05	0.18

Briefly, MC-SVD solves the minimization problem $\frac{1}{2}\|X - AB^T\|_F^2 + \frac{\lambda}{2}(\|A\|_F^2 + \|B\|_F^2)$ for A and B where $\|\cdot\|_F$ is the Frobenius norm by setting to 0 the missing values. Once estimated, AB^T can approximate the original matrix X, including the missing values. This is applied to the 2-dimensional "slice" of countries-FSI for each year.

BTF acts in a similar way but using a tensorial decomposition of the 3-dimensional tensors that stacks all the annual slices together so that the imputation process involves information coming from a temporal dimension as well.

To assess imputation performances and to choose the best method, we test the algorithm in three settings. In the first, referred to as *Original* or setting a, we consider the whole dataset made of 119 countries by 17 variables for 8 years for a total of 16 184 entries. It contains 8% of missing values, thus we randomly remove some additional values representing 10%, 20% and 30% of the initial dataset. In the second, referred to as *No missing* or setting b, we drop all entries with missing values and apply the same incremental sampling procedure on the remaining subset. In the last, referred to as *Some missing* or setting c, we drop all countries with at least 3 missing values for any year and apply again the incremental sampling procedure on the remaining subset. Furthermore, we fit the two methods, MC-SVD and BTF, on the previous 3 cases (a, b and c) with different sampling percentages and we evaluate the Mean Absolute Reconstruction Error (MARE) on the excluded observations as follows:

$$MARE = \frac{1}{M}\sum_{i}^{M}|x_{excluded} - x_{reconstructed}|$$

where M is the total number of excluded values. Moreover, we check the sensitivity to the original percentage of missing values by comparing the MARE based on the *No missing* and *Some missing* settings with the one based on the *Original* setting.

After having imputed missing data, in order to ensure the adequate sample size suitable for the presented methodologies, we run the Kaiser–Meyer–Olkin test (Kaiser, 1970) resulting in the large score of 81.9% and 82.7% for MC-SVD and BTF respectively. Moreover, we check for stationarity of each FSI-country pair over the time span. We perform standard Augmented Dickey–Fuller and Ljung–Box test and since some non-stationarity is revealed, we integrate all time series with lag 1, in order not to sacrifice too many observations. Additionally, we run the Im–Pesaran-Shin test (Im et al., 2003) obtaining p-values $p \ll 0.01$ for both model specifications, i.e. "individual intercepts" and "individual intercepts and trends" for the underlying Augmented Dickey–Fuller test, implying the acceptance of alternative hypothesis of stationarity for the input variables time-series. Therefore, the final dataset consists of 17 variables for 119 countries and 7 lagged years.

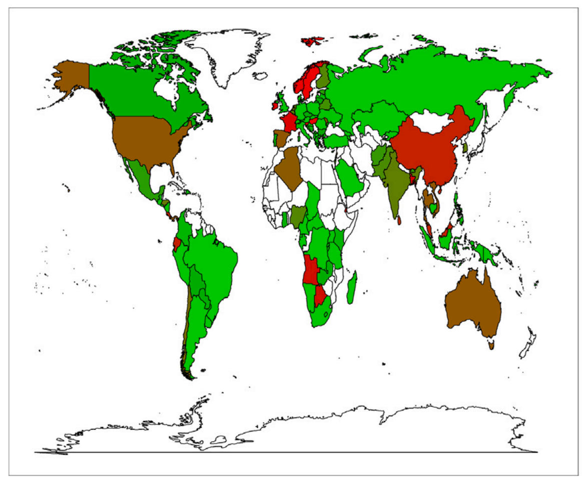


Fig. 1. Map of analyzed countries and missing values percentage distribution by color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Data analysis

As described in Section 5, we assess the cross-validation error of the data imputation techniques. Figure B.8 in Appendix B reports the results of the imputation performance for both techniques. The blue shaded bars in the upper row represent the average reconstruction error for different percentage of additional missing values for each of the three settings. Whiskers on top of each bar show the scaled magnitude of maximum value of reconstruction error. Bars on the lower row represent the percentage variation of the average reconstruction error of the *No missing* (setting b) and *Some missing* (setting c), settings compared to the *Original* (setting a). Green bars signal that the imputation technique has a lower average reconstruction error. The figure shows that when comparing, setting b and setting c with setting a, the method is performing better when considering data with less missing values, as expected. On the other hand, red bars mean that the technique fails in improving the reconstruction performance on subsets with less missing values. Overall, we find that Bayesian Tensor Factorization performs better.

Consequently, we standardize the dataset for each year and then we apply the PCA method described in Section 4. Results for the PCA approach can be found in Table 3 where the average variance explained by loadings over all years is reported, as well as the average R^2 both on the whole dataset and on subsets with values trimmed for the 95th

Table 2
Correlation matrix of input variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	−0.12*															
3	−0.18*	0.82*														
4	−0.10*	0.48*	0.61*													
5	0.00	0.00	−0.06	−0.08*												
6	0.07*	−0.91*	−0.75*	−0.45*	0.03											
7	−0.10*	0.61*	0.52*	0.32*	−0.09*	−0.58*										
8	−0.15*	0.92*	0.76*	0.44*	0.00	−0.84*	0.66*									
9	−0.15*	0.03	0.04	0.13*	0.02	−0.01	0.00	0.03								
10	0.05	−0.89*	−0.75*	−0.45*	−0.16*	0.87*	−0.51*	−0.80*	−0.02							
11	−0.05	0.03	−0.02	0.05	−0.03	−0.01	0.04	0.00	0.04	−0.01						
12	0.16*	−0.06	−0.08*	0.01	−0.05	0.03	0.02	−0.04	0.05	0.04	0.62*					
13	0.49*	−0.06	−0.15*	−0.09*	0.11*	0.06	0.01	−0.03	−0.20*	0.06	−0.04	0.05				
14	0.49*	−0.05	−0.12*	−0.11*	0.09*	0.06	0.02	−0.01	−0.15*	0.07*	−0.07*	0.00	0.91*			
15	−0.08*	0.00	0.00	−0.08*	0.04	0.02	−0.05	−0.03	−0.05	0.00	−0.19*	−0.35*	0.00	0.07*		
16	−0.17*	0.02	0.02	−0.05	0.02	0.03	−0.04	−0.01	−0.04	0.01	−0.19*	−0.42*	0.02	0.11*	0.88*	
17	−0.09*	0.03	0.02	0.00	−0.02	0.04	0.04	0.02	−0.08*	0.04	0.03	−0.24*	0.10*	0.17*	0.43*	0.57*

Legend is below: 1 'Emerging Markets Bond (EMB) Capital to assets', 2 'Customer deposits to total non interbank loans', 3 'EMB Foreign currency liabilities to total liabilities', 4 'EMB Foreign currency loans to total loans', 5 'EMB Personnel expenses to non interest expenses', 6 'Interest margin to gross income', 7 'Liquid assets to short term liabilities', 8 'Liquid assets to total assets', 9 'Net open position of foreign exchange to capital', 10 'Non interest expenses to total income', 11 'Non performing loans net of capital provisions', 12 'Non performing loans to total gross loans', 13 'Regulatory capital to risk weighted assets', 14 'Regulatory tier 1 capital to risk weighted assets', 15 'Return on assets', 16 'Return on equity' and 17 'Sectorial distribution of loans residents'.
*p-val < 0.05.

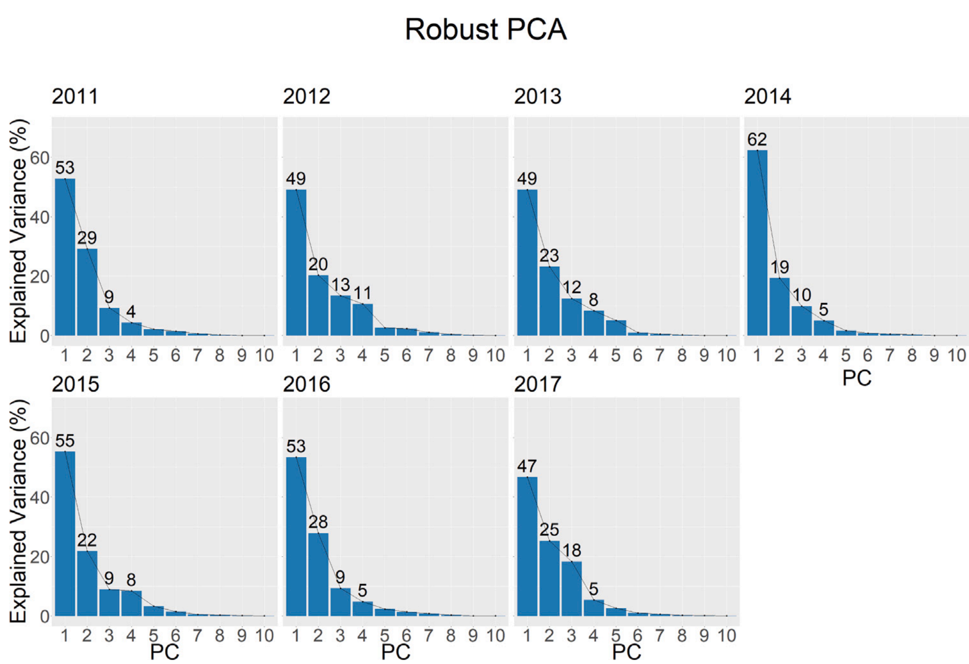
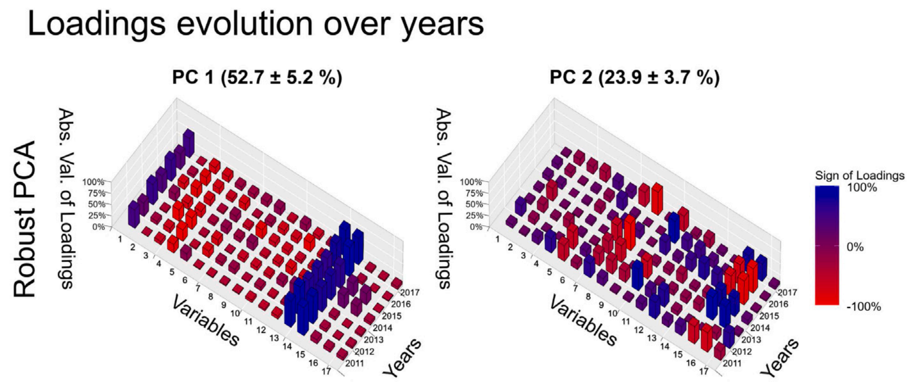


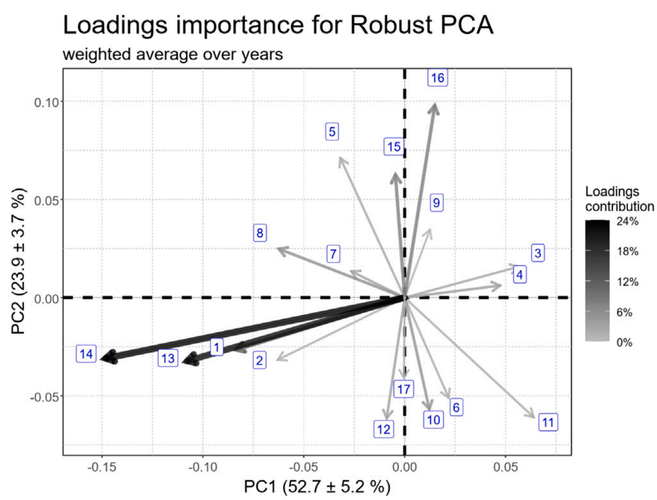
Fig. 2. Scree plot for Robust PCA method for each year. The first two components account for an average 76% of cumulative explained variance.

and 99th percentiles to check for outliers impact. In our context R^2 means the ratio of the amount of variance explained by our retained components over the total variance contained in the original variables. Moreover, we run the Im–Pesaran–Shin test on the PCA index and $p - values \ll 0.01$ for all model specifications ensure its stationarity. The stationarity is important because we can infer that the changes over time, which the index is expected to capture, can be statistically robust and not caused by any trend in the data or mean-reversion effects.

The results show that the robust PCA method performed best regardless of the employed data (full data set, 1% trimmed and %5 trimmed), both in terms of explained variance and of R^2. Given the achieved values, we retain only the first two principal components, which account for a cumulative explained variance of 76% (such choice allows also for a visually interpretable FSIND index). Fig. 2 shows the scree plots of the variance explained by the loadings using the robust PCA method only along the seven considered years. On average the first component accounts for the 50% of the total variability, the only significant exception is represented by 2014 which spikes at more than 60%. Fig. 3(a) reports the loadings of the first two principal components and their variations along the years. Each principal component can be seen as a weighted average of the original 17 variables where the weights are the loadings, which are normalized between 1 and −1. Therefore, loadings disclose how much each component is influenced by the original variables. The first component has a stable behavior over the years, mainly driven by Emerging Markets Bond (EMB) Capital to assets (1), Regulatory capital to risk weighted assets (13) and Regulatory tier 1 capital to risk weighted assets (14) with a strong positive correlation (in blue) and EMB Foreign currency liabilities to total liabilities (3) and EMB Foreign currency loans to total loans (4) with a minor negative correlation (in red). The second component has a year-varying behavior for the most important variables such as Liquid assets to short term liabilities (7), Liquid assets to total assets (8), Non interest expenses to



(a) Loading importance over years for Robust PCA method, i.e. magnitude of weights of the linear combination that defines each component. Blue shaded bars represent positive contribution to each component loading while red shaded bars a negative one. The higher the bar the more the original variable contributes to the loading.



(b) Combined effect of loadings for the first two components for Robust PCA method. Values over years are averaged and weighted by yearly loadings importance. Darker and thicker arrows represent the original variables that contributes the most to loadings importance for each of the two PCA components.

Fig. 3. Legend is below: 1 'Emerging Markets Bond (EMB) Capital to assets', 2 'Customer deposits to total non interbank loans', 3 'EMB Foreign currency liabilities to total liabilities', 4 'EMB Foreign currency loans to total loans', 5 'EMB Personnel expenses to non interest expenses', 6 'Interest margin to gross income', 7 'Liquid assets to short term liabilities', 8 'Liquid assets to total assets', 9 'Net open position of foreign exchange to capital', 10 'Non interest expenses to total income', 11 'Non performing loans net of capital provisions', 12 'Non performing loans to total gross loans', 13 'Regulatory capital to risk weighted assets', 14 'Regulatory tier 1 capital to risk weighted assets', 15 'Return on assets', 16 'Return on equity' and 17 'Sectorial distribution of loans residents'. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

total income (10), Return on assets (15) and Return on equity (16). The first component takes into account for the long-term economic drivers, thus less subject to variation over the relatively short time period considered, whereas the second component accounts for short term indicators usually more responsive to market shocks. Thus far, we can conclude that the 2-dimensional index should be able to discriminate Financial Soundness based on two different but complementary factors. Fig. 3(b) shows the bi-plot referred to the two first components in which values over years are averaged and weighted by yearly loading importance. Figure C.11 in the Appendix details the yearly deviation from the average; percentage variations are between −4% and 4% that is small enough to consider the average values employed in Fig. 3(b) adequately representative of the overall index pattern. Figures from C.9 to C.13 in the Appendix report the scree plots and loading importance plots for other implemented PCA methods. FSIND can be found at https://github.com/AleBitetto/FSIND.

Once the continuous FSIND is estimated, we find the optimal threshold to binarize our index. As described in Section 4, we test thresholds for both components ranging from −1.5 to 1.5 with a step of 0.5 (for a total of 7×7 combinations). We use as target variable for our validation models five macroeconomic indices from World Development Indicators (WDI)[2] namely: bank non-performing loans to total gross loans ratio (percent), Consumer Price Index, GDP per capita annual growth (percent), gross domestic savings (percent of GDP), unemployment (percent of total labor force), population annual growth (percent).

First, we evaluate the stability of outliers using the extreme values of Absolute Percentage Error (APE) on predicted target values, according to the Generalized Extreme Studentized Deviate test (Rosner, 1983). We subsequently check their consistency by counting the total number of detected outliers for each threshold and their shared percentage across all thresholds. Secondly, we compare the performance of the models for each threshold according to APE in order to assess stability. As a further comparison, we also add APE performance of models fitted

[2] http://wdi.worldbank.org/table.

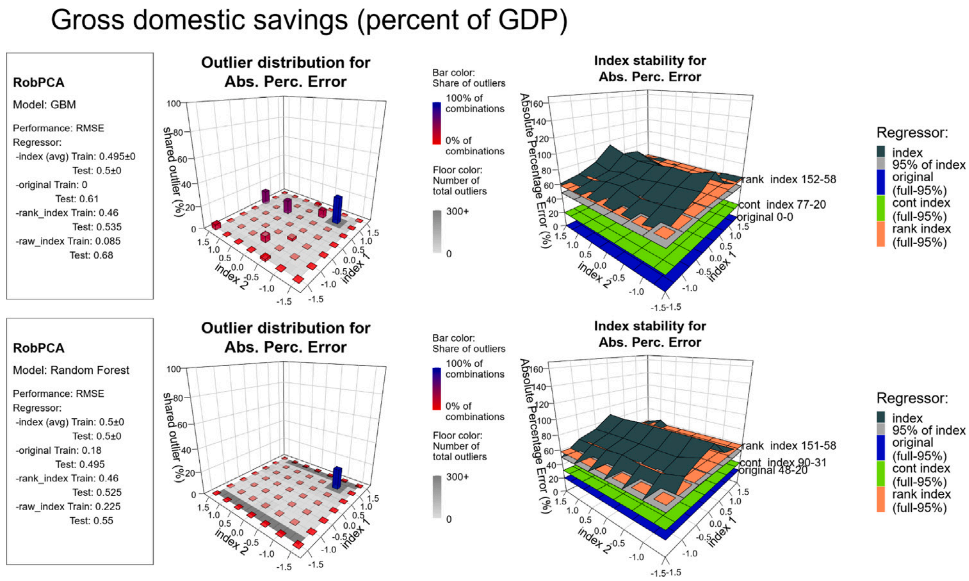


Fig. 4. Index validation for gross domestic savings. The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: floor color shows the total amount of detected outlier and bars height measure the percentage of shared outliers among all threshold combinations. The right surface plots depict how predicted values performance, i.e. Average Percentage Error (APE), changes over the input regression settings: dark and light gray surfaces represent the binary FSIND performance over all dataset and trimmed top 5th quantile values respectively, blue surface is for the original 17 FSI variables, green surface is for the continuous FSIND and orange is for the ranked FSIND. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

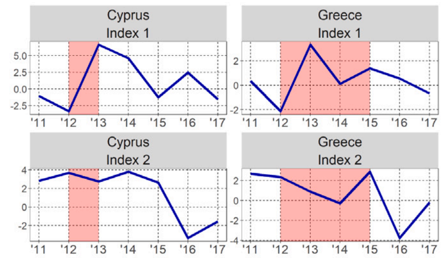
Table 3
Results based on the three different PCA. The first two principal component are provided and evaluated in terms of mean explained variance, mean R^2 and mean R^2 trimmed the top 1st and 5th percentile.

Method	Number of PC	Mean Expl.Variance	Mean R^2	Mean R^2 on 99th	Mean R^2 on 95th	Im-Pesaran-Shin test
PCA	1	$22.2 \pm 6.1\%$	$22.2 \pm 6.1\%$	$26.6 \pm 15.3\%$	$33.3 \pm 20.6\%$	$\ll 0.01$
	2	$37.9 \pm 9.8\%$	$37.9 \pm 9.8\%$	$43.6 \pm 14.8\%$	$50.6 \pm 15.5\%$	$\ll 0.01$
RobPCA	1	$52.7 \pm 5.2\%$	$98.2 \pm 0.7\%$	$98.4 \pm 0.6\%$	$98.7 \pm 0.5\%$	$\ll 0.01$
	2	$76.5 \pm 5.3\%$	$99.1 \pm 0.4\%$	$99.2 \pm 0.3\%$	$99.3 \pm 0.3\%$	$\ll 0.01$
RobSparPCA	1	$22.3 \pm 6.1\%$	$15.8 \pm 2.5\%$	$21.9 \pm 8.3\%$	$28.1 \pm 4\%$	$\ll 0.01$
	2	$38 \pm 9.9\%$	$28.2 \pm 4.1\%$	$36.9 \pm 8.2\%$	$47.7 \pm 6\%$	$\ll 0.01$

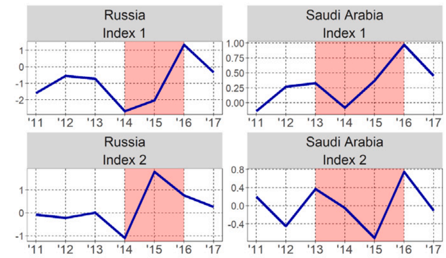
using as regressors (a) the initial dataset variables used to build the FSIND (17 FSI) (*original* case), (b) the continuous FSIND (*cont index* case) and (c) the continuous FSIND discretized into intervals, i.e. we create two categorical variables, one for each FSIND dimension, with 8 values each (*rank index* case) corresponding to the ranges identified by the same 8 candidate thresholds.

We use the gross domestic savings (percent of GDP) as our first validation variable. Fig. 4 shows the results. The left panel reports the training vs test fitting performance, i.e. Root Mean Squared Error (RMSE), in order to check model robustness and avoid overfitting: small gaps between train and test indicates an absence of overfitting. The central bar plots display outliers stability: the gray scale on the floor highlights the total count of outliers for each threshold combinations, the darker the higher. The height and color of bars represent the percentage of shared outliers, i.e. how many observations are marked as outlier in every threshold combinations. Higher bars with darker floor color mean that the maximum amount of detected outliers are equally shared across all combinations, resulting in the index ability to detect and isolate outliers, regardless of the threshold combination. Finally, right surface plots depict how predicted values performance, i.e. APE, changes over the aforementioned input regression settings: dark and light gray surfaces represent the binary FSIND performance over all dataset and trimmed top 5th quantile values respectively, blue surface is for the original 17 FSI variables, green surface is for the continuous FSIND and orange is for the ranked FSIND. Low surface height means good predicted values accuracy and flat surface means performance stability over different threshold combinations. Figure D.14 to D.18 in the Appendix report the results for the remaining target variables. Since the resulting performance proved to be stable across all threshold values for all tested variables, we use the 0 threshold for both components of FSIND.

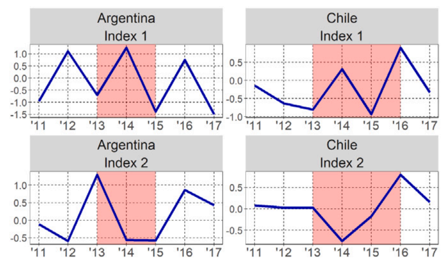
Fig. 5 shows the evolution over years of the continuous FSIND for selected countries. In particular, Fig. 5(a) shows that Greece and Cyprus have similar patterns as their financial systems are interdependent. The FSIND trend rises during the 2012–2015 bail-out of Greece and drops in the 2016 in response to the country's recovery. The FSIND trend also rises during the 2012–2013 financial crisis in Cyprus, as highlighted by the spike of the first component of the index. In Fig. 5(b) Saudi Arabia and Russia show similar patterns between their two indices and between each other because both countries are key oil exporters affected by similar risks and world events. Indeed, raise of both components matches the 2014–2016 Russian financial crisis as well as the 2014–2016 Saudi crisis due to the collapse of oil price and its impact on GDP. Similar behavior is shown in Fig. 5(c) because both Argentina and Chile suffer from structural deficiencies and economic turmoils. FSIND is able to capture the Chilean financial and institutional crisis, between 2013 and 2016, caused by a drop in copper price, of which Chile is a major exporter. Also Poland, Ukraine and Russia have the same behavior because they are all economically interdependent through the energy distribution network, common cultural origins and share similar geopolitical concerns as shown in Fig. 5(e). A spike on the first component of FSIND outlines the 2011–2012 crisis in Poland as a consequence of the European global one. It also shows the 2013–2015 Ukrainian debt crisis while struggling to cope with the fallout from Crimea's



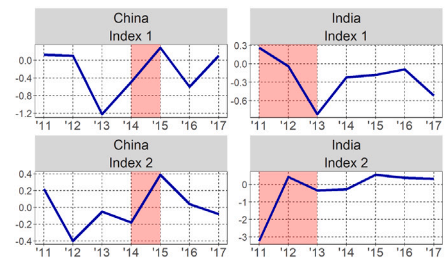
(a) Greece vs Cyprus



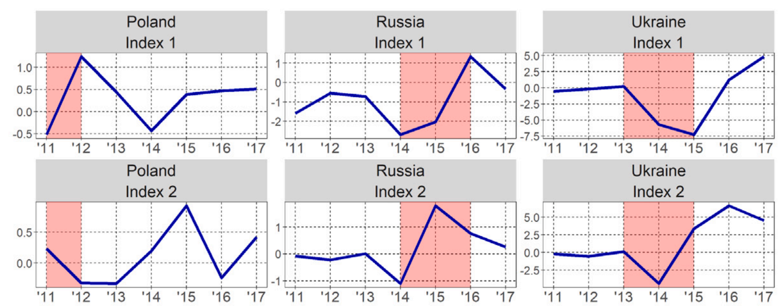
(b) Saudi Arabia vs Russia



(c) Argentina vs Chile



(d) India vs China



(e) Poland vs Ukraine vs Russia

Fig. 5. Continuous FSIND evolution over years for some countries with similar pattern due to common cultural, political and financial background. Financial crisis are shaded in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

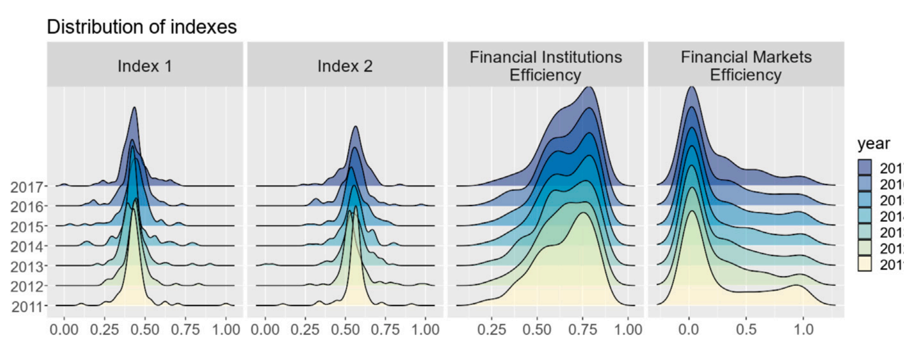


Fig. 6. Comparison of the two FSIND dimensions with Financial Institution Efficiency and Financial Markets Efficiency index distributions.

annexation by Russia and continuing war with pro-Russian separatists. Fig. 5(d) shows that India and China have similar patterns for index 1 and closely for index 2 as they both have to deal with similar problems of overpopulation, environmental pollution and both followed rapid credit expansion policies in recent years to accommodate the need of fast income growth and inequality reduction. Indeed, the increasing trend of the second component of the index captures the 2011–2013 Indian stock-market crisis, followed by Moody's rating downgrade and high inflation as well as the stock-market crash that struck China in 2014–2015, clearly highlighted by spikes in both FSIND component. Figure D.19 to D.22 in the Appendix report the complete list of all countries.

We compare the country ranking defined by our 2-dimensional FSIND with the Financial Institution Efficiency (FIE) and Financial Markets Efficiency (FME). First, we compare the distribution of each single index, as shown in Fig. 6. Then, for each FSIND component and for both FIE and FME, we evaluate 20 quantiles, i.e. one every 5%, and split the index values in 17 rolling bins, each of which spans between five quantiles, i.e. a 20% range, and a shift factor of 5%. For example, the first bin contains observations, i.e. countries, with index values between the 0th and 20th quantiles, the second between 5th and 25th ones and

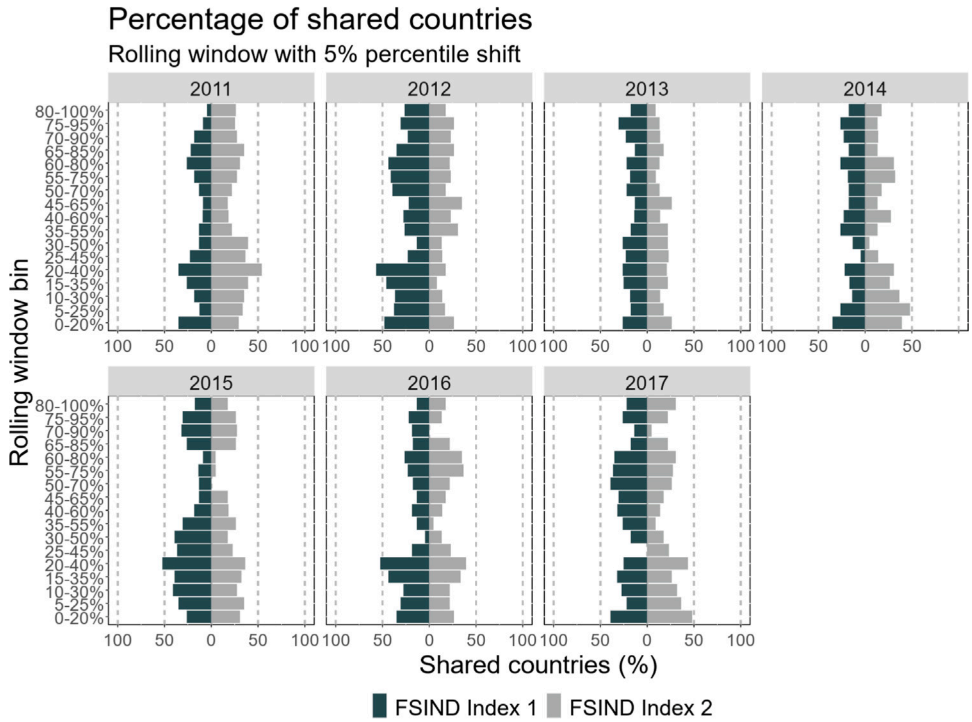


Fig. 7. Ranking ability comparison for FSIND components and Financial Markets Efficiency (FME) index. For each FSIND component and FME, 20 quantiles, i.e. one every 5% are evaluated, splitting the index values in 17 rolling bins, each of which spans between five quantiles, i.e. a 20% range, and a shift factor of 5%. For each of the corresponding bin of the three indexes the percentage of the shared countries with FSIND first (in black) and second (in gray) component.

so on. Finally, for each of the corresponding bin of the four indexes we evaluate the percentage of shared countries, checking how many countries have the same ranking. Fig. 7 reports the results based on the FME comparison, showing an average percentage of shared countries close to 50% over the years, meaning that both FSIND components have a good matching with the reference index. Figure E.23 in the Appendix shows a less matching percentage with the FSIND components, mainly due to the strong market-based relationship highlighted by the most important loadings, i.e. index drivers, described above.

7. Conclusions

In this paper, we address the challenging issue of assessing the overall soundness of financial institutions in a country and thereby contributing to the understanding of global financial stability considerations. Predicting financial instability is necessary for providing warnings against impending abnormalities and for taking precautionary action. The ability to detect complex and nonlinear relationships, not fully biased by a priori theory/beliefs, has contributed to the increasing use of data-driven methods as an advance over the traditional statistical models. We have contributed to this strand of research by developing a new data-driven financial soundness index for 119 countries, with very good results. However, we believe that our measure of financial system soundness should be used in combination with existing model-based and theory-driven approaches (Athey & Imbens, 2019). Mixed approaches that include pure algorithmic techniques that do not make specific assumptions about the data generating process, and parametric regression techniques have advantages for they can embed the analysis in a theoretical context. They can allow the overcoming of the problem of reliability of the predictive model, by using advanced data-driven techniques, while at the same time allowing the estimation of distance-to-default using standard regression techniques.

We use data from 119 developed and developing countries and a fully model-based and data-driven methodological approach to produce a financial soundness index (FSIND) at the country level. Our data includes seventeen financial soundness indicators from the IMF's core indicators set during the 2010–2017 period.

We address the problem of missing data and misaligned information by applying a process of complete assessment of data quality and multilevel data imputation. Subsequently, we apply a dimension-reduction model based on alternative robust PCA algorithms to produce different versions of our index. First, we produce a synthetic binary measure of our FSIND index that can be used to identify a stable vs unstable financial system in each country. Secondly, we produce a synthetic continuous measure of the FSIND index.

We validate the FSIND index by selecting suitable thresholds based on the criterion of best performance obtained by applying alternative regression models. We subsequently use the FSIND index as a key regressor to predict outcome variables, such as bank non-performing loans to total gross loans ratio (percent), Consumer Price Index, GDP per capita annual growth (percent), gross domestic savings (percent of GDP), unemployment (percent of total labor force), population annual growth (percent). The results show that our approach to index construction summarizes well the contribution of key factors affecting the soundness of financial institutions and captures well the dynamics of the financial system position over time.

Our index has a meaningful interpretation. It uses credible information on financial institution characteristics and accounts adequately its relative contribution to aggregate variation. From the inspection of the country specific patterns of the index, we find out its ability of detecting economic and financial crisis periods. Moreover, the binary and continuous versions of our index facilitate alternative uses aiming at classification or ranking assessment of financial systems around the world.

Our validation process has shown that the index makes robust predictions. However, more future work would be needed to assess its full prediction potential. We are aware of the limitations of the present results. For instance, we have not fully considered the interactions along the temporal dimension since we evaluate each year's effect separately from that of the others. Future analysis should take advantage of native temporal models, which consider and elicit the whole temporal horizon.

Declaration of competing interest

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

Acknowledgments

This research has received funding from the European Union's Horizon 2020 research and innovation program "PERISCOPE: Pan European Response to the ImpactS of COvid-19 and future Pandemics and Epidemics", under the Grant Agreement No. 101016233, H2020-SC1-PHE-CORONAVIRUS-2020-2-RTD.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.irfa.2022.102451.

References

Abdou, H. A., Abdallah, W. M., Mulkeen, J., Ntim, C. G., & Wang, Y. (2017). Prediction of financial strength ratings using machine learning and conventional techniques. *Investment Management and Financial Innovations*, *14*, 194–211.

Abdou, H., Tsafack, M., Ntim, C., & Baker, R. (2016). Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*, [ISSN: 0950-7051] *103*, 89–103. http://dx.doi.org/10.1016/j.knosys.2016.03.023.

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. P. (2012). *Measuring systemic risk*: *CEPR discussion paper, DP8824*.

Adrian, T., & Brunnermeier, M. K. (2016). Covar. *American Economic Review*, *106*(7), 1705–1741.

Alaka, H. A., Oyedele, L., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, *94*, 164–184, URL https://uwe-repository.worktribe.com/output/870618.

Alessi, L., Antunes, A., Babecký, J., Baltussen, S., Behn, M., Bonfim, D., Bush, O., Detken, C., Frost, J., Guimarães, R., Havránek, T., Joy, M. M., Kauko, K., Matj, J., Monteiro, N., Neudorfer, B., Peltonen, T., Rodrigues, P. M. M., Rusnák, M., ... igraiová, D. (2011). *Comparing different early warning systems: results from a horse race competition among members of the macro-prudential research network*. European Central Bank, URL https://www.ecb.europa.eu/events/pdf/conferences/140623/Vasicek-et-al_Comparing-Different-Early-Warning-Systems.pdf?6c50558fcb0b2fd010f6a2c3e02ab286.

Alessi, L., & Detken, C. (2011a). Quasi real time early warning indicators for costly asset price boom/bust cycles: a role for global liquidity. *European Journal of Political Economy*, *257*(3), 20–533.

Alessi, L., & Detken, C. (2011b). 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.

Alessi, L., & Detken, C. (2018a). Identifying excessive credit growth and leverage. *Journal of Financial Stability*, *35*, 215–225.

Alessi, L., & Detken, C. (2018b). Identifying excessive credit growth and leverage. *Journal of Financial Stability*, *35*, 215–225, Network models, stress testing and other tools for financial stability monitoring and macroprudential policy design and implementation.

Allen, F., & Gale, D. (2000). *Comparing financial systems*. Cambridge, MA: The MIT Press.

Arakelian, V., Dellaportas, P., Savona, R., & Vezzoli, M. (2019). Sovereign risk zones in europe during and after the debt crisis. *Quantitative Finance*, *19*(6), 961–980.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725. http://dx.doi.org/10.1146/annurev-economics-080217-053433.

Avesani, R. G., Pascual, A. G., & Li, J. (2006). *A new risk indicator and stress testing tool: a multifactor nth-to-default cds basket*: *IMF working paper, 6*, (105).

Babus, A. (2016). The formation of financial networks. *Rand Journal of Economics*, *47*(2), 239–272.

Baele, L. (2005). Volatility spillover effects in european equity markets. *Journal of Financial and Quantitative Analysis*, *40*(2), 373–401.

Basel Committee on Banking Supervision (2018). *Implications of fintech developments for banks and bank supervisors, consultative document*: *BIS report*, URL https://www.bis.org/bcbs/publ/d415.htm.

Battiston, S., Caldarelli, G., May, R. M., Roukny, T., & Stiglitz, J. E. (2016). The price of complexity in financial networks. *Proceedings of the National Academy of Sciences*, *113*(36), 10031–10036. http://dx.doi.org/10.1073/pnas.1521573113.

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., Hommes, C., Jaeger, C., May, R., & Scheffer, M. (2016). Complexity theory and financial regulation. *Science*, *351*(6275), 818–819. http://dx.doi.org/10.1126/science.aad0299.

Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, *104*(3), 535–559.

Bitetto, A., Cerchiello, P., & Mertzanis, C. (2021). A data-driven approach to measuring epidemiological susceptibility risk around the world. *Scientific Reports*, 11.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., & Simsek, O. (2020). *Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach*: *Bank of England working papers 848*, Bank of England, URL https://ideas.repec.org/p/boe/boeewp/0848.html.

Borio, C. E. V., & Drehmann, M. (2009). *Towards an operational framework for financial stability: 'fuzzy' measurement and its consequences*: *BIS working paper, 284*.

Brave, S., & Butters, A. (2011). Monitoring financial stability: A financial conditions index approach. *Economic Perspectives*, *35*(1), 22–43.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

Candes, E. J., Li, X., Ma, Y., & Wright, J. (2009). Robust principal component analysis?

Casabianca, E., Catalano, M., Forni, L., Giarda, E., & Passeri, S. (2019). *An early warning system for banking crises: from regression-based analysis to machine learning techniques*: *Marco fanno working papers 0235*, Dipartimento di Scienze Economiche Marco Fanno.

Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., & Oliveira, L. I. A. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, *55*, 194–211.

Cevik, E. I., Dibooglu, S., & Kutan, A. M. (2013). Measuring financial stress in transition economies. *Journal of Financial Stability*, *9*, 597–611.

Chan-Lau, J. A., Espinosa, M., Giesecke, K., & Sole, J. A. (2009). *Assessing the systemic implications of financial linkages*: *IMF global financial stability report, 2*.

Clements, M., & Hendry, D. (1998). *Forecasting economic time series*. Cambridge University Press.

Consoli, S., Recupero, D. R., & Saisana, M. (2021). Data science for economics and finance: Methodologies and applications.

Corsetti, G., Pericoli, M., & Sbracia, M. (2005). 'Some contagion, some interdependence': More pitfalls in tests of financial contagion. *Journal of International Money and Finance*, *24*(8), 1177–1199.

Creane, S., Goyal, R., Mubarak, A. M., & Sab, R. (2006). Measuring financial development in the middle east and north africa: A new database. *IMF Staff Papers*, *53*, 479–511.

Demyanyk, Y., & Hasan, I. (2009). *Financial crises and bank failures: a review of prediction methods*: *Bank of finland research discussion paper, 35*.

Devi, S., & Radhika, Y. (2018). A survey on machine learning and statistical techniques in bankruptcy prediction. *International Journal of Machine Learning and Computing*, *8*, 133–139.

Dornbusch, R., Park, Y. C., & Claessens, S. (2000). Contagion: Understanding how it spreads. *World Bank Research Observer*, *15*(2), 177–197.

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

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

Eidenberger, J., Neudorfer, B., Sigmund, M., & Stein, I. (2014). *What predicts financial (in)stability? A Bayesian approach*: *Discussion papers 36/2014*, Deutsche Bundesbank, URL https://ideas.repec.org/p/zbw/bubdps/362014.html.

Erichson, N. B., Zheng, P., Brunton, K. M. S. L., Kutz, J. N., & Aravkin, A. Y. (2018). Sparse principal component analysis via variable projection.

European Central Bank (2005). Measurement challenges in assessing financial stability. *Financial Stability Review*, 131–141.

European Central Bank (2009). Box 1: A global index of financial turbulence. *Financial Stability Review*, 21–23.

Financial Stability Board (2017). *Artificial intelligence and machine learning in financial services-market developments and financial stability implications*: *FSB report*, URL https://www.fsb.org/wp-content/uploads/P011117.pdf.

Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: A comparative approach. *Journal of Financial Stability*, *4*(2), 149–164.

Frankel, J., & Wei, S.-J. (2004). *Managing macroeconomic crises*: *NBER working papers 10907*, National Bureau of Economic Research, Inc.

Gabriele, C. (2019). *Learning from trees: a mixed approach to building early warning systems for systemic banking crises*: *Working papers 40*, European Stability Mechanism, URL https://ideas.repec.org/p/stm/wpaper/40.html.

Gadanecz, B., & Jayaram, K. (2008). Measures of financial stability – a review. In *Proceedings of the ifc conference on measuring financial innovation and its impact* (pp. 365–380). Bank for International Settlements (Ed.). (31).

Geraci, M. V., & Gnabo, J.-Y. (2018). Measuring interconnectedness between financial institutions with bayesian time-varying vector autoregressions. *Journal of Financial and Quantitative Analysis*, *53*(3), 1371–1390. http://dx.doi.org/10.1017/S0022109018000108.

Glasserman, P., & Young, H. P. (2016). Contagion in financial networks. *Journal of Economic Literature*, *54*(3), 779–831. http://dx.doi.org/10.1257/jel.20151228.

Gosh, S., & Gosh, A. (2002). *Structural vulnerabilities and currency crises*: *IMF working paper*.

Guidolin, M., Hansen, E., & Pedio, M. (2019). Cross-asset contagion in the financial crisis: A Bayesian time-varying parameter approach. *Journal of Financial Markets*, *45*(C), 83–114.

Hakkio, C., & Keeton, W. (2009). Financial stress: What is it, how can it be measured, and why does it matter? *Federal Reserve Bank of Kansas City - Economic Review*, 5–50.

Hastie, T., Mazumder, R., Lee, J. D., & Zadeh, R. (2015). Matrix completion and low-rank svd via fast alternating least squares.

Hausenblas, V., Kubicová, I., & sanovská, J. Le. (2015). Contagion risk in the czech financial system: A network analysis and simulation approach. *Economic Systems*, [ISSN: 0939-3625] *39*(1), 156–180. http://dx.doi.org/10.1016/j.ecosys.2014.07.001, URL https://www.sciencedirect.com/science/article/pii/S0939362515000047. Symposium: Youth Labour Market and the Great Recession.

Hawkesby, C. (2000). Maintaining financial system stability: the role of macroprudential indicators. *Bulletin of the Reserve Bank of New Zealand*, *63*(2), 38–52.

Holmfeldt, M., Rydén, A., Strömberg, L., & Strömqvist, M. (2009). How has the stress on the financial markets developed? – an index-based discussion. *Riksbank Economic Commentaries*, 13, URL http://archive.riksbank.se/Upload/Dokument_riksbank/Kat_publicerat/Ekonomiska%20kommentarer/2009/ek_kom_no13_09eng.pdf.

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

Huang, X., Zhou, H., & Zhu, H. B. (2009). A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance*, *33*(11), 2036–2049.

Illing, M., & Liu, Y. (2006). Measuring financial stress in a developed country: An application to canada. *Journal of Financial Stability*, *2*, 243–265.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*, 53–74.

International Monetary Fund (2019). The financial sector assessment program (fsap). URL https://www.imf.org/en/About/Factsheets/Sheets/2016/08/01/16/14/Financial-Sector-Assessment-Program.

Islami, M., & Kurz-Kim, J. R. (2013). *A single composite financial stress indicator and its real impact in the euro area*: *Bundesbank discussion paper, 31*.

Joy, M., Rusnák, M., mídková, K., & Vaíek, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics*, *22*(1), 44–67.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401–415.

Kaminsky, G. (2006). Currency crises: Are they all the same? *Journal of International Money and Finance*, *25*(3), 503–527, URL https://EconPapers.repec.org/RePEc:eee:jimfin:v:25:y:2006:i:3:p:503-527.

Khan, S. A., & Ammad-ud-din, M. (2016). Tensorbf: an r package for bayesian tensor. URL https://www.biorxiv.org/content/biorxiv/early/2016/12/29/097048.full.pdf.

Kim, H., Cho, H., & Ryu, D. (2020). Corporate default predictions using machine learning: Literature review. *Sustainability*, *12*(16).

Kliesen, K. L., & Smith, D. C. (2010). *Economic synopses*: *Vol. 2*, *Measuring financial market stress*. Federal Reserve Bank of St. Louis.

L, A., P, B., & R, S. (2019). Anatomy of a sovereign debt crisis: Cds spreads and real-time macroeconomic data. [ISSN: 2467-2203] (KJ-AE-19-003-EN-N (online)). (online).

Laeven, L., & Valencia, F. (2012). *Systemic banking crises: an update*: *IMF working paper, 12*, (163).

Lane, P. R. (2019). Macrofinancial stability and the euro. *IMF Economic Review*, *67*, 424–442.

Lang, J. H., Peltonen, T. A., & Sarlin, P. (2018). *Working paper series 2182*, *A framework for early-warning modeling with an application to banks*. European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20182182.html.

Lin, W. Y., Hu, Y. H., & Tsai, C. F. (2012). Machine learning in financial crisis prediction: A survey. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, *42*(4), 421–436.

Liu, Y., Yang, M., Wang, Y., Li, Y., Xiong, T., & Li, A. (2022). Applying machine learning algorithms to predict default probability in the online credit market: Evidence from china. *International Review of Financial Analysis*, *79*, Article 101971.

Lo Duca, M., & Peltonen, T. A. (2011). *CB working paper series*: *Vol. 1311*, *Macro-financial vulnerabilities and future financial stress assessing systemic risks and predicting systemic events*.

Louzis, D. P., & Vouldis, A. T. (2011). A financial systemic stress index for greece. In *First conference of the macro-prudential research (MaRs) network of the European system of central banks in frankfurt am main*.

Manasse, P., & Roubini, N. (2009). Rules of thumb for sovereign debt crises. *International Journal of Finance & Economics*, *78*(2), 192–205.

Manasse, P., Savona, R., & Vezzoli, M. (2016). Danger zones for banking crises in emerging markets. *International Journal of Finance & Economics*, *21*(4), 360–381, URL https://ideas.repec.org/a/wly/ijfiec/v21y2016i4p360-381.html.

Manasse, P., Schimmelpfennig, A., & Roubini, N. (2003). *Predicting sovereign debt crises*: *IMF working papers 2003/221*, International Monetary Fund, URL https://EconPapers.repec.org/RePEc:imf:imfwpa:2003/221.

Masson, P. (1999). Contagion:: macroeconomic models with multiple equilibria. *Journal of International Money and Finance*, *18*(4), 587–602.

Mertzanis, C. (2018). Complexity, big data and financial stability. *Quantitative Finance and Economics*, *2*(3), 637–660.

Minoiu, C., Kang, C., Subrahmanian, V., & Berea, A. (2015). Does financial connectedness predict crises? *Quantitative Finance*, *15*(4), 607–624.

Minoiu, M. C., & Reyes, M. J. A. (2011). *A network analysis of global banking: 1978–2009*: *IMF working papers*, International Monetary Fund.

Pericoli, M., & Sbracia, M. (2003). A primer on financial contagion. *Journal of Economic Surveys*, *17*(4), 571–608.

Quagliariello, M. (2009). *Stress-testing the banking system*. Cambridge University Press.

Quinn, D., Schindler, M., & Toyoda, A. M. (2011). Assessing measures of financial openness and integration. *IMF Economic Review*, *59*, 488–522.

Ray-Bing, C., Kuo-Jung, L., Yi-Chi, C., & Chi-Hsiang, C. (2017). On the determinants of the 2008 financial crisis: a Bayesian approach to the selection of groups and variables. *Studies in Nonlinear Dynamics & Econometrics*, *21*(5), 1–17.

Rose, K. A., & Spiegel, M. M. (2012). Cross-country causes and consequences of the 2008 crisis: Early warning. *Japan and the World Economy*, *24*(1), 1–16.

Rosner, B. (1983). Percentage points for a generalized esd many-outlier procedure. *Technometrics*, *25*, 165–172.

Samitas, A., Kampouris, E., & Kenourgios, D. (2020). Machine learning as an early warning system to predict financial crisis. *International Review of Financial Analysis*, *71*(C).

Sarlin, P., & Peltonen, T. (2013). Mapping the state of financial stability. *Journal of International Financial Markets, Institutions and Money*, *26*(C), 46–76.

Savona, R., & Vezzoli, M. (2012). Multidimensional distance-to-collapse point and sovereign default prediction. *Intelligent Systems in Accounting, Finance and Management*, *19*(4), 205–228, URL https://EconPapers.repec.org/RePEc:wly:isacfm:v:19:y:2012:i:4:p:205-228.

Savona, R., & Vezzoli, M. (2015). Fitting and forecasting sovereign defaults using multiple risk signals. *Oxford Bulletin of Economics and Statistics*, *77*(1), 66–92. http://dx.doi.org/10.1111/obes.12052, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12052.

Schuldenzucker, S., Seuken, S., & Battiston, S. (2020). Default ambiguity: Credit default swaps create new systemic risks in financial networks. *Management Science*, *66*(5), 1981–1998. http://dx.doi.org/10.1287/mnsc.2019.3304.

Segoviano, M. A., & Goodhart, C. (2009). *Banking stability measures*: *IMF working paper, 9*, (4).

Slingenberg, J. W., & Haan, J. de. (2011). *Forecasting financial stress*: *De Nederlandsche bank working paper, 292*.

Tanaka, K., Kinkyo, T., & Hamori, S. (2016). Random forests-based early warning system for bank failures. *Economics Letters*, *148*, 118–121.

The Basel Committee on Banking Supervision (2013). *Principles for effective risk aggregation and risk reporting*. Basel, Switzerland: Bank for International Settlement.

Tunyi, A. A., & Ntim, C. (2016). Location advantages, governance quality, stock market development and firm characteristics as antecedents of african m & as. *Journal of International Management*, *22*(2), 147–167.

Vermeulen, R., Hoeberichts, M., Vasícek, B., Zigraiová, D., Smídková, K., & de Haan, J. (2015). Financial stress indices and financial crises. *Open Economies Review*, *26*, 383–406.

Ward, F. (2017). Spotting the danger zone: Forecasting financial crises with classification tree ensembles and many predictors. *Journal of Applied Econometrics*, *32*(2), 359–378.