UNIVERSITY OF PAVIA

DOCTORAL THESIS

The role of Explainable Artificial Intelligence in risk assessment: a study on the economic and epidemiologic impact

---

Author:
Alessandro BITETTO

Supervisor:
Prof. Paola CERCHIELLO

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

XXXIV CYCLE
Electronics, Computer Science and Electrical Engineering

March 25, 2022

Declaration of Authorship

I, Alessandro BITETTO, declare that this thesis titled, “The role of Explainable Artificial Intelligence in risk assessment: a study on the economic and epidemiologic impact” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Abstract

The growing application of black-box Artificial Intelligence algorithms in many real-world application is raising the importance of understanding how the models make their decision. The research field that aims to "open" the black-box and to make the predictions more interpretable, is referred as eXplainable Artificial Intelligence (XAI). Another important field of research, strictly related to XAI, is the compression of information, also referred as dimensionality reduction. Having a synthetic set of few variables that captures the behaviour and the relationships of many more variables can be an effective tool for XAI as well. Thus, the contribution of the present thesis is the development of new approaches in the field of explainability, working on the two complementary pillars of dimensionality reduction and variables importance. The convergence of the two pillars copes with the aim of helping decision makers with the interpretation of the results.

This thesis is composed of seven chapters: an introduction and a conclusion plus five self contained sections reporting the corresponding papers. Chapter 1 proposes a PCA-based method to create a synthetic index to measure the condition of a country's financial system, providing policy makers and financial institutions with a monitoring and policy tool that is easy to implement and update. In chapter 2, a Dynamic Factor Model is used to produce a synthetic index that is able to capture the time evolution of cross-country dependencies of financial variables. The index is proved to increase the accuracy in predicting the ease in accessing to financial funding. In chapter 3, a set of variables covering health, environmental safety infrastructures, demographic, economic and institutional effectiveness is used to test two methodologies to build an Epidemiological Susceptibility Risk index. The predictive power of both indexes is tested on forecasting task involving Macroeconomic variables. In chapter 4, the credit riskiness of Small Medium Enterprises (henceforth SMEs) is assessed by testing and assessing the increase of performance of a machine learning historical random forest model compared to an ordered probit model. The relevance of each variable in predicting SME credit risk is assessed by using Shapley values. In chapter 5, a dataset of Italian unlisted firms provides evidence of the importance of using market information when assessing the credit risk for SMEs. A non-linear dimensionality reduction technique is applied to assign market volatility from listed peers and to evaluate Merton's probability of default (PD). Results show the increase in accuracy of predicting the default of unlisted firms when using the evaluated PD. Moreover, the way PD affects the defaults is explored by assessing its contribution to the predicted outcome by the means of Shapley values.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS 2 Stage Least Square

AE Auto-Encoder with Multilayer Perceptron

AE-LSTM Auto-Encoder with Long-Short Term Memory

AIC Akaike information criterion

APE Absolute Percentage Error

AUC Area Under the Curve

BTF Bayesian Tensor Factorization

CMP Conditional Mixed-Process

DD Distance to Default

DFM Dynamic Factor Model

E-Net Regularized OLS Elastic-Net

ESR Epidemiological Susceptibility Risk

FA Factor Analysis

FIE Financial Institution Efficiency

FME Financial Markets Efficiency

FSI Financial Soundness Indicators

FSIND Financial Soundness Index

HRF Historical Random Forest

IV instrumental variables

MARE Mean Absolute Reconstruction Error

MARS Multivariate Adaptive Regression Spline

MC-SVD Matrix Completion with Low Rank SVD

MSME Micro Small Medium Enterprise

NN single layer Neural Network

PCA Principal Component Analysis

PD Probability of Default

PFI Permutation Feature Importance

RE Reconstruction Error

RF Random Forest

RMSE Root Mean Square Error

ROC Receiver Operating Characteristic

SHAP SHapley Additive exPlanations

SME Small Medium Enterprise

SVM Support Vector Machine

UMAP Uniform Manifold Approximation and Projection

VAR Vector Auto Regressive

VIF Variance Inflation Factor

XAI eXplainable Artificial Intelligence

General introduction

Artificial Intelligence, in its broad definition and meaning, is becoming an integral part of many real-world application. The main reasons for its spread are the exponential increase in availability and amount of data to be processed, the improvements in computing resources (e.g. GPUs, TPUs, cloud computing, etc) and the development of more complex algorithms (mainly based on Artificial Neural Networks). Nowadays, application of Artificial Intelligence affects a variety of decision making processes, ranging from finance, medicine, robotics, agriculture, (ciber-)security and many more.

In such a continuously evolving environment and the expansion of applications in new fields, the understanding of how these so called "black-box" models make their decision has a crucial role. The research field that aims to "open" the black-box and to make the predictions more interpretable is referred as eXplainable Artificial Intelligence (XAI). From a legal point of view, the introduction of regulation such as the European General Data Protection Regulation (GDPR) and the American Algorithmic Accountability Act, raised the concern of having a set of mandatory tools to make the models as transparent as possible to the customers, clearly stating any possible drawback and excluding any possibility of bias. From an ethical point of view, applications such as medical screening or security raised the problem of understanding the drivers of models' predictions so to avoid any kind of discrimination and possible social inequalities. Finally, aside from legal and ethic issues, a better knowledge of how models make their decision clearly has the added value for any users to leverage the information and to increase the performances. Explainability techniques can be classified according to several criteria:

- Intrinsic or post-hoc: distinction whether interpretability is achieved by restricting the complexity of the model (intrinsic) or by applying methods that analyze the model after training (post hoc). Example of intrinsic models are machine learning algorithms that are considered interpretable due to their simple structure, such as linear regression family (OLS, regularized, GLM) or decision trees. Post hoc techniques examples are Permutation Features Importance or Shapley values.

- Feature summary statistic: methods which provide summary statistics for each variable. Some methods return a single number per feature, such as feature importance, or a more complex result, such as the pairwise feature interaction strengths, which consist of a number for each feature pair.

- Feature summary visualization: feature summary statistics can be visualized. Some feature summaries are actually only meaningful if they are visualized and a table would be a wrong choice, such as for the partial dependence plot which are curves that show a feature and the average predicted outcome.

- Data point: this category includes all methods that return data points (already existing or newly created) to make a model interpretable. For example, counterfactual explanations explains the prediction of a data instance finding a similar data point by changing some of the features for which the predicted outcome changes in a relevant way.

- Approximation: black-box models can be approximated (globally or locally) with a more interpretable model. For example LIME locally approximates data points by fitting a regularized linear model such as LASSO.

- Model-specific or model-agnostic: model-specific interpretation techniques are limited to specific model classes. For example, the interpretation of regression coefficients in a linear model. Model-agnostic methods can be used with regards to any model and are applied after the model has been trained (post hoc). These agnostic methods usually work by analyzing feature input and output pairs only. Shapley values are model-agnostic tools.

- Local or global: this class entails the concept whether the interpretation method explain an individual prediction or the entire model behaviour.

For a more complete overview please refer to (Islam et al., 2021, Molnar, 2019).

Another important field of research, strictly related to XAI, is the compression of information, also referred as dimensionality reduction. The abundance of available data as well as their interconnectedness tipically lead to huge number of variables to be analyzed and to be fed into the models. Although modern algorithms can handle even thousands of different variables, the selection or the summarising into smaller set of variables play a fundamental role in increasing the performance of the models and in helping the interpretation and visualization of both inputs and results. In this context, having a synthetic set of few variables that captures the behaviour and the relationships of many more variables can be an effective tool for XAI. Finally, a more general paradigm of dimensionality reduction is the construction of embeddings, low-dimensional vector representations that encapsulate the variation of the input variables. Embeddings can be evaluated on any type of data, overcoming the limitation of techniques that can be applied only on tabular data (i.e. in the Euclidean domain) and extending the application to unstructured data such as text sequences, graphs and trees (usually referred to non-Euclidean domains). Dimensionality reduction techniques can be divided into three main categories:

- Features selection: methods which only keep the most important variables based on some relevance criteria and there is no transformation applied to the set of features. These techniques are mostly directly related to XAI. Backward Elimination or Forward Selection are some examples.

- Linear methods: also known as Matrix Factorization techniques. The idea is to apply some linear transformation to the set of input features to create new features, either keeping the same number of variables or reducing the total number, i.e. summarising. The general approach in applying these techniques is based on factorizing the input matrix of data (observation in rows, features in columns) into two low-rank matrixes: one represents the compact representation of the data and contains the new synthetic variables, the other represents the archetypes, i.e. some kind of basic constituent blocks that can be used to reconstruct the input. Applying different types of constraints on this two matrixes leads to the creation of several alternative methods such as Principal Component Analysis (PCA), Factor Analysis (FA), Linear Autoencoders (Artificial Neural Networks), Latent Dirichelet Allocation (LDA) and Non-Negative Matrix Factorization.

- Non-linear methods: also known as Manifold Learning or Neighbour Graphs. The idea is to construct a graph from the data (or to use an already provided one, dependending on the nature of the data) and then to embed it into a tabular representation. The construction of the graph can be achieved defining some metrics criteria, e.g. distances among observations, and creating links according to given thresholds or similarity algorithms, e.g. shortest path length as for the ISOMAP method. The resulting adjacency matrix is then factorized by its eigenvalues and the eigenvectors of the biggest eigenvalues are finally used to create a low-dimensional embedding, such as in the Spectral Embedding method. Improving the way the graph links are created leads to different methods: t-distributed Stochastic Neighbor Embedding (t-SNE) for example uses an k-Nearest Neighbours algorithms weighted by a kernel with bandwidth adapted to the k neighbours. Instead, Uniform Manifold Approximation and Projection (UMAP) creates a secondary graph, lying in a low-dimensional space, fitting some transformation function f that links points in the original graph to their representation in the low-dimensional space. The function tries to preserve some topological property of the original graph, such as centrality measures, links' strength and shortest path distance. Then, the k-Nearest Neighbours algorithms is applied to the low-dimensional graph and the evaluated links are transformed back to the original graph by the inverse of function f.

The contribution of this thesis to the existing literature is from both a methodological and empirical point of view, in the sense that new methodologies have been designed while trying to solve real-world problems. This thesis focuses on the development of new approaches to the topic of explainability, working on the complementary pillars of dimensionality reduction and variables importance. The convergence of the two pillars copes with the aim of helping decision makers with the interpretation of the results. In particular, the summarising of information is applied to the construction of indexes to measure the riskiness of both financial systems and epidemic outbreaks. On the other hand, the combination of dimensionality reduction techniques in the pre-processing of variables to be fed into models and the used of XAI algorithms on models' predictions, lead to the increase of performances and to the identification of the most relevant and impactful drivers.

The contributions are presented in five self contained chapters. In chapter 1, IMF's financial soundness indicators is used to measure the condition of a country's financial system around the world. Different versions of principal component analysis (PCA) are tested to deal with the presence of strong cross-sectional and time dependence in the data due to unobserved common factors. A synthetic data-driven index is produced and provides policy makers and financial institutions with a monitoring and policy tool that is easy to implement and update. In chapter 2, the methodology described in chapter 1 is improved by the means of a Dynamic Factor Model (DFM), resulting in a synthetic index able to capture the time evolution of cross-country dependencies of IMF's variables. The index is used as additional variable in a dataset consisting of 76 developing countries spanning from 2010 to 2018 with annual frequency and an ordered probit model is fitted to predict the ease in accessing to financial funding. The addition of the index resulted in the improvement in predictions' accuracy. In chapter 3, a set of 17 time-varying variables for 206 countries during the 2010-2019 period, covering health, environmental safety and transport infrastructures, demographic, economic and institutional effectiveness is used to test two methodologies to build an Epidemiological Susceptibility Risk Index: a PCA-based approach is used to create a low dimensional indicator as a weighted average of input variables for every year separately, whereas a Dynamic Factor Model is used to estimate latent factors, capturing the mutual interactions of countries and variables over all years. Finally, the predictive power of both indexes is tested on forecasting task involving usual Macroeconomic variables such as GDP. In chapter 4, the credit riskiness of SMEs is assessed by testing and comparing a classic parametric approach fitting an ordered probit model with a non-parametric one calibrating a machine learning historical random forest (HRF) model. Results provide evidence that a dynamic HRF approach outperforms the traditional ordered probit model, highlighting how advanced estimation methodologies that use machine learning techniques can be successfully implemented to predict SME credit risk. Moreover, the relevance of each variable in predicting SME credit risk is assessed by using Shapley values. In chapter 5, a set of 22 accounting variables of 10k Italian unlisted firms provides evidence of the importance of using market information when assessing the credit risk for SMEs. A methodology to match each firm to a listed peer is developed so to evaluate its probability of default (PD) according to Merton's model. A non-linear dimensionality reduction technique is applied to map the accounting variables to a simpler and synthetic representation so to apply clustering methods and assign market volatility from the most similar listed peers. The increase in accuracy of predicting the default of unlisted firm when using the evaluated PD, is tested with both linear/non-linear and parametric/non-parametric models. Finally, the way PD affects the defaults is explored by assessing its contribution to the predicted outcome by the means of Shapley values.

Chapter 1

A data-driven approach to measuring financial soundness

1.1 Introduction

The importance of a sound financial system for sustained economic growth is well documented (Allen and 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.

Measuring financial stability is a formidable challenge (Gadanecz and 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 and 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.

One way to deal with the challenges of complexity and timely execution of the financial system's risk assessment is to implement a data-driven approach to measuring financial soundness. 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. While no data-driven algorithms exist that can lead to fully optimal assessments of financial soundness, this approach has considerable advantages, such as avoiding post-hoc rationalization, as is typical for economic narrative-based methods.

In this paper, we construct a financial soundness index (FSIND) using a data-driven methodological approach and information provided by the International Monetary Fund's (IMF) Financial Soundness Indicators (FSIs). Given the nature of the measurement problem, we choose principal component analysis (PCA) to deal 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. 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 (Illing and Liu, 2006, Hakkio and Keeton, 2009). 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. The FSIND index is available for 119 countries for the 2010-2017 period. We enrich the index by including information from supplementary variables that take into account geographical and cultural dimensions. 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 is related to the new wave of research on stress-testing and early warning models aimed at assessing financial stability (Alessi and Detken, 2011b, K.Rose and Spiegel, 2012, Lo Duca and Peltonen, 2013, Drehmann and Juselius, 2014). 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 (2018) 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 paper contributes to the financial stability literature by creating a new financial soundness index for 119 countries that is fully data-driven, tested and validated. The data "speak" by means of an unsupervised statistical learning technique, namely the Principal Component Analysis (hereafter PCA). This technique 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 Hakkio and Keeton (2009), Kliesen and Smith (2010), Brave and Butters (2011) 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. Finally, our analysis 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.

The paper is organized as follows: in Section 1.2 we discuss the literature on financial soundness index and relative approaches, in Section 1.3 we present the methodology detailing the Principal Component Analysis, a more robust version employed and the validation index techniques. In Section 1.4 we extensively discuss the data and the preprocessing phase, in Section 1.5 results will be shown and discussed and conclusions are given in Section 1.6.

1.2 Relevant literature

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 and 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 and 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.

The reduced-form approach measures financial system risk by using a quantile regression to calculate conditional values at risk (CoVaR) (Adrian and 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 interrelatedness 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, 2019a). 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 approaches to financial stability assessment have been variously used to produce early warning signals, macro-stress tests and financial stability indicators (Lin et al., 2012, Alessi et al., 2011, Cavalcante et al., 2016, Quagliariello, 2009, Demyanyk and Hasan, 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.

Other studies analyzed the financial stability problem in a cross-country setting. However, these studies have had to deal with the considerable challenges of missing values, insignificant variables across countries (while significant at the country level) and difficulties in the assessment of control variables. 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 de 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).

Overall, financial stability tend to have been single-country based and use less sophisticated estimation methods, often devoid of additional qualitative information. Our analysis goes further by creating a financial soundness index for both developed and developing countries that is fully model-based and data-driven, properly tested and validated, and accounts for the role of qualitative factors. In so doing, we focus on the efficiency of the machine learning 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.

1.3 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:

$$C_1 = w_1 Y_1 + \ldots + w_p Y_p \tag{1.1}$$

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 $k << p$ 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:

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

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:

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

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:

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

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.

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

1.4 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 1.1 and Table 1.2 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 A.1. 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 Figure 1.1. 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 and Ammad-ud-din, 2016)

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 16184 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

Table 1.1: 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 forex 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

Table 1.2: Correlation matrix of input variables. 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 forex 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'.

	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*

$*$ p-val $<$ 0.05

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.

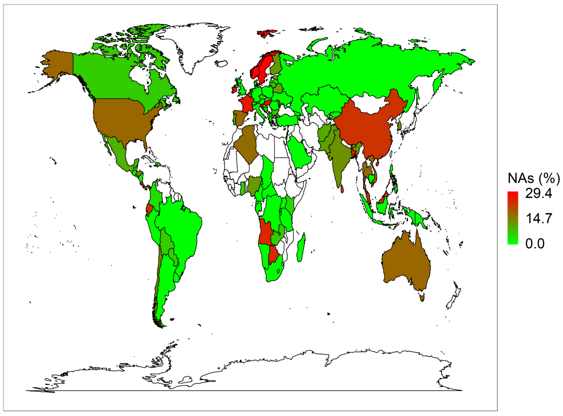


Figure 1.1: Map of analyzed countries and missing values percentage distribution by color.

1.5 Data Analysis

As described in Section 1.4, we assess the cross-validation error of the data imputation techniques. Figure A1 in A.2 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 1.3. Results for the PCA approach can be found in Table 1.3 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 95$^{\text{th}}$ and 99$^{\text{th}}$ 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). Figure 1.2 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%. Figure 1.3a 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 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. Figure 1.3b shows the bi-plot referred to the two first components in which values over years are averaged and weighted by yearly loading importance. Figure A4 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 figure 1.3b adequately representative of the overall index pattern. Figures from A2 to A6 in the Appendix report the scree plots and loading importance plots for other implemented PCA methods.

Table 1.3: 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 1^{st} and 5^{th} percentile.

Method	Number of PC	Mean Expl. Variance	Mean R^2	Mean R^2 on 99^{th}	Mean R^2 on 95^{th}	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	$38 \pm 9.9\%$	$28.2 \pm 4.1\%$	$36.9 \pm 8.2\%$	$47.7 \pm 6\%$	$\ll 0.01$

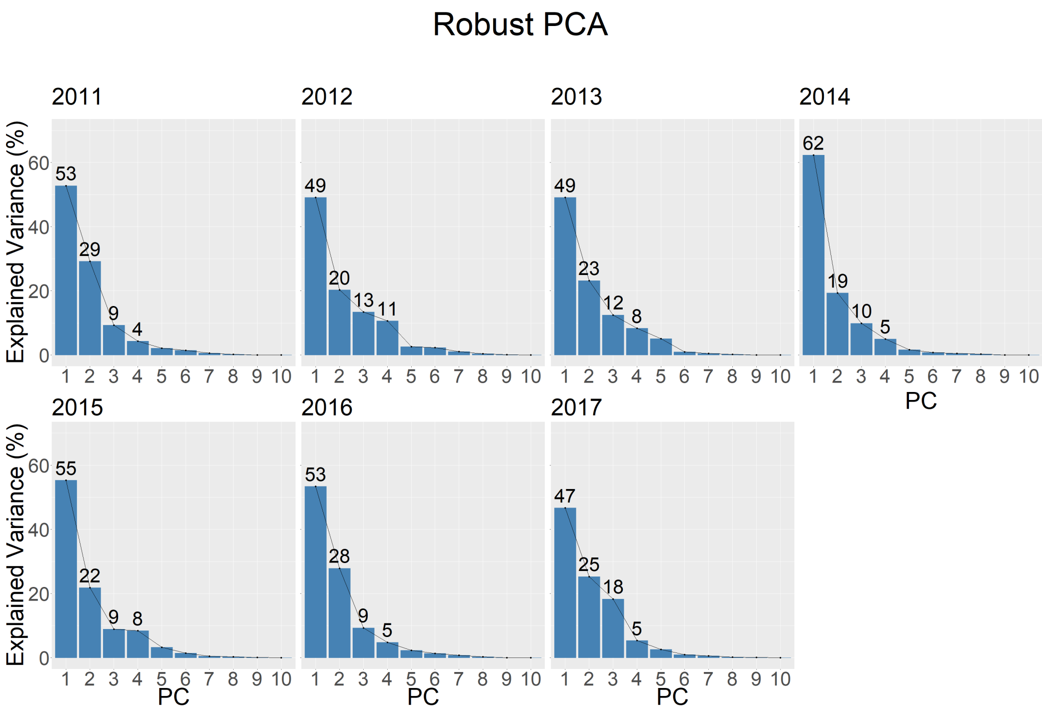
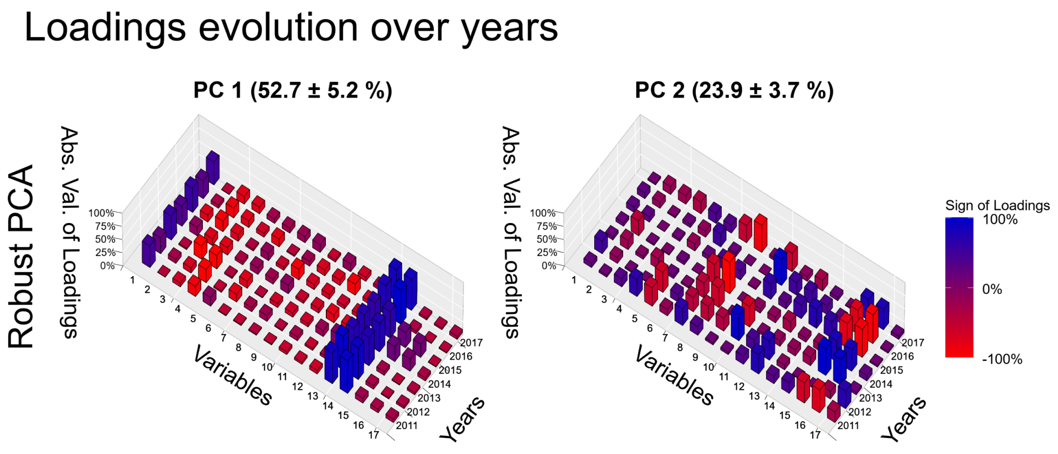
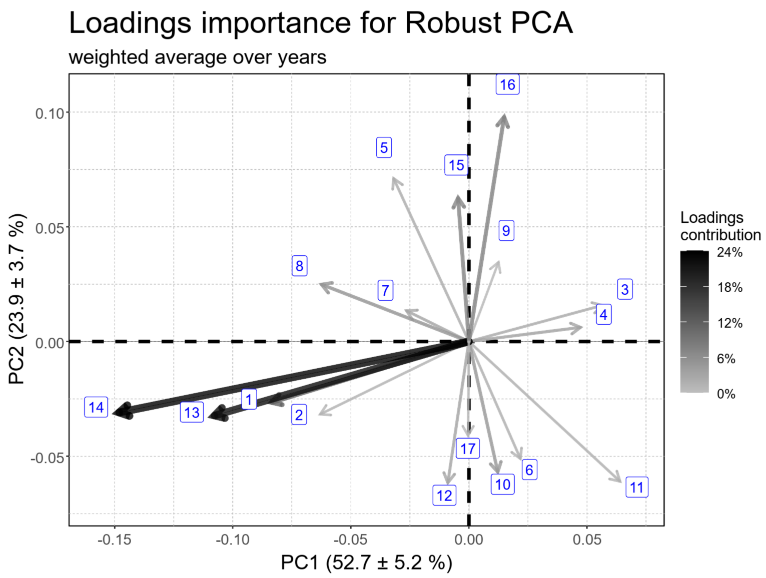


Figure 1.2: Scree plot for Robust PCA method for each year. The first two components account for an average 76% of cumulative explained variance.



(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.

Figure 1.3: 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 forex 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'.

Once the continuous FSIND is estimated, we find the optimal threshold to binarize our index. As described in Section 1.3, 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 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. Figure 1.4 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 grey 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 grey surfaces represent the binary FSIND performance over all dataset and trimmed top 5^{th} 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 A7 to A11 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.

Figure 1.5 shows the evolution over years of the continuous FSIND for selected countries. In particular, Figure 1.5a 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 Figure 1.5b 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 Figure 1.5c 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

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

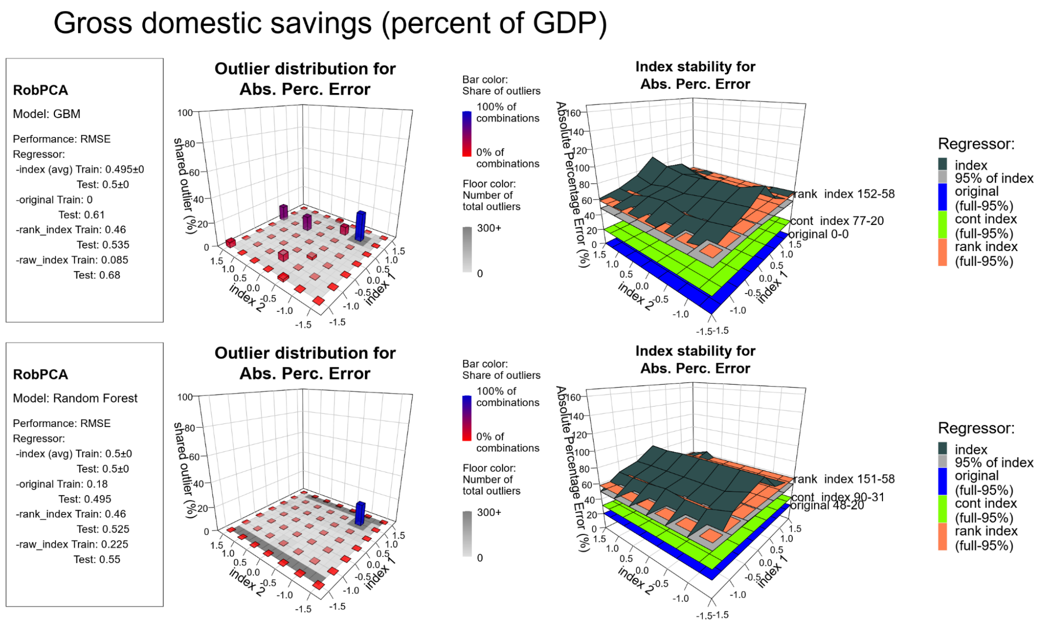
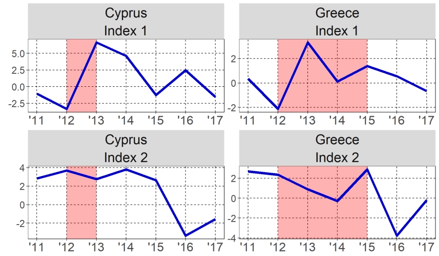
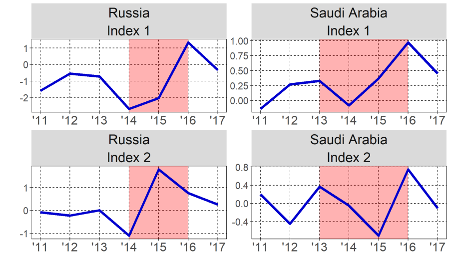


Figure 1.4: Index validation for gross domestic savings. The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: floor color shows the 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 grey surfaces represent the binary FSIND performance over all dataset and trimmed top 5$^{\text{th}}$ 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.

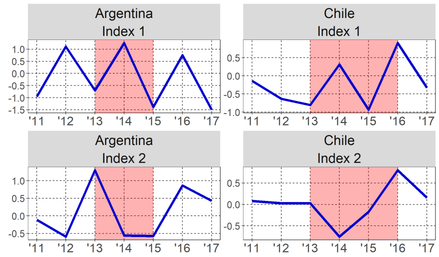
energy distribution network, common cultural origins and share similar geopolitical concerns as shown in Figure 1.5e. A spike on the first component of FSIND outlines the 2011-2012 crisis in Poland as a consequence of the European global one. It also shows the 2013-2015 Ukrainian debt crisis while struggling to cope with the fallout from Crimea's annexation by Russia and continuing war with pro-Russian separatists. Figure 1.5d 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 A12 to A15 in the Appendix report the complete list of all countries.



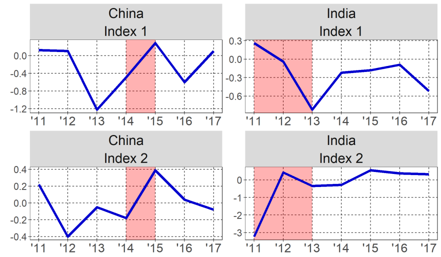
(a) Greece vs Cyprus



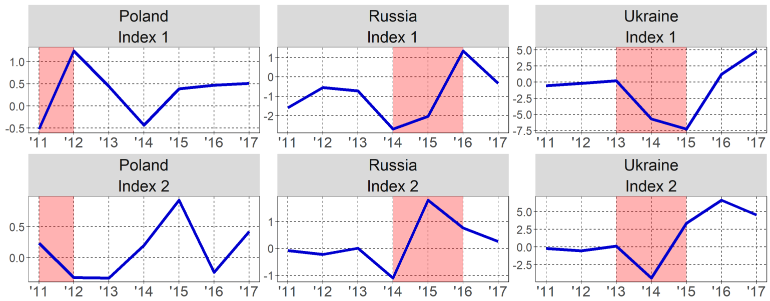
(b) Saudi Arabia vs Russia



(c) Argentina vs Chile



(d) India vs China



(e) Poland vs Ukraine vs Russia

Figure 1.5: 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.

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 Figure 1.6. 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 0^{th} and 20^{th} quantiles, the second between 5^{th} and 25^{th} ones and 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. Figure 1.7 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 A16 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.

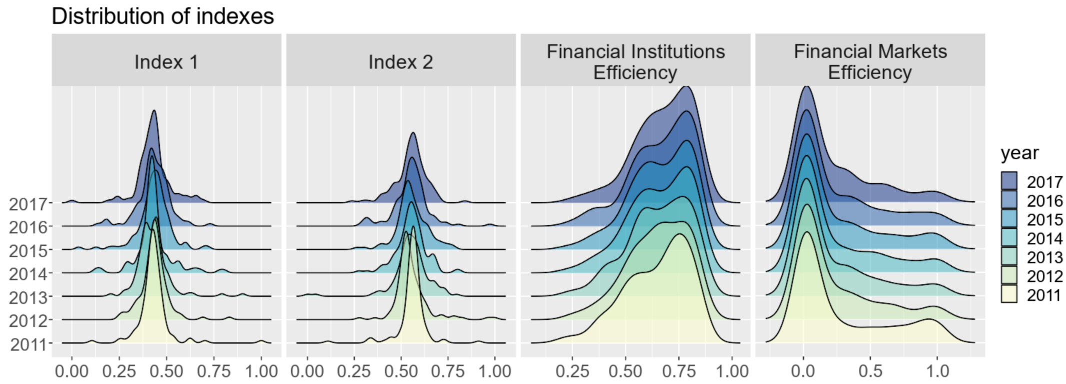


Figure 1.6: Comparison of the two FSIND dimensions with Financial Institution Efficiency and Financial Markets Efficiency index distributions.

1.6 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. 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.

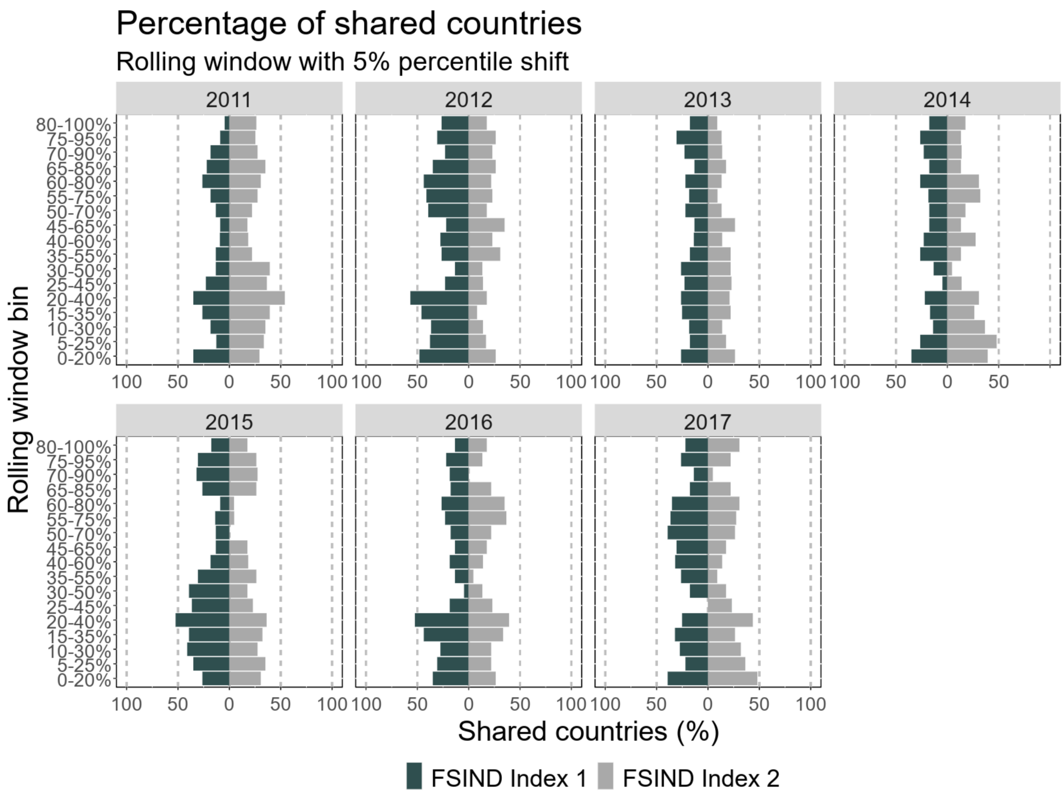


Figure 1.7: Ranking ability comparison for FSIND components and Financial Markets Efficiency (FME) index. For each FSIND component and FME, 20 quantiles, i.e. one every 5% are evaluated, splitting the index values in 17 rolling bins, each of which spans between five quantiles, i.e. a 20% range, and a 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 grey) component.

Chapter 2

Do sound financial systems improve the financing constraints of firms? Evidence from developing countries

2.1 Introduction

Financing constraints affect corporate investment and growth. The determinants of the financing constraints of firms have therefore been an important issue in the finance literature (Fazzari et al., 1988; 2000, Kaplan and Zingales, 1997; 2000, Brown et al., 2012a, Ferrando et al., 2017). While a key consideration for firms across the globe, financing constraints are especially important for firms operating in developing countries due to specific financial and institutional reasons (i.e., inadequate collateral, inefficient governance, etc.) (Claessens, 2006, Beck and Laeven, 2006, Beck et al., 2005, Karlan and Morduch, 2010).

The determinants of financing constraints of firms have been analyzed within a framework that considers the role of the specific characteristics of firms as well as that of country-level economic factors and institutions. Indeed, prior studies have documented the importance of firms' size and age for understanding their financing constraints (Devereux and Schiantarelli, 1990, Schiantarelli, 1995, Oliner and Rudebusch, 1992, Schiffer and Weder di Mauro, 2001, Klapper et al., 2006, Kuntchev et al., 2013, Petersen and Rajan, 1994, Beck et al., 2005). Other studies emphasized the role of the sectoral characteristics of firms (Hall et al., 2000, Abor, 2007). Yet other studies, stressed the importance of ownership structure (Harrison and McMillan, 2003, Colombo, 2001, Clarke et al., 2006, Mertzanis, 2017). Some studies explored the effects of firm location on their access to finance (Berger and Udell, 1995b, Gilbert, 2008). Finally, some studies explored the role of the firms' legal incorporation status (Harhoff and Korting, 1998, Cassar, 2004, Abor, 2007). While using diverse methodologies and data, these and subsequent studies have broadly documented the important role of firm-specific characteristics in explaining financing constraints.

However, the sole focus on the specific characteristics of firms has not been sufficient for understanding the causes of their financing constraints (Kaplan and Zingales, 1997), redirecting the focus of research to the broader association between the financial system, financial access and economic growth. Especially influential have been the studies that documented robust associations between financial depth and economic growth (Levine, 1997, Levine et al., 2000a), financial depth and corporate finance (Demirgüç-Kunt and Maksimovic, 1998, Rajan and Zingales, 1998, Levine et al., 2000b), financial depth and income inequality (Beck et al., 2007, Honohan, 2006) as well as the role of financial liberalization (Laeven, 2003), among others. However, the link between finance and economic growth is weak (Rodrik and Subramanian, 2009, Stiglitz, 2000).

Another strand of literature focused on the role of institutions in affecting directly and indirectly the financing conditions and behavior of firms. For example, some studies emphasized the role the role of historical origins and political systems (Acemoglu et al., 2001, Acemoglu and Robinson, 2005), other studies stressed the role of the individual legal rights (North, 1990), and further studies analyzed the role of culture and religion (Guiso et al., 2006) and social capital (Putnam, 2000). Other studies documented the important role of social fractionalisation (Alesina and La Ferrara, 2005) and family ties (Mertzanis, 2019), whilst some studies emphasized the need to properly differentiate among institutions for explaining firms' external finance decisions (Knack and Xu, 2017).

The recent global financial crisis stressed the role of alternative financial markets and infrastructure for understanding the financing constraints of firms. For example, the latter are found to be affected by the developments in the interbank markets that condition the ability of the banking system to extent credit to firms (Ivashina and Scharfstein, 2010, Campello et al., 2010, Duchin et al., 2010). The supply-side credit effects are found to be more pronounced in bank-dependent firms (Leary, 2009), in firms financed by short-term debt and trade credit (Akbar et al., 2013), in firms with large institutional holdings (Erkens et al., 2012) and where banks were inadequate capitalized (Paravisini, 2008).

A related post-crisis literature stressed the role of financial stability and macroprudential policies regulation policy. Credit availability to firms was linked inter alia to the macroprudential policies (Ayyagari et al., 2018, Yarba and Guner, 2020), the bank capital requirements (Fisera et al., 2019, Fang et al., 2020, Gopalakrishnan et al., 2021), the financial supervision structure (Mertzanis, 2020) and the central bank-imposed liquidity constraints (Ananou et al., 2021). While there are several studies that measure financial stability see surveys by Acharya et al. (2012), Gadanecz and Jayaram (2008)), they do not directly assess the effects of these measures on the financing decisions of firms. Empirical studies mainly focus on developed countries and tend to stress the effect of financial stability policy on bank lending policies subject to the characteristics of banks and less to the characteristics and financing needs of individual firms. The evidence of the effect of financial stability considerations on the financial of firms in less developed countries is very limited and focuses on single countries only (European Central Bank, 2005, Gray et al., 2007).

We contribute to this strand of literature by developing a novel country measure of financial soundness that captures policy considerations and using it to explain the financing constraints of individual firms in medium and low-income countries. We use a data-driven statistical approach to combine the individual country-level financial soundness indicators (FSIs) produced by the International Monetary Fund (IMF) and produce a financial soundness index (FSIND). Subsequently, using micro data from the World Bank's Enterprise Surveys, we explore the impact of our financial soundness index on the financing constraints of 63,000 firms in 76 low- and middle-income countries during 2010-2018. We analyze the extent to which macroprudential and other regulatory policies (as captured by the FSIs) influence the extent to which firms experience access to finance as an obstacle to their business operations. The use of firm-level data to study the effects of our country-level financial soundness index has the advantage of mitigating reverse causality bias since it is unlikely that individual firms' decisions will influence macro policies and of allowing the inclusion of country-year fixed effects to control for the impact of omitted variable bias. The results show that the financial soundness index is a broadly robust predictor of financing constraints of firms in developing countries. We find that firms operating in countries with higher levels of financial system soundness experience lower financing constraints. The effect is stronger for older and larger firms, which are publicly listed. Interesting the effect is insignificant for subsidiary and state-owned firms. These results are robust to controlling for country, year and industry shocks through country-year-sector fixed effects, allowing for heterogeneous effects of other industry and country-level factors that influence firm behavior. Our analysis deploys several sensitivity tests to check the predictive robustness of the FSIND under alternative variable measures, sample structures and estimation methods as well as using the Oster test (Oster, 2019) to determine the relevance of the selected macro-group variables relative to unobserved ones thereby assessing their impact on the change of the coefficients' value.

A major concern with our empirical findings is about the potential endogeneity that may affect the observed relation among financial stability at the country level and individual firms' access to finance. Our country-level index captures general policy conditions and policies and not firm-level decisions. Further, it is possible that country-level economic and other factors affect simultaneously the financial soundness conditions and the individual firm decisions creating confounding bias. While our data captures within-firm variation and across-firm differential effects, the potential endogeneity may still be a problem. We address the endogeneity concern by applying alternative variable measurement, alternative sample structures and instrumental variable analysis.

Our analysis adds to the literature in several ways. First, extending prior studies (Claessens et al., 2013, Cerutti et al., 2017, Akinci and Olmstead-Rumsey, 2018), we develop a new measure of the health and soundness of financial institutions, markets and households using credible IMF information that accounts for the incidence of macroprudential and other policies that assess and monitor the strengths and vulnerabilities of the financial system as a whole. Our measure is not limited by its dependence on the conditions and prudential ratios of individual financial institutions alone (Čihák and Schaeck, 2010, Cihak et al., 2012), but it reflects the broader, combined financial conditions, including compliance with international financial sector standards and codes, and the outcome of stress tests. We construct of financial soundness index that is fully data-driven, tested and validated. The data 'speak' by means of an unsupervised statistical learning technique, which makes neither a priori assumptions on the relationship among the independent variables nor a subjective decision on the variables to be possibly dropped.

Second, our analysis uses information on a large number of diverse firms operating in all the important medium and low-income countries. We analyze the controlling effect of a wide range of firm-specific characteristic to account for firm-level heterogeneity. Our analysis contributes to the large literature on the role of firm-specific characteristics in explaining the financing constraints of firms (Petersen and Rajan, 1994, Berger and Udell, 1995b, Beck et al., 2008) and on the micro effects of macroprudential policies (Ayyagari et al., 2018, Yarba and Guner, 2020). Moreover, our analysis covers a large number of developing countries, where financial stability information is scant and fragmented, mostly based on single country measures (European Central Bank, 2005, Gray et al., 2007), using a consistent and uniform measure. In this respect, our analysis helps elucidate the challenging tradeoff between financial stability and economic dynamism in developing countries. Third, our analysis contributes to the literature on the effect of aggregate economic shocks (Gertler and Gilchrist, 1994, Chodorow-Reich, 2014, Ananou et al., 2021) and institutional factors on the financial decisions of firms.

In what follows, Section 2.2 reviews the related literature; Section 2.3 describes the construction of the FSIND, the data and the empirical methodology used for the analysis; Section 2.4 explores the predictive power of FSIND and other firm-specific characteristics on the financing constraints of firms and applies various sensitivity tests; Section 2.5 contains different endogeneity tests; and finally section 2.6 concludes the paper.

2.2 Relevant literature

The 2007-8 financial crisis raised the need for macroprudential analysis. The latter has been seen as important for identifying vulnerabilities in the financial system as a whole, which in turn required improved information on the soundness of financial systems. The paucity of data in this area, and a lack of dissemination and cross-country comparability have been recognized as key stumbling blocks. In response, the International Monetary Fund (IMF) has worked closely with national agencies and regional and international institutions to develop a set of Financial Soundness Indicators (FSIs), which monitor the financial sector's current health condition (International Monetary Fund, 2019b). The soundness of a country's financial system has attracted the researchers' attention and it is directly linked to financial stability considerations (Restoy, 2017). It is especially important for developing countries, where the financial systems are less developed, firms suffer from inadequate credit access and information quality, and financial inclusion is a key policy consideration.

The soundness of a country's financial system is intertwined with its macroprudential policies. Claessens et al. (2013) classify different types of macroprudential policies according to their purpose. Some focus on dampening an expected credit boom or credit crunch and they are more cyclical in nature. Others focus on increasing the resilience of the financial sector, using capital or provisioning requirements, and they are more capital-driven. Subsequently, some policies focus more on the conditions of financial institutions whilst others focus more directly on borrowers. Thus, depending on the phase of the business cycle and the choice of financial policies in different counties, the overall configuration of macroprudential policies and financial soundness will differ among countries. Cerutti et al. (2017) document the use of various macroprudential policies in 119 countries over the period of 2000–13 and find that macroprudential policies are associated with lower country credit growth. Their effects are less in open and financially more developed countries. Akinci and Olmstead-Rumsey (2018) use quarterly data to construct an index that measures the tightening and easing of macroprudential policies in 57 countries and show that these policies are used in tandem with bank reserve requirements, capital flow management measures, and monetary policy. Lim et al. (2011) study a smaller subset of 49 countries and find that macroprudential policies are associated with reductions in the procyclicality of credit and firm leverage. Edge and Liang (2019) stress the role of the establishment of Financial Stability Committees (FCSs) as a tool for financial risk mitigation and use the interaction between FCSs and regulatory agencies in 58 countries to analyze the drivers of financial stability. Their results show that, after controlling for the severity of the financial crisis, countries with stronger FSCs are more likely to use the countercyclical measures of credit growth, especially relative to countries where a bank regulator or the central bank has the authority to set counter-cyclical policy. Fendoglu (2017) constructs a macroprudential policy stance index based on the IMF's detailed survey on macroprudential policy actions and he finds that an overall tightening in the macroprudential policy stance is effective in containing both the credit cycles per se and the impact of portfolio inflows on the credit cycles.

Financial soundness considerations affect the indirect financing of firms through the banking system and especially the conditions of financial institutions and the regulation of capital requirements. For example, Chodorow-Reich (2014) examines the impact of credit supply disruptions associated with the crisis and finds bigger effects among small firms, due information asymmetries caused by frictions. Firms that had pre-crisis relationships with less healthy lenders had a lower likelihood of obtaining a loan during the crisis, paid a higher interest rate on the loan, and reduced employment. Degryse et al. (2015) use data that comprise geographical information of bank branches or headquarters and analyze the effect of banks' financial conditions (leverage, core deposits, etc.) on their provision of credit on SMEs before and during the financial crisis. They document a significant association between banks' financial conditions and firms' access to credit, which is affected by the firms' proximity to branches and headquarters as well as the phasing of the crisis. Fisera et al. (2019) analyze the effect of Basel III rules on the financing constraints of small- and medium-size enterprises in developing countries. They find that higher capital requirements are associated with a negative effect on firms' access to finance, especially those that have limited access to the financial system (only a bank account). Gopalakrishnan et al. (2021) analyze the effects of Basel regulations on risk-sensitive assets on the debt financing choices of firms. Using a difference-in-difference analysis of firms in 52 countries, they find that low-rated firms experience a reduction in credit availability, which is further associated with lower investment and lower dividend payout to shareholders. They also find that the effect is stronger in countries that allow banks to implement internal ratings systems. Calem et al. (2020) analyse the impact of several prudential policies on the supply of credit in the US. They find a negative effect of stricter stress-test regulation on the amount of mortgage credit. They also find that the share of speculative-grade loan origination decreased with higher bank regulation. Fang et al. (2020) show that higher bank capital requirements are associated with lower firms' access to credit. They use quarterly data for 14 Peruvian banks and several model specifications to address concerns about the endogeneity of capital requirements. They find that the capital requirement effect is stronger during periods of lower economic growth and that banks with low levels of liquidity, capitalization and profitability, are more reactive to changes in capital requirements. Desai et al. (2004) analyze how multinational firms capitalize their affiliate firms around the world and show that, in response to prudential policies, these affiliates substitute internal borrowing for expensive external financing thereby alleviating their financing constraints. Ananou et al. (2021) focus on the role of central bank-imposed liquidity constraints. They find that bank liquidity shortages during the global financial crisis of 2007-2009 led to the introduction of liquidity regulations (Liquidity Balance Rule) in the Netherlands, the impact of which was an increase in corporate credit due to higher inflow of retail deposits and equity injections.

On the other hand, the soundness of a country's financial system can have a direct impact on firms' capability and willingness to demand finance. The effect operates through the general level of uncertainty, which is a source of destabilization that affects individual firm behavior. Mac an Bhaird et al. (2016) examine the effects of the perception of a loan application rejection by firms in 9 European countries. They find that the transmission of macro-financial uncertainty effects through the banking system may lead to higher levels of firms' discouragement in applying for loans. They highlight the importance of capital market regulation and enforcement mechanisms in mitigating the negative effects of higher uncertainty on firms borrowing discouragement. Becchetti and Trovato (2002) argue that, in conditions of uncertainty, younger and the smaller firms are least likely to lower their demand for external finance because they may have higher growth potential, which they need more finance to secure.

The impact of macroprudential policies may also be affected by the characteristics of the financial intermediation structure. Dabla-Norris et al. (2015) use firm-level data from emerging markets and a general equilibrium model based on game theory, to identify constraints to financial inclusion. They find that macroprudential policies influence the size of participation costs and of collateral thereby affecting firms' access to finance. Their results also show that alleviation of financial frictions are associated with differential impact on firms across countries, due to country-specific characteristics that determine the connections and balance between inequality and financial inclusion. Mertzanis (2020) explores the impact of financial supervision structure on firms' financing constraints in 48 developing countries. He suggests that decentralized structures of prudential supervision are associated with more binding financing constraints of firms in high-income developing countries and less binding ones in market-based financial systems. Ehigiamusoe and Samsurijan (2021) provide evidence that a stable macro-financial environment and higher levels of regulatory quality are necessary conditions for enhancing the role of finance in accelerating economic growth in developing countries. They also find that the mitigating effect turns negative beyond a certain level of finance in the economy.

In this paper, we extent this line of research by examining the effect of a novel composite financial soundness indicator at the country level, based on combined information from the IMF's financial soundness indicators across countries, on the financing constraints of the individual firms operating in those countries. The nature of information used for its construction makes our indicator a reasonable proxy of the state of macroprudential policies of countries (Claessens et al., 2013, Cerutti et al., 2017, Akinci and Olmstead-Rumsey, 2018). Our study follows other studies that experimented with the construction of composite measures of financial stability. For example, Van den End (2006) and Nelson and Perli (2007) argue that the complexity of financial intermediation makes general financial market indicators valuable inputs to measuring financial stability. Similarly, Hawkins and Klau (2000), Nelson and Perli (2007), Gray et al. (2007), Illing and Liu (2006), used alternative combinations of aggregate financial variables and different aggregation models to produce different aggregate financial stability indicators. The interest in constructing financial stability indicators has also been extended to central banks (Bank of England, 2008, Sveriges Riksbank, 2008). In the next section, we explain the construction of the FSIND.

2.3 Data and Methodology

2.3.1 Construction of the Financial Soundness Index

We construct our financial soundness index as a synthetic aggregation of country-level information provided by the IMF's financial soundness indicators. Synthetic measures are typically based on assumptions made by experts regarding the choice of weights. These assumptions are subjective by nature and therefore the associated synthetic indices may be questionable, leading to debate on what is a robust financial indicator to consider. Prior studies have implemented various methods for producing synthetic financial indexes, which can be broadly grouped into econometric methods and statistical learning methods. The former comprises inter alia the studies by Moccero et al. (2014), Opschoor et al. (2014), Mamatzakis and Tsionas (2020), Huang et al. (2021). Those papers typically employ Vector Autoregressive or GARCH models to naturally elicit the temporal evolution of the considered financial variables. The latter comprise studies that use dimension reduction techniques like Principal Component Analysis or Factorial Analysis, such as Kabundi and Mbelu (2017), Ahamed and Mallick (2019), Saha and Dutta (2020).

In this paper, we use a data-driven statistical approach to construct our financial soundness index based on country-level information included in the 17 financial soundness indicators (FSIs) produced by the IMF during 2010-2018 that cover 140 developed and developing countries. Tables B4 and Table B1 in the Appendix present the summary statistics of the index's constituent variables 1 to 17 and their pairwise correlations. Unfortunately, some countries have missing values of the 17 indicators and years. As a result, we restrict our analysis to 76 countries from 2010 to 2018, selected with an incidence of missing values not exceeding 30%. Table B2 in the Appendix provides the selected countries and the associated percent of missing values. Since the presence of many missing values could considerably impact the quality and reliability of results, we carry out missing values treatment and imputation. In our sample, 16 countries show a percent of missing values between 20-29%. Thus, we apply two alternative data imputation methods: a Matrix Completion with Low Rank SVD method (MC-SVD) (Hastie et al., 2015) and Bayesian Tensor Factorization (BTF) method (Khan and Ammad-ud-din, 2016). Briefly, MC-SVD solves the minimisation 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. B.2 describes the assessment of the reconstruction performance for the two imputation techniques. Overall, we find that Bayesian Tensor Factorisation performs better.

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 independent variables time-series. Consequently, we remove differences in magnitude among the independent variables by standardising the values, i.e. we subtract the mean and divide by the standard deviation. Having all variables on the same reference scale is crucial for unbiased estimation when applying dimensionality reduction techniques.

Then we take advantage of a statistical methodology to build the index following the dimensionality reduction approach: Factor Analysis (FA). FA models the measurement of latent variables, seen through the relationships they cause in a set of Y variables. The model is represented by a set of equations $Y_i = b_i F_i + u_i, i = 1, \ldots, p$, where Y_i are the original variables, F_i are the latent factors and b_i, u_i are the parameters of the combination. Recalling that our dataset has three dimensions, *Country*, *Variable* and *Time*, we evaluate a temporal dependent version of FA called Dynamic Factor Model (DFM), modelling country/variable interactions for all the available years within the same model. Given the $p \times n$ matrix $\mathbf{X}$, the model assumes that there exist some $k \times n$ factors $\mathbf{F}$ such that their mutual interaction over time can be expressed by a $k \times k$ interaction matrix $\mathbf{A}$ and the observed variable can be expressed as a linear function of the factors themselves through a $p \times k$ loading matrix $\mathbf{C}$. The problem can be solved as a system of equations:

$$\begin{cases} \mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathcal{N}(0, \mathbf{Q}) \\ \mathbf{X}_t = \mathbf{C}\mathbf{F}_t + \mathcal{N}(0, \mathbf{R}) \end{cases} \tag{2.1}$$

where $\mathcal{N}$ is the normal probability distribution and $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrix of the residuals of each equation in (2.1), respectively. Due to the short time series of the independent variables, this model cannot be fitted considering all countries together as the resulting system of equations (2.1) is under-determined. Thus, we deal with the problem as follows: first, following Holmes et al. (2018), we fit DFM for each country, obtaining the factor matrices $\mathbf{F}^i$, the factor interactions $\mathbf{A}^i$ and the factor loadings $\mathbf{C}^i$, $i = 1, \ldots, n$. Second, we fit a Vector Auto Regressive (VAR) model in order to get $\hat{\boldsymbol{A}}$ 1-year lag matrix that incorporates cross-countries interactions of $\mathbf{A}^i$. We implement the model using *R* package `sparsevar` because this calibration problem has too many parameters to estimate relative to the number of observations, thus requiring a sparse approach. Then, we use Kalman Filter to get smoothed factors $\widehat{\mathbf{F}^i}$ using $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}} = diag(\boldsymbol{C}^i)$, that is to get latent factors that incorporate cross-countries interactions. Briefly, Kalman filter re-estimates the factor matrix $\mathbf{F}$ iterating the two equations in (2.1) until the error between the predicted observed variables $\hat{\boldsymbol{X}}$ and the true one is minimised. We implement the model using *R* package `FKF`. We assume $\hat{\boldsymbol{C}}$ to be diagonal in order not to double-count correlations within the observed variables and because cross-country interactions are already modelled through the VAR. Moreover, the described procedure depends upon two hyper-parameters: the sparsity coefficient α of the VAR and the correlation structure of the residuals for Kalman filter. Thus, we simulate synthetic factors $\widetilde{\mathbf{F}}$

with different combinations of number of observed variables, countries, years, latent factors $\mathbf{F}$, and we generate the corresponding $\mathbf{X}_t$ given different combination of $\mathbf{A}$, defined by α, and $\mathbf{C}$, randomly generated, using equation (2.1). Then, for each of the previous combination and correlation structure of residuals $\mathbf{Q}$, we apply the described algorithm and assess the reconstruction error on the fitted factors $\widetilde{\mathbf{F}}$ with the simulated factors $\mathbf{F}$. The optimal parameters found are $\alpha = 0.2$ and a diagonal structure. The final index, hereinafter referred to as Financial Soundness Index (FSIND), will be represented by the k-dimensional factor matrix F. One of the goals is to select the optimal number of components k as a trade-off between the maximal explained variance and the smallest value of components k. We produce a k-dimensional continuous FSIND per country-year pair. Afterwards, we evaluate the R^2 on both the whole dataset and subsets with values trimmed for the 95$^{\text{th}}$ and 99$^{\text{th}}$ percentiles in order to check for the impact of outliers. 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. We fit the DFM model with one and two factors as well under the assumption of interactions between factors, i.e. estimated $\hat{\boldsymbol{A}}$, and no interactions, i.e. $\hat{\boldsymbol{A}} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. Table B5 in the Appendix reports the results. Models with no factors' interactions have low performance, meaning that cross-countries effects are relevant in order to capture the intrinsic relationship within the data. In fact, the normalised entries of the estimated interaction matrix $\hat{\boldsymbol{A}}$ turn out to rather large, ranging into $[-0.76, 0.75]$. Moreover, the use of two factors provides very small improvements on the performances compared to the single factor version in both model settings. Therefore, we prefer to retain only the single factor model, which explains at its minimum an R^2 of 65% and because the possibility of building up our FSIND index considering just one component eases the interpretation, the relative employment and the subsequent monitoring. Additionally, we run the Im-Pesaran-Shin test on the FSIND 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. B.3 reports the interpretation of the relative importance of the DFM loadings and their impact on each country. Finally, the maps in Figure B4 in the Appendix report the global distribution of FSIND index over years for each country. For sake of clarity, we recall that high values of FSIND are reported for less riskier countries, on the contrary high values correspond to riskier and unstable countries.

2.3.2 Description of Data

To identify the causal effect, we use firm-level data from the Enterprise Surveys carried out by the World Bank (ES hereafter). The basic dataset includes 105,665 non-financial firms located in 76 middle- and low-income countries during 2007-2018. The collection of the ES data is based on successive rounds of surveys. These survey rounds are essentially independent collections of cross-section data, where only few firms systematically appear throughout the successive surveys. The data panel structure is therefore unbalanced but it has the advantage of containing consistent information based on standardized response across all survey years and countries. The data have the important strength of representing diverse firms by size, industrial sector, incorporation status, location of operations, and other specific to them characteristics. The responses reflect the firms' experience of firm performance given the surrounding business environment. To contain self-selection bias, the ES data use random samples of representative firms with different characteristics, which the collectors update for each country and properly bring them into consistent form. The ES data have the reasonable drawback of whether they truly reflect firm behavior. However, at the absence of high-quality census data in most developing countries, survey data include information that directly reflects the firms' knowledge, which may convey more valuable information on their true experience. This limits the chance of inverse causation, for changes in a country's digital adoption progress resulting from changes in an individual firm's performance are highly improbable. We test different control variables and model designs to ensure that improper specification does not affect the causal effect. We further deal with potential asymmetry of information problems by applying proper clustering of estimated standard errors.

The outcome variable in our analysis is the firms' experience of financing constraints (ACCESS). Based on the ES description, it is the response of firms to the survey question: "How problematic is financing for the operation and growth of your business?". The response varies between zero (no constraint), one (minor constraint), two (moderate constraint), three (major constraint) and four (very severe constraint). Thus, ACCESS is an ordinal variable within the range $[0,4]$. However, it is possible that these answers may not capture all reality as well as that some firms may report financing constraints while they are not actually constrained by them but only facing temporary liquidity distress. Therefore, one must be cautious of this behavioral bias and interpret the results carefully. Alternative measures of financing constraints are typically based on balance-sheet information (Almeida et al., 2021). We acknowledge some disadvantages associated with the subjective nature of our measure of financing constraints. However, our measure has certain advantages. First, it captures both financing availability and financing cost (interest rates, fees and collateral requirements). Second, it comprises all alternative forms of external financing that are common and often indistinguishable in developing countries (bank financing, equity financing, trade/supplier finance, informal finance, etc.). Third, paradoxically it may better reflect reality. Claessens and Tzioumis (2006) argue that balance-sheet information in many developing countries is low quality, inconsistent and mostly unaudited. Instead, information based on micro-survey data reflecting directly firms' views may be more valuable at least with regards to the developing countries. Bouton and Tiongson (2010) document a significant association between subjective appraisals of credit market constraints and objectively measurable indicators. Finally, survey information may better capture firms' decisions in conditions of uncertainty. Table 2.1 reports the average value of ACCESS and the range of FSIND across countries. Given the temporal evolution of the latter, we prefer not to display the average values which would not properly reflect the fluctuations over time.

Firms from Estonia, Israel, Thailand and Sweden have low average level of ACCESS, meaning minor constraint in accessing finance, whereas firms from Afghanistan and several African countries (Ghana, Angola, Tanzania, etc.) show high average level of ACCESS, resulting in major constraint. Regarding FSIND, we observe a differentiated pattern with some countries characterized by small variations like Sweden, Croatia, Thailand and Israel. While countries like Colombia, Slovak Republic, Portugal and Argentina show extremely high levels of variability with big drop in the relative financial soundness followed by partial recovery periods. We can identify a negative relationship between ACCESS and FSIND: Figure 2.1 shows the distribution of FSIND values for each level of ACCESS, which is indeed characterized by a negative relationship. We can also notice a more significant drop in the values of FSIND as we move from ACCESS level 2 trough 4, signaling a higher sensitivity of the index to upper levels of the financial access constraints distribution.

However, while the FSIND is expected to affect firms' financing constraints, its effect is not directly observed. Therefore, we control for other characteristics of firms that could mitigate the effect. These include the firm's age (AGE), size (SIZE), sector of activity (SECTOR), location of operations (LOCATION), foreign ownership (OWNFOR), state ownership (OWNGOV), whether the firm is an exporter of goods and services (EXPORT) and whether the firm is a local subsidiary of a foreign firm (SUBSID). The ES provide the data for all the firm-level controls. Many studies have documented the significant effect of the specific characteristics of firms on their financing constraints (Beck and Laeven, 2006, Mertzanis, 2019).

Moreover, we use country-level controls to capture the role of economic and institutional factors in mitigating the FSIND effect on firm's financing constraints. We use the World Development Indicators for macro-economic variables, the World Bank's Country Policy and Institutional Assessments (CPIA) for institutional variables, the Center of Government (COG) for political variables and the Global Financial Inclusion (GFI) database for financial access variables. Table 2.2 presents the summary statistics of the variables used in the analysis. Table B3 in the Appendix presents their pairwise correlations. The correlations and the VIF value do not show severe collinearity between the FSIND and among the firm-specific variables and we therefore include them all in the regression analysis.

Table 2.1: Comparison between the ACCESS average values and minimum-maximum range of FSIND and total number of considered firms in each country.

Country	Mean ACCESS	FSIND range	Total firms	Country	Mean ACCESS	FSIND range	Total firms	Country	Mean ACCESS	FSIND range	Total firms
Afghanistan, Islamic Republic of	2.37	[-3.14 , 4.83]	412	Estonia	0.41	[-5.13 , 2.53]	281	Panama	0.90	[-4.57 , 1.91]	373
Albania	0.74	[-2.33 , 2.63]	368	Georgia	1.06	[-4.23 , 3.62]	368	Papua New Guinea	0.74	[-4.61 , 4.67]	76
Angola	2.45	[-2.39 , 2.64]	343	Ghana	2.52	[-2.04 , 2.30]	722	Paraguay	1.17	[-3.83 , 3.35]	734
Argentina	1.91	[-8.48 , 9.65]	2,033	Guatemala	1.44	[-3.76 , 4.90]	938	Peru	1.19	[-3.75 , 4.01]	2,007
Armenia, Republic of	1.72	[-2.67 , 4.15]	370	Honduras	1.53	[-3.35 , 2.00]	695	Philippines	0.79	[-5.11 , 2.68]	1,292
Bangladesh	1.81	[-3.38 , 2.04]	1,448	Hungary	0.79	[-6.10 , 5.22]	317	Poland	1.07	[-2.80 , 1.31]	545
Belarus	0.99	[-1.63 , 1.88]	364	India	1.16	[-3.86 , 3.00]	9,255	Romania	1.49	[-2.60 , 2.55]	543
Bhutan	1.13	[-5.01 , 1.25]	262	Indonesia	1.19	[-4.55 , 2.72]	1,319	Russian Federation	1.32	[-3.87 , 4.02]	4,092
Bolivia	1.40	[-3.64 , 2.15]	719	Israel	0.53	[-3.89 , 3.13]	489	Rwanda	1.67	[-1.76 , 2.66]	249
Bosnia and Herzegovina	1.23	[-5.25 , 1.77]	371	Kazakhstan	0.86	[-2.49 , 4.89]	581	Slovak Republic	1.05	[-4.22 , 4.22]	275
Botswana	1.48	[-1.55 , 1.13]	275	Kenya	1.39	[-1.93 , 4.07]	778	Slovenia	1.20	[-4.21 , 4.11]	281
Bulgaria	0.96	[-3.44 , 4.71]	298	Kosovo, Republic of	2.01	[-2.96 , 2.99]	209	Solomon Islands	1.17	[-1.52 , 5.00]	161
Burundi	1.92	[-1.53 , 3.47]	167	Kyrgyz Republic	1.21	[-2.38 , 4.02]	279	Sri Lanka	1.60	[-1.06 , 5.40]	599
Cambodia	1.22	[-2.10 , 4.90]	828	Latvia	1.11	[-3.25 , 2.58]	343	Sweden	0.61	[-2.22 , 1.18]	605
Cameroon	2.13	[-2.67 , 4.56]	361	Lebanon	1.76	[-2.95 , 1.83]	569	Tanzania	2.33	[-4.93 , 5.56]	782
Central African Republic	2.09	[-4.73 , 2.57]	161	Lesotho	2.17	[-4.28 , 2.62]	160	Thailand	0.58	[-8.49 , 8.03]	993
Chile	1.37	[-2.67 , 3.39]	1,034	Lithuania	0.91	[-2.43 , 2.63]	274	Trinidad and Tobago	1.78	[-5.14 , 2.74]	380
China, P.R.: Mainland	0.81	[-7.11 , 9.89]	2,683	Macedonia, FYR	1.32	[-4.40 , 4.82]	370	Turkey	0.73	[-5.14 , 1.71]	1,330
Colombia	1.72	[-2.52 , 3.44]	1,937	Madagascar	1.24	[-2.56 , 1.96]	347	Uganda	1.79	[-2.15 , 2.09]	747
Costa Rica	2.06	[-2.92 , 2.16]	542	Malaysia	1.37	[-6.08 , 4.10]	1,011	Ukraine	1.33	[-3.35 , 2.60]	994
Croatia	1.28	[-1.05 , 2.40]	370	Mexico	1.51	[-2.29 , 3.46]	1,471	Uruguay	1.24	[-2.18 , 3.08]	941
Czech Republic	1.08	[-4.51 , 2.63]	261	Moldova	0.65	[-4.69 , 1.41]	361	Vietnam	0.90	[-5.94 , 6.11]	962
Djibouti	1.11	[-10.40 , 74.68]	274	Namibia	2.01	[-5.83 , 5.91]	587	West Bank and Gaza	2.02	[-3.84 , 4.44]	442
Dominican Republic	1.37	[-3.75 , 2.11]	723	Nicaragua	1.14	[-7.08 , 2.66]	668	Zambia	1.86	[-3.92 , 1.66]	715
Ecuador	1.47	[-1.09 , 2.67]	729	Nigeria	1.55	[-4.29 , 2.59]	2,589				
El Salvador	1.46	[-5.08 , 2.22]	1,069	Pakistan	1.31	[-1.99 , 4.56]	1,216				

Table 2.2: List of variables used to predict ACCESS, with sources, aggregation level, total number of observations and descriptive summary statistics. Top table reports numeric variables, bottom table reports ordinal variables, i.e. variables with discrete values such as ranking.

Variable	Description	Source	Aggregation Level	Obs	Mean	S.D.	Min	P25	Median	P75	Max
FSIND	FSIND index	Authors	Country	64,717	-0.16	2.57	-10.33	-1.74	-0.26	1.61	5.61
GDPCAP	1-Y lag of GDP per capita (constant 2000 US$)	WDI	Country	63,894	5901.67	6726.12	242.85	1544.62	3692.97	8947.74	53408.79
INFLDFL	Inflation, GDP deflator (annual %)	WDI	Country	63,894	5.73	6.97	-3.85	2.98	3.98	6.57	52.99
LENDINT	1-Y lag of Lending interest rate (%)	WDI	Country	60,757	12.73	6.23	4.59	8.46	10.88	15.65	56.13
FININD	1-Y lag of IMF: financial institutions index	WDI	Country	63,894	0.24	0.17	0.03	0.13	0.2	0.28	0.98
FINDEP	1-Y lag of Financial system deposits to gdp (%)	WDI	Country	63,894	43.49	27.54	8.33	24.07	36.78	60.46	222.97
GENDEQ	1-Y lag of CPIA gender equality rating (1=low to 6=high)	CPIA	Country	49,032	3.76	0.7	1.5	3	3.82	4.5	5
BILHUM	CPIA building human resources rating (1=low to 6=high)	CPIA	Country	49,032	3.82	0.33	2.5	3.5	4	4	4.5
FISPOL	1-Y lag of CPIA fiscal policy rating (1=low to 6=high)	CPIA	Country	49,032	3.74	0.49	2.5	3.5	3.5	4	5
STABDEM	Stability of Democratic Institutions rating	COG	Country	27,918	6.09	2.32	1	5	6.5	7.5	10
LIMLEND	Limitations on lending to the government (%)	COG	Country	18,853	0.71	0.17	0.01	0.62	0.73	0.8	1
OUTLOAN	Outstanding loans from commercial banks (% of GDP)	GFI	Country	64,634	2209.51	28898.53	3.04	24.78	37.89	50.39	530000
NUMBRW	log of Household Borrowers	GFI	Country	36,974	3e+07	7.8e+07	3980	760000	3100000	9e+06	4.9e+08
AGE	log of the years since the firm's establishment	ES	Firm	63,894	3.08	0.53	1.1	2.71	3.04	3.37	5.39
EXPORT	percent of firm's sales directly exported	ES	Firm	63,894	7.44	21.48	0	0	0	0	100
OWNFOR	percent of firm's stock owned by foreign investors	ES	Firm	63,894	6.82	23.23	0	0	0	0	100
OWNGOV	percent of firm's stock owned by the state	ES	Firm	63,894	0.55	5.66	0	0	0	0	100

							Level				
Variable	Description	Source	Aggregation Level	Obs	Mean	S.D.	0	1	2	3	4	5
ACCESS	access to finance (0-4), 4=highest difficulty	ES	Firm	63,894	1.33	1.25	34.7%	23.1%	22.7%	13.3%	6.2%	
SIZE	1=Small(<20),2=Medium(20-99),3=Large(100 And Over)	ES	Firm	63,894	1.78	0.77		43.5%	35.4%	21.1%		
LOCATION	1=capital city,2=city with over 1 million,3=city btwn 1/4 and 1 million, 4=city btwn 50K and 250K,5=city with less than 50K	ES	Firm	63,427	2.79	1.19		11.5%	36.9%	23.8%	16%	11.7%
SUBSID	0=independent firm,1=subsidiary of a larger firm	ES	Firm	63,894	0.19	0.39	81.5%	18.5%				
LISTED	whether the firm is listed in an exchange, 1=yes, 0=no	ES	Firm	63,894	0.04	0.2	95.9%	4.1%				
VETOPWR	Legislature Veto Power	COG	Country	18,588	0.86	0.34	13.6%	86.4%				

Notes: Macro-economic variables are collected from World Development Indicators (WDI) https://databank.worldbank.org/source/world-development-indicators, institutional governance variables are collected from Country Policy And Institutional Assessment Primary (CPIA) https://datacatalog.worldbank.org/dataset/country-policy-and-institutional-assessment, political variables are collected from Center of Government (COG) https://www.worldbank.org/en/topic/governance/brief/center-of-government-global-solution-group and financial access variables are collected from Global Financial Inclusion (GFI) https://datacatalog.worldbank.org/dataset/global-financial-inclusion-global-findex-database. World Bank's Enterprise Surveys (ES) variables can be found athttps://www.enterprisesurveys.org/en/enterprisesurveys.

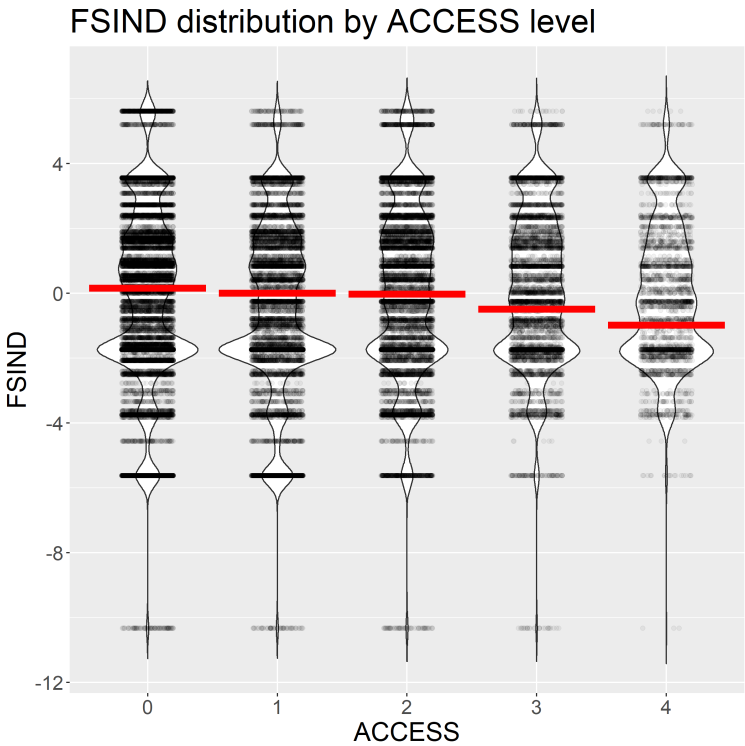


Figure 2.1: Distribution of FSIND values for each level of ACCESS. Dots represent each firms' FSIND value and red bars represent FSIND average values.

In our sample, most firms are small and medium size rather than large size, they are exporters of goods and services and operate in the large urban than rural areas. Most firms are private, non-listed firms and a minority of them are subsidiaries of foreign companies, owned by domestic and foreign owners, with only few of them owned by the state. We subsequently match firm-level information with our country-level FSIND and other economic and institutional information. However, our FSIND has missing values for some countries, which reduces our full sample to 76 countries during 2010-2018 with a rate of missing values not exceeding 30% and about 63,894 firms. Our sample has an unbalanced panel structure, which led us to apply two alternative missing value imputation methods, i.e., the MC-SVD and BTF methods, so as to take advantage of the temporal dependence of the variables between years. Thus, we standardize numerical variables and rescale ordinal variables to the range $[0,1]$ and use country-mean value imputation for numeric variables and country-median value imputation for ordinal variables.

2.3.3 Identification strategy and estimation model

Identifying a causal effect running from the country-level FSIND to the firm-level financial behavior is challenging due to the possible presence of unobserved countrywide factors that are simultaneously linked with both the digital adoption conditions and firm performance. We include alternative model specifications to reduce this possibility. As a first step to causal identification, we include fixed effects at country, year, and sector levels. Country effects control for time-invariant conditions in a firm's country. Year effects control for time-varying shocks, which affect the behavior of all firms in our sample (e.g., technological shocks). Sector effects control for any time-invariant and industry-specific conditions (i.e., competition, regulation) that affect firm performance. Rajan and Zingales (1998) found that financing constraints of firms are stronger in sectors that require more external finance. Carreira and Lopes (2016) show that firms in the service sector suffer from more severe financial constraints than those in manufacturing. Deploying both fixed effects and diverse firm-specific variables could control for some of the unobserved influence on financing constraints. This implies that we identify the trend of our FSIND only from changes between consecutive years within the same country, as shown in the dynamic map in Figure B4 in the Appendix. We attempt to capture the potentially remaining omitted-variable bias by using endogeneity analysis later in the document.

Since the outcome variable is an ordinal one, we use an ordered probit model and the maximum likelihood estimator for estimating the regression (Greene, 2012). In this setting, we measure the key predictor variable at the country level whereas we measure the outcome variable at the firm level. Moulton (1990) identified the statistical bias that results from the attempt to measure the effect of aggregate policy variables on micro units. Consequently, we cluster the standard errors at the country level. Our setting also implies that in a given country and year, there are several different firm-level observations per one key predictor observation. The error term of the estimation might be large since it is difficult to fit all the outcome points at the same time, thereby inducing a more conservative estimate of the effect of the key predictor variable. We also test for the impact of outliers and data imbalances by capping the maximum number of firms in each country and removing countries with extreme values. We perform sensitivity and endogeneity analysis based on the use of alternative measures of the key variables and alternative estimation methods using instrumental variables. After assessing the model's stability with respect to the sample, we check for its robustness by including additional control variables. The estimation model assumes that the firms' response is described by the following equation:

$$ACCESS_{if_jt} = \beta_0 + \beta_1 FSIND_{it} + \boldsymbol{\beta_2 X}_{if_jt} + \boldsymbol{\beta_3 K}_{it} + u_{if_jt} \tag{2.2}$$

where $ACCESS_{if_jt}$ is the underlying probability that the firm f_j ($f_j = 1,\ldots,M_i$) among all M_i firms in country i ($i = 1,\ldots,N$) and year t ($t = 1,\ldots,T$) perceives access to finance to be no, low, moderate, major or severe constraint; $FSIND_{it}$ is the index of financial soundness of country i and year t; $\boldsymbol{X}_{if_jt}$ is the vector of firm-specific control variables per firm f_j in country i and year t; and $\boldsymbol{K}_{it}$ is the vector of country-level control variables per country i and year t. The term u_{if_jt} is the composite error term component that comprises the sum of η_i, λ_t and ε_{it}, where η_i accounts for unobservable country-specific effects, λ_t accounts for year-specific effects and the ε_{it} is a disturbance parameter that is assumed to vary across countries and years. Note that when analyzing ordinal data with a probit model, there is no equivalent statistic to the OLS based R^2 to evaluate the goodness-of-fit. The model estimates are maximum likelihood ones obtained through an iterative process. Similarly, unlike the OLS case, the coefficients of the probit estimation should be interpreted as changes in conditional probability of the outcome variable following changes in the regressors. Finally, we are well aware of the difficulty in interpreting the observed correlations as causal effects. We therefore interpret our results as strength of association rather than causation, and the use of the words "prediction" or "impact" or "effect" is made to simplify exposition.

2.4 Analysis of the results

2.4.1 Baseline Results

Table 2.3 presents the results of the probit model. The first column shows the estimates of the baseline model. The FSIND is statistically significant in the whole sample, documenting a negative association between financial soundness conditions and firms' access to finance across countries. The second and third columns show the estimates after splitting the sample into high- and low-income countries by the median level of GDP per capita. The remaining columns show the estimates for each level of firm size. Fafchamps and Labonne (2017) show that splitting sample delivers more predictive power. The improvement operates through a lower likelihood that relevant hypotheses are left untested. The FSIND is negative throughout and significant in the whole sample, for small firms and for those operating in low-income countries. It appears that financial stability considerations are relatively more important in affecting the financing constraints of small-size firms in less developed countries. These results appear to be in line with Laeven (2003) where the analysis of a panel of 400 firms from 13 developing countries shows that liberalization has an impact on the financial constraints of small firms, whereas large firms do not experience any change. Moreover, older firms, with strong foreign ownership and operating in large urban areas face lower financing constraints. Publicly listed firms also experience lower constraints. Further, Table 2.4 reports the marginal effects for each predicted level of ACCESS and Figure 2.2 highlights the marginal probability of ACCESS compared with the increase of FSIND. Thus, changes in FSIND appear to be associated with stronger marginal effects at the higher levels of financing constraints. As a consequence, the effects of policies that improve financial stability will be relatively more beneficial for the finance-hungry firms.

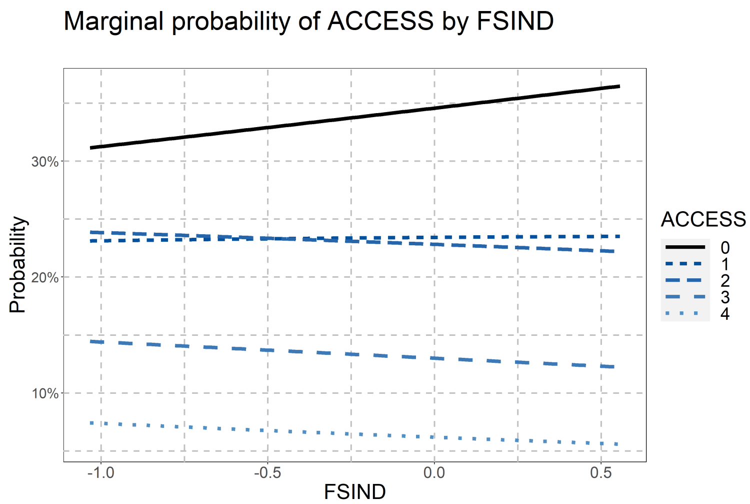


Figure 2.2: Marginal probability of each level of ACCESS compared with the increase of FSIND.

2.4.2 Sensitivity tests

We then run several sensitivity tests to ensure the stability of the estimated coefficients and make our findings robust against potential measurement error. A first sensitivity test is run by splitting the sample in Table 2.3. A second sensitivity test focuses on the clustering specification of standard errors. The unbalanced panel nature of the dataset which contains different number of firms in each country can cause heteroskedasticity bias in the estimation of the coefficients (Abadie et al., 2017). Thus, we test for potential bias in the coefficients by using alternative clustering specifications. Table 2.5 shows the effect of the different specifications. Given the negligible difference between the alternative clustering approaches, we maintain our clustering strategy at the country level. A third sensitivity test focuses on the measurement of the ordinal outcome variable, ACCESS. The five levels of the outcome variable may cause model overfit, i.e. the model could adapt too closely to relationships between each

Table 2.3: Predicting ACCESS with ordinal probit model - OLS.

Variable	Baseline	High Income Countries	Low Income Countries	Small Firms	Medium Firms	Large Firms
FSIND	-0.0961**	0.00705	-0.143***	-0.151**	-0.0636	-0.101
	(0.0391)	(0.0834)	(0.0377)	(0.0759)	(0.0454)	(0.0698)
LISTED	-0.112***	-0.184***	-0.0684	-0.104	-0.148***	-0.0922***
	(0.0409)	(0.0610)	(0.0451)	(0.0797)	(0.0570)	(0.0357)
AGE	-0.763***	-0.544*	-1.043***	-0.755***	-0.617***	-0.847***
	(0.169)	(0.279)	(0.206)	(0.273)	(0.171)	(0.221)
SIZE	-0.0631***	-0.102***	-0.0252			
	(0.0136)	(0.0152)	(0.0180)			
SUBSID	-0.0290	-0.0143	-0.0417	-0.0601	0.00327	-0.0282
	(0.0317)	(0.0487)	(0.0380)	(0.0416)	(0.0465)	(0.0293)
LOCATION	0.180**	0.263*	0.104	0.174**	0.234**	0.121
	(0.0910)	(0.147)	(0.0970)	(0.0864)	(0.114)	(0.0993)
EXPORT	-0.0759*	-0.102	-0.0378	-0.0494	-0.0497	-0.0674*
	(0.0424)	(0.0744)	(0.0380)	(0.102)	(0.0609)	(0.0385)
OWNFOR	-0.244***	-0.250***	-0.244***	-0.274***	-0.239***	-0.207***
	(0.0279)	(0.0476)	(0.0364)	(0.0500)	(0.0394)	(0.0383)
OWNGOV	-0.292	-0.0221	-0.433*	-0.205	-0.331	-0.288
	(0.190)	(0.0991)	(0.248)	(0.137)	(0.289)	(0.200)
Observations	64,717	32,029	32,688	28,201	23,014	13,502
Pseudo R^2	0.0437	0.0431	0.0429	0.0522	0.0382	0.0343
Wald χ^2	10161.42	12908.46	8773.01	16670.74	7511.31	6223.87
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Std. Err.	Country	Country	Country	Country	Country	Country

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country. The bottom part of the table reports which fixed effects are used in each model specification. First column reports the baseline model, second and third report the countries' income-based subset models and last three columns report the firms' size-based subset models. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

single level and the independent variables leading to loss of generalization power (Agresti, 2012). For this reason, we test the model's ability to generalize the causal effect between outcome and independent variables by transforming the five levels ACCESS variable into a binary one. We examine the effect of FSIND on different specifications for the binary transformation of ACCESS, grouping the levels above a selected threshold tr, with $tr = 1,2,3,4$: all levels above tr will be assigned the label "1" and the remaining ones will be assigned the label "0". Table 2.6 shows that the FSIND coefficient is negative and significant in all binary specifications except for the $tr = 1$ case, where it becomes positive and not significant. The reason for the latter behavior may be the following: the small number of firms with "0" ACCESS level affects the distribution of the outcome variable to be predicted, resulting in a coefficient quite close to zero and with no statistical significance (Agresti, 2012). A fourth sensitivity test is concerned with the structure of the sample. It focuses on the impact of data imbalances and outliers. Since the number of firms in each country differs substantially, we cap the total number of firms at different levels, randomly selecting the countries to be retained and averaging the coefficients' estimation over 10 sampling trials. The first four columns of Table 2.7 report the results after using capping limits of 850, 900, 950 and 1000

Table 2.4: Marginal effects for baseline model - OLS.

	ACCESS				
	0	1	2	3	4
FSIND	0.0337***	0.0139***	-0.00658***	-0.0204***	-0.0206***
	(0.00499)	(0.00270)	(0.00139)	(0.00340)	(0.00327)
LISTED	0.0628***	0.0260***	-0.0123***	-0.0381***	-0.0385***
	(0.0149)	(0.00734)	(0.00331)	(0.00957)	(0.00983)
AGE	0.271**	0.112**	-0.0529**	-0.164*	-0.166**
	(0.138)	(0.0551)	(0.0231)	(0.0930)	(0.0765)
SUBSID	0.0382***	0.0158***	-0.00746***	-0.0231***	-0.0234***
	(0.0126)	(0.00544)	(0.00260)	(0.00721)	(0.00839)
LOCATION	-0.0747**	-0.0309**	0.0146*	0.0453***	0.0457**
	(0.0317)	(0.0150)	(0.00779)	(0.0172)	(0.0217)
EXPORT	0.0499	0.0206	-0.00974	-0.0302*	-0.0305
	(0.0306)	(0.0137)	(0.00696)	(0.0178)	(0.0196)
OWNFOR	0.0542***	0.0224***	-0.0106***	-0.0328***	-0.0332***
	(0.0153)	(0.00835)	(0.00391)	(0.0103)	(0.00986)
OWNGOV	0.116***	0.0481***	-0.0227**	-0.0705***	-0.0712***
	(0.0359)	(0.0186)	(0.00953)	(0.0212)	(0.0244)

Notes: The table reports the marginal effects for each predicted level of ACCESS and their standard error (in parentheses) in the baseline setting of Table 2.3. Estimation method is OLS with standard errors clustered by firm's country. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

firms in each country respectively. The capping limits have been selected taking as a reference the interquartile range of the distribution of the total number of firms in each country. The FSIND coefficient remains negative and significant for all the considered levels. The last column shows that FSIND remains negative and significant after excluding all countries that have extreme values in one or more independent variables, namely Lesotho, New Guinea, China, India and Russia. Extreme values and relative countries have been selected according to a thresholding of the 5^{th} and 95^{th} percentiles of each independent variables.

2.5 Endogeneity and Robustness Analysis

2.5.1 Endogeneity Analysis

In order to identify the causal effect, we use a cross-section of data capturing the individual firms' experience of financing constraints for multiple years. This limits the possibility of reverse causality: observing a change in the financial soundness conditions of a country as a result of a change in a firm's experience of constraints in obtaining external finance is unlikely. We also experiment with different specifications of models and control variables to ensure that our causal effect does not suffer from improper specification. We further deal with potential asymmetry of information problems by applying proper clustering of estimated standard errors. Despite the inclusion of fixed effects controlling for invariant country factors, our estimates will not produce unbiased assessments of the FSIND effect on firms' financing constraints, because of the possible presence of unobserved factors affecting the financial soundness conditions and the financing constraints of firms simultaneously. For example, countries that in recent years may have experienced improving conditions for firms' access to finance may have also implemented policies that improved the health conditions of financial institutions and markets (e.g., improved prudential ratios, governance institutions, etc.) during the same period, thereby increasing the soundness of the financial system as a whole. This possibility means that the covariance term $\mathrm{Cov}(FSIND_{it}, u_{if_jt})$ is non-zero, because

Table 2.5: Predicting ACCESS with ordinal probit model - OLS.

Variable	Standard Errors	Robust Standard Errors	Country-Year Clustering	Country Clustering
FSIND	-0.0961**	-0.0961***	-0.0961***	-0.0961**
	(0.0399)	(0.0352)	(0.0254)	(0.0391)
LISTED	-0.112***	-0.112***	-0.112***	-0.112***
	(0.0227)	(0.0227)	(0.0401)	(0.0409)
AGE	-0.763***	-0.763***	-0.763***	-0.763***
	(0.0892)	(0.0900)	(0.162)	(0.169)
SIZE	-0.0631***	-0.0631***	-0.0631***	-0.0631***
	(0.00631)	(0.00636)	(0.0135)	(0.0136)
SUBSID	-0.0290**	-0.0290**	-0.0290	-0.0290
	(0.0121)	(0.0120)	(0.0314)	(0.0317)
LOCATION	0.180***	0.180***	0.180**	0.180**
	(0.0204)	(0.0205)	(0.0893)	(0.0910)
EXPORT	-0.0759***	-0.0759***	-0.0759*	-0.0759*
	(0.0216)	(0.0219)	(0.0412)	(0.0424)
OWNFOR	-0.244***	-0.244***	-0.244***	-0.244***
	(0.0202)	(0.0206)	(0.0287)	(0.0279)
OWNGOV	-0.292***	-0.292***	-0.292	-0.292
	(0.0814)	(0.0857)	(0.190)	(0.190)
Observations	64,717	64,717	64,717	64,717
Pseudo R^2	0.0437	0.0437	0.0437	0.0437
Wald χ^2	7713.42	8319.34	82695.62	53591.33
Year effects	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with different standard errors estimations. The bottom part of the table reports which fixed effects are used in each model specification. First and second columns report the classical and robust standard errors estimation. Third and fourth columns report the country-year and country clustering for standard errors estimation. The *, ** and *** symbols denote the p-values at 10th, 5th and 1st significance level, respectively.

even if it is conditional on the fixed effects, the FSIND might be endogenous to financing constraints decisions. For this reason, we will use instrumental variables (IV) methods that deploy the 2SLS estimator to check the robustness of our estimates. We apply the IV method under four different groups of control variables (Table 2.8).

We use the log number of household borrowers in a country (NUMBRW) as the external instrument in the IV analysis. NUMBRW correlates highly with the FSIND index (-0.461) and poorly with ACCESS (-0.0152). Santoso and Sukada (2009) document the importance of household borrowing risk for financial stability. The IMF provides the data. We use the instrumental variables two-stage least square (2SLS) model with the following specification:

$$\begin{aligned} FSIND_{it} &= \gamma_0 + \gamma_1 NUMBRW_{it} + e_{it} \\ ACCESS_{if_jt} &= \beta_0 + \beta_1 FSIND_{it} + \boldsymbol{\beta_2 X}_{if_jt} + \boldsymbol{\beta_3 K}_{it} + u_{if_jt} \end{aligned} \tag{2.3}$$

Table 2.6: Predicting ACCESS with binary probit model - OLS.

Variable	0 vs 1 2 3 4	0 1 vs 2 3 4	0 1 2 vs 3 4	0 1 2 3 vs 4
FSIND	0.0159	-0.167***	-0.186**	-0.170*
	(0.0694)	(0.0595)	(0.0722)	(0.103)
LISTED	-0.119**	-0.134***	-0.124**	-0.0170
	(0.0471)	(0.0432)	(0.0542)	(0.0657)
AGE	-1.040***	-0.790***	-0.547**	-0.321
	(0.259)	(0.175)	(0.244)	(0.217)
SIZE	-0.0295**	-0.0771***	-0.0993***	-0.0984***
	(0.0142)	(0.0202)	(0.0174)	(0.0166)
SUBSID	-0.0180	-0.0356	-0.0414	-0.0569
	(0.0303)	(0.0424)	(0.0352)	(0.0364)
LOCATION	0.207*	0.189*	0.115*	0.106
	(0.117)	(0.110)	(0.0659)	(0.0778)
EXPORT	-0.0866**	-0.0665	-0.0782	-0.0590
	(0.0377)	(0.0560)	(0.0549)	(0.0746)
OWNFOR	-0.255***	-0.269***	-0.237***	-0.192***
	(0.0370)	(0.0285)	(0.0333)	(0.0509)
OWNGOV	-0.408	-0.189	-0.121	-0.0431
	(0.248)	(0.164)	(0.200)	(0.265)
Observations	64,717	64,717	64,717	64,717
Pseudo R^2	0.0784	0.0766	0.0858	0.0805
Wald χ^2	8749.12	9520.18	12392.67	15987.78
Year effects	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	Yes	Yes
Clustered Std. Err.	Country	Country	Country	Country

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country. The bottom part of the table reports which fixed effects are used in each model specification. All columns report the results of the binary probit model when the ACCESS variable is grouped into a binary variable splitting levels above and below a certain threshold. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

where $\boldsymbol{K}$ is the matrix whose columns are the country-level control variables per country i and year t described in Table 2.8. Tables from 2.9 to 2.12 report the results of 2SLS model for each of the four groups. The first stage results show that the external instruments are robust. The IV results document that the FSIND remains a significant and negative predictor of firms' financing constraints. More specifically, after controlling for the impact of macroeconomic and monetary conditions (Table 2.9), higher values of the FSIND index are associated with lower financing constraints of firms. The lower is the 1-Year lag of lending interest rate, the stronger is the effect. Further, we control for the impact of financial conditions (Table 2.10), the effect of our FSIND index remains negative and significant. The lower is the volume of outstanding loans made by commercial banks, the stronger is the beneficial effect on firms' financing constraints. These findings appear to be in line with the view that monetary stability may be necessary but not a sufficient condition for financial stability (Borio and Lowe,

Table 2.7: Predicting ACCESS with ordinal probit model - OLS.

Variable	Cap to 850	Cap to 900	Cap to 950	Cap to 1000	Outlier Countries Excluded
FSIND	-0.1188 *	-0.1247 *	-0.1293 *	-0.1362 **	-0.148 ***
	-0.0656	-0.062	-0.0644	-0.0638	-0.0565
LISTED	-0.1269 ***	-0.1294 ***	-0.1261 ***	-0.1263 ***	-0.140 ***
	-0.0352	-0.0347	-0.0342	-0.0338	-0.0319
AGE	-0.2669 ***	-0.2693 ***	-0.2724 ***	-0.2707 ***	-0.242***
	-0.0524	-0.0517	-0.0511	-0.0505	-0.0476
SIZE	-0.1424 ***	-0.1407 ***	-0.1403 ***	-0.1403 ***	-0.152 ***
	-0.0166	-0.0164	-0.0162	-0.016	-0.155
SUBSID	-0.0913 ***	-0.0894 ***	-0.0864 ***	-0.0862 ***	-0.0817 ***
	-0.0153	-0.0151	-0.0149	-0.0147	-0.0138
LOCATION	-0.1020 ***	-0.0999***	-0.1012 ***	-0.1008 ***	-0.0798 ***
	-0.0253	-0.0249	-0.0246	-0.0244	-0.0233
EXPORT	-0.1047 ***	-0.1070 ***	-0.1090 ***	-0.1098 ***	-0.129 ***
	-0.027	-0.0266	-0.0262	-0.0259	-0.0246
OWNFOR	-0.0429 ***	-0.0419 ***	-0.0417 ***	-0.0402 ***	-0.0347 ***
	-0.0124	-0.0123	-0.0121	-0.012	-0.0115
OWNGOV	-0.0805 *	-0.0799 *	-0.0773 *	-0.0816 *	-0.0941 **
	-0.0417	-0.0411	-0.0406	-0.0402	-0.0377
Observations	41,953	43,003	44,032	44,931	48,451
Pseudo R^2	0.0512	0.051	0.0507	0.0504	0.0493
Wald χ^2	12462.56	8345.89	7923.23	6765.71	5529.37
Year effects	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	Yes	Yes	Yes
Clustered Std. Err.	Country	Country	Country	Country	Country

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country. The bottom part of the table reports which fixed effects are used in each model specification. First four columns report the estimation when capping the total number of firms in each country to different thresholds. Last column reports the estimation when removing countries with extreme values in one or more independent variables, namely Lesotho, New Guinea, China, India and Russia. The *, ** and *** symbols denote the p-values at 10th, 5th and 1st significance level, respectively.

Table 2.8: List of groups of variables used as controls.

Group	Variable	Description
Macro-economic	GDPCAP	1-Year lag of GDP per capita
	INFLDFL	Inflation deflator
	LENDINT	1-Year lag of Lending interest rate
Financial access	FININD	1-Year lag of IMF's Financial Institutions Index
	FINDEP	1-Year lag of financial system deposits to GDP ratio
	OUTLOAN	Outstanding loans from commercial banks
Institutional governance	GENDEQ	1-Year lag of gender equality rating
	BILHUM	Building human resources rating
	FISPOL	1-Year lag of fiscal policy rating
Political	STABDEM	Stability of democratic institutions rating
	LIMLEND	Percentage of limitations on lending to the government
	VETOPWR	Legislature Veto Power rating

2002, Borio et al., 2003). In this view, financial risks may grow beneath the surface of low-inflation. Excessive focus on monetary stability, as a condition for maintaining expectations of long-term economic growth, may in turn cause corporate indebtedness and discrepancies between prices of asset with varying maturities perpetuating financial instability (Shirakawa, 2012). Moreover, after controlling for the impact of social conditions (Table 2.11), the effect of our FSIND index also remains negative and significant. The beneficial effect on firms' financing constraints is stronger, the higher is the level of gender equality in the country and the lower is the public's perception of fiscal policy fairness in the country. Ozili (2020) argues that social activism has had adverse effects on financial stability in the post-2008 era in developing countries. He also finds that gender equality and environmental sustainability advocacy have improved financial stability. Finally, after controlling for the impact of political institutions (Table 2.12), the effect of our FSIND index on financing constraints again remains negative and significant. The beneficial effect on firms' financing constraints is stronger, the higher is the stability of democratic institutions, and the stronger is the veto powers of a country's legislature, whilst any restrictions imposed upon the level of government lending may weakly mitigate the effect of the FSIND on financing constraints. Beck et al. (2020) stress the complex link between politics and finance, and Funke et al. (2016) find that financial crises cause a decrease of government majorities and an increase in political polarization leading to policy uncertainty. In all control groups, the results remain robust to alternative model settings regarding the inclusion of firm size and industry fixed effects to capture firm and industry heterogeneity. Our IV estimations show that the firm-specific characteristics of firms remain significant controls. Generally, older firms which are subsidiaries of multinational firms are associated with lower financing constraints. Export-oriented firms which are located in larger cities also experience lower financing constraints. Finally, firms with considerable foreign and government ownership stakes appear to experience lower constraints. Overall, our analysis shows that the FSIND effect is not significantly affected by unobserved factors which strengthens its statistical independence and explanatory relevance. The Kleibergen-Paap rk LM statistic allows us to reject the null hypothesis that our model is under-identified ($p < 0.001$) (Kleibergen and Paap, 2006). As a result, there is some reason for confidence in the validity of our chosen instruments. The Portmanteau statistic shows the absence of auto-correlation (Inoue and Solon, 2006, Wursten, 2018). Since the two-step process of the 2SLS model can be affected by data imbalances, we also use an alternative IV estimation procedure based on the conditional mixed-process (CMP) that fits both equations described in (2.3) simultaneously (Roodman, 2011). Tables B6 to B9 in the Appendix report the results of CMP model for each of the four groups. The statistical significance of the Arellano-Bond ρ coefficient (Arellano and Bond, 1991) indicates that the null hypothesis of no endogeneity is rejected, which justifies the use of the IV methods.

2.5.2 Further Robustness checks

In order to further check the robustness of our results to omitted variable bias, we implement a novel technique that uses the Oster test (Oster, 2019). The distinctive feature of this technique is that it allows for a "full adjustment" by exploiting information not only on coefficient movements after the inclusion of new controls, but also on movements in R^2 values so as to compute bounding values for the treatment effect. The test proposes that, if a regression coefficient changes only a little when new controls are added, any remaining bias is likely to be small. Whereas if the coefficient changes considerably, there may still be a substantial omitted variable bias, undermining confidence in the coefficient estimate. Two key parameters specify the relationship between observable and unobservable variables selection and the maximum amount of variation which can be explained by the model. The first parameter δ defines the importance of the unobservable variables relative to the observable ones in influencing the outcome variable. When $\delta = 1$ the observable and the unobservable variables are equally important and affect the coefficient β in the same direction; when$0 < \delta < 1$ the unobserved variables are less important than the observed ones; the opposite holds when $\delta > 1$. The second parameter, R^2_{max} is the maximum R^2 under the full model where all (observed and unobserved) variables are included. This can be as high as 1 if the outcome variable is measured without error ($u = 0$), but it cannot be smaller than the R^2 obtained from the

Table 2.9: Predicting ACCESS with ordinal probit model with instrumental variables and macro-economic controls - 2SLS.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.0519***	-0.0384***	-0.0471***	-0.0338***	-0.0485***	-0.0368***	-0.0505***	-0.0351***
	(0.0113)	(0.0105)	(0.0105)	(0.0103)	(0.0102)	(0.0102)	(0.0117)	(0.0106)
LISTED	-0.0362***	-0.0364***	-0.0297***	-0.0298***	-0.0359***	-0.0361***	-0.0293***	-0.0295***
	(0.00860)	(0.00859)	(0.00882)	(0.00881)	(0.00868)	(0.00867)	(0.00877)	(0.00876)
AGE	-0.260***	-0.262***	-0.178**	-0.179**	-0.235***	-0.237***	-0.197***	-0.199***
	(0.0720)	(0.0723)	(0.0716)	(0.0718)	(0.0715)	(0.0717)	(0.0722)	(0.0725)
SUBSID	-0.0185**	-0.0188**	-0.0155*	-0.0157*	-0.0208**	-0.0210**	-0.0123	-0.0126
	(0.00879)	(0.00883)	(0.00841)	(0.00844)	(0.00896)	(0.00900)	(0.00822)	(0.00824)
LOCATION	0.0521**	0.0506**	0.0506**	0.0493**	0.0534**	0.0522**	0.0487**	0.0472**
	(0.0225)	(0.0234)	(0.0230)	(0.0238)	(0.0227)	(0.0235)	(0.0228)	(0.0236)
EXPORT	-0.0414***	-0.0413***	-0.0198	-0.0196	-0.0317**	-0.0316**	-0.0290**	-0.0289**
	(0.0146)	(0.0146)	(0.0145)	(0.0145)	(0.0152)	(0.0152)	(0.0138)	(0.0138)
OWNFOR	-0.0749***	-0.0749***	-0.0665***	-0.0665***	-0.0740***	-0.0740***	-0.0666***	-0.0665***
	(0.00852)	(0.00849)	(0.00821)	(0.00818)	(0.00852)	(0.00850)	(0.00832)	(0.00829)
OWNGOV	-0.0994***	-0.0993***	-0.0913**	-0.0912**	-0.0976***	-0.0975***	-0.0925**	-0.0924**
	(0.0336)	(0.0337)	(0.0355)	(0.0355)	(0.0328)	(0.0328)	(0.0366)	(0.0367)
GDPCAP		0.0321		0.0350		0.0295		0.0386
		(0.0292)		(0.0263)		(0.0262)		(0.0294)
INFLDFL		0.0382*		0.0308		0.0331*		0.0361*
		(0.0201)		(0.0194)		(0.0191)		(0.0206)
LENDINT		-0.232***		-0.202**		-0.199**		-0.238***
		(0.0826)		(0.0811)		(0.0787)		(0.0848)
First stage results								
NUMBRW	8.156**	5.438***	8.159**	5.443***	8.155**	5.439***	8.160**	5.443***
	(3.965)	(1.645)	(3.964)	(1.645)	(3.965)	(1.645)	(3.964)	(1.645)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.013	0.013	0.012	0.012	0.011	0.011	0.015	0.015
F-stat	19.72	28.10	14.03	27.21	16.57	31.27	17.09	24.59
Kleibergen-Paap	6.04**	6.01**	5.96**	5.91**	5.96**	5.91**	5.93**	6.01**
Portmanteau	1	1	1	1	1	1	1	1
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and 2SLS of Eq. (2.3) for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Null hypothesis of Kleibergen-Paap test is the under-identification of the model. Null hypothesis of Portmanteau test is the absence of auto-correlation. The *, ** and *** symbols denote the p-values at 10th, 5th and 1st significance level, respectively.

controlled regression. Both δ and R^2_{max} are unknown parameters to be chosen given the particular context and econometric model. The higher the value of δ associated to a variable, the more relevant that variable is. Table 2.13 shows the results after applying the Oster test. As further robustness check, we also multiply the R^2_{max} value by an arbitrary number $\pi > 1$ which implies the relaxing of the test's assumptions by an increase in the values of the R^2_{max}. On the contrary, small multiplying effects on R^2_{max} are more restrictive.

In Table 2.13, we observe that all the control variables present values of $|\delta|$ higher than the threshold, even in the most restrictive case of $\pi = 1$, confirming once again the validity and robustness of the analysis.

2.6 Conclusions

The impact of financial instability on corporate finance is a key policy question that reflects the effectiveness of macroprudential policies. There have been various research efforts to

Table 2.10: Predicting ACCESS with ordinal probit model with instrumental variables and financial access controls - 2SLS.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.0519***	-0.0630***	-0.0471***	-0.0565***	-0.0485***	-0.0587***	-0.0505***	-0.0609***
	(0.0113)	(0.0194)	(0.0105)	(0.0185)	(0.0102)	(0.0180)	(0.0117)	(0.0202)
LISTED	-0.0362***	-0.0361***	-0.0297***	-0.0296***	-0.0359***	-0.0358***	-0.0293***	-0.0293***
	(0.00860)	(0.00857)	(0.00882)	(0.00880)	(0.00868)	(0.00865)	(0.00877)	(0.00874)
AGE	-0.260***	-0.262***	-0.178**	-0.179**	-0.235***	-0.237***	-0.197***	-0.199***
	(0.0720)	(0.0719)	(0.0716)	(0.0714)	(0.0715)	(0.0714)	(0.0722)	(0.0721)
SUBSID	-0.0185**	-0.0186**	-0.0155*	-0.0155*	-0.0208**	-0.0208**	-0.0123	-0.0123
	(0.00879)	(0.00885)	(0.00841)	(0.00846)	(0.00896)	(0.00901)	(0.00822)	(0.00826)
LOCATION	0.0521**	0.0521**	0.0506**	0.0506**	0.0534**	0.0535**	0.0487**	0.0488**
	(0.0225)	(0.0226)	(0.0230)	(0.0231)	(0.0227)	(0.0228)	(0.0228)	(0.0228)
EXPORT	-0.0414***	-0.0413***	-0.0198	-0.0198	-0.0317**	-0.0316**	-0.0290**	-0.0290**
	(0.0146)	(0.0147)	(0.0145)	(0.0146)	(0.0152)	(0.0153)	(0.0138)	(0.0138)
OWNFOR	-0.0749***	-0.0749***	-0.0665***	-0.0666***	-0.0740***	-0.0740***	-0.0666***	-0.0666***
	(0.00852)	(0.00857)	(0.00821)	(0.00825)	(0.00852)	(0.00857)	(0.00832)	(0.00837)
OWNGOV	-0.0994***	-0.0993***	-0.0913**	-0.0912**	-0.0976***	-0.0974***	-0.0925**	-0.0924**
	(0.0336)	(0.0337)	(0.0355)	(0.0355)	(0.0328)	(0.0328)	(0.0366)	(0.0367)
FININD		-0.182		-0.247		-0.206		-0.225
		(0.442)		(0.401)		(0.395)		(0.452)
FINDEP		-0.179		-0.125		-0.155		-0.149
		(0.258)		(0.236)		(0.231)		(0.266)
OUTLOAN		632.9*		549.2		578.1*		606.9
		(364.3)		(342.2)		(336.7)		(373.0)
First stage results								
NUMBRW	8.156**	5.517***	8.159**	5.523***	8.155**	5.516***	8.160**	5.524***
	(3.965)	(1.968)	(3.964)	(1.967)	(3.965)	(1.968)	(3.964)	(1.967)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.013	0.013	0.012	0.012	0.011	0.011	0.015	0.015
F-stat	19.72	15.09	14.03	9.994	16.57	11.32	17.09	13.38
Kleibergen-Paap	5.92**	5.54*	5.89*	5.38	5.87*	5.53	6.03**	5.41
Portmanteau	1	1	1	1	1	1	1	1
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and 2SLS of Eq. (2.3) for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Null hypothesis of Kleibergen-Paap test is the under-identification of the model. Null hypothesis of Portmanteau test is the absence of auto-correlation. The *, ** and *** symbols denote the p-values at 10th, 5th and 1st significance level, respectively.

conceptualize and measure financial stability. Following this line of research, our paper constructed a synthetic index of the financial system's soundness for 76 low- and middle-income countries during 2010-2018 using the IMF's financial soundness indicators as constituent elements. The index accounts for the incidence of macroprudential and other policies that assess and monitor the strengths and vulnerabilities of the financial system as a whole.

Our financial soundness index differs from previous approaches in that it is not limited by its dependence on the conditions and prudential ratios of individual financial institutions alone, but it reflects the broader, combined financial conditions, including compliance with international financial sector standards and codes, and the outcome of stress tests. Moreover, our index is fully data-driven, tested and validated by means of an unsupervised statistical learning technique, which makes neither a priori assumptions on the relationship among the input variables nor a subjective decision on the chosen variables.

Subsequently, we used the index to predict the financing constraints of individual non-financial firms in middle- and low-income countries. We control for the effect of the specific characteristics of firms and the influence of economic and institutional country-level factors.

Table 2.11: Predicting ACCESS with ordinal probit model with instrumental variables and institutional governance controls - 2SLS.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.0519***	-0.0526***	-0.0471***	-0.0477***	-0.0485***	-0.0491***	-0.0505***	-0.0512***
	(0.0113)	(0.0114)	(0.0105)	(0.0106)	(0.0102)	(0.0103)	(0.0117)	(0.0118)
LISTED	-0.0362***	-0.0358***	-0.0297***	-0.0293***	-0.0359***	-0.0355***	-0.0293***	-0.0289***
	(0.00860)	(0.00866)	(0.00882)	(0.00886)	(0.00868)	(0.00873)	(0.00877)	(0.00882)
AGE	-0.260***	-0.260***	-0.178**	-0.177**	-0.235***	-0.235***	-0.197***	-0.197***
	(0.0720)	(0.0720)	(0.0716)	(0.0716)	(0.0715)	(0.0715)	(0.0722)	(0.0722)
SUBSID	-0.0185**	-0.0185**	-0.0155*	-0.0155*	-0.0208**	-0.0208**	-0.0123	-0.0123
	(0.00879)	(0.00879)	(0.00841)	(0.00841)	(0.00896)	(0.00896)	(0.00822)	(0.00822)
LOCATION	0.0521**	0.0534**	0.0506**	0.0517**	0.0534**	0.0546**	0.0487**	0.0501**
	(0.0225)	(0.0225)	(0.0230)	(0.0231)	(0.0227)	(0.0228)	(0.0228)	(0.0228)
EXPORT	-0.0414***	-0.0414***	-0.0198	-0.0198	-0.0317**	-0.0317**	-0.0290**	-0.0291**
	(0.0146)	(0.0146)	(0.0145)	(0.0145)	(0.0152)	(0.0152)	(0.0138)	(0.0137)
OWNFOR	-0.0749***	-0.0750***	-0.0665***	-0.0667***	-0.0740***	-0.0741***	-0.0666***	-0.0668***
	(0.00852)	(0.00857)	(0.00821)	(0.00825)	(0.00852)	(0.00856)	(0.00832)	(0.00836)
OWNGOV	-0.0994***	-0.0995***	-0.0913**	-0.0914***	-0.0976***	-0.0977***	-0.0925**	-0.0927**
	(0.0336)	(0.0336)	(0.0355)	(0.0354)	(0.0328)	(0.0327)	(0.0366)	(0.0366)
GENDEQ		-1.093***		-0.947***		-0.974***		-1.073***
		(0.267)		(0.279)		(0.269)		(0.280)
BILHUM		-1.171		-1.447		-1.262		-1.368
		(0.911)		(0.948)		(0.932)		(0.930)
FISPOL		1.616***		1.447***		1.463***		1.611***
		(0.246)		(0.248)		(0.239)		(0.256)
First stage results								
NUMBRW	8.156**	8.148**	8.159**	8.152**	8.155**	8.148**	8.160**	8.153**
	(3.965)	(3.963)	(3.964)	(3.961)	(3.965)	(3.962)	(3.964)	(3.962)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.013	0.013	0.012	0.013	0.011	0.011	0.015	0.015
F-stat	19.72	19287.6	14.03	15279.6	16.57	17131.3	17.09	18178.0
Kleibergen-Paap	5.93**	5.89*	5.94**	5.91**	5.86*	5.89*	5.88*	5.95**
Portmanteau	1	1	1	1	1	1	1	1
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and 2SLS of Eq. (2.3) for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Null hypothesis of Kleibergen-Paap test is the under-identification of the model. Null hypothesis of Portmanteau test is the absence of auto-correlation. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

We carry out sensitivity analysis to contain measurement error and we use endogeneity analysis to correct for omitted variable bias. We further apply the Oster test to obtain more robust results.

Our results show that our financial soundness index is a negative and significant predictor of the firm's financing constraints across countries. The results remain broadly stable after splitting the sample by income level, and controlling for firm size and sector of activity. It appears that financial stability considerations are relatively more important in affecting the financing constraints of small-size firms in less developed countries. Our results hold after carrying out endogeneity analysis using IV methods and remain robust to additions of different groups of country-level control variables.

While the analysis needs to be further extended and tested in different data samples and settings, it emerges that financial stability considerations and the associated macroprudential policies are important interventions for improving firms' access to finance, especially of smaller firms in less developed countries.

Table 2.12: Predicting ACCESS with ordinal probit model with instrumental variables and political controls - 2SLS.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.0519***	-0.0336***	-0.0471***	-0.0296***	-0.0485***	-0.0317***	-0.0505***	-0.0312***
	(0.0113)	(0.00612)	(0.0105)	(0.00582)	(0.0102)	(0.00594)	(0.0117)	(0.00598)
LISTED	-0.0362***	-0.0363***	-0.0297***	-0.0302***	-0.0359***	-0.0359***	-0.0293***	-0.0298***
	(0.00860)	(0.00878)	(0.00882)	(0.00901)	(0.00868)	(0.00887)	(0.00877)	(0.00895)
AGE	-0.260***	-0.276***	-0.178**	-0.194***	-0.235***	-0.249***	-0.197***	-0.215***
	(0.0720)	(0.0751)	(0.0716)	(0.0746)	(0.0715)	(0.0746)	(0.0722)	(0.0753)
SUBSID	-0.0185**	-0.0212**	-0.0155*	-0.0183**	-0.0208**	-0.0237**	-0.0123	-0.0147*
	(0.00879)	(0.00910)	(0.00841)	(0.00869)	(0.00896)	(0.00928)	(0.00822)	(0.00847)
LOCATION	0.0521**	0.0496**	0.0506**	0.0482**	0.0534**	0.0510**	0.0487**	0.0462*
	(0.0225)	(0.0235)	(0.0230)	(0.0241)	(0.0227)	(0.0237)	(0.0228)	(0.0238)
EXPORT	-0.0414***	-0.0451***	-0.0198	-0.0236	-0.0317**	-0.0350**	-0.0290**	-0.0331**
	(0.0146)	(0.0148)	(0.0145)	(0.0147)	(0.0152)	(0.0155)	(0.0138)	(0.0138)
OWNFOR	-0.0749***	-0.0752***	-0.0665***	-0.0670***	-0.0740***	-0.0741***	-0.0666***	-0.0671***
	(0.00852)	(0.00870)	(0.00821)	(0.00839)	(0.00852)	(0.00870)	(0.00832)	(0.00852)
OWNGOV	-0.0994***	-0.0937**	-0.0913**	-0.0868**	-0.0976***	-0.0921***	-0.0925**	-0.0878**
	(0.0336)	(0.0366)	(0.0355)	(0.0381)	(0.0328)	(0.0356)	(0.0366)	(0.0394)
STABDEM		-0.171***		-0.183***		-0.166***		-0.192***
		(0.0501)		(0.0498)		(0.0497)		(0.0498)
LIMLEND		-0.116*		-0.0921		-0.0958		-0.114
		(0.0627)		(0.0649)		(0.0586)		(0.0706)
VETOPWR		0.0293***		0.0250***		0.0268***		0.0274***
		(0.00727)		(0.00663)		(0.00669)		(0.00727)
First stage results								
NUMBRW	8.156**	7.004**	8.159**	7.015**	8.155**	7.007**	8.160**	7.013**
	(3.965)	(2.928)	(3.964)	(2.928)	(3.965)	(2.929)	(3.964)	(2.927)
Observations	39,383	37,739	39,383	37,739	39,383	37,739	39,383	37,739
Pseudo R^2	0.013	0.014	0.012	0.013	0.011	0.011	0.015	0.016
F-stat	19.72	30.88	14.03	29.71	16.57	34.00	17.09	28.06
Kleibergen-Paap	5.92**	6.42**	5.86*	6.55**	5.86*	6.37**	5.96**	6.44**
Portmanteau	1	1	1	1	1	1	1	1
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and 2SLS of Eq. (2.3) for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Null hypothesis of Kleibergen-Paap test is the under-identification of the model. Null hypothesis of Portmanteau test is the absence of auto-correlation. The *, ** and *** symbols denote the p-values at 10th, 5th and 1st significance level, respectively.

Table 2.13: Assessing the relevance of macro-control variables.

Group	Variable		$\pi = 1$ $\lvert\delta\rvert$	$\pi = 1.01$ $\lvert\delta\rvert$	$\pi = 1.02$ $\lvert\delta\rvert$	$\pi = 1.03$ $\lvert\delta\rvert$	$\pi = 1.04$ $\lvert\delta\rvert$	$\pi = 1.05$ $\lvert\delta\rvert$
		R^2 baseline	$R^2_{max} = 0.121$	$R^2_{max} = 0.122$	$R^2_{max} = 0.123$	$R^2_{max} = 0.124$	$R^2_{max} = 0.126$	$R^2_{max} = 0.127$
Macro-economic	GDPCAP	0.121	5.842	0.229	0.116	0.078	0.059	0.047
	INFLDFL	0.121	4.64	0.231	0.118	0.079	0.06	0.048
	LENDINT	0.121	15.826	1.608	0.843	0.571	0.432	0.347
		R^2 baseline	$R^2_{max} = 0.121$	$R^2_{max} = 0.122$	$R^2_{max} = 0.123$	$R^2_{max} = 0.124$	$R^2_{max} = 0.125$	$R^2_{max} = 0.127$
Financial Access	FININD	0.121	2.485	0.289	0.153	0.104	0.079	0.063
	FINDEP	0.121	3.576	0.303	0.158	0.106	0.08	0.065
	OUTLOAN	0.121	26.359	0.279	0.139	0.093	0.07	0.056
		R^2 baseline	$R^2_{max} = 0.121$	$R^2_{max} = 0.122$	$R^2_{max} = 0.123$	$R^2_{max} = 0.125$	$R^2_{max} = 0.126$	$R^2_{max} = 0.127$
Institutional governance	GENDEQ	0.121	12.275	0.474	0.241	0.161	0.121	0.097
	BILHUM	0.121	6.013	0.901	0.485	0.332	0.252	0.203
	FISPOL	0.121	1.008	0.16	0.086	0.059	0.045	0.036
		R^2 baseline	$R^2_{max} = 0.112$	$R^2_{max} = 0.113$	$R^2_{max} = 0.114$	$R^2_{max} = 0.115$	$R^2_{max} = 0.116$	$R^2_{max} = 0.118$
Political	STABDEM	0.112	21.289	0.974	0.496	0.333	0.25	0.201
	LIMLEND	0.112	91.159	1.963	0.987	0.659	0.495	0.396
	VETOPWR	0.112	32.977	1.018	0.515	0.344	0.259	0.207

Notes: The table reports results of Oster test in order to state the relevance of each variable compared to unobserved variables and assess the impact on the change of the coefficients' value. For example, $\delta = 2$ means that the unobservable variables would need to be twice as important as the observable ones to shrink the coefficient to zero. The higher the value of δ, the more relevant is that variable. The R^2 of the model with both observable and unobservable variables is required for the calculation of δ. Multiplying the R^2 value by a number $\pi > 1$ relaxes the test's assumptions, allowing the model to take into account the errors in estimation due to poor specification power of the observed variables.

Chapter 3

A data-driven approach to measuring epidemiological susceptibility risk around the world

3.1 Introduction

During the past year, the Covid-19 pandemic has infected more than 100 million people and caused more than 2 million deaths in more than 200 countries around the world. The associated real and social costs are huge. Some estimates raise the global real cost of the Covid-19 pandemic for the next few years to several USD trillion (The International Monetary Fund, 2020). A great concern has been the virus' spread to countries with weaker epidemics management systems. Thus, knowing how countries with different degrees of preparedness have responded to the pandemic is important for assessing cross-country epidemiological risk and optimally deploying resources in support of this global health emergency. This is critical knowledge of globally susceptible populations, with several countries reporting infection levels exceeding their average historical levels. These policy concerns have remained valid during all phases of the Covid-19 pandemic and especially during the process of gradual adjustment of the lockdown restrictions. The question of country preparedness has surfaced again following the pandemic's evolution (The World Health Organization, 2020d).

The question of countries' preparedness to manage epidemiological risk must be addressed from a long-term perspective. It is likely that the world will continue to face epidemic risks, which many countries are still ill positioned to manage. In addition to climate change and urbanization, global population displacement and migration—now happening in nearly every corner of the world—create favorable conditions for the emergence and spread of new pathogens. Countries also face an increasing potential threat of accidental or deliberate release of deadly engineered pathogens, which could cause even greater harm than a naturally occurring pandemic. Scientific advances that help in fighting epidemic diseases have also allowed pathogens to be engineered or recreated in laboratories. Meanwhile, cross-country disparities in capacity and inattention to biological threats have exacerbated preparedness gaps. Measuring country preparedness emerges as a key real policy challenge for both countries and organizations.

We contribute to addressing this policy challenge by creating an index of epidemiological susceptibility risk (ESR) for 168 countries. Various real and non-real factors affect the extent to which a country is susceptible to epidemiological risk. We produce a new epidemiological preparedness measure that relies on objective information that facilitates policy choices. We build on previous studies and our index information accounts for the role of environmental, health, transport and communications infrastructures; real activity; demographics; and governance institutions. We deal with the complexity of these factors by implementing a fully data-driven approach to measuring their influence on epidemiological risk. In contrast to previous studies (Rivers et al., 2019, Polonsky et al., 2019, Mertzanis and Papastathopoulos, 2021), our fully data-driven approach produces results that provide a better evidence basis to support reasoning and decision. While there are no data-driven algorithms that can lead to fully optimal assessments of risk, our approach has considerable advantages, such as avoiding the subjective weight determination and the need for post-hoc rationalization. Evidence shows that data-driven models offer better predictive accuracy in epidemiological research than knowledge-based ones (Rajabi et al., 2014). Given the complexity of the problem, we choose different versions of principal component analysis (PCA) as well as dynamic factor models (DFM) to deal with the presence of strong cross-section dependence in the data due to unobserved common factors. We conduct extensive in-sample model evaluations of 168 countries covering 17 indicators during the period 2010-2019. Our results show that the robust PCA method explains more than 90% of total variability, whilst the DFM explains about 76% of the total variability.

Our paper contributes to the literature in the following ways: it builds on previous studies by proposing a substantially improved index of epidemiological susceptibility risk that is fully model-based and data-driven, tested and validated according to advanced statistical techniques (see section Results). We use alternative analytical estimation models based on unsupervised statistical learning methods, which make neither a priori assumptions on the relationship among the input variables nor a subjective decision on the variables to be possibly dropped. Further, our data-driven approach does not need to define a target variable, thereby avoiding a further risk of subjectivity. The only model assumption lays on the number of components built on the original variable space reflecting the desired level of captured variability and predictive ability. Moreover, the new coordinates must, by construction, lie on a linear space and be orthogonal (i.e., uncorrelated). No correlation ensures that each new principal component or dynamic factor describes a specific and unknown in advance latent phenomenon through the linear combination of the initial variables. We produce the index values with different methods, which allow policy makers to assess country preparedness according to specific needs and objectives.

Moreover, our paper contributes to the multifaceted literature on the conceptualization and measurement of epidemiological risk taking a long-term perspective (Gupta et al., 2018). Indeed, most studies focus on epidemics forecasting and they do not explicitly consider the preparedness question. The key novelty of our ESR measure is the consideration of long-term, policy-relevant conditions, and not merely of the temporary incidence of diseases, affecting the contagion of epidemics. Our ESR index is not meant to predict the short-term transmission of epidemic outbreaks but rather assess the long-term risk of epidemic contagion, largely reflecting the effect of policy. Finally, our analysis 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 learning algorithm (often ensemble models do the job), the dataset, the selection of leading variables and the preprocessing phase in general play a key role in producing accurate assessments. We have placed special emphasis on these aspects in our analysis.

Most efforts to contain the spread and effects of epidemics use the results of prediction models (Rivers et al., 2019, Polonsky et al., 2019). The prediction of the Covid-19 behavior has deployed sophisticated methods that include big data, social media information, stochastic models and data science/machine learning techniques along with medical (symptomatic and asymptomatic) parameters (Shinde et al., 2020, Nikolopoulos et al., 2021). However, prediction accuracy is limited due to the short period of data availability, data suitability, lockdown policies, difficulties in tracking the movement of people, changes in the incubation period and mutation of the virus, but also inappropriate algorithms and models.

The prediction of an epidemic establishes an alarm, which calls for a decision on what policy measures to undertake. The decision must be based on appropriate optimization of the prediction parameters, the likelihood of epidemic spread and its potential impact. Thus, it can be very complex and difficult, especially for locations with large and dense populations or critical infrastructure. Epidemics managers must factor prediction uncertainty into their decision-making models. However, while prediction methods have improved considerably and can handle increasing levels of complexity (Reich et al., 2019, Spreco et al., 2018, Debellut et al., 2018), prediction is essentially a short-term research enterprise. Instead, the overall preparedness of a country is a crucial long-term factor that guides the making of optimal decisions in response to an epidemic prediction.

The emergence of various epidemic outbreaks in the recent years led to the formulation of various country preparedness approaches that use different information and data aggregation methods. We briefly survey the most important ones. The Global Health Security Index (GHSI) represents a comprehensive assessment and benchmarking of health security and related capabilities of the countries that participate in the WHO's International Health Regulations. The GHSI is a joint project of the Nuclear Threat Initiative, the Johns Hopkins Center for Health Security, and The Economist Intelligence Unit (Johns Hopkins University Centre for Health Security, 2019). The GHSI provides a measure of a country's preparedness based on the capacity gaps of countries in their potential response to epidemics (T.Craig et al., 2020). However, the GHSI has been first published in 2019 and therefore it does not provide historical data to be used in thorough real research. Further, the GHSI is too broad and includes global catastrophic and biological hazards, which on the one hand endows it with a broad coverage capacity but, on the other hand, make it less flexible and less suitable for a tool of prediction of epidemic-driven real outcomes. Najmul (2020) find insignificant correlation between the GHSI and the incidence of Covid-19. After multiple testing, they suggest the inclusion of information on demographics and the reappraisal of its aggregation methodology. Razavi et al. (2020) argue that, while very comprehensive, the GHSI scoring may not be suitable for determining priorities and comparing countries with one another, calling for a further refinement of the index process that rationalizes the index's extensive focus on developed countries and health-related variables and its weighting methodology.

A related effort to assess country preparedness is the Joint External Evaluation (JEE) assessment tool. The latter is an externally validated, voluntary and collaborative assessment of 19 technical blocs of information necessary to validate the countries' capacity to detect and respond to public health risks (The World Health Organization, 2017). Unlike the GHSI, which allows inter-country comparisons, the JEE is a formal component of the WHO's Monitoring and Evaluation Framework, which all UN member states must implement. The JEE is not designed for making inter-country comparisons, but instead it is a technical tool for providing support to WHO member countries in setting quantified baseline thresholds for assessing progress. Shahpar (2019) use the average of the JEE's 19 technical areas for benchmark/comparison and argue that the JEE represents an initial effort at policy coordination that requires more global collaboration and prioritization of intervention. Garfield et al. (2019) tested the effectiveness of the JEE tool in a few African countries and found a high level of correspondence between score and policy text at the country level but also considerable differences in actual country responses relative to the benchmark JEE scores. They propose a better alignment of the JEE measures with the timing and depth of the country responses, which also reflect the contribution of international assistance in these areas.

Moreover, the Joint Research Centre (JRC), the European Commission's science and knowledge service, has cooperated with the World Health Organization to produce the Index for Risk Management (INFORM) (Doherty et al., 2018). The latter is a composite indicator that identifies countries at risk of humanitarian crisis and disaster that would overwhelm national response capacity and would be more likely to require international assistance. The INFORM model is based on risk concepts published in scientific literature and envisages three dimensions of risk: hazards and exposure, vulnerability, and lack of coping capacity.

Risk components factored into the analysis include natural disasters, socioeconomic factors, such as inequality and aid dependency, and institutional capacity, such as built environment and access to health care. However, the INFORM framework does not adequately capture the effect of biological hazards (i.e., epidemic outbreaks). The INFORM Annual meeting 2017 in Rome agreed to proceed by incorporating ancillary information from the WHO epidemiological risk initiative relating to health components to improve the overall INFORM index (The INFORM Annual Meeting Report, 2017). The index measures a wide variety of hazard risks and less so epidemiological ones and its multi-level and complex construction also makes it less flexible and less suitable for use as a policy tool.

Another comprehensive effort to develop a preparedness index was expended by the U.S. Center for Disease Control and Prevention (CDCP). Following the emergence of various national hazards, the CDCP produced the National Health Security Preparedness Index at the U.S. state level to measure the preparedness (NHSPI, 2015). The NHSPI uses information from six broad domains of national health security (NHSPI, 2015, CDCP, 2014). The domains are the management of incident and information, the delivery of health-care services, the improvement of occupational and environmental health conditions, the management of countermeasures, community engagement and planning conditions, and the surveillance of health security conditions. After reviewing these occupational and environmental health domains, we observe no inclusion of indicators of occupational health and safety but only measures of environmental health. Overall, while the NHSPI is comprehensive, it covers only one country (the U.S.) for only a few years. Moreover, we do not find evidence of using the NHSPI to predict real outcomes in the US economy.

Furthermore, E.Marcozzi et al. (2020) present a Hospital Medical Surge Preparedness Index (HMSPI) that can be used to systematically evaluate health care facilities across the U.S. states regarding their capacity to handle patient surges during disasters. The index aims to ensure that the US health care delivery system is poised to respond to mass casualty events by assessing the ability of victims to access health care (Kaji et al., 2008) as well as resolving weaknesses and reinforcing strengths in hospital and emergency management planning and capacity (Simiyu et al., 2014). The HMSPI uses four domains of surge capacity: staff, supplies, space, and integrated systems, and their subcomponents. However, the HMSPI is a static measure and of interest mainly to the US researchers.

Finally, Mertzanis and Papastathopoulos (2021) propose a composite index of epidemiological susceptibility risk, which they use to predict tourist flows around the world. They use information on time-varying, policy-relevant factors, such as infrastructure; demographics, real activity and institutions, which they standardize and combine based on a standard PCA method to produce a continuous value index, using equal weights. While their index proves a significant predictor of tourist flows, their methodological approach is a rather simple one depriving their index from its full predictive potential. The authors acknowledge the need for using more sophisticated dimensionality reduction methods to achieve better results. Table 3.1 provides a summary of key previous efforts to develop alternative composite measures of country preparedness to epidemiological risk. We acknowledge that other studies exist, mainly in epidemiological research field, that have measured aspects of epidemiological risk. However, we refer more directly to those that have had important policy implications.

A common characteristic of the above preparedness measures is that they are composite indicators (CIs). Some indices measure preparedness using mostly health-related information, whilst others extend their coverage to include information on relevant disasters and crises, others focus on the role of environmental factors, and yet others take into consideration real and institutional factors. Thus, while structurally different, these indices capture complementary aspects of epidemiological risk manifestation. As a result, some of them may be more suitable for measuring long-term country likelihood to suffer from the outbreak of epidemics, others could better measure long-term country preparedness to respond

Table 3.1: Comparison of alternative measures of country preparedness to epidemiological risk.

Index name	Coverage	Source
Global Health Security Index (GHSI)	Composite index, covering 195 WHO member countries, available since 2019. It measures country preparedness to respond to epidemics based on capacity gaps	The Johns Hopkins Center for Health Security & the Economist Intelligence Unit
Joint External Evaluation (JEE) Assessment Tool	Composite index, covering 195 WHO member countries, available since 2005. It measures policy gaps relative to benchmark in responding to public health risks	WHO: IHR Monitoring and Evaluation Framework
National Health Security Preparedness Index	Composite index, covering the USA only, available since 2015. It measures management efficiency in responding to public health risks	The Center for Disease Control and Prevention (CDCP)
Index for Risk Management (INFORM)	Composite index, covering 191 countries, available since 2019 (version covering epidemic risk). It measures the extent to which countries are at risk of humanitarian crisis and disaster that would overwhelm national response capacity.	Joint Research Centre (JRC), European Commission.
Hospital Medical Surge Preparedness Index	Composite index, covering the USA only, available since 2015. It measures the ability of health care facilities to handle patient surges during disasters	E.Marcozzi et al. (2020)
Epidemiological susceptibility risk index	Composite index, covering 188 countries during 2000-2019. It measures the extent to which countries are susceptible to epidemiological risk broadly accounting for health, economic and institutional factors.	Mertzanis and Papastathopoulos (2021)

effectively to epidemic outbreaks, whilst others may be more suitable to assess the long-term effects of epidemic outbreaks on the economy. Alternative composite measures can only capture different structural and time-relevant aspects of a phenomenon. They should therefore be properly integrated in a broader framework that considers their general and environmental repercussions (Morand, 2020). Moreover, the construction involves stages where subjective judgments need to be made on the selection of indicators, the treatment of missing values, the choice of aggregation process and the weights of the indicators, etc. The unavoidable subjectivity involved in their construction may undermine their credibility and therefore it is important to identify the sources of subjectivity. However, the absence of an objective way to determine weights and the aggregation methods should not compromise their validity provided that the overall construction process is transparent (Nardo et al., 2005). This paper proposes a data-driven approach, which overcomes potential subjectivity bias in weight selection, takes into consideration dynamic effects and provides a better understanding of the complexity in approximating epidemic effects. After all, evidence-based evaluation of national epidemic management programs is critical to their future success (Koplan et al., 1999).

The conception of our ESR index originated in our observation that the spread of COVID-19 differed among countries. We observed that some countries fared better than others in containing the spread, regardless of their level of development, which was mainly the result of policy choices. The index we propose, measures country susceptibility to epidemiological risk for the 2010-2019 period based on complete annual country level data. It is worth noticing that, it may not be suitable to measure the incidence of Covid19 outbreak on a daily basis, not least because the pandemic has emerged in the last year, for which data is only partly available. Our index may be better suited to capture the impact of long-term time-varying structural factors on the contagion of epidemic outbreaks and their effect on the economy. Our index construction reflects our effort to include relevant policy variables. To this end, it reflects the importance of infrastructure, demographics, real activity and governance (Najmul, 2020, Razavi et al., 2020, Mertzanis and Papastathopoulos, 2021).

The literature on epidemiological risk provides justification for these factors. First, quality health care infrastructure facilitates the timely detection and monitoring of infectious people in time and space, and therefore the successful containment of the epidemic (Morse, 2007). Global coordination increases monitoring efficiency. Moreover, quality health care infrastructure helps improve productivity and employment and hence production resilience, general stability and social inclusion (Boyce and Brown, 2019). Adequate financing of health care infrastructure contributes decisively to its effectiveness (Kruk and Freedman, 2008).

Second, an effective communications infrastructure improves market surveillance, raises public awareness of epidemics risks and facilitates the swift private and public responses by assembling and broadcasting suitable information (Rainwater-Lovett et al., 2016). A new survey finds that about 53 percent of adults in the U.S. say that the internet has been essential for dealing with the pandemic, whilst 34 percent describe it as "important, but not essential" (Pew Research Center, 2020).

Third, an effective transportation infrastructure facilitates the monitoring and control of infectious population but also the response and timely provision of necessary care (Meyer and Elrahman, 2019). This is especially important with respect to passenger aviation that unavoidably contributes to the spread of an epidemic. Hufnagel et al. (2004) found a significant association between heterogeneity in airline connectivity networks and epidemic predictability.

Fourth, an effective infrastructure securing clean water and sanitation services is necessary for containing the speed and spread of epidemics and induces the health care sector's response to adhere to high sanitary standards (D.Phelps et al., 2017). During epidemic outbreaks, the transmission of diseases occurs through both access to local water distribution facilities and the availability of man-made or natural water resources and sanitation systems. The OECD (2020) argues that enhancing environmental health through better air quality, water and sanitation, waste management, along with efforts to safeguard biodiversity, will reduce the vulnerability of communities to the effects of epidemics. KWR (2020) found that screening for Covid-19 at municipal wastewater plants in the Netherlands contributed to a better monitoring of its spread.

Fifth, demographics is also important. The increasing life expectancy and decreasing fertility rates change the patterns of consumption thereby affecting the dynamic of epidemics. For instance, Geard et al. (2015) argue that declining fertility rates are associated with an older mean age of disease infection that affects the spread of epidemics, depending on vaccination and other policy measures. Further, the rising urbanization rate globally affects epidemics in two ways (Neiderud, 2014): it causes improvements in health infrastructure in urban areas, but also provides a fertile ground for the emergence of new pathogens due to tighter human encounter. Population density is generally associated with a faster and wider spread of epidemics (Tarwater and Martin, 2001, Li et al., 2018).

Sixth, real activity also affects the spread of epidemics. Relman et al. (2020) report the views of different experts on how travel, trade and conflict move people, animals and plants globally affecting the transmission of diseases. Adda (2015) finds that booms increase people's mobility among different transmission venues (ports, airports, etc.) and interpersonal interaction thereby contributing to a wider and faster spread of epidemics. Suhrcke et al. (2011) argue that real downturns cause higher urbanization and congestion of people seeking jobs, worsening living and health care access conditions of living, which in turn lead to adverse epidemic effects. Kafertein (1997) argued that the rapid concentration of global food trade in a few multinational corporations increased the transmission of foodborne diseases. Lang (2001) stressed the effects of mass production and logistics procedures on the spread of infectious diseases.

Finally, institutional governance matters. Quah (2007) and Pritchett et al. (2013) document from different perspectives how institutional governance, exerted through various social interactions, social coordination and risk management policies, affect the spread of epidemics. However, the capacity of governance institutions develops differently among countries, subject to political influence, uncertainty or conflict (Gayer et al., 2007). The OECD (2010) argues that higher human capital improves governance and health outcomes through stronger social capital networks, employment prospects and psychological responses.

3.2 Results

After the imputation of missing values (see section C.2 in the Appendix), we standardize the dataset for each year and then we apply first the PCA method in all different versions, as described in section Results. Table 3.2 reports the results of the different PCA versions. We report the average variance explained by loadings across years, as well as the average R^2 on both the whole dataset and subsets with values trimmed for the $95th$ and $99th$ percentiles in order 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 Augmented Dickey–Fuller 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 the employed data (full data set, 1% trimmed and %5 trimmed). Accordingly, we retain only the first principal component, which explains at its minimum a remarkable 87% of the total variance and therefore renders the resulting ESR index visually interpretable. Figures C3-C5 in the Appendix report the scree plots of the variance explained by the loadings among all PCA methods and Figure C6 shows the relative importance of the loadings. This includes the percent of variance explained by the first principal component of each PCA method per year.

Table 3.2: Results from Robust PCA. Mean is evaluated over years. Mean Explained Variance is evaluated from the eigenvalues of PCA, R^2 is reported for the full dataset and for the $99th$ and $95th$ percentiles. Augmented Dickey–Fuller test for stationarity of the ESR index as well.

Method	Number of PC	Mean Explained Variance	Mean R^2	Mean R^2 on 99th	Mean R^2 on 95th	Augmented Dickey–Fuller
PCA	1	$49.9 \pm 0.9\%$	$49.9 \pm 0.9\%$	$57.3 \pm 1.1\%$	$65.3 \pm 0.9\%$	$\ll 0.01$
RobPCA	1	$87 \pm 0.9\%$	$94.8 \pm 0.3\%$	$95.4 \pm 0.2\%$	$96.5 \pm 0.2\%$	$\ll 0.01$
RobSparPCA	1	$50.2 \pm 0.9\%$	$28.5 \pm 3\%$	$33.6 \pm 3.6\%$	$38.2 \pm 4.5\%$	$\ll 0.01$

Then we apply the DFM method, as described in section Results, which depends upon two hyper-parameters: the sparsity coefficient α of the VAR and the correlation structure of the residuals for Kalman filter. Thus, we simulate synthetic factors $\widetilde{\mathbf{F}}$ with different combinations of number of observed variables, countries, years, latent factors $\mathbf{F}$, and we generate the corresponding $\mathbf{y}_t$ given different combination of $\mathbf{A}$, defined by α, and $\mathbf{C}$, randomly generated, using equation (3.1). Then, for each of the previous combination and correlation structure of residuals $\mathbf{Q}$, we apply the described algorithm and assess the reconstruction error on the fitted factors $\widetilde{\mathbf{F}}$ with the simulated factors $\mathbf{F}$. The optimal parameters found are $\alpha = 0.2$ and a diagonal structure. Afterwards, we evaluate the R^2 on DFM model. Table 3.3 reports the DFM results. In this case, the poorer performance is due to the small size of the dataset compared to the number of parameters, despite mitigated with sparseness. Moreover, the estimated interactions factor in $\hat{\boldsymbol{A}}$ turns out to be very small (values range in $[-0.06, 0.05]$), so we assume to be valid the no interactions setting, which has produced the highest R^2 (73.6%).

We run the Augmented Dickey–Fuller test also on the DFM based index obtaining p-values $\ll 0.01$ for both model specifications and ensuring its stationarity as for the PCA case. Figure C7 in the Appendix shows the relative importance of the loadings for the DFM model with interpretation.

Table 3.3: Results for DFM. R^2 is reported for the full dataset and for the $99th$ and $95th$ percentiles. We also report Augmented Dickey–Fuller test for stationarity of the ESR index.

Method	Number of Factors	R^2	R^2 on 99th	R^2 on 95th	Augmented Dickey–Fuller
DFM with interactions	1	−204.5%	−43.8%	7.7%	$\ll 0.01$
DFM without interactions	1	−405.4%	38.6%	73.6%	$\ll 0.01$

As robustness check, we compare the two ESR index values generated by the competing methods in terms of predictive power within a supervised analysis setting. To this end, we use the following macro-economic variables: real GDP per capita, government consumption (percent of total), price level of capital formation, trade volume, unemployment rate and outstanding loans of commercial banks. We standardize the target variables before fitting the algorithms to make the results comparable. We use both linear and non-linear data-driven learning algorithms to capture potential non-linearity effects in the data. We use alternatively the learning techniques of Random Forest, Regularized OLS (Elastic-Net), Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, Multivariate Adaptive Regression Spline (MARS) and a single layer Neural Network (NN). All the hyper-parameters are tuned with Bayesian Optimization and a 5-fold cross-validation. When fitting Elastic-Net with a single regressor, we use the OLS regression. Final performances are evaluated using a further 5-folds cross-validation and the average test set Root Mean Square Error (RMSE) is considered. The seed used to select the cross-validation folds has been kept fixed for all algorithms in order to ensure reproducible results. We provide examples of the comparison results. Table 3.4 (available in the Appendix) shows the RMSE percent increase in predicting Unemployment rate with the single index as regressor compared to the RMSE obtained with all 17 original variables. RMSE of models which are fitted considering ESR index solely tends to increase as we would reasonably expect. However, RMSE increases are always within one standard deviation bound suggesting that a much simplified analysis based on 1 unique index is significant and largely satisfies the parsimony principle. Table 3.4 clearly shows that Random Forest has the lowest RMSE by employing the original 17 variables (0.079) and further the ESR index based on the DFM approach presents the minimum RMSE (0.447). Complete results for all the fitted regressions are reported in C.5 of the Appendix.

Table 3.4: RMSE in predicting Unemployment rate using continuous index as regressor. RMSE for regression with original variables is reported in parenthesis.

	RMSE index (RMSE original)	
Algorithm	DFM	Robust PCA
Elastic-Net	0.999(0.859)	0.995(0.859)
MARS	1(0.583)	0.924(0.583)
Random Forest	0.447(0.079)	0.7(0.079)
Single Layer NN	0.994(0.31)	0.932(0.31)
SVM-RBF	1.024(0.083)	0.936(0.083)

Further, we can provide useful visual insights by exploring the temporal evolution of the ESR index values for each country in a world map. The map in Figure 3.1 reports the global distribution of the ESR index for DFM methods (the PCA one is available in the Appendix).

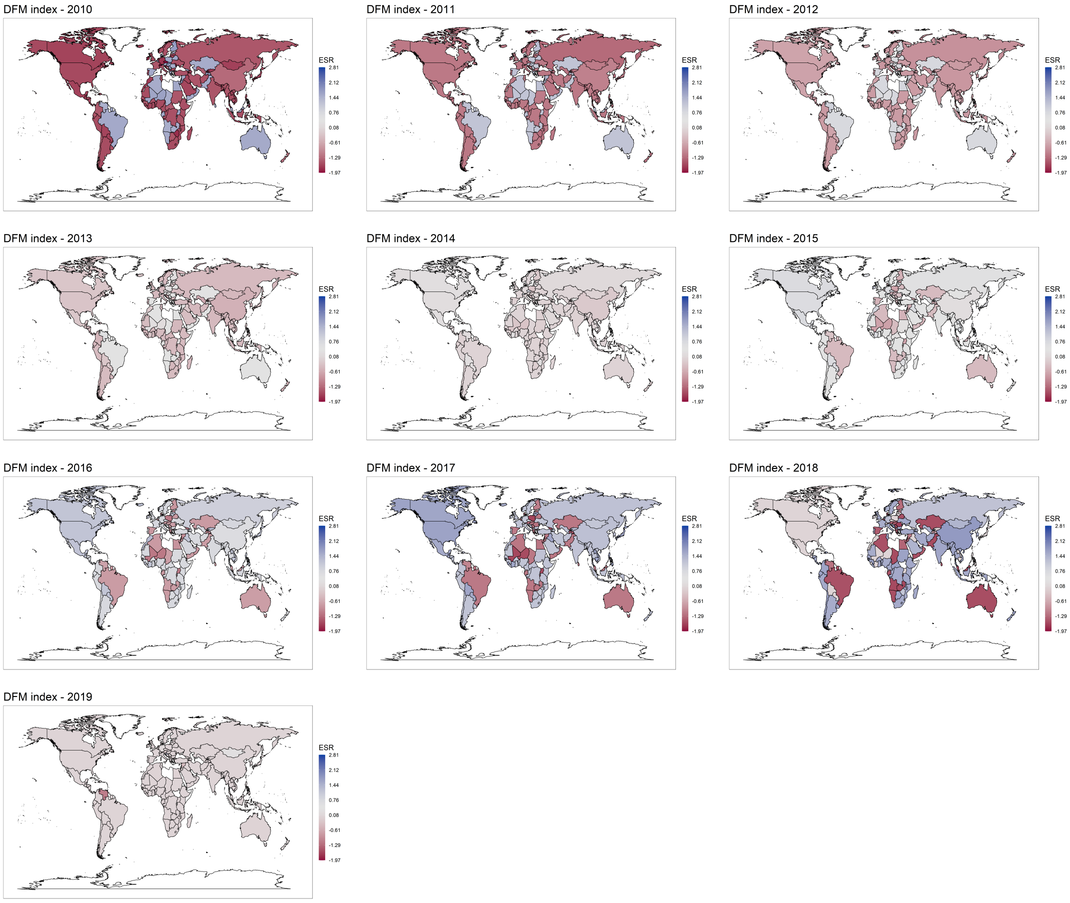


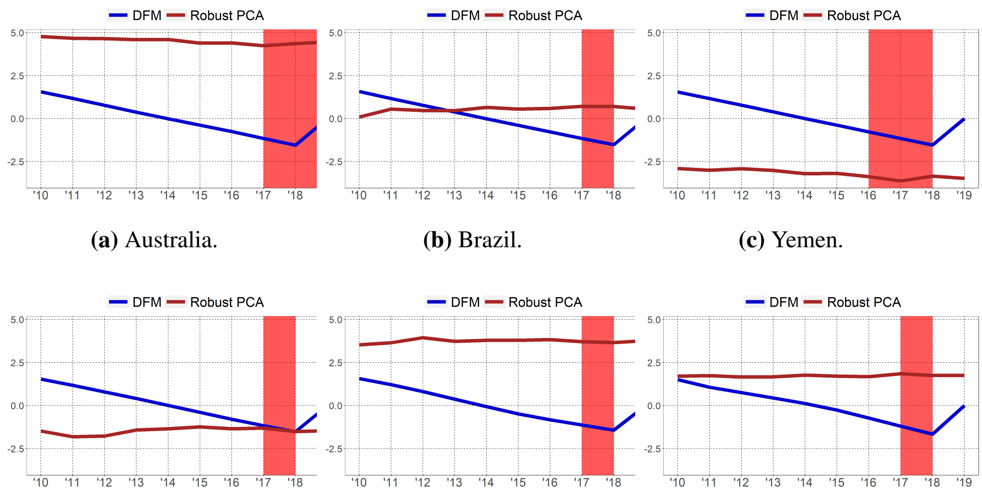
Figure 3.1: DFM index evolution over years.Shades of red color refer to riskier countries, while shades of blue to safer ones.

Indeed, the native characteristic of DFM of properly modeling the temporal dynamics is reflected in the world map which presents more variability in the colour change compared to PCA. Finally, Figure 3.2 shows the evolution over time of the ESR index for some individual countries, comparing the PCA and DFM methods. The PCA index is quite stable over time, whilst the DFM index captures the time dynamics of underlying latent factors. For example, Figure 3.2a shows that our index can capture the abnormal increase of Influenza cases in 2018-19 in Australia. In Figure 3.2b ESR index highlights the Zika virus outbreak of 2018 in Brazil. In Figure 3.2c the index underlines the Cholera spread between 2016 and 2018 in Yemen. Cholera outbreak in 2018 is captured for Algeria as well as shown in Figure 3.2d. Similarly Figure 3.2e and Figure 3.2f show how the index is able to capture the abnormal Influenza spread of 2018 and the increase of Measles case in 2018 in Spain and Romania respectively. Figure C14 to C17 in the Appendix provide the detailed evolution of the ESR index per country during the 2010-2019 period using both PCA and DFM methods. In order to support the previous insights, we checked the Spearman correlation between our proposed ESR and the historical incidence of a number of diseases extracted from World Health Organization: HIV, Malaria, Tubercolosis (TBC) and Tropical Neglected Diseases (NTD). Table 3.5 reports the countries whose ESR index has the highest correlation with the corresponding disease's evolution over the years. Only results for the DFM approach are reported. Results show the goodness of the proposed index. We can notice many high and significant correlations for all over the world countries (European, South American, African and Asian ones). The analysis suggest that the ESR index can play an important role in signaling pandemic outbreak periods thus helping regulators and countries in improving preparedness and recovery plans. Moreover, by looking at Figure 3.2, we can spot the temporal evolution of both the indexes and it emerges clearly how sensitive the ESR index is to epidemic outbreaks (particularly the DFM based one).

Table 3.5: Correlation between ESR index and the historical disease incidence for HIV, Malaria, Tubercolosis (TBC) and Tropical Neglected Diseases (NTD). Only results for the DFM appoach and for the top highly correlated countries are reported.

Country	HIV	Malaria	TBC	NTD
Angola	**1***	**0.98***	**1***	**0.5***
Argentina	**0.94***		**0.67***	0.3*
Brazil	0.21*	0.43*	0.37*	**0.92***
Dominican Republic	**0.99***	0.32*	**0.85***	0.38*
France	**0.88***		0.09*	0.45
Indonesia		**0.93***	**0.97***	**0.95***
Netherlands	**0.98***		**0.83***	0.28*
Nigeria	**1***	0.47		**0.57***
Pakistan	**0.95***	**0.97***	**0.97***	0.1*

$*$ p-val < 0.05



(a) Australia. **(b)** Brazil. **(c)** Yemen.

(d) Algeria. **(e)** Spain. **(f)** Romania.

Figure 3.2: Index evolution over years for some countries. Disease outbreaks are shaded in red.

3.3 Discussion

Epidemic outbreaks are extreme events that are perceived by the population to be more frequent and severe, mainly due to the increased globalisation and interconnections. The COVID-19 pandemic is an extreme risk event that has unfolded with tremendous speed and breadth. Epidemics cause huge real costs for institutions and countries. It is therefore important to evaluate the extent to which countries can identify and manage epidemiological risks adequately. Despite significant improvements in infrastructure and governance worldwide, many countries remain unprepared to adequately identify and manage epidemiological risks. In this study, we have proposed a country preparedness evaluation framework that countries and institutions could use to manage the contagion and consequences of epidemic risks. The framework is based on the development of a composite indicator, which we call epidemiological susceptibility risk index (ESR), for 168 countries during 2010-2019.

In constructing our ESR measure, we use objective and regularly reproduced information that accounts for the role of infrastructure, real activity, demographics and governance institutions. This integrated view of measuring epidemiological risk is in line with the general directions proposed by the WHO. We complement previous efforts at assessing country preparedness by proposing a methodological framework that makes the assessment of preparedness more policy-driven and expanded around the world. Importantly, our proposed framework uses a data-driven approach to constructing the index that utilizes both PCA and DFM methods and their variants for achieving dimensionality reduction. The results show that, after accounting for data characteristics and missing values, the robust PCA method shows very good performance whereby the first dimension explains about 90% of total variability. However, the nature of its construction prevents it from capturing properly the temporal latent dynamic of the data. We therefore use the alternative DFM method for this purpose. Albeit somewhat less efficient in comparative terms (the first component explains about 76% of the total variability), the DFM method must be considered as the benchmark model since it properly models the temporal dynamics, which are important in capturing epidemic outbreaks across a wide range of countries during the 10 available years. Our ESR index is fully data-driven and does not allow for arbitrary and subjective choice of weights that could impair its predictive efficiency.

This framework and index could provide the basis for developing risk assessments of epidemiological risk contagion after the outbreak of an epidemic but also for ongoing monitoring of its spread and social and real effects. It would also allow for useful comparisons in country preparedness and performance. This framework and index could be used by firms to assess likely real consequences of epidemics and could therefore have managerial implications. For example, in addition to help managing epidemiological risk, the framework could be useful in aligning country and corporate policy to environmental sustainability considerations and responsible behavior. Further, it takes into consideration ongoing regulatory initiatives that stress the importance of non-financial risks due to climate change.

Finally, our framework could be revised and extended towards various directions to support decision making. One way to improve it is to increase the data series availability mindful of the missing data problem using more advanced techniques. Another way to extend it includes the addition of new relevant dimensions that may capture other aspects of epidemiological risk. As research on the sources and spread of Covid-19 continues, new information is being revealed, which might inform the re-construction of our ESR index. Another way would be to apply alternative data dimensionality reduction techniques and compare the predictive results. The extensive check on the index's predictive power remains to be accomplished by applying it to diverse real-world situations.

3.4 Methods

3.4.1 Sources of data

The preceding literature provides the broad directions and information for constructing the epidemiological susceptibility risk index (ESR). The index broadly captures the effects of the above-described building blocks of epidemiological risk. Following previous studies, we select objective and periodically reproducible variables that, given the relevant literature, best capture the extent to which a country may be susceptible to epidemiological risk and for which there is adequate and ongoing country coverage. The index does not model restrictions per se, but the objective outcome of restrictions in terms of people and products. Our initial dataset includes the values of 17 time-varying variables for 206 countries during the 2010-2019 period, classified in seven groups to construct the ESR index: health infrastructure; environmental safety infrastructure; transport infrastructure; communications infrastructure; demographics; economic activity; and governance institutions. To capture health infrastructure effects, we use (1) the value of health expenditure per capita (current USD); (2) the index value of health care access and quality; (3) the response rate to public health hazards; (4) the number of physicians per 1,000 people; and (5) the number of hospital beds per 1,000 people. To capture transport infrastructure effects, we use (6) the (inverse of the) number of air passengers as a percent of total population. To capture demographic effects, we use (7) the number of urban populations as a percent of total population; (8) the number of people per Km2 of land (population density); and (9) the population of 65+ years of age as a percent of total population. To capture environmental safety infrastructure effects, we use (10) the number of people using safely managed drinking water services as a percent of total population; (11) the number of people using safely managed sanitation services as a percent of total population. To capture relevant real activity effects, we use (12) the value of trade in services as a percent of total trade and (13) the value of trade in goods as a percent of total trade. To capture communications infrastructure, we use (14) the number of individuals using the internet as a percent of total population. Finally, to capture institutional effectiveness, we use (15) the extent of human capital development; (16) the value of government effectiveness indicator and (17) the value of the rule of law indicator. The World Health Organization (WHO) database provides the data for variables (1) to (4); the World Development Indicators (WDI) database provides the data for variables (5) to (15); the Penn Tables (PT) database provides the data for variable (16) and the Worldwide Governance Indicators (WGI) database provides the data for variables (17) to (18). For sake of clarity, we stress that 3 out of the 17 considered variables are different in terms of measurement levels. Indeed, the human capital index (13), the value of government effectiveness indicator (16) and the value of the rule of law indicator (17) are indexes built upon other variables. However, this does not imply problems in the model specifications since they allow to take into account a wider range of information without adding more noise and keeping compact the model. A similar approach was followed by Cevik et al. (2013), Creane et al. (2006), Brave and Butters (2011) and Sales et al. (2012).

Tables C1 and C2 in the Appendix present the summary statistics of the index's constituent variables Var1 to Var17 and their pairwise correlations. 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 84.5%. Moreover, 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.

Higher values of these variables are associated with a lower risk of a country being susceptible to epidemiological contagion or, alternatively, they indicate better preparedness to manage these risks. While there are other relevant variables, the selected variables reflect factors and conditions that the literature has highlighted; they are objectively (not perceived) measured across countries, exhibit a low incidence of missing values and they are reproducible on a periodic basis. We did not include time-invariant factors (e.g., culture, religion, genetics) for we intend the index to capture mainly policy-relevant dynamic influences. For the same reason, we did not include time-varying factors relating to the environment conditions (e.g., temperature, rainfall) and slowly changing institutional factors (e.g., legal systems). We believe these factors should act as external controls mediating the predictive effectiveness of the ESR index on real behavior rather than being constituent elements of the index itself. We do acknowledge the limitation of choosing certain variables than others or many more, but we had to draw the line somewhere. We do believe there is room for future improvements in the index's conceptualization and construction. An advantage of this construction is that our ESR index is mainly a policy-based and not a perceptions-based measure, which allows us to explore its effects on real behavior largely devoid of perceptions, which would make it more severely prone to endogeneity.

3.4.2 Dimensionality reduction

The aim of our analysis is to extract a synthetic indicator that summarizes at best the relationship among variables in a lower dimensional space. We apply two alternative but complementary statistical methodologies to reduce dimensionality and construct the index: Principal Component Analysis (*PCA*) and Factor Analysis (*FA*). *PCA* aims at creating new variables from a larger set of observed covariates, where each one is a linear combination of the Y original variables. The model is represented by the equation $C = w_1 Y_1 + \ldots + w_i Y_p$, where C is the new principal component, Y_i are the original variables and w_i are the weights of the linear combination for $i = 1, \ldots, p$.

FA, on the other hand, models the measurement of latent variables, seen through the relationships they cause in a set of Y variables. The model is represented by a set of equations $Y_i = b_i F_i + u_i, i = 1, \ldots, p$, where Y_i are the original variables, F_i are the latent factors and b_i, u_i are the parameters of the combination.

Recalling that our dataset has three dimensions, *Country*, *Variable* and *Time*, we use PCA to model country/variable interaction for each year whereas FA to model country/time interaction, for all variables. Thus, using PCA, we create a low dimensional (1 way) indicator, explaining the maximum variance of the data and considering each year separately. Whereas, using FA, we estimate a single latent component able to capture the temporal interactions among the original variables. We describe the application of each dimensionality reduction method below in more detail.

We evaluate PCA on each year separately, producing T models. To ensure the stability and robustness of results, we apply and compare three different PCA techniques: regular PCA, Robust PCA and Robust Sparse PCA. PCA aims at finding new and wise linear combinations of the original data, in a way that the amount of explained variance of the data is maximised. Those combinations are mathematically constrained to be mutually orthogonal (that is uncorrelated) 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 $k << p$ such that the projected data matrix $W = XC^T$, also called scores matrix, will have dimension $n \times k$. The maximization problem is stated as follows:

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

where $\|\cdot\|_F$ is the Frobenius norm. We implement the model using R function `prcomp`. Since we do not rely on the classical PCA but, rather, we seek for 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 maximization problem is stated as follows:

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

where $\|L\|_*$ is the nuclear norm and λ is a penalization term. Following the procedure of Candes et al. (2009), once fitted, $\mathbf{L}$ can be used as a proxy for $\mathbf{X}$ with the extreme values excluded. Finally, following Erichson et al. (2018), we produce both a robust estimation and a sparse representation of the principal components by adding a sparsity constraint on the matrix C. The associated maximization problem is stated as follows:

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

ψ and ϕ are regularizing functions (i.e. LASSO or Elastic Net).

3.4.3 Dynamic Factor Model

Moreover, we evaluate a temporal dependent version of FA called Dynamic Factor Model (DFM), using all the available years within the same model. Given the $p \times n$ matrix $\mathbf{X}$, the model assumes that there exist some $k \times n$ factors $\mathbf{F}$ such that their mutual interaction over time can be expressed by a $k \times k$ interaction matrix $\mathbf{A}$ and the observed variable can be expressed as a linear function of the factors themselves through a $p \times k$ loading matrix $\mathbf{C}$. The problem can be solved as a system of equations:

$$\begin{cases} \mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathcal{N}(0,\mathbf{Q}) \\ \mathbf{X}_t = \mathbf{C}\mathbf{F}_t + \mathcal{N}(0,\mathbf{R}) \end{cases} \tag{3.1}$$

where $\mathcal{N}$ is the normal probability distribution and $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrix of the residuals of each equation in (3.1), respectively. Due to the short time series of the input variables, this model cannot be fitted considering all countries together as the resulting system of equations (3.1) is under-determined. Thus, we deal with the problem as follows: first, following Holmes et al. (2018), we fit DFM for each country, obtaining the factor matrices $\mathbf{F}^i$, the factor interactions $\mathbf{A}^i$ and the factor loadings $\mathbf{C}^i$, $i = 1,\ldots,n$. Second, we fit a Vector Auto Regressive (VAR) model in order to get $\hat{\boldsymbol{A}}$ 1-year lag matrix that incorporates cross-countries interactions of $\mathbf{A}^i$. We implement the model using *R* package `sparsevar` (Vazzoler et al., 2016) because this calibration problem has too many parameters to estimate relative to the number of observations, thus requiring a sparse approach. Finally, we use Kalman Filter to get smoothed factors $\widehat{\mathbf{F}^i}$ using $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{C}} = diag(\boldsymbol{C}^i)$, that is to get latent factors that incorporate cross-countries interactions. Briefly, Kalman filter re-estimates the factor matrix $\mathbf{F}$ iterating the two equations in (3.1) until the error between the predicted observed variables $\hat{\boldsymbol{X}}$ and the true one is minimized. We implement the model using *R* package `FKF` (Luethi et al., 2018). We assume $\hat{\boldsymbol{C}}$ to be diagonal in order not to double-count correlations within the observed variables and because cross-country interactions are already modelled through the VAR.

In both cases (PCA and DFM), the final index ESR will be represented by the scores matrix W and the factor matrix F respectively, both k-dimensional. One of the goal is to select the optimal number of components k as a trade-off between the maximal explained variance and the smallest value of components k.

3.4.4 Validation

Applying a dimensionality reduction technique by merely maximising the amount of explained variance with the smallest set of components, could be misleading and conduct to hardly interpretable results. Thus, once identified the most reliable results, we compare the fitting power of the produced indexes to a baseline benchmark. We accordingly estimate several parametric and non-parametric regression models to produce comparisons of the produced ESR index with the original set of variables. We use, as target variable, the following macro-economic variables: real GDP per capita, government consumption (percent of total), price level of capital formation, trade volume, unemployment rate, outstanding loans of commercial banks. Our validation process aims at demonstrating the relevance of the new index in representing the information conveyed by the original component variables. If the modeling ability of the composite ESR index, measured by the root mean square error (RMSE), is comparable to the original one based on the initial variables, we can conclude that the produced indicator is not only satisfactory according to the chosen dimension reduction technique but also effective in terms of predictive power within a simplified framework.

3.5 Data availability

The World Health Organization (WHO) data can be found at https://www.who.int/data/collections; the World Development Indicators (WDI) data can be found at https://databank.worldbank.org/source/world-development-indicators; the Penn Tables (PT) data can be found at https://www.rug.nl/ggdc/productivity/pwt/?lang=en and the Worldwide Governance Indicators (WGI) data can be found at https://databank.worldbank.org/source/worldwide-governance-indicator

Chapter 4

Machine Learning and Credit Risk: Empirical Evidence from SMEs

4.1 Introduction

Determining corporate credit ratings is a well-known topic theoretically and empirically, both in the financial academic literature and in the industry (Altman, 1980, Louzada et al., 2016b, Blöchlinger and Leippold, 2018). Within this topic, the corporate martket can be viewed as being composed of different segments. Among the latter, SMEs represent a large segment of the corporate market in several economies. As such, SME credit ratings have recently drawn the attention of academics and policy makers. This attention has led to a fervent debate on how to reach accurate estimation of SME credit risk. In this debate, the key features of SMEs are their informational opaqueness, greater perceived risk, and reliance on soft information-intensive relationship banking (OECD, 2020, Berger and Udell, 1995a, Claessens et al., 2005). In this regard, the importance to incorporate soft information in SME credit risk assessment has been acknowledged by regulators. As a matter of fact, regulators have introduced the internal ratings-based (IRB) approach, that allows banks to include qualitative soft information when assessing corporate credit risk (Bank for international settlements, 2006, Cucinelli et al., 2018). However, the successful implementation of the internal ratings-based (IRB) approach is challenged by the presence of severe communication frictions; the latter limit the successful "hardening" and transmission of soft information across large banking organizations (Stein, 2002, Liberti and Petersen, 2018, Filomeni et al., 2020a, Filomeni et al., 2020b), challenging the traditional role of relationship banking. These communication frictions are even exacerbated when banks engage in M&A activity, that leads to the creation of large banking conglomerates mostly relying on transactional (rather than relationship) banking (Ferri and Pesic, 2017, Berger et al., 2005). This has spurred us to investigate alternative methodologies to evaluate SME credit risk, mostly based on hard information.

Except for a few studies implementing alternative methodologies (Fantazzini and Figini, 2009, Moscatelli et al., 2019a), the literature has been mainly focused on the types of information a financial intermediary should use in assessing SME credit risk. This occurs at the expense of testing the performance of advanced statistical and machine learning techniques. Indeed, the high predictive capability of advanced methodologies (mostly based on hard information) would challenge the role of soft information and mitigate those communication frictions that hamper the successful "hardening" and transmission of soft information.

To fill this gap, this paper tests two alternative approaches grounded in both statistical learning and machine learning, and compares their respective capability in predicting SME credit risk. Specifically, we compare a classic parametric approach fitting an ordered probit model with a non-parametric one calibrating a machine learning Historical Random Forest (HRF) approach.

Our objective is to provide an alternative methodology that allows to reach accurate SME credit risk evaluation, by overcoming issues related to the transmission of soft information.

In this regard, we add to the existing studies by testing and comparing the performance of parametric versus non-parametric methodologies. However, differently from the extant literature, this paper is the first one that applies a dynamic Historical Random Forest (HRF) approach. Moreover, we further contribute by assessing the relevance of each variable to predict SME credit risk, through the use of Shapley values.

By way of preview, our results provide novel evidence that a dynamic Historical Random Forest (HRF) approach outperforms the traditional ordered probit model in assessing SME credit risk. This highlights how advanced estimation methodologies, based on machine learning techniques and mostly on hard information, can be successfully implemented in predicting SME credit risk.

To reach our research objective, we employ a unique and proprietary dataset comprising granular firm-level data on a panel of 810 Italian SMEs over the time period 2015-2017. Particular relevance is attributed to SME credit ratings. The latter are assigned to SMEs by an insurance company in the context of a revolving trade receivables securitization program initiated by a large European investment bank in favour of some of its most valuable corporate clients. Indeed, SME credit ratings are firstly produced by the insurance company and then used by the bank. In this way, the latter can assess the credit risk of the acquired portfolio of securitized trade receivables originated by its valuable corporate clients. Securitization data are matched with accounting information on our sample of 810 Italian SMEs retrieved from Orbis database. The below-explained analogy between insurance and banking SME credit ratings makes the former suitable for the purpose of our study. Indeed, both banking and insurance SME credit ratings are based both on hard and soft information. On the one hand, the former are based on relationship-intensive soft information collected directly and indirectly through continuous and personal bank-firm interactions. On the other hand, the latter are based on both proprietary soft information (i.e., client information, special investigation teams) and private and publicly-available hard information (i.e., partnerships, registered payment defaults, credit reference agencies, accounting data, payment performance data, network of risk information).

Our research question represents a matter of concerns to policy makers, since inaccurate credit risk measurement could threaten the stability of the banking sector, undermining the pivotal intermediation role played by banks in the economy. This assumes even greater relevance in light of the current COVID-19 crisis. Indeed, in periods of financial distress, an accurate credit risk assessment would allow banks to better forecast ex-ante corporate default probability.

The remainder of the paper is structured as follows. In Section 4.2 we review the existing literature. In Section 4.3 we present the empirical methodology. In Section 4.4 we describe data. In Section 4.5 we present and discuss our results. Finally, in Section 4.6, we conclude.

4.2 Related literature and our contribution

Within the existing literature, the application of alternative methodologies for estimating SME credit ratings, such as data mining techniques, tree based methodology, AI (Lin et al., 2009, Olmeda and Fernandez, 1997, De Andrés et al., 2005) or other hybrid methods (Ahn et al., 2000, Hsieh, 2004) have become widespread (Falavigna, 2006 for a detailed discussion). More recently, the latest wave of digitalization in financial markets, i.e., Fintech, has contributed to an unprecedented technological development and an increase in the number and variety of new statistical methodologies applied to the financial sector. Indeed, banks have started to explore the implementation of advanced estimation techniques for SME credit risk evaluation, although the adoption of machine learning and AI algorithms is still not fully permitted by regulators (Bussmann et al., 2020). As a matter of fact, machine learning techniques can introduce biases in lending behavior at the risk of financial inclusion and may entail issues related to consumer protection, ethics, privacy, and transparency in the eyes of supervisors and policy makers (Bazarbash, 2019). Indeed, machine learning can be harder to interpret and explain to the various stakeholders (Financial Stability Board, 2017, World Bank Group, 2019). Therefore, SME credit rating estimation has gained a renewed attention lately, also thanks to the availability of new statistical techniques and different data sources that complement the basic information available on SMEs with the aim to reach a more accurate assessment of SME credit risk.

On the one hand, we start from the existing literature and follow a path of continuity with Moscatelli et al. (2019a) and Fantazzini and Figini (2009) in terms of comparison between two types of default forecasting techniques, i.e., statistical (parametric approach) and machine learning (ML) models (non-parametric approach). Moscatelli et al. (2019a), using data on financial and credit behavioral indicators for Italian non-financial firms, present better forecasting performance with the employment of ML models, although this gain is minimal when high quality information, i.e., credit behavioral features, is added to training data and becomes negligible if the dataset is small. Overall, their results suggest that ML-based credit allocation rule results in lower credit losses for lenders. Fantazzini and Figini (2009) apply Random Survival Forests to compare their relative performance to a standard logit model and find that, while the latter outperforms the former in terms of out-of-sample accuracy, the opposite holds for in-sample accuracy.

On the other hand, we depart from the existing studies and provide a novel contribution to this stream of literature along three dimensions. Firstly, we extend Moscatelli et al. (2019a) data comparison in terms of model discriminatory power by making use of granular micro-level data collected from a large European bank and an international insurance company. Secondly, while previous studies have applied static credit scoring models to analyze the key determinants of firm credit ratings, we apply a static and dynamic modelling framework. Specifically, dynamics are introduced to analyze persistence in credit rating and compare the predictive power of the two approaches, i.e., ordered probit and Historical Random Forest (HRF). Thirdly, the lack of explainability in models with high prediction performance, i.e. ML models, has been addressed with an innovative model-agnostic interpretation approach of results known as SHAP (SHapley Additive exPlanations). Specifically, as reported in previous works (Fantazzini and Figini, 2009), while permutation feature importance helps in making comparisons among features easily, it does neither show how much each feature weights nor identify the impact of features with medium permutation importance. In this regard, the Shapley explainer is crucial to correctly understand the positive or negative contribution of a feature value to the difference between the actual and the mean prediction. This contribution extends the notion of permutation feature importance and SHAP to a static and dynamic setting for an ordered probit model and Historical Random Forest (HRF) approach.

4.3 Methodology

Given the longitudinal nature of the data, a comparison of models has been performed along two dimensions: a *static* versus a *dynamic* framework and a *parametric* versus a *non-parametric* approach. In the static setting the target rating at time t is regressed with balance sheet and securitization variables at the same time t, whilst in the dynamic setting both target and independent variables at time s, with $s < t$, are added as additional regressors. Given the ordinal nature of the target variable, an ordered probit has been selected as parametric model and the Historical Random Forest as non-parametric one. The impact of standalone balance sheets and securitization variables has been further evaluated aside of the set including all variables, adding a third dimension to the comparison analysis.

The target rating has been firstly modeled through the following static ordered probit model:

$$y_{it} = \mathbf{X}_{it}\beta + \alpha_i + \varepsilon_{it},$$

where $y_{it} \in [2,9]$ is an observed index of credit quality for the i-th firm at t-th quarter, $i = 1,\dots,N$ and $t = 1,\dots,T$, $\mathbf{X}_{it}$ indicates a vector $1 \times k$, where $k = 21$, of explanatory variables for i-th firm at time t, β is a $k \times 1$ vector of unknown parameters to be estimated, α_i is a firm-specific and time invariant component and ε_{it} is the disturbance term which is assumed to be normally distributed.

Several studies pointed out that rating changes tend to exhibit serial correlation (Carty and Fons, 1994, Gonzalez et al., 2004) and that agencies seem to be slow to react to new information (Odders-White and Ready, 2006). Therefore, the model has been extended to the dynamic framework, adding the lagged values of the dependent variable. The resulting model can be interpreted as a first-order Markov process and, following Wooldridge (2005), Contoyannis et al. (2004) and Greene and Hemsher (2008), is defined as:

$$y_{it} = X_{it}\beta + y_{i(t-1)}\gamma + y_{i0}\delta + \alpha_i + \varepsilon_{it},$$

where $y_{i(t-1)}$ indicates the i-th firm rating in the previous quarter, γ represents the parameters linked to rating in the previous quarter, y_{i0} is the first available firm rating, at time $t = 0$. Both static and dynamic version of model have been implemented using *R* package `oglmx` (Carroll, 2018).

Random forest (RF), introduced by Breiman (2001), is a non-parametric learning method based on the ensemble of decision trees, which represents one of the state-of-the-art machine learning method for prediction and classification (Capitaine et al., 2019). The static version of the model uses the classic implementation of RF where the target variable y_{it} of the i-th firm at quarter t is predicted by the dependent variable X_{it} as described in the probit model. Given the ordinal nature of the target variable, the classification version of RF has been used.

The first approaches dealing with longitudinal and clustered data involved tree-based methods (Segal, 1992, Hajjem et al., 2014, Sela and Simonoff, 2012) and are based on the idea of iterating between fixed and random part and estimating the parameters via Expectation Maximization (EM) algorithm. All these approaches represent semi-parametric fixed effects model in which the non-parametric part is evaluated through RF. The main contribution of this paper regards the application of an innovative random forest algorithm based on historical trees, suitable for longitudinal data and implemented in the *R* package `htree` (Sexton, 2018).

Let us consider longitudinal data with n individuals, the i-th individual having n_i observations over time. Specifically, data is assumed to be of the form

$$z_{ij} = (y_{ij}, t_{ij}, x_{ij}),$$

with $i = 1,\dots,n$ and $j = 1,\dots,n_i$, where y_{ij} represent the response of the i-th individual at time t_{ij} and x_{ij} the vector of predictors at time t_{ij}. The method applies both with regular and irregular sampling in time, i.e., the number of observations can be different for each subject. HRF estimates the response variable y_{ij} using the concurrent observations and all preceding observations of the i-th individual at (but not including) time t_{ij}. Node splitting follows the standard approach of RF, e.g. minimizing the Gini impurity or the Cross-Entropy for classification or Root Mean Square Error for regression, except for the fact that the number of observations of an historical predictor will vary according to i and j. In particular, a summary function first transforms all previous values of a predictor and is denoted as $s(\eta;\bar{z}_{ijk})$ where η represent parameters of the function and $\bar{z}_{ijk}$ denotes the set of historical values of the k-th element of z_{ij}. Then, node splitting is done by minimizing the following expression over the vector (k, μ, c, η):

$$\operatorname{argmin} \sum_{(ij) \in \text{Node}} (y_{ij} - \mu_L I(s(\eta; \bar{z}_{ijk}) < c) - \mu_R I(s(\eta; \bar{z}_{ijk}) \geq c))^2,$$

where $I(\cdot)$ is the indicator function, μ_L and μ_R are the weighted number of samples reaching node in the left and right split, respectively and c is the splitting threshold. Setting $n_{ij}(\eta)$ as the number of observation of the i-th individual in the time window $[t_{ij} - \eta_1, t_{ij} - \eta_2]$, the set of possible value of η_1 and η_2 is determined by the difference in time between within-individual successive observations in the data, both provided by the user or selected among the quantiles of the corresponding distribution. When a split is attempted on a historical predictor, a sample of this set is taken upon which the best split is selected and the size of this set can be defined by the user as well. The summary function can be defined in different ways, according to its set of parameters η. For example:

- "frequency"
$$s(\eta; \bar{z}_{ijk}) = \sum_{h: t_{ij} - \eta_1 \leq t_{i_h} < t_{ij}} I(z_{ihk} < \eta_2)$$
- "normalized frequency"
$$s(\eta; \bar{z}_{ijk}) = \sum_{h: t_{ij} - \eta_1 \leq t_{i_h} < t_{ij}} \frac{I(z_{ihk} < \eta_2)}{n_{ij}(\eta)}$$
- "average"
$$s(\eta; \bar{z}_{ijk}) = \sum_{h: t_{ij} - \eta_1 \leq t_{i_h} < t_{ij}} \frac{z_{ihk}}{n_{ij}(\eta)}$$

If $n_{ij}(\eta) = 0$, the summary function is set to zero.

In order to evaluate the performance of both models, to optimize the hyperparameter of HRF and to select the optimal subset of variables of the probit model, a set of evaluation metrics has been taken into consideration. First of all the confusion matrix has been used to assess the accuracy of the prediction of each rating class and the F_1-score was selected as an aggregated metric. Moreover, the difference of performances on train and validation set must be minimized when tuning the hyperparameters so that overfitting can be avoided. Therefore, a weighting adjustment on the F_1-score has been selected among the following:

- $F_{1\,\text{ratio}} = F_{1\,\text{test}} + \frac{F_{1\,\text{test}}}{\Delta F_{1\,\text{train-test}}}$
- $F_{1\,\text{harmonic}} = \frac{2}{\frac{1}{F_{1\,\text{test}}} + \frac{1}{\Delta F_{1\,\text{train-test}}}}$
- $F_{1\,\text{cross-entropy}} = -F_{1\,\text{test}}^{\gamma} \log(1 - F_{1\,\text{test}}) - (1 - \Delta F_{1\,\text{train-test}})^{\gamma} \log(\Delta F_{1\,\text{train-test}}), \gamma \geq 1$

The most efficient weighting resulted to be $F_{1\,\text{cross-entropy}}$ with $\gamma = 4$.

Validation of model performances and train/validation set splitting of the data have been evaluated with a variable-length rolling-window temporal approach. In particular, given that the maximum number of available quarters is 10 and the variable amount of total quarters for each firm, a validation set of the 2 most recent quarters and a train set of all remaining quarters have been chosen. As the minimum number of available quarters for each firm is 7 and a minimum number of observations in each train set has been fixed to 2, the final number of folds used in the cross-validation is 4.

HRF hyperparameters tuning has been performed by means of a Bayesian Optimization approach through *R* package `rBayesianOptimization` (Yan, 2016).

A comparison of explanatory power of all combination of models, framework and set of variables has been added to the predictive power one using two relevant state-of-the-art techniques: Permutation Feature Importance (PFI) and SHAP values. The aim of both methods is to estimate the importance of each variable determining the most relevant ones for the prediction.

In the PFI the importance of each feature is evaluated by computing the gain in model's prediction error after shuffling feature's values. A feature is considered relevant for model's prediction if the prediction error increases after permuting its values, otherwise, if model error remains unchanged, its contribution is not important. As proposed by Fisher et al. (2018), the algorithm for a generic model f can be defined as:

Algorithm 1: Permutation Feature Importance

```
Input: Trained model f, feature matrix X, target vector y, performance metric
       P(y, f)
Estimate the original model performance P_orig = f(y, X);
foreach feature j = 1, ..., p do
    Generate feature matrix X_perm by permuting feature j in the data X;
    Estimate P_perm = f(y, X_perm) based on the predictions of the permuted data;
    Evaluate PFI_j = P_perm / P_orig. Alternatively, the difference can be used:
     PFI_j = P_perm - P_orig;
    return PFI_j;
end
Sort features by descending PFI
```

Shapley values represent the marginal contribution of each feature to the prediction of a given data point. The feature values for instance x behave like players in a game where the prediction is the payout. As described in Shapley (1953), the Shapley value Φ_j of a feature value x_j, is defined by means of a value function *val* of actors in S and represents its contribution to the prediction, weighted and summed across all possible coalitions:

$$\Phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} (val(S \cup \{x_j\}) - val(S))$$

where S denotes a subset of features, x represents the feature values of the instance of interest and p the number of features and $val_x(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in S:

$$val_x(S) = \int \hat{f}(x_1,\ldots,x_p) d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X))$$

Estimating the Shapley values for more than a few features becomes computationally infeasible since all possible coalitions of feature values need to be considered with and without feature j. A Monte-Carlo sampling was proposed by Strumbelj and Kononenko (2014):

$$\hat{\Phi}_j = \frac{1}{M} \sum_{m=1}^{M} (\hat{f}(x_{+j}^m) - \hat{f}(x_{-j}^m))$$

where $\hat{f}(x_{+j}^m)$ represents the prediction for the instance of interest x but with a random permutation of features (taken from a random data point z) except for j-th feature. The vector x_{-j}^m is identical to x_{+j}^m, but the value for feature j is randomized as well from the sampled z. The algorithm for a generic model f can be defined as:

Algorithm 2: Shapley value

Output: Shapley value for the value of the j-th feature
Input : Number of iterations M, instance of interest x, feature index j, data matrix X, and machine learning model f

foreach $m = 1, \ldots, M$ **do**
  Draw random instance z from data matrix X;
  Choose a random permutation o of the feature values;
  Order instance x: $x_O = (x_{(1)}, \ldots, x_{(j)}, \ldots, x_{(p)})$;
  Order instance z: $z_O = (z_{(1)}, \ldots, z_{(j)}, \ldots, z_{(p)})$;
  Construct two new instances:
  - With feature j: $x_{+j} = (x_{(1)}, \ldots, x_{(j-1)}, x_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
  - Without feature j: $x_{-j} = (x_{(1)}, \ldots, x_{(j-1)}, z_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$

  Compute marginal contribution: $\Phi_j^m = \hat{f}(x_{+j}) - \hat{f}(x_{-j})$;
  return Φ_j^m;
end
Compute Shapley value as the average: $\Phi_j(x) = \frac{1}{M} \sum_{m=1}^{M} \Phi_j^m$

This procedure needs to be repeated for each feature of interest in order to get all the Shapley values. Among the advantages of Shapley values over the other methods, in first place there is the efficiency property, i.e., the difference between prediction and average prediction is fairly distributed among features. It is important to remark that the SHAP values have been computed for this multiclass problem in order to investigate, for each class, how the predictors bring up or down the probability of belonging to a certain class, compared to the average probability of this class for the full data.

4.3.1 Statistical assessment of differences

After the evaluation of the quality of a multiple learned classifiers and consequent classification of new samples with unknown class labels, the statistical comparison of classifiers is needed to assess the statistical differences between the results obtained by different algorithms in different instances of problems, datasets, etc. The typical sequence of analysis involves, firstly, using a test that compares simultaneously all the considered algorithms in order to test the presence of any algorithm that behaves differently. Then, if the null hypothesis is rejected (i.e., if the results show globally significant differences), the next step is analyzing which pairwise combinations are different by implementing post hoc tests.

Firstly, classical non-parametric Friedman test (Friedman, 1937) has been implemented. Sometimes observations do not meet measurement requirements and in order to avoid assumptions about the underlying populations, nonparametric statistical tests would be appropriate, like Friedman's test. The latter represents an alternative nonparametric procedure to the parametric two-way analysis of variance and it is used to detect differences in treatments across multiple test attempts. The computational procedure involves ranking each row together, ordering the rows values in decreasing order and calculate the average rank for each column. To compare two columns, the formula is the following:

$$z = \frac{(R_i - R_j)}{\sqrt{\frac{k(k+1)}{6N}}}$$

where R_i is the average rank obtained from the Friedman test for the i-th column, k represents the number of columns and N the number of blocks sets both used for comparison purposes.

The underlying idea is to compare the accuracy of different classifiers using different data; as a consequence, columns will be the classifiers and rows the datasets.

Then, the corresponding post hoc tests for Friedman have been implemented, correcting p-values for multiple testing (Bergmann and Hommel's correction procedure). The latter applies a correction based on a list of possible hypothesis testing and amplifies the test power by considering only exhaustive sets of hypothesis (i.e., hypothesis that can be simultaneously true).

4.4 Data

Data used in this paper can be divided into two main categories: securitization (SEC) vs accounting (BS) data. Data on securitization transactions have been collected from a large European bank which plays a leading role in the niche of revolving trade receivables' securitization programmes. Data on accounting figures have been collected from Orbis database, developed by Bureau Van Dijk (a Moody's analytics company), by matching the VAT code for each given borrowing firm[1]. Among securitization data, a peculiar role is played by a measure considered as the expression of a given borrower's credit quality: the credit rating. The credit rating, attributed by the insurance company to each SME, analyses the given SME's financial health and creditworthiness to predict its default risk based on both proprietary soft information (i.e., client information, special investigation teams) and private and publicly-available hard information (i.e., partnerships, registered payment defaults, credit reference agencies, accounting data, payment performance data, network of risk information)[2]. Therefore, insurance rating provides an objective and quantifiable means by which a company's degree of credit risk can be assessed. More specifically, the credit rating, i.e., our target variable, is a factor variable with eight categories, ranging from 2 till 9. Through this rating, the insurance company is capable of distinguishing between high-risk and low-risk clients by assigning, respectively, high and low insurance ratings. The insurance rating assigned to each SME is categorized in a numeric scale ranging from 2 to 9 according to the given firm credit risk. The higher the number, the worst the credit rating. Credit ratings evolution over time is showed in Fig. 4.1, highlighting an overall persistent behavior for all classes of risk. Proprietary credit rating data have been linked to Orbis' firm-level balance sheet statements and profit and loss accounts for the analyzed time period. More specifically, the dataset consists of 34 variables for 810 Italian firms collected from Q1 2015 to Q2 of 2017. In particular, 6 numerical securitization variables (hereinafter referred as SEC) were provided by the Insurance Company with quarterly frequency and 28 (25 numerical and 3 categorical) balance-sheet variables (hereinafter referred as BS) were collected by Orbis platform with annual frequency. Annual values are repeated over all quarters of each year. Furthermore, Nace Rev. 2 has been used to classify firms' main sector (NACE) and main division (Industry). Geolocalization variables have been extracted through Google Maps API and have been linked to each SME present in the dataset to control for unobserved heterogeneity in the given SME's industry and location. Tab. D1 reports the definition of the variables used in the empirical analysis and some descriptive statistics.

[1]The database construction process played a crucial role in making such an empirical analysis possible, despite being time-consuming due to the required manual input of proprietary micro-level data, properly integrated with additional accounting data collected from Orbis database.

[2]Scores are not permanent and can be affected by different factors. There are several ways to increase low scores and possibly lower premiums. To begin, a consumer will benefit by improving his or her credit rating and paying bills on time, as well as reducing debt. Also, limiting the number of insurance claims filed over a certain period can help boost an insurance score.

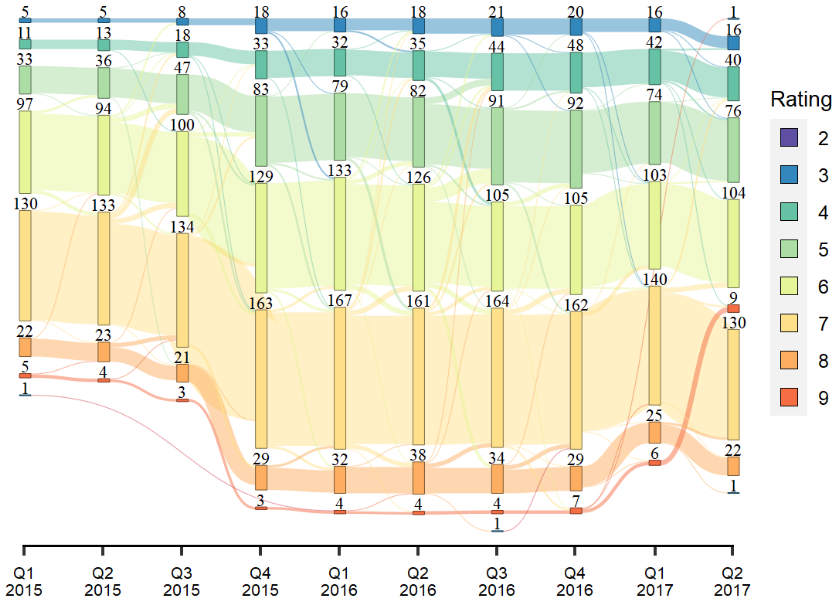


Figure 4.1: Rating evolution over time.

4.4.1 Dataset construction and cleaning

The initial raw dataset contained a panel data of 810 firms in the time period that spans from the first quarter of 2015 to the second quarter of 2017. Data were then cleaned, checked for outliers, redundant data and missing values.

Firstly, SMEs with too many missing values were excluded from the analysis, specifically those SMEs with a large number of null observations with respect to Orbis variables and with null securitization data were not considered in the analysis, given the different industrial sector, size and number of employees of the firms (we removed 254 firms with more than 10% of missing variables and 92 firms with null securitization data). Moreover, not all firms were available for the entire time period because some entered the insurance company portfolio after Q1 2015 and some left before Q2 2017. Therefore, given the high percentage of missing, some quarters of SEC variables were excluded as well. However, missing values (aka NA) for the cleaned dataset are still present, although with a low incidence (less than 10%). Details on the strategy used for handling NA are reported in Appendix D.

Secondly, data cleaning involved checking for holes in credit ratings. In this regard, the exploratory data analysis resulted in 1,484 missing values for the target variable (including also zero values, which are meaningless). In the overall dataset, the percentage of missing is 18.3%, with highest percentages of missing values showed at the beginning and at the end of the considered time period, i.e., the first three quarters of 2015 and the first two quarters of 2017. Moreover, some gaps within the time series were founded and replaced with recovered additional data.

Thirdly, the distribution of the time difference between consecutive years has been investigated in order to understand if computing the average to recover additional missing data was an appropriate procedure. In this regard, the stability of the distributions proved that this approach was a suitable one to follow.

Fourthly, the variables have been normalized with respect to a different set of features, according to the securitization or accounting nature of the specific data in order to obtain a normalized set of predictors between 0 and 1. However, some extreme values have been deliberately left in the dataset to reflect the extreme characteristics of some SMEs with respect to the normalized range and to avoid having a too lean dataset in terms of number of observations.

Lastly, outliers have been removed based on inter-quantile range (α–quantile and $(1-\alpha)$–quantile) but, in order to keep values of variables with small variance, if the distance between maximum and minimum value was less than an arbitrary value of tolerance, then no outlier has been removed.

The whole dataset cleaning process resulted in a final dataset comprising 534 Italian SMEs in the time period that spans from the first quarter of 2015 to the second quarter of 2017, from our original panel data composed of 810 Italian SMEs.

4.4.2 Additional variables and transformation

Dataset has been augmented with additional variables created from the original ones and, given the different size of companies, both BS and SEC variables have been normalized. Moreover categorical variables have been converted into dummies, dropping the $n-th$ level in order to avoid multicollinearity. In particular:

- adding log10 of annual *Turnover* and quarterly *Outstanding*;
- adding annual ratio *Delinquency* and *Delinquency90* and *Outstanding*;
- adding quarterly ratio *Outstanding_Invoices* and *Outstanding_Portfolio* of *Outstanding* and *InvoicesCount* and *PortfolioCount*, respectively. Both ratios, being still an average amount of money, are further normalized by annual *Purchase*;
- adding annual binary dummy *Liquidity Tension*, 1 if *Collectionperioddays* is greater than *Creditperioddays*, 0 otherwise;
- adding annual binary dummy *Dummy_Delinquency*, 1 if *Delinquency* is greater of equal then the average *Delinquency* of all firms, 0 otherwise;
- adding annual binary dummy *Delinquency Severe*, 1 if *Delinquency* is greater of equal then the average *Delinquency* of all firms with a two standard deviations confidence, 0 otherwise;
- adding regional dummy *Region* aggregating each city into 4 geographical macro-areas: North-East, North-West, Center and South+Islands;
- normalizing all BS variables by *TotalAsset*;
- normalizing all SEC variables by *Outstanding*.

Finally, correlation among the BS and SEC variables has been checked and 9 variables have been removed according to a Variance Inflation Factor (VIF) value above 5. A complete list of variables description and corresponding statistics is reported in Tab. 4.1. Dummy and categorical variables distribution by each rating is reported in Tab. 4.2. Geographical distribution of firms is showed in Fig. 4.2.

Final dataset consisted of 464 firms and 21 variables, 6 SEC and 15 BS, and was then treated according to an unbalanced panel data structure, resulting in 3,009 rows.

Table 4.1: List of final variables.

Variable	Description	Mean	Stdev	Median	Minimum	Maximum	Removed due to VIF
Rating	Rating score, 2 means low risk	5.1091	1.2443	5	2	9	
Purchase	Accounting of Cash and Credit purchases	1.4914	0.9555	1.3062	0.0168	6.4811	
Current liabilities	Company's debts or obligations that are due to be paid to creditors within one year	0.5480	0.1996	0.5457	0.0383	2.0324	
Current ratio	Comparison a firm's current assets to its current liabilities	0.0136	0.0089	0.0121	0.0008	0.1920	x
Delinquency	Dummy variable equal to 1 if the firm misses a scheduled payment on an invoice, otherwise equal to 0	0.0162	0.1043	0	0	1	
EBIT	Company's net income before income tax expense and interest expenses are deducted	0.0485	0.0881	0.0400	-1.4438	0.6867	
Fixed assets	Long-term tangible piece of property or equipment that a firm owns and uses in its operations	0.3316	0.2169	0.3045	0.0014	0.9833	x
Collections	Amount of invoices currently sold to the bank	2.7146	90.7025	0.7687	0	5520.7044	
Liquidity	Company's ability to pay off current debt obligations without raising external capital	0.0104	0.0078	0.0091	0.0008	0.1594	
Outstanding	Amount of securitization transactions in which the borrowing firm is involved, expressing its economic exposure in logarithmic scale (base 10)	4.1590	1.9485	4.6758	0	7.1786	
Turnover	Annual sales volume net of all discounts and sales taxes in logarithmic scale (base 10)	4.5325	0.8341	4.4351	2.8520	6.9362	
LT Debt	Debt with maturities greater than 12 months	0.0911	0.1035	0.0540	0	0.5163	
Asset Turnover	Sales revenue divided by capital employed	0.0753	0.1769	0.0412	0.0008	3.5299	x
New Receivables	Monetary amount of receivables sold to the bank with respect to a given borrowing firm at the current invoices' transfer	0.2060	0.2355	0.1634	0	1	
Outstanding_Invoices	Amount of securitization transactions in which the borrowing firm is involved divided by total number of invoices and annual Purchase	0.1332	0.3392	0.0002	0	1	x
Outstanding_Portfolio	Amount of securitization transactions in which the borrowing firm is involved divided by total number of portfolios and annual Purchase	0.1454	0.3464	0.0017	0	1	x
Profit Margin	Percentage of sales turned into profits	0.0240	0.0643	0.0174	-0.7288	0.5611	
Profit per employee	Net Income for the past twelve months (LTM) divided by the current number of Full-Time Equivalent employees	0.0047	0.0502	0.0004	-0.0178	1	
ROA	Net income divided by total assets	0.0318	0.0887	0.0210	-0.3528	1.9188	
ROCE	Company's earnings before interest and tax (EBIT) divided by its capital employed	0.0782	0.2240	0.0710	-7.3081	0.8548	x
ROE	Fiscal year net income divided by total equity	0.0822	0.6385	0.0831	-13.7168	9.7300	
Solvency_A	Firm's capacity to meet its long-term financial commitments	0.2834	0.1843	0.2468	-0.7866	0.9333	
Tangibles	Assets that have a physical value	0.2477	0.1931	0.2121	0	0.9797	
Working Capital	Difference between a company's current assets and its current liabilities	0.1372	0.2414	0.1195	-1.7193	1.0661	
Delinquency Severe	Dummy variable equal to 1 if delinquency_outstandingpost is larger of equal than +2 standard deviations from the mean of all clients						
Delinquency 90	Dummy variable equal to 1 if Scaduto90 (i.e., payments overdue by more than 90 days evaluated on average by ID) is larger than 0, otherwise equal to 0						
Liquidity Tension	Dummy variable equal to 1 if Collectionperioddays (number of days it takes to turn accounts receivable into cash) is larger than Creditperioddays (number of days that a customer is allowed to wait before paying an invoice), otherwise equal to 0						x
NACE	Statistical Classification of Economic Activities in the European Community						
Industry	Industrial classification variable reflecting main division within main section of NACE						x
Region	Geographical macro-areas						

Table 4.2: Dummy and categorical variables distribution by each rating.

		Rating							
Variable		2	3	4	5	6	7	8	9
NACE									
Manufacturing		20%	23%	28%	34%	35%	31%	18%	0%
Wholesale and retail trade; repair of motor vehicles and motorcycles		57%	61%	54%	53%	54%	59%	82%	100%
Accommodation and food service activities		22%	8%	15%	10%	7%	8%	0%	0%
Agriculture, forestry and fishing		2%	7%	1%	1%	2%	1%	0%	0%
Other		1%	6%	3%	2%	2%	3%	0%	0%
REGION									
North-East		24%	35%	42%	38%	26%	20%	8%	0%
North-West		61%	35%	28%	23%	26%	25%	31%	0%
Center		6%	20%	16%	16%	19%	29%	12%	25%
South+Islands		9%	10%	15%	22%	30%	26%	49%	75%
DUMMY									
Delinquency Severe	0	74%	90%	92%	98%	100%	100%	100%	100%
	1	26%	10%	8%	2%	0%	0%	0%	0%
Delinquency 90	0	47%	29%	60%	74%	82%	82%	43%	25%
	1	53%	41%	40%	26%	18%	18%	57%	50%
Liquidity Tension	0	76%	69%	68%	54%	50%	56%	59%	100%
	1	24%	31%	32%	46%	50%	44%	41%	0%

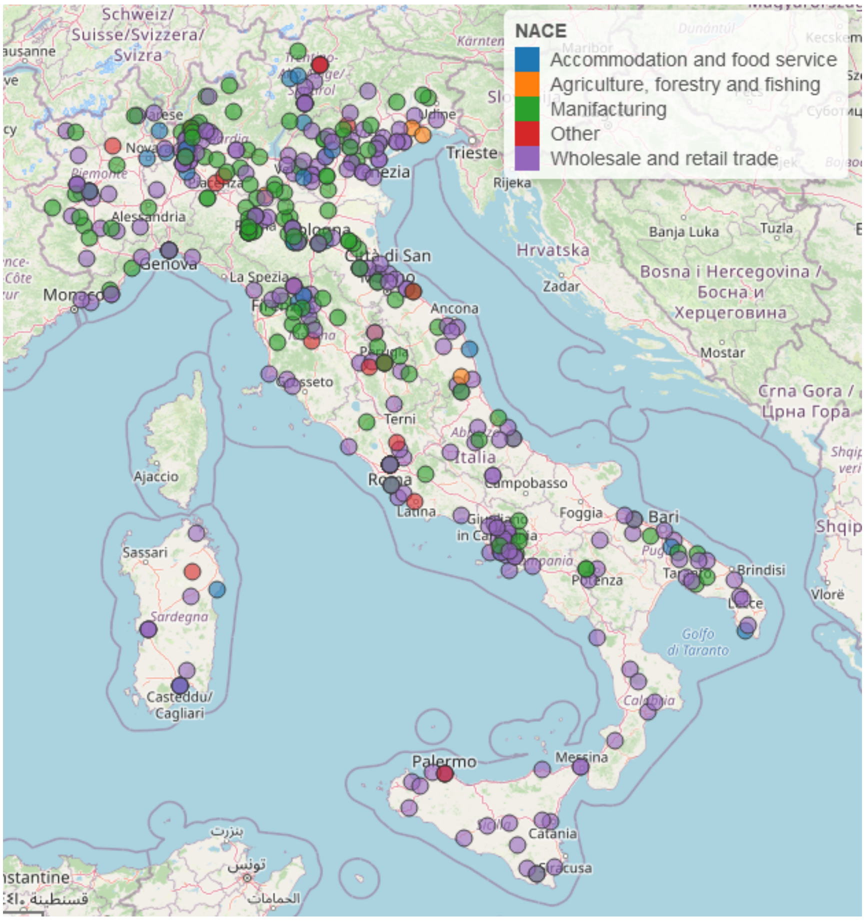


Figure 4.2: Geographical distribution of firms.

4.5 Empirical Analysis

In this section, comparison of classification performance of models along three dimension is presented. As introduced in section (4.3), models have been distinguished according to *static* versus *dynamic* framework, *parametric* versus *non-parametric* approach and *BS* vs *SEC* set of predictors. Model evaluation has been made in terms of macro-averaged F_1-score on both in-sample and out-of-sample predictions. At the end, an optimal set of predictors, combining both *BS* and *SEC* set, has been identified according to feature importance evaluated on the last dimension of comparison analysis. Based on the latter, classification performance has been presented.

4.5.1 Model evaluation

A set of evaluation metrics has been used (D.2) in order to obtain an optimal combination of hyperparameters (HRF model) and variables (PB model). Respectively, F_1 cross-entropy metric has been maximized during cross-validation phase to avoid ovefitting; Akaike information criterion (AIC) has been minimized during best subset selection to obtain a stable function of predictors.

Tab. D2 shows the optimal hyperparameters set and the selected set of predictors with reference to SEC set of predictors. Based on the shown model architecture, a summary of classification performance with regards to both the training and validation sample is shown in Tab. 4.3. F_1-score with regards to HRF and dynamic PB can be highlighted as the lowest ones compared to the case without temporal dependence. Previous rating states, together with predictors' history, seem to be necessary for a correct classification of insurance credit risk. Dynamic PB outperforms the other estimated models, with 80% - 70% F_1-score respectively on train and test set. Tab. D3 reports the proposed models for the BS set of predictors, for both dynamic and static version.

A first difference between SEC and BS set that can be pointed out is the number of predictors, lower in the former with respect to the latter; this could influence the analysis resulting into different conclusions for this set of variables. Analyzing the classification performance on BS set in Tab. 4.3, HRF outperforms the other models in terms of overall classification performance with an F_1-score of 90% for train set and 70% for test set. If compared with SEC case, the persistence of rating history leads to more accurate prediction in the PB case but does not have a significant influence in the accuracy of HRF predictions.

Based on the results in terms of variable importance and best subset selection aimed at reducing the high dimensional feature space and obtaining an optimal group of features, a final model has been implemented by combining both BS and SEC set of variables. Before fitting the models, preliminary analysis has been implemented in order to check for correlation and collinearity. The results report significant correlation between Outstanding and Turnover, that leads to the removal of Outstanding since a measure of firm's financial exposure has already been selected with Delinquency and having a metric for firms' size in terms revenues seems to be useful in the determination of credit rating. Even if the regional and industrial classification variables have a significant effect on the target only for one specific category and the impact cannot be confirmed in terms of importance, both regional and industrial classification variables have been kept in order to have an insight about the economic framework. To summarize, the following variables have been selected for the final set: *Collections, New Receivables, Delinquency, Turnover, Solvency_A, Working Capital, LT Debt, Current liabilities, Liquidity, NA- CE, Region*. In terms of classification performance (4.3), HRF outperforms the other models, with good performance on a macro level (90%-70% F_1-score on train and test set). Comparing the final model with the BS set of variables, no relevant differences can be reported since the performances are almost the same. This could be due to the high

Table 4.3: Macro-averaged F_1-score on training and test sample for all set of predictors.

			F_1-score		
Model	Version	Sample	BS	SEC	BS + SEC
HRF	Static	Train	0.919	0.4105	0.9611
		Test	0.677	0.3417	0.6986
	Dynamic	Train	0.9154	0.7480	0.9014
		Test	0.7361	0.5519	0.7326
PB	Static	Train	0.4634	0.4579	0.4609
		Test	0.4529	0.4284	0.4543
	Dynamic	Train	0.8148	0.7901	0.799
		Test	0.7407	0.7346	0.74487

numerosity of variables for the BS set that allows to reach the performances of the optimal set. To conclude, the autoregressive behavior of the models is necessary to reduce the misclassification cost.

4.5.2 Model explanation

In this section, explainability capabilities of both HRF and PB have been compared using PFI and SHAP values together with marginal effects. On one side, the change in probability correlated to each predictor has been explored in order to understand the sign of the effect on each class of the target variable; on the other side, more complex relationships have been investigated through SHAP values. As a result, the most relevant features, in terms of relative importance, have been selected in order to implement the final credit-scoring model. According to classification performance, feature importance figures with reference to the best statistical models for the three set of variables have been reported in D.3. As classifiers based on probit model do not seem to work better than random choice (i.e. accuracy metrics less than 50%), results in terms of feature importance are meaningless.

With regards to PFI, relative importance has been computed as difference between the original and the permuted F_1-score then averaged and normalized over the sum of the absolute values of all the obtained permutation metrics. This procedure results in a range of values between 0%-100%, with a negative score when a random permutation of a feature's value results in better performance metric and high importance score when a feature is more sensitive to random shuffling, i.e., it is more "important" for prediction. In the process of selecting the most important predictors, the features are considered, individually, in terms of relative importance ranking and, on an aggregated level, in terms of total percentage of relative importance carried by the features in top position. Related figures are presented on a macro-level (aggregated for all Rating classes) and distinguished according to time dependence. The latter distinction has been carried out when both models (static and dynamic version) report accuracy metrics higher than 50% on test set. Otherwise, only one case has been analysed.

PFI helps to easily make comparisons between features but it does not tell how each feature matter and does not allow the identification of the impact of features with medium permutation importance. The Shapley explainer is crucial to correctly understand why a model predicts a given class for a given ID on a given time period (single row-prediction pair), since it goes through the input data, row-by-row and feature-by-feature, changing its values to identify how the base prediction differs holding all else equal for that row and, as a consequence, explains how this prediction was reached. The contribution of each variable towards the single row-prediction compared to the base prediction for the full data set is called Shapley value (*phi*). On a multiclass perspective, SHAP will output a separate matrix for each class prediction for the given row in order to understand how, for each class, the predictors bring down or up the probability of belonging to that specific class. The Shapley values of each feature have been aggregated in two ways based on the average contribution computed by feature and grouped according to rating classes with the aim of investigating how each feature impacts, on average, on the predicted probability of each class compared to the average probability of this class for the full dataset. Given the best performance of Random forest algorithm as an ensemble of historical classification trees and the slow computational procedure in calculating the Shapley values, figures with reference to only dynamic HRF model have been reported in D.3 for comparison with PFI.

Starting from BS set of variables, Fig. D1 shows the importance ranking in terms of PFI for both static and dynamic version of the HRF model. Relative importance values slightly vary if the autoregressive behavior is considered in the model or not. With reference to the static version of the model, Turnover can be observed in the top position with 47% of relative importance value, followed by Solvency (17%), Working Capital (7%), LT Debt (6%) and Liquidity (5%) with lower order of magnitude. On an aggregated level, the previous features represent almost the 90% of relative feature importance over the total of 20 considered predictors. Considering autoregressive behavior of the target and predictors, the LagRating shows a slightly negative relative importance of 1%, carrying no positive effect on model prediction error if compared to balance sheet informations in the top of the graph. Importance of considerable order is reported by the same previous set of BS variables, with the addition of Profit Margin (10%). Quantitative variables are more important than qualitative ones, i.e. NACE and Region, since each dummy has a frequency that affects its importance value. Opposite behaviour is reported from Probit model (Fig. D2), where autoregressive behaviour seems to carry 90% of relative importance on model prediction error. As a result, the other variables reports negligible relative importance scores.

Furthermore, for PB model, marginal effects can be analyzed in order to investigate the change in probability when the predictor variable increases by one unit. According to Tab. D5 , which reports the PB models distinguished according to autoregressive behavior, it can be noticed that Rating class 6 represents a threshold for change of sign of partial derivatives, allowing the interpretation of results by distinguishing between low-risk (3,4,5) and high-risk (6,7). Regarding the key indicators to the financial solvency of the company, i.e., Current liabilities, LT Debt and Working capital, an increment of these metrics implies positive impact on the probability of belonging to high-risk classes. Higher long and short term financial obligations reflect higher debt and, consequently, higher risk. Together with debt, Working capital has the same effects on rating classes because, being computed as the difference between shares and the sum of trade credit and payables, the positive sign could reflect the weight of trade payables in the short term that, in the case of SMEs, is particularly high and could result in higher risk. Furthermore, high working capital is flagged as having liquidity issues, since a company is not effectively reallocating capital into higher growth. Contrarily, Liquidity, ROA, Tangibles, Collections e Turnover show negative marginal effects in correspondence of the riskiest classes, since high values for liquidity, profitability and size measures represent a signal of solid financial and operational performance. As a consequence, a rise in these metrics is associated with higher probability of belonging to low-risk classes. Specifically, high liquidity implies better ability of the company to meet its short-term obligations on time, resulting in lower debt and, consequently risk; associated with a healthy profile, the efficiency of the management and the annual sales volume as signals of firm expansion and consolidated business model.

Regarding categorical variables, regional classification seems to have a significant impact on the predictive power, since the marginal effects show a high level of risk in Southern Italy, possibly caused by the different economic context compared to Northern Italy. Belonging to some specific classes (3-4-5) does not imply temporal persistence of those rating classes over time; on the contrary, belonging to all classes (except Rating 7) seems to increase the probability of having a rating score of Rating 6, given probably the high numerosity of the latter. In addition, all the lagged classes show positive impact on the riskiest Rating 7. However, given the complexity of the classification problem at hand, defining the target variable as binary (4.5.3) allows to understand that there exists a persistence in belonging to low-risk and high-risk classes over time.

Furthermore, Shapley values (Fig. D3) confirm previous results, highlighting high average contribution of Turnover, together with Solvency, Profit margin, Working capital and Liquidity. As expected, heterogeneous contribution is carried by aforementioned features with respect to Rating classes, with highest impact on the largest ones (i.e. Rating class 4,5 and 6).

Following the same computational procedure for the SEC set of variables, it can be noticed that a relevant role is played by LagRating (97%) within HRF modelling framework, followed by slightly positive scores of Outstanding and New Receivables in the determination of rating score (Fig. D4). Securitization variables reports negligible contribution if compared to time dependence. Same conclusions could be reported for Probit model (Fig. D5).

Delinquency and Outstanding represent metrics of economic exposure of the firms under investigation; the former with respect to missed payments and, the latter, to securitization transactions in which the borrowing firm is involved. These metrics are directly linked to the level of risk reported by each firm. On the other side, New Receivables measures trade balance credits in terms of volume and lengthening of deadlines. Higher position of trade credit could reflect liquidity drainage, i.e., less investment availability.

As mentioned before, the partial derivatives (Tab. D6) highlight class 6 as threshold for change of sign, and, as a consequence, an increase in New Receivables and Delinquency results into a positive effect on the probability of belonging to the riskiest rating classes. The opposite behavior is showed by Outstanding and Collections. With regards to the lagged dummies of the target variable, same conclusions as for the BS set can be extracted.

Shap results allow to grasp individual contribution of securitization variables (Fig. D6). Delinquency and Delinquency 90 report the highest average impact on Rating classes, being relatively important in HRF classifier.

Among the combined set of variables, Turnover shows a predominant role with a relative importance of about 30% on a macro level, followed by Solvency, Working Capital, LT Debt and Liquidity, reaching a total relative importance of 70% for the dynamic case and 80% for the static one (Fig. D7). Within the optimal set of variables, the time contribution is in the lowest position, with slightly negative PFI; the selected predictors enable to better differentiate between classes without allowing for the persistence of credit history. The BS set of variables overcomes, in terms of PFI, the SEC predictors. On the contrary, PB model highlights 92% contribution of time dependence (Fig. D8). In line with the previous results, the partial derivatives of Tab. D7 highlight conclusions already mentioned for the distinct set of variables (BS and SEC); it is worth noticing the high levels of significance for all the marginal effects.

The Shapley values report, overall, the same importance ranking for the selected set of predictors; Turnover, Working Capital, Liquidity, LT Debt and Current liabilities show the highest magnitude in terms of average impact by feature and class (Fig. D9). The magnitude of the features' effect is smaller for SEC set compared to BS one; the latter has relevant impact on the classes with largest number of observations (4,5,6). Specifically, the combination of Current liabilities, Liquidity, Solvency and Turnover plays a significant role for the identification of the extreme classes, bringing up or down the probability of belonging to that specific class.

4.5.3 Assessment of differences and robustness checks

According to the methodology presented in 4.3.1, statistical comparison of classifiers has been implemented to assess significant differences between the results obtained in the previous section. Firstly, macro-weighted balanced accuracy obtained by the previous algorithms in the three different datasets (i.e., BS, SEC and BS+SEC) has been imported; then, differences have been tested on algorithm (further divided based on autoregressive behavior) and dataset level. Since results obtained from Friedman test show globally significant differences on algorithm level, the next step involves analyzing which pairwise combinations are different. The p-value matrix generated when doing all the pairwise comparisons show significant differences at 0.05 level for the PB model based on different time dimension, highlighting the temporal component as statistically significant discriminant between algorithms (Tab. 4.4).

Additional checks have been performed to test the robustness of previous findings, in particular alternative formulation of the target variable. The latter test attempts to reduce the multiclass problem to multiple (or single) binary classification problems (one class vs the others or high-risk vs low-risk class) in order to check the accuracy of results in comparison to the ordinal formulation of the target variable.

Given the complexity of the classification problem at hand and the subtlety of the different behaviors that the classifiers exhibit, the ordinal scale has been converted to dichotomous variable. Firstly, a formulation Rating 7 vs ALL has been implemented, resulting into poor performances given the imbalanced nature of the dataset with respect to the tails. Then, the target variable has been defined as High-risk Rating (class 6 and 7) compared to all the other classes, in order to check if the models are able to more accurately price a risk and differentiate between lower and higher insurance risks. The descriptive analysis highlights a balanced distribution of observations in the two groups for both the considered set of predictors. Overall, the alternative formulation of the target variable affects positively the SEC case since the classification metrics are slightly higher (+0.1) compared to the ordinal one. For the other set, the performances are almost the same, except for the PB case where the metrics are better with the binary target. The selected set of variables, for PB model, is the same and the marginal effects of the binary cases reflect exactly the duality into the sign of the partial derivatives for ordinal case, since the threshold that highlights the change of sign is class 6. Summarizing, the binary formulation simplifies the classification problem at hand and results in slightly higher performances together with same explainable conclusions as for individual risk.

Table 4.4: Corrected p-value matrix using Bergmann and Hommel's correction procedure generated when doing all the pairwise comparisons.

		PB		HRF	
		Static	Dynamic	Static	Dynamic
PB	Static		**0.02**	0.52	0.21
	Dynamic	**0.02**		0.12	0.52
HRF	Static	0.52	0.12		0.52
	Dynamic	0.21	0.52	0.52	

4.6 Conclusions

By employing a unique and proprietary dataset comprising granular firm-level securitization and accounting data on a panel of 810 Italian SMEs over the time period 2015-2017, this paper tests two alternative approaches grounded in statistical learning and machine learning frameworks and compares their respective capability in predicting SME credit risk. Specifically, we compare a classic parametric approach fitting an ordered probit model with a non-parametric one calibrating a machine learning Historical Random Forest (HRF) approach. Both models are implemented according to a static and a dynamic framework. Moreover, we further assess the relevance of each variable to predict SME credit risk, through the use of Shapley values.

Our results provide evidence that the dynamic Historical Random Forest (HRF) approach outperforms the traditional ordered probit model in assessing SME credit risk. This shows that advanced machine learning methodologies can be successfully adopted by banks to predict SME credit risk, highlighting the opportunity to complement traditional methods with more advanced estimation techniques that rely on machine learning.

Our research question represents a matter of concerns to policy makers, since inaccurate credit risk measurement could threaten the stability of the banking sector, undermining the pivotal intermediation role played by banks in the economy. This assumes even greater relevance in light of the current COVID-19 crisis. Indeed, in periods of financial distress, an accurate credit risk assessment would allow banks to better forecast ex-ante corporate default probability.

This paper paves the way for future and unforeseeable research in this area. Future extensions of this work could involve not only applying alternative machine learning methods, but also testing whether the latter could successfully predict and "harden" soft information, thus eventually substituting for the traditional role of relationship banking in small business lending.

Chapter 5

Understanding corporate default using Random Forest: The role of accounting and market information

5.1 Introduction

What are the factors affecting corporate default risk? The aim of banks' core business, as in their intrinsic nature, is to perform accurate assessment of borrowers' capability to repay their debt. This activity is performed by collecting information about a given borrower from different sources. The type of information a bank should use when assessing credit risk has been a matter of concern for policy makers since inaccurate credit risk measurement could threaten the stability of the banking sector and undermine the pivotal intermediation role played by banks in the economy. In this regard, banks' need to implement reliable credit risk models to timely and precisely forecast business failure is imperative to reach appropriate lending decisions and, eventually, to engage in corrective action.

When focusing on the predictions of default risk of micro-, small- and mid-sized enterprises (MSMEs), it is crucial to adopt a credit risk assessment model that takes into account their peculiarities in order to provide a reliable prediction of default. Indeed, MSMEs have specific characteristics which are not similar to those of larger firms on which the existent literature on default prediction modeling has mainly been based (Norden and Weber, 2010, Peel et al., 1986, Hol, 2007). In this regard, MSME lending suffers from more severe agency problems, exhibits higher default risk, has lower accounting quality and is more informationally opaque (Burgstahler et al., 2006). Given the importance of MSMEs for market economies, it is imperative to implement credit assessment models specifically addressed to MSMEs with the objective to minimize expected and unexpected losses as accurately as possible. To this purpose, the objective of this paper is to develop a credit risk model for MSMEs that takes into account, in addition to accounting measures, market information obtained from comparable publicly listed companies. Motivated by the findings of the relevant literature on the assessment and forecasting of corporate default risk, we exploit a unique and proprietary dataset comprising over 10,136 Italian firms and their 113 co-operative banks over the period 2012–2014 to estimate multivariate forecasting models on the incidence of corporate default by using both market and accounting information. Given the unlisted nature of our Italian micro-, small, and mid-sized enterprises (MSMEs), we estimate the Merton's Probability of Default (PD) based on market information obtained from those publicly listed and deemed as comparable by a data-driven clustering approach, avoiding any a-priori assumption of mapping by size, industry and number of employees[1].

[1] Hence, we argue that our modeling approach for the evaluation of the market risk of MSMEs is not prone to estimation or misspecification error. Instead, we argue that this is the only feasible modeling approach for capturing the market risk of firms for which no market data exist.

The paper contributes to the literature along two dimensions. Firstly, our hybrid credit scoring models, which use a combination of market and accounting information, provide better default predictions for unlisted firms when compared with the respective predictive power of models which only use accounting or market information. Although several papers have applied Merton model to private companies (Rikkers and Thibeault, 2009, Andrikopoulos and Khorasgani, 2018, Falkenstein et al., 2000), our study, to the best of our knowledge, represents the first attempt to introduce a new credit risk modeling approach that encompasses market information for predicting corporate defaults of unlisted companies. Indeed, we show that the estimated Merton's Probability of Default (PD) credit risk measure has incremental predictive power on corporate default when added into a multivariate predictive regression model which already includes accounting information. A possible economic interpretation of this finding relies on the fact that the nature of MSMEs as quite risky companies with a lot of time variation and sudden shifts in their risk profile (Islam and Tedford, 2012) leads market information to better capture the corporate default risk of these firms. In fact, market data respond more quickly to new information about borrowers' creditworthiness when compared to more sluggishly-responding accounting measures. In support of this claim, Islam and Tedford (2012) finds that SMEs usually face three types of risks: operational, occupational and economic. The first involves the loss of production and its capability, the second comprises the risks associated with employees' health, safety and well-being, and the third is affected by financial penalties resulting from the first two as well as compensation claims and damage to reputation. Monetary factors and accounting measures alone can ignore many issues that impact the long-term competitive advantage of the company and the reasons that can lead to business demise. Furthermore, there are multiple internal and external causes contributing to failure and none of them seem to dominate in leading the firm to default. The former encompass poor business competencies, high cost pressure, fraud, poor quality of products and services and private domain, whereas the latter involve government policies, natural disaster, global economic downturns and increased industry competition (Kucher et al., 2020). As explained in Islam (2008), all these risks can cause loss of market share and eventually put the organization out of business. Following the advice of Virdi (2005) on the importance for MSMEs of managing risk in a professional and structured way, our hypothesis relies on the importance of market volatility in capturing the effects of the aforementioned risks. In particular, employees' dissatisfaction and strikes are usually reported by media and may affect the firm's reputation causing shocks on their stock prices. Similarly, internal policy, low level of enterprise culture and dubious choices of management board can reduce the stakeholder and shareholder trust resulting in stock downturn. Moreover, shortage of goods and machinery breakdown and national or international government policies can lead to analogous effects. Although the data we use are annual, short-term shocks or rare events can still be captured by asset volatility, regardless of the overall price trend as well. We therefore conclude that the assets volatility can be a reliable proxy for several types of risk and can be accurately mapped through a panel of representative peers spanning over different industrial sectors and firm sizes.

The second dimension involves the implementation of predictive models and their explainability. In the last years, many works on the application of Machine Learning (ML) model to economic problems have been published (Mullainathan and Spiess, 2017, Akbari et al., 2021, Avramov et al., 2021, Olson et al., 2021). Specifically, Kim et al. (2020) report a survey on ML applications to credit default prediction. Generally, linear classification models, such as linear discriminant analysis (LDA) or logistic regression (T., 2001, E.I. and G., 2007, J. and V., 2014, S. et al., 2015), show lower prediction ability rather than non-linear and non-parametric models, such as Random Forest (RF) or Boosted Trees (BT) (Zhu et al., 2019, Barbaglia et al., 2021). However, most of these studies restrict their analysis on the increase of performances compared to linear models and do not investigate the relevance of the input variables and their effects on the predictions. An attempt on the explanation of the overall importance of the input variables can be found in Moscatelli et al. (2019b). Moreover, Barbaglia et al. (2021), Albanesi and Vamossy (2019) show an increasing attention on local explanation of predictions, i.e. understanding how each variable's value can impact the predicted outcome, justified by the need of a more deep investigation of the complex non-linear correlations captured by the advanced models. Our work contributes to this new stream of research (usually referred to eXplainable Artificial Intelligence) because we implement both a non-linear parametric and non-parametric ML algorithms and we go beyond the prediction of corporate defaults and implement an advanced methodology that involves the use of two state-of-the-art techniques so to evaluate the importance of the variables on the predictions: Permutation Feature Importance (Fisher et al., 2018) explains the overall variables' relevance, whereas Shapley Additive Explanations (Lundberg et al., 2020) provide the contribution of each variable's values to the predicted probability of default for a single observations. In addition, we implement a sophisticated clustering technique that, to the best of our knowledge, is the first application of Artificial Neural Networks to compress the information of financial ratios so to map unlisted MSMEs to a pool of listed ones.

The studies of Falkenstein et al. (2000), Rikkers and Thibeault (2009) and Andrikopoulos and Khorasgani (2018) are closely related to our work, providing evidence that hybrid models which incorporate both market and accounting information have significant predictive power on corporate defaults of unlisted firms. Despite our findings are supportive of the results reached in those studies, our study differs in several aspects. Firstly, in contrast with Falkenstein et al. (2000) and Rikkers and Thibeault (2009), we derive the market value of unlisted firms by collecting market data from comparable publicly-listed companies that operate in the same industry matched in terms of size, industry, and number of employees by a data-driven approach ("comparable approach") rather than by KMV's private firm model which makes use of the industry average market value of equity. The latter has the drawback of being exposed to considerable variation over time (Falkenstein et al., 2000) or by the present value methodology of cash flows for valuing a company (Rikkers and Thibeault, 2009, Chen and Liao, 2005, J.F. et al., 2001). The rationale behind choosing a "comparable approach" stems from the benefits highlighted by the extant literature in the field of corporate finance when adopting a comparability method in equity valuation of private unlisted companies using industry-level data (Andrikopoulos and Khorasgani, 2018, Baker and Ruback, 1999, Alford, 1992, McCarthy, 1999). Secondly, differently from Rikkers and Thibeault (2009), we exploit a unique proprietary dataset comprising granular data on a sample of 10,136 Italian unlisted MSMEs operating within 113 Italian co-operative credit banks over the period 2012-2014, rather than by relying on data collected from a single bank. Lastly, our sample comprises Italian unlisted companies operating in the manufacturing sector, as in Andrikopoulos and Khorasgani (2018), and also in the service industry.

One policy implication resulting from our findings is that banks can potentially integrate their hybrid credit scoring methodologies with market information for credit risk assessments, with the purpose of increasing the accuracy of forecasting corporate defaults for unlisted firms. This would allow banks to expand the spectrum of information used in credit risk measurements helping them to enhance their internal hybrid credit scoring by including both accounting and market information on the credit quality of a given borrower. Thus, results reported in this paper could be very helpful for forward-looking financial risk management frameworks (Breden, 2008, Rodriguez Gonzalez et al., 2018).

The remainder of this paper is organized as follows. Section 5.2 briefly reviews the relevant literature. Section 5.3 discusses the data and Section 5.4 presents the econometric methodology. Section 5.5 illustrates the empirical results and Section 5.6 concludes.

5.2 Determinants of corporate default: a review of the literature

Up to date, most of the literature on corporate credit risk modeling has focused on both accounting-based approaches and structural market-based models.

5.2.1 Accounting-based models

The power of accounting-based techniques to predict default risk has been already widely explored by the extant literature. Such methodologies include all the statistical techniques that elaborate quantitative information about a borrower, i.e. financial ratios and statement data, into a numerical score reflecting the credit quality and predictive of the default probability of the borrower itself (Beaver, 1966, Altman, 1968, Ohlson, 1980, Edminster, 1972, Blum, 1974, Grice and Ingram, 2001, Pindado et al., 2008, Louzada et al., 2016a). In this regard, there are several studies investigating the efficacy of accounting-based credit scores to predict future corporate default. Calabrese et al. (2016) show that accounting-based information has significant in-sample and out-of-sample forecasting power on the timing of bankruptcy of Italian MSMEs over the period 2006-2011. Beaver (1966) examines whether financial ratios predict subsequent business bankruptcy. E.I. and G. (2007) analyze the most relevant variables in forecasting the company's future credit quality and construct a default prediction model. Peel et al. (1986) expand the variable set used in forecasting default and find that non-conventional variables computed from UK companies' annual reports and statements can significantly enhance the predictive power of more conventional models. Keasey and Watson (1987) implement a similar model to predict small business failure. Charalambous et al. (2004) apply logistic regressions and neural networks to develop bankruptcy forecasting models and assess the power of cash flow variables to forecast UK corporate default. Lin et al. (2012) test a number of different accounting-based risk models to predict UK small business bankruptcy. Overall, these models seem to suggest that the more relevant accounting variables are added to the model, the better its corporate default forecasting power becomes. These models remain the most widely used methodology for the prediction of default of unlisted companies, although they exhibit some disadvantages. The main disadvantages are the fact that ratios might correlate with each other affecting the estimates, and when using comparative ratio analyses, differences between firms in terms of methods of accounting or methods of operations have to be taken into account (Stickney and Weil, 1997). In this regard, we attempt to overcome those disadvantages by adopting a structural form model approach, i.e., the Merton's model, whose principle is that a firm defaults if the company's asset value falls below the default boundary. Moreover, the models we use can deal with multicollinearity between input variables and thus provide robust estimations.

However, another stream of literature suggests that credit risk models which incorporate both accounting and non-accounting information, i.e. creditors' legal action and audit reports, have a higher predictive power of corporate default risk (Altman et al., 2010). In this regard, Bhimani et al. (2010) provide further support to the findings in Altman et al. (2010) by showing that, for a large sample of privately-owned Portuguese firms, accounting and non-accounting information is a significant predictor of corporate default. Along the same lines, Dierkes et al. (2013), by examining the forecasting performance of credit models on the default risk of privately-owned small firms, find that non-accounting business information improves default forecasting performance when added into the information variable set of default predictive models that use only accounting information. Fiordelisi et al. (2014) find that, especially for small firms, bank-firm relationship information is a significant determinant of corporate default risk (Volk, 2012, Qian et al., 2015). In a similar analysis aimed at assessing corporate default risk of Italian firms, Foglia et al. (1998) show that multiple bank-firm relationships increase the likelihood of default. Moreover, Norden and Weber (2010) find that borrowers' checking account activity and credit line usage are significant early warning signals of default for a sample of German firms. By using loan-level data from a large Chinese bank, Qian et al. (2015) show that delegation of authority to line units increases the predictive power of internal ratings on borrowers' credit risk. Greater accountability to loan officers increases soft information production and affects the effort of the loan officer in evaluating a borrower's credit risk. This increased production of soft information is likely to be hardened into the bank's internal ratings which result in scores having a stronger effect on the terms of loan contracts and predictive power on ex-post loan outcomes (Gropp and Guettler, 2018, Liberti and Mian, 2009, Liberti and Petersen, 2017, Brown et al., 2012b).

5.2.2 Market-based models

In addition to the aforementioned accounting-based models, structural market-based models in the form of Merton-type models have been applied to predict corporate bankruptcy. Specifically, the Merton's Distance to Default (DD) model, which is based on observable equity market data, has been extensively used for the estimation and prediction of corporate default risk for listed firms in the US equity market (Bharath and Shumway, 2008, Byström, 2006, Vassalou and Xing, 2004). Moreover, other studies have compared the corporate default predictive power of accounting- and market-based models. Overall, these empirical studies have extensively shown the incremental predictive information content of the Merton model on corporate default predictions with respect to alternative models primarily based on accounting ratios like the Altman (1968) z-score model (Agarwal and Taffler, 2008, Altman, 1968, J. and V., 2014, Das et al., 2009, Doumpos et al., 2015, Hillegeist et al., 2004). Among these, Das et al. (2009) show that the accounting-based models of corporate default perform similarly to the market-based models of credit risk assessments when it comes to predict the distress risk of an international sample of 2860 quarterly Credit Default Swap (CDS) spreads, while Doumpos et al. (2015), when forecasting the bankruptcy risk of a panel of European listed firms for the period 2002-2012, find that the inclusion of the Merton's market-based Distance-to-Default (DD) measure into the information variable set which is composed only by accounting-based financial ratios, adds significant predictive information content. In a similar vein, Tinoco and Wilson (2013) find that corporate default risk can be more accurately predicted when models simultaneously incorporate accounting, stock-market and macroeconomic information, rather than using different types of information in isolation. However, those studies do not integrate, but rather compare, accounting with market information in forecasting corporate default.

5.3 Data

We use two sources of information for our analysis: a proprietary one, consisting of granular information on over 10,136 Italian unlisted micro-, small, and mid-sized enterprises (MSMEs)s, and a public one, comprising data on comparable publicly listed companies, i.e., peers.

5.3.1 MSME data

We exploit a unique and disaggregated dataset on an unbalanced panel sample of 10,136 firms and 113 cooperative credit banks, for a total of 19,743 firm-year observations over the period 2012–2014. Specifically, we consider firms with less than 250 employees and revenue at most of 50 million. We selected a subset of 22 financial ratios out of 30 removing the ones showing high partial correlation with many other ratios. Therefore, some ratios with not so low correlation with at most one other ratio are still kept because the models we use for the predictions are robust to multicollinearity. Tables 5.1 and E1 in the Appendix report the complete list of variables with description and statistics and their pairwise correlations, respectively. The target variable we want to predict is a binary flag indicating whether the firm defaults (1) or not (0). In our context, the flag of default is assigned when the client becomes insolvent in the last 12 months following loan disbursement, and with a past due of at least 180 days. Moreover, we control for additional categorical variables, describing time-invariant characteristics of our unlisted firms, such as the region to which the firm belongs, firm size and industry. Table 5.2 reports the list of control variables used in the analysis and their distribution over the two target classes.

Table 5.1: List of input variables for MSMEs dataset.

Variable	Description	Mean	St.Dev.	Min	5th perc	Median	95th perc	Max
1 - Oth Reven on Reven	Other revenues on revenues	0.03	0.05	0	0	0.01	0.19	0.19
2 - Deprec on Costs	Depreciation on costs	0.06	0.08	0	0	0.03	0.26	0.34
3 - Pay to Bank on Assets	Payables to banks on current assets	0.83	1.5	0	0	0.47	2.73	11.25
4 - Cashflow on Reven	Cash flow on revenues	0.08	0.08	0.01	0.01	0.06	0.26	0.41
5 - Fixed Asset Cov	Fixed asset coverage	1.15	1.99	0.07	0.07	0.57	4.89	11.17
6 - Labor Cost on Reven	Labor cost on revenues	0.56	0.32	0	0	0.61	1.03	1.03
7 - ST Pay on Due to Bank	Short-term payables on amounts due to banks	2.05	2.46	0.16	0.21	1	9.49	9.49
8 - Tot Debt on ST Debt	Total debt on short-term debts	2.3	2.04	1	1	1.67	5.79	13.35
9 - Tot Debt on Net Worth	Total debt on net worth	7.92	10.2	0.35	0.48	3.73	36.5	41.94
10 - Pay to Suppl on Net Worth	Payables to suppliers on Net worth	2.69	3.48	0.04	0.12	1.01	13.01	13.01
11 - Pay to Suppl on Tot Debt	Payables to suppliers on Total debt	0.4	0.22	0.02	0.07	0.36	0.84	0.84
12 - Inventory Duration	Inventory duration	0.78	1.09	0.02	0.03	0.5	2.68	7.16
13 - Quick Ratio	Quick ratio	1.41	1.1	0.04	0.22	1.18	3.42	6.54
14 - Debt Burden Index	Debt burden index	0.4	0.38	0.01	0.02	0.23	1	1
15 - Fin Int on Reven	Financial inrerest on revenues	0.02	0.02	0	0	0.02	0.08	0.1
16 - Fin Int on Added Val	Financial inrerest on added value	0.08	0.07	0.01	0.01	0.05	0.25	0.25
17 - Net Worth on LT Eqt/Pay	Net worth on long-term equity and payables	0.49	0.31	0.05	0.06	0.48	1	1
18 - Net Worth on NW+Invent	Net worth on net worth and inventories	0.64	0.3	0.07	0.1	0.7	1	1
19 - ROA	Return on Assets	0.02	0.07	-0.1	-0.1	0.01	0.17	0.27
20 - ROD	Return on Debt	0.03	0.01	0	0	0.02	0.05	0.05
21 - Working Cap Turnover	Working capital turnover	2.3	2.25	0.25	0.55	1.77	5.78	18.32
22 - Turnover	Turnover normalized by Total Assets	1.16	0.74	0.07	0.2	1.02	2.84	3.17

Table 5.2: List of control variables for MSMEs dataset.

Variable	Target										
FIRM SIZE		Large	Medium	Micro	Small	TOTAL					
	0	2.4%	9.1%	54.7%	27.1%	93.2%					
	1	0.2%	0.8%	3.8%	2%	6.8%					
	TOTAL	2.6%	9.9%	58.5%	29%						
DUMMY INDUSTRY		Manufacturing	Services	TOTAL							
	0	33.1%	60.1%	93.2%							
	1	2.2%	4.6%	6.8%							
	TOTAL	35.3%	64.7%								
INDUSTRIAL SECTOR		Food & Accommodation	Energy supply	Entertainment	Information & communication	Manufacturing	Professional, scientific and technical	Real estate	Trade	Transportation	TOTAL
	0	4.5%	1.3%	1.4%	4.1%	33.1%	8.7%	6.9%	29.2%	4%	93.2%
	1	0.6%	0.1%	0.1%	0.2%	2.2%	0.6%	0.7%	2.1%	0.3%	6.8%
	TOTAL	5%	1.4%	1.4%	4.4%	35.3%	9.3%	7.6%	31.3%	4.3%	
REGION		Central	Islands	North-east	North-west	South	TOTAL				
	0	1.4%	2.3%	52%	26.2%	11.4%	93.2%				
	1	0.1%	0.3%	3.6%	1.8%	0.9%	6.8%				
	TOTAL	1.5%	2.6%	55.6%	28%	12.3%					
FIRM TYPE		Enterprises	SEO	Small business	TOTAL						
	0	75.7%	3.4%	14.1%	93.2%						
	1	5.9%	0.1%	0.8%	6.8%						
	TOTAL	81.6%	3.5%	14.9%							

5.3.2 Peers data

We select a panel of 40 Italian listed firms, evenly distributed in manufacturing and services sector. We collect accounting figures from Orbis database, developed by Bureau Van Dijk (a Moody's analytics company), by matching the VAT code for each given peer firm[2]. The accounting figures are used to reconstruct and match or proxy the 22 financial ratios of the

[2]The database construction process played a crucial role in making such an empirical analysis possible, despite being time-consuming due to the required manual input of proprietary micro-level data, properly integrated with additional accounting data collected from Orbis database.

MSMEs dataset. Moreover, daily stock prices are collected from Refinitiv Eikon database and are used to compute the annual assets volatility of comparable publicly-listed companies. Table E2 in the Appendix reports the statistics for the 22 variables as well as for the volatility, total assets and total liabilities used as inputs in the Merton's model formula, as described in Section 5.4.1. Figure E1 in the Appendix depicts the comparison of the 22 variables between the two datasets, showing that the selected peers are adequately representative of our sample of unlisted micro-, small, and mid-sized enterprises (MSMEs).

5.4 Methodology

The aim of this paper is to assess the impact of market information, i.e., the Merton's probability of default (PD), in predicting corporate default risk of unlisted firms, in addition to accounting-based measures. Our analysis can be summarized into three steps. Firstly, we match each MSME to one or a group of peers and evaluate its firm-wise PD. Section 5.4.1 recalls how the PD is evaluated following the Merton's model and Section 5.4.2 describes the peers-to-firm matching procedure, consisting of a low dimensional representation of the 22 variables space and its subsequent clustering. Secondly, we predict corporate default by calibrating different classification models, both using financial variables as predictors (baseline) and including the PD (extended). Section 5.4.3 shows the calibration of the models and the differences of models' performance between the baseline and extended cases. Lastly, we investigate which predictor contributed the most to predict corporate default, by the means of feature importance techniques. Section 5.4.4 reports the estimation of the contribution of each variable to the predicted class (default or non-default) for both the baseline and extended cases.

5.4.1 Estimation of the Merton model

We estimate the Merton model (Merton, 1974) of corporate default risk for our sample of MSMEs. According to the Merton model, the corporate default takes place when the company is unable to pay off its debts, or when the current market of assets falls below the market value of liabilities. For this reason, the market value of equity of the MSME is treated as a call option on the asset value of the MSME with strike price equal to the market value of debt[3]. The MSME asset value process follows a Geometric Brownian motion as shown in Equation (5.1) below:

$$dA_t = rA_t dt + \sigma_A A_t dz \tag{5.1}$$

where A_t is the firms market value of assets and σ_A is the volatility of assets. r is one-year maturity risk-free rate of return, which we choose to be the yield of the 1-year maturity domestic government bond with 1-year maturity[4]. Since the market value of equity is treated as a call option, the company's equity value at maturity (which is the end of each yearly period in our model), the company's equity E_t at maturity (at the end of each yearly period in our model) is priced as shown below:

$$E_t = rA_t \Phi(d_1) - Le^{-rT}\Phi(d_2) \tag{5.2}$$

where A_t is firm's assets and L is firm's liabilities (which are assumed to be constant for each yearly period). T is the time to maturity which in our model is equal to one year equal to 1

[3]For modeling issues, it is assumed that the market value of debt (or liabilities) is equal to the book value (or accounting value) of total liabilities of the MSME. Moreover, the market value of debt (liabilities) is assumed to remain constant during each yearly period.

[4]We obtain yearly time series data for the 1-year domestic government bond yield for the time period covering 2009 to 2014. The yearly Italian government bond yield data are downloaded from Thomson Reuters.

year ($T = 1$), r is the risk-free interest rate with one-year maturity (the 1-year government bond rate) and Φ is the cumulative standard normal distribution function. Since default is treated as a European call option, then the values d_1 and d_2 are given by the following formulas:

$$d_1 = \frac{\ln A_0/L + (r + \sigma_A^2/2)T}{\sigma_A \sqrt{T}} \tag{5.3}$$

$$d_2 = d_1 - \sigma_A \sqrt{T} \tag{5.4}$$

According to the assumptions of the model, the value of the firm's equity is a function of the value of firm's assets and time, so, it follows from Ito's lemma that:

$$\sigma_E = \frac{A}{E}\left(\frac{dE}{dA}\right)\sigma_A \tag{5.5}$$

where σ_A is the volatility of assets and σ_E the volatility of firms' equity value. Solving Equations (5.3) to (5.5) allows to evaluate A and σ_A which are the inputs for the calculation of the Distance to Default (DD) measure, given in Equation (5.6):

$$DD = \frac{\ln A_0 + (r + \sigma_A^2/2)T - \ln L}{\sigma_A \sqrt{T}} \tag{5.6}$$

The resulting Probability of Default (PD) is given in Equation (5.7) below:

$$PD = \Phi(-DD) \tag{5.7}$$

where DD is the Distance to Default measure given in Equation (5.6).

5.4.2 Matching unlisted firms with peers

Since there are no market data available for our sample of unlisted MSMEs, we proxy the market value of equity of unlisted MSMEs with those of their comparable publicly-listed companies. As for the latter, the market value of equity is computed as the daily product of their share price multiplied by the number of shares outstanding. Our implicit assumption made for the estimation of the Merton's Probability of Default (PD) and Distance-to-Default (DD) is that those MSMEs which have similar size, number of employees, and industry sector with our Italian peers share the same risk profile and belong to the same (market) risk class of the latter[5]. In order to render the matching procedure as accurate as possible, we opt for a clustering approach: we find the optimal number of clusters in the MSME dataset and then we assign each peer to the most similar cluster by minimizing the average distance from all firms in the cluster.

Given that the high number of variables of the MSME dataset can affect the clustering algorithm, we apply several dimensionality reduction techniques to have a condensed representation of the original data, hereinafter referred to "embedding". The main idea is to find a function $f : \mathbb{R}^n \mapsto \mathbb{R}^k$, with $k \ll n$, that can project the original high-dimensional data $X \in \mathbb{R}^n$, e.g. $n = 22$ for MSMEs, into a low-dimensional one $E = f(X) \in \mathbb{R}^k$ trying to preserve the mutual distance between points, both locally and globally (Gracia et al., 2014). The embedding E has the same number of observations of X but less variables. The inverse of f, $f^{-1} : \mathbb{R}^k \mapsto \mathbb{R}^n$ can be used to project back the embedding E so to get the reconstruction $\hat{X}$

[5]Our assumption on the (market) risk classes goes back to the Modigliani and Miller (1958) risk class assumption according to which firms with similar characteristics and balance sheet data belong to the same 'risk class'.

of the original X and the reconstruction error (RE) can be defined as follows:

$$RE = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij} - \hat{X}_{ij})^2$$

where N is the total number of observations, and X_{ij} and $\hat{X}_{ij}$ are the i-th row and j-th column elements of X and $\hat{X}$, respectively. The reconstruction error decreases as k increases and can be used to find the optimal value of k, as a trade-off of keeping both RE and k small enough.

Given the panel structure of our MSME data, we evaluate the embedding both at the firm-year level, getting different low-dimensional coordinates for each firm-year pair, and at the firm level, getting a shared low-dimensional coordinate for each firm-year pair. The former approach evaluates a time-variant embedding whereas the latter estimates an "average" embedding on the trend of each firm over the years. Thus, in the former we have 22 variables, in the latter we have $22x3 = 66$ variables, as we reshape the dataset so to have variables-year pairs as new input variables. We tested three different dimensionality reduction techniques on both type of dataset: Robust Principal Component Analysis (RobPCA) (Candes et al., 2009), Auto-Encoder with Multilayer Perceptron (AE) (Kramer, 1991) and Auto-Encoder with Long-Short Term Memory (AE-LSTM) (Cho et al., 2014). RobPCA builds the embedding by creating a linear combination of the original variables, where each combination is the new coordinate. AE is a particular architecture of Artificial Neural Network that mixes linear combination of variable with their non-linear transformations so to overcome the linearity limitation of RobPCA. AE-LSTM is an extension of AE that includes an auto-regressive terms in order to take into account time dependence of variables. In this way, AE-LSTM treats each batch of observations of the same firm over the years as a single input.

Being the embedding evaluated, we perform the clustering of the data using the k-means algorithm (MacQueen, 1967) with the Euclidean distance. We test different number of clusters and we find the optimal value based on Davies-Bouldin index (Davies and Bouldin, 1979) and Silhouette coefficient (Rousseeuw, 1987). The higher the former and the closer to 1 the latter, the better is the clustering. After selecting the best clustering, we apply the embedding function on peers' dataset so to have the same low-dimensional representation and to assign each peer to the closest cluster. The closeness is intended as the minimum average euclidean distance from all observation in the cluster.

Although the lower dimension (six) of the embedding resulted in a better performance for the clustering, data cannot be visualized. Therefore, we make use of another dimensionality reduction technique that is better suitable for visualization rather than for clustering. Thus, applying the Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018, McInnes et al., 2018) we can visualize the clusters into a 3-dimensional space.

Finally, we provide each MSME observation with its respective PD. As described in Section 5.4.1, PD can be calculated with Equation (5.7) after evaluating DD with Equation (5.6), using the total assets A, the total liabilities L and the assets volatility σ_A. We evaluate the PD with two approaches. In the first (named *average-PD*), with evaluate the average $\bar{A}$, $\bar{L}$ and $\bar{\sigma_A}$ over all peers in the same cluster for each year and use them to evaluate the average DD. So, we have $kx3$ different DD values, one for each year-cluster pair, where k is the optimal number of clusters. The DD is then matched with each MSME observation by year-cluster. In the second approach (named *pointwise-PD*), we use each MSME observation's A and L and the average year-cluster peers $\bar{\sigma_A}$ to have a firm-year level DD.

5.4.3 Prediction of default

After assigning the PD to all our unlisted MSMEs, we calibrate three different models to predict the binary target, (1) for defaulted firm and (0) otherwise. Each model is calibrated with the set of 22 variables (*baseline*) and with the addition of the PD (*extended*). First, we inspect the distribution of each input variable with respect to the target variable. Figure E2 in the Appendix shows similar behavior of the input variables for both subset of defaulted and non-defaulted firms, meaning that the overall relation between each predictor and the target is weak because there is no clear polarization in the distributions. Thus, we expect low prediction performances when using classical linear models because they estimate coefficients that should discriminate between the 0s and the 1s in the entire distribution of input variables. Moreover, the true relationship between input and target variable may be non-linear. Therefore, we opt for a non-linear and piecewise model, the Multivariate Adaptive Regression Spline (MARS) (Friedman et al., 2009), that estimates multiple polynomial relationships in different partition intervals of each input variable. So, the model can be seen as an ensemble of sub-models that are estimated in each combination of partitions in which input variables can be divided. For example, if we split the input domain into quartiles of each variable, MARS estimates a polynomial function for observations whose input variables are in the lowest quartile of the corresponding distributions, and so on for all possible variable-quartile combinations. As MARS is a parametric algorithm, meaning that we have to define a structure of each estimation function, e.g. polynomial, we test also a non-parametric model, the Random Forest (RF) (Breiman, 2001). RF is an ensemble of decision trees that partition the input domain with nested binary splitting aiming to maximise the discrimination of all target values. Each branch of the tree contains a set of hierarchical rules, e.g. values of a certain variable greater or less than a fixed threshold, so that (possibly) all observations satisfying each chain of rules have the same target value, i.e. 0 or 1. The estimation function of RF is then a combination of rules that can approximate non-linear relationships between input and target variables. Nonetheless, we use a regularized linear model, i.e. Elastic-Net, as a benchmark. As noticed in Section 5.3, the presence of few variables with moderate correlation won't affect the models' performances because the ensemble nature of MARS and RF and the regularization feature of Elastic-Net are suitable to deal with multicollinearity. Each model has a set of hyper-parameters that must be defined before the calibration. For example, MARS requires the maximum degree of polynomials to be fitted, RF requires the number of decision trees to be estimated. We find the optimal value of hyper-parameters by the means of Bayesian Optimization with a 5-fold Stratified Cross-Validation[6] performance estimation. Given the imbalanced nature of the data, as described in Section 5.3, we use the F1-score as a class-specific performance metric, so to highlight the importance of predicting the rarest 1-labelled targets, and the Area Under the Receiver Operating Characteristic Curve (AUC) as an overall performance metric. Moreover, each model has been calibrated with the additional constraint of weights for each observation, i.e. penalizing the prediction error on 1s more than the error on 0s. Both F1-score and weighting help the calibration procedure to prevent overfitting to a certain extent, allowing the model to have a good generalization power[7]. Furthermore, we include control variables in both *baseline* and *extended* case in order to assess models' robustness to time-invariant characteristics of the observations. Finally, we investigate the persistence of target values over time, i.e. we examine the impact of clients that changed their outcome over the years, both from defaulted to recovered and vice-versa. Table E3 in the Appendix reports the number of clients that changed over time. In order to assess the impact of this phenomenon, we compare the distribution of the input variables subject to clients' behavior, and we calibrate the models both on the entire dataset and on the dataset where we remove the clients that changed the outcome over the years. We find that models' performances are not affected by the inclusion of target-switching clients, resulting in the robustness of our results to this phenomenon. Figure E3 in the Appendix

[6] Stratification is performed with respect to both target variable and control variables, when included

[7] This means that the model has similar performances on both data used for calibration and new observations

shows the distribution of relative changes over the years of each input variable splitted by clients' behavior, i.e., clients that are persistent over time and clients that do not exhibit such a behavior.

5.4.4 Importance of variables

We explore which input variable contributed most in each model predictions, focusing on the changes when the PD is added. For this reason, we evaluate the predictive power of the variables using two state-of-the-art techniques for feature importance: Permutation Feature Importance (PFI) and Shapley Additive Explanations (SHAP). PFI evaluates the importance of the i-th variable by comparing the performance, e.g. F1-score, of the model that predicts the observations used for the calibration against the performance of the model that predicts the same observations where the values of the i-th column are shuffled (Fisher et al., 2018). In this way the correlation between the i-th variable and all the others is broken thus removing the influence of that variable on the model predictions. If the change in performance is negligible, the i-th variable is not important for the model. SHAP is based on Shapley values, a method from coalitional game theory which provides a way to fairly distribute the payout among the players by computing average marginal contribution of each player across all possible coalitions (Shapley, 1953, Osborne and Rubinstein, 1994). SHAP, proposed by Lundberg et al. (2020), uses Shapley values to evaluate the difference of the predicted value of a single observation, comparing the prediction of all possible combinations of variables that include the i-th variable against the ones that don't. The differences are then averaged and the positive or negative change in the prediction is used as variable importance. For example, if the model predicts the probability of default, SHAP evaluates, for a single observation, which variable contributed most in increasing or decreasing the final probability. In this way, exploiting the additive property of Shapley values, it is possible to estimate the impact of all variables on the final predicted value, for every single observation. PFI provides a global measure of importance, measuring the impact of all observations together. Moreover, it measures the changes of a global performance. SHAP, on the other way, provides a local measure of importance, measuring the impact of variables for every single observation. However, taking the average of the absolute values of each observation's SHAP, it is still possible the get a global measure of the average importance of the variables. Instead, taking the average of the Shapley values rather than their absolute value, provides an average effect of each variable on the predictions. E.3 illustrates both techniques in details.

5.5 Results

As described in Section 5.4.2, we firstly find the embedding that minimizes the reconstruction error. Table 5.3 reports the optimal dimension k, the reconstruction error of the different algorithms and the R^2. In our context R^2 means the ratio of the amount of variance explained by the embedding over the total variance contained in the original data and represents a proxy of how much intrinsic information within the data is preserved in the transformation. The embedding resulting from AE with the firm-year level approach performed best showing the lowest reconstruction error and the highest R^2. Methods evaluated with firm level approach performed worst and won't be included in the following analysis.

Then, we look for the optimal number of clusters. Table 5.4 reports the performance of the clustering on each embedding as well as the comparison with the clusters found on the original high-dimensional data. We select five clusters identified on the AE embedding. Moreover, we apply the UMAP algorithm to visualize the clusters into a 3-dimensional space. Figure 5.1 depicts the five clusters for all observations (small points) as well as the matched peers (bold spheres), showing a good separation, even if there is small overlapping between

Table 5.3: Results of dimensionality reduction. Reconstruction Error and its proportion with the average absolute value of the input data is reported for all methods as well as R^2.

Input level	Rows	Columns	Method	Input Dimension	Embedding Dimension	Reconstruction Error (% of Avg Abs Input)	$R2^2$
Firm-year	Firm-year pairs	Variables	AE	19,743 x 22	19,743 x 6	0.1418 (20%)	98%
			RobPCA	19,743 x 22	19,743 x 9	0.2033 (30.6%)	95.70%
Firm (batch of years)	Firms	Variables	AE-LSTM	10,136 x 22	10,136 x 10	0.2138 (31.8%)	94.60%
Firm	Firms	Variables-year pairs	AE	10,136 x 66	10,136 x 32	0.2391 (35.9%)	91.30%
			RobPCA	10,136 x 66	10,136 x 15	0.3857 (58%)	84.80%

Table 5.4: Results of clustering. Davies-Bouldin index and Silhouette coefficient are reported for clusters evaluated on both embedding and original data. Davies-Bouldin is a positive number, the smaller the better the separation between the clusters. Silhouette coefficient is bounded between -1 and 1, where -1 means overlapping of clusters and 1 perfect separation. Only top two results for each method are reported.

		Dimension		Davies-Bouldin		Silhouette	
Method	Clusters	Original	Embedding	Original	Embedding	Original	Embedding
AE	5	22	6	0.36	0.1	0.45	0.84
	4			0.43	0.13	0.37	0.55
RobPCA	4	22	9	0.57	0.15	0.39	0.69
	3			0.69	0.22	0.3	0.41
AE-LSTM	3	22	10	0.8	0.23	0.07	0.17
	2			0.93	0.37	-0.13	0.02

the yellow and green cluster and few blue peers are mapped close to the red ones. On the contrary, Figure E4 in the Appendix shows the UMAP projection of original data clustered with the optimal number of clusters found for the embedding, i.e. five., where the clusters are clearly overlapping and most of the peers are misplaced. It is worth pointing out that the figure doesn't show the real distribution of the data, that actually lies in a 22-dimensional space, and the clustering has been performed onto the 6-dimensional space of the AE embedding. Therefore, the distances between the plotted points cannot be used in a proper clustering algorithm. Finally, the clusters are used to assign the PD to each client matching year-cluster pairs, choosing among a total of $5x3 = 15$ different values.

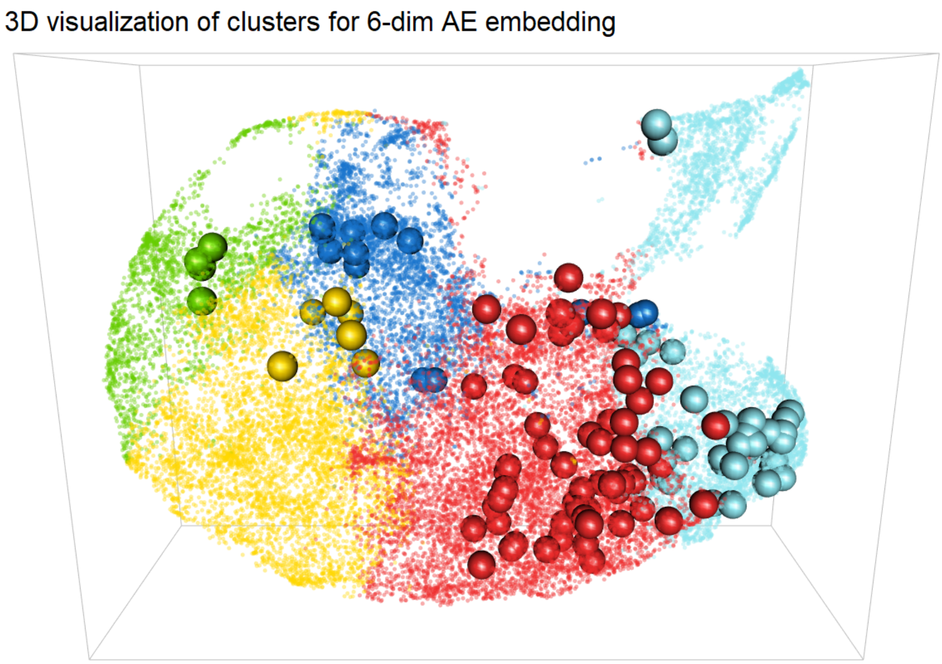


Figure 5.1: 3D visualization of five clusters for the 6-dimensional AE embedding. Visual embedding is evaluated with UMAP algorithm. Small points are MSMEs observations, bold spheres are peers' observations.

Being the PD assigned, we calibrate the prediction models. The following results refer to the PDs evaluated with the *pointwise-PD* approach described in Section 5.4.2 because it performed better than the *average-PD* one, although the findings described below still hold robust. Figure 5.2 shows the distribution of PDs compared with the corresponding target values. PD seems to be a reliable indicator for the outcome of the target variable.

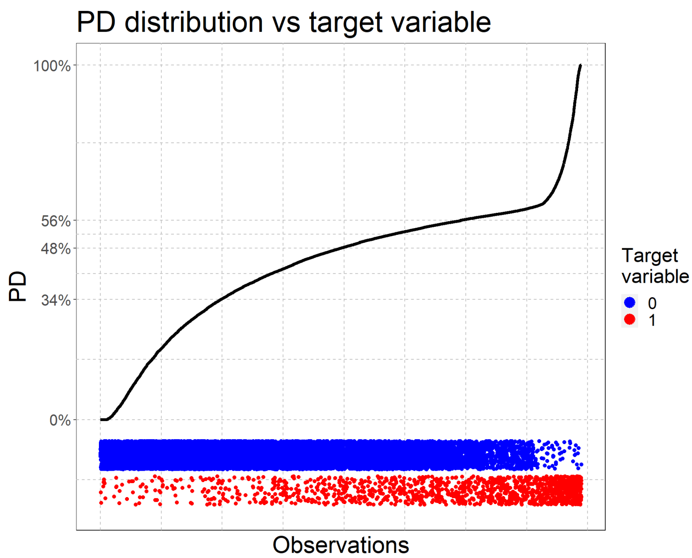


Figure 5.2: Distribution of PDs compared with the corresponding target values. y-axis reports quartiles of PD values.

We tune the parameters of each model with the Stratified Cross-Validation and we calibrate the models with the optimal parameters on the entire dataset, so to have a single model[8] to be used for feature importance evaluation. In Table 5.5 we report performance on cross-validation folds and on the entire dataset for each model as well a comparison between the models trained with the input variables only and the ones with the addition of PD. Random Forest is the only model with good performances, being able to capture the different local separation of the data, as discussed in Section 5.4.3. Nevertheless, all models show an improvement on class-specific performance, i.e. F1-score for the defaulted class, and on the AUC when the PD is included as predictor. Table E4 in the Appendix reports the results of the models with controls for fixed effects, showing stability of performances and a resulting robustness of the models. Finally, Figures from E5 to E10 in the Appendix show the comparison of F1-score and ROC curves of all models and fixed effects.

Table 5.5: F1-score and AUC for Elastic-Net, MARS and Random Forest calibrated on dataset with input variables only and with the addition of PD. Values refer to performance of model calibrated on the entire dataset. Values in parenthesis refer to average performance of validation folds of Cross-Validation.

	F1 (Cross-Val)		AUC (Cross-Val)	
Algorithm	Baseline	With PD	Baseline	With PD
Elastic-Net	30.7% (30.1±1.7%)	35.1% (35.1±1.5%)	79.8% (79.6±0.6%)	82% (81.7±0.8%)
MARS	36% (33.8±1.4%)	40% (37.5±0.6%)	82.5% (81.7±0.6%)	84.2% (82.8±0.8%)
Random Forest	89.5% (85.1±1.7%)	95.8% (91.4±1.2%)	89.8% (85.4±1.1%)	96.1% (91.7±0.7%)

Finally, we explore the feature importance for all models. PFI and SHAP are evaluated on model calibrated with input variables and with the addition of PD. Figure 5.3 shows the PFI of Random Forest model, where the changes of F1-score are normalized. PD is the second most important variable, slightly below the financial interest on revenues. Figures 5.4a and 5.4b show the effect of input variables on the predicted probabilities[9] of Random Forest model, for each observation predicted as 1 and 0, respectively. The color of the points ranges from red, meaning that the observation has low value for the specific variable, to blue, meaning high values for the same variable. The position on the horizontal axis represents the contribution of the variable in increasing or decreasing the predicted probability of each observation. Values on the left column reports the average absolute change in predicted probability over all observations and the normalized values, in parenthesis. PD is on the top two most important variables and we can check the expected impact on the predicted probability: for defaulted observations, high values of PD (blue) result in a major increase of probability, whereas for non-defaulted observations low values of PD (red) result in a major decrease of probability. The accounting variables, as well as the PD, exhibit the expected effect on the predicted probability, e.g. lower return on assets (ROA) and working capital turnover increase the predicted probability whereas lower financial interests decrease the latter. Figures 5.5a and 5.5b show the average signed effect of input variables on the predicted probabilities for all observations predicted as 1 and 0, respectively. In both cases, PD is on

[8] In the k-fold Cross-Validation, k models are calibrated on $k-1$ fold and the performances on the k-th fold are then averaged.

[9] All three classification models predict probabilities in $[0,1]$. If the probability is above 0.5, the observation is classified as defaulted (1), non defaulted (0) otherwise.

the top two most important variables and increases the predicted probability for defaulted observations while reducing the probability for non-defaulted observations. We see that PFI and SHAP agree on the importance of PD, supporting its added value already measured with the increase of performance of the models. Although both techniques lead to the same conclusion, it is worth noting the complementary contribution to model interpretability: PFI provides a synthetic cumulative measure of the relative importance of the variables, whereas SHAP provides insights on the magnitude and the direction of the effect of the variables on each observation, similarly to the explanation of linear regression coefficients. Figures from E11 to E16 in the Appendix report the PFI and SHAP variable importance for Elastic-Net and MARS models, leading to similar results, supporting the relevance of the addition of PD as a predictor.

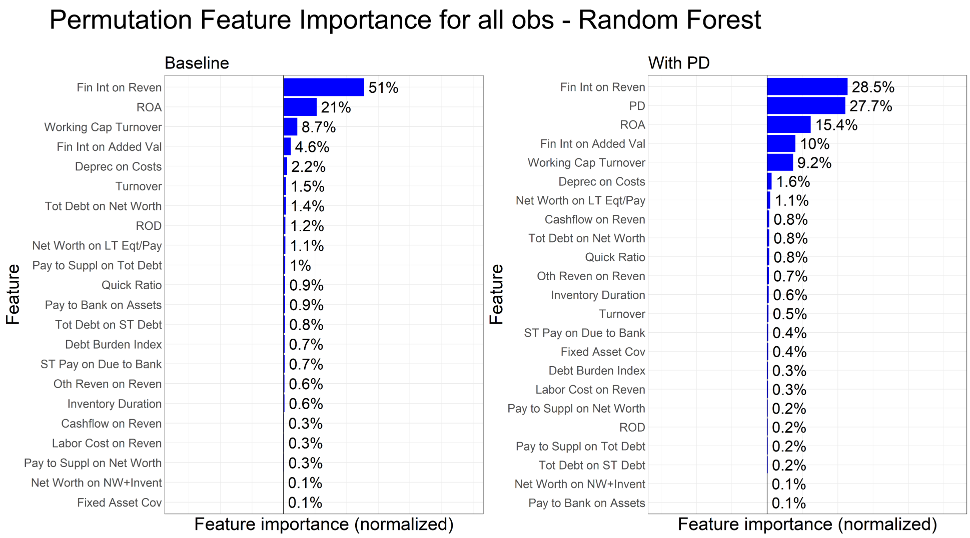


Figure 5.3: Permutation Feature Importance for Random Forest model, comparing variable importance of model calibrated with input variables and with the addition of PD. Normalized changes of F1-score are used to rank the variables.

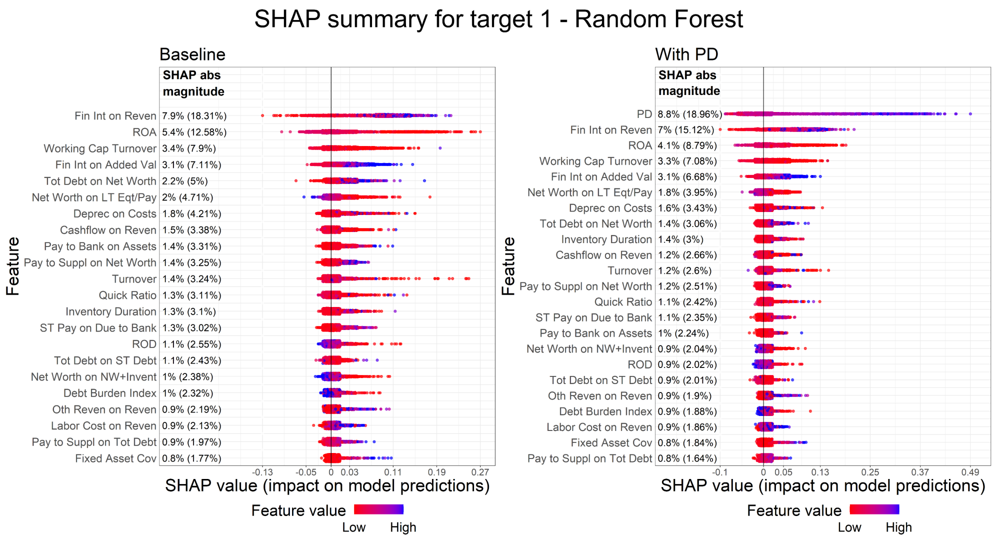
5.6 Conclusions

By exploiting a unique and proprietary dataset on a sample of 10,136 Italian micro-, small, and mid-sized enterprises (MSMEs) operating with 113 cooperative banks over the period 2012–2014, this paper investigates the role of market information in predicting corporate default for unlisted firms. The status of bank's clients is predicted by the means of three empirical models, i.e., logistic Elastic-Net, Multivariate Adaptive Regression Spline (MARS), and Random Forest (RF). We calculate the proxy of market-based Merton's PD credit risk measure by using market data of comparable publicly-listed companies to proxy for the asset price volatility of our unlisted firms. Specifically, the matching procedure between unlisted and their comparable publicly-listed firms is implemented by the means of dimensionality reduction and clustering technique. Moreover, we further evaluate each variable importance in predicting corporate default through the use of Shapley values.

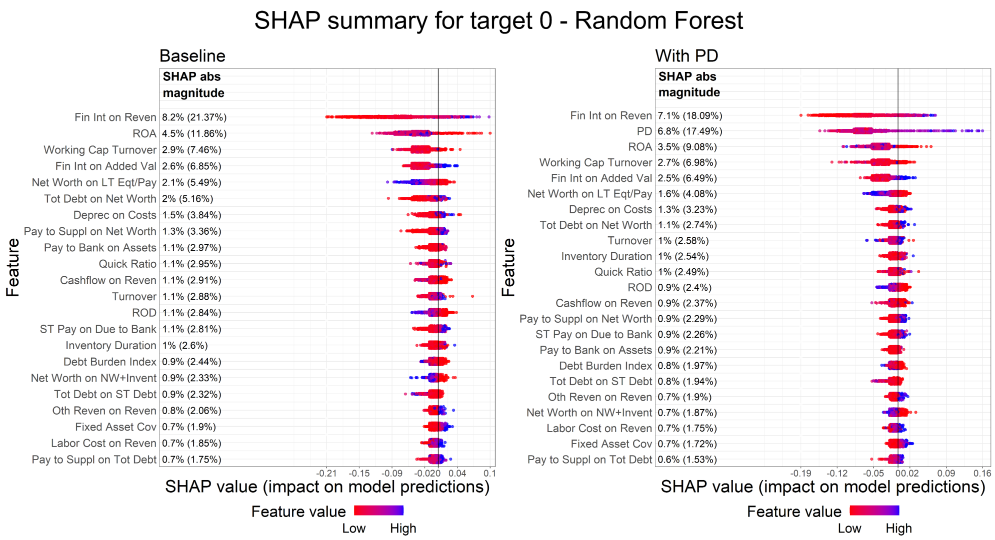
Our results provide novel evidence that market information represents a crucial indicator in predicting corporate default of unlisted firms. Indeed, we show a significant improvement of the model performance, both on class-specific (F1-score for defaulted class) and overall metrics (Area Under the Curve) when using market information in credit risk assessment, in addition to accounting information. Moreover, by taking advantage of global and local variable importance techniques we prove that the increase in performance is effectively attributable to market information, highlighting its relevant effect in predicting corporate default.

Our study makes important inferences for policy implications. Indeed, our findings shed new light on the opportunity for banks to potentially integrate their hybrid credit scoring methodologies with market information for credit risk assessments, with the purpose of increasing the accuracy of forecasting corporate defaults for unlisted firms. Thus, the results of this paper could be very helpful for forward-looking financial risk management frameworks (Breden, 2008, Rodriguez Gonzalez et al., 2018).

Future extensions stemming from this work could involve not only applying alternative prediction models so to provide further evidence on the importance of market information to predict corporate default of unlisted firms, but also testing the impact of synthetic information extracted with the dimensionality reduction technique when replacing the original financial ratios. Testing different clustering technique and exploring the distribution of the clusters could also lead to new insights on clients' behavior and their connections with the market, through the mapping with their publicly-listed peers.

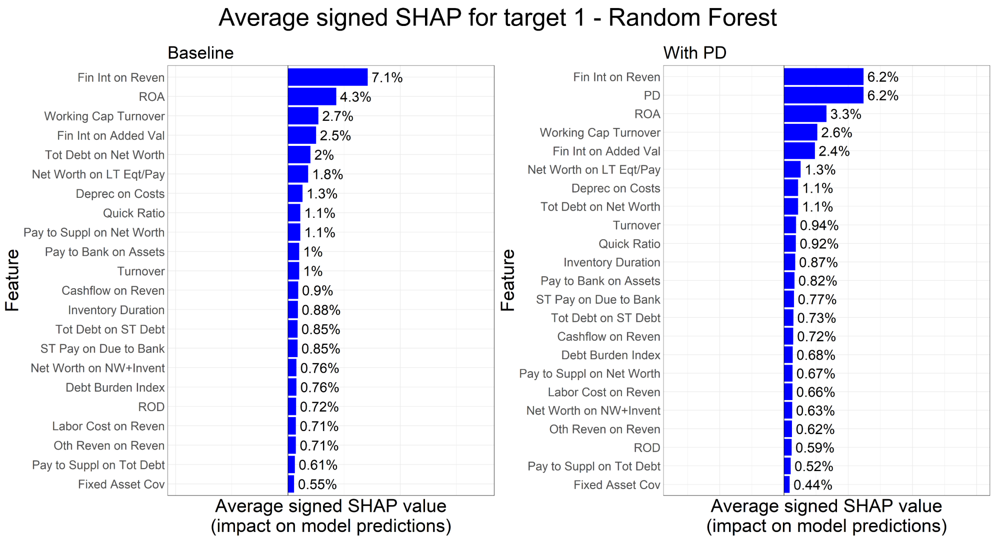


(a) Defaulted clients.

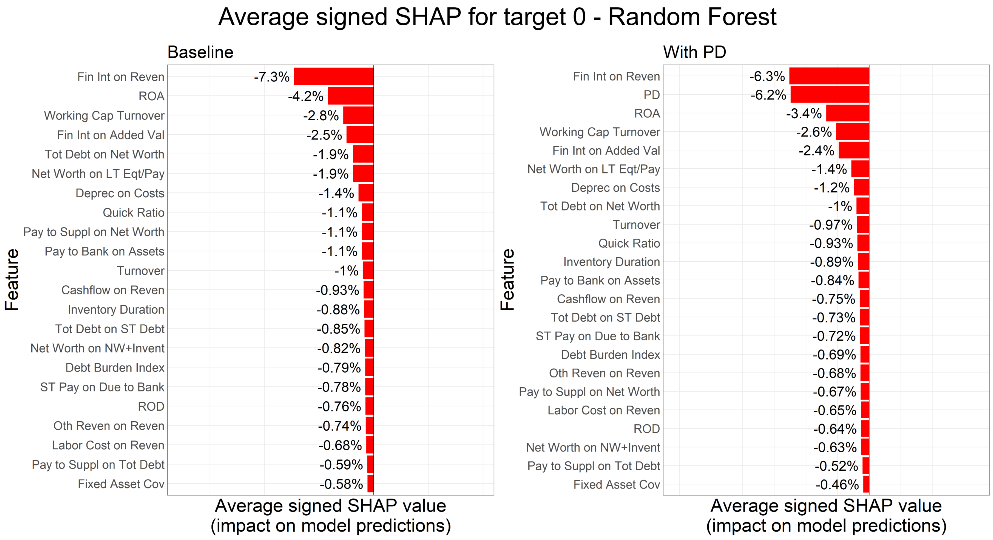


(b) Non-defaulted clients.

Figure 5.4: SHAP effects on predicted probability for Random Forest model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. The color of the points ranges from red, meaning that the observation has low value for the specific variable, to blue, meaning high values for the same variable. The position on the horizontal axis represents the contribution of the variable in increasing or decreasing the predicted probability of each observation. Values on the left column reports the average absolute change in predicted probability over all observations and the normalized values, in parenthesis.



(a) Defaulted clients.



(b) Non-defaulted clients.

Figure 5.5: SHAP average signed effect for Random Forest model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. Bars report the average effect of input variables on the predicted probabilities for all observations predicted as 1 and 0, respectively.

Chapter 6

Concluding remarks

The motivation of this thesis is grounded on the growing attention on explainability in Artificial Intelligence applications. In this context, the understanding of how these so called "black-box" models make their decision has a crucial role. Apart from legal and ethical issues, exploring the interpretation of the drivers of the predictions can increase the performances and give useful insights on which features should be leveraged. Another key aspect of model interpretability is the compression of information. This affects the model performance because, although modern algorithms can handle even thousands of different variables, the selection or the summarising into smaller set of variables make the computations more efficient and robust. Moreover, having a synthetic set of few variables that captures the behaviour and the relationships of many more variables can be an effective tool for XAI.

In Chapter 1, we address the challenging issue of assessing the overall soundness of financial institutions in a country and thereby contributing to the debate on the understanding of global financial stability. We apply a dimensionality-reduction model based on alternative Robust PCA algorithms to produce different versions of our index. We subsequently validate the index using it as a key regressor to predict macro-economic variables. 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 in detecting economic and financial crisis periods. 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.

In Chapter 2, we tested the predictive power of Financial Soundness Index (FSI) built in our previous work in determining the relationship with constraints for financial access. Data consisted of 76 developing countries spanning from 2010 to 2018 with annual frequency. We fitted an ordered probit model with ease in accessing to financial funding (from 1 - easiest to 4 - hardest) as target variable and macro-economic indicators as predictors. We included fixed effect terms for year, country and industrial sector in order to account for all the source of variability. Finally, we conducted an extensive and robust analysis in order to assess the research hypothesis under investigation: our financial soundness measure is a predictor of the access to finance regardless any considered control and instrumental variables. Results show that our financial soundness index is a negative and significant predictor of the firm's financing constraints across countries. It appears that financial stability considerations are relatively more important in affecting the financing constraints of small-size firms in less developed countries. While the analysis needs to be further extended and tested in different data samples and settings, it emerges that financial stability considerations and the associated macroprudential policies are important interventions for improving firms' access to finance, especially of smaller firms in less developed countries.

In Chapter 3, we complement previous efforts at assessing country preparedness to epidemiological risk by proposing a methodological framework that makes the assessment of preparedness more policy-driven and expanded around the world. We apply both principal component analysis (PCA) and dynamic factor model (DFM) to deal with the presence of strong cross-section dependence in the data. The index could provide the basis for developing risk assessments of epidemiological risk contagion after the outbreak of an epidemic but also for ongoing monitoring of its spread and social and economic effects. It would also allow for useful comparisons in country preparedness and performance. This index could be used by organisations to assess likely economic consequences of epidemics and could therefore have managerial implications. We aim to improve the results by increasing the data series availability mindful of the missing data problem using more advanced techniques, also including new relevant dimensions that may capture other aspects of epidemiological risk. As research on the sources and spread of Covid-19 continues, new information is being revealed, which might inform the fined tuned construction of an advanced version of our index. Finally, in future works we should apply alternative data dimensionality reduction techniques and compare the predictive results. The extensive check on the index's predictive power remains to be accomplished by applying it to diverse real-world situations.

In Chapter 4, we use a dataset of granular firm-level securitization and accounting data on a panel of Italian SMEs and we test two alternative approaches grounded in statistical learning and machine learning frameworks, comparing their respective capability in predicting SME credit risk. We further assess the relevance of each variable to affect models' predictions, through the use of Shapley values. Results provide evidence that the dynamic Historical Random Forest (HRF) approach outperforms the traditional ordered probit model, showing that advanced machine learning methodologies can be successfully adopted by banks to predict SME credit risk. This highlights the opportunity to complement traditional methods with more advanced estimation techniques that rely on machine learning. Our research question represents a matter of concerns to policy makers, since inaccurate credit risk measurement could threaten the stability of the banking sector, undermining the pivotal intermediation role played by banks in the economy. This assumes even greater relevance in light of the current COVID-19 crisis. Future extensions of this work could involve not only applying alternative machine learning methods, but also testing whether the latter could successfully predict and "harden" soft information, thus eventually substituting for the traditional role of relationship banking in small business lending.

In Chapter 5, we use a sample of Italian micro, small, and mid-sized enterprises (MSMEs) and we investigate the role of market information in predicting corporate default for unlisted firms. We calibrate three models to predict the status of bank's clients, i.e., linear models, Multivariate Adaptive Regression Spline (MARS), and Random Forest (RF). We calculate the proxy of market-based Merton's PD credit risk measure by using market data of comparable publicly-listed companies to proxy for the asset price volatility of our unlisted firms. Specifically, the matching procedure between unlisted and their comparable publicly-listed firms is implemented by the means of dimensionality reduction and clustering technique. Moreover, we further evaluate each variable importance in predicting corporate default through the use of Shapley values. Results provide novel evidence that market information represents a crucial indicator in predicting corporate default of unlisted firms. Indeed, we show a significant improvement of the model performance when using market information in credit risk assessment, in addition to accounting information. Moreover, by taking advantage of global and local variable importance technique we prove that the increase in performance is effectively attributable to market information. Our study makes important inferences for policy implications because our findings shed new light on the opportunity for banks to potentially integrate their hybrid credit scoring methodologies with market information for credit risk assessments, with the purpose of increasing the accuracy of forecasting corporate defaults for unlisted firms. Future extensions stemming from this work could involve not only applying alternative prediction models so to provide further evidence on the importance of market information, but also testing the impact of synthetic information extracted with the dimensionality reduction technique when replacing the original financial ratios. Testing different clustering technique and exploring the distribution of the clusters could also lead to new insights on clients' behavior and their connections with the market, through the mapping with their publicly-listed peers.

This thesis is focused on both methodological and empirical aspects, in the sense that new methodologies have been designed and applied to real-world problems. The contribution of this thesis is in the development of new approaches tackling the topic of explainability, working on the complementary pillars of dimensionality reduction and variables importance with the aim of helping decision makers with the interpretation of the results.

Future research should extend and improve the methodological frameworks presented so far. Moreover, further interesting domains of application are already under examination, such as the unstructured data used in modelling the interactions between cities during the last Covid-19 pandemic, with the objective of evaluating the impact on the economic system and the consequences of restriction policies.

Bibliography

Abadie, A., Athey, S., Guido, W. I., and Wooldridge, J., 2017. When should you adjust standard errors for clustering? *NBER Working Paper*, 24003.

Abor, J., 2007. Industry classification and the capital structure of ghanaian smes. *Studies in Economics and Finance*, 24(3):207–219.

Acemoglu, D. and Robinson, J. A., 2005. The role of institutions in growth and development. *Review of Economic and Institutions*, 1.

Acemoglu, D., Johnson, S., and Robinson, J. A., 2001. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91:1369–1401.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. P., 2012. Measuring systemic risk. *CEPR Discussion Paper*, DP8824.

Adda, J., 2015. Economic activity and the spread of viral diseases. evidence from high frequency data. *IZA Discussion Paper*, 9326.

Adrian, T. and Brunnermeier, M. K., 2016. Covar. *Am Econ Rev*, 106(7):1705–1741.

Agarwal, V. and Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, 32: 1541–1551.

Agresti, A. *Categorical data analysis. Third Edition.* Wiley, 2012.

Ahamed, M. M. and Mallick, S., 2019. Is financial inclusion good for bank stability? international evidence. *Journal of Economic Behavior & Organization*, 157(C):403–427.

Ahn, B. S., Cho, S. S., and Kim, C., 2000. The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems With Applications*, 18:65–74.

Akbar, S., Rehman, S. u., and Ormrod, P., 2013. The impact of recent financial shocks on the financing and investment policies of uk private firms. *International Review of Financial Analysis*, 26(C):59–70.

Akbari, A., Ng, L., and Solnik, B., 2021. Drivers of economic and financial integration: A machine learning approach. *Journal of Empirical Finance*, 61:82–102.

Akinci, O. and Olmstead-Rumsey, J., 2018. How effective are macroprudential policies? an empirical investigation. *Journal of Financial Intermediation*, 33(C):33–57.

Albanesi, S. and Vamossy, D. F., 2019. Predicting consumer default: A deep learning approach. *National Bureau of Economic Research Working Paper*, 26165.

Alesina, A. and La Ferrara, E., 2005. Ethnic diversity and economic performance. *Journal of Economic Literature*, 43:762–800.

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

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., Matějů, J., Monteiro, N., Neudorfer, B., Peltonen, T., Rodrigues, P. M. M., Rusnák, M., Schudel, W., Sigmund, M., Stremmel, H., Šmídková, K., Van Tilburg, R., Vašíček, B., and Žigraiová, 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.

Alford, A. W., 1992. The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, 30:94–108.

Allen, F. and Gale, D., 2000. Comparing financial systems. *The MIT Press, Cambridge, MA*.

Almeida, H., Campello, M., and Weisbach, M. S., 2021. The cash flow sensitivity of cash: Replication, extension, and robustness. *Fisher College of Business Working Paper*, 2021-03-002.

Altman, E. I., 1980. Commercial bank lending: process, credit scoring, and costs of errors in lending. *Journal of Financial and Quantitative Analysis*, pages 813–832.

Altman, E., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23:589–609.

Altman, E., Sabato, G., and Wilson, N., 2010. The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk*, 6.

Ananou, F., Chronopoulos, D. K., Tarazi, A., and Wilson, J., 2021. Liquidity regulation and bank lending. *Journal of Corporate Finance*, 69(C):S0929119921001188.

Andrikopoulos, P. and Khorasgani, A., 2018. Predicting unlisted smes' default: Incorporating market information on accounting-based models for improved accuracy. *The British Accounting Review*, 50:559–573.

Arellano, M. and Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58:277–97.

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

Avramov, D., Li, M., and Wang, H., 2021. Predicting corporate policies using downside risk: A machine learning approach. *Journal of Empirical Finance*, 63:1–26.

Ayyagari, M., Beck, T., and Soledad Martinez Peria, M., 2018. The micro impact of macroprudential policies: Firm-level evidence. *IMF Working Paper*, WP/18/267.

Baker, M. and Ruback, R. S., 1999. Estimating industry multiples. *Working Paper Harvard University Cambridge*.

Bank for international settlements, 2006. International convergence of capital measurement and capital standards: a revised framework. *Bank for international settlements*.

Bank of England, 2008. Financial stability report. *NBER Working Papers*, (23).

Barbaglia, L., Manzan, S., and Tosetti, E., 07 2021. Forecasting Loan Default in Europe with Machine Learning*. *Journal of Financial Econometrics*.

Bazarbash, M., 2019. Fintech in financial inclusion: machine learning applications in assessing credit risk.

Beaver, W., 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, 4:71–111.

Becchetti, L. and Trovato, G., 2002. The determinants of growth for small and medium sized firms: The role of the availability of external finance. *Small Business Economics*, 19(4): 291–306.

Beck, T. and Laeven, L., 2006. Institution building and growth in transition economies. *Journal of Economic Growth*, 11:157–186.

Beck, T., Demirguc-Kunt, A., and Maksimovic, V., 2005. Financial and legal constraints to growth: Does firm size matter? *The Journal of Finance*, 60(1):137–177.

Beck, T., Saka, O., and Volpin, P., 2020. Financial stability: does social activism matter? *Finance and politics: New insights.* URL https://voxeu.org/article/finance-and-politics-new-insights.

Beck, T., Demirgüç-Kunt, A., and Levine, R., 2007. Finance, inequality and the poor. *Journal of Economic Growth*, 12(1):27–49.

Beck, T., Demirguc-Kunt, A., and Maksimovic, V., 2008. Financing patterns around the world: Are small firms different? *Journal of Financial Economics*, 89(3):467–487.

Berger, A. N. and Udell, G. F., 1995a. Relationship lending and lines of credit in small firm finance. *Journal of business*, pages 351–381.

Berger, A. N. and Udell, G. F., 1995b. Relationship lending and lines of credit in small firm finance. *The Journal of Business*, 68(3):351–381.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C., 2005. Does function follow organizational form? evidence from the lending practices of large and small banks. *Journal of Financial Economics*, 76:237–269.

Bharath, S. and Shumway, T., 2008. Forecasting default with the merton distance to default model. *Review Financial Studies*, 21:1339–1369.

Bhimani, A., Gulamhussen, M. A., and Lopes, S. D. R., 2010. Accounting and non-accounting determinants of default: an analysis of privately-held firms. *Journal of Accounting and Public Policy*, 29:517–532.

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

Blöchlinger, A. and Leippold, M., 2018. Are ratings the worst form of credit assessment except for all the others? *Journal of Financial and Quantitative Analysis*, 53(1):299–334.

Blum, M., 1974. Failing company discriminant-analysis. *Journal of Accounting Research*, 12:1–25.

Borio, C. and Lowe, P., 2002. Asset prices, financial and monetary stability: Exloring the nexus. *BIS Working Paper*, 114.

Borio, C. E. V. and Drehmann, M., 2009. Towards an operational framework for financial stability: 'fuzzy' measurement and its consequences. *BIS Working Paper*, 284.

Borio, C., English, W., and Filardo, A. A tale of two perspectives: old or new challenges for monetary policy? In Settlements, B. f. I., editor, *Monetary policy in a changing environment*, volume 19, pages 1–59. Bank for International Settlements, 2003.

Bouton, L. and Tiongson, E. R., 2010. Subjective perceptions of financing constraints: How well do they reflect credit market conditions? *Emerging Markets Review*, 11(2):98–105.

Boyce, T. and Brown, C., 2019. Economic and social impacts and benefits of health systems. *World Health Organization, Regional Office*.

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

Breden, D., 2008. Monitoring the operational risk environment effectively. *Journal of Risk Management in Financial Institutions*, 1:156–164.

Breiman, L., 2001. Random forests. *Machine Learning*, 45:5–32.

Brown, J. R., Martinsson, G., and Petersen, B. C., 2012a. Law, stock markets, and innovation. *Journal of Finance, Forthcoming*.

Brown, M., Westerfeld, S., Schaller, M., and Heusler, M., 2012b. Information or insurance? on the role of loan officer discretion in credit assessment. *MoFiR Working Paper*, 67.

Burgstahler, D., Hail, L., and Leuz, C., 2006. The importance of reporting incentives: Earnings management in european private and public firms. *The Accounting Review*, 81: 983–1016.

Bussmann, N., Giudici, P., Marinelli, D., and Papenbrock, J., 2020. Explainable ai in fintech risk management. *Frontiers in Artificial Intelligence*, 3:26.

Byström, H., 2006. Merton unraveled: A flexible way of modeling default risk. *Journal of Alternative Investments*, 8:39–47.

Calabrese, R., Marra, G., and Osmetti, S., 2016. Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. *Journal of the Operational Research Society*, 67:604–615.

Calem, P., Correa, R., and Lee, S. J., 2020. Prudential policies and their impact on credit in the united states. *Journal of Financial Intermediation*, 42:100826.

Campello, M., Graham, J. R., and Harvey, C., 2010. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics*, 97(3):470–487.

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

Capitaine, L., Genuer, R., and Thiébaut, R., 2019. Random forests for high-dimensional longitudinal data. *Statistical Methods in Medical Research*.

Carreira, C. and Lopes, L., 2016. Collecting new pieces to the regional knowledge spillovers puzzle: high-tech versus low-tech industries. *GEMF Working Papers*.

Carroll, N., 2018. Estimation of ordered generalized linear models. URL https://CRAN.R-project.org/package=oglmx.

Carty, L. and Fons, J., 1994. Measuring changes in corporate credit quality. *Special report, Moody's*.

Cassar, G., 2004. The financing of business start-ups. *Journal of Business Venturing*, 19(2): 261–283.

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

CDCP, 2014. Board of scientific counselors bsc meeting: Summary report/record of the proceedings. *Centers for Disease Control and Prevention.* URL http://www.cdc.gov/phpr/science/documents/bsc_ophpr_meeting_minutes_04_07_14.pdf.

Cerutti, E., Claessens, S., and Laeven, L., 2017. The use and effectiveness of macroprudential policies: New evidence. *Journal of Financial Stability*, 28(C):203–224.

Cevik, E. I., Dibooglu, S., and 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., and Sole, J. A., 2009. Assessing the systemic implications of financial linkages. *IMF Global Financial Stability Report*, 2.

Charalambous, C., Neophytou, E., and Charitou, A., 2004. Predicting corporate failure: empirical evidence for the uk. *European Accounting Review*, 13:465–497.

Chen, T. and Liao, H., 2005. A multi-period corporate credit model-an intrinsic valuation approach.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, page 1724–1734.

Chodorow-Reich, G., 2014. The employment effects of credit market disruptions: Firm-level evidence from the 2008-9 financial crisis. *The Quarterly Journal of Economics*, 129(1): 1–59.

Cihak, M., Demirguc-Kunt, A., Feyen, E., and Levine, R. Benchmarking financial systems around the world. Policy Research Working Paper Series 6175, The World Bank, 2012.

Claessens, S. and Tzioumis, K., 2006. Measuring firms access to finance. *Conference: Access to Finance, Building Inclusive Financial Systems*.

Claessens, S., 2006. Access to financial services: A review of the issues and public policy objectives. *World Bank Research Observer*, 21(2):207–240.

Claessens, S., Krahnen, J., and Lang, W. W., 2005. The basel ii reform and retail credit markets. *Journal of Financial Services Research*, 28(1-3):5–13.

Claessens, S., Ghosh, S. R., and Mihet, R., 2013. Macro-prudential policies to mitigate financial system vulnerabilities. *Journal of International Money and Finance*, 39:153–185.

Clarke, G. R. G., Xu, L. C., and fu Zou, H., 2006. Finance and income inequality: What do the data tell us? *Southern Economic Journal*, 72(3):578–596.

Colombo, E., 2001. Determinants of corporate capital structure: Evidence from hungarian firms. *Applied Economics*, 33:1689–1701. ISSN 0003-6846. doi: 10.1080/00036840010015057.

Contoyannis, P., Jones, A., and Rice, N., 2004. The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, 19:473–503.

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

Cucinelli, D., Di Battista, M. L., Marchese, M., and Nieri, L., 2018. Credit risk in european banks: The bright side of the internal ratings based approach. *Journal of Banking & Finance*, 93:213–229.

Dabla-Norris, E., Ji, Y., Townsend, R., and Unsal, D. F., 2015. Identifying constraints to financial inclusion and their impact on gdp and inequality: A structural framework for policy. *IMF Working Paper*.

Das, S., Hanouna, P., and Sarin, A., 2009. Accounting-based versus market-based cross-sectional models of cds spreads. *Journal of Banking and Finance*, 33:719–730.

Davies, D. L. and Bouldin, D. W., 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, page 224–227.

De Andrés, J., Landajo, M., and Lorca, P., 2005. Forecasting business profitability by using classification techniques: A comparative analysis based on a spanish case. *European Journal of Operational Research*, 167(2):518–542.

Debellut, F., Hendrix, N., Ortiz, J. R., Lambach, P., Neuzil, K. M., and Bhat, N., 2018. Forecasting demand for maternal influenza immunization in low- and lower-middle income countries. *PLoS One*, 13.

Degryse, H., Matthews, K., and Zhao, T., 2015. Smes and access to bank credit: Evidence on the regional propagation of the financial crisis in the uk. *CESifo Working Paper Series*, (5424).

Demirgüç-Kunt, A. and Maksimovic, V., 1998. Law, finance, and firm growth. *The Journal of Finance*, 53(6):2107–2137.

Demyanyk, Y. and Hasan, I., 2009. Financial crises and bank failures: a review of prediction methods. *Bank of Finland Research Discussion Paper*, 35.

Desai, M. A., Foley, C. F., and Hines Jr., J. R., 2004. A multinational perspective on capital structure choice and internal capital markets. *The Journal of Finance*, 59(6):2451–2487.

Devereux, M. and Schiantarelli, F. *Investment, Financial Factors, and Cash Flow: Evidence from U.K. Panel Data*, pages 279–306. University of Chicago Press, January 1990.

Dierkes, M., Erner, C., Langer, T., and Norden, L., 2013. Business credit information sharing and default risk of private firms. *Journal of Banking and Finance*, 37:2867–2878.

Doherty, B., Marin-Ferrer, M., and Vernaccini, L., 2018. The inform epidemic risk index.

Doumpos, M., Niklis, D., Zopounidis, C., and Andriosopoulos, K., 2015. Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from european listed firms. *Journal of Banking and Finance*, 50:599–607.

D.Phelps, M., Azman, A. S., Lewnard, J. A., Antillon, M., Simonsen, L., Andreasen, V., and andV. E. Pitze, P. K. M. J., 2017. The importance of thinking beyond the water supply in cholera epidemics. a historical urban case study. *PLoS Neglected Tropical Diseases*, 11: 1–15.

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

Duchin, R., Ozbas, O., and Sensoy, B. A., 2010. Costly external finance, corporate investment, and the subprime mortgage credit crisis. *Journal of Financial Economics*, 97(3): 418–435.

Edge, R. M. and Liang, J. N., 2019. New financial stability governance structures and central banks. *Finance and Economics Discussion Series (FEDS)*.

Edminster, R., 1972. An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, 7:1477–1493.

Ehigiamusoe, K. U. and Samsurijan, M. S., 2021. What matters for finance-growth nexus? a critical survey of macroeconomic stability, institutions, financial and economic development. *International Journal of Finance & Economics*, 26(4):5302–5320.

E.I., A. and G., S., 2007. Modelling credit risk for smes: Evidence from the us market. *The Journal of Business*, 43:332–357.

E.Marcozzi, D., Lawler, R. P. J. V., French, M. T., Mecher, C., Baehr, J. P. N. E., and Browne, B. J., 2020. Development of a hospital medical surge preparedness index using a national hospital survey. *Health Services and Outcomes Research Methodology*, 20:60–83.

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

Erkens, D. H., Hung, M., and Matos, P., 2012. Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide. *Journal of Corporate Finance*, 18(2):389–411.

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

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

Fafchamps, M. and Labonne, J., 2017. Using split samples to improve inference on causal effects. *Political Analysis*, 25(4):465–482.

Falavigna, G. Models for Default Risk Analysis: Focus on Artificial Neural Networks, Model Comparisons, Hybrid Frameworks. CERIS Working Paper 200610, Institute for Economic Research on Firms and Growth - Moncalieri (TO) ITALY -NOW- Research Institute on Sustainable Economic Growth - Moncalieri (TO) ITALY, 2006.

Falkenstein, E., Boral, A., and Carty, L., 2000. Riskcalctm for private companies: Moody's default model. *Moody's Investor Service*.

Fang, X., Jutrsa, D., Soledad Martinez Peria, M., Presbitero, A. F., and Ratnovski, L., 2020. Bank capital requirements and lending in emerging markets: The role of bank characteristics and economic conditions. *Journal of Banking & Finance*, page 105806.

Fantazzini, D. and Figini, S., 2009. Random survival forests models for sme credit risk measurement. *Methodology and Computing in Applied Probability*, 11:29–45.

Fazzari, S. M., Hubbard, G. R., and Petersen, B. C., 1988. Financing constraints and corporate investment. *Brooking Papers of Economic Activity*, 1:141–206.

Fazzari, S. M., Hubbard, G. R., and Petersen, B. C., 2000. Invesmtent - cash flow sensitivities are useful: A comment on kaplan and zingales. *The Quarterly Journal of Economics*, pages 695–705.

Fendoglu, S., 2017. Credit cycles and capital flows: Effectiveness of the macroprudential policy framework in emerging market economies. *Journal of Banking & Finance*, 79(C): 110–128.

Ferrando, A., Popov, A., and Udell, G. F., 2017. Sovereign stress and smes' access to finance: Evidence from the ecb's safe survey. *Journal of Banking & Finance*, 81(C):65–80.

Ferri, G. and Pesic, V., 2017. Bank regulatory arbitrage via risk weighted assets dispersion. *Journal of Financial Stability*, 33(C):331–345. doi: 10.1016/j.jfs.2016.10.006.

Filomeni, S., Udell, G. F., and Zazzaro, A., 2020a. Hardening Soft Information: Does Organizational Distance Matter? *European Journal of Finance*. doi: https://doi.org/10.1080/1351847X.2020.1857812.

Filomeni, S., Udell, G. F., and Zazzaro, A., 2020b. Communication frictions in banking organizations: Evidence from credit score lending. *Economics Letters*, 195C(109412).

Financial Stability Board, 2017. Artificial intelligence and machine learning in financial services: Market developments and financial stability implications.

Fiordelisi, F., Monferrà, S., and Sampagnaro, G., 2014. Relationship lending and credit quality. *Journal of Financial Services Research*, 46:295–315.

Fisera, B., Horvath, R., and Melecky, M., 2019. Basel iii implementation and sme financing: Evidence for emerging markets and developing economies. *World Bank Policy Research Working Paper*, 9069.

Fisher, A., Rudin, C., and Dominici, F., 2018. odel class reliance: Variable importance measures for any machine learning model class, from the 'rashomon' perspective. URL `http://arxiv.org/abs/1801.01489`.

Foglia, A., Laviola, S., and Reedtz, P. M., 1998. Multiple banking relationships and the fragility of corporate borrowers. *Journal of Banking and Finance*, 22:1441–1456.

Friedman, J., Hastie, T., and Tibshirani, R., 2009. The elements of statistical learning. *Springer*.

Funke, M., Schularick, M., and Trebesch, C., 2016. Going to extremes: Politics after financial crises, 1870–2014. *European Economic Review*, 88(C):227–260.

Gadanecz, B. and Jayaram, K., 2008. Measures of financial stability – a review. *Bank for International Settlements (Ed.), Proceedings of the IFC Conference on Measuring Financial Innovation and Its Impact.*, (31):365–380.

Garfield, R., Bartee, M., and Mayigane, L. N., 2019. Validating joint external evaluation reports with the quality of outbreak response in ethiopia, nigeria and madagascar. *The BMJ Global Health*, 4:1–8.

Gayer, M., Legros, D., Formenty, P., and Connolly, M. A., 2007. Conflict and emerging infectious diseases. *Emerging Infectious Diseases*, 13:1625–31.

Geard, N., Glass, K., McCaw, J. M., McBryde, E. S., Korb, K. B., Keeling, M. J., and McVernon, J., 2015. The effects of demographic change on disease transmission and vaccine impact in a household structured population. *Epidemics*, 13:56–64.

Gertler, M. and Gilchrist, S., 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, 109(2):309–340.

Gilbert, G. N., 2008. Researching social life.

Gonzalez, F., Haas, F., Johannes, R., Persson, M., Toledo, L., Violi, R., Wieland, M., and Zins, C., 2004. Market dynamics associated with credit ratings. a literature review. *Occasional Paper 16, European Central Bank.*

Gopalakrishnan, B., Jacob, J., and Mohapatra, S., 2021. Risk-sensitive basel regulations and firms' access to credit: Direct and indirect effects. *Journal of Banking & Finance*, 126.

Gracia, A., Gonzalez, S., Robles, V., and Menasalvas, E., 2014. A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Information Sciences*, 270: 1–27.

Gray, D. F., Merton, R., and Bodie, Z., 2007. New framework for measuring and managing macrofinancial risk and financial stability. *NBER Working Papers*, (13607).

Greene, W. and Hemsher, D., 2008. Modeling ordered choices: A primer and recent developments. *Working Paper 26, New York University, Leonard N. Stern School of Business, Department of Economics*.

Greene, W. H. *Econometric Analysis. Seventh Edition.* Pearson, Boston., 2012.

Grice, J. and Ingram, R., 2001. Tests of the generalizability of altman's bankruptcy prediction model. *Journal of Business Research*, 54:53–61.

Gropp, R. and Guettler, A., 2018. Hidden gems and borrowers with dirty little secrets: Investment in soft information, borrower self-selection and competition. *Journal of Banking and Finance*, 87:26–39.

Guiso, L., Sapienza, P., and Zingales, L., 2006. Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20:23–48.

Gupta, V., Kraemer, J. D., Katz, R., Jha, A. K., Kerry, V. B., Sane, J., Ollgren, J., and Salminen, M. O., 2018. Analysis of results from the joint external evaluation, examining its strength and assessing for trends among participating countries. *Journal of Global Health*, 8:1–9.

Hajjem, A., Bellavance, F., and Larocque, D., 2014. Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84:1313–1328.

Hakkio, C. and 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*, page 5–50.

Hall, G., Hutchinson, P., and Michaelas, N., 2000. Industry effects on the determinants of unquoted smes' capital structure. *International Journal of the Economics of Business*, 7 (3):297–312.

Harhoff, D. and Korting, T., 1998. Lending relationships in germany - empirical evidence from survey data. *Journal of Banking & Finance*, 22(10-11):1317–1353.

Harrison, A. and McMillan, M., 2003. Does direct foreign investment affect domestic credit constraints? *Journal of International Economics*, 61(1):73–100.

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

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

Hawkins, J. and Klau, M., 2000. Measuring potential vulnerabilities in emerging market economies. *BIS Working Paper*, 91.

Hillegeist, S., Keating, E., Cram, D., and Lundstedt, K., 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9:5–34.

Hol, S., 2007. The influence of the business cycle on bankruptcy probability. *International Transactions in Operational Research*, 14:75–90.

Holmes, E. E., Ward, E. J., and Scheuerell, M. D., 2018. Analysis of multivariate time-series using the marss package. URL https://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf.

Holmfeldt, M., Rydén, A., Strömberg, L., and 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.

Honohan, P., 2006. Cross-country variation in household access to financial services. *The World Bank, Trinity College Dublin and CEPR*.

Hsieh, N.-C., 2004. An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27(4):623 – 633.

Huang, R., Kale, S., Paramati, S. R., and Taghizadeh-Hesary, F., 2021. The nexus between financial inclusion and economic development: Comparison of old and new eu member countries. *Economic Analysis and Policy*, 69:1–15.

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

Hufnagel, L., Brockmann, D., and Geisel, T., 2004. Forecast and control of epidemics in a globalized world. *PNAS*, 101:15124–15129.

Čihák, M. and Schaeck, K., 2010. How well do aggregate prudential ratios identify banking system problems? *Journal of Financial Stability*, 6(3):130–144.

Illing, M. and 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., and Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115:53–74.

Inoue, A. and Solon, G., 2006. A portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory*.

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

International Monetary Fund, 2019b. The financial soundness indicators and the imf. URL `https://www.imf.org/external/np/sta/fsi/eng/fsi.htm`.

Islam, A. and Tedford, D., 2012. Risk determinants of small and medium-sized manufacturing enterprises (smes) - an exploratory study in new zealand. *Journal of Industrial Engineering International*, 8.

Islam, M. A., 2008. Risk management in small and medium-sized manufacturing organization in new zealand. *PhD Thesis, Department of Mechanical Engineering. The University of Auckland*.

Islam, S. R., Eberle, W., Ghafoor, S. K., and Ahmed, M., 2021. Explainable artificial intelligence approaches: A survey. *arXiv.org*. URL `https://arxiv.org/abs/2101.09429`.

Islami, M. and Kurz-Kim, J. R., 2013. A single composite financial stress indicator and its real impact in the euro area. *Bundesbank Discussion Paper*, 31.

Ivashina, V. and Scharfstein, D., 2010. Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97(3):319–338.

J., B. and V., A., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? a comprehensive test. *Journal of Banking and Finance*, 40:432–442.

J.F., W., J.A., S., and B.A., J., 2001. Takeovers, restructuring, and corporate governance. *Pearson*.

Johns Hopkins University Centre for Health Security, 2019. Forecasting demand for maternal influenza immunization in low- and lower-middle income countries. *Global Health Security Index*. URL `https://www.ghsindex.org/`.

Johnson, D. R. and Young, R., 2011. Toward best practices in analyzing datasets with missing data, comparisons and recommendations. *Journal of Marriage and Family*, 73:926–45.

Kabundi, A. and Mbelu, A., 2017. Estimating a timevarying financial conditions index for south africa. *Working Papers, South African Reserve Bank*, 8008.

Kafertein, F. K., 1997. Foodborne disease control: A transnational challenge. *Emerging Infectious Diseases*, 3:503–510.

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

Kaji, A. H., Langford, V., and Lewis, R. J., 2008. Assessing hospital disaster preparedness: A comparison of an on-site survey, directly observed drill performance, and video analysis of teamwork. *Annals of Emergency Medicine*, 52:195–201.

Kaplan, S. and Zingales, L., 2000. Investment-cash flow sensitivities are not valid measures of financing constraints. *The Quarterly Journal of Economics*, 115(2):707–712.

Kaplan, S. N. and Zingales, L., 1997. Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics*, 112(1):169–215.

Karlan, D. and Morduch, J. Access to finance. volume 5, chapter Chapter 71, pages 4703–4784. Elsevier, 2010.

Keasey, K. and Watson, R., 1987. Non-financial symptoms and the prediction of small company failure: A test of argenti's hypotheses. *Journal of Business Finance and Accounting*.

Khan, S. A. and 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., and Ryu, D., 2020. Corporate default predictions using machine learning: Literature review. *Sustainability*, 12(16).

King, G., Honaker, J., Joseph, A., and Scheve, K., 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95:49–69.

Klapper, L., Laeven, L., and Rajan, R., 2006. Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics*, 82(3):591–629.

Kleibergen, F. and Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133:97–126.

Kliesen, K. L. and Smith, D. C., 2010. Measuring financial market stress. *Federal Reserve Bank of St. Louis, Economic Synopses*, 2.

Knack, S. and Xu, L. C., 2017. Unbundling institutions for external finance: Worldwide firm-level evidence. *Journal of Corporate Finance*, 44(C):215–232.

Koplan, J. P., Milstein, R., and Wetterhall, S., 1999. Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report*, 48:1–40.

Kramer, M. A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37:233–243.

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

Kruk, M. E. and Freedman, L. P., 2008. Assessing health system performance in developing countries. a review of the literature. *Health Policy*, 85:263–76.

Kucher, A., Mayr, S., Mitter, C., Duller, C., and Feldbauer-Durstmüller, B., 2020. Firm age dynamics and causes of corporate bankruptcy: age dependent explanations for business failure. *Review of Managerial Science*, 14:633–661.

Kuntchev, V., Ramalho, R., Rodriguez-Meza, J., and Yang, J. S. What have we learned from the enterprise surveys regarding access to credit by SMEs ? Policy Research Working Paper Series 6670, The World Bank, October 2013. URL https://ideas.repec.org/p/wbk/wbrwps/6670.html.

KWR, 2020. What we learn about the corona virus through waste water research. URL https://www.kwrwater.nl/en/actueel/what-can-we-learn-about-the-corona-virus-throughwaste-water-research.

Laeven, L. and Valencia, F., 2012. Systemic banking crises: An update. *IMF Working Paper*, 12(163). URL www.imf.org/external/pubs/ft/wp/2012/wp12163.pdf.

Laeven, L., 2003. Does financial liberalization reduce financing constraints? *Financial Management*, 32(1):5–34.

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

Lang, T., 2001. Trade, public health and food. *International Cooperation in Health*, pages 81–108.

Leary, M. T., 2009. Bank loan supply, lender choice, and corporate capital structure. *The Journal of Finance*, 64(3):1143–1185.

Levine, R., 1997. Financial development and economic growth: Views and agenda. *Journal of Economic Literature*, 35(2):688–726.

Levine, R., Loayza, N., and Beck, T., August 2000a. Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46(1):31–77.

Levine, R., Loayza, N., and Beck, T., August 2000b. Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, 46(1):31–77.

Li, R., Richmond, P., and Roehner, B. M., 2018. Effect of population density on epidemics. *Physica A: Statistical Mechanics and its Applications*, 510:713–724.

Liberti, J. and Mian, A., 2009. Estimating the effect of hierarchies on information use. *Review Financial Studies*, 22:4057–4090.

Liberti, J. and Petersen, M., 2017. Information: Hard and soft. *Review of Corporate Finance Studies*, 8:1–41.

Liberti, J. M. and Petersen, M. A., 2018. Information: Hard and Soft. *The Review of Corporate Finance Studies*, 8(1):1–41.

Lim, C., Columba, F., Costa, A., Kongsamut, P., Otani, A., Saiyid, M., Wezel, T., and Wu, X., 2011. Macroprudential policy: What instruments and how are they used? lessons from country experiences. *IMF Working Paper*, 11/238.

Lin, S.-W., Shiue, Y.-R., Chen, S.-C., and Cheng, H.-M., 2009. Applying enhanced data mining approaches in predicting bank performance: A case of taiwanese commercial banks. *Expert Systems with Applications*, 36(9):11543 – 11551.

Lin, W. Y., Hu, Y. H., and 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.

Lo Duca, M. and Peltonen, T. A., 2011. Macro-financial vulnerabilities and future financial stress assessing systemic risks and predicting systemic events. *CB Working Paper Series*, 1311.

Lo Duca, M. and Peltonen, T. A., 2013. Assessing systemic risks and predicting systemic events. *Journal of Banking and Finance*, 37(7):2183–2195.

Louzada, F., Ara, A., and Fernandes, G., 2016a. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, 21:117–134.

Louzada, F., Ara, A., and Fernandes, G. B., 2016b. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, 21(2):117 – 134.

Louzis, D. P. and Vouldis, A. T., 2011. A financial systemic stress index for greece. *First Conference of the Macro-prudential Research (MaRs) network of the European System of Central Banks in Frankfurt am Main*.

Luethi, D., Erb, P., and Otziger, S. *FKF: Fast Kalman Filter*, 2018. URL `https://CRAN.R-project.org/package=FKF`. R package version 0.1.5.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I., 2020. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2:2522–5839.

Mac an Bhaird, C., Vidal, J. S., and Lucey, B., 2016. Discouraged borrowers: Evidence for eurozone smes. *Journal of International Financial Markets, Institutions and Money*, 44 (C):46–55.

MacQueen, J. B., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297.

Mamatzakis, E. C. and Tsionas, M. G., 2020. A bayesian panel stochastic volatility measure of financial stability. *International Journal of Finance & Economics*.

McCarthy, E., 1999. Pricing ipos: Science or science fiction? *Journal of Accountancy*, 188: 51–58.

McInnes, L., Healy, J., and Melville, J., February 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*.

McInnes, L., Healy, J., Saul, N., and Grossberger, L., 2018. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861.

Merton, R. C., 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470.

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

Mertzanis, C., 2019. Family ties, institutions and financing constraints in developing countries. *Journal of Banking & Finance*, 108:105650.

Mertzanis, C., 2020. Financial supervision structure, decentralized decision-making and financing constraints. *Journal of Economic Behavior & Organization*, 174:13–37.

Mertzanis, C. and Papastathopoulos, A., 2021. Epidemiological susceptibility risk and tourist flows around the world. *Annals for Tourism Research*, 86.

Mertzanis, C., 2017. Ownership structure and access to finance in developing countries. *Applied Economics*, 49(32):3195–3213.

Meyer, M. D. and Elrahman, O. A., 2019. transportation and public health. an integrated approach to policy, planning, and implementation. *The National Academy of Sciences, Engineering and medicine.* URL https://www.elsevier.com/books/transportation-and-public-health/elrahman/978-0-12-816774-8.

Moccero, D. N., Pariès, M. D., and Maurin, L., 2014. Financial conditions index and identification of credit supply shocks for the euro area. *International Finance*, 17(3):297–321.

Modigliani, F. and Miller, M., 1958. The cost of capital, corporate finance, and the theory of investment. *American Economic Review*, 3(48):261–297.

Molnar, C. *Interpretable Machine Learning*. 2019. https://christophm.github.io/interpretable-ml-book/.

Morand, S., 2020. Emerging diseases, livestock expansion and biodiversity loss are positively related at global scale. *Biological Conservation*, 248.

Morse, S. S., 2007. Global infectious disease surveillance and health intelligence. *Health Affairs*, 26:1069–77.

Moscatelli, M., Narizzano, S., Parlapiano, F., and Viggiano, G. Corporate default forecasting with machine learning. Temi di discussione (Economic working papers) 1256, Bank of Italy, Economic Research and International Relations Area, 2019a.

Moscatelli, M., Narizzano, S., Parlapiano, F., and Viggiano, G., 2019b. Corporate default forecasting with machine learning. *Bank of Italy Working Paper*, 1256.

Moulton, B. R., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, 72(2):334–338.

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

Najmul, H., 2020. The global health security index and joint external evaluation score for health preparedness are not correlated with countries' covid-19 detection response time and mortality outcome. *Epidemiology and Infection*, 148:1–8.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E., 2005. Handbook on constructing composite indicators: methodology and user guide. *OECD Statistics Working Paper STD/DOC*.

Neiderud, C. J., 2014. How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology and Epidemiology*, 5:1–9.

Nelson, W. R. and Perli, R., 2007. Selected indicators of financial stability. *Irving Fisher Committee's Bulletin on Central Bank Statistics*, 23:92–105.

NHSPI, 2015. The national health security preparedness index. *The BMJ Global Health*. URL http://www.nhspi.org/about/.

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., and Vasilakis, C., 2021. Forecasting and planning during a pandemic: Covid-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*, 290:99–115.

Norden, L. and Weber, M., 2010. Credit line usage, checking account activity, and default risk of bank borrowers. *Review of Financial Studies*, 23:3665–3699.

North, D., 1990. Institutions, institutional change and economic performance. *Cambridge: Cambridge University Press.*

Odders-White, E. and Ready, M., 2006. Credit ratings and stock liquidity. *Review of Financial Studies*, 19:119–157.

OECD, 2020. Financing smes and entrepreneurs: An oecd scoreboard. special edition: The impact of covid-19.

Ohlson, J., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18:109–131.

Oliner, S. D. and Rudebusch, G. D., November 1992. Sources of the Financing Hierarchy for Business Investment. *The Review of Economics and Statistics*, 74(4):643–654.

Olmeda, I. and Fernandez, E., 1997. Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction. *Computational Economics*, 10(4):317–335.

Olson, L. M., Qi, M., Zhang, X., and Zhao, X., 2021. Machine learning loss given default for corporate debt. *Journal of Empirical Finance*, 64:144–159.

Opschoor, A., van Dijk, D., and van der Wel, M., 2014. Predicting volatility and correlations with financial conditions indexes. *Journal of Empirical Finance*, 29:435–447.

Osborne, M. J. and Rubinstein, A., 1994. A course in game theory. *MIT press*.

Oster, E., 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics, Taylor & Francis Journals*, 37:187–204.

Ozili, P. K., 2020. Financial stability: does social activism matter? *Journal of Financial Regulation and Compliance*, 28:183–214.

Paravisini, D., 2008. Local bank financial constraints and firm access to external finance. *The Journal of Finance*, 63(5):2161–2193.

Peel, M., Peel, D., and Pope, P., 1986. Predicting corporate failure - some results for the uk corporate sector. *Omega*, 14:5–12.

Petersen, M. A. and Rajan, R. G., 1994. The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, 49(1):3–37.

Pew Research Center, 2020. The role of the internet during the covid-19 outbreak. *Pew Research*. URL https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-covid-19-

Pindado, J., Rodrogues, L., and de la Torre, C., 2008. Estimating financial distress likelihood. *Journal of Business Research*, 61:995–1003.

Polonsky, J. A., Baidjoe, A., Kamvar, Z. N., Cori, A., Durski, K., and Edmunds, W. J., 2019. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 374.

Pritchett, L., Woolcock, M., and Andrews, M., 2013. Looking like a state, techniques of persistent failure in state capability for implementation. *Journal of Development Studies*, 49:1–18.

Putnam, R., 2000. Bowling alone: The collapse and revival of american community. *Simon & Schuster*.

Qian, J., Strahan, P. E., and Yang., Z., 2015. The impact of incentives and communication costs on information production and use: Evidence from bank lending. *Journal of Finance*, 70:1457–1493.

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

Quah, S. R., 2007. Crisis preparedness: Asia and the global governance of epidemics. *Washington, D.C.: The Brookings Institution.*

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

Rainwater-Lovett, K., Rodriguez-Barraquer, I., and Moss, W. J., 2016. Tracking viruses with smartphones and social media. in katze, m. g., korth, m. j., lyn, g. nathanson, n. (eds) viral pathogenesis. *Basics to Systems Biology.*

Rajabi, M., Mansourian, A., Pilesjo, P., Hedefalk, F., Groth, R., and Bazmani, A., 2014. Comparing knowledge-driven and data-driven modeling methods for susceptibility mapping in spatial epidemiology: a case study in visceral leishmaniasis. in huerta, j., schade, s., granell, c. (eds) connecting a digital europe through location and place. *Proceedings of the AGILE'2014. International Conference on Geographic Information Science*, pages 3–6.

Rajan, R. G. and Zingales, L., 1998. Financial dependence and growth. *NBER Working Paper*, 5758.

Razavi, A., Erondu, N. A., and Okereke, E., 2020. The global health security index: what value does it add? *The BMJ Global Health*, 5:1–3.

Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., and Moore, E., 2019. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. *Proc. Natl. Acad. Sci. U. S. A.*, 116:3146–3154.

Relman, D. A., Choffnes, E. R., and Mack, R., 2020. Infectious disease movement in a borderless world: Workshop summary. *Forum on Microbial Threats; Institute of Medicine of the National Academies*.

Restoy, F., 2017. Financial soundness indicators - looking beyond the lessons learned from the crisis. *Keynote address. Users' Workshop on Financial Soundness Indicators, International Monetary Fund*. URL https://www.bis.org/speeches/sp170427.htm.

Rikkers, F. and Thibeault, A. E., 2009. A structural form default prediction model for smes, evidence from the dutch market. *Multinational Finance Journal*, 13:229–264.

Rivers, C., Chretien, J. P., Riley, S., Pavlin, J. A., Woodward, A., and Brett-Major, D., 2019. Using outbreak science to strengthen the use of models during epidemics. *Nat. Commun.*, 10.

Rodriguez Gonzalez, M., Basse, T., and Kunze, F., 2018. Early warning indicator systems for real estate investments: Empirical evidence and some thoughts from the perspective of financial risk management. *ZVersWiss*, 107:387–403.

Rodrik, D. and Subramanian, A., 2009. Why did financial globalization disappoint? *IMF Staff Papers*, 56(1):112–138.

Roodman, D., 2011. Fitting fully observed recursive mixed-process models with cmp. *Stata Journal*, 11(2):159–206.

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

Rousseeuw, P. J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, pages 53–65.

S., T., Y., Y., and H., G., 2015. Variable selection and corporate bankruptcy forecasts. *Journal of Banking and Finance*, 52:89–100.

Saha, M. and Dutta, K. D., 2020. Nexus of financial inclusion, competition, concentration and financial stability. *Competitiveness Review: An International Business Journal Incorporating Journal of Global Competitiveness*, 31:669–692.

Sales, A. S., Areosa, W. D., and Areosa, M. B. M., 2012. Some financial stability indicators for brazil. *Banco Central do Brasil Working Paper*, 287.

Santoso, W. and Sukada, M. Risk profile of households and the impact on financial stability. In Settlements, B. f. I., editor, *Household debt: implications for monetary policy and financial stability*, volume 46, pages 58–74. Bank for International Settlements, 2009.

Schiantarelli, F. Financial constraints and investment: A critical review of methodological issues and international evidence. Boston College Working Papers in Economics 293., Boston College Department of Economics, 1995. URL https://EconPapers.repec.org/RePEc:boc:bocoec:293.

Schiffer, M. and Weder di Mauro, B. Firm size and the business environment: Worldwide survey results. Working papers, World Bank - International Finance Corporation, 2001.

Segal, M. R., 1992. Tree-structured methods for longitudinal data. *Journal of the American Statistical Association*, 87:407–418.

Segoviano, M. A. and Goodhart, C., 2009. Banking stability measures. *IMF Working Paper*, 9(4).

Sela, R. J. and Simonoff, J. S., 2012. Re-em trees: A new data mining approach for longitudinal data. *Machine Learning*, 86:169–207.

Sexton, J., 2018. Historical tree ensembles for longitudinal data. URL https://CRAN.R-project.org/package=htree.

Shahpar, C., 2019. Protecting the world from infectious disease threats: now or never. *BMJ Global Health*, 4.

Shapley, L. S., 1953. A value for n-person games. *Contributions to the Theory of Games 2.28*, pages 307–317.

Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, I., and Hassanien, A. E., 2020. Forecasting models for coronavirus disease covid-19: A survey. *SN Computer Science 1*, 197.

Shirakawa, M., 2012. Deleveraging and growth: is the developed world following japan's long and winding road? *Lecture at the London School of Economics and Political Science, Asia Research Centre*.

Simiyu, C. N., Odhiambo-Otieno, G., and Okero, D., 2014. Capacity indicators for disaster preparedness in hospitals within nairobi county, kenya. *Pan Afr Med Journal*, 18:349.

Slingenberg, J. W. and de Haan, J., 2011. Forecasting financial stress. *De Nederlandsche Bank Working Paper*, 292.

Spreco, A., Eriksson, O., Dahlstrom, O., Cowling, B. J., and Timpka, T., 2018. Evaluation of nowcasting for detecting and predicting local influenza epidemics, sweden, 2009-2014. *Emerging Infect. Dis.*, 24:1868–1873.

Stein, J. C., 2002. Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, LVII(5):1891–1921.

Stickney, C. P. and Weil, R. L., 1997. Financial accounting: An introduction to concepts, methods, and uses. *Dryden Press Series in Accounting*.

Stiglitz, J., 2000. Capital market liberalization, economic growth, and instability. *World Development*, 28(6):1075–1086.

Strumbelj, E. and Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems 41.3*, pages 647–665.

Suhrcke, M., Stuckler, D., Suk, J. E., Desai, M., Senek, M., McKee, M., Tsolova, S., Basu, S., Abubakar, I., and Hunter, P., 2011. The impact of economic crises on communicable disease transmission and control, a systematic review of the evidence. *PLoS ONE*.

Sveriges Riksbank, 2008. Financial stability report. *NBER Working Papers*, (2008:1).

T., S., 2001. Forecasting bankruptcy more accurately: a simple hazard model. *The Journal of Business*, 74:101–124.

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

Tarwater, P. M. and Martin, C. F., 2001. Effects of population density on the spread of disease. *Complexity*, 6:29–36.

T.Craig, A., Heywood, A. E., and Hall, J., 2020. Risk of covid-19 importation to the pacific islands through global air travel. *Epidemiology and Infection*, 148.

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

The INFORM Annual Meeting Report, 2017. URL https://docs.google.com/document/d/13B8L7_XQLWNxp_vZa8Dbow2cWLUOo2UGoSN083yKSZs/edit?usp=sharing.

The International Monetary Fund, 2020. Exceptional times, exceptional action. *Opening Remarks for Spring Meetings Press Conference*. URL https,//www.imf.org/en/News/Articles/2020/04/15/sp041520-exceptional-times-exceptional-action.

The OECD, 2010. Social capital, human capital and health: What is the evidence?

The OECD, 2020. Environmental health and strengthening resilience to pandemics.

The World Health Organization, 2017. Joint external evaluation jee. *Zoonotic Diseases Action Package Conference.*

The World Health Organization, 2020d. Emergencies preparedness, response. sars-cov-2 variants. URL https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/.

Tinoco, M. H. and Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30:394–419.

Van den End, J., 2006. Indicator and boundaries of financial stability. *Working Paper DNB*, 97.

Vassalou, M. and Xing, Y., 2004. Default risk in equity returns. *Journal of Finance*, 59: 831–868.

Vazzoler, S., Frattarolo, L., and Billio, M. *sparsevar: A Package for Sparse VAR/VECM Estimation*, 2016. URL https://CRAN.R-project.org/package=sparsevar. R package version 0.0.10.

Vermeulen, R., Hoeberichts, M., Vasícek, B., Zigraiová, D., Smídková, K., and de Haan, J., 2015. Financial stress indices and financial crises. *Open Econ. Rev.*, 26:383–406.

Virdi, A. A., 2005. Risk management among smes–executive report of discovery research. *The Consultation and Research Centre of the Institute of Chartered Accountants in England and Wales*.

Volk, M., 2012. Estimating probability of default and comparing it to credit rating classification by banks. *Economic and Business Review*, 14:299–320.

Wooldridge, J., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20:39–54.

World Bank Group. Credit Scoring Approaches Guidelines. Technical report, 2019.

Wursten, J., 2018. Testing for serial correlation in fixed-effects panel models. *Stata Journal*.

Yan, Y., 2016. rbayesianoptimization: Bayesian optimization of hyperparameters. URL https://CRAN.R-project.org/package=rBayesianOptimization.

Yarba, I. and Guner, Z. N., 2020. Uncertainty, macroprudential policies and corporate leverage : Firm-level evidence. *Central Bank Review*, 20(2):33–42.

Zhu, L., Qiu, D., Ergu, D., Ying, C., and Liu, K., 2019. A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162:503–513.

A A data-driven approach to measuring financial soundness

A.1 List of countries

Table A1: Complete list of selected countries and relative missing values count and percentage over total number of observations.

Country	Missing values	Country	Missing values	Country	Missing values
Albania	-	Tanzania	-	Lesotho	15 (11%)
Argentina	-	Turkey	-	Pakistan	15 (11%)
Armenia	-	Uganda	-	Belgium	16 (11.8%)
Austria	-	Ukraine	-	Finland	16 (11.8%)
Brazil	-	U. K.	-	Kuwait	16 (11.8%)
Brunei	-	Uruguay	-	Nigeria	16 (11.8%)
Burundi	-	Italy	2 (1.5%)	Singapore	16 (11.8%)
Cambodia	-	Switzerland	3 (2.2%)	India	17 (12.5%)
Cameroon	-	Cyprus	4 (2.9%)	Korea Rep	17 (12.5%)
Ctr Afr Rep	-	Eswatini	4 (2.9%)	Solomon	17 (12.5%)
Chad	-	Latvia	4 (2.9%)	Honduras	18 (13.2%)
Macao	-	Seychelles	4 (2.9%)	Netherlands	18 (13.2%)
Congo	-	Colombia	5 (3.7%)	Chile	20 (14.7%)
Croatia	-	Hong Kong	6 (4.4%)	Lebanon	22 (16.2%)
Denmark	-	Fiji	6 (4.4%)	Algeria	23 (16.9%)
El Salvador	-	Kenya	6 (4.4%)	Australia	24 (17.6%)
Eq. Guinea	-	Tonga	6 (4.4%)	Moldova	24 (17.6%)
Gabon	-	Vanuatu	6 (4.4%)	Panama	24 (17.6%)
Georgia	-	Ghana	7 (5.1%)	San Marino	24 (17.6%)
Germany	-	Bolivia	8 (5.9%)	Spain	24 (17.6%)
Guatemala	-	Bosnia	8 (5.9%)	Thailand	24 (17.6%)
Indonesia	-	Canada	8 (5.9%)	United States	24 (17.6%)
Kazakhstan	-	Czech Republic	8 (5.9%)	Vietnam	24 (17.6%)
Kyrgyz Rep	-	Dominican Rep	8 (5.9%)	Sri Lanka	31 (22.8%)
Macedonia	-	Greece	8 (5.9%)	China	32 (23.5%)
Madagascar	-	Kosovo	8 (5.9%)	Costa Rica	32 (23.5%)
Malta	-	Luxembourg	8 (5.9%)	Ecuador	32 (23.5%)
Mauritius	-	Paraguay	8 (5.9%)	Malaysia	32 (23.5%)
Namibia	-	Portugal	8 (5.9%)	Angola	34 (25%)
Nicaragua	-	Trinidad Tobago	8 (5.9%)	Botswana	34 (25%)
P. N. Guinea	-	West Bank	8 (5.9%)	Gambia	34 (25%)
Peru	-	Zambia	8 (5.9%)	Bangladesh	36 (26.5%)
Philippines	-	Bulgaria	10 (7.4%)	France	36 (26.5%)
Poland	-	Lithuania	10 (7.4%)	Ireland	37 (27.2%)
Romania	-	Estonia	12 (8.8%)	Djibouti	40 (29.4%)
Russia	-	Mexico	12 (8.8%)	Hungary	40 (29.4%)
Rwanda	-	Afghanistan	13 (9.6%)	Norway	40 (29.4%)
Saudi Arabia	-	Bhutan	13 (9.6%)	Slovenia	40 (29.4%)
Slovak Rep	-	Belarus	14 (10.3%)	Sweden	40 (29.4%)
South Africa	-	Israel	14 (10.3%)		

A.2 Missing values imputation methodology

Figure A1: Missing values imputation methodologies are tested in three settings. In the first (named *Original*, setting a) the whole dataset contains 8% of missing values, and additional values representing 10%, 20% and 30% of the initial dataset are randomly removed. In the second (named *No missing*, setting b) all entries with missing values are dropped from the whole dataset and the same incremental sampling procedure is applied on the remaining subset. In the last (named *Some missing*, setting c) all countries with at least 3 missing values for any year are dropped and the incremental sampling procedure is again applied on the remaining subset. The blue shaded bars in the upper row represent the average precentage reconstruction error (MAPE) for different percentage of additional missing values for each of the three settings. Whiskers on the top of each bar shows the scaled magnitude of maximum value of reconstruction error as well its numeric value, the relative magnitude R of MAPE compared to the average value of the original dataset and the relative magnitude RM of the maximum percentage reconstruction error compared to the average value of the original dataset. Bars on the lower row represent the percentage variation of the average reconstruction error of b) *No missing* and c) *Some missing* settings compared to a) *Original* setting, i. e. $MARE/MARE_{\text{Orig}} - 1$. Green bars mean that the imputation technique has a lower average reconstruction error when applied on the subset with no original missing values, setting b, and on the subset with some original missing values, setting c, compared to the average reconstruction error when applied on the full dataset with all original missing values, setting a.

A.3 Scree and loading plot for PCA methods

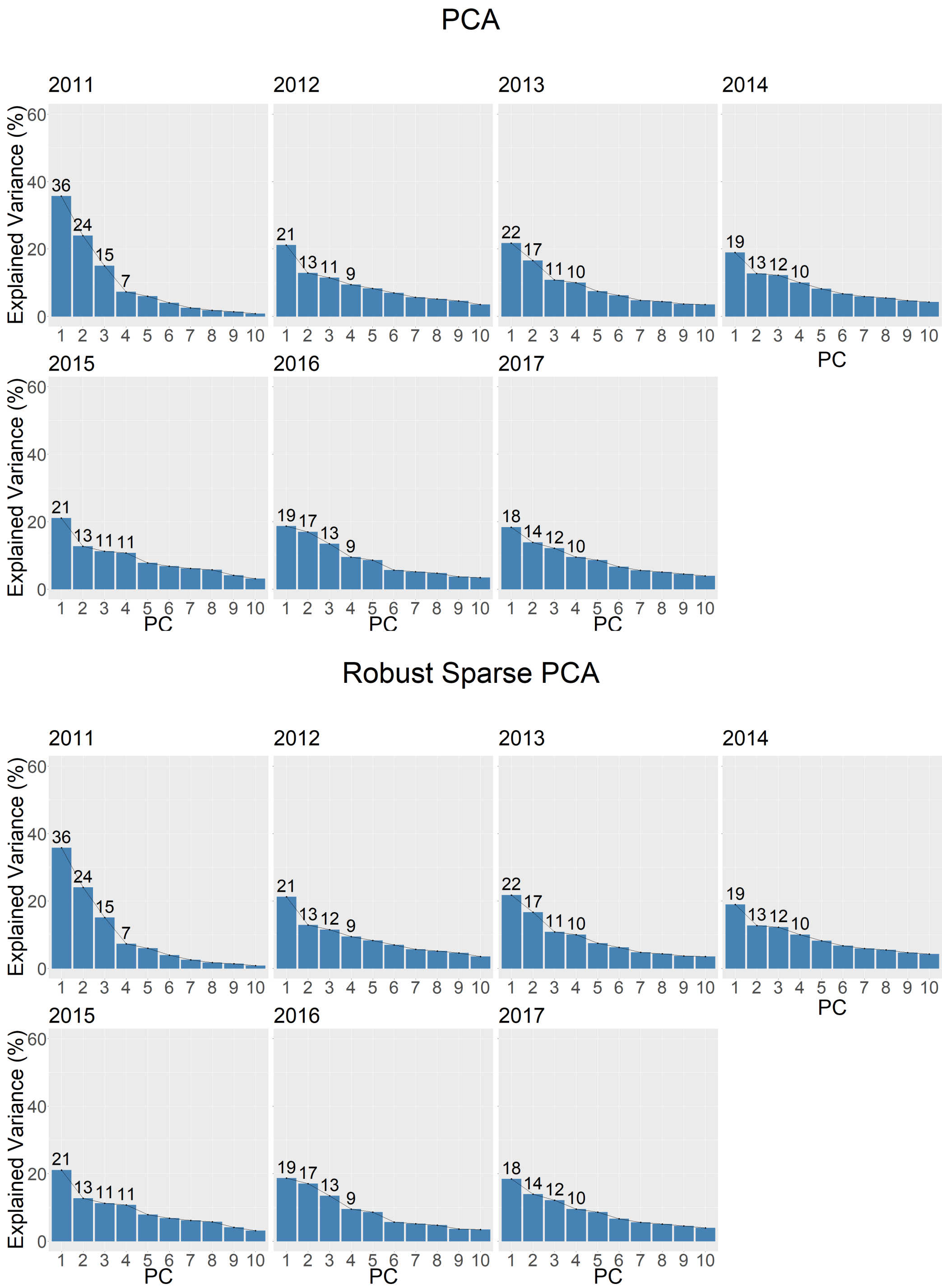


Figure A2: Scree plot for PCA and Sparse Robust PCA method for each year. The first two components account for an average of 38% of cumulative explained variance for both methods.

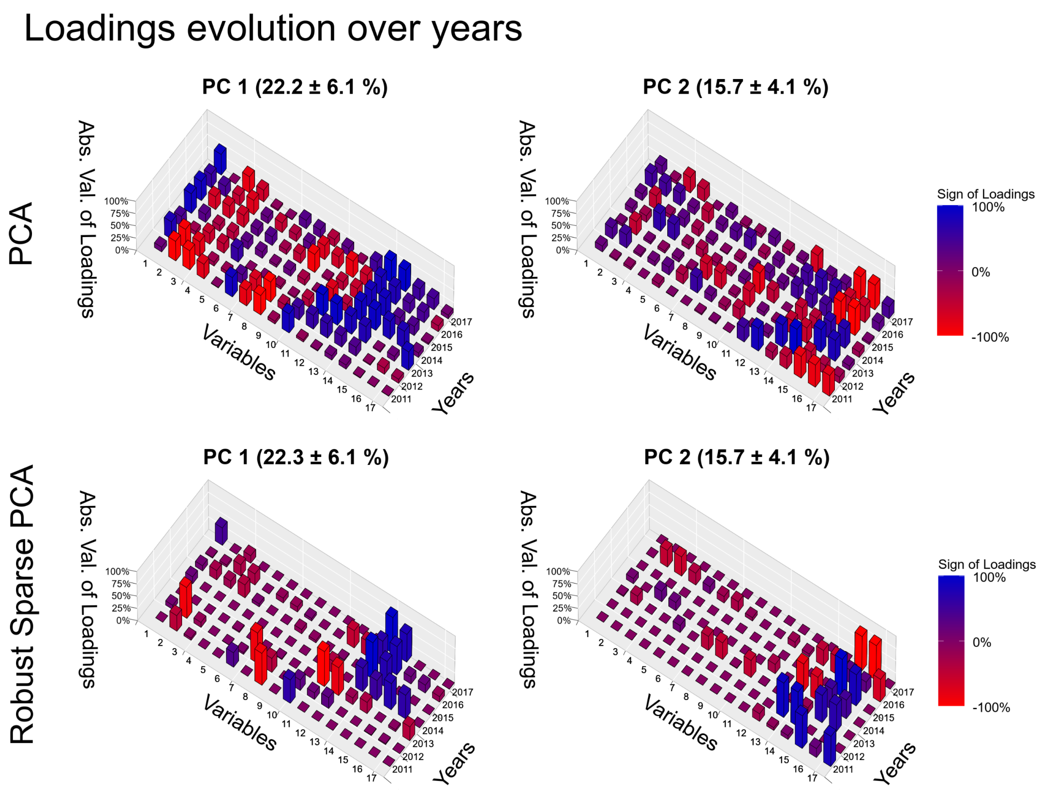


Figure A3: Loading importance over years for Robust PCA method, i.e. magnitude of weights of the linear combination that defines each component. Blue shaded bars represent positive contribution to each component loading while red shaded bars negative one. The higher the bar the more the original variable contributes to the loading. 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 forex 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'.

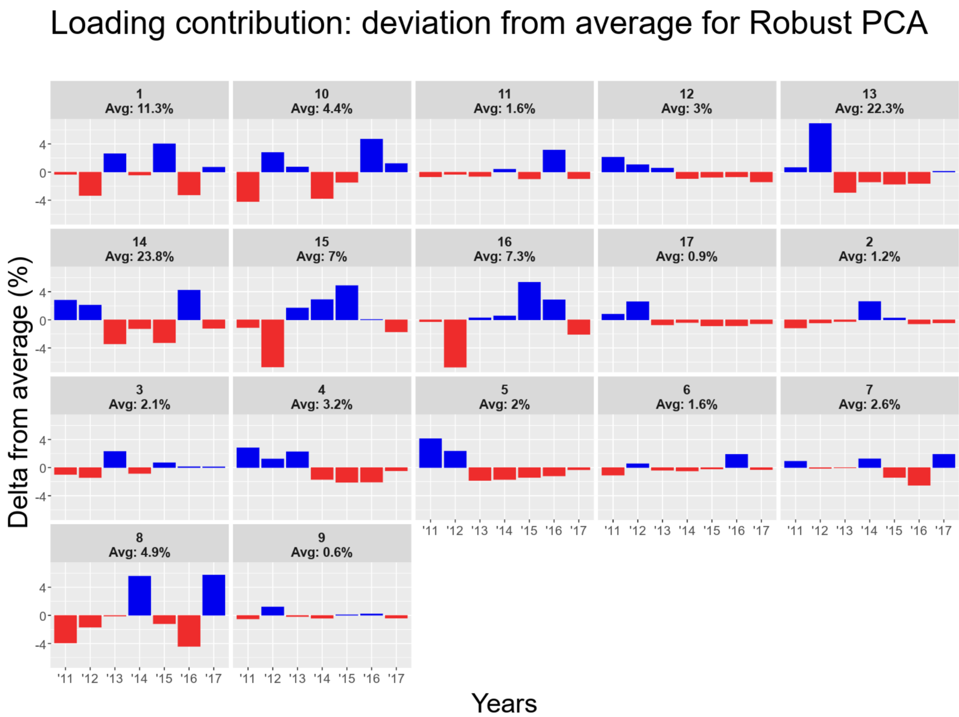
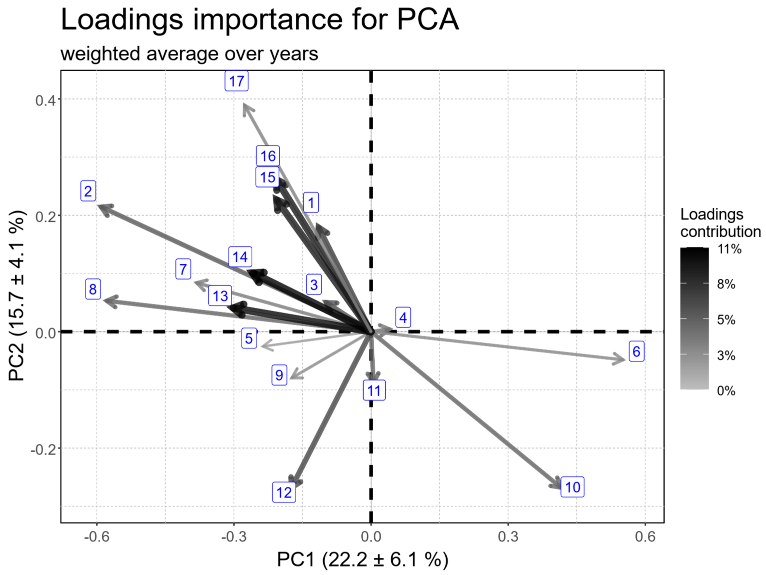
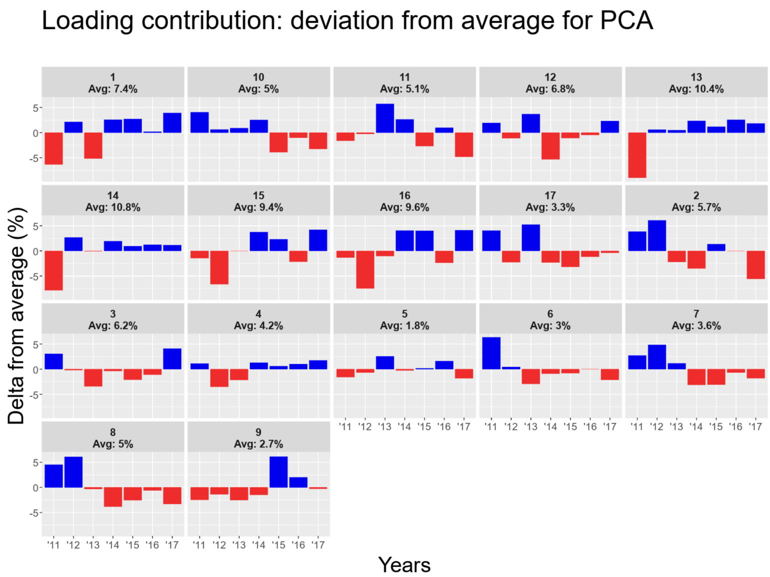


Figure A4: Deviation from average of loading importance for each year. For each original variable and each year the positive or negative difference of loading importance from the weighted average reported in Figure 1.3b. Maximum percentage deviation from the average is 4%. Legend is below: 1 'Emerging 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 forex 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'.

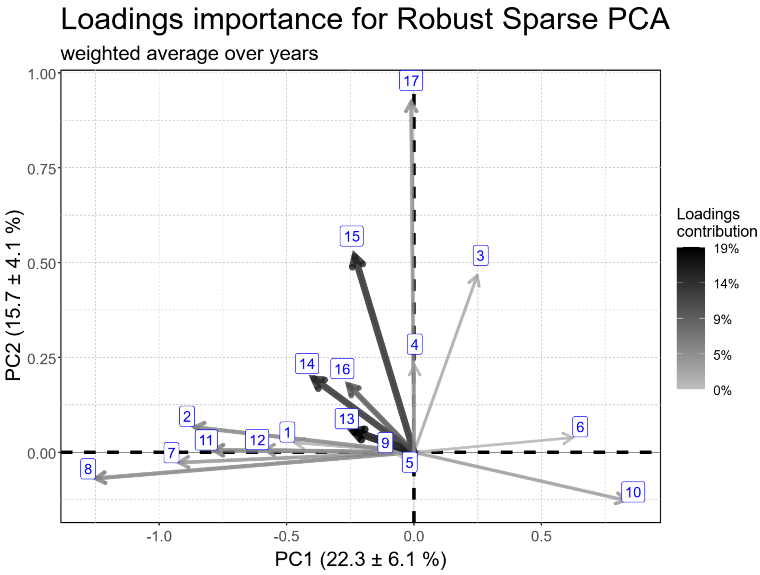


(a) Combined effect of loadings for the first two components for Robust PCA method. Values over years are averaged and weighted by yearly loadings importance. Darker and thicker arrows represent the original variables that contributes the most to loadings importance for each of the two PCA components.

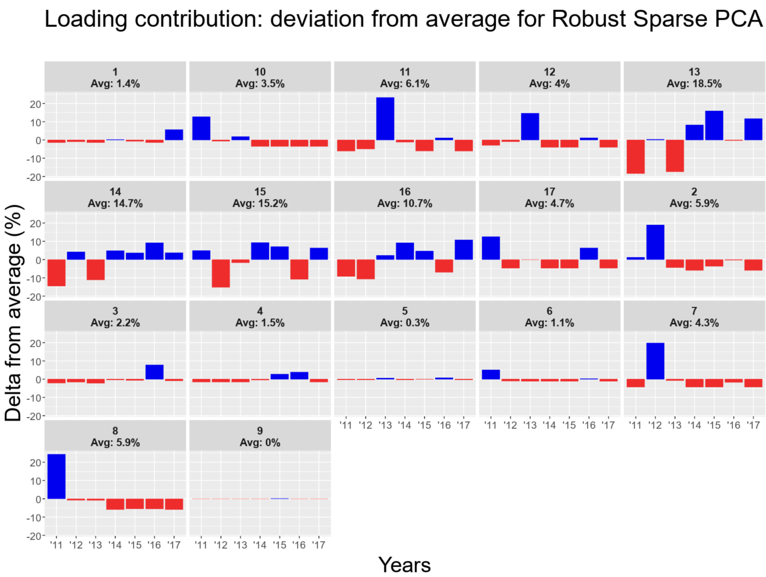


(b) Deviation from average of loading importance for each year. For each original variable and each year the positive or negative difference of loading importance from the weighted average reported in Figure A5a. Maximum percentage deviation from the average is 5%.

Figure A5: 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 forex 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'.



(a) 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.



(b) Deviation from average of loading importance for each year. For each original variable and each year the positive or negative difference of loading importance from the weighted average reported in Figure A6a. Maximum percentage deviation from the average is 20%.

Figure A6: 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 forex 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'.

A.4 Binary FSIND validation

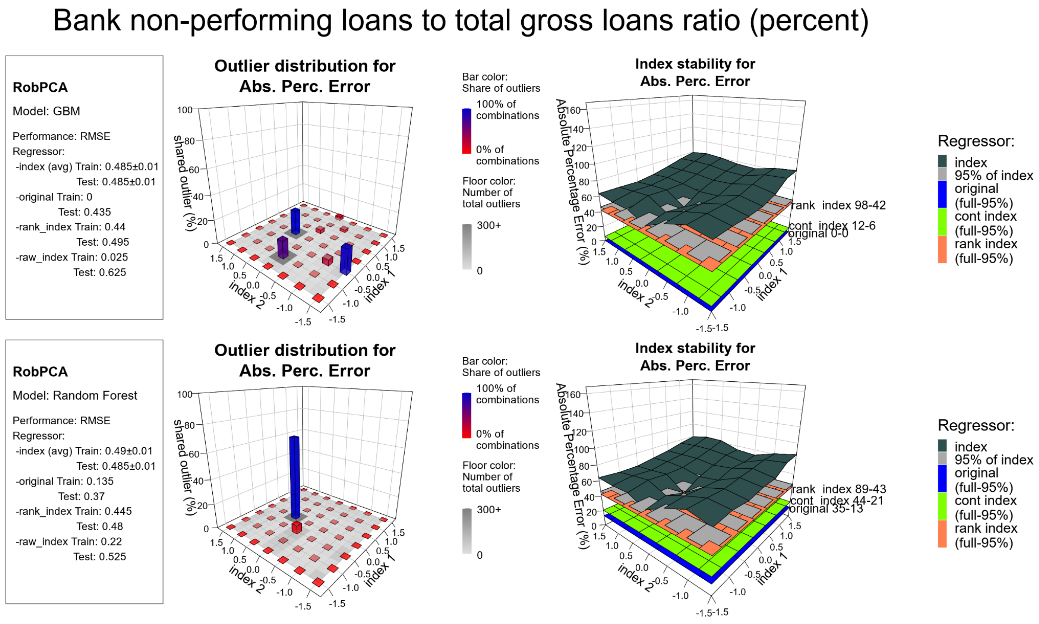


Figure A7: Index validation for bank non-performing loans to total gross loans ratio (percent). The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold 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 grey 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.

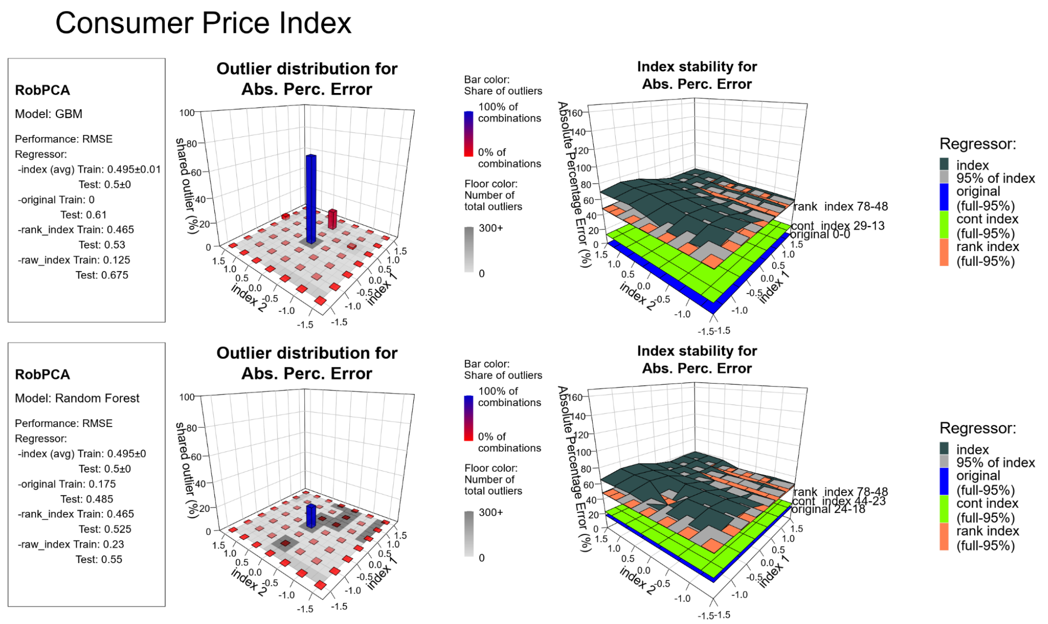


Figure A8: Index validation for Consumer Price Index. The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: floor color shows the 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 grey 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.

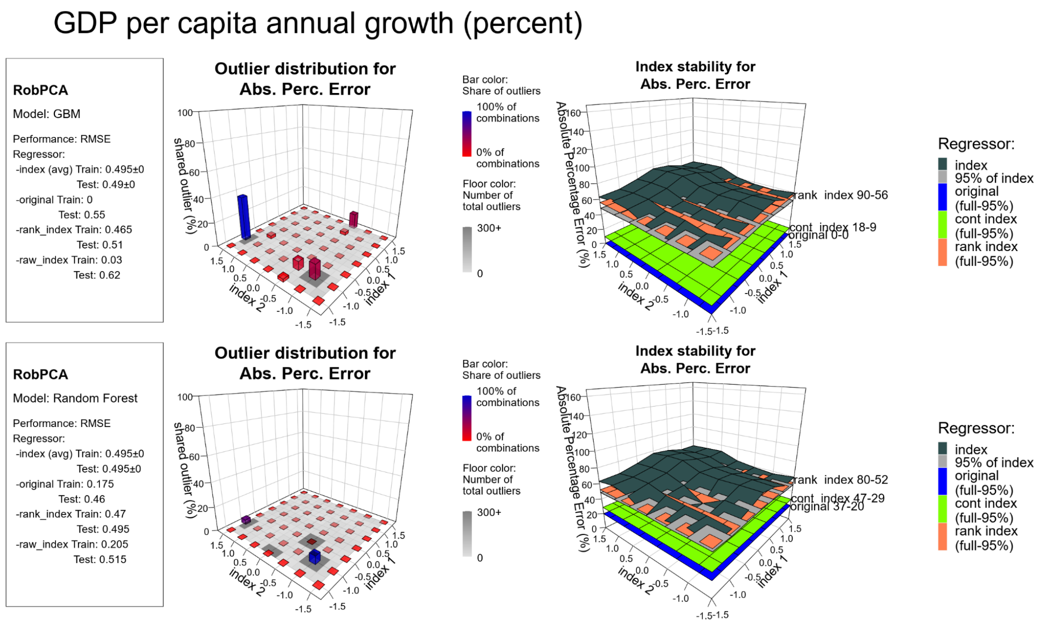


Figure A9: Index validation for GDP per capita annual growth (percent). The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: floor 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 grey 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.

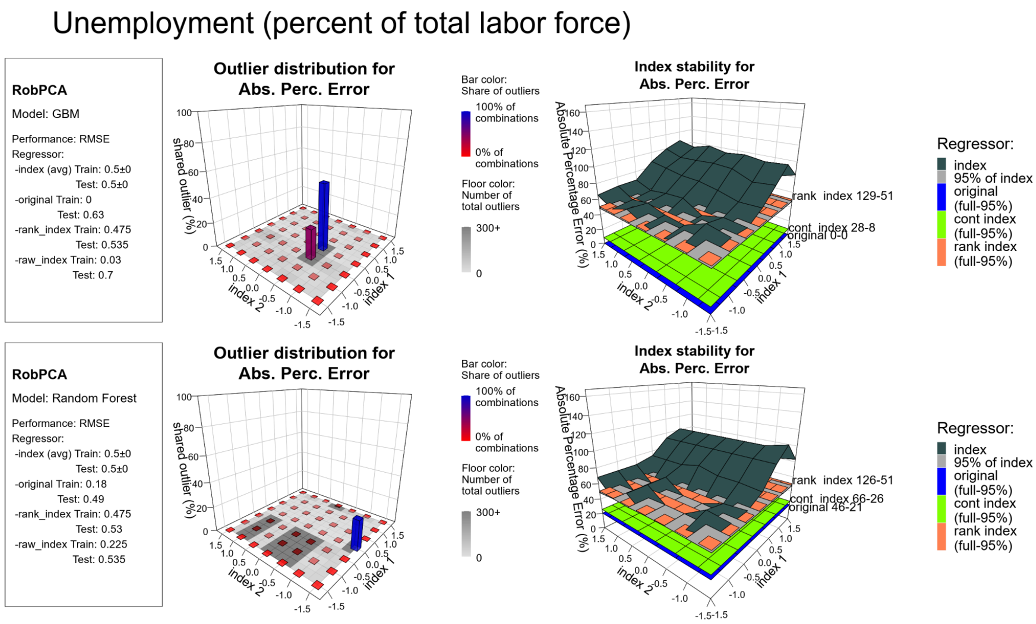


Figure A10: Index validation for unemployment (percent of total labor force). The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: 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 grey 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.

Population annual growth (percent)

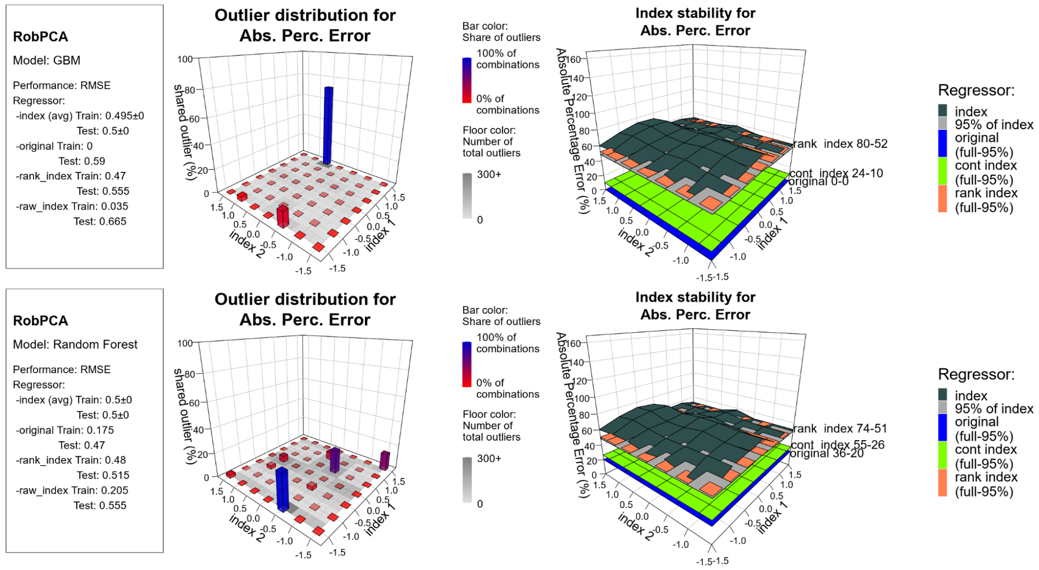


Figure A11: Index validation for population annual growth (percent). The left panel reports models fitting performance, i.e. Root Mean Square Error (RMSE) for train and test set, for all input regressors settings. The central bar plots display outlier stability for all threshold combinations: floor color 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 grey 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.

A.5 FSIND evolution over years

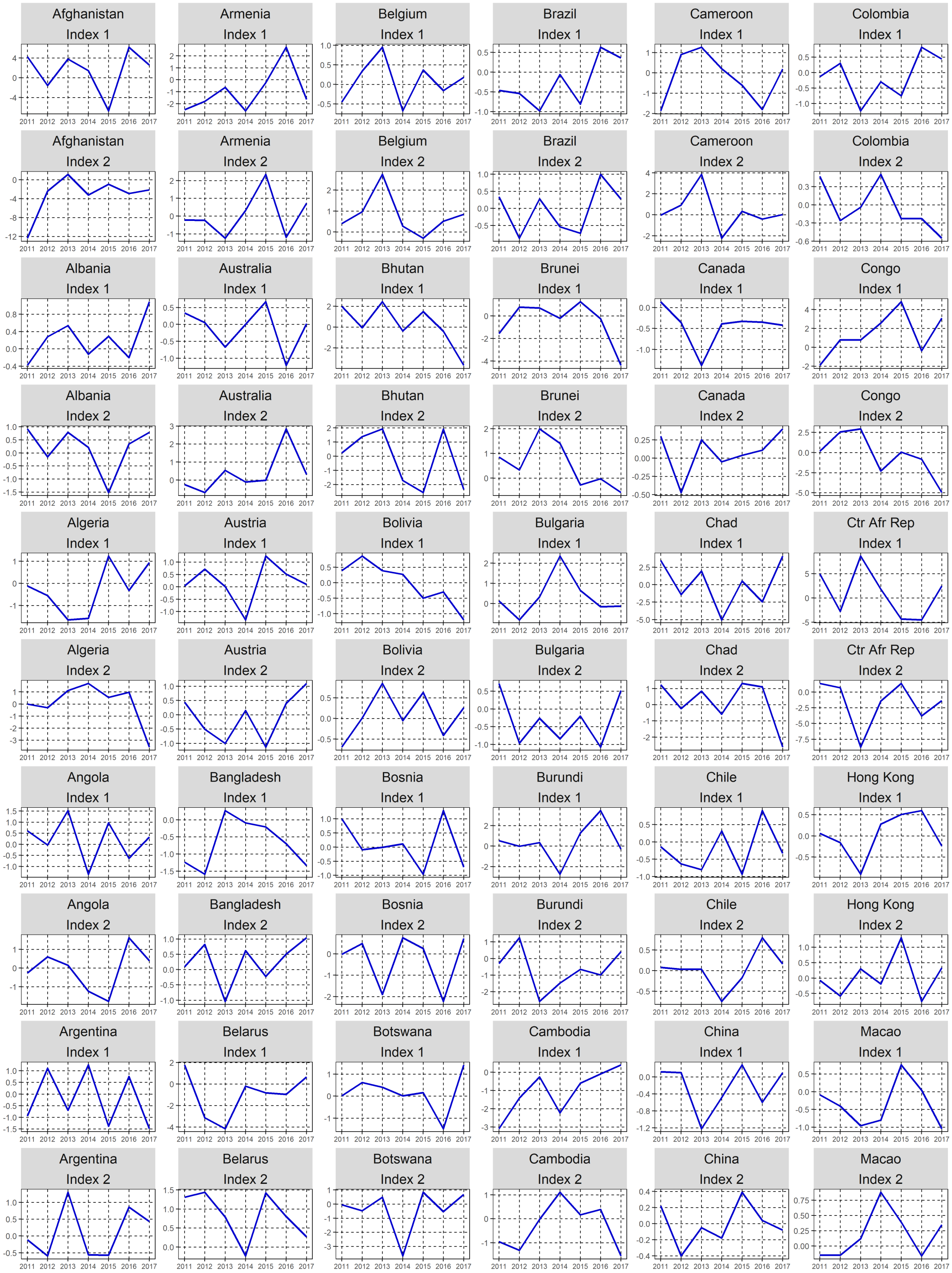


Figure A12: Index evolution over years

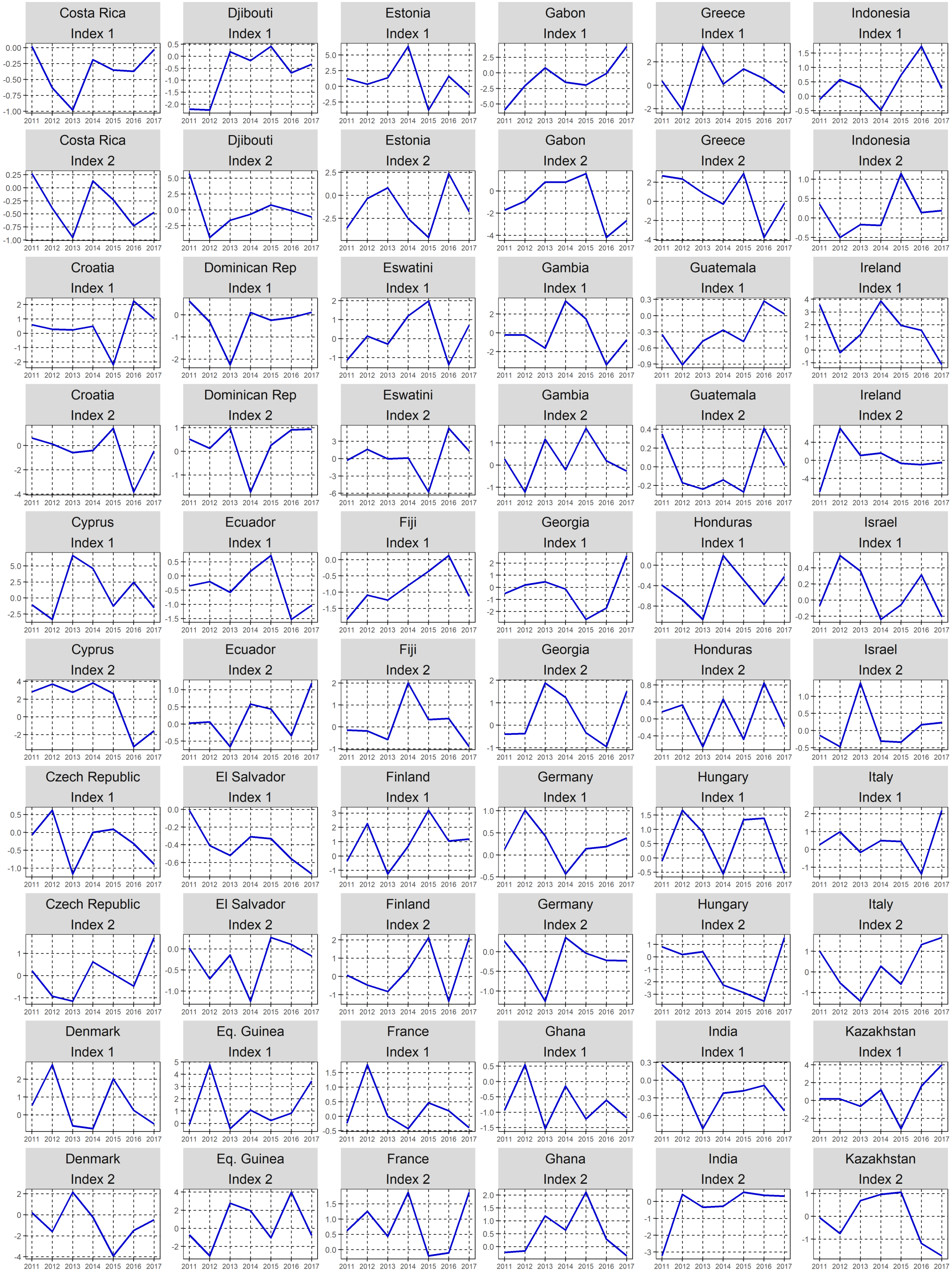


Figure A13: Index evolution over years

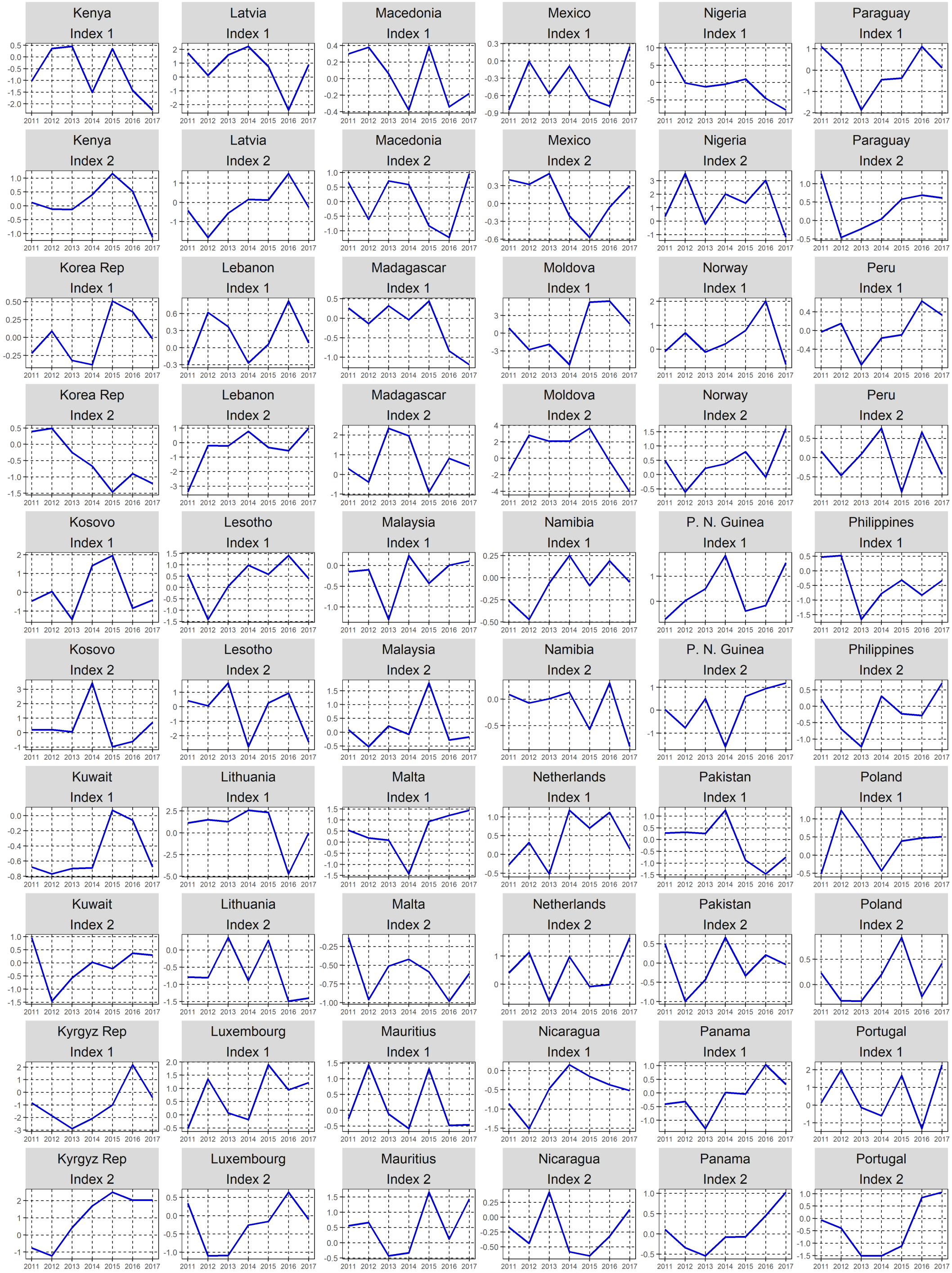


Figure A14: Index evolution over years

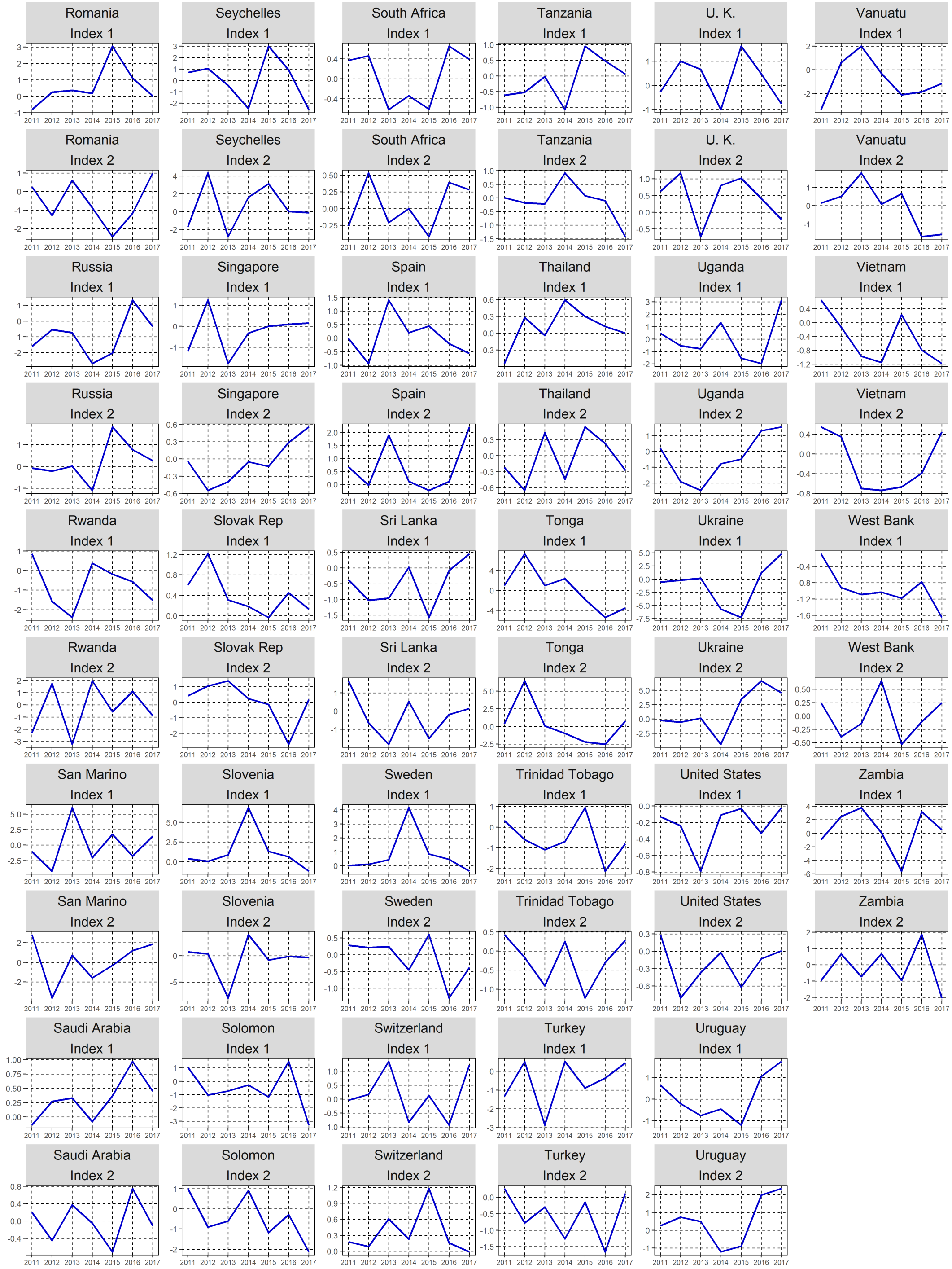


Figure A15: Index evolution over years

A.6 FSIND index comparison

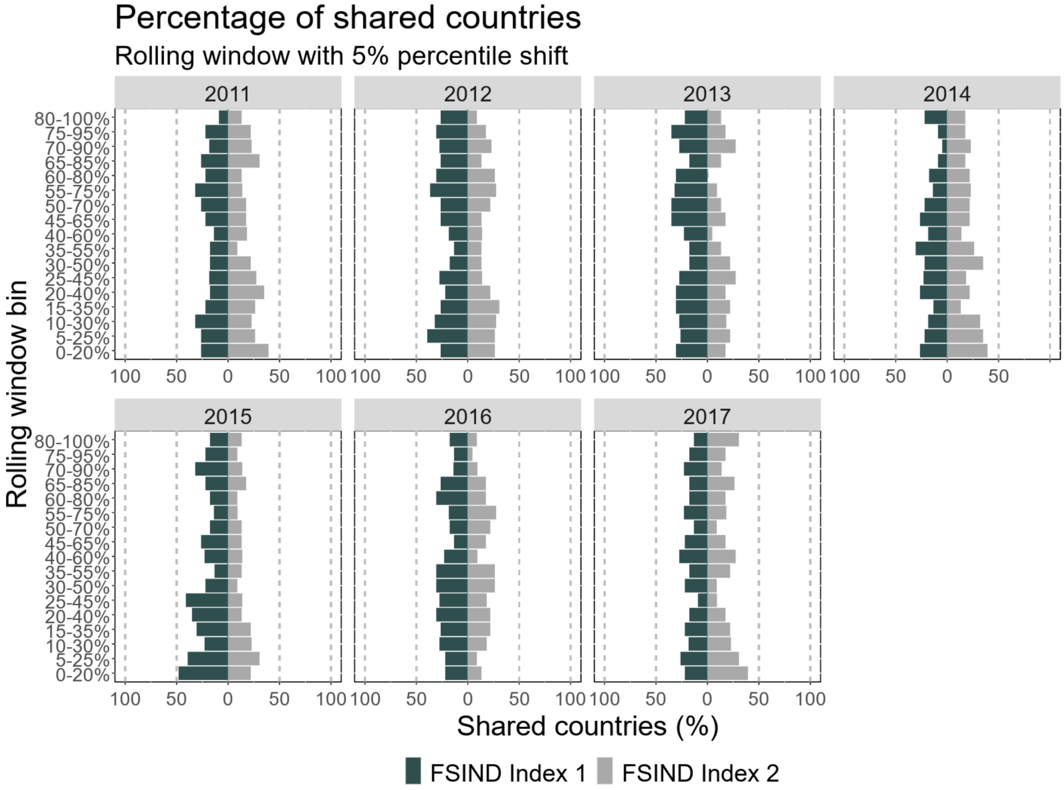


Figure A16: Ranking ability comparison for FSIND components and Financial Institutions Efficiency (FIE) index. For each FSIND component and FIE, 20 quantiles, i.e. one every 5% are evaluated, splitting the index values in 17 rolling bins, each of which spans between five quantiles, i.e. a 20% range, 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 grey) component.

B Do sound financial systems improve the financing constraints of firms? Evidence from developing countries

B.1 FSIND index and ACCESS prediction dataset

B.2 Missing values imputation methodology

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 76 countries by 17 variables for 8 years for a total of 16184 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

Table B1: Correlation matrix of independent variables for DFM index evaluation. Variable Inflation Factor (VIF) is reported below, showing low collinearity between regressors, as well as p-values significance level legend. Variables' 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 forex 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'.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1																
2	0.0071	1															
3	0.0674*	0.164***	1														
4	0.004	0.1305***	0.9271***	1													
5	-0.181***	-0.0464	0.0936**	0.0707*	1												
6	0.0201	0.0352	0.06*	0.0746**	0.3416***	1											
7	0.0451	0.0259	-0.1703***	-0.1903***	-0.1338***	-0.0487	1										
8	0.0508	0.494***	0.1057***	0.0731**	-0.1228***	0.0529	0.2813***	1									
9	-0.0895**	0.2558***	0.2176***	0.2886***	-0.0129	0.0309	0.0724*	0.2063***	1								
10	0.0052	-0.0845**	0.0544	0.0639*	-0.1198***	0.3292***	0.226***	0.0423	-0.0612*	1							
11	-0.1553***	-0.0914***	-0.0662*	-0.0303	0.0787**	-0.0278	-0.0538	-0.0387	-0.0642*	0.0643*	1						
12	0.1221***	0.0074	-0.0431	3e-04	-0.1157***	-0.027	0.0451	0.1241***	-0.0454	0.0949***	0.7802***	1					
13	0.6206***	0.1822***	0.0268	-0.0395	-0.1914***	-0.0562*	0.0655*	0.1964***	-0.1176***	0.037	-0.1335***	0.064*	1				
14	0.6058***	0.2206***	0.0204	-0.0482	-0.2114***	-0.052	0.0893***	0.246***	-0.1135***	0.0641*	-0.0635*	0.1287***	0.9388***	1			
15	0.3928***	0.1213***	-0.1286***	-0.1891***	-0.0812**	-0.0186	-0.0316	0.0208	-0.0183	-0.1454***	-0.4035***	-0.3634***	0.2648***	0.2157***	1		
16	0.0955***	0.0728**	-0.1167***	-0.1639***	-0.0155	0.0189	-0.0194	-0.0227	0.042	-0.1289***	-0.4353***	-0.399***	0.0699**	0.0281	0.875***	1	
17	0.4719***	0.1293***	-0.073*	-0.1446***	-0.091**	0.1896***	0.0191	-0.0532	-0.052	-0.0363	-0.2283***	0.0136	0.1445***	0.1073***	0.3822***	0.2655***	1
VIF	2.654																

* p<0.1, ** p<0.05, *** p<0.01

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. Figure B1 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.

B.3 Loadings Plot for DFM method and FSIND evolution over years.

In this appendix we report loadings of the DFM approach. As described in Section 2.3 the loadings $\mathbf{C}^i$ for the i-th country are stacked into the diagonal matrix $\mathbf{C}$, whereas the cross-country interactions are introduced by the matrix $\hat{\boldsymbol{A}}$ estimated with VAR. Our setting forces the $\mathbf{C}^i$ to be constant so we can estimate loadings for each country-variable pair. Therefore, for ease of visualisation, figure B2 reports the distribution of the loadings for each independent variable over the 76 countries, representing the average trend over the years. The bimodal shape of all distributions implies a clear discriminative power of the index between

Table B2: Complete list of countries for FSIND evaluation and relative missing values count and percentage over total number of observations. The "x" indicates whether the country is matched in the subset used to predict the ACCESS value.

Country	Missing values	ACCESS Dataset	Country	Missing values	ACCESS Dataset	Country	Missing values	ACCESS Dataset
Albania	-	x	Tanzania	-	x	Lesotho	15 (11%)	x
Argentina	-	x	Turkey	-	x	Pakistan	15 (11%)	x
Armenia, Republic of	-	x	Uganda	-	x	Belgium	16 (11.8%)	
Austria	-		Ukraine	-	x	Finland	16 (11.8%)	
Brazil	-		U. K.	-		Kuwait	16 (11.8%)	
Brunei	-		Uruguay	-	x	Nigeria	16 (11.8%)	x
Burundi	-	x	Italy	2 (1.5%)		Singapore	16 (11.8%)	
Cambodia	-	x	Switzerland	3 (2.2%)		India	17 (12.5%)	x
Cameroon	-	x	Cyprus	4 (2.9%)		Korea Rep	17 (12.5%)	
Central African Republic	-	x	Eswatini	4 (2.9%)		Solomon Islands	17 (12.5%)	x
Chad	-		Latvia	4 (2.9%)	x	Honduras	18 (13.2%)	x
Macao	-		Seychelles	4 (2.9%)		Netherlands	18 (13.2%)	
Congo	-		Colombia	5 (3.7%)	x	Chile	20 (14.7%)	x
Croatia	-	x	Hong Kong	6 (4.4%)		Lebanon	22 (16.2%)	x
Denmark	-		Fiji	6 (4.4%)		Algeria	23 (16.9%)	
El Salvador	-	x	Kenya	6 (4.4%)	x	Australia	24 (17.6%)	
Eq. Guinea	-		Tonga	6 (4.4%)		Moldova	24 (17.6%)	x
Gabon	-		Vanuatu	6 (4.4%)		Panama	24 (17.6%)	x
Georgia	-	x	Ghana	7 (5.1%)	x	San Marino	24 (17.6%)	
Germany	-		Bolivia	8 (5.9%)	x	Spain	24 (17.6%)	
Guatemala	-	x	Bosnia and Herzegovina	8 (5.9%)	x	Thailand	24 (17.6%)	x
Indonesia	-	x	Canada	8 (5.9%)		United States	24 (17.6%)	
Kazakhstan	-	x	Czech Republic	8 (5.9%)	x	Vietnam	24 (17.6%)	x
Kyrgyz Republic	-	x	Dominican Republic	8 (5.9%)	x	Sri Lanka	31 (22.8%)	x
Macedonia, FYR	-	x	Greece	8 (5.9%)		China, P.R.: Mainland	32 (23.5%)	x
Madagascar	-	x	Kosovo, Republic of	8 (5.9%)	x	Costa Rica	32 (23.5%)	x
Malta	-		Luxembourg	8 (5.9%)		Ecuador	32 (23.5%)	x
Mauritius	-		Paraguay	8 (5.9%)	x	Malaysia	32 (23.5%)	x
Namibia	-	x	Portugal	8 (5.9%)		Angola	34 (25%)	x
Nicaragua	-	x	Trinidad and Tobago	8 (5.9%)	x	Botswana	34 (25%)	x
Papua New Guinea	-	x	West Bank and Gaza	8 (5.9%)	x	Gambia	34 (25%)	
Peru	-	x	Zambia	8 (5.9%)	x	Bangladesh	36 (26.5%)	x
Philippines	-	x	Bulgaria	10 (7.4%)	x	France	36 (26.5%)	
Poland	-	x	Lithuania	10 (7.4%)	x	Ireland	37 (27.2%)	
Romania	-	x	Estonia	12 (8.8%)	x	Djibouti	40 (29.4%)	x
Russian Federation	-	x	Mexico	12 (8.8%)	x	Hungary	40 (29.4%)	x
Rwanda	-	x	Afghanistan, Islamic Republic of	13 (9.6%)	x	Norway	40 (29.4%)	
Saudi Arabia	-		Bhutan	13 (9.6%)	x	Slovenia	40 (29.4%)	x
Slovak Republic	-	x	Belarus	14 (10.3%)	x	Sweden	40 (29.4%)	x
South Africa	-		Israel	14 (10.3%)	x			

Table B3: Correlation matrix of variable used to predict ACCESS. Variable Inflation Factor (VIF) is reported below, showing very low collinearity between regressors, as well as p-values significance level legend.

	ACCESS	FSIND	LISTED	AGE	SIZE	SUBSID	LOCATION	EXPORT	OWNFOR	OWNGOV
ACCESS	1									
FSIND	-0.0268***	1								
LISTED	-0.0315***	0.0052	1							
AGE	-0.0253***	0.0095**	0.0864***	1						
SIZE	-0.0903***	-0.0591***	0.1246***	0.2399***	1					
SUBSID	-0.0288***	-0.0426***	0.0982***	0.069***	0.1952***	1				
LOCATION	-0.0341***	0.1188***	-0.0594***	-0.1033***	-0.101***	-0.0192***	1			
EXPORT	-0.0451***	-0.0017	0.033***	0.0532***	0.2371***	0.079***	0.0297***	1		
OWNFOR	-0.046***	-0.0025	0.0811***	0.0154***	0.1672***	0.1334***	-0.0241***	0.1894***	1	
OWNGOV	-0.0234***	-0.0294***	0.057***	0.0207***	0.0498***	0.0282***	-0.003	0.0182***	-0.0008	1
VIF	1.064									

* p<0.1, ** p<0.05, *** p<0.01

less risky countries and riskier ones. Figure B3, instead, reports the contribution of independent variables on the loading for each country. Blue shaded points represent the positive contribution of the variables to each loading while red shaded points represent the negative one. The bigger the points the more the independent variable contributes to the loading.

Table B4: List of variables used to build the FSIND index, with sources, aggregation level, total number of observations and descriptive summary statistics.

Variable	Source	Aggregation Level	Obs	Mean	S.D.	Min	P25	Median	P75	Max
1 - EMB Capital to assets (%)	FSI	Country	1,127	10.28	3.57	1.49	7.57	10.02	12.37	24.85
2 - EMB Customer deposits to total non interbank loans (%)			1,077	120.73	56.5	29.01	89.3	111.71	131.83	626.93
3 - EMB Foreign currency liabilities to total liabilities (%)			997	30.61	24.87	0	10.18	23.96	49.26	100
4 - EMB Foreign currency loans to total loans (%)			1,014	28.75	26.26	0	8.03	22.7	43.79	100.06
5 - EMB Personnel expenses to non interest expenses (%)			1,097	44.17	12.04	5.29	36.8	44.03	51.14	91.58
6 - Interest margin to gross income (%)			1,169	59.01	18.4	-294.33	51.58	60.4	68.81	142.77
7 - Liquid assets to short term liabilities (%)			1,111	69.13	61.11	10	34.58	48.99	78.71	690.37
8 - Liquid assets to total assets (%)			1,140	27.92	13.03	4.99	18.82	25.77	33.77	74.97
9 - Net open position of forex to capital (%)			969	9.57	36.74	-95.43	0.14	2.67	8.66	407.97
10 - Non interest expenses to total income (%)			1,169	58.17	17.88	-303.46	49.57	57.14	66.34	115.79
11 - Non performing loans net of capital provisions (%)			1,169	18.78	38.28	-51.61	3.64	9.08	20.38	413.56
12 - Non performing loans to total gross loans (%)			1,167	6.81	7.4	0	2.22	4.05	9.31	54.54
13 - Regulatory capital to risk weighted assets (%)			1,171	17.67	4.83	1.75	14.67	16.83	19.3	42.2
14 - Regulatory tier 1 capital to risk weighted assets (%)			1,166	15.43	4.86	2.18	12.3	14.39	17.31	40.3
15 - Return on assets (%)			1,169	1.5	1.8	-25.61	0.76	1.38	2.24	10.28
16 - Return on equity (%)			1,166	13.22	21.93	-505.64	8.18	14.05	20.34	65.4
17 - Sectoral distribution of loans residents (%)			1,063	87.85	16.05	20.67	83.32	94.9	99.25	100

Table B5: Results for DFM methods with different number of factors and factors interactions. R^2 is reported for the full dataset and for the 99$^{\text{th}}$ and 95$^{\text{th}}$ percentiles. We also report Im-Pesaran-Shin test for stationarity on the FSIND index.

Factors Interactions	Number of Factors	R^2	R^2 on 99$^{\text{th}}$	R^2 on 95$^{\text{th}}$	Im-Pesaran-Shin test
No	1	35.7%	36.5%	39.4%	$\ll 0.01$
No	2	39.9%	42.9%	44.3%	$\ll 0.01$
Yes	1	64.1%	66.5%	69.7%	$\ll 0.01$
Yes	2	66.4%	67.7%	70.3%	$\ll 0.01$

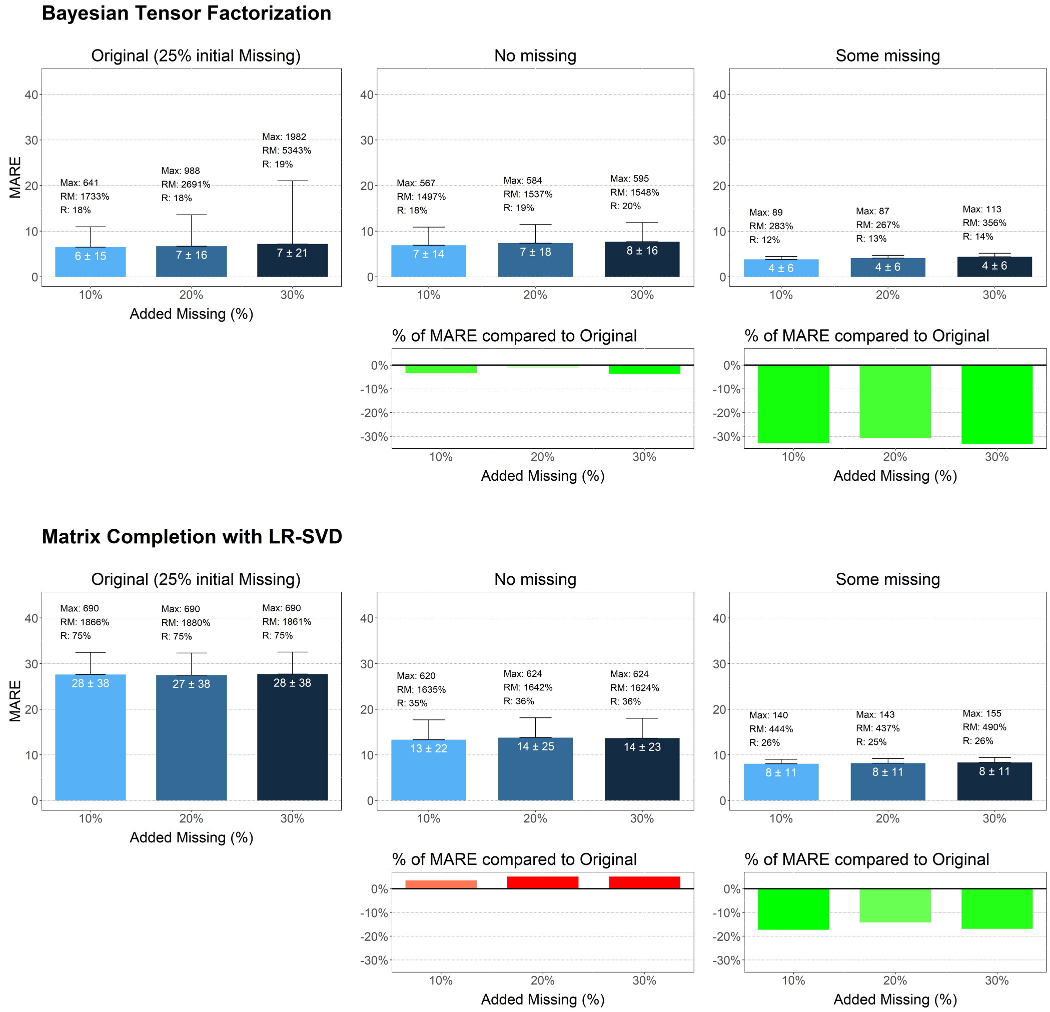


Figure B1: Missing values imputation methodologies are tested in three settings. In the first (named *Original*, setting a) the whole dataset contains 8% of missing values, and additional values representing 10%, 20% and 30% of the initial dataset are randomly removed. In the second (named *No missing*, setting b) all entries with missing values are dropped from the whole dataset and the same incremental sampling procedure is applied on the remaining subset. In the last (named *Some missing*, setting c) all countries with at least 3 missing values for any year are dropped and the incremental sampling procedure is again applied on the remaining subset. The blue shaded bars in the upper row represent the average percentage reconstruction error (MAPE) for different percentage of additional missing values for each of the three settings. Whiskers on the top of each bar shows the scaled magnitude of maximum value of reconstruction error as well its numeric value, the relative magnitude R of MAPE compared to the average value of the original dataset and the relative magnitude RM of the maximum percentage reconstruction error compared to the average value of the original dataset. Bars on the lower row represent the percentage variation of the average reconstruction error of b) *No missing* and c) *Some missing* settings compared to a) *Original* setting, i. e. $MARE\,/\,MARE_{\text{Orig}} - 1$. Green bars mean that the imputation technique has a lower average reconstruction error when applied on the subset with no original missing values, setting b, and on the subset with some original missing values, setting c, compared to the average reconstruction error when applied on the full dataset with all original missing values, setting a.

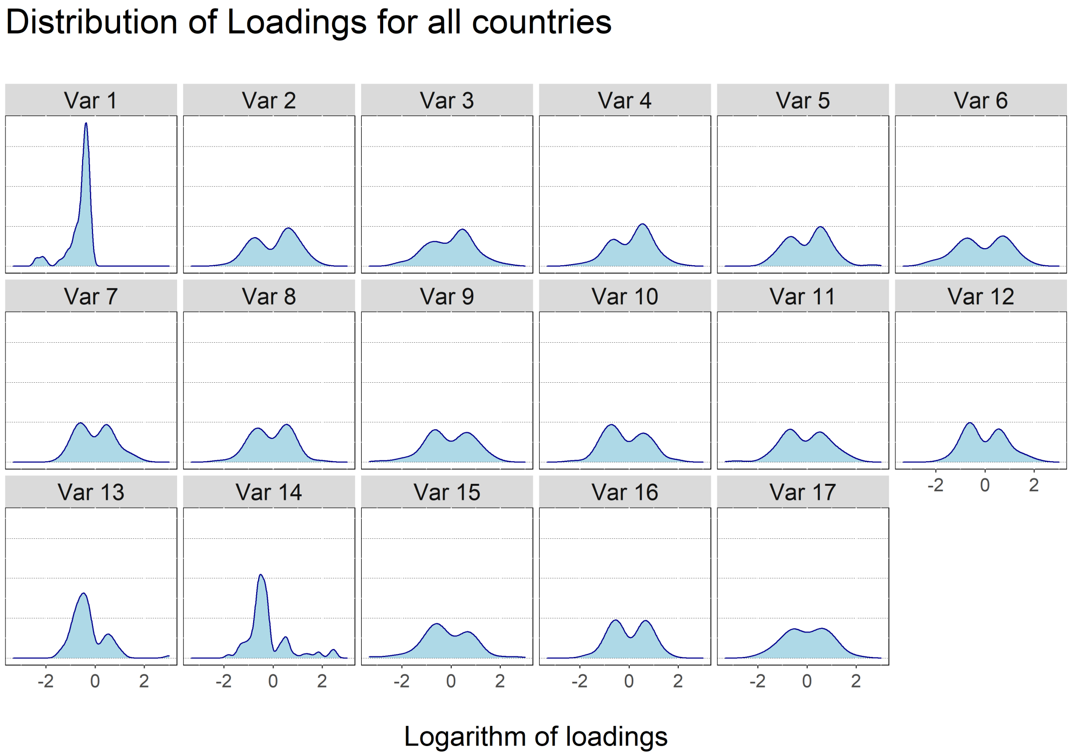


Figure B2: Loadings distribution over all countries for each independent variable. On x-axis is reported the logarithm of loading values. Variables' legend is below:
1 'Emerging Markets Bond (EMB) Capital to assets', 2 'Customer deposits to total non interbank loans', 3 'EMB Foreign currency liabilities 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 forex 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'.

B.4 Robustness test for ACCESS prediction models

C A data-driven approach to measuring epidemiological susceptibility risk around the world

C.1 List of variables and countries

Table C1: List of used variable. Sources are World Health Organization (WHO), World Bank's Development Indicators (WDI), Penn Tables (PT) and World Bank's Worldwide Governance Indicators (WGI).

Variable	Description	Source	Total Obs.	Missing Values	Min	Max	Mean	Median	Standard Deviation
var1	health care exenditure per capita	WHO	1,680	523 (31%)	12.64	10,014.71	1,077.66	317.86	1,821.28
var2	health care access and quality	WHO	1,680	20 (1.2%)	28.60	93.60	62.97	62.55	16.49
var3	response level (%) to public health hazards	WHO	1,680	670 (40%)	0.00	100.00	66.15	73.00	30.61
var4	num of physicians per 1000 people	WDI	1,680	941 (56%)	0.00	6.11	2.01	2.05	1.42
var5	num of hospital beds per 1000 people	WDI	1,680	1175 (70%)	0.10	13.40	3.25	2.70	2.31
var6	num of air passengers to population ratio	WDI	1,680	397 (24%)	0.00	34.53	1.18	0.29	2.87
var7	num of urban pop (% of total)	WDI	1,680	168 (10%)	10.64	100.00	58.91	59.48	22.19
var8	num of people per Km2 (pop density)	WDI	1,680	177 (11%)	1.75	7,953.00	231.38	81.13	808.73
var9	num of people age 65% (% of total)	WDI	1,680	177 (11%)	0.69	27.58	8.38	6.20	5.93
var10	num of people using drinking water services (% of pop)	WDI	1,680	340 (20%)	33.05	100.00	86.51	94.72	16.70
	num of people using safely-managed								

Figure B3: Contribution of independent variables on loadings for all countries. Blue shaded points represent the positive contribution of the variables to each loading while red shaded points represent the negative one. The bigger the points the more the independent variable contributes to the loading. Variables' 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 forex 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'.

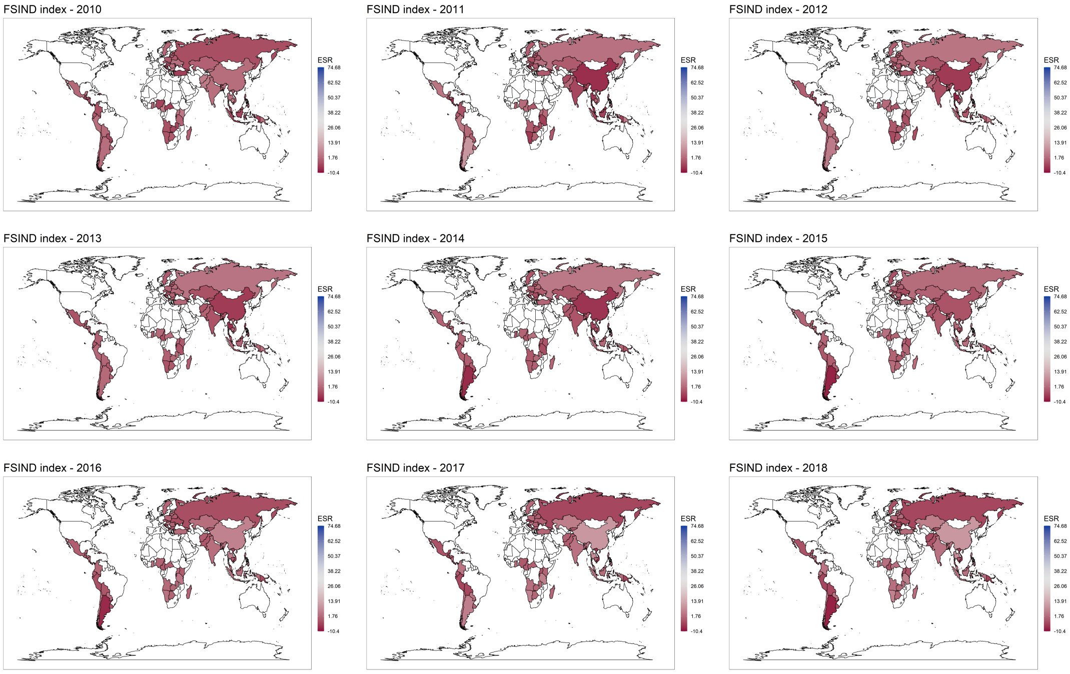


Figure B4: FSIND index evolution over years. Shades of red color refer to riskier countries, while shades of blue to safer ones.

Table C2: Correlation matrix of input variables.
var1 is health care exenditure per capita, var2 is health care access and quality, var3 is response level (%) to public health hazards, var4 is num of physicians per 1000 people, var5 is num of hospital beds per 1000 people, var6 is num of air passengers to population ratio, var7 is num of urban pop (% of total), var8 is num of people per Km2 (pop density), var9 is num of people age 65% (% of total), var10 is num of people using drinking water services (% of pop), var11 is num of people using safely-managed drinking water services (% of pop), var12 is num of people using safely-managed sanitation services (% of pop), var13 is human capital index, var14 is num of people using the internet (% of pop), var15 is value of trade (% GDP), var16 is government effectiveness index, var17 is rule of law index.

	var1	var2	var3	var4	var5	var6	var7	var8	var9	var10	var11	var12	var13	var14	var15	var16
var2	0.66*															
var3	0.35*	0.5*														
var4	0.57*	0.75*	0.44*													
var5	0.32*	0.52*	0.27*	0.64*												
var6	0.33*	0.32*	0.17*	0.21*	0.01											
var7	0.48*	0.7*	0.39*	0.58*	0.31*	0.22*										
var8	-0.04	0.15*	0.13*	-0.04	-0.02	0.2*	0.19*									
var9	0.61*	0.79*	0.38*	0.76*	0.67*	0.13*	0.46*	0.07*								
var10	0.41*	0.79*	0.42*	0.65*	0.37*	0.22*	0.64*	0.12*	0.61*							
var11	0.43*	0.53*	0.22*	0.5*	0.31*	0.26*	0.5*	0.14*	0.49*	0.36*						
var12	0.38*	0.29*	0.22*	0.27*	0.28*	0.16*	0.29*	0.14*	0.27*	-0.09*	0.72*					
var13	0.53*	0.67*	0.43*	0.54*	0.46*	0.16*	0.5*	0.12*	0.62*	0.57*	0.35*	0.34*				
var14	0.63*	0.86*	0.51*	0.66*	0.48*	0.36*	0.69*	0.15*	0.69*	0.73*	0.53*	0.31*	0.6*			
var15	0.12*	0.31*	0.06*	0.16*	0.19*	0.32*	0.29*	0.55*	0.23*	0.26*	0.2*	0.14*	0.19*	0.32*		
var16	0.71*	0.81*	0.47*	0.6*	0.39*	0.36*	0.58*	0.24*	0.72*	0.66*	0.4*	0.28*	0.63*	0.79*	0.37*	
var17	0.73*	0.76*	0.42*	0.57*	0.37*	0.38*	0.53*	0.22*	0.69*	0.58*	0.41*	0.32*	0.56*	0.75*	0.37*	0.95*

$*$ p-val < 0.05

Table B6: Predicting ACCESS with ordinal probit model with instrumental variables and macro-economic controls - CMP.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.1119***	-0.0335**	-0.0312***	-0.0118*	-0.091***	-0.074**	-0.0897***	-0.0298*
	(0.0169)	(0.0181)	(0.0169)	(0.0181)	(0.0170)	(0.0182)	(0.0169)	(0.0181)
LISTED	-0.0757***	-0.039***	-0.0134***	-0.0423***	-0.0884***	-0.027***	-0.0536***	-0.0169***
	(0.00821)	(0.00821)	(0.00825)	(0.00825)	(0.00822)	(0.00822)	(0.00824)	(0.00824)
AGE	-0.1875***	-0.3031***	-0.389***	-0.1519***	-0.1283***	-0.4007***	-0.1875***	-0.0183***
	(0.0301)	(0.0301)	(0.0309)	(0.0309)	(0.0300)	(0.0301)	(0.0309)	(0.0309)
SUBSID	-0.011***	-0.0055***	-0.0363***	-0.001***	-0.0228***	-0.0096***	-9e-04***	-0.0064***
	(0.00414)	(0.00414)	(0.00418)	(0.00419)	(0.00413)	(0.00414)	(0.00419)	(0.00419)
LOCATION	0.0044***	0.103***	0.0546***	0.0595***	0.0396***	0.0555***	0.0216***	0.0057***
	(0.00744)	(0.00756)	(0.00745)	(0.00757)	(0.00745)	(0.00757)	(0.00745)	(0.00757)
EXPORT	-0.0558***	-0.0906***	-0.0481***	-0.0324***	-0.0172***	-0.0119***	-0.0312***	-0.0693***
	(0.00709)	(0.00709)	(0.00718)	(0.00718)	(0.00703)	(0.00703)	(0.00723)	(0.00723)
OWNFOR	-0.0499***	-0.1327***	-0.1206***	-0.0423***	-0.0158***	-0.0212***	-0.1576***	-0.0818***
	(0.00645)	(0.00645)	(0.00652)	(0.00652)	(0.00646)	(0.00646)	(0.00651)	(0.00651)
OWNGOV	-0.2248***	-0.1236***	-0.0363***	-0.1805***	-0.1278***	-0.0722***	-0.0851***	-0.1911***
	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)
GDPCAP		0.0258		0.0471		0.0725		0.0632
		(0.0334)		(0.0335)		(0.0335)		(0.0334)
INFLDFL		0.081		0.0626		0.0132		0.0712
		(0.0343)		(0.0343)		(0.0343)		(0.0343)
LENDINT		-0.5656*		-0.1841		-0.2102		-0.0901*
		(0.136)		(0.136)		(0.136)		(0.136)
NUMBRW	18.0828***	12.9832***	4.2844***	13.5466***	11.9256***	1.9514***	13.7929***	2.4601***
	(0.142)	(0.112)	(0.142)	(0.112)	(0.142)	(0.112)	(0.142)	(0.112)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
F-stat	96.88	93.11	95.24	91.57	96.83	93	95.59	91.97
ρ	0.0137*	0.0089	0.0125*	0.0082	0.0125*	0.0086	0.0138*	0.0084
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and Conditional Mixed Process for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Arellano-Bond ρ indicates the magnitude of auto-correlation and its significance level. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

Table C3: Complete list of selected countries and relative missing values count and percentage over total number of observations.

Country	Missing Values	Country	Missing Values	Country	Missing Values
Nigeria	91 (6.7%)	Philippines	276 (20.3%)	Slovak Republic	336 (24.7%)
Sri Lanka	92 (6.8%)	Costa Rica	277 (20.4%)	Latvia	337 (24.8%)
Armenia	103 (7.6%)	St. Vincent and the Grenadines	278 (20.4%)	Serbia	337 (24.8%)
Lao PDR	105 (7.7%)	Mali	280 (20.6%)	Spain	337 (24.8%)
Mongolia	106 (7.8%)	Yemen, Rep.	280 (20.6%)	Austria	341 (25.1%)
Bolivia	113 (8.3%)	Guinea-Bissau	283 (20.8%)	Trinidad and Tobago	342 (25.1%)
Honduras	117 (8.6%)	China	285 (21%)	Belgium	347 (25.5%)
Moldova	122 (9%)	Indonesia	285 (21%)	Tunisia	347 (25.5%)
Nicaragua	123 (9%)	Liberia	286 (21%)	Eswatini	348 (25.6%)
Sierra Leone	129 (9.5%)	Croatia	288 (21.2%)	Romania	350 (25.7%)
Tanzania	130 (9.6%)	Ecuador	289 (21.2%)	Qatar	354 (26%)
Mauritania	134 (9.9%)	Malaysia	288 (21.2%)	Mauritius	355 (26.1%)
Benin	138 (10.1%)	Chile	292 (21.5%)	Kazakhstan	357 (26.2%)
India	139 (10.2%)	Hungary	292 (21.5%)	Bulgaria	359 (26.4%)
Kenya	142 (10.4%)	Singapore	293 (21.5%)	Malta	359 (26.4%)
Togo	141 (10.4%)	Djibouti	296 (21.8%)	Fiji	362 (26.6%)
Cote d'Ivoire	146 (10.7%)	Malawi	298 (21.9%)	Turkey	362 (26.6%)
Cameroon	150 (11%)	Sweden	300 (22.1%)	Luxembourg	363 (26.7%)
Burundi	151 (11.1%)	Peru	302 (22.2%)	Uruguay	365 (26.8%)
Mozambique	151 (11.1%)	Egypt, Arab Rep.	303 (22.3%)	Ukraine	367 (27%)
Tajikistan	152 (11.2%)	Brazil	305 (22.4%)	Finland	369 (27.1%)
Georgia	159 (11.7%)	South Africa	305 (22.4%)	Botswana	373 (27.4%)
Burkina Faso	163 (12%)	Thailand	304 (22.4%)	Denmark	372 (27.4%)
Niger	163 (12%)	Iran, Islamic Rep.	310 (22.8%)	Lebanon	372 (27.4%)
Bangladesh	165 (12.1%)	Switzerland	310 (22.8%)	Israel	375 (27.6%)
Angola	169 (12.4%)	Dominica	311 (22.9%)	Oman	376 (27.6%)
Rwanda	169 (12.4%)	Canada	313 (23%)	Portugal	376 (27.6%)
Zimbabwe	168 (12.4%)	Lithuania	313 (23%)	Norway	377 (27.7%)
Sudan	170 (12.5%)	Argentina	316 (23.2%)	Saudi Arabia	379 (27.9%)
Vietnam	170 (12.5%)	Jordan	315 (23.2%)	Germany	381 (28%)
Senegal	175 (12.9%)	Uzbekistan	315 (23.2%)	Iceland	382 (28.1%)

Table B7: Predicting ACCESS with ordinal probit model with instrumental variables and financial access controls - CMP.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.1197***	-0.0914***	-0.1086***	-0.1072***	-0.1111***	-0.0795***	-0.0147***	-0.0453***
	(0.0169)	(0.0213)	(0.0169)	(0.0213)	(0.0170)	(0.0213)	(0.0169)	(0.0213)
LISTED	-0.0208***	-0.0539***	-0.0157***	-0.0382***	-0.0475***	-0.0333***	-0.0167***	-0.0294***
	(0.00821)	(0.00821)	(0.00825)	(0.00826)	(0.00822)	(0.00822)	(0.00824)	(0.00825)
AGE	-0.1184***	-0.2736***	-0.1316***	-0.3285***	-0.0306***	-0.0458***	-0.0304***	-0.2891***
	(0.0301)	(0.0301)	(0.0309)	(0.0309)	(0.0300)	(0.0301)	(0.0309)	(0.0309)
SUBSID	-0.007***	-0.0182***	-0.0135***	-0.0057***	-0.043***	-0.0223***	-0.0255***	-0.0066***
	(0.00414)	(0.00415)	(0.00418)	(0.00420)	(0.00413)	(0.00414)	(0.00419)	(0.00420)
LOCATION	0.0879***	0.0246***	0.1071***	0.1262***	0.0861***	0.0628***	0.098***	0.0118***
	(0.00744)	(0.00747)	(0.00745)	(0.00748)	(0.00745)	(0.00748)	(0.00745)	(0.00747)
EXPORT	-0.0479***	-0.0045***	-0.0148***	-0.0318***	-0.0286***	-0.0747***	-0.0494***	-0.0415***
	(0.00709)	(0.00710)	(0.00718)	(0.00718)	(0.00703)	(0.00703)	(0.00723)	(0.00723)
OWNFOR	-0.0191***	-0.1141***	-0.0142***	-0.1223***	-0.1403***	-0.0774***	-0.1082***	-0.0324***
	(0.00645)	(0.00645)	(0.00652)	(0.00652)	(0.00646)	(0.00646)	(0.00651)	(0.00651)
OWNGOV	-0.221***	-0.2404***	-0.1825***	-0.1875***	-0.0954***	-0.1727***	-0.0769***	-0.0414***
	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)
FININD		-0.3638		-0.1719		-0.3295		-0.2239
		(0.348)		(0.348)		(0.348)		(0.348)
FINDEP		-0.2076		-0.1764		-0.1506		-0.1791
		(0.250)		(0.251)		(0.251)		(0.250)
OUTLOAN		1281.4853		1189.7615		812.5037		388.4608
		(540.2)		(540.4)		(540.7)		(539.7)
NUMBRW	2.0887***	5.9549***	17.4143***	2.3762***	7.4528***	7.386***	5.7145***	10.9275***
	(0.142)	(0.120)	(0.142)	(0.120)	(0.142)	(0.120)	(0.142)	(0.120)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
F-stat	96.88	93.08	95.24	91.55	96.83	92.98	95.59	91.94
ρ	0.0137*	0.0159**	0.0125*	0.0147*	0.0125*	0.0145*	0.0138*	0.0161**
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and Conditional Mixed Process for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Arellano-Bond ρ indicates the magnitude of auto-correlation and its significance level. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

In table C4 we report the distribution over time of the missing values quota, as to evaluate the impact of missing data imputation. It clearly emerges the highest quota for the last two available years.

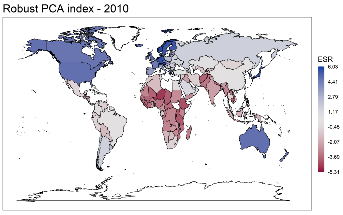
The maps in Figure C1 report the evolution of the ERS index over years based on PCA.

Table B8: Predicting ACCESS with ordinal probit model with instrumental variables and institutional governance controls - CMP.

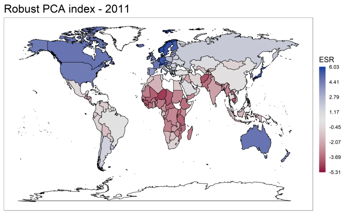
Variable	1	2	3	4	5	6	7	8
FSIND	-0.0705***	-0.1108***	-0.0273***	-0.1161***	-0.1077***	-0.0576***	-0.0603***	-0.1072***
	(0.0169)	(0.0169)	(0.0169)	(0.0169)	(0.0170)	(0.0170)	(0.0169)	(0.0169)
LISTED	-0.0262***	-0.0074***	-0.0617***	-0.0054***	-0.0432***	-0.01***	-0.0231***	-0.0387***
	(0.00821)	(0.00821)	(0.00825)	(0.00826)	(0.00822)	(0.00822)	(0.00824)	(0.00825)
AGE	-0.2608***	-0.6282***	-0.3576***	-0.3919***	-0.3287***	-0.2656***	-0.3228***	-0.287***
	(0.0301)	(0.0301)	(0.0309)	(0.0309)	(0.0300)	(0.0300)	(0.0309)	(0.0309)
SUBSID	-0.0188***	-0.0196***	-0.0381***	-0.0304***	-0.0369***	-0.0424***	-0.0173***	-0.0268***
	(0.00414)	(0.00414)	(0.00418)	(0.00418)	(0.00413)	(0.00413)	(0.00419)	(0.00419)
LOCATION	0.0284***	0.0098***	0.1174***	0.0863***	0.008***	0.0591***	0.0093***	0.0675***
	(0.00744)	(0.00748)	(0.00745)	(0.00749)	(0.00745)	(0.00749)	(0.00745)	(0.00748)
EXPORT	-0.0646***	-0.0372***	-0.0127***	-0.0458***	-0.0336***	-0.0782***	-0.0558***	-0.0472***
	(0.00709)	(0.00709)	(0.00718)	(0.00718)	(0.00703)	(0.00703)	(0.00723)	(0.00723)
OWNFOR	-0.1716***	-0.054***	-0.0085***	-0.1442***	-0.0568***	-0.0422***	-0.1277***	-0.1086***
	(0.00645)	(0.00645)	(0.00652)	(0.00652)	(0.00646)	(0.00646)	(0.00651)	(0.00651)
OWNGOV	-0.1209***	-0.2237***	-0.138***	-0.0584***	-0.2188***	-0.2324***	-0.0836***	-0.1088***
	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.0249)
GENDEQ		-0.8597*		-1.3421		-0.8981*		-1.9112*
		(0.582)		(0.582)		(0.583)		(0.582)
BILHUM		-1.1051		-1.4269		-2.9455		-2.6702
		(1.627)		(1.627)		(1.629)		(1.625)
FISPOL		0.5901***		3.6012***		2.8655***		3.1127***
		(0.517)		(0.517)		(0.518)		(0.517)
NUMBRW	3.7613***	11.1949***	16.3328***	16.3608***	11.1223***	19.3083***	5.2505***	3.5525***
	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)	(0.142)
Observations	39,383	39,383	39,383	39,383	39,383	39,383	39,383	39,383
Pseudo R^2	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
F-stat	96.88	95.62	95.24	94	96.83	95.54	95.59	94.38
ρ	0.0137*	0.014*	0.0125*	0.0128*	0.0125*	0.0128*	0.0138*	0.0141*
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and Conditional Mixed Process for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Arellano-Bond ρ indicates the magnitude of auto-correlation and its significance level. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

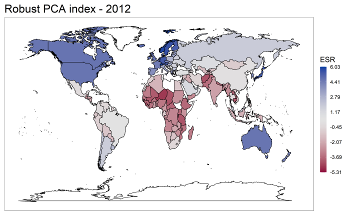
Robust PCA index - 2010



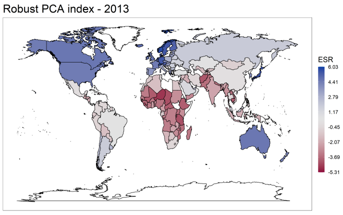
Robust PCA index - 2011



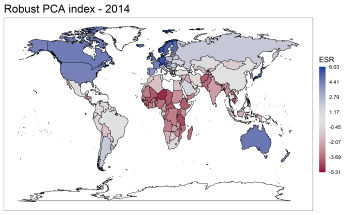
Robust PCA index - 2012



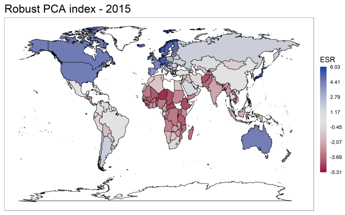
Robust PCA index - 2013



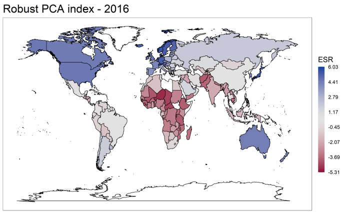
Robust PCA index - 2014



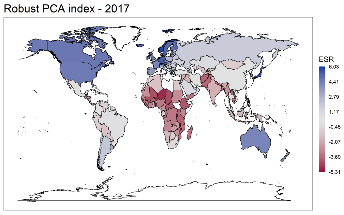
Robust PCA index - 2015



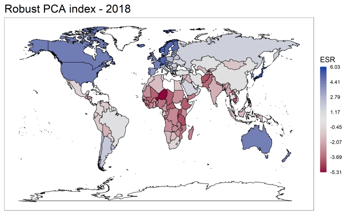
Robust PCA index - 2016



Robust PCA index - 2017



Robust PCA index - 2018



Robust PCA index - 2019

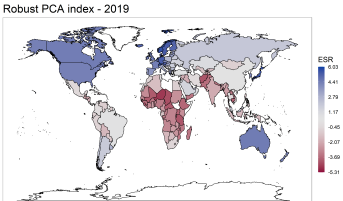


Table B9: Predicting ACCESS with ordinal probit model with instrumental variables and political controls - CMP.

Variable	1	2	3	4	5	6	7	8
FSIND	-0.0706***	-0.0138**	-0.087***	-0.0489*	-0.0221***	-0.0767**	-0.0866***	-0.0045*
	(0.0169)	(0.0162)	(0.0169)	(0.0162)	(0.0170)	(0.0162)	(0.0169)	(0.0162)
LISTED	-0.0648***	-0.0561***	-0.0233***	-0.0208***	-0.0216***	-0.0855***	-0.0297***	-0.0033***
	(0.00821)	(0.00841)	(0.00825)	(0.00845)	(0.00822)	(0.00842)	(0.00824)	(0.00844)
AGE	-0.4045***	-0.2139***	-0.0863***	-0.4144***	-0.3044***	-0.088***	-0.1647***	-0.3881***
	(0.0301)	(0.0309)	(0.0309)	(0.0316)	(0.0300)	(0.0308)	(0.0309)	(0.0317)
SUBSID	-0.0379***	-0.0012***	-0.0255***	-0.0025***	-0.0367***	-0.0212***	-0.0273***	-0.0056***
	(0.00414)	(0.00425)	(0.00418)	(0.00431)	(0.00413)	(0.00425)	(0.00419)	(0.00432)
LOCATION	0.0293***	0.1225***	0.0858***	0.1161***	0.0375***	0.0823***	0.0984***	0.0061***
	(0.00744)	(0.00766)	(0.00745)	(0.00767)	(0.00745)	(0.00767)	(0.00745)	(0.00766)
EXPORT	-0.0951***	-0.1007***	-0.0336***	-0.0421***	-0.055***	-0.0131***	-0.0425***	-0.0372***
	(0.00709)	(0.00725)	(0.00718)	(0.00734)	(0.00703)	(0.00718)	(0.00723)	(0.00738)
OWNFOR	-0.1411***	-0.1239***	-0.0329***	-0.0237***	-0.0147***	-0.128***	-0.0392***	-0.0621***
	(0.00645)	(0.00657)	(0.00652)	(0.00665)	(0.00646)	(0.00658)	(0.00651)	(0.00664)
OWNGOV	-0.1625***	-0.0655***	-0.1042***	-0.1225***	-0.1702***	-0.1357***	-0.1383***	-0.0954***
	(0.0249)	(0.0253)	(0.0249)	(0.0253)	(0.0249)	(0.0253)	(0.0249)	(0.0253)
STABDEM		-0.2007*		-0.4422*		-0.1479*		-0.3294*
		(0.101)		(0.101)		(0.101)		(0.100)
LIMLEND		-0.2046		-0.2171		-0.2342		-0.0309
		(0.104)		(0.104)		(0.104)		(0.104)
VETOPWR		0.0504		0.0342		0.0637		0.0574
		(0.0254)		(0.0254)		(0.0255)		(0.0254)
NUMBRW	16.7708***	6.7573***	13.369***	15.4917***	5.8635***	14.2124***	9.3071***	16.6192***
	(0.142)	(0.117)	(0.142)	(0.117)	(0.142)	(0.117)	(0.142)	(0.117)
Observations	39,383	37,739	39,383	37,739	39,383	37,739	39,383	37,739
Pseudo R^2	0.15	0.14	0.15	0.14	0.15	0.14	0.15	0.15
F-stat	96.88	89.34	95.24	87.51	96.83	89.08	95.59	88.08
ρ	0.0137*	0.0028	0.0125*	0.0026	0.0125*	0.0027	0.0138*	0.0026
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector effects	Yes	Yes	No	No	No	No	Yes	Yes
Size effects	No	No	Yes	Yes	No	No	Yes	Yes
Clustered Std. Err.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports coefficients and their standard error (in parentheses). The outcome variable is ACCESS and all variables are defined in Table 2.2. Data span over the period 2010-2018 for 76 countries. Estimation method is OLS with standard errors clustered by firm's country and Conditional Mixed Process for instrumental variables. The bottom part of the table reports which fixed effects are used in each model specification. Specifications (1), (3), (5) and (7) report the results for the model without the control variables and different combinations of fixed effects. Specifications (2), (4), (6) and (8) report the results for the model with control variables and different combinations of fixed effects. Arellano-Bond ρ indicates the magnitude of auto-correlation and its significance level. The *, ** and *** symbols denote the p-values at 10^{th}, 5^{th} and 1^{st} significance level, respectively.

Table C4: Missing values over years.

Year	Total Observations	Missing Values
2010	22,848	4,367 (19.1%)
2011	22,848	3,959 (17.3%)
2012	22,848	4,273 (18.7%)
2013	22,848	4,072 (17.8%)
2014	22,848	4,019 (17.6%)
2015	22,848	4,494 (19.7%)
2016	22,848	4,404 (19.3%)
2017	22,848	4,245 (18.6%)
2018	22,848	7,478 (32.7%)
2019	22,848	8,218 (36.0%)

C.2 Missing values imputation methodology and pre-processing

Imputation of missing data

We next assess the data quality and completeness and address the problem of missing data. There are various imputation methods that are suitable for different data sets and conditions (Johnson and Young, 2011). There is a trade-off between data availability and the construction of a comprehensive composite index. We stress that the goal of our index construction is not to create artificial data series. We do not use the individual series as standalone predictors. Instead, we combine them to produce a single composite index that reflects consistently epidemiological susceptibility risk, rather than data availability. King et al. (2001) argue that imputation of missing data and their combination into aggregate indices is highly common in social sciences research because the nature of measured phenomena is associated with incomplete records. Typically, more data is available for larger countries and in recent years.We therefore restricted our sample to the 2010-2019 period and 168 countries (81.5% of all countries) in which the missing data tolerance rate does not exceed 40%, which gives a total of 28,560 country-year observations. Table C3 presents the full list of the sample countries and their rate of missing data, whilst Table C4 presents the missing data incidence by year.

Since the presence of many missing values can extremely impact the quality and the reliability of results, we set an operational protocol of missing values treatment and imputation. In our final sample, 114 out of 168 countries show a rate of missing data between 20-39%. To address the missing values problem that would make possible the application of robust data aggregation methods, we test two different data imputation techniques: Matrix Completion with Low Rank SVD (MC-SVD) proposed by Hastie et al. (2015) and Bayesian Tensor Factorization (BTF) proposed by Khan and Ammad-ud-din (2016).

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 on the 2-dimensional "slice" of countries-variables for each year. Subsequently, we apply the BTF method, which in addition uses a tensorial decomposition of the 3-dimensional tensors that stack 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 setting (named *Original*) we consider the whole dataset made of 168 countries and the 17 constituents variables over 10 years for a total of 28,560 entries. The full sample has 25% of missing values, thus we randomly remove some additional values, representing 10%, 20% and 30% of the initial dataset. In the second setting (named *No missing*) we drop all entries with missing values and apply the same incremental sampling procedure on the remaining subset. In the last setting (named *Some missing*) 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 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 on *No missing* and *Some missing* with the one on *Original*. Figure C2 shows bar plot of MARE values for all settings for each increasing percentage of added missing values. Bar whiskers are scaled value of

$max(MARE)$, defined as:

$$RM = \frac{max(MARE)}{\text{Average value of Original matrix}}$$

In order to grasp the magnitude of the impact of MARE we also report its ratio R with the average value of the non-missing entries of original matrix:

$$R = \frac{MARE}{\text{Average value of Original matrix}}$$

Figure C2: Testing missing values imputation methodologies. Blue bars report the Mean Absolute Reconstruction Error (MARE), green/red bars report the percent decrease/increase of MARE compared to the one evaluated on the *Original* setting.

Finally, for the *No missing* and *Some missing* setting we show the green/red bar plot reporting the percent decrease/increase, respectively, of MARE compared to the one evaluated in the *Original* setting so to evaluate the impact of missing data in the matching entries subset. BTF has lower MARE and higher percent decrease compared to MC-SVD implying a better data reconstruction ability and reliability.

Normalization of data

We remove differences in magnitude among the input variables by standardising the values, i.e. we subtract the mean and divide by the standard deviation. Having all variables on the same reference scale is crucial for unbiased estimation when applying dimensionality reduction techniques. Standardisation relates country performance of a variable as a bounded (by unitary standard deviation) variation from an average value (set to zero by definition) across all countries and years, which facilitates variable aggregation expressed in different measurement units. Further, when applying dimensionality reduction methods, component weights can have a significant effect on the overall composite indicator and country rankings. Several weighting techniques exist (Nardo et al., 2005). Some are based on statistical models (e.g., factor analysis), whilst others are based on participatory methods (e.g., analytical hierarchy process). Regardless of the method used, weights are essentially value judgments. However, our data-driven approach overcomes the problem of arbitrary and subjective choice of weights that could constrain the index's predictive efficacy.

C.3 Scree Plot and Loadings Plot for PCA method

In this Appendix we report scree plots and loadings of all the competing PCA approaches: Original PCA, Robust PCA, Robust Sparse PCA. If we pay attention to loadings results available in C6, we can notice that Original PCA and Robust PCA are very similar to each other, while Robust Sparse PCA appears different for several variables (namely var2, var3, var6, var8, var12, var13, var14) because by construction it aims to a sparse and parsimonious representation. In the Robust PCA almost all the variables have a meaningful positive contribution to the first Principal Component, that constitutes our ESR index (var8 (num of people per Km2) and var15 (value of trade as % of GDP) appear to be less significant).

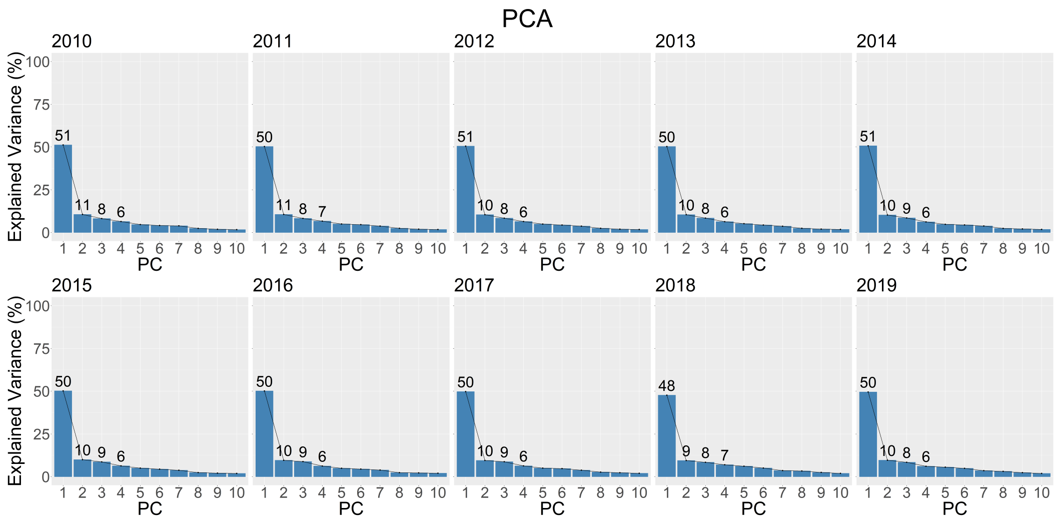


Figure C3: Scree plot for PCA method.

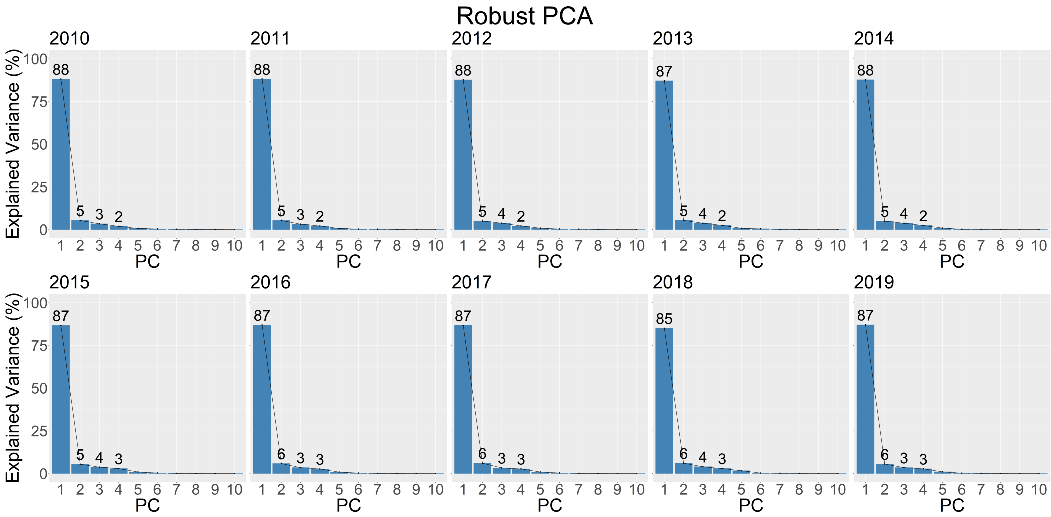


Figure C4: Scree plot for Robust PCA method.

Figure C3 clearly shows how important is the first component whatever year we take into account. Such result has several important implications: PCA proves that there exists a strong latent component which is highly connected to almost all the variables. Moreover, the possibility of building up our ESR index considering just one component eases the interpretation, the relative employment and the subsequent monitoring.

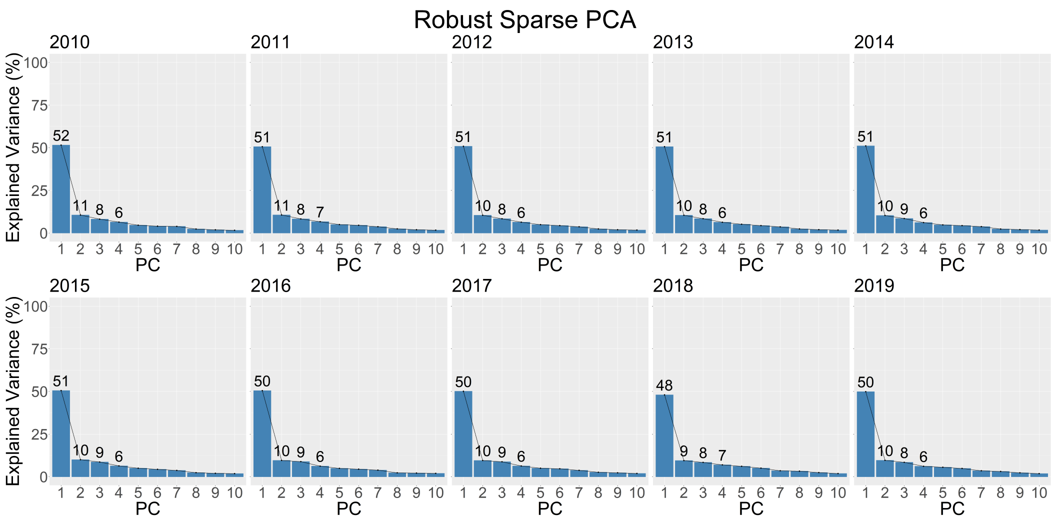


Figure C5: Scree plot for Robust Sparse PCA method.

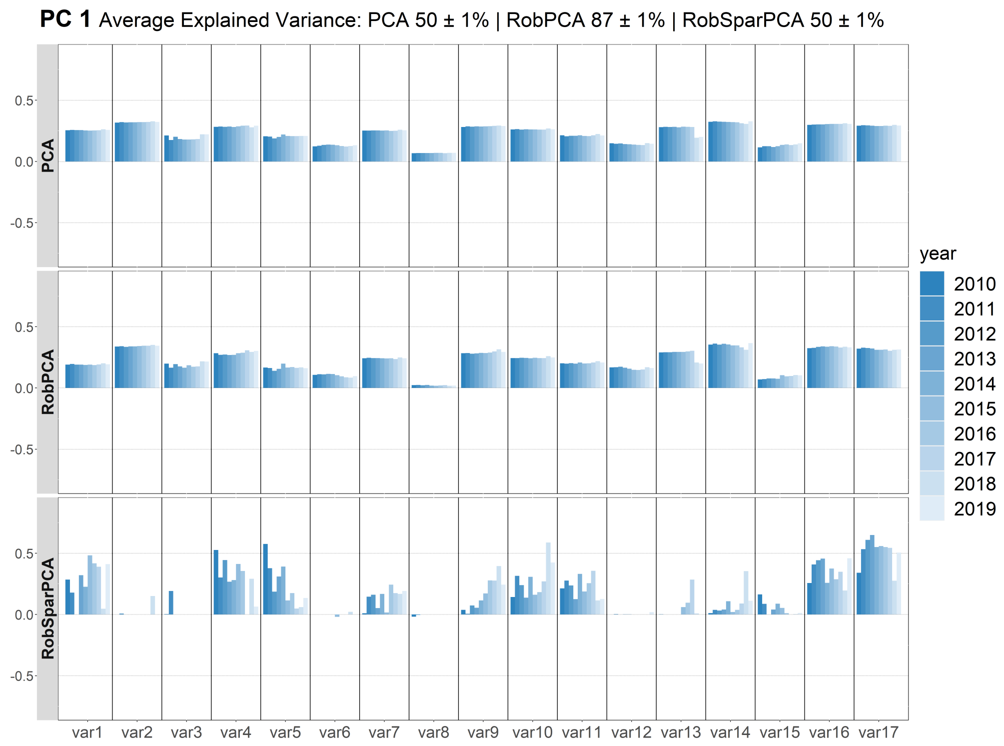


Figure C6: Loading plot for all PCA methods.

C.4 Loadings Plot for DFM method

The loadings $\mathbf{C}^i$ for the i-th country are stacked into the diagonal matrix $\mathbf{C}$, whereas the cross-country interactions are introduced by the matrix $\hat{\boldsymbol{A}}$ estimated with VAR. Our setting force the $\mathbf{C}^i$ to be constant so we can estimate loadings for each country-variable pair. Therefore, for ease of visualization, figure C7 reports the distribution of the loadings for each input variable over the 168 countries, representing the average trend over the years. The bimodal shape of all distributions implies a clear discriminative power of the index between less risky countries and riskier ones.

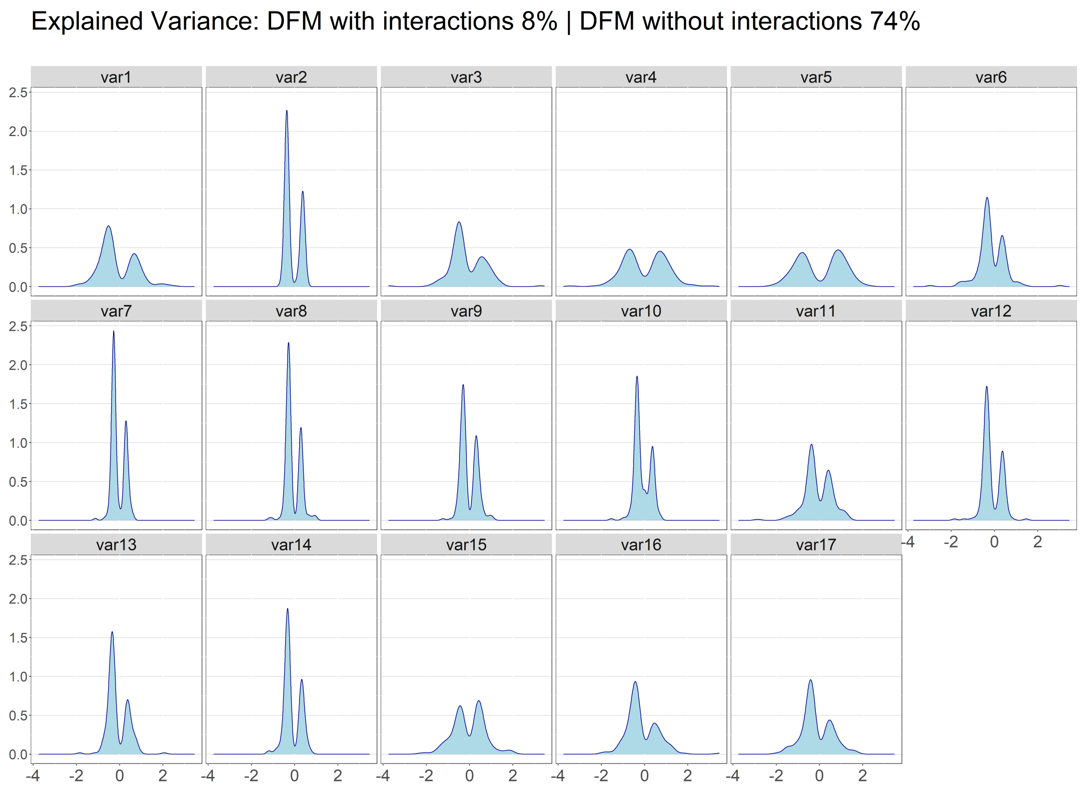


Figure C7: Loading plot for DFM method. On x-axis is reported the logarithm of loading values.

C.5 Index Robustness Check

Our robustness check is performed by using the ESR index as an input variable in supervised regressions. The aim is to evaluate the fitting power of the summary index compared to the original variables in modeling some relevant macro economic indicators. From Figure C8 through Figure C13 we report percent increase of RMSE in predicting macro economic indicators of interest (Unenployment, Real GDP per capita, Share of government consumption, Price level of capital information, Trade Volume, Outstanding Loans of Commercial banks) due to the employment of the ESR index. The graphs report comparison between regressions with the single continuous ESR index as regressor and the one with original variables. In this way we assess how much the RMSE increases by substituting 17 variables with our summary index. In table C5 we report numerical results for the regressions above described.

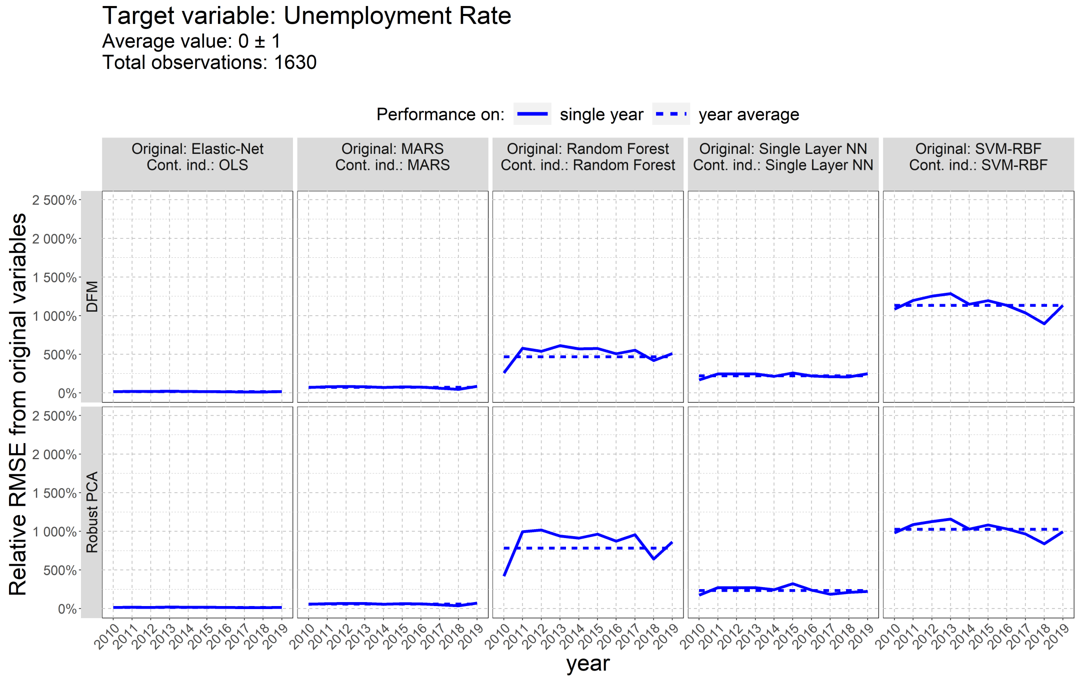


Figure C8: RMSE percent increase in predicting Unemployment rate. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

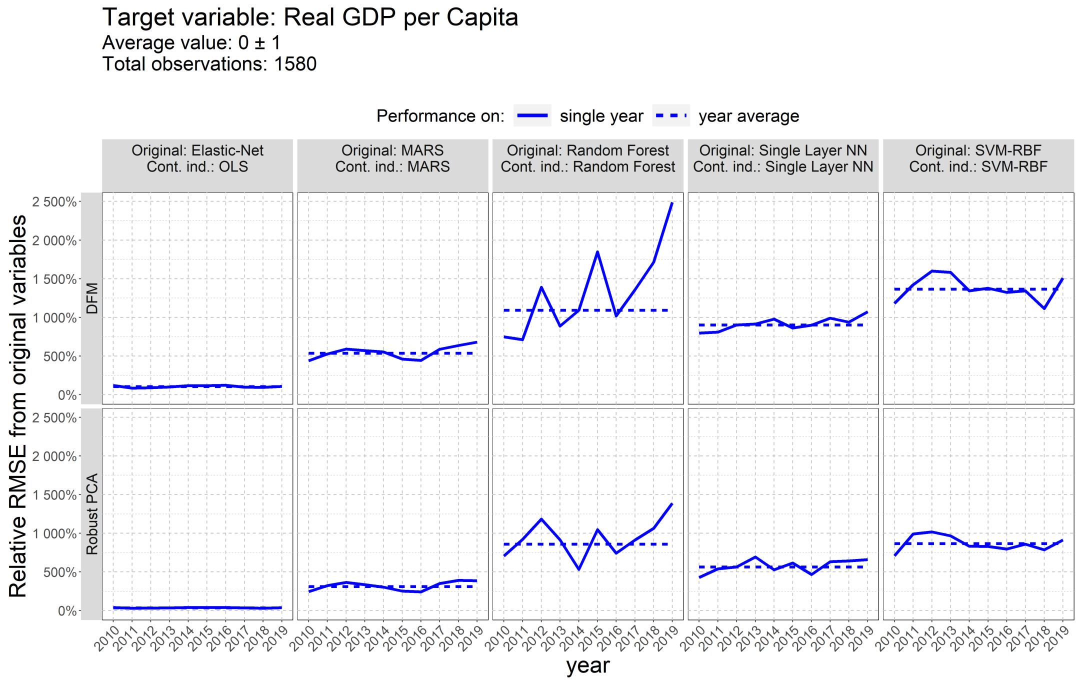


Figure C9: RMSE percent increase in predicting Real GDP per capita. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

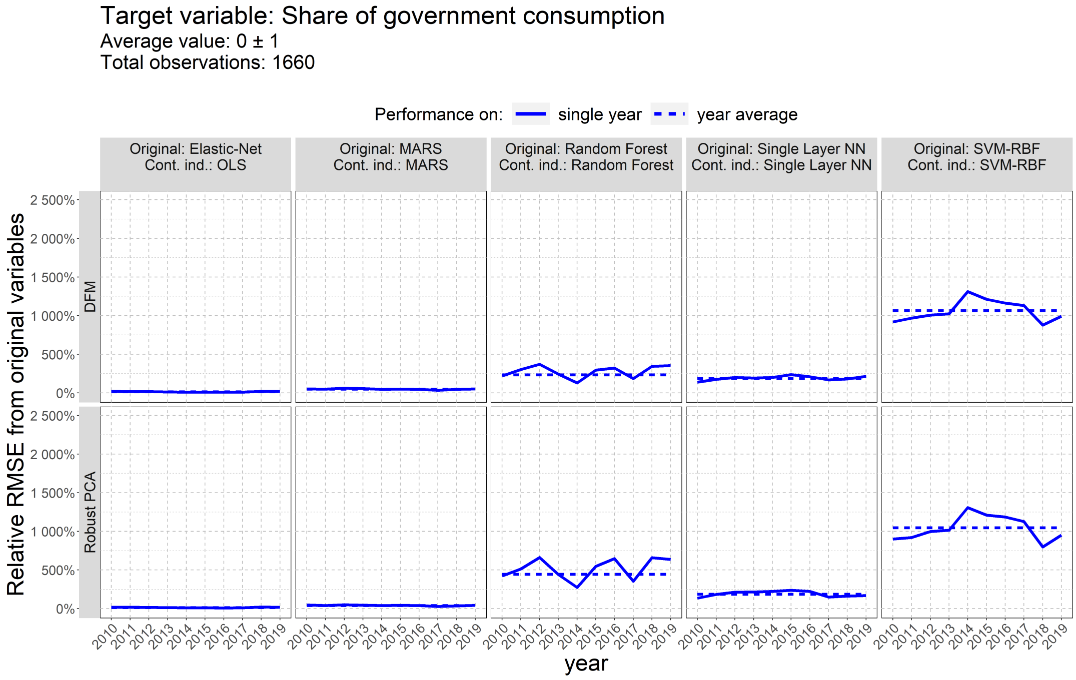


Figure C10: RMSE percent increase in predicting Share of government consumption. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

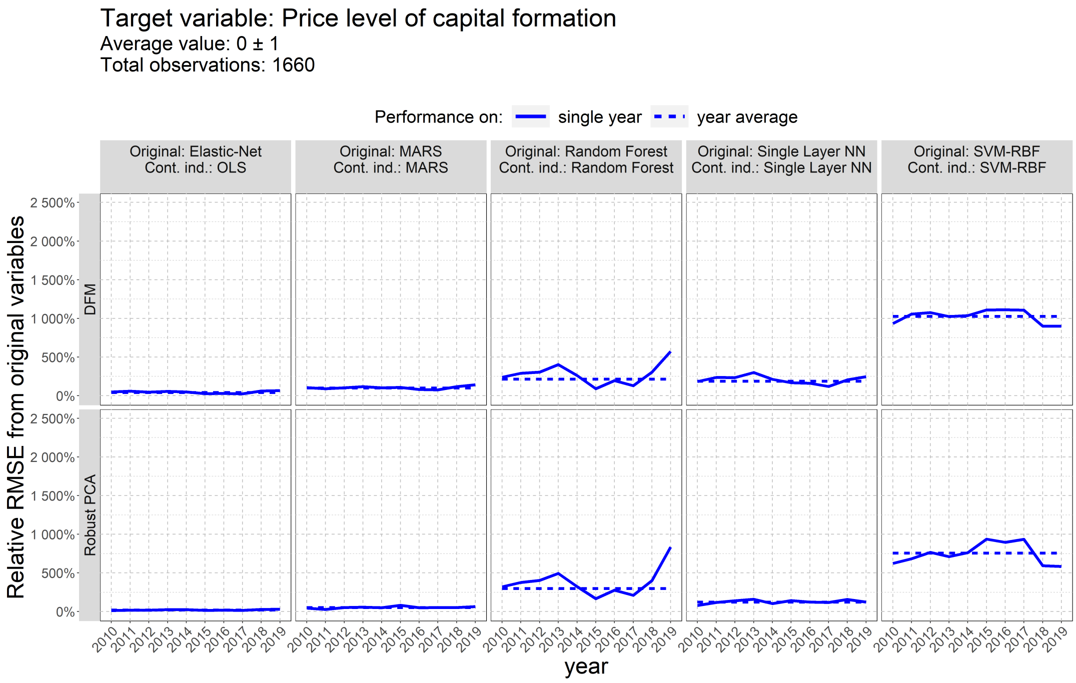


Figure C11: RMSE percent increase in predicting Price level of capital formation. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

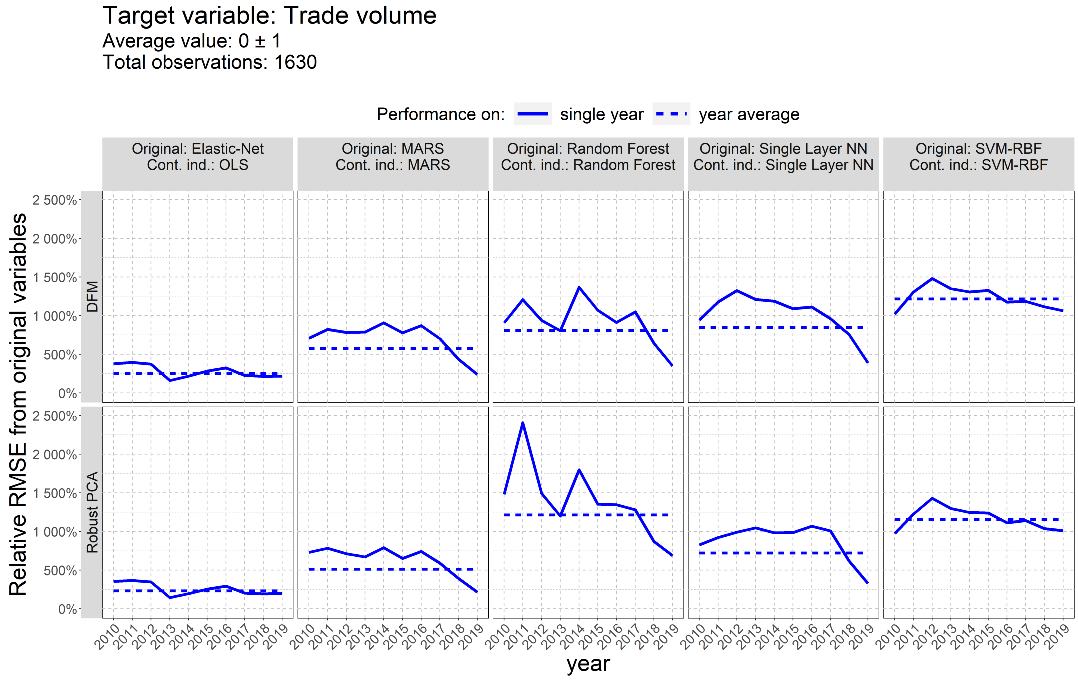


Figure C12: RMSE percent increase in predicting Trade volume. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

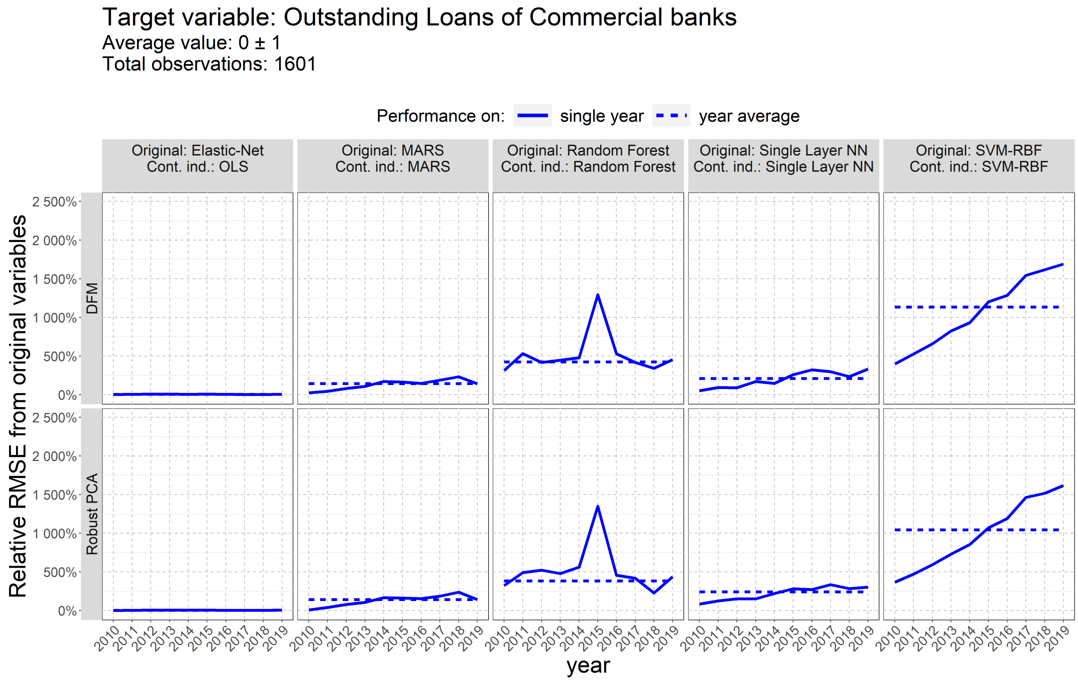


Figure C13: RMSE percent increase in predicting Outstanding Loans of Commercial banks. Comparison between regression with the single continuous index as regressor and the one with original variables. Solid lines show the single year metrics, dashed lines show the full dataset, i.e. average over years, metric.

Table C5: RMSE in predicting macro-economic variables with continuous index as regressor. RMSE for regression with original variables is reported in parenthesis.

	RMSE index (RMSE original)					
Target variable	Outstanding Loans of Commercial banks		Price level of capital formation		Real GDP per Capita	
Algorithm	DFM	Robust PCA	DFM	Robust PCA	DFM	Robust PCA
Elastic-Net	0.998(0.962)	1(0.962)	1(0.705)	0.839(0.705)	1(0.492)	0.663(0.492)
MARS	0.995(0.409)	0.986(0.409)	1(0.502)	0.764(0.502)	0.987(0.155)	0.634(0.155)
Random Forest	0.854(0.163)	0.914(0.163)	0.432(0.137)	0.549(0.137)	0.479(0.04)	0.395(0.04)
SVM-RBF	1.001(0.081)	1(0.081)	1.005(0.089)	0.764(0.089)	1.027(0.07)	0.664(0.07)
Single Layer NN	0.991(0.321)	0.976(0.321)	0.994(0.347)	0.768(0.347)	0.997(0.099)	0.663(0.099)

Target variable	Share of government consumption		Trade volume	
Algorithm	DFM	Robust PCA	DFM	Robust PCA
Elastic-Net	1(0.887)	0.999(0.887)	1(0.283)	0.935(0.283)
MARS	1(0.679)	0.948(0.679)	1(0.148)	0.909(0.148)
Random Forest	0.445(0.133)	0.719(0.133)	0.437(0.048)	0.637(0.048)
SVM-RBF	0.992(0.085)	0.977(0.085)	1.011(0.077)	0.954(0.077)
Single Layer NN	0.994(0.35)	0.973(0.35)	0.995(0.105)	0.926(0.105)

C.6 Index evolution over years

From C14 through C17 we report the evolution across time of the ESR index based on the two competing techniques for the different considered countries. It clearly emerges the higher sensitivity of the ESR index based on the DFM approach to the temporal dynamics which are explicitly modelled. PCA instead produces a rather flat pattern in line with the no direct modelling of the available years.

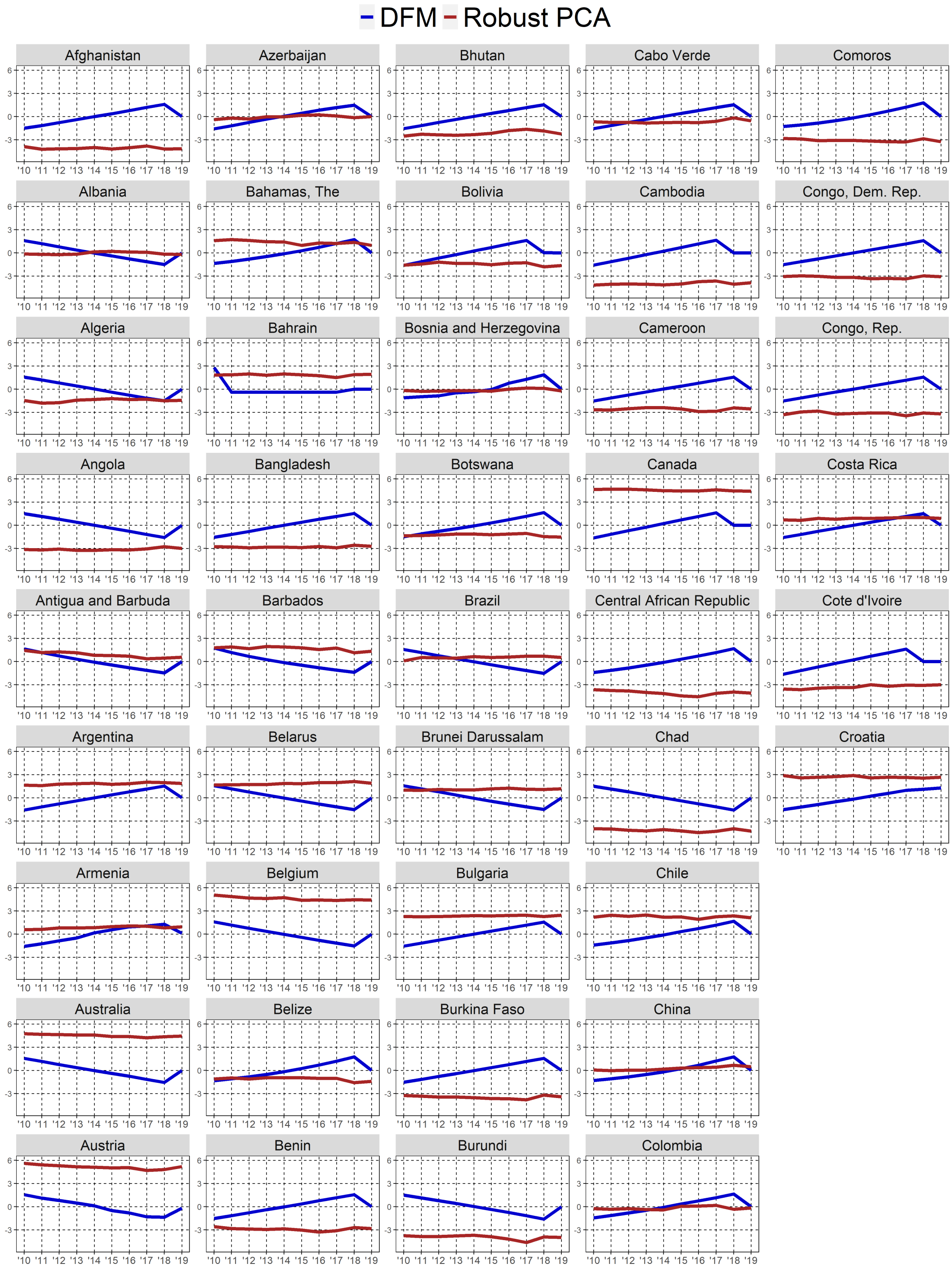


Figure C14: Index evolution over years.

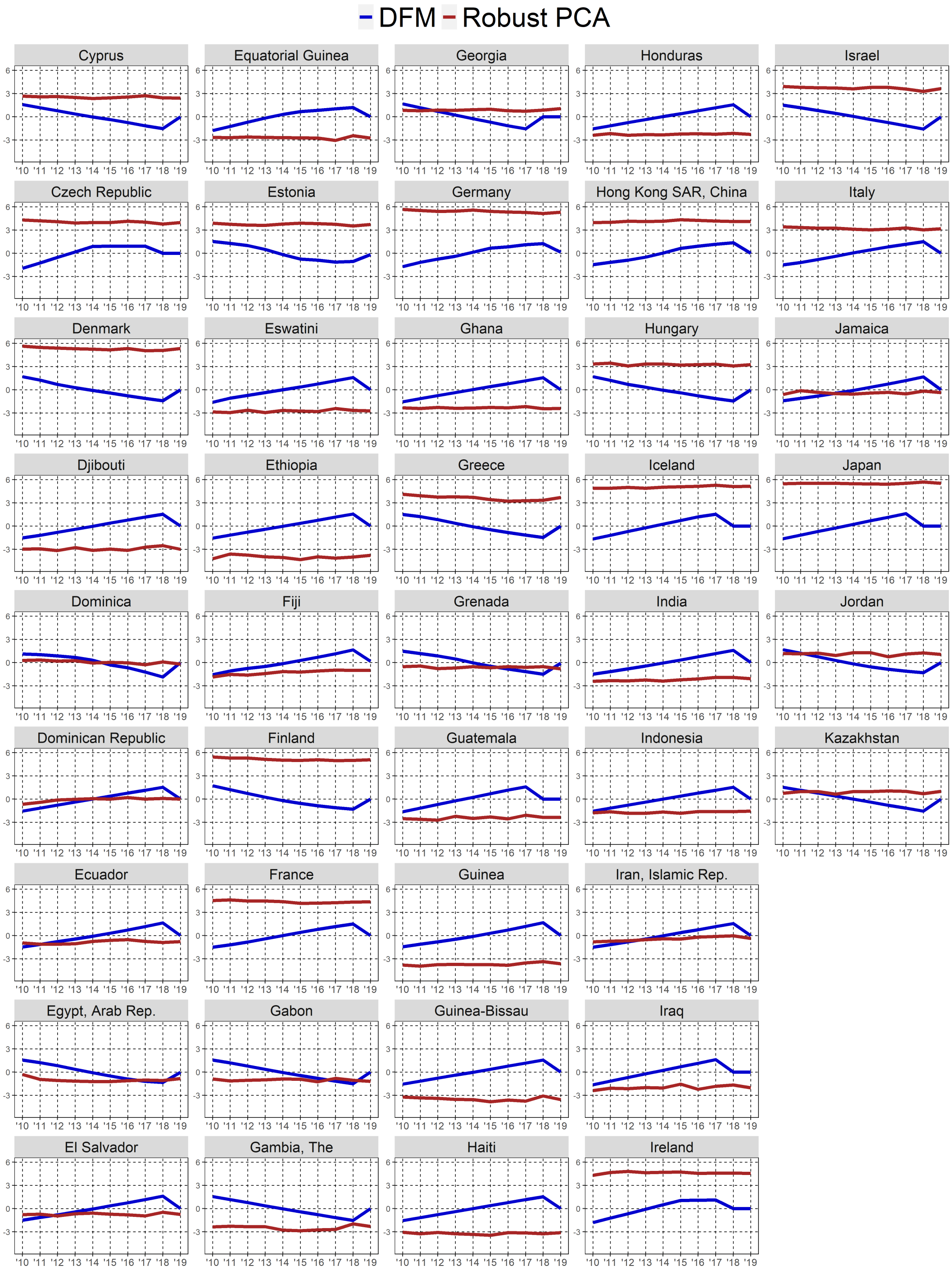


Figure C15: Index evolution over years.

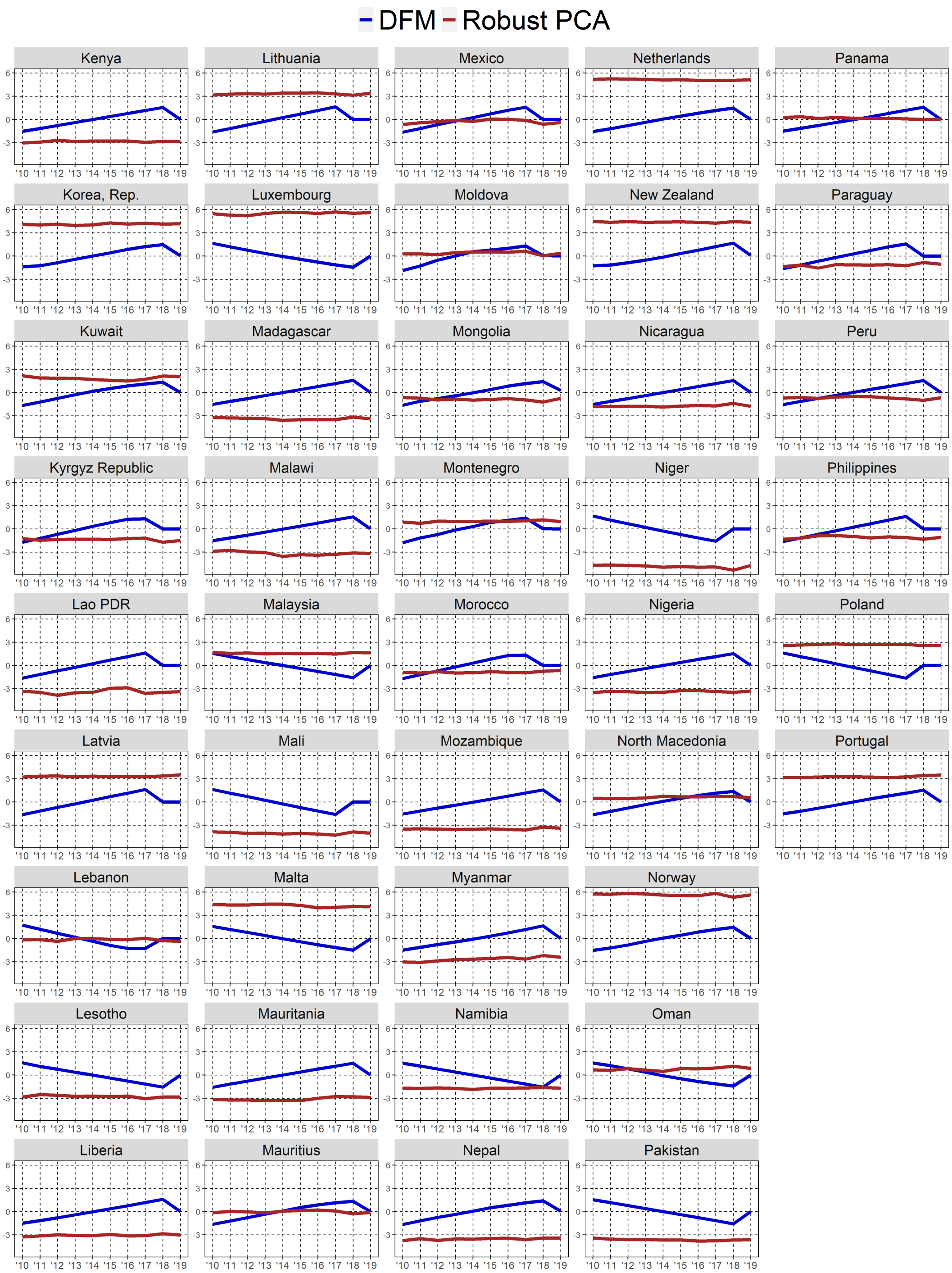


Figure C16: Index evolution over years.

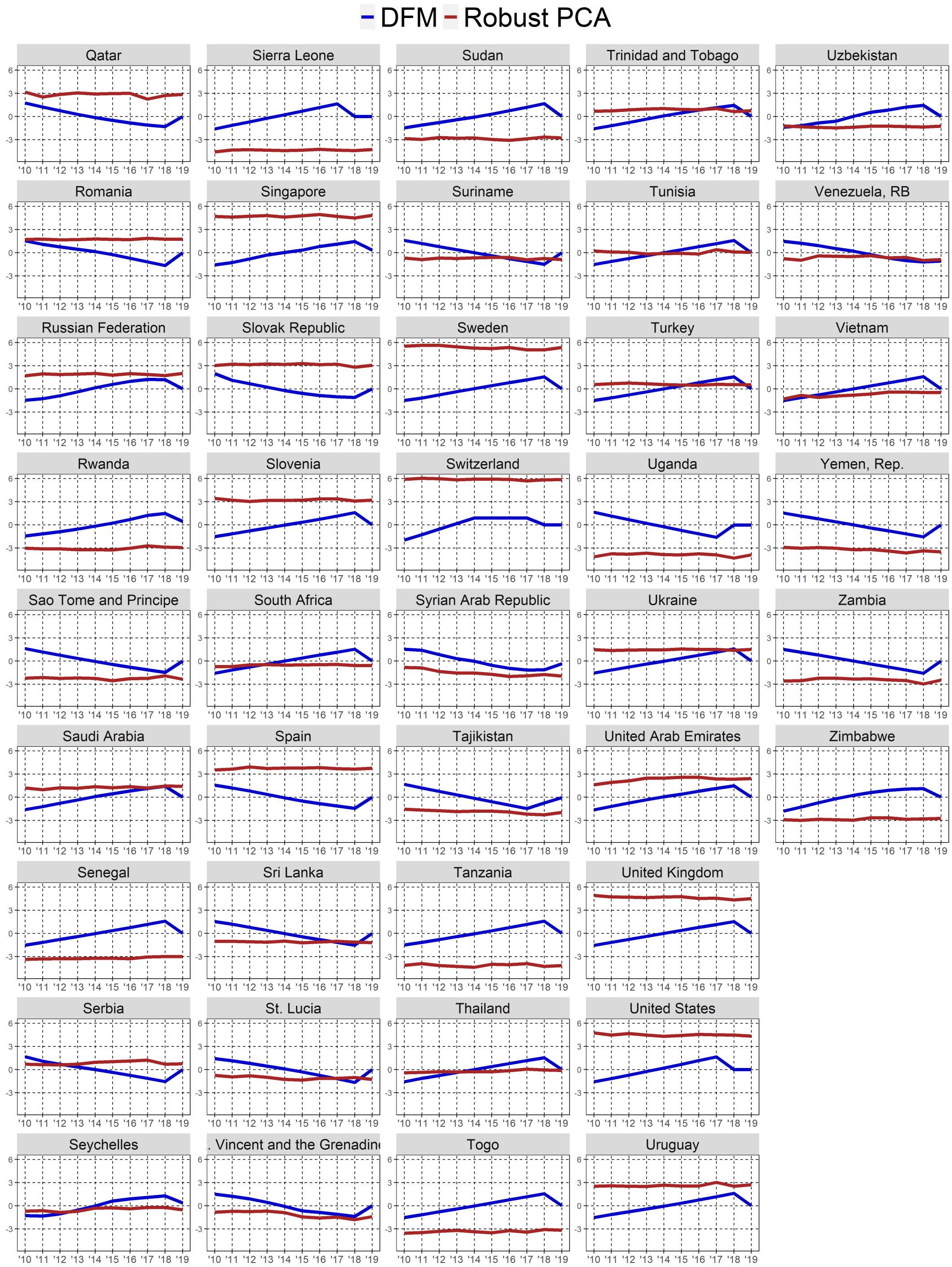


Figure C17: Index evolution over years.

D Machine Learning and Credit Risk: Empirical Evidence from SMEs

D.1 List of raw variables

Table D1: List of initial variables.

Variable	Type	Missing %	Minimum	Maximum	Mean	St Dev	Unique values	Source	Frequency	Action
Industry	Cat	0%					22			
NACE	Cat	0%					11			
City	Cat	0%					375			
Purchase_2015		4%	242,000	6,038,375,000	157,916,200	468,969,700				
Purchase_2016		3%	236,000	6,277,094,000	154,080,200	452,602,000				
Purchase_2017		6%	18,000	6,497,610,000	174,069,100	508,482,300				
Collectionperioddays2015		2%	0	264	52	42				
Collectionperioddays2016		0%	0	290	51	43				
Collectionperioddays2017		1%	0	265	50	40				
Creditperioddays2015		2%	0	171	52	29				
Creditperioddays2016		0%	0	174	53	32				
Creditperioddays2017		1%	0	557	57	45				
Current liabilities2015		2%	116,078	1,762,623,000	65,453,870	180,990,000				
Current liabilities2016		0%	142,224	1,716,557,000	64,732,280	176,089,600				
Current liabilities2017		1%	142,548	1,746,051,000	67,123,170	175,378,800				
Current ratio2015		2%	0	19	1	1				
Current ratio2016		0%	0	9	1	1				
Current ratio2017		2%	0	7	1	1				
EBIT2015		2%	-97,272,000	423,930,000	6,082,113	29,011,610				
EBIT2016		0%	-73,706,000	402,693,000	6,155,327	28,772,780				
EBIT2017		1%	-114,775,000	410,921,000	6,409,824	30,289,210				
Fixed assets2015		2%	2,039	3,255,230,000	72,978,520	271,180,300				
Fixed assets2016		0%	500	3,422,961,000	74,910,600	280,711,700				
Fixed assets2017		1%	500	4,578,240,000	79,002,090	314,053,100				
Liquidity2015		2%	0	16	1	1				
Liquidity2016		0%	0	7	1	1				
Liquidity2017	Num	1%	0	6	1	1		Orbis	Annual	
LT Debt2015		2%	0	743,361,000	13,335,680	51,390,820				
LT Debt2016		0%	0	516,854,000	13,735,920	49,286,210				
LT Debt2017		1%	0	1,378,198,000	16,004,280	75,481,730				
Asset Turnover2015		2%	0	353	8	22				
Asset Turnover2016		1%	0	308	7	16				
Asset Turnover2017		2%	0	231	7	15				
Profit Margin2015		2%	-40	25	2	4				
Profit Margin2016		0%	-67	29	2	6				
Profit Margin2017		2%	-73	56	2	7				
Profit per employee2015		2%	-69,693	566,486	18,088	43,258				
Profit per employee2016		1%	-89,917	480,993	17,743	41,255				
Profit per employee2017		2%	-76,612	273,365	16,557	33,493				
ROA2015		2%	-25	23	3	4				
ROA2016		0%	-35	31	3	5				
ROA2017		2%	-35	47	3	6				
ROCE2015		2%	-731	84	8	35				
ROCE2016		1%	-120	85	8	15				
ROCE2017		2%	-364	81	6	28				
ROE2015		2%	-529	973	11	53				
ROE2016		1%	-309	95	9	26				
ROE2017		3%	-837	94	3	62				
Solvency_L2015		2%	-0	91	28	18				
Solvency_L2016		0%	-10	92	29	18				
Solvency_L2017		1%	-79	93	29	20				

Variable	Type	Missing %	Minimum	Maximum	Mean	St Dev	Unique values	Source	Frequency	Action
Tangibles2015		2%	0	3,041,447,000	46,639,800	186,760,700				
Tangibles2016		0%	0	3,257,302,000	49,759,500	199,723,300				
Tangibles2017		1%	0	4,388,377,000	53,987,410	240,358,800				
TotalAsset2015		2%	238,723	4,807,100,000	138,682,700	425,440,600				
TotalAsset2016		0%	305,390	5,641,500,000	141,434,300	446,524,100				
TotalAsset2017		1%	297,820	6,122,933,000	148,336,500	468,131,000				
Turnover2015		2%	661,365	8,315,389,000	226,622,000	683,864,800				
Turnover2016		0%	250,000	8,688,413,000	231,283,200	697,866,600				
Turnover2017		1%	250,000	8,896,700,000	243,088,200	720,414,600				
Working Capital2015		2%	-470,089,000	401,200,000	-1,101,782	53,508,490				
Working Capital2016		0%	-526,333,000	437,500,000	-1,357,159	57,046,520				
Working Capital2017	Num	1%	-532,052,000	417,400,000	-1,516,843	57,449,010		Orbis	Annual	
EBITDA_2015		89%	-6,573,000	381,351,000	12,709,390	51,550,450				Removed
EBITDA_2016		89%	-6,708,000	417,812,000	13,301,330	55,816,760				Removed
EBITDA_2017		89%	-9,458,000	425,655,000	13,246,350	55,360,140				Removed
Gearing2015		22%	0	998	158	173				Removed
Gearing2016		18%	0	993	151	163				Removed
Gearing2017		20%	0	987	158	172				Removed
Interestcover2015		15%	-65	980	34	101				Removed
Interestcover2016		9%	-87	743	33	82				Removed
Interestcover2017		12%	-81	841	38	102				Removed
Solvency_L2015		14%	0	99	35	24				Removed
Solvency_L2016		14%	0	99	35	24				Removed
Solvency_L2017		16%	1	100	36	25				Removed
InvoicesCount_03_2015		100%								Removed
InvoicesCount_03_2016		0%	0	15,442	232	1,263				
InvoicesCount_03_2017		0%	0	33,175	246	1,631				
InvoicesCount_06_2015		100%								Removed
InvoicesCount_06_2016		0%	0	17,894	249	1,402				
InvoicesCount_06_2017		0%	0	21,945	220	1,280				
InvoicesCount_09_2015		100%								Removed
InvoicesCount_09_2016		0%	0	20,218	229	1,325				
InvoicesCount_12_2015		0%	0	20,375	240	1,349				
InvoicesCount_12_2016		0%	0	20,781	243	1,399				
Collections_03_2015		7%	0	3,767,527	51,407	267,665				
Collections_03_2016		0%	0	4,549,522	70,702	322,386				
Collections_03_2017		7%	0	6,827,631	101,783	458,061				
Collections_06_2015		3%	0	4,913,739	80,929	378,060				
Collections_06_2016	Num	0%	0	3,131,328	48,427	229,664		Insurance	Quarterly	
Collections_06_2017		8%	0	6,838,858	95,551	440,323				
Collections_09_2015		2%	0	5,525,285	74,247	366,442				
Collections_09_2016		0%	0	6,368,652	66,993	378,023				
Collections_12_2015		0%	0	5,685,703	71,191	357,369				
Collections_12_2016		0%	0	8,502,823	93,058	497,841				
Delinquency90_032015		7%	0	574,535	2,482	25,639				
Delinquency90_032016		0%	0	987,342	6,087	55,590				
Delinquency90_032017		7%	0	802,891	3,897	40,401				
Delinquency90_062015		3%	0	1,053,269	3,206	45,242				
Delinquency90_062016		0%	0	792,409	3,152	33,447				
Delinquency90_062017		8%	0	461,054	2,584	22,676				
Delinquency90_092015		2%	0	1,184,993	3,745	46,382				
Delinquency90_092016		0%	0	902,860	3,357	39,019				
Delinquency90_122015		0%	0	1,461,948	3,511	54,612				
Delinquency90_122016		0%	0	653,700	2,520	30,138				

Variable	Type	Missing %	Minimum	Maximum	Mean	St Dev	Unique values	Source	Frequency	Action
New Receivables_03_2015		7%	0	3,212,298	70,264	277,914				
New Receivables_03_2016		0%	0	3,153,487	62,513	258,509				
New Receivables_03_2017		0%	0	3,371,893	63,498	262,418				
New Receivables_06_2015		3%	0	2,869,105	67,241	278,122				
New Receivables_06_2016		0%	0	2,968,038	72,552	295,227				
New Receivables_06_2017		0%	0	3,854,462	68,858	286,216				
New Receivables_09_2015		2%	0	4,501,308	74,487	321,610				
New Receivables_09_2016		0%	0	5,217,448	74,585	324,415				
New Receivables_12_2015		0%	0	3,560,522	83,456	311,808				
New Receivables_12_2016		0%	0	5,279,336	93,226	378,014				
Outstanding_03_2015		7%	0	12,163,100	394,611	1,317,095				
Outstanding_03_2016		0%	0	14,320,530	367,375	1,296,520				
Outstanding_03_2017		5%	0	14,515,050	443,584	1,521,907				
Outstanding_06_2015		3%	0	10,712,350	327,825	1,160,684				
Outstanding_06_2016	Num	0%	0	14,682,840	390,294	1,443,342		Insurance	Quarterly	
Outstanding_06_2017		7%	0	14,497,480	479,817	1,526,787				
Outstanding_09_2015		2%	0	11,777,300	326,774	1,163,738				
Outstanding_09_2016		0%	0	14,052,470	363,672	1,343,489				
Outstanding_12_2015		0%	0	14,598,040	379,470	1,346,056				
Outstanding_12_2016		0%	0	15,085,840	383,250	1,444,328				
PortfolioCount_03_2015		7%	0	10	2	2				
PortfolioCount_03_2016		0%	0	13	2	2				
PortfolioCount_03_2017		0%	0	12	2	2				
PortfolioCount_06_2015		3%	0	10	2	2				
PortfolioCount_06_2016		0%	0	12	2	2				
PortfolioCount_06_2017		0%	0	12	2	2				
PortfolioCount_09_2015		2%	0	10	2	2				
PortfolioCount_09_2016		0%	0	11	2	2				
PortfolioCount_12_2015		0%	0	13	2	2				
PortfolioCount_12_2016		0%	0	11	2	2				

D.2 Performance

Table D2: Model architecture for SEC set of predictors.

Model	Version	Hyperparameters or Selected set of predictors
HRF	Static	Mtry = 5; Ntrees = 10; Nodesize = 100
	Dynamic	Mtry = 4; Ntrees = 141; Nodesize = 89; Method = "mean0"
PB	Static	New Receivables+Outstanding+Delinquency
	Dynamic	Collections+Outstanding+Delinquency+LagRating

Table D3: Model architecture for BS set of predictors.

Model	Version	Hyperparameters or Selected set of predictors
HRF	Static	Mtry = 14; Ntrees = 500; Nodesize= 1
	Dynamic	Mtry = 6; Ntrees = 50; Nodesize= 3; Method = “meanw0”
PB	Static	Current liabilities + Liquidity ratio + LT Debt + ROA+ Tangibles + Working Capital + Purchase + Turnover + Region + NACE
	Dynamic	Current liabilities + Liquidity + LT Debt + Working Capital + Purchase + EBIT + Turnover + Region + LagRating

Table D4: Model architecture for BS+SEC set of predictors.

Model	Version	Hyperparameters or Selected set of predictors
HRF	Static	Mtry = 5; Ntrees = 500; Nodesize= 1
	Dynamic	Mtry = 5; Ntrees = 50; Nodesize= 3; Method = “freqw”
PB	Static	Collections + New Receivables + Delinquency + Turnover + Solvency_A + Working Capital + LT Debt + Current liabilities + Liquidity
	Dynamic	Collections + New Receivables + Delinquency + Turnover + Solvency_A + Working Capital + LT Debt + Current liabilities + Liquidity + LagRating

Table D5: Table of PB marginal effects for BS variables.

Model	Historical	Variables	Marginal effects				
			y = 3	y = 4	y = 5	y = 6	y = 7
PB	Static	Current liabilities	-0.2871 (****)	-0.5251 (****)	-0.1829 (****)	0.7660 (****)	0.2292 (****)
		Liquidity	1.1073 (ns)	2.0251 (ns)	0.7056 (ns)	-2.9543 (ns)	-0.8837 (***)
		LT Debt	-0.3299 (****)	-0.6034 (****)	-0.2102 (****)	0.8802 (***)	0.2633 (***)
		ROA	0.3262 (****)	0.5966 (****)	0.2079 (****)	-0.8702 (****)	-0.2603 (****)
		Tangibles	0.0233 (ns)	0.0425 (ns)	0.0148 (ns)	-0.0620 (ns)	-0.018 (ns)
		Working Capital	-0.0569 (****)	-0.1042 (****)	-0.0363 (****)	0.1520 (****)	0.0455 (****)
		acquisti	0.0131 (***)	0.0240 (**)	0.0083 (**)	-0.0349 (**)	-0.010 (**)
		Turnover	0.0584 (****)	0.1068 (****)	0.0372 (****)	-0.1558 (****)	-0.0466 (****)
		R1	0.0168 (*)	0.0295 (*)	0.0091 (*)	-0.0431 (*)	-0.0123 (*)
		R2	0.0057 (ns)	0.0102 (ns)	0.0033 (ns)	-0.0149 (ns)	-0.0043 (ns)
		R3	-0.0233 (**)	-0.0464 (**)	-0.0210 (*)	0.0675 (**)	0.0232 (**)
		N1	-0.0279 (ns)	-0.0540 (ns)	-0.0227 (ns)	0.0785 (ns)	0.0260 (ns)
		N2	-0.0225 (ns)	-0.0405 (ns)	-0.0136 (ns)	0.0591 (ns)	0.0175 (ns)
		N3	-0.0042 (ns)	-0.0079 (ns)	-0.0029 (ns)	0.0116 (ns)	0.8485 (ns)
		N4	0.0116 (ns)	0.0199 (ns)	0.0055 (ns)	-0.0291 (ns)	0.6509 (ns)
	Dynamic	Current liabilities	-0.0509 (****)	-0.3914 (****)	-0.2904 (****)	0.7069 (****)	0.0259 (****)
		Liquidity	0.2106 (ns)	1.6164 (ns)	1.1993 (ns)	-2.9192 (ns)	-0.1071 (ns)
		LT Debt	-0.0582 (****)	-0.4471 (****)	-0.3317 (****)	0.8074 (****)	0.0296 (****)
		Working Capital	-0.0114 (***)	-0.0878 (**)	-0.0651 (**)	0.1586 (**)	0.0058 (**)
		acquisti	0.0018 (*)	0.0142 (*)	0.0105 (*)	-0.0257 (*)	-0.0009 (*)
		ebit	0.0409 (**)	0.3143 (**)	0.2332 (**)	-0.5676 (**)	-0.0208 (**)
		Turnover	0.0099 (****)	0.0757 (****)	0.0562 (****)	-0.1368 (****)	-0.005 (****)
		R1	0.0020 (ns)	0.0151 (ns)	0.0107 (ns)	-0.0269 (ns)	-0.0009 (ns)
		R2	0.0023 (ns)	0.0169 (ns)	0.0118 (ns)	-0.0299 (ns)	-0.0011 (ns)
		R3	-0.0034 (*)	-0.0274 (ns)	-0.0231 (ns)	0.0518 (ns)	0.0021 (ns)
		LagRating_4	-0.0157 (****)	-0.1522 (****)	-0.2306 (****)	0.3613 (****)	0.0371 (****)
		LagRating_5	-0.0314 (****)	-0.2405 (****)	-0.3465 (****)	0.5251 (****)	0.0933 (****)
		LagRating_6	-0.1026 (****)	-0.4008 (****)	-0.3555 (****)	0.6173 (****)	0.2417 (****)
		LagRating_7	-0.0264 (****)	-0.2293 (****)	-0.5007 (****)	-0.2166 (****)	0.9730 (****)

Table D6: Table of PB marginal effects for SEC variables.

Model	Version	Variables	Marginal effects y = 3	y = 4	y = 5	y = 6	y = 7
PB	Static	New Receivables	-0.052 (**)	-0.046 (**)	-0.0132 (**)	0.0727 (**)	0.0384 (**)
		Outstanding	0.0242 (****)	0.0217 (****)	0.0062 (****)	-0.0341 (****)	-0.0180 (****)
		Delinquency	-0.2319 (****)	-0.2080 (***)	-0.0593 (****)	0.3267 (****)	0.1725 (****)
	Dynamic	Collections	0.0042 (ns)	0.0164 (ns)	0.0084 (ns)	-0.0271 (ns)	-0.0021 (ns)
		Outstanding	0.0039 (****)	0.0158 (****)	0.0081 (****)	-0.0259 (****)	-0.0019 (****)
		Delinquency	-0.0542 (**)	-0.2143 (**)	-0.1103 (**)	0.3519 (**)	0.0269 (**)
		LagRating_4	-0.0342 (****)	-0.1718 (****)	-0.1851 (****)	0.3291 (****)	0.0621 (****)
		LagRating_5	-0.0656 (****)	-0.2709 (****)	-0.2998 (****)	0.4692 (****)	0.1670 (****)
		LagRating_6	-0.1835 (****)	-0.4003 (****)	-0.2866 (****)	0.4951 (****)	0.3755 (****)
		LagRating_7	-0.0541 (****)	-0.2563 (****)	-0.4291 (****)	-0.2386 (****)	0.9769 (****)

Table D7: Table of PB marginal effects for BS+SEC variables.

Model	Historical	Variables	Marginal effects y = 3	y = 4	y = 5	y = 6	y = 7
PB	Static	Delinquency	-0.1739 (****)	-0.2928 (****)	-0.0834 (****)	0.4206 (****)	0.1295 (****)
		Turnover	0.0659 (****)	0.1110 (****)	0.0316 (****)	-0.1594 (****)	-0.0491 (****)
		Working Capital	-0.0884 (****)	-0.1488 (****)	-0.0424 (****)	0.2138 (****)	0.0658 (****)
		LT Debt	-0.3978 (****)	-0.6697 (****)	-0.1908 (****)	0.9619 (****)	0.2963 (****)
		Current liabilities	-0.2948 (****)	-0.4963 (****)	-0.1414 (****)	0.7129 (****)	0.2196 (****)
		Liquidity	1.9809 (***)	3.3348 (***)	0.9503 (***)	-4.7904 (***)	-1.4757 (***)
	Dynamic	Delinquency	-0.0399 (**)	-0.2439 (**)	-0.1881 (**)	0.7069 (**)	0.4012 (**)
		Turnover	0.0141 (****)	0.0861 (****)	0.0482 (****)	-2.9192 (****)	-0.1416 (****)
		Working Capital	-0.0179 (****)	-0.1095 (****)	-0.0613 (****)	0.8074 (****)	0.1802 (****)
		LT ebt	-0.0850 (****)	-0.5190 (**)	-0.2905 (**)	0.1586 (**)	0.8537 (**)
		Current liabilities	-0.0639 (*)	-0.3902 (*)	-0.2184 (*)	0.0257 (*)	0.6418 (*)
		Liquidity	0.4997 (**)	3.0508 (**)	1.077 (**)	-0.5676 (**)	-5.0185 (**)
		LagRating_4	-0.0211 (****)	-0.1522 (****)	-0.1881 (****)	0.3613 (****)	0.0385 (****)
		LagRating_5	-0.0386 (****)	-0.2405 (****)	-0.2978 (****)	0.5251 (****)	0.0959 (****)
		LagRating_6	-0.1111 (****)	-0.4008 (****)	-0.3212 (****)	0.6173 (****)	0.2354 (****)
		LagRating_7	-0.0331 (****)	-0.2293 (****)	-0.4708 (****)	-0.2166 (****)	0.9537 (****)

D.3 Feature importance

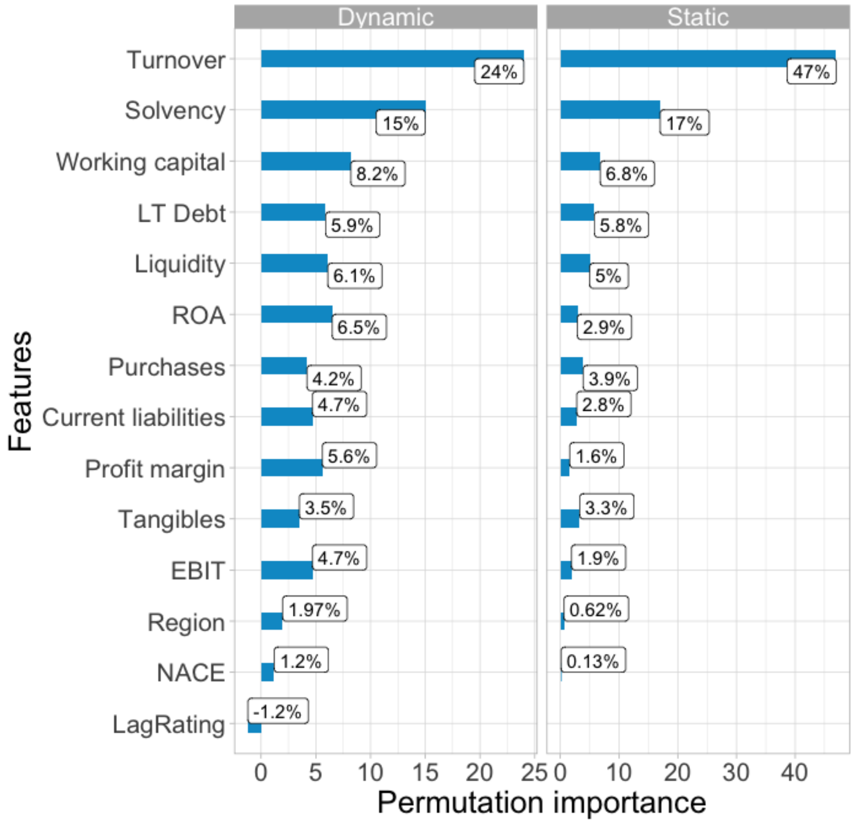


Figure D1: Macro-averaged relative permutation importance for HRF model for BS set (dynamic and static case).

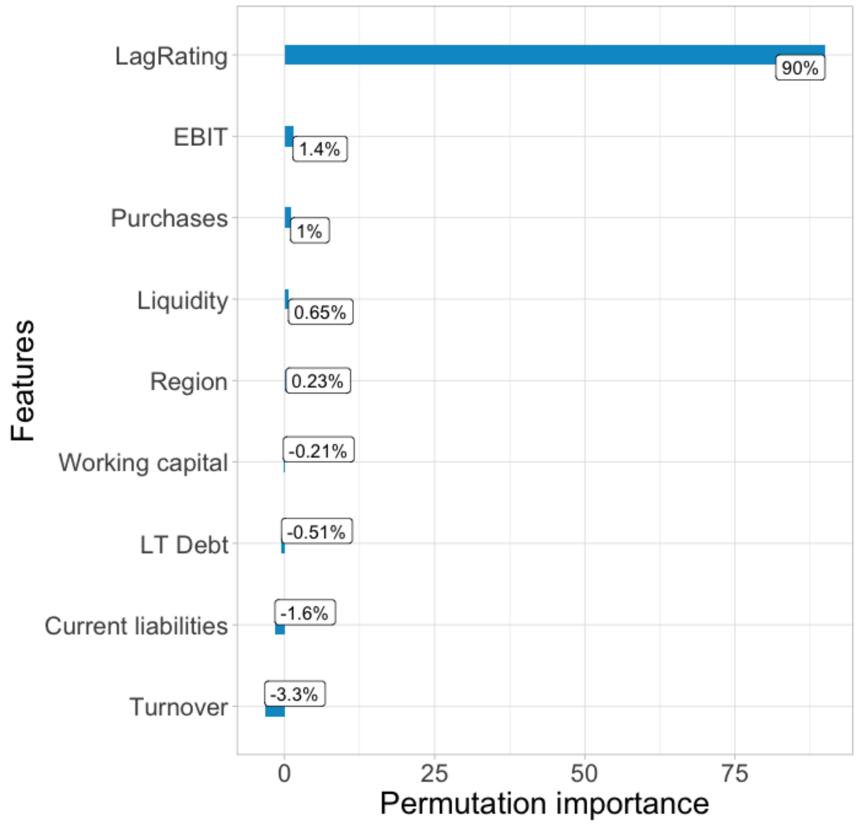


Figure D2: Macro-averaged relative permutation importance for PB model for BS set (dynamic case).

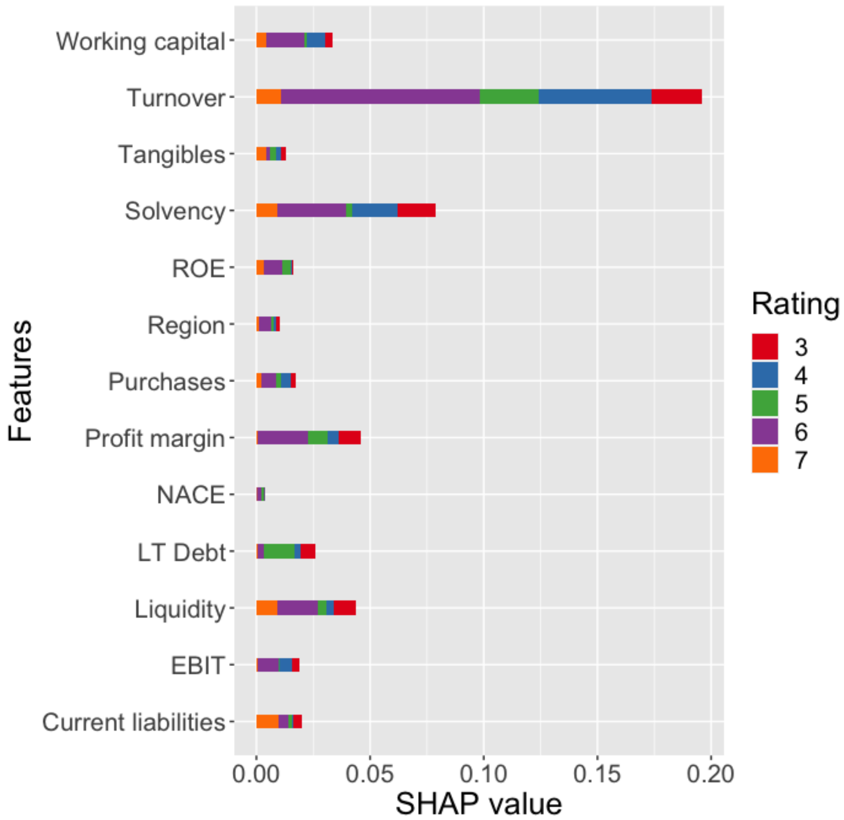


Figure D3: SHAP value (average impact of predictors for each class) for dynamic HRF model with regards to BS set.

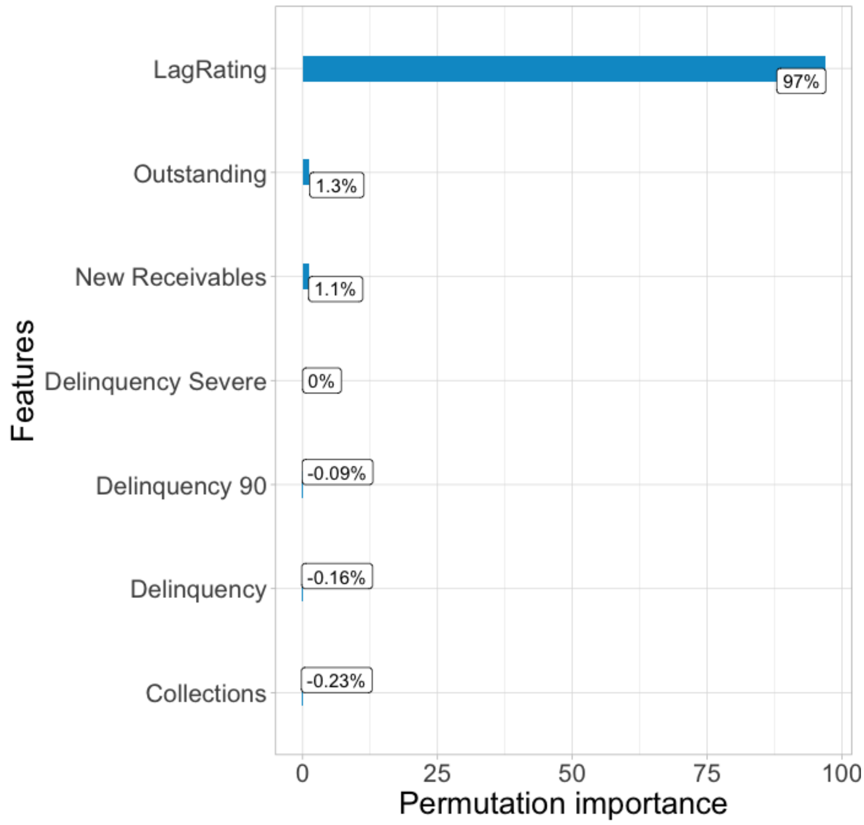


Figure D4: Macro-averaged relative permutation importance for HRF model for SEC set (dynamic case).

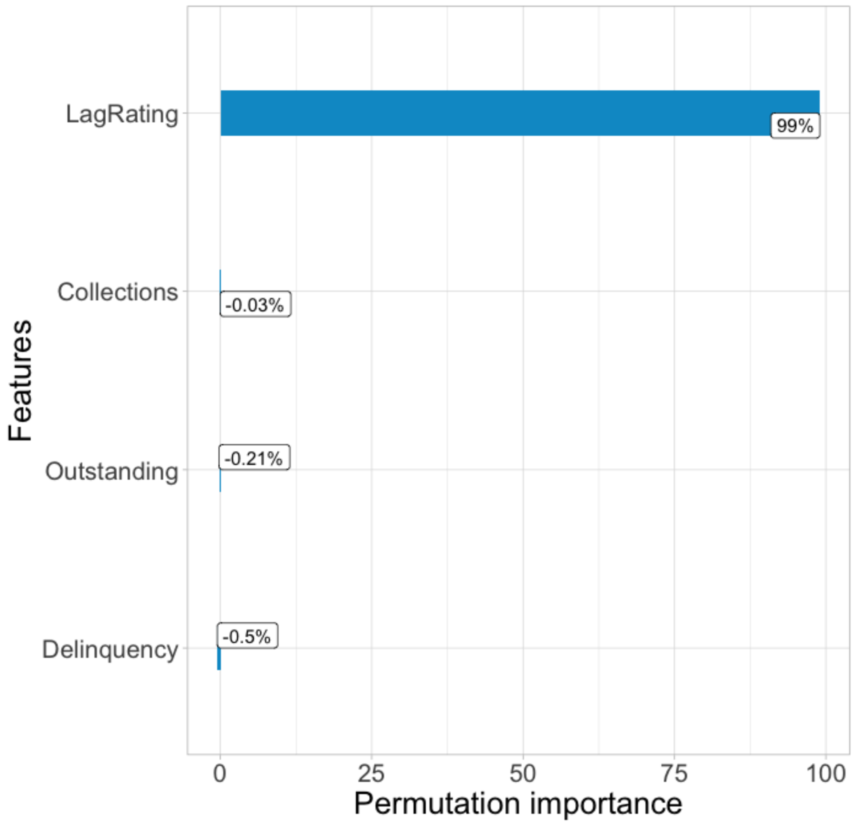


Figure D5: Macro-averaged relative permutation importance for PB model for SEC set (dynamic case).

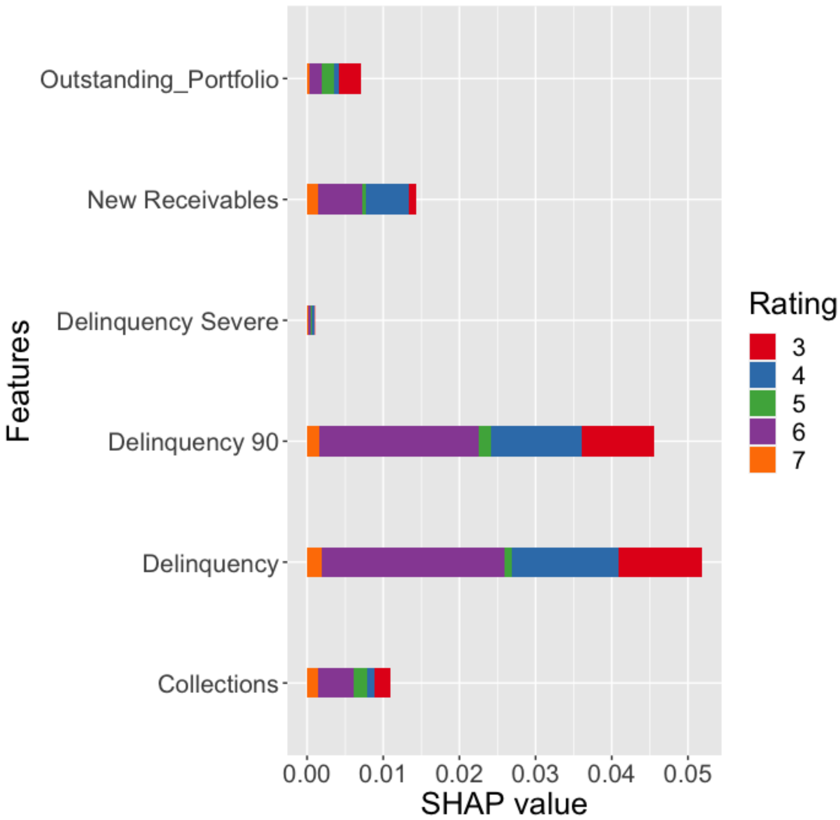


Figure D6: SHAP value (average impact of predictors for each class) for dynamic HRF model with regards to SEC set.

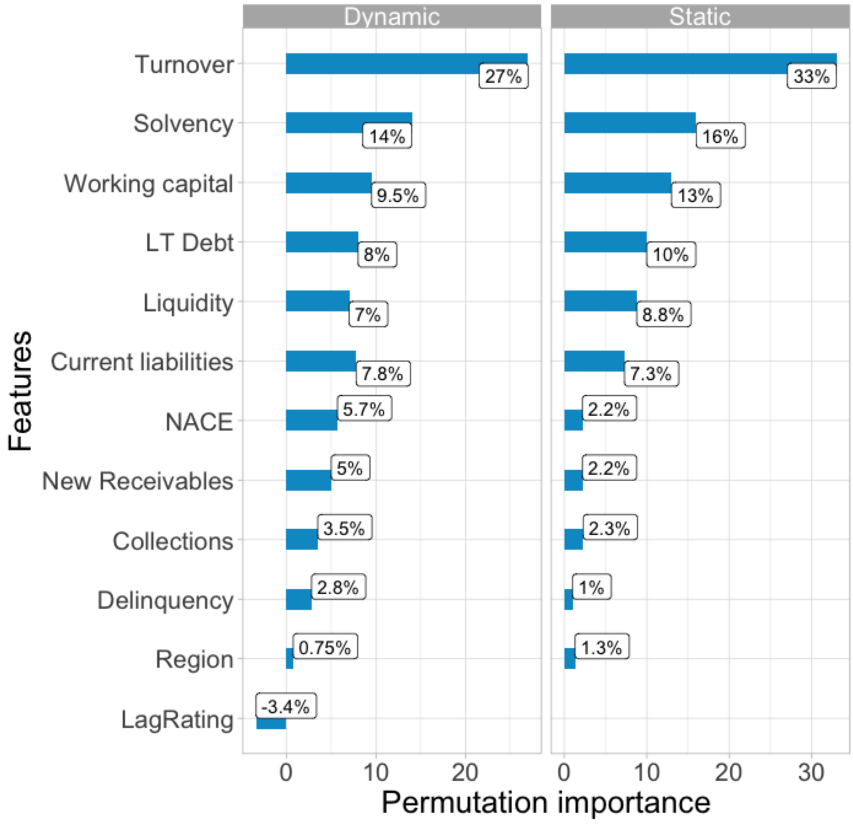


Figure D7: Macro-averaged relative permutation importance for HRF model for SEC+BS set (dynamic and static case).

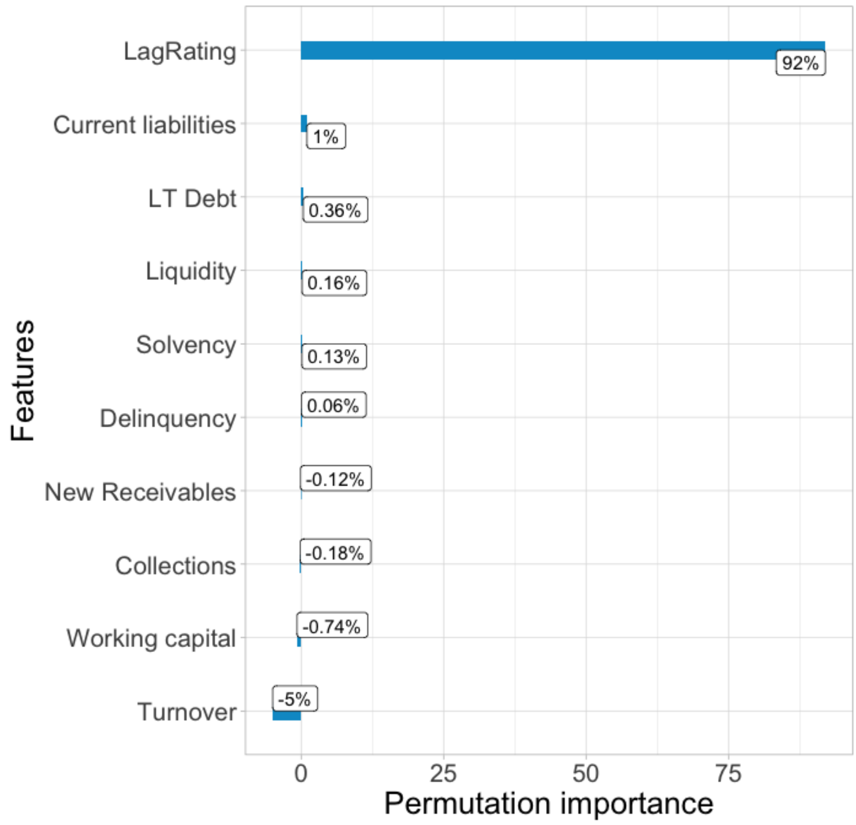


Figure D8: Macro-averaged relative permutation importance for PB model for SEC+BS set (dynamic case).

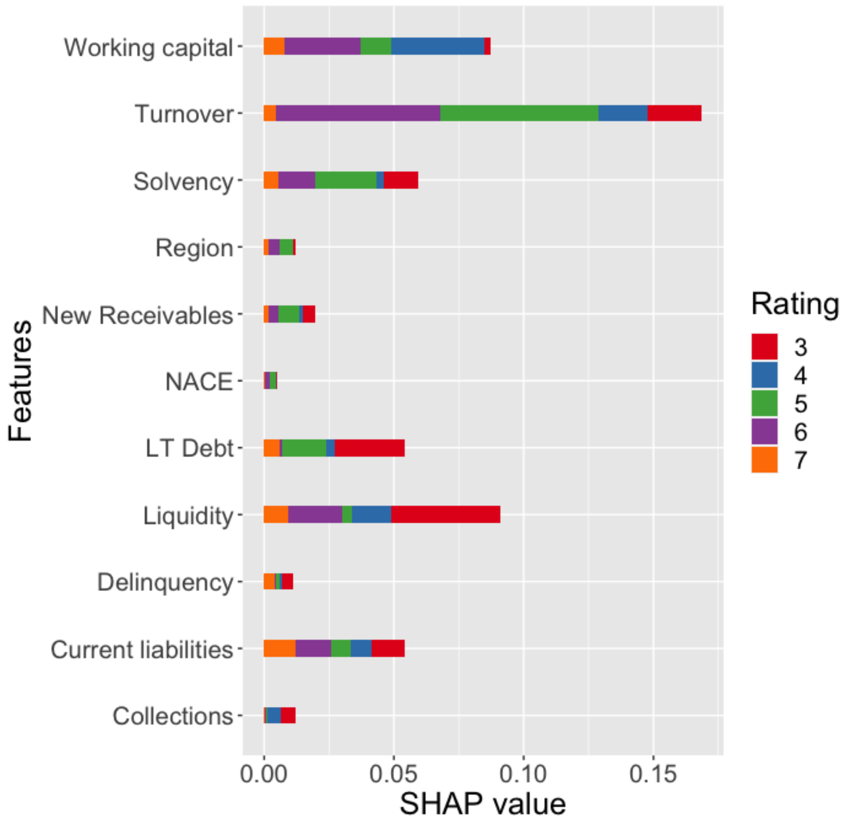


Figure D9: SHAP value (average impact of predictors for each class) for dynamic HRF model with regards to BS+SEC set.

D.4 Missing Values handling

In order to impute missing values for BS variable the following approach was used:

- for each BS variable, evaluate the average percentage increase/decrease of consecutive years:

$$\Delta_{t+1,t} = \frac{1}{N}\sum_{i=1}^{N} \frac{BS_i^{t+1}}{BS_i^t} - 1,$$

  where $t = 2015, 2016$ is the reference year and N it the total number of observations.

- impute missing value for t-th year given the $(t+1)$-th year by:

$$BS_i^t = \frac{BS_i^{t+1}}{1+\Delta_{t+1,t}}, \quad t = 2015, 2016$$

  for single missing year and impute value for t-th year given the $(t+2)$-th year by:

$$BS_i^t = \frac{BS_i^{t+2}}{(1+\Delta_{t+1,t})(1+\Delta_{t+2,t+1})}$$

  for double consecutive missing years.

Missing values for leading or trailing quarters of 2015 and 2017, respectively, are allowed for SEC variables given the unbalanced panel nature of the data.

E Understanding corporate default using Random Forest: The role of accounting and market information

E.1 Dataset

Table E1: Correlation matrix of input variables for MSMEs. Legend is below:
1 is 'Oth Reven on Reven', 2 is 'Deprec on Costs', 3 is 'Pay to Bank on Assets', 4 is 'Cashflow on Reven', 5 is 'Fixed Asset Cov', 6 is 'Labor Cost on Reven', 7 is 'ST Pay on Due to Bank', 8 is 'Tot Debt on ST Debt', 9 is 'Tot Debt on Net Worth', 10 is 'Pay to Suppl on Net Worth', 11 is 'Pay to Suppl on Tot Debt', 12 is 'Inventory Duration', 13 is 'Quick Ratio', 14 is 'Debt Burden Index', 15 is 'Fin Int on Reven', 16 is 'Fin Int on Added Val', 17 is 'Net Worth on LT Eqt/Pay', 18 is 'Net Worth on NW+Invent', 19 is 'ROA', 20 is 'ROD', 21 is 'Working Cap Turnover', 22 is 'Turnover'

	1	2	3	4	5	6	7	8	9	10	11
1	1										
2	0.19***	1									
3	0.11***	0.38***	1								
4	0.16***	0.52***	0.2***	1							
5	-0.05***	-0.18***	-0.16***	-0.01**	1						
6	-0.1***	-0.14***	-0.13***	-0.38***	-0.08***	1					
7	-0.04***	-0.13***	-0.28***	-0.06***	0.14***	0.05***	1				
8	0.1***	0.31***	0.55***	0.2***	-0.11***	-0.16***	-0.36***	1			
9	0.03***	-0.09***	0.06***	-0.21***	-0.2***	0.1***	-0.02**	0.03***	1		
10	-0.01	-0.19***	-0.11***	-0.26***	-0.13***	0.11***	0.17***	-0.15***	0.82***	1	
11	-0.11***	-0.32***	-0.4***	-0.13***	0.21***	0	0.55***	-0.52***	-0.03***	0.29***	1
12	0.21***	0	-0.04***	-0.01	0.06***	-0.09***	-0.07***	0.11***	0.05***	0.01**	-0.1***
13	-0.04***	0.05***	-0.12***	0.16***	0.18***	-0.09***	-0.16***	0.35***	-0.15***	-0.21***	-0.19***
14	0	-0.06***	0.04***	-0.11***	-0.02***	0.2***	-0.09***	0.03***	0.12***	0.08***	-0.06***
15	0.25***	0.37***	0.45***	0.23***	-0.17***	-0.2***	-0.29***	0.46***	0.1***	-0.07***	-0.42***
16	0.05***	0.02**	0.25***	-0.09***	-0.1***	-0.09***	-0.27***	0.31***	0.2***	0.11***	-0.25***
17	-0.05***	0.02**	-0.21***	0.24***	0.28***	-0.12***	0.33***	-0.33***	-0.57***	-0.47***	0.34***
18	-0.02***	0.29***	0.12***	0.36***	0.04***	-0.16***	0.05***	0.04***	-0.45***	-0.47***	-0.07***
19	-0.08***	-0.04***	-0.12***	0.52***	0.19***	-0.29***	0.08***	-0.09***	-0.21***	-0.16***	0.14***
20	-0.01	0.06***	0.13***	0.06***	-0.14***	-0.08***	-0.3***	0.22***	0.02**	-0.08***	-0.32***
21	-0.15***	-0.02**	0.34***	0	-0.08***	0.02***	0.02**	0.03***	-0.06***	-0.06***	0
22	-0.17***	-0.4***	-0.25***	-0.25***	0.2***	0.09***	0.18***	-0.26***	0.02**	0.14***	0.3***

	12	13	14	15	16	17	18	19	20	21	22
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12	1										
13	-0.17***	1									
14	0.21***	-0.07***	1								
15	0.31***	0	0.26***	1							
16	0.24***	-0.05***	0.36***	0.6***	1						
17	-0.08***	0.06***	-0.05***	-0.27***	-0.38***	1					
18	-0.4***	0.31***	-0.18***	0	-0.23***	0.43***	1				
19	-0.16***	0.15***	-0.31***	-0.2***	-0.26***	0.31***	0.23***	1			
20	0.02***	0.11***	0.18***	0.52***	0.46***	-0.21***	-0.02***	-0.03***	1		
21	-0.28***	-0.17***	-0.1***	-0.16***	-0.11***	0.04***	0.13***	0.13***	0.1***	1	
22	-0.3***	-0.07***	-0.11***	-0.47***	-0.2***	0.08***	-0.09***	0.25***	-0.11***	0.41***	1

Table E2: List of input variables for peers dataset.

Variable	Description	Mean	St.Dev.	Min	5th perc	Median	95th perc	Max
1 - Oth Reven on Reven	Other revenues on revenues	0.03	0.08	0	0	0.01	0.08	0.93
2 - Deprec on Costs	Depreciation on costs	0.08	0.11	0	0	0.05	0.31	0.72
3 - Pay to Bank on Assets	Payables to banks on current assets	-1.48	9.64	-90.58	-10.6	0.19	2.74	16.84
4 - Cashflow on Reven	Cash flow on revenues	-3.4	41.62	-526.34	-0.31	0.05	0.26	0.71
5 - Fixed Asset Cov	Fixed asset coverage	14.52	137.75	-0.19	0.49	1.13	2.86	1727.34
6 - Labor Cost on Reven	Labor cost on revenues	-0.06	10.79	-125.98	-0.05	0.69	1.43	37.86
7 - ST Pay on Due to Bank	Short-term payables on amounts due to banks	26.86	81.79	0.51	0.97	4.31	100	924.29
8 - Tot Debt on ST Debt	Total debt on short-term debts	1.73	1.05	1.01	1.07	1.38	3.37	7.28
9 - Tot Debt on Net Worth	Total debt on net worth	2.42	9.67	-72.91	0.24	1.61	6.72	68.4
10 - Pay to Suppl on Net Worth	Payables to suppliers on Net worth	0.71	2.28	-6.31	0.04	0.35	1.84	17.8
11 - Pay to Suppl on Tot Debt	Payables to suppliers on Total debt	0.28	0.17	0.02	0.04	0.25	0.59	0.75
12 - Inventory Duration	Inventory duration	0.79	1.15	0	0	0.5	2.26	7.13
13 - Quick Ratio	Quick ratio	1.25	1.07	0.09	0.3	1	2.47	9.41
14 - Debt Burden Index	Debt burden index	0.28	3.07	-16.8	-1.65	0.16	1.58	30.5
15 - Fin Int on Reven	Financial inrerest on revenues	3.2	38.5	0	0	0.02	0.39	486.94
16 - Fin Int on Added Val	Financial inrerest on added value	-0.19	3.05	-28.69	0	0.07	0.7	5.86
17 - Net Worth on LT Eqt/Pay	Net worth on long-term equity and payables	0.62	0.48	-3.82	0.25	0.7	0.96	0.99
18 - Net Worth on NW+Invent	Net worth on net worth and inventories	0.75	0.37	-2.92	0.42	0.76	1	2.35
19 - ROA	Return on Assets	0	0.09	-0.48	-0.18	0.01	0.09	0.2
20 - ROD	Return on Debt	0.11	0.31	-0.15	-0.04	0	1	1
21 - Working Cap Turnover	Working capital turnover	2.18	5.61	0	0.13	1.25	5.43	69.26
22 - Turnover	Turnover normalized by Total Assets	0.8	0.46	0	0.1	0.78	1.64	2.12
Total Assets	Total Assets (EUR Mln)	201.85	329.77	4.91	9.45	72.93	775.71	1621.96
Total Liabilities	Total Liabilities (EUR Mln)	66.82	243.67	0	0	7.42	118.94	1742.64
Volatility	Assets Volatility	0.52	0.82	0.01	0.04	0.21	2.31	4.18

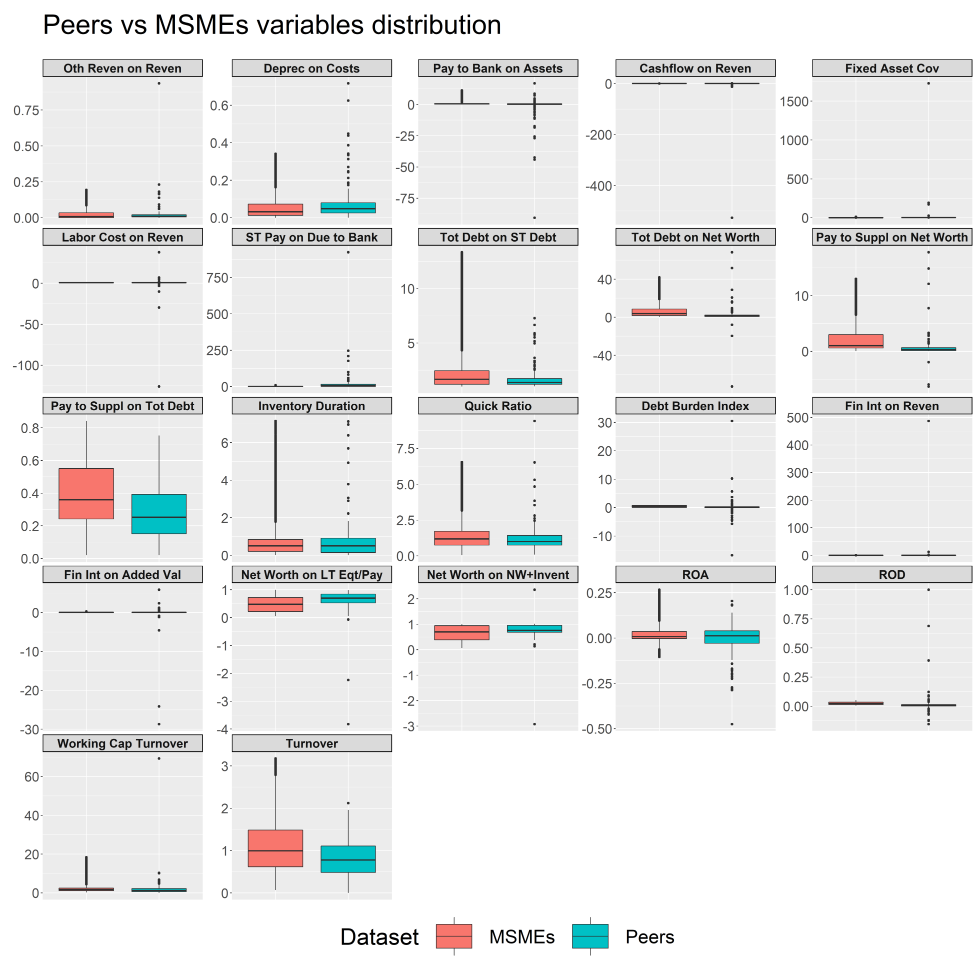


Figure E1: Distribution of input variables for Peers and MSMEs.

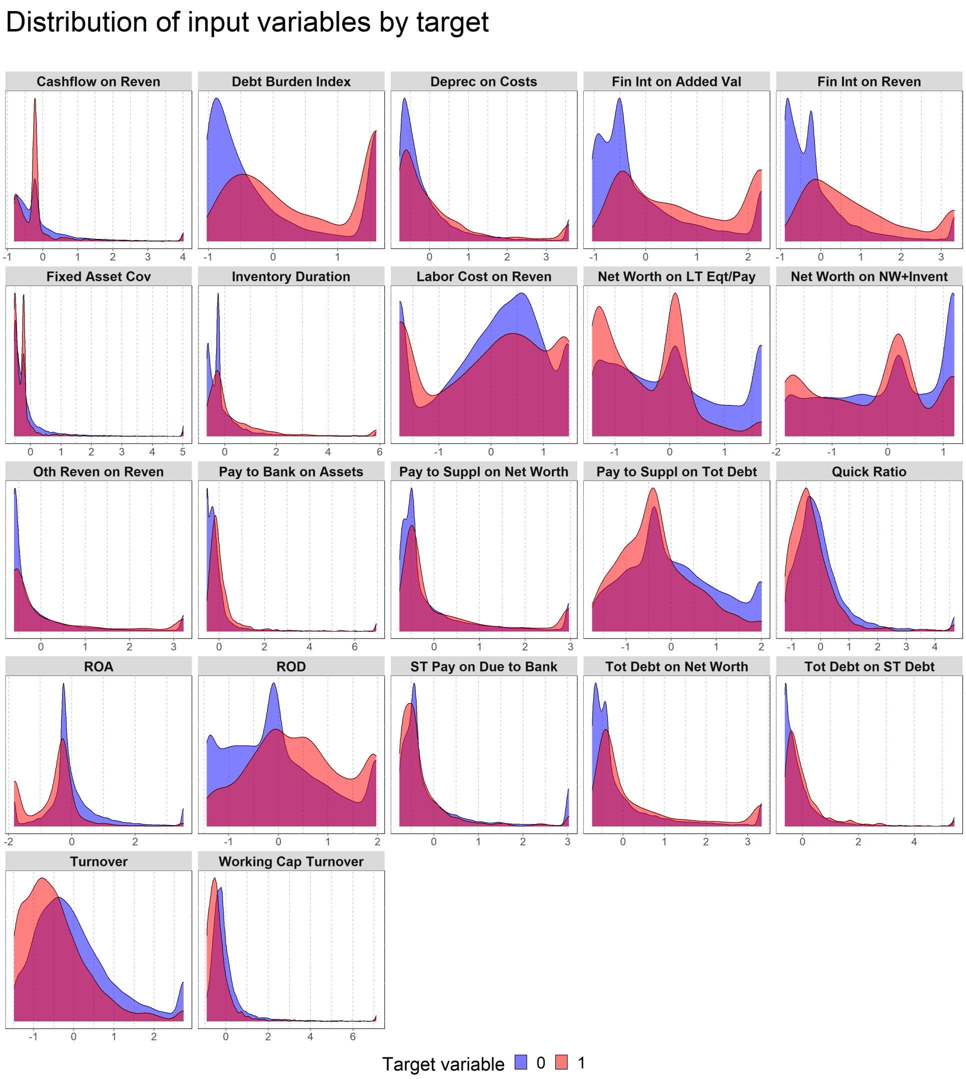


Figure E2: Distribution of input variables for MSMEs splitted by target variable.

Table E3: Distribution of clients that are persistent over time, i.e. target is always 0 or 1, compared with clients that move from 0 to 1 and vice-versa.

Target	Total clients	Total banks
0	17,943	9,228
1	876	446
0 (0->1)	388	388
0 (1->0)	74	74
1 (0->1)	388	
1 (1->0)	74	
Total	19,743	10,136

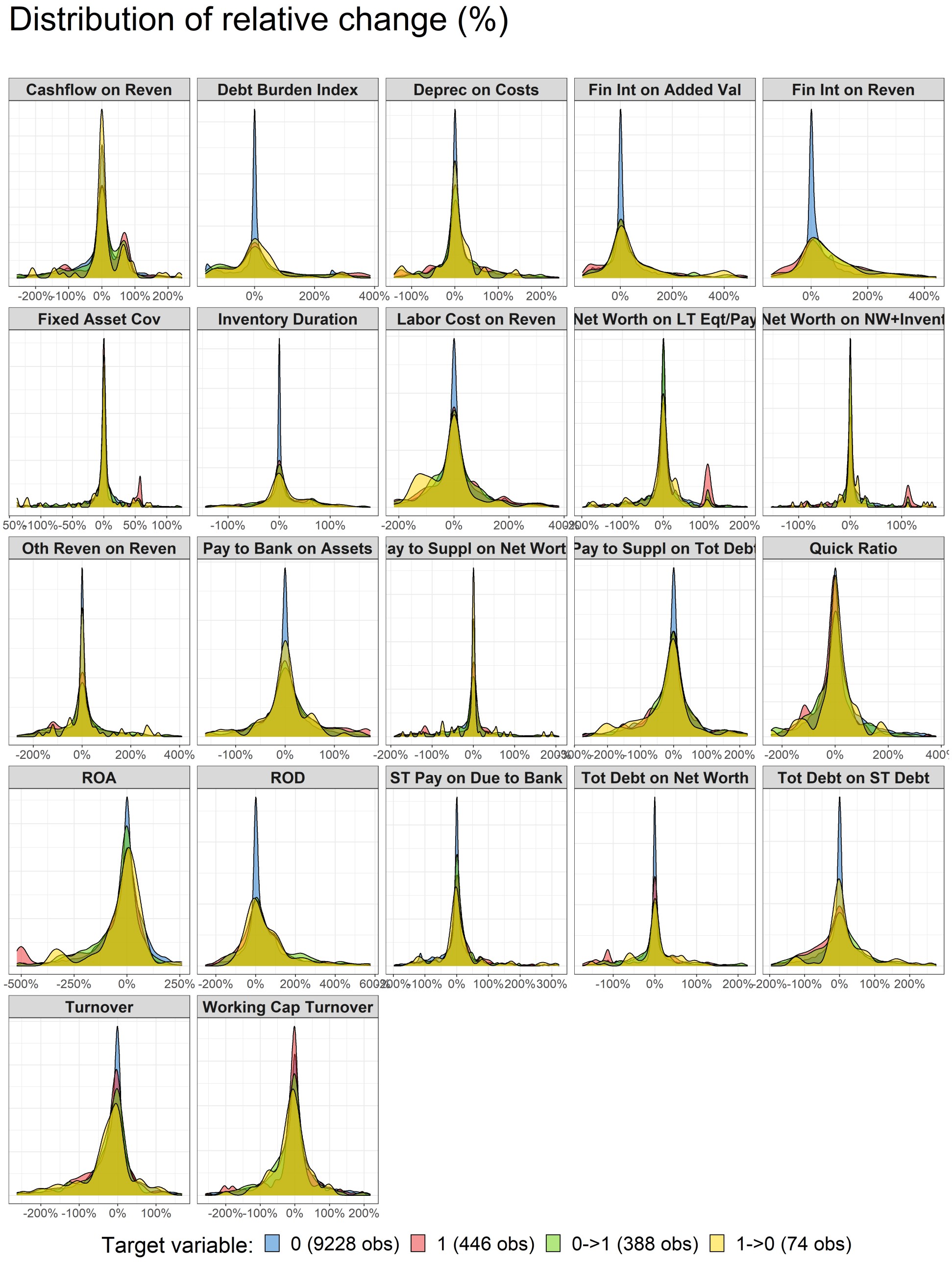


Figure E3: Distribution of relative changes over the years of each input variable divided by clients' behavior. Blue and red distributions represent the clients with persistent target of 0 and 1, respectively, green and yellow distributions represent the clients that moved from 0 to 1 and vice-versa, respectively.

E.2 Results

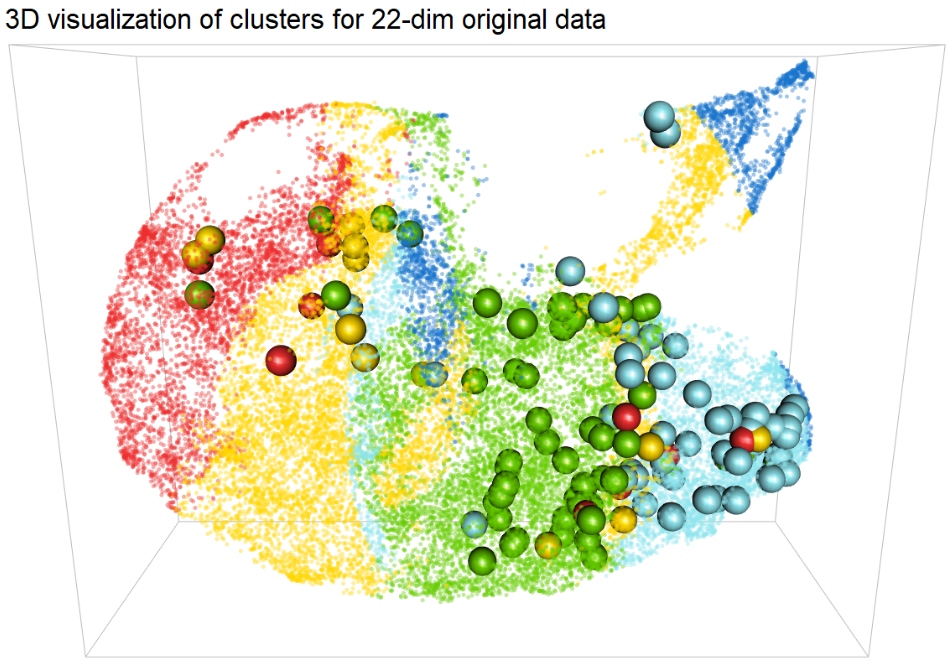


Figure E4: 3D visualization of five clusters for the 22-dimensional original data. Visual embedding is evaluated with UMAP algorithm. Small points are MSMEs observations, bold spheres are peers' observations.

Table E4: F1-score and AUC for Elastic-Net, MARS and Random Forest calibrated on dataset with input variables only and with the addition of PD and with or without controls for fixed effects. Values refer to performance of model calibrated on the entire dataset. Values in parenthesis refer to average performance of validation folds of Cross-Validation.

		F1 (Cross-Val)		AUC (Cross-Val)	
Control	Algorithm	Baseline	With PD	Baseline	With PD
	Elastic-Net	30.7% (30.1±1.7%)	35.1% (35.1±1.5%)	79.8% (79.6±0.6%)	82% (81.7±0.8%)
No control	MARS	36% (33.8±1.4%)	40% (37.5±0.6%)	82.5% (81.7±0.6%)	84.2% (82.8±0.8%)
	Random Forest	89.5% (85.1±1.7%)	95.8% (91.4±1.2%)	89.8% (85.4±1.1%)	96.1% (91.7±0.7%)
	Elastic-Net	30.7% (30.7±1.3%)	35.1% (35±3%)	79.8% (79.5±1%)	82% (81.8±1.3%)
Dummy Industry	MARS	34.2% (34.4±1.8%)	38.8% (37.5±2.8%)	82.4% (81.9±0.4%)	83.8% (83.2±1.2%)
	Random Forest	90.5% (87.3±1.9%)	95.9% (93.4±2.8%)	90.8% (87.6±1%)	96.2% (93.7±1.6%)
	Elastic-Net	30.9% (30.8±0.9%)	35.3% (35.3±1.4%)	79.9% (79.8±1.6%)	82.5% (82.4±1.2%)
Firm Size	MARS	37.3% (35.4±0.8%)	41.3% (39.3±2.3%)	83.4% (82.2±1.2%)	84.5% (83.3±1.3%)
	Random Forest	90.7% (88.5±2.7%)	96% (91.3±1.9%)	91% (88.8±1.5%)	96.3% (91.6±1.6%)
	Elastic-Net	30.8% (30.8±1.2%)	35.4% (35.1±1.7%)	79.8% (79.6±1.2%)	82.2% (82.1±1.3%)
Firm Type	MARS	36.2% (34.6±2.1%)	40.5% (37.7±3.3%)	82.9% (81.8±1.3%)	84.7% (82.8±1.4%)
	Random Forest	89.5% (87±1.4%)	96.1% (91.5±2.8%)	89.8% (87.3±1%)	96.4% (91.8±1.2%)
	Elastic-Net	31.3% (31.3±1.7%)	35.4% (34.9±1.5%)	80.1% (80±2%)	82.3% (82±1.6%)
Industrial Sector	MARS	34.3% (33.8±2%)	40.3% (36.9±2.8%)	82.4% (81.9±1.7%)	84.6% (82.3±1.5%)
	Random Forest	93.4% (90.2±1.6%)	97.3% (94.7±2%)	93.6% (90.4±1.5%)	97.6% (95±1.4%)
	Elastic-Net	30.9% (30.7±1.6%)	35.1% (35±2.6%)	79.8% (79.6±1.9%)	82.1% (81.9±2.1%)
Region	MARS	34.3% (34.1±1.5%)	37% (36.6±2.7%)	82.4% (82±2.3%)	83.8% (83.1±2.2%)
	Random Forest	92.4% (89.5±1.4%)	97.5% (95.3±2.7%)	92.7% (89.8±2.2%)	97.8% (95.5±2.2%)

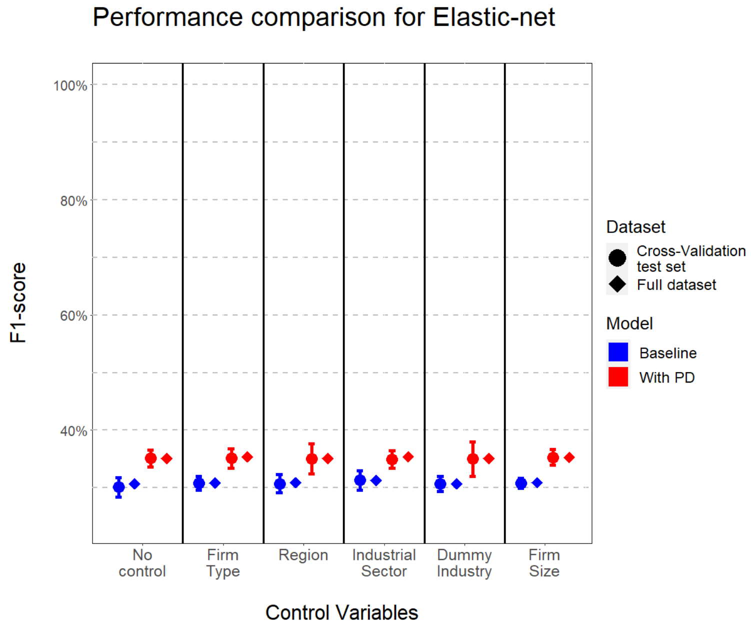


Figure E5: Comparison of F1-score for Elastic-Net model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

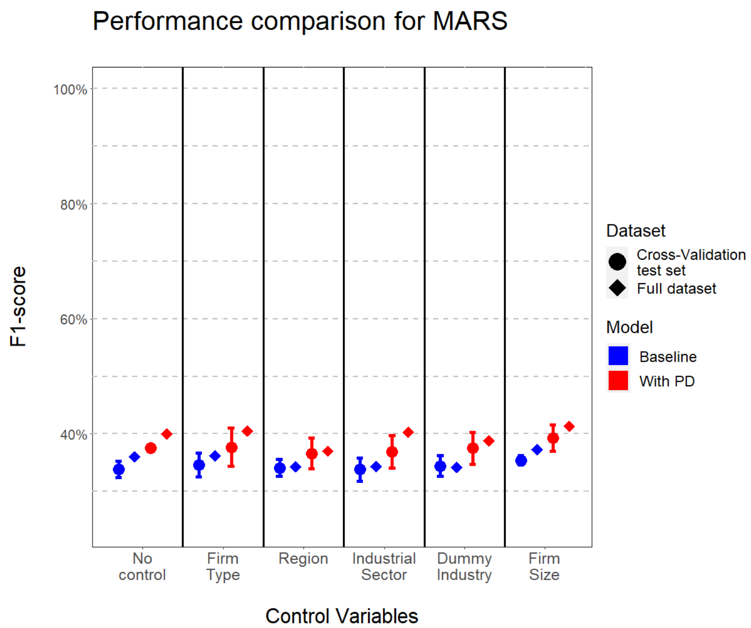


Figure E6: Comparison of F1-score for MARS model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

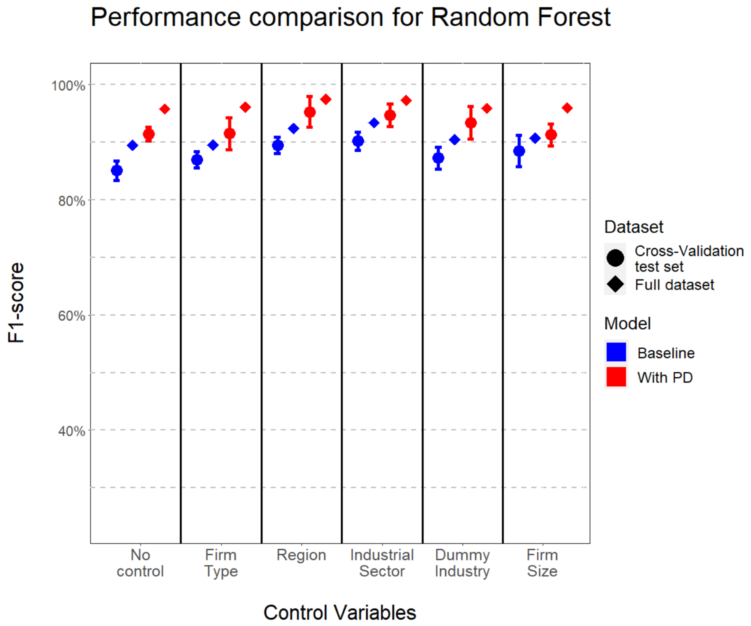


Figure E7: Comparison of F1-score for Random Forest model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

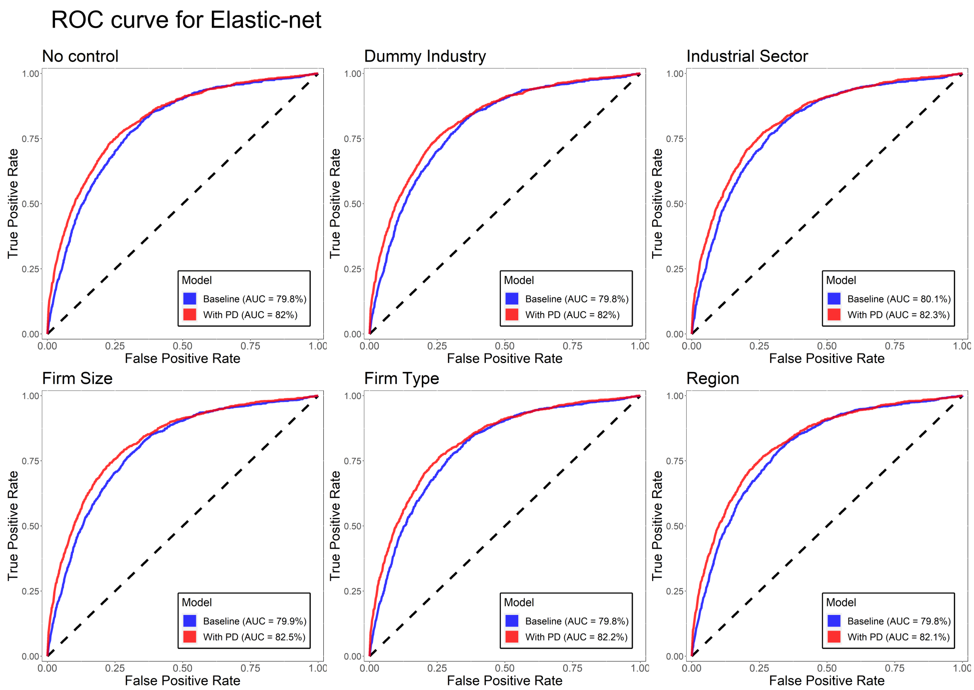


Figure E8: Comparison of ROC curves for Elastic-Net model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

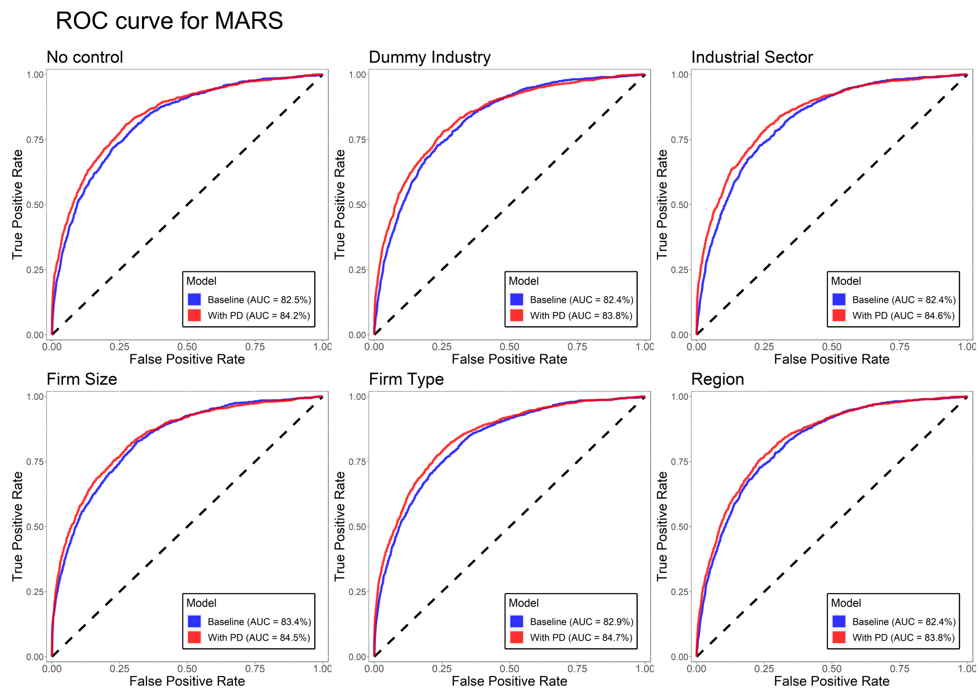


Figure E9: Comparison of ROC curves for MARS model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

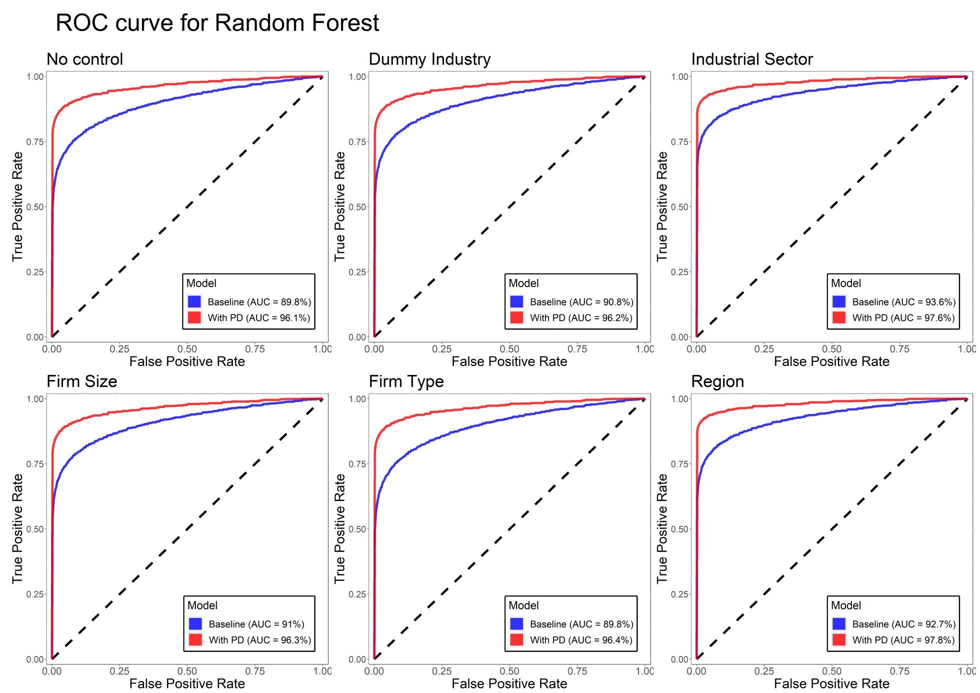


Figure E10: Comparison of ROC curves for Random Forest model for models calibrated with input variables only and with the addition of PD, as well as with or without controls for fixed effects.

E.3 Feature importance

Explainability capabilities all models PB have been compared using Permutation Feature Importance (PFI) and Shapley Additive Explanations (SHAP). The change in models' performances and in the probability correlated to each predictor has been explored in order to understand the sign of the effect on each class of the target variable.

PFI evaluates the importance of each variable by computing the gain in model's prediction error after shuffling feature's values. A feature is considered relevant for model's prediction if the prediction error increases after permuting its values, otherwise, if model error remains unchanged, its contribution is not important. As proposed by Fisher et al. (2018), the algorithm for a generic model f can be defined as:

Algorithm 3: Permutation Feature Importance

Input: Trained model f, feature matrix X, target vector y, performance metric $P(y, f)$

Estimate the original model performance $P_{\text{orig}} = f(y, X)$;
foreach *feature* $j = 1, \dots, p$ **do**
  Generate feature matrix X_{perm} by permuting feature j in the data X;
  Estimate $P_{\text{perm}} = f(y, X_{\text{perm}})$ based on the predictions of the permuted data;
  Evaluate $\text{PFI}_j = P_{\text{perm}} / P_{\text{orig}}$. Alternatively, the difference can be used: $\text{PFI}_j = P_{\text{perm}} - P_{\text{orig}}$;
  return PFI_j;
end
Sort features by descending PFI

Shapley values represent the marginal contribution of each feature to the prediction of a given data point. The feature values for instance x behave like players in a game where the prediction is the payout. As described in Shapley (1953), the Shapley value Φ_j of a feature value x_j, is defined by means of a value function val of actors in S and represents its contribution to the prediction, weighted and summed across all possible coalitions:

$$\Phi_j(val) = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{x_j\}) - val(S))$$

where S denotes a subset of features, x represents the feature values of the instance of interest and p the number of features and $val_x(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in S:

$$val_x(S) = \int \hat{f}(x_1, \dots, x_p) d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X))$$

Estimating the Shapley values for more than a few features becomes computationally infeasible since all possible coalitions of feature values need to be considered with and without feature j. A Monte-Carlo sampling was proposed by Strumbelj and Kononenko (2014):

$$\hat{\Phi}_j = \frac{1}{M} \sum_{m=1}^{M} (\hat{f}(x_{+j}^m) - \hat{f}(x_{-j}^m))$$

where $\hat{f}(x_{+j}^m)$ represents the prediction for the instance of interest x but with a random permutation of features (taken from a random data point z) except for j-th feature. The vector x_{-j}^m is identical to x_{+j}^m, but the value for feature j is randomized as well from the sampled z. The algorithm for a generic model f can be defined as:

Algorithm 4: Shapley value

Output: Shapley value for the value of the j-th feature
Input : Number of iterations M, instance of interest x, feature index j, data matrix X, and machine learning model f

foreach $m = 1, \ldots, M$ **do**
  Draw random instance z from data matrix X;
  Choose a random permutation o of the feature values;
  Order instance x: $x_O = (x_{(1)}, \ldots, x_{(j)}, \ldots, x_{(p)})$;
  Order instance z: $z_O = (z_{(1)}, \ldots, z_{(j)}, \ldots, z_{(p)})$;
  Construct two new instances:
  - With feature j: $x_{+j} = (x_{(1)}, \ldots, x_{(j-1)}, x_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
  - Without feature j: $x_{-j} = (x_{(1)}, \ldots, x_{(j-1)}, z_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$

  Compute marginal contribution: $\Phi_j^m = \hat{f}(x_{+j}) - \hat{f}(x_{-j})$;
  return Φ_j^m;
end
Compute Shapley value as the average: $\Phi_j(x) = \frac{1}{M} \sum_{m=1}^{M} \Phi_j^m$

This procedure needs to be repeated for each feature of interest in order to get all the Shapley values. Among the advantages of Shapley values over the other methods, in first place there is the efficiency property, i.e., the difference between prediction and average prediction is fairly distributed among features.

Figures from E11 to E16 report the PFI and SHAP variable importance for Elastic-Net and MARS models, calibrated with input variables and with the addition of PD as a predictor.

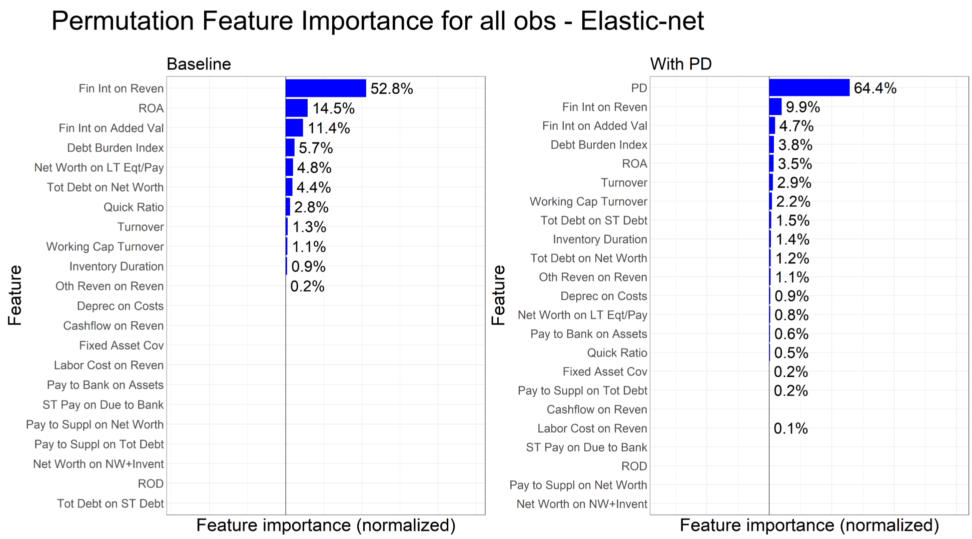
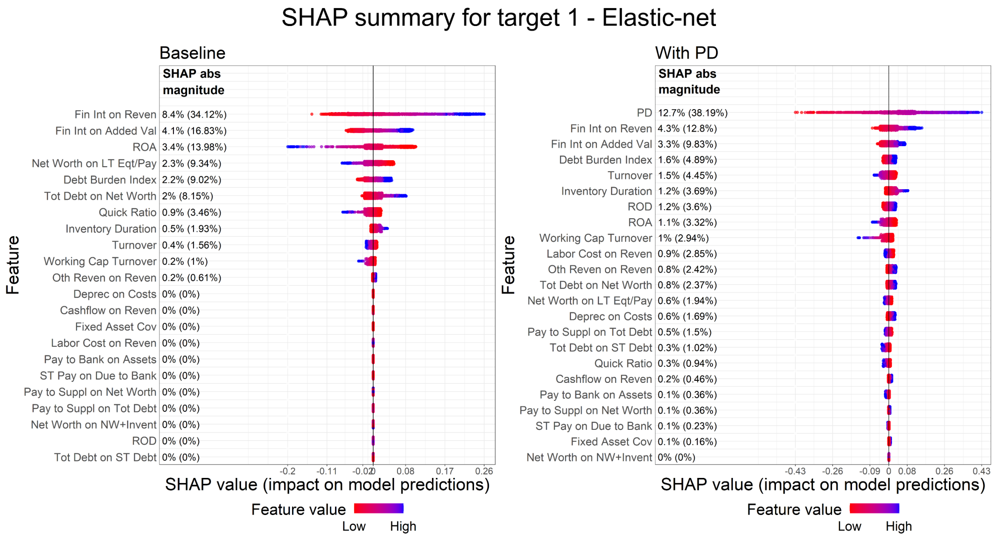
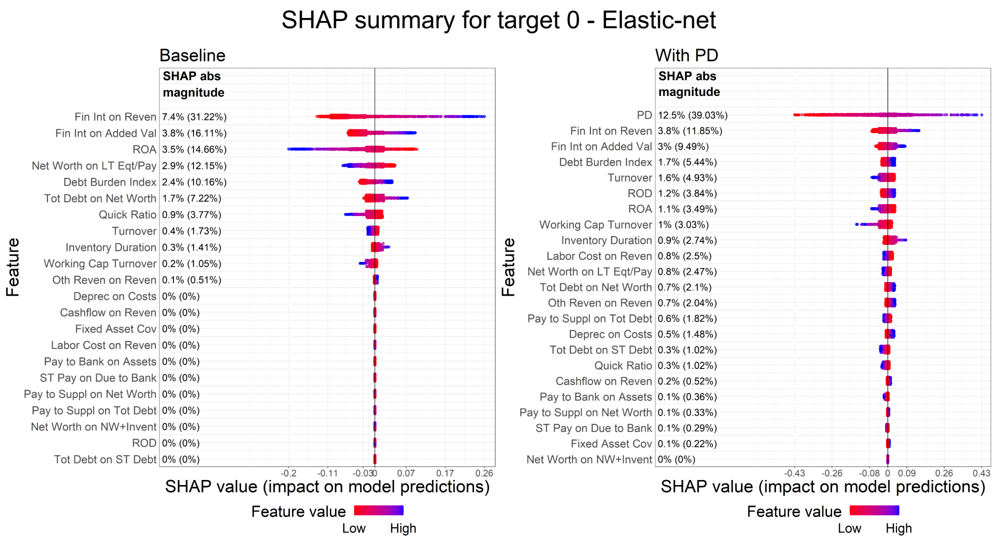


Figure E11: Permutation Feature Importance for Elastic-Net model, comparing variable importance of model calibrated with input variables and with the addition of PD. Normalized changes of F1-score are used to rank the variables.

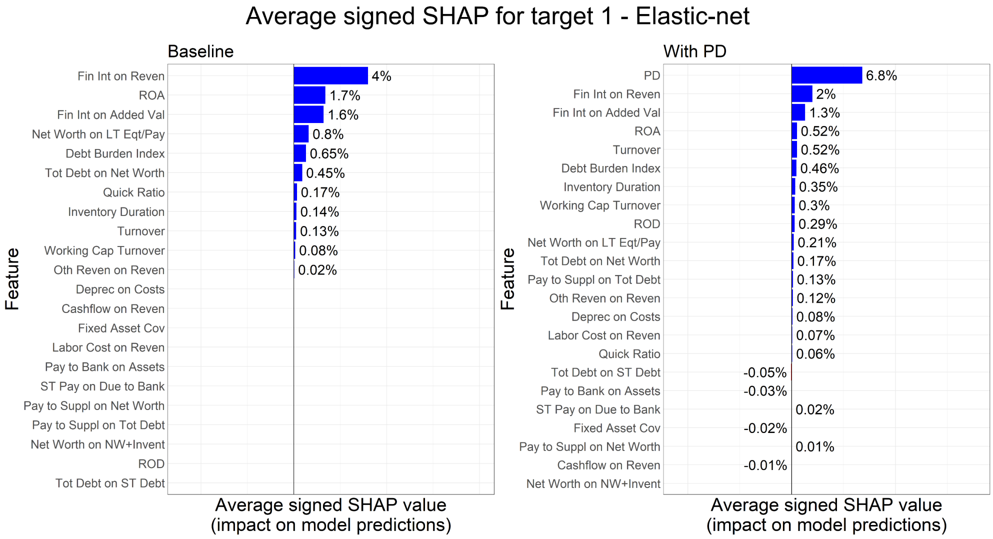


(a) Defaulted clients.

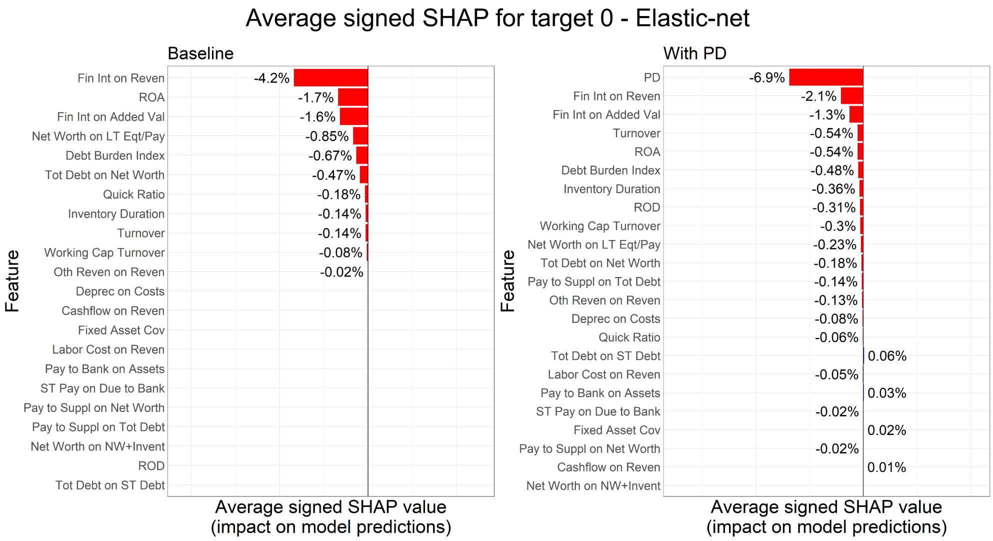


(b) Non-defaulted clients.

Figure E12: SHAP effects on predicted probability for Elastic-Net model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. The color of the points ranges from red, meaning that the observation has low value for the specific variable, to blue, meaning high values for the same variable. The position on the horizontal axis represents the contribution of the variable in increasing or decreasing the predicted probability of each observation. Values on the left column reports the average absolute change in predicted probability over all observations and the normalized values, in parenthesis.



(a) Defaulted clients.



(b) Non-defaulted clients.

Figure E13: SHAP average signed effect for Elastic-Net model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. Bars report the average effect of input variables on the predicted probabilities for all observations predicted as 1 and 0, respectively.

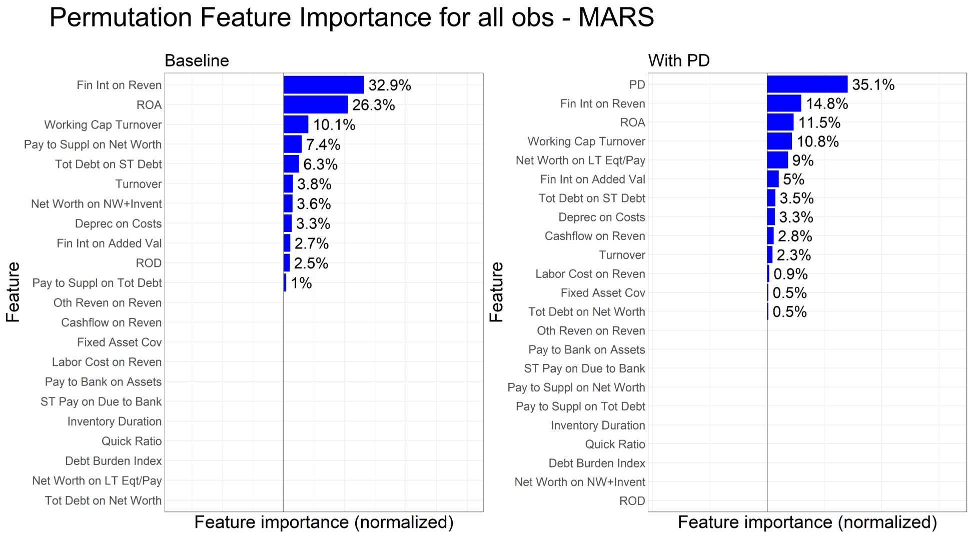
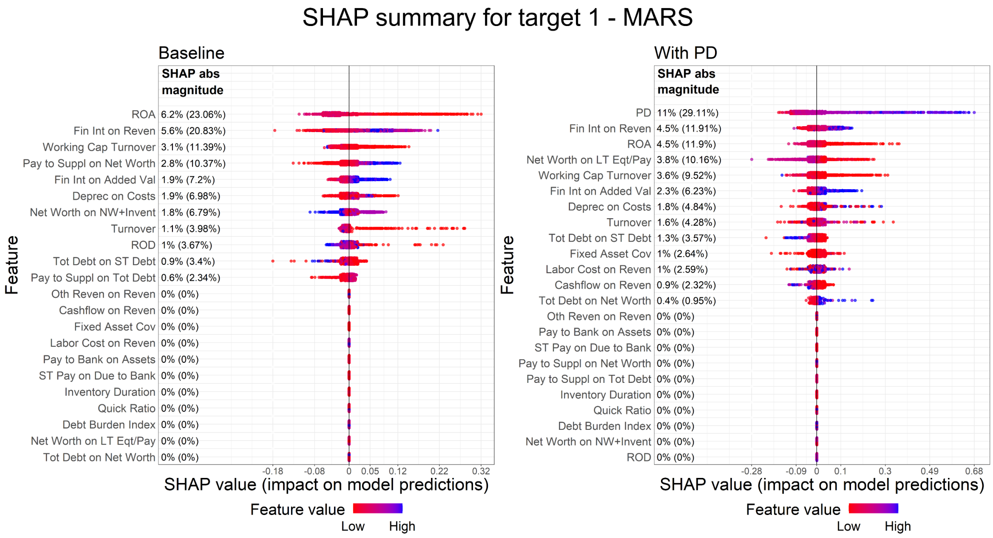
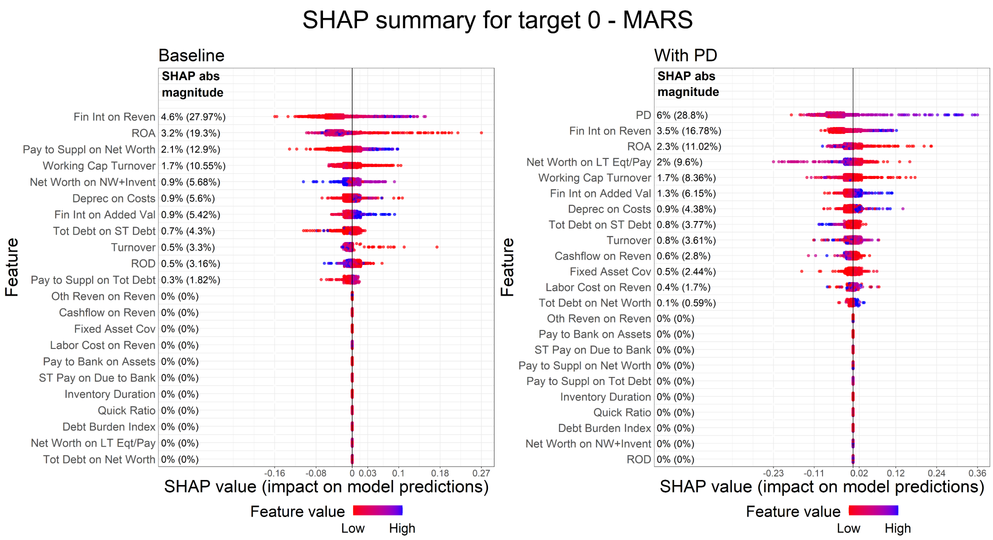


Figure E14: Permutation Feature Importance for MARS model, comparing variable importance of model calibrated with input variables and with the addition of PD. Normalized changes of F1-score are used to rank the variables.

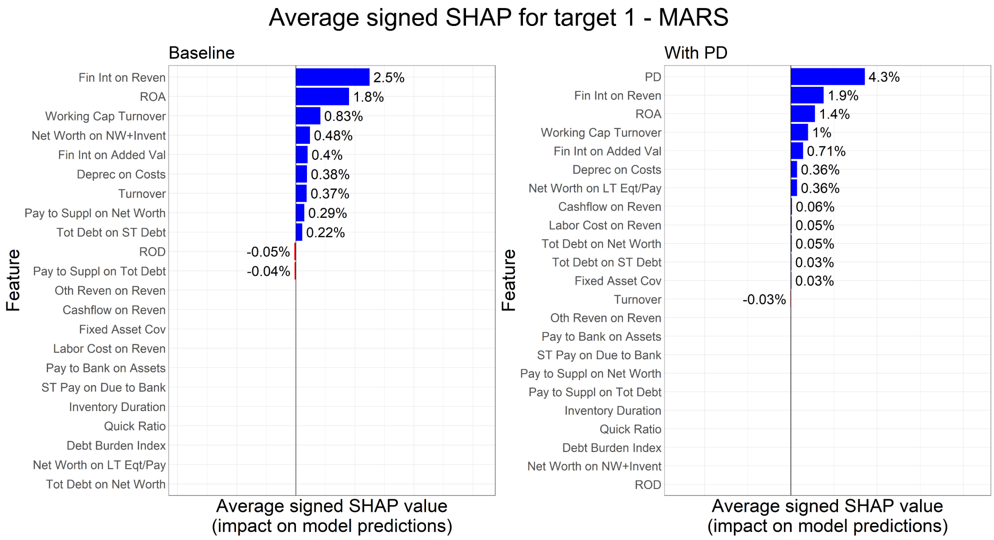


(a) Defaulted clients.

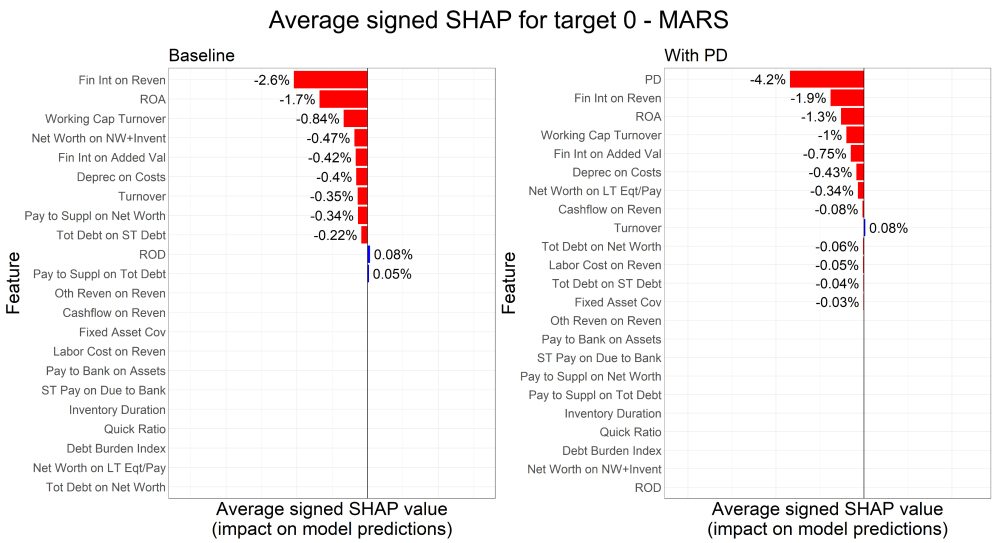


(b) Non-defaulted clients.

Figure E15: SHAP effects on predicted probability for MARS model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. The color of the points ranges from red, meaning that the observation has low value for the specific variable, to blue, meaning high values for the same variable. The position on the horizontal axis represents the contribution of the variable in increasing or decreasing the predicted probability of each observation. Values on the left column reports the average absolute change in predicted probability over all observations and the normalized values, in parenthesis.



(a) Defaulted clients.



(b) Non-defaulted clients.

Figure E16: SHAP average signed effect for MARS model and defaulted (top) and non-defaulted (bottom) observations only, comparing variable importance of model calibrated with input variables and with the addition of PD. Bars report the average effect of input variables on the predicted probabilities for all observations predicted as 1 and 0, respectively.