A data-driven approach to financial soundness around the world

Alessandro Bitetto
University of Pavia

Paola Cerchiello
University of Pavia

Charilaos Mertzanis
Abu Dhabi University

5th Winter School on Data Analytics
Higher School of Economics

November 20-22, 2020
Nizhny Novgorod, Russia

IMPORTANCE OF SOUND FINANCIAL SYSTEMS

- **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 saving for productive use,**
  - **providing information on existing and potential investment opportunities,**
  - **improving corporate governance; and**
  - **facilitating trading, diversification, and risk management.**
- **The 2008 financial crisis has underscored the importance of**
  - **financial system resiliency in providing finance to the economy during the business cycle; and**
  - **limiting the types of financial/real imbalances that develop during times of prosperity.**
- **Financial stability also interacts with monetary stability**
  - **The financial stability trilemma (Schoenmaker, 2009) states that (a) a stable financial system, (b) an integrated financial system and (c) national financial stability policy, are incompatible. Any two of the three objectives can be combined, but not all three.**
  - **This poses considerable problems for understanding the role & effects of financial stability.**

MEASURING FINANCIAL STABILITY 1

- **Measuring financial stability isn't easy (Gadanecz & Jayaram, 2009).**
  - **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.**
- **Assessing financial stability therefore requires the consideration of the diverse macroeconomic, structural and institutional aspects of the financial system (ECB, 2008).**
- **Why?**
  - **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.**
  - **Moreover, the maintaining of strong macroeconomic performance and effective monetary policy at the national level requires sound financial institutions.**
  - **Finally, central banks and governments are paying increasing attention to monitoring the health and efficiency of financial institutions and markets and to macroeconomic and institutional developments that pose potential risks to financial stability. Thus, CBs need proper metrics for performing effective monitoring.**

MEASURING FINANCIAL STABILITY 2

- **Need to consider both quantitative and qualitative approaches.**

- **There is a need for implementing a net risk approach that combines both the quantitative and qualitative aspects of financial vulnerabilities, based on proper information, which includes:**
  - **quantitative evaluation of all risks faced by financial institutions**
  - **qualitative adjustment for institutional factors.**

- **This approach helps to better assess the extent to which the risks are adequately managed through:**
  - **market discipline and internal governance in an institution, and**
  - **supervisory frameworks in the financial system as a whole.**

- **Synthesizing this info produces an overall risk assessment for individual institutions and an overall stability assessment for the financial system.**

- **The nature of government intervention in the economy, payment culture and insolvency regime, credit/deposit guarantees, the quality of supervision and regulation, moral hazard, corporate governance, and management quality all affect the overall incentive structure for maintaining financial stability**

MEASURING FINANCIAL STABILITY 2

Adopting a data-driven approach to measuring financial stability

- Several methods have been proposed to assess financial stability.
- The literature is driven by economic theories without specific data analytics and objective assumptions on the relationship among variables and the outcome.
- We use a fully data-driven approach to measure financial soundness of countries in the sense that we let the data "speak" by means of an unsupervised statistical learning technique - Principal Component Analysis (hereafter PCA).
- PCA makes neither a-priori assumptions on the relationship among the input variables nor a subjective decision on the important variables. PCA does not need to define a target variable to be modelled, avoiding a further level of subjectivity.
- The PCA maximizes the intrinsic explained variance of the input variables, identifying new coordinates (PCs) according to which the data vary the most. In this way, we can focus on which new coordinates can capture any change in the data.
- The only model assumption lays on the number of components built on the original variables based on the desired level of captured variability and explanatory ability.
- Moreover, by construction, the new coordinates must lie on a linear space and be mutually orthogonal (i.e. independent). Such independence ensures that each new PC is describing a specific and not known in advance latent phenomenon through the linear combination of the initial variables.

OUR CONTRIBUTION

- **Adopting a data-driven approach to measuring financial stability**

- **In particular our paper contributes to the literature in the following ways:**

- **Proposing a new Financial Soundness Index (henceforth FSIND) fully model-based, data-driven & validated according to advanced statistical techniques;**
- **Building up the FSIND index by using IMF's Financial Soundness Indicators generated to measure and monitor the strengths and weaknesses of the financial system in its member countries;**
- **Considering the wider and most comprehensive dataset filled with 140 countries either developed or developing ones (unlike single countries);**
- **Enriching the information set with institutional variables that take into account geographical and cultural dimensions per country;**
- **Conducting an extensive model validation and sensitivity analysis that makes the model robust and statistically sound;**
- **Providing the policymakers with an index in double format, both continuous and dichotomous, so to allow the choice of the most informative expression according to the needs and the policies to be put in place.**

SOME REVIEW OF THE LITERATURE 1

Approaches to measuring financial stability

- **The measurement and monitoring of the soundness of the financial institutions helps detect early on the potential buildup of systemic risk that may lead to a crisis.**
- **Several methods have been used to assess financial stability. Each has its advantages, disadvantages and limitations.**
- **In recent years, policy makers and researchers have focused on various statistical indicators that epitomize and describe the vulnerabilities of the financial system to appraise its stability. Among the commonly used quantitative methods for financial stability assessment are early warning systems, macro-stress testing, and financial stability indices (Alessi et al., 2011; Alessi & Detken, 2018; Lin et al., 2012).**
- **Further, the approaches to developing stability measures have changed over time as the focus has moved from the micro-prudential to the macro-prudential dimension of financial stability.**
- **Guérineaua and Léon (2019) argue that the financial stability issue in low-income countries has received less attention in recent years, assuming that they have been less touched by the global financial crisis compared to other economies.**
- **These authors however focus more narrow on the banking system (in)stability and on the credit information sharing power by analyzing 159 countries equally divided into developing and developed ones for a time window 2008 and 2014.**

SOME REVIEW OF THE LITERATURE 2

Comparative studies

- Comparing our paper to related ones, we notice that many **financial stress indices are produced for specific countries following their peculiarities and economic characteristics** (Hakkio & Keeton, 2009; Holmfeldt et al., 2009; Cardarelli et al., 2011; Slingenberg & Haan, 2011; Holló et al., 2012)
- Such efforts, although interesting and often accurate, **lack generalization power**, so they can hardly be applied worldwide by policymakers.
- Papers analyzing multi-countries data (Guérineaua and Léon, 2019; Cevik et al., 2013; Islami & Kurz-Kim, 2013) are more challenging and difficult to implement. Our paper relates to those but also differs in substantial ways.
- However, adopting a **cross-country approach has its drawbacks**:
  - several missing values,
  - non significant variables (albeit significant at country level)' and
  - difficulties in control variables assessment.
- Despite the difficulties, we argue that, in a globalized and interconnected economy, it is not useful to focus solely on country specific analysis for it provides policy makers with limited and shortsighted tools.
- Finally, there are no papers using data-driven methods to produce financial indices

DATA

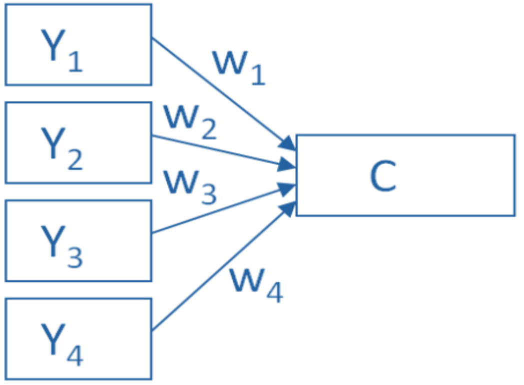
Data structure

- The IMF disseminates data on selected FSIs. As of today, 140 jurisdictions report FSIs on a regular basis. Our paper uses data from 140 countries covering 17 systemic indicators for the period 2006-2017, plus other variables. Our FSIs include:

Financial Soundness Indicators	Other institutional / geo variables
1. Assets to gross domestic product (GDP)	1. Latitude Static
2. Assets to total financial system assets	2. Longitude
3. Earnings to interest and principal expenses	3. Individualism vs collectivism
4. Interest margin to gross income	4. Indulgence vs restraint
5. Liquid assets to short term liabilities	5. Long term orientation vs short term normative orientation
6. Liquid assets to total assets (liquid asset ratio)	6. Masculinity vs femininity
7. Net open position in foreign exchange to capital	7. Power distance
8. Non-financial corporations sector	8. Uncertainty avoidance
9. Non-interest expenses to gross income	
10. Non-performing loans net of capital provisions	
11. Non-performing loans to total gross loans	
12. Other financial corporations	
13. Regulatory tier 1 capital to risk-weighted assets	
14. Return on assets	
15. Return on equity	
16. Sectoral distribution of loans to total loans	
17. Total debt to equity	

Background

- Our analysis extracts synthetic indicators that summarize at best the relationship among variables in a lower dimensional space. A widely-shared method used for dimensionality reduction is Principal Component Analysis (PCA).
- Briefly, PCA aims at creating one or more new component from a larger set of measured variables, where each component is a linear combination of the *Y* original variables (see Figure 1). The model is represented by the equation $C = w_1Y_1 + . . . + w_4Y_4$, where *C* is the new index, Y_i are the original variables and w_i are the weights of the linear combination.
- Our dataset has three dimensions: ***Variable, Country***, ***Time***, which exhibit interdependencies (over time), so we decided to apply the PCA dimensionality reduction technique so as to model country/variable interaction for each year.



METHODOLOGY 2

PCA approach used

- **We use 3 alternative PCA methods:**
- **PCA. It finds the directions on which, projecting the original data, the latter's total variance is maximized. Those directions are constrained to be mutually orthogonal and are called Principal Components (PCs) or loadings. Given a n *x* p data matrix X, where n is the number of obs and p is the number of vars, we find the k *x* p PC matrix C, with usually k < p, such that the projected data $W = XC^T$, also called scores, will have dimension n *x* k. The maximization problem can be seen as:**
- **Minimize C $\|X - XCC\|_F$, subject to $C^TC = I$**

where ||.||F is the Frobenius norm. We implement the model using R package *prcomp*.

- **Robust PCA. To produce a robust estimation of PCs (that deals with corrupted obs), we 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 data anomalies. The problem can be solved by:**
- **Minimize L,S $\|L\|_* + \lambda \|S\|1$, subject to $L + S = X$**
- **where $\|L\|\cdot$ is the nuclear norm and λ is a penalization term. Once fitted, L is used as a proxy for X but cleaned up by extreme values. We implement the model described by Candes et al. (2009).**
- **Robust Sparse PCA. As a further improvement, we produce a robust and sparse representation of the PCs by adding a sparsity constraint on matrix C. The problem can be then solved by:**
- **Minimize C;W $\|X-WC^T\| - S\|^2_F + \psi(C) + \varphi(W) + \lambda\|S\|1$, subject to $C^TC = I$**
- **And ψ & φ are regularizing functions (LASSO or Elastic Net) described by Erichson et al. (2018).**

METHODOLOGY 3

PCA approach – dealing with missing values

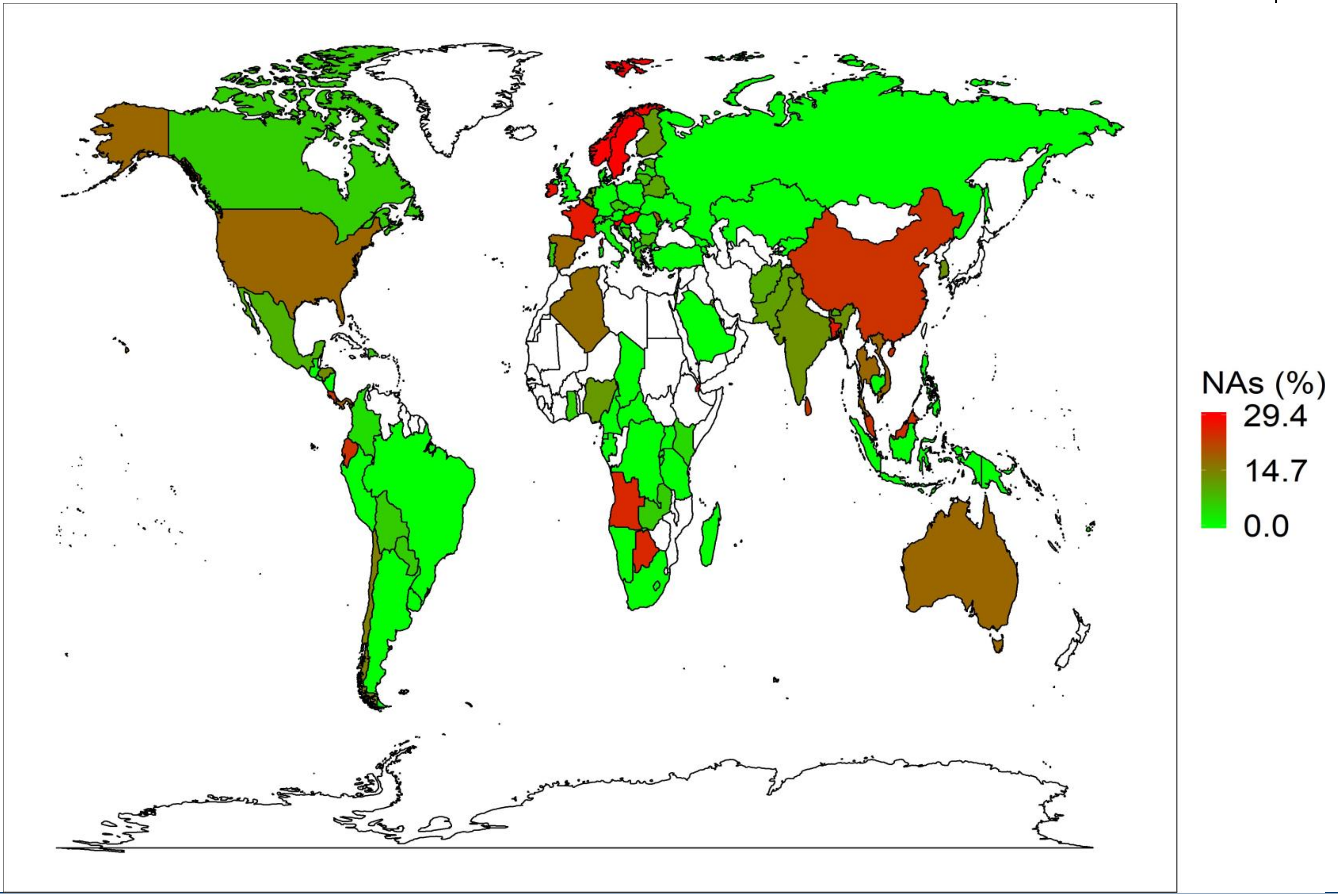
- **Some countries presented many missing values, as well as fewer than expected years. Thus, we decided to restrict our analysis to a subset of 119 countries from 2010 to 2017, choosing a missing value tolerance level not exceeding 30%.**
- **We report the list of countries and relative percentage of missing values in Appendix A. Since the presence of missing values affects the quality and reliability of results, we set a protocol of missing values treatment and imputation.**
- **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%. An overview of selected countries and their missing values percentage is shown in Figure 1 (next slide).**
- **In order to deal with missing values in a robust manner, we applied 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-uddin, 2016)**

METHODOLOGY 4

- **PCA approach – missing values**



METHODOLOGY 5

PCA approach – missing values

- **The BTF approach uses a tensorial decomposition of the 3-dimensional tensors that stacks all the annual slices together so that the imputation process involves information coming from both cross-sectional and temporal dimensions.**
- **To assess best performance, we applied the imputation algorithm in three settings.**
- **In the first (Original), we consider the whole dataset made of 119 countries, 25 vars for 8 years: a total of 23800 obs. It contains 8% of missing values, so we randomly remove some additional values representing 10%, 20% and 30% of the initial dataset.**
- **In the second (No missing), we drop all entries with missing values and apply the same incremental sampling procedure on the remaining subset.**
- **In the third (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, where M = total num of excluded values.**
- **Moreover, we check the sensitivity to the original percentage of missing values by comparing the MARE of the No missing and Some missing with the one on Original.**

METHODOLOGY 6

PCA approach – dealing with stationarity

- **After imputing missing data, we check for stationarity of each FSI-country pair over the time span. We perform standard Augmented Dickey-Fuller and Ljung-Box tests and since some non-stationarity is revealed, we integrate all time series with lag 1, in order not to sacrifice too many obs. The final dataset includes 25 vars for 119 countries and 7 lagged years.**

Method	Number of PC	Mean Explained Variance	Mean R2	Mean R2 on 99th	Mean R2 on 95th
PCA	1	15.7 ± 4%	15.7 ± 4%	16.5 ± 8.4%	19.4 ± 11.8%
	2	27.9 ± 6%	27.9 ± 6%	28.9 ± 8.9%	34.3 ± 10%
RobPCA	1	40.6 ± 3.5%	94 ± 0.9%	94.5 ± 0.9%	95.7 ± 0.9%
	2	64.9 ± 4.8%	96.4 ± 0.8%	96.7 ± 0.8%	97.3 ± 0.7%
RobSparPCA	1	15.8 ± 4%	10.7 ± 2.4%	12.8 ± 4.1%	14.8 ± 2.7%
	2	28 ± 6%	20.1 ± 3.3%	23.1 ± 3.2%	28.2 ± 2.7%

- **The Table presents results of PCA approach. We report the average variance explained by loadings for all years and the average R^2 both on the whole dataset and on subsets with values trimmed for the 95th & 99th percentiles to check for outliers.**
- **The RobPCA approach performs best**
- **We check only the first two principal components so that the index is interpretable.**

METHODOLOGY 7

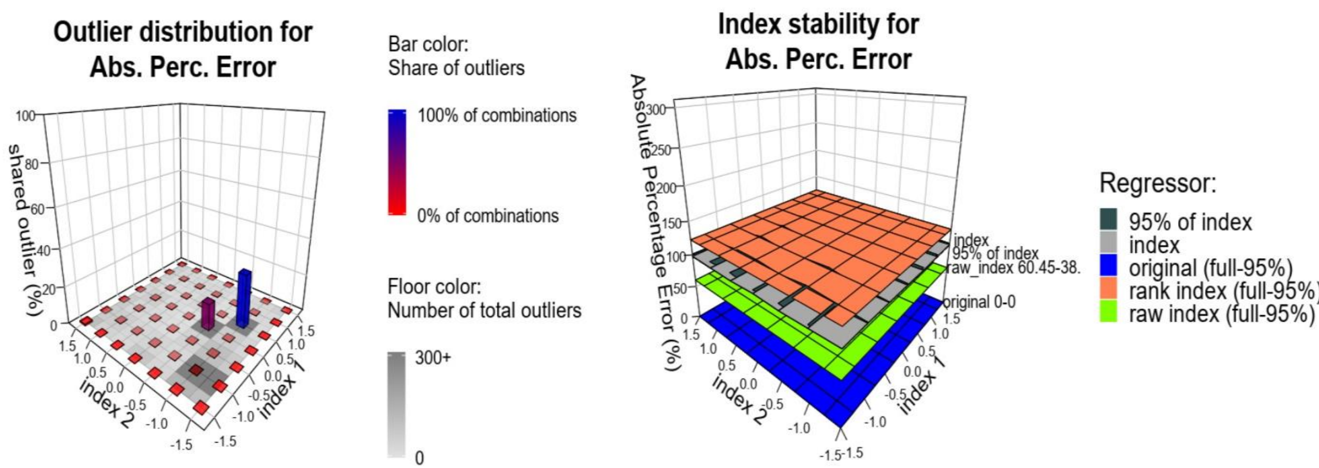
- **PCA approach – dealing with outliers**
- **We assess the outliers stability:**
  - **First, we evaluate outliers based on the extreme values of the Absolute Percentage Error (APE) on prediction according to the Generalized Extreme Studentized Deviate test (Rosner, 1983), and**
  - **Then, in order to check consistency, we count the total number of detected outliers for each threshold and their shared percentage over all thresholds.**
  - **Further, in order to assess outlier stability, we compare the performance of the models for each threshold according to APE.**
- **Finally, we also check predictive performance of the fitted models using as regressors the initial dataset used to build the FSIND (FSI + Hofstede + Geographical details) (original case), the continuous FSIND (raw index case) and the continuous FSIND discretized according to the selected thresholds, i.e. 2 categorical variables with 8 levels each (rank index case).**

PCA approach – validation

- **Index validation using the domestic credit to private sector (% of GDP):**

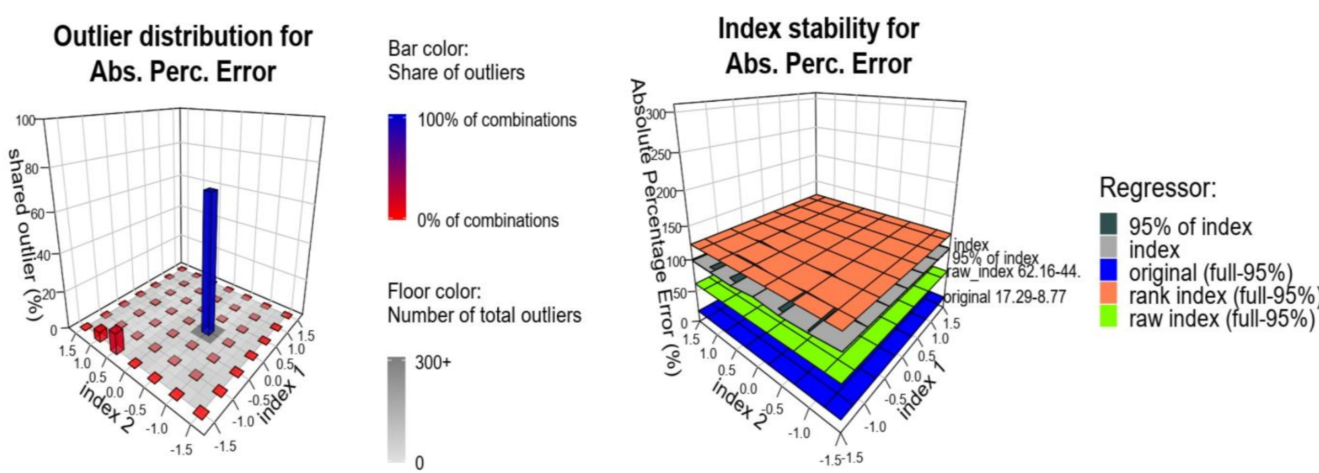
RobPCA

Model: GBM
Performance: RMSE
Regressor:
-index (avg) Train: 1±0 Test: 1±0
-original Train: 0 Test: 0.38
-rank_index Train: 0.95 Test: 1.06
-raw_index Train: 0.28 Test: 1.3



RobPCA

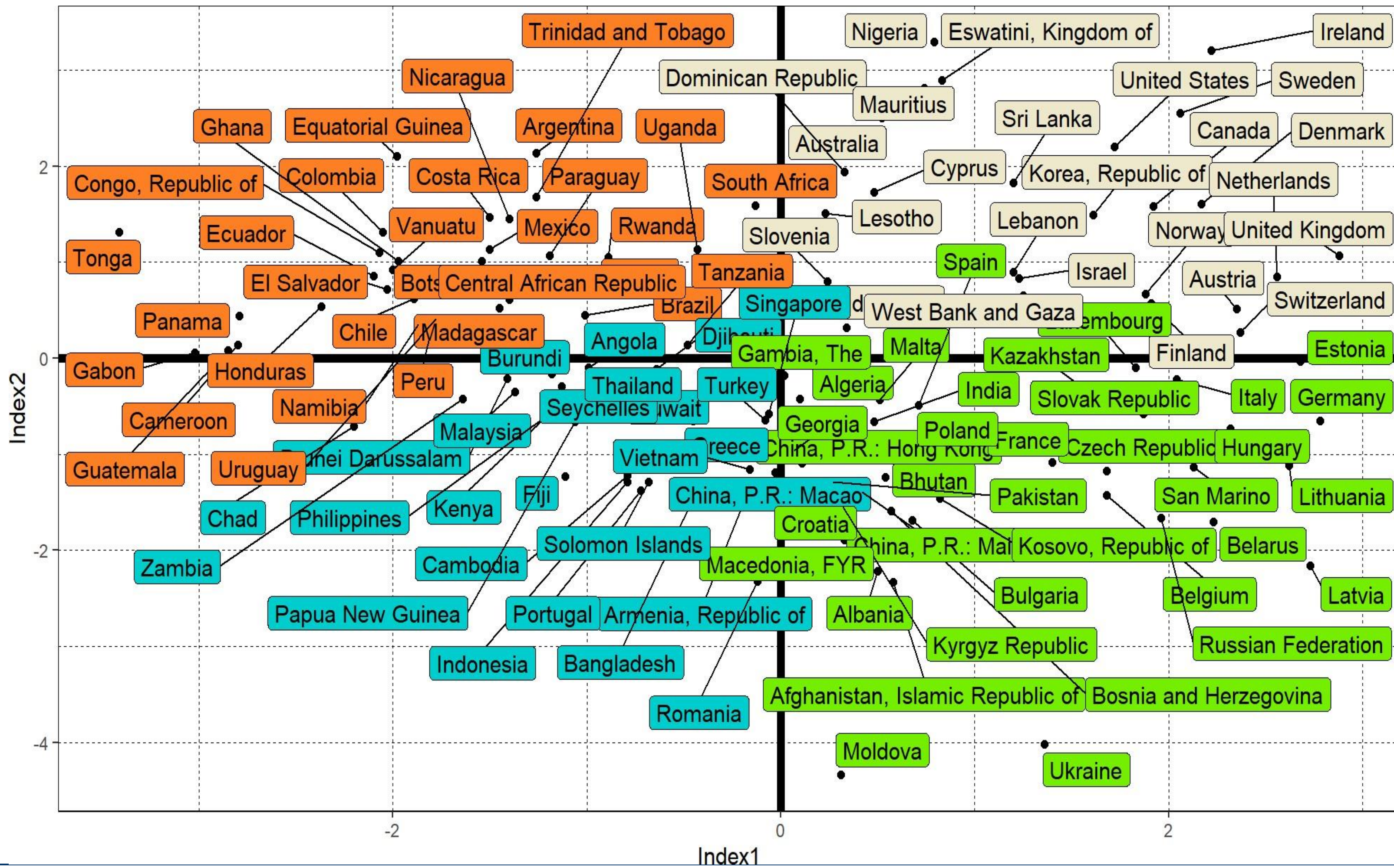
Model: Random Forest
Performance: RMSE
Regressor:
-index (avg) Train: 1±0 Test: 1±0
-original Train: 0.09 Test: 0.24
-rank_index Train: 0.95 Test: 1.05
-raw_index Train: 0.43 Test: 1.1



RESULTS 1

THE FINANCIAL SOUNDNESS INDEX - FSIND

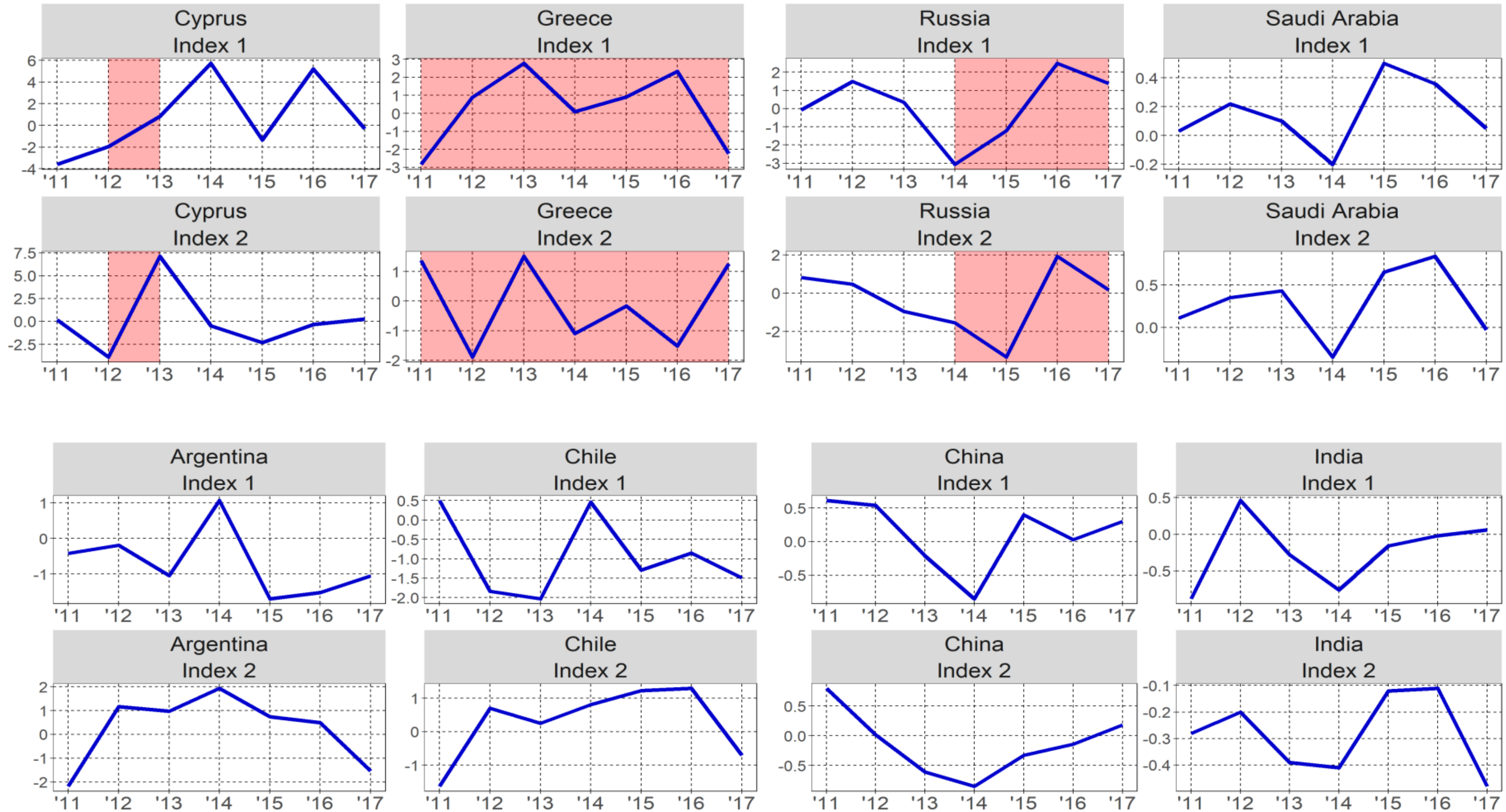
- **Country classification of predicted PCA outcomes for year 2014**



RESULTS 2

THE FINANCIAL SOUNDNESS INDEX - FSIND

- **Sample comparative country performance of FSIND (binary vs continuous)**



CONCLUSIONS

THE FINANCIAL SOUNDNESS INDEX - FSIND

- We assess the financial soundness for most of worldwide countries through a fully data-driven approach. We use IMF's financial indicators, in particular the core set filled up of 17 indexes for the period 2006-2017, plus Hofstede's cultural dimensions and geographical information (i.e. latitude and longitude).
- As is typical, there are missing values or not aligned information. Thus, before applying proper statistical models, we conducted a complete process of data quality assessment and data imputation.
- After data imputation, we proceed with a dimensions reduction model (robust PCA) and produced both a continuous and a binary index FSID (stable vs unstable), for each country.
- As for the initial data quality, we paid attention on the index validation by selecting suitable thresholds on the basis of the best performance obtained by several competing regression models.
- The index plays the role of the covariate and the target variable is a macro economic quantity, like GDP or Non performing loans. Results show that our index can summarize and capture well the dynamics of the economy.

Thank you