A data-driven approach to measuring financial soundness throughout the world

Alessandro Bitetto[1], Paola Cerchiello[1], Charilaos Mertzanis[2], Ernst Wit[3]

[1] University of Pavia, Italy,
[2] Abu Dhabi University, United Arab Emirates,
[3]Università della Svizzera Italiana, Switzerland

DATA ANALYTICS 2020
Nice, October 25-29th, 2020

Motivation

Motivation

- Macroeconomic variables are often used to assess financial stability for countries by the means of synthetic indices based on expert-judgement assumptions of Financial Institutions, e.g. weighted average
- However, for their subjective nature, all indices can be questionable and can lead to endless debate on which one should be used as a robust financial indicator
- The aim of this work is to create a objective, thus data-driven, alternative index

Data

Dataset

- Annual Financial Soundness Indicators (FSI) provided by International Monetary Fund (IMF) ranging from 2007 to 2017 and for most of worldwide countries, including both strong and developing economies for a total of 119 countries and 17 FSI.
- 6 Hofstede Indicators (Individualism, Masculinity, etc.) for each country, fixed for all years
- 2 Geographical Indicator (Latitude and Longitude)
- Final dataset has $n = 119$ countries with $p = 25$ variables for $T = 7$ years

Methodology

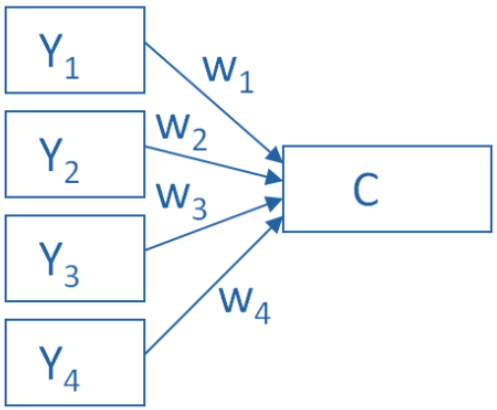
Methodology used

- As the data have 3 dimensions, *Country*, *Variables* and *Time*, two complementary techniques have been used:
  - Principal Component Analysis (PCA) to model country/variables interaction, for each year
  - Factor Analysis (FA) to model country/time interaction, for all variables

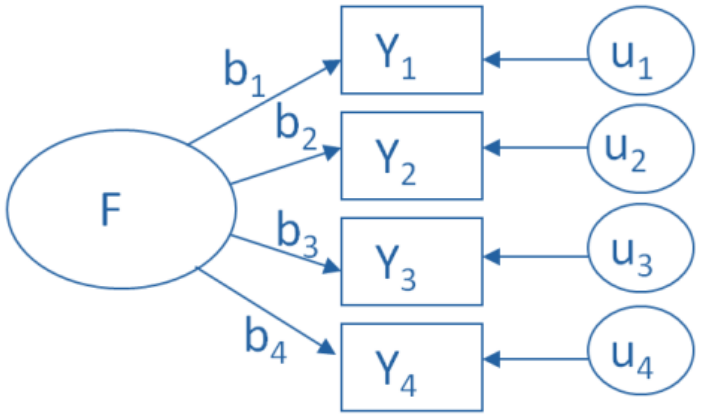
Methodology used

(a) PCA aims to create one or more index variables from a larger set of measured variables, where each index is a linear combination of the Y original variables

(b) FA models the measurement of latent variables, seen through the relationships they cause in a set of Y variables.



(a) *PCA:* the model is an equation $C = w_1 Y_1 + \ldots + w_4 Y_4$



(b) *FA:* the model is a set of equations $Y_i = b_i F_1 + u_i, i = 1, \ldots, 4$

PCA - details

- The following PCA techniques have been tested for each year:
  - **PCA**
  - **Robust PCA**: decompose M by solving

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

    where $\|L\|_*$ is the nuclear norm
  - **Robust Sparse PCA**: minimize

$$f(A, B) = \frac{1}{2}\|X - XBA^T - S\|_F^2 + \psi(B) + \gamma\|S\|_1$$

    where B is the sparse loading matrix, A is orthonormal, ψ is a regulizer (i.e. LASSO or Elastic Net) and S captures outliers
- Robust PCA performed best with an average (over years) Explained Variance of $46 \pm 3\%$ for the first 2 PC

FA - details

- Due to small depth of each FSI time series the following approach has been used:
  - Fit a Dynamic Factor Model

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

    for each of n country, obtaining *factor matrices* F^i, *factor interactions* A^i and *factor loadings* C^i, $i = 1, \ldots, n$
  - Fit a Vector Auto Regressive (VAR) model in order to get $\hat{A}$ lag-1 matrix that incorporates cross-countries interaction of A^i
  - Use Kalman Filter to get smoothed factors $\widehat{F^i}$ using $\hat{A}$ and $\hat{C} = diag(C^i)$ in order to get latent factors that incorporates cross-countries interactions
- Optimal number of factors has been set to 2 with Y-reconstruction error validation

Index evaluation

The final index, hereinafter referred as Financial Soundness Index (FSIND), will be constructed by:

- $(n \cdot T) \times 2$ scores matrix in PCA approach
- $(n \cdot T) \times 2$ factor matrix in DFM approach

Index Validation

Binary Index

- Both methodologies produce continuous value for the 2 components of the index
- In order to get a binary index, the following procedure has been followed:
  - set a threshold and get the binary index, i.e. 0 or 1
  - perform a regression task where target is a economic variable (such as *GDP* or *Non Performing Loans*) and regressors are the 2 binary using different partitioning algorithm, such as *Random Forest* and *Gradient Boosting Machine*
  - evaluate prediction accuracy and outliers for different threshold
- Robust threshold has been set to 0 for both indices

Results and future work

Index comparison

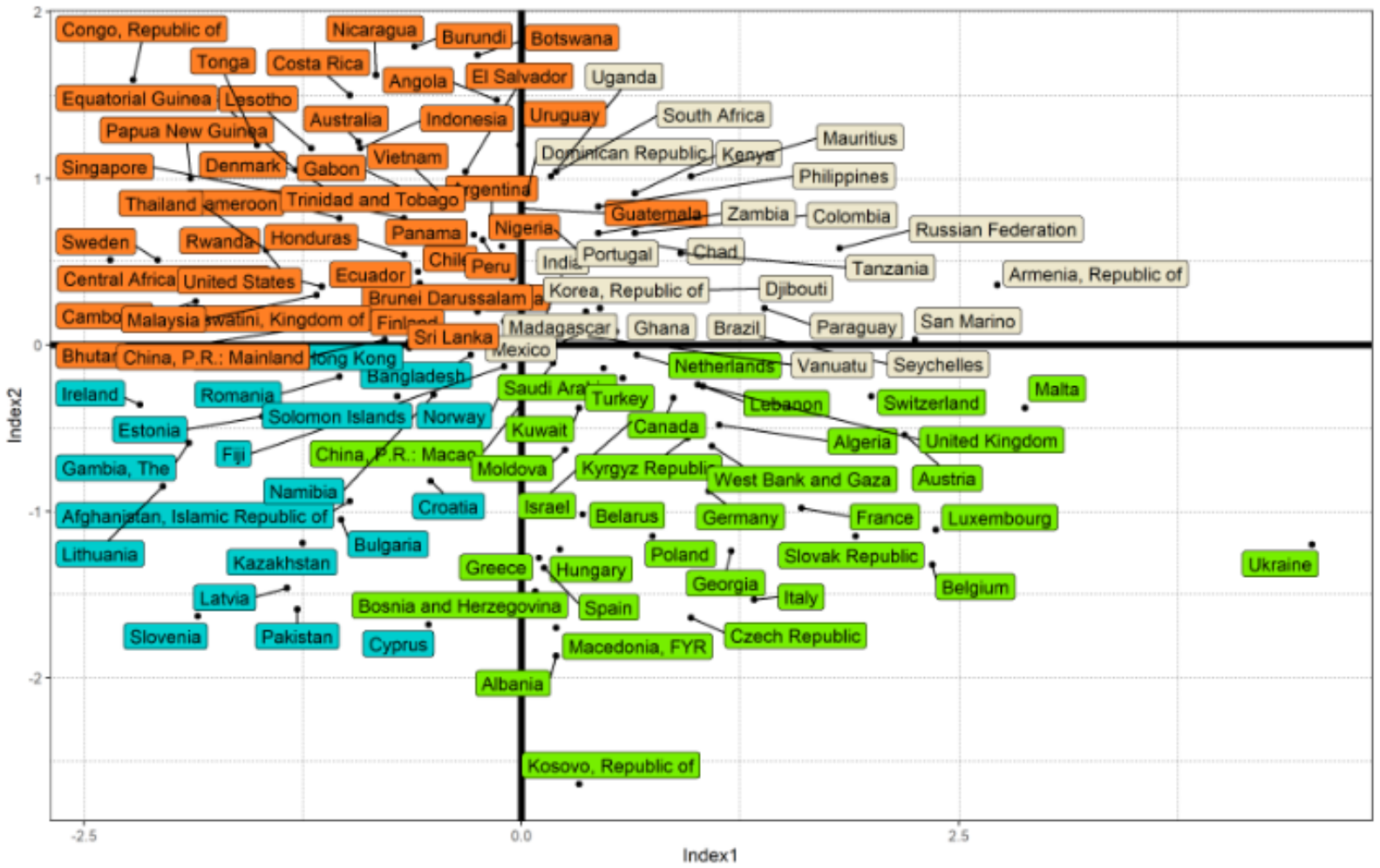


Figure: Robust PCA index for 2014

Index comparison

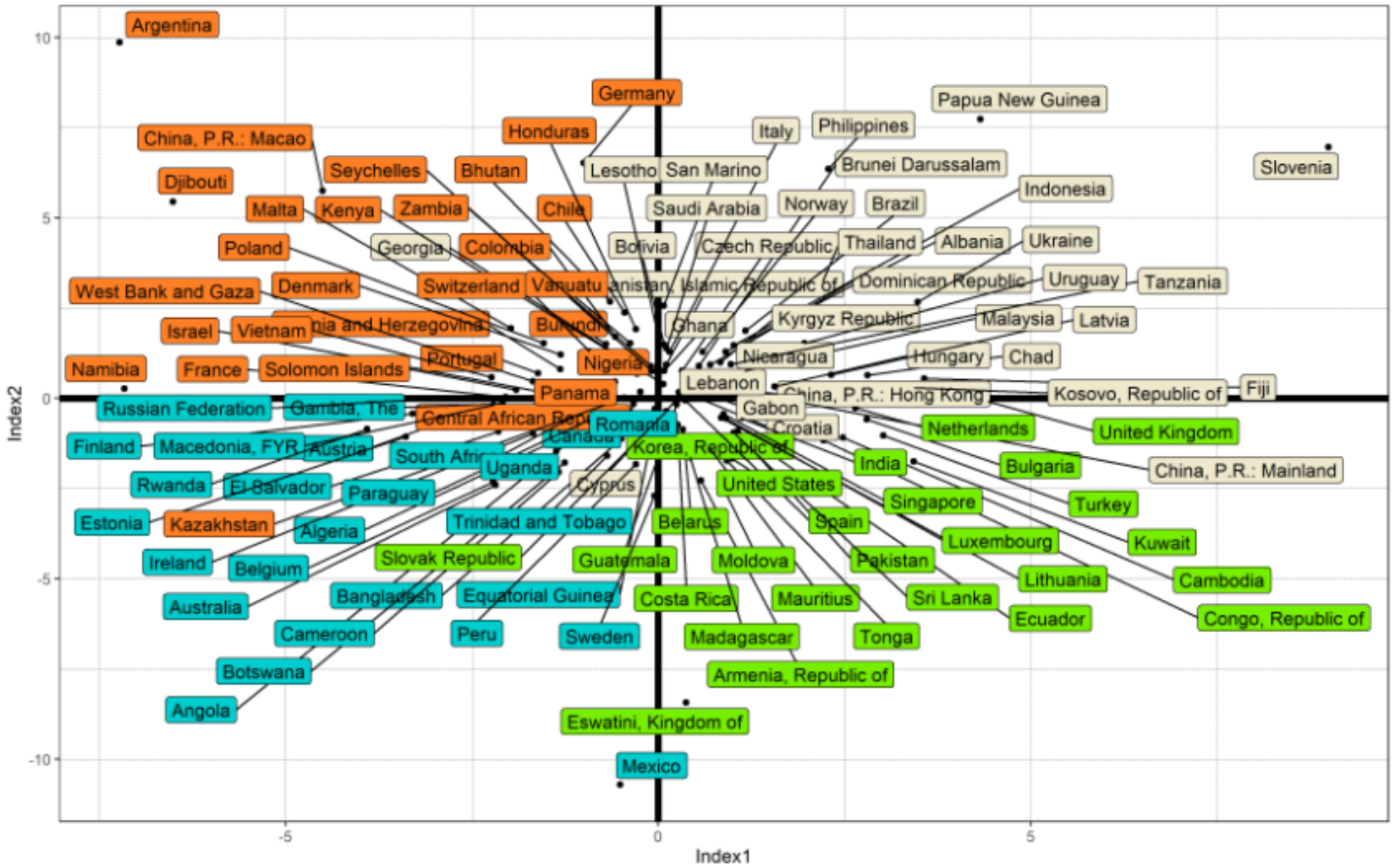


Figure: DFM index for 2014

Index comparison

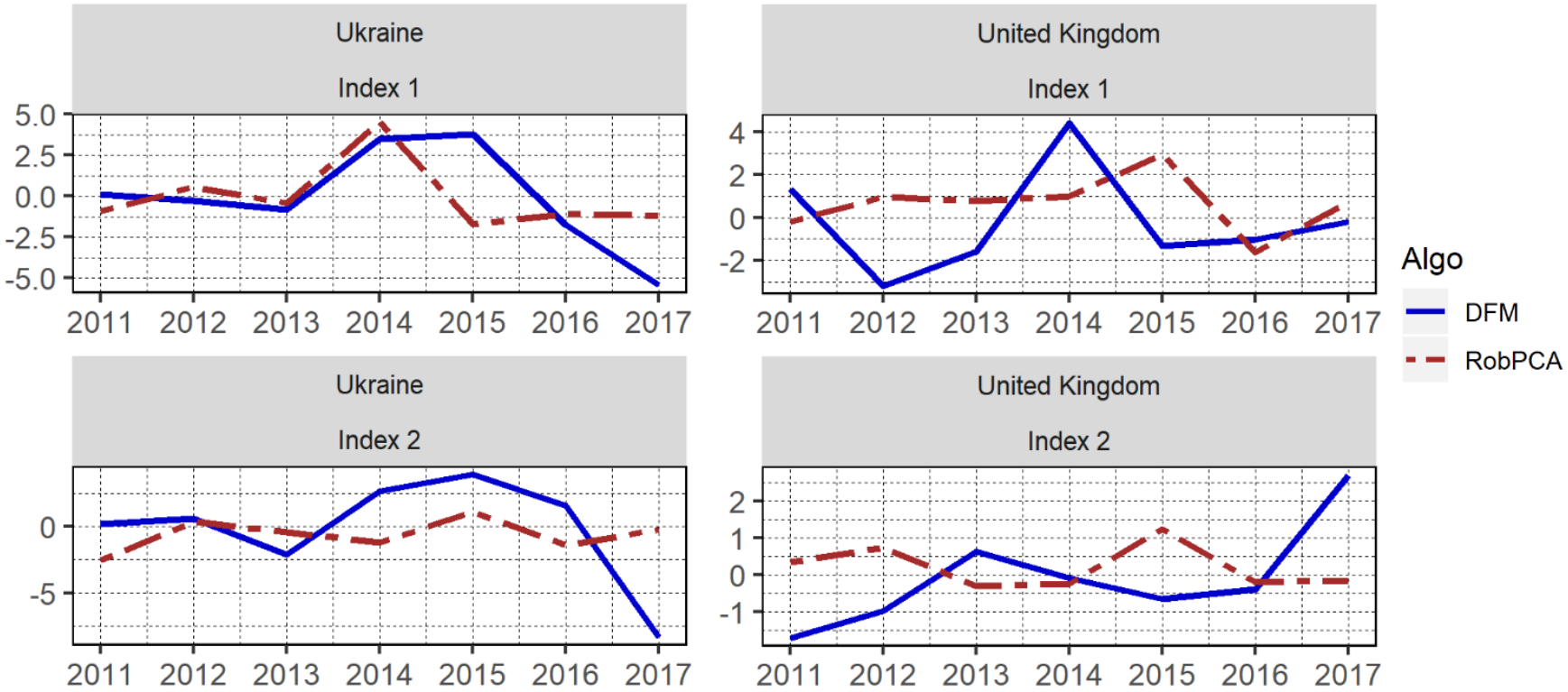


Figure: Index evolution over years for Ukraine (both methods agree) and United Kingdom (methods disagree)

Application

FSIND predictive power has been tested on a regression task:

- Dataset consisted of 53 annual macro-economic variables ranging from 2007 to 2017 for $\sim 65,000$ firms from worldwide countries, matched with the perimeter used to build our index.
- Target variable is the ordinal indicator of ease in accessing to financial funding, from 1, easiest to 4, hardest
- An ordered probit model is used as baseline model to predict the target variable, given a subset (selected by significance level of coefficients) of the initial 53 variables
- FSIND are used as additional predictors: 2-dimensional PCA version and 2-dimensional DFM version are used on both continuous and binary form

Application results

ACCESS	Baseline	OLS	Probit RE	IV 2SLS	IV GMM
DFM_CONT1	−0.136**	−0.0487***	−0.166**	−0.230***	−0.129**
	(0.0552)	(0.0155)	(0.0722)	(0.0847)	(0.0541)
LISTED	−0.0971***	−0.0247***	−0.117***	-0.0157	-0.00844
	(0.0293)	(0.00766)	(0.0348)	(0.00978)	(0.0117)
AGE	−0.265***	−0.0742***	−0.234***	−0.0787***	−0.0910***
	(0.0415)	(0.0112)	(0.0503)	(0.0131)	(0.0154)
SIZE	−0.137***	−0.0401***	−0.0534***	−0.0364***	−0.0449***
	(0.0131)	(0.00351)	(0.0160)	(0.00425)	(0.00503)
SECTOR	−0.119***	−0.0325***	−0.114***	−0.0279***	−0.0281***
	(0.00959)	(0.00259)	(0.0118)	(0.00327)	(0.00391)
SUBSID	−0.0456***	−0.0131***	−0.0526***	-0.00302	−0.0124***
	(0.0120)	(0.00320)	(0.0147)	(0.00396)	(0.00464)
LEGAL	0.0489***	0.0124*	0.0395	0.0241***	0.0336***
	(0.0257)	(0.00689)	(0.0315)	(0.00844)	(0.0101)
LOCATION	0.176***	0.0454***	0.183***	0.0570***	0.0671***
	(0.0206)	(0.00560)	(0.0252)	(0.00683)	(0.00844)
EXPORT	−0.111***	−0.0285***	−0.126***	−0.0195***	−0.0229***
	(0.0217)	(0.00588)	(0.0255)	(0.00648)	(0.00837)
AUDIT	-0.0135	-0.00407	−0.0378***	0.00306	0.00477
	(0.0101)	(0.00272)	(0.0124)	(0.00337)	(0.00393)
MANAGEXP	−0.370***	−0.0940***	−0.649***	−0.0670***	−0.0880***
	(0.0699)	(0.0186)	(0.0832)	(0.0226)	(0.0263)
L_MANAGEXP	0.262***	0.0734***	0.283***	0.0245	0.0499**
	(0.0576)	(0.0155)	(0.0700)	(0.0189)	(0.0218)
Year Effect	Yes	Yes	No	Yes	Yes
Country Effect	Yes	Yes	Yes	Yes	Yes
Chi-square	8131.98	9958.76	467.67	21201.72	59707.78
Sargan j				140.35	
(p-value)				(0.0033)	
Hansen					0.089191
(p-value)					(0.7652)
N.	64717	64717	64717	43243	31942

Future work

- FSIND must be compared with other economical indices and find meaningful economical explanation
- FSIND robustness should be futher tested with alternative approaches, such as tensor decomposition
- Additional methodologies involving *Network Theory* should be tested for comparison:
  - *Factorial Graphical Model* for a time-independent estimation
  - *Time Series Chain Graphical Model* for time-dependent estimation

  and centrality measures could be used as FSIND weights