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

Alessandro Bitetto\textsuperscript{1}, Paola Cerchiello\textsuperscript{1}, Charilaos Mertzanis\textsuperscript{2}, Ernst Wit\textsuperscript{3}

\textsuperscript{1} University of Pavia, Italy,
\textsuperscript{2} Abu Dhabi University, United Arab Emirates,
\textsuperscript{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.
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
(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
    
    \[
    \text{minimize } \|L\|_* + \lambda \|S\|_1 \\
    \text{subject to } L + S = M
    \]
    
    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, $\psi$ is a regularizer (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
Due to small depth of each FSI time series the following approach has been used:

- Fit a Dynamic Factor Model

\[
\begin{align*}
F_t^i &= A^i F_{t-1}^i + \mathcal{N}(0, Q^i) \\
y_t^i &= C^i F_t^i + \mathcal{N}(0, R^i)
\end{align*}
\]

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 \( \hat{F}^i \) using \( \hat{A} \) and \( \hat{C} = \text{diag}(C^i) \) 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
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
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 an 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
Figure: Robust PCA index for 2014
Figure: DFM index for 2014
Index comparison

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

<table>
<thead>
<tr>
<th>ACCESS</th>
<th>Baseline</th>
<th>OLS</th>
<th>Probit RE</th>
<th>IV 2SLS</th>
<th>IV GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFM_CONT1</td>
<td>-0.136**</td>
<td>-0.0487***</td>
<td>-0.166**</td>
<td>-0.230***</td>
<td>-0.129**</td>
</tr>
<tr>
<td></td>
<td>(0.0552)</td>
<td>(0.0155)</td>
<td>(0.0722)</td>
<td>(0.0847)</td>
<td>(0.0541)</td>
</tr>
<tr>
<td>LISTED</td>
<td>-0.0971***</td>
<td>-0.0247***</td>
<td>-0.117***</td>
<td>-0.0157</td>
<td>-0.00844</td>
</tr>
<tr>
<td></td>
<td>(0.0293)</td>
<td>(0.00766)</td>
<td>(0.0348)</td>
<td>(0.00978)</td>
<td>(0.0117)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.265***</td>
<td>-0.0742***</td>
<td>-0.234***</td>
<td>-0.0787***</td>
<td>-0.0910***</td>
</tr>
<tr>
<td></td>
<td>(0.0415)</td>
<td>(0.0112)</td>
<td>(0.0503)</td>
<td>(0.0131)</td>
<td>(0.0154)</td>
</tr>
<tr>
<td>SIZE</td>
<td>-0.137***</td>
<td>-0.0401***</td>
<td>-0.0534***</td>
<td>-0.0364***</td>
<td>-0.0449***</td>
</tr>
<tr>
<td></td>
<td>(0.0131)</td>
<td>(0.00351)</td>
<td>(0.0160)</td>
<td>(0.00425)</td>
<td>(0.00503)</td>
</tr>
<tr>
<td>SECTOR</td>
<td>-0.119***</td>
<td>-0.0325***</td>
<td>-0.114***</td>
<td>-0.0279***</td>
<td>-0.0281***</td>
</tr>
<tr>
<td></td>
<td>(0.00959)</td>
<td>(0.00259)</td>
<td>(0.0118)</td>
<td>(0.00327)</td>
<td>(0.00391)</td>
</tr>
<tr>
<td>SUBSID</td>
<td>-0.0456***</td>
<td>-0.0131***</td>
<td>-0.0526***</td>
<td>-0.00302</td>
<td>-0.0124***</td>
</tr>
<tr>
<td></td>
<td>(0.0120)</td>
<td>(0.00320)</td>
<td>(0.0147)</td>
<td>(0.00396)</td>
<td>(0.00464)</td>
</tr>
<tr>
<td>LEGAL</td>
<td>0.0489***</td>
<td>0.0124*</td>
<td>0.0395</td>
<td>0.0241***</td>
<td>0.0336***</td>
</tr>
<tr>
<td></td>
<td>(0.0257)</td>
<td>(0.00689)</td>
<td>(0.0315)</td>
<td>(0.00844)</td>
<td>(0.0101)</td>
</tr>
<tr>
<td>LOCATION</td>
<td>0.176***</td>
<td>0.0454***</td>
<td>0.183***</td>
<td>0.0570***</td>
<td>0.0671***</td>
</tr>
<tr>
<td></td>
<td>(0.0206)</td>
<td>(0.00560)</td>
<td>(0.0252)</td>
<td>(0.00683)</td>
<td>(0.00844)</td>
</tr>
<tr>
<td>EXPORT</td>
<td>-0.111***</td>
<td>-0.0285***</td>
<td>-0.126***</td>
<td>-0.0195***</td>
<td>-0.0229***</td>
</tr>
<tr>
<td></td>
<td>(0.0217)</td>
<td>(0.00588)</td>
<td>(0.0255)</td>
<td>(0.00648)</td>
<td>(0.00837)</td>
</tr>
<tr>
<td>AUDIT</td>
<td>-0.0135</td>
<td>-0.00407</td>
<td>-0.0378***</td>
<td>0.00306</td>
<td>0.00477</td>
</tr>
<tr>
<td></td>
<td>(0.0101)</td>
<td>(0.00272)</td>
<td>(0.0124)</td>
<td>(0.00337)</td>
<td>(0.00393)</td>
</tr>
<tr>
<td>MANAGEXP</td>
<td>-0.370***</td>
<td>-0.0940***</td>
<td>-0.649***</td>
<td>-0.0670***</td>
<td>-0.0880***</td>
</tr>
<tr>
<td></td>
<td>(0.0699)</td>
<td>(0.0186)</td>
<td>(0.0832)</td>
<td>(0.0226)</td>
<td>(0.0263)</td>
</tr>
<tr>
<td>p_Managexp</td>
<td>0.262***</td>
<td>0.0734***</td>
<td>0.283***</td>
<td>0.0245</td>
<td>0.0499**</td>
</tr>
<tr>
<td></td>
<td>(0.0576)</td>
<td>(0.0155)</td>
<td>(0.0700)</td>
<td>(0.0189)</td>
<td>(0.0218)</td>
</tr>
</tbody>
</table>

Year Effect: Yes
Country Effect: Yes
Chi-square: 8131.98
Sargan j (p-value): 140.35
Hansen (p-value): 0.089191
N: 64717

October 25–29th, 2020
Future work

- FSIND must be compared with other economical indices and find meaningful economical explanation.
- FSIND robustness should be further 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.