ISSN: 2281-1346

UNIVERSITÀ DI PAVIA

Department of Economics and Management

DEM Working Paper Series

Dealing with dimension reduction in financial panel data

Alessandro Bitetto
(University of Pavia)

Paola Cerchiello
(University of Pavia)

Charilaos Mertzanis
(Abu Dhabi University)

 # 207 (06-22)

Via San Felice, 5
I-27100 Pavia

https://economiaemanagement.dip.unipv.it/it

Dealing with dimension reduction in financial panel data

Abstract

In this paper, we present a fully data-driven statistical approach to building a synthetic index based on intrinsic information of the considered ecosystem, namely the financial one. Among the several methods made available in the literature, we propose the employment of a Dynamic Factor Model approach which allows us to fully and correctly compare observations at hand in space and time. We contribute to the research field by offering a statistically sound methodology which goes beyond state of the art techniques on dimension reduction, mainly based on Principal Component Analysis. We adopt a country by country fitting strategy to elicit the inner country specific characteristics and then we combine results together by means of a Vector Autoregressive and Kalman filter approach. To this aim, we analyze a set of 17 Financial Soundness Indicators provided by the International Monetary Fund ranging from 2006 to 2017 for 140 countries that span the globe, including both strong and developing economies.

Keywords: Financial stability, Financing constraints, Data-driven, Dynamic Factor Model, State-space model, dimension reduction

1 Introduction

In many diversified application fields, there is a great demand for summary indicators able to bring useful and comprehensive information on the phenomenon under analysis. A common way to summarise information from a large enough set of variables is to create synthetic indexes based on assumptions taken by domain experts, typically resulting in the usage of synthetic measures like for example the weighted average. However, these measures are subjective by nature and in this sense they can be questionable, leading to endless debate on which one should be used as a robust indicator.

In the literature, we can list many papers which have proposed methods for calculating summary indexes. They can be mainly grouped into two classes: econometric approaches and statistical learning methods. The former comprises, among the others, the attempts of [1–4]. Those papers typically employ Vector Autoregressive or GARCH models to naturally elicit the temporal evolution of the considered variables. On the other hand, statistical learning methods mainly based on dimension reduction techniques like Principal Component Analysis or Factorial Analysis, have received similar attention [5–7].

Narrowing the scope of this paper onto the financial field, many attempts have been proposed in the scientific community aimed at the production of reliable indicators useful as predictors for further analysis too. Most research efforts at creating specific measures of financial stability, while applying various methods and models, have mainly focused on specific countries accounting for specific peculiarities and economic characteristics. For instance, [8] developed a daily financial stress index for the Canadian financial system and proposed alternative approaches to aggregating individual stress indicators into a composite stress index. Their index comprises eleven financial market variables, which they aggregate using weights determined by the relative size of the market to which each indicator pertains compared to a broad measure of total credit in the economy. [9] proposed the Kansas City Financial Stress Index (KCFSI) for the US economy. This index uses eleven financial market variables for 1990-2007, each of which captures one or more key features of financial stress. Applying the same methodology, [10] aggregated 18 weekly financial market indicators into the St. Louis Fed's Financial Stress Index (STLFSI). Likeswise, [11] produced the Cleveland Financial Stress Index (CFSI) by integrating 11 daily financial market indicators form the debt, equity, foreign exchange and banking markets. They normalized the raw indicators by transforming the series values into the corresponding CDF values. The transformed indicators are then aggregated into the composite indicator by applying time-varying credit weights, which are proportional to the quarterly financing flows through the four markets. [12] proposed a financial stress index for the Colombian economy, using measures of financial institutions' profitability, liquidity and probability of default. In particular, they used capital, liquidity, credit risk and return ratios monthly data (nine variables in total) from January 1995 to November 2008 for a heterogeneous bunch of financial institutions, including commercial banks, mortgage banks, commercial financial companies and financial cooperatives. The authors generated and combined information from different indicators through different quantitative methods, such as a variance-equal weight, principal component and count data methods. [13] proposed the composite indicator of systemic stress (CISS). They applied portfolio theory to compose five market-specific sub-indices based on fifteen individual financial stress measures. The CISS index takes into account the time-varying cross-correlations between the sub-indices as well. Applied to the Euro area, the CISS follows a systemic risk perspective, which assigns more weight on situations in which financial stress prevails simultaneously in several market segments. The authors establish critical CISS levels beyond which financial stress negatively affects real economic activity. [14] proposed a financial stability index for the Malaysian economy based on a dynamic factor model that uses fifteen diverse financial measures, ranging from non-performing loans to crude oil price and private capital funding, etc. The authors tested the predictive power of their index against the Malaysian business cycle and used it to examine the effect of credit expansion on the stability of the Malaysian financial system during April 1997 to December 2011. [15] use sixteen IMF financial soundness indicators during the period 2003-2013 to study financial stability in Israel. The approximate financial stability by using the economic resilience (ER) index produced by the International Institute for Management Development (IMD). The authors applied three different methods, namely principal component analysis, regression and hybrid models, and concluded that the PCA method were the most effective ones. [16] construct a banking system stability index (BSSI) for Nigeria using a weighted combination of the banking soundness index, the banking vulnerability index and the economic climate index during 2007:Q1 trough 2012:Q2. The BSSI index performed well in predicting the domestic financial crisis and the therefore the authors proposed it as an early warning tool. [17] focuses on nine Central Eastern and Southeastern European countries and develops a financial instability index (FII) using monetary, financial and foreign exchange data during 1996 to 2012. In order to construct the index, they cluster variables into five groups and subsequently use the quantile distribution to identify periods of financial stability versus financial instability. Finally, they use the resulting weighted average to design their index and fit a panel GMM regression to identify the most influential variables on the index evolution. [18] focuses on the role of monetary policy on financial stability. The authors analyzed the impact of the European Central Bank's decisions by measuring the difference between the optimal and real interest rates during 1999-2011 on a quarterly basis. They applied a stochastic reduced-form model that uses information on inflation and the GDP growth rate to build a financial stability index. They applied their index to predict the 2007-8 crisis and suggested an increase of the interest rate during 2010-11 as balancing mechanism to contain inflation. [19] focuses on the Brazilian financial system and uses quarterly macroeconomic indicators during 1995:Q1 to 2011:Q4 to build the broad financial stability indicator (BFSI) and the specific financial stability indicator (SFSI), both of which are shown to predict three Brazilian financial crisis episodes. They applied a principal factor method based on unobserved factors to construct the BFSI and an OLS regression of three main financial market indicators to construct the SFSI. They also applied a business cycle decomposition method that uses the co-movement of financial and real indicators to assess the driving role of financial vs. real factors, respectively. [20] construct a financial conditions index (FCI) for the USA using a hundred financial indicators from early 1970s to late 2010s. They use both a PCA and a dynamic factor analysis on time-series to estimate a weighted average value, which is a threshold for assessing financial stability vs. instability. They validate the FCI by regressing it against macroeconomic variables and including high-frequency non-financial measures of economic activity. [21] of the Federal Reserve Board produced a weekly "financial fragility indicator" for the USA computed in two steps using twelve market-based financial stress measures. In the first step, the standardized inputs are first reduced to three summary indicators: the level factor (variance-equal weighted average), the rate-of-change factor (rolling eight-week percentage change in the level factor) and the correlation factor (percentage of total variation in the individual stress variables explained by the first principal component over a rolling 26-week window). In the second step, they computed the financial fragility indicator as the fitted probability from a logit model with the three summary indicators as explanatory variables and a binary pre-defined crisis indicator as the dependent variable. Following a similar approach, [22] computed a similar weekly financial stability index for the Eurozone based on 16 financial variables. He used only the level and the rate of change as explanatory variables in a probit regression, but the correlation coefficient turns significant. He then identified crisis events for the computation of the binary indicator on the basis of a keyword-search through relevant parts of the ECB Monthly Bulletin. [23] pursued an approach similar to the one by [20] to construct a "financial market stress indicator" for Germany and the Euro area. The index comprises 23 and 22 raw stress factors, respectively, covering the banking sector, securities markets and foreign exchange conditions. [24] constructed a monthly "Financial Systemic Stress Index" for Greece, in which they aggregated 14 individual stress measures based on financial market data and monthly bank balance sheet data into five sub-indices using portfolio-theoretic measures (i.e. cross-correlations) computed through a multivariate GARCH model. They applied principal component analysis at the sub-index level, and the sub-indices were normalized using logistic transformation. A drawback of the approaches discussed above, while interesting and often accurate, is that they are focused on individual countries and thus lack generalization power. As a result, using these approaches, policymakers and practitioners cannot effectively analyze the complex interconnectedness of the global financial system.

In this paper, we want to present a fully data-driven statistical approach to building a synthetic index based on intrinsic information of the financial system. To this aim, we analyze a set of 17 Financial Soundness Indicators (FSI)[1] provided by the International Monetary Fund ranging from 2006 to 2017 for 140 countries that span the globe, including both strong and developing economies. Among the several methods available in the literature, as described above, we propose the employment of a Dynamic Factor Model approach (hereafter DFM) which allows us to fully compare countries in space and time. Differently from other papers which employ DFM, we adopt a country by country fitting strategy to elicit the inner country specific characteristics and then we combine results together by means of a Vector Autoregressive and Kalman filter approach. Such strategy is requested by the size of the dataset (140 countries $\times$ 17 variables $\times$ 11 years) and by identifiability constraints.

The paper is organized as follows: in Section 2 we present our modelling strategy, in Section 3 we briefly describe the data, in Section 4 we discuss our findings and in Section 5 we draw conclusions.

2 Methodology

We take advantage of a statistical methodology to build the index following a dimensionality reduction approach based on 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

[1] https://data.imf.org/?sk=51B096FA-2CD2-40C2-8D09-0699CC1764DA

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:

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

where $\mathscr{N}$ is the normal probability distribution and $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrix of the residuals of each equation in Eq. 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 Eq. 1 is under-determined. Thus, we deal with the problem as follows: first, following [25], 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 Eq. 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 (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^{th} and 99^{th} percentiles in order to check for the impact of outliers. In our context, in analogy with the classical R^2, we compute the RSS term as the squared residuals given after the reconstruction step using only the retained principal components and the TSS term as the total variance contained in the original variables. We fit the DFM

Table 1: 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

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.

3 The Data

We analyze a set of 17 Financial Soundness Indicators (FSI)[2] provided by the International Monetary Fund ranging from 2006 to 2017 for 140 countries that span the globe, including both strong and developing economies. Tables 1 and Table 2 present the summary statistics of the index's constituent variables 1 to 17 and their pairwise correlations.

Before employing the methodology explained in Section 2, we assess the data quality and cope with issues related to the presence of missing data. In this way, we can assure that results are not biased by low levels of quality and inconsistencies.

Indeed, some countries have missing values for the 17 pairs variables/years. As a result, we restrict our analysis to 116 countries, selected with an incidence of missing values not exceeding 30%. In our sample, 16 countries show a percentage of missing values between 20-29%. Thus, we apply two alternative data imputation methods: a Matrix Completion with Low Rank SVD method (MC-SVD) [26] and Bayesian Tensor Factorization (BTF) method [27]. 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

[2] https://data.imf.org/?sk=51B096FA-2CD2-40C2-8D09-0699CC1764DA

Table 2: 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

"slice" of countries-variables 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. Overall, we find that Bayesian Tensor Factorisation performs better.

4 Results

In this section we discuss results obtained by the employment of the DFM approach explained in section 2.

Table 3 reports the R^2 both on the whole dataset and on subsets with values trimmed for the 95th and 99th percentiles to check for outliers' impact. In our context, in analogy with the classical R^2, we compute the RSS term as the squared residuals given after the reconstruction step using only the selected number of factors and the TSS term as the total variance contained in the original variables. 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 be rather large, ranging between $[-0.76, 0.75]$. 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 a 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 [28] on the FSIND index which results into a $p-value \ll 0.01$ for all model specifications ensuring 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.

When dealing with dimensionality reduction approaches, a key role is played by loadings which interpretation in terms of relative importance and impact on each country can induce insights and speculations on the effects of the original variables.

Table 3: Results for DFM methods with different number of factors and factors interactions. R^2 is reported for the full dataset and for the 99^{th} and 95^{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^{th}	R^2 on 95^{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%	$\ll 0.01$

As described in Section 2 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 a 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 1 reports the distribution of the loadings for each independent variable over the 116 countries, representing the average trend across the years. The bimodal shape of all distributions implies a clear discriminative power of the index between less risky countries and riskier ones. Figure 2, instead, reports the contribution of independent variables on the loading for each country. Blue shaded points represent the positive contributions of the variables to each loading while red shaded points represent the negative ones. The bigger the points, the more the original variables contribute to the loading.

For sake of robustness and comparability, we fit an alternative dimension reduction technique which commonly represents a baseline: Principal Component Analysis. Such approach does not embed naturally any temporal dynamics differently from the DFM but it represents anyhow a well established benchmark. To overcome such issue of no proper temporal component elicitation, we fit the PCA approach employing all the countries together year by year. Results for the PCA can be found in Table 4 where we report the average R^2 both on the whole dataset and on subsets with values trimmed for the 95^{th} and 99^{th} percentiles.

The reader can clearly notice that the PCA approach performs worse in terms of Mean R^2 in any configuration. This stresses, once more, how important is the proper modelling of the temporal dynamics.

Table 4: Results based on the three different PCA. The first two principal component are provided and evaluated in terms of mean R^2 and mean R^2 trimmed the top 1^{st} and 5^{th} percentile.

Number of PC	Mean R^2	Mean R^2 on 99^{th}	Mean R^2 on 95^{th}	Im-Pesaran-Shin test
1	$22.2 \pm 6.1\%$	$26.6 \pm 15.3\%$	$33.3 \pm 20.6\%$	$\ll 0.01$
2	$37.9 \pm 9.8\%$	$43.6 \pm 14.8\%$	$50.6 \pm 15.5\%$	$\ll 0.01$

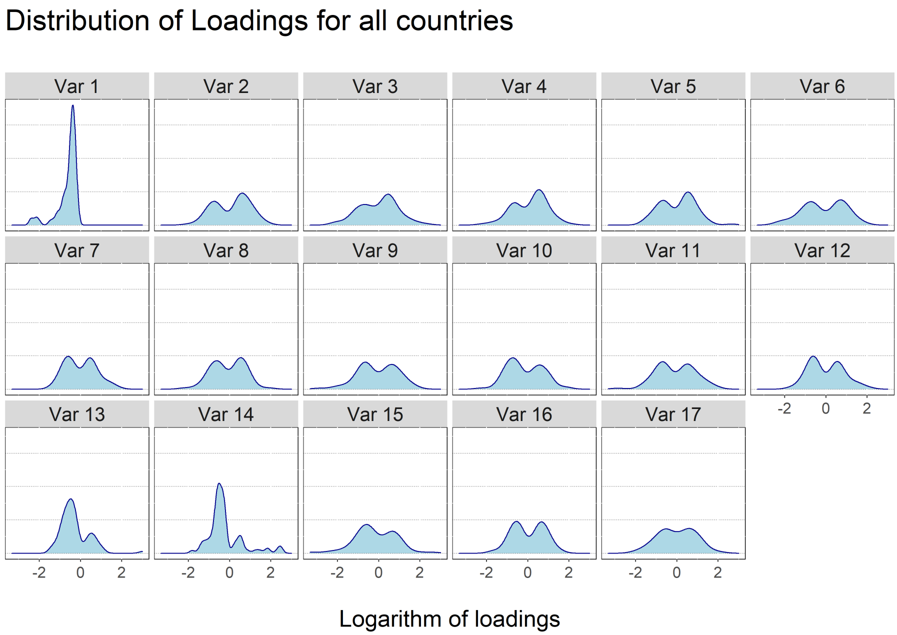

Figure 1: Loadings distribution over all countries for each independent variable. On x-axis is reported the logarithm of loading values. Variables' legend is below:
1 'Emerging Markets Bond (EMB) Capital to assets', 2 'Customer deposits to total non interbank loans', 3 'EMB Foreign currency liabilities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

5 Final Remarks

In this paper, we address an important topic relevant in many applications fields. It is common to be exposed to the management and analysis of several variables measured through time and space. In this sense, it plays a crucial role the employment of approaches to reducing the features space by creating one or more summary indexes which can be profitably used both for descriptive and predictive purposes. A common benchmark for dimension reduction is Principal Component Analysis which, despite its simplicity in the fitting process and in the interpretation, lacks of adaptability and accuracy when the available data presents strong temporal and spatial dimensions.

To overcome such limitation, we propose to employ a Dynamic Factor Model which naturally exploits any temporal dynamics and represents a valid alternative for the building up of summary indexes neither subjective nor fully experts-driven. To fully describe the potentials of the DFM, we take advantage of a financial application where it is relevant to produce an index that leverages on 17 financial indicators, able to measure the financial soundness of all the countries all over the world. We contribute to the research field, not only by proposing a new fully data driven financial soundness indicator, but also by offering a statistically sound methodology which goes beyond state of the art techniques and largely outperforms baseline approaches (namely PCA). The latter, indeed, fails in properly modelling the temporal evolution of the units of interest (in the present paper the countries) and in building a summary indicator representative enough of the variability contained in the original variables. We underline that the approach is general enough to be applied to contexts different from the present one.

The methodology can be improved in a number of ways. For example, non linear approaches could be employed and compared: neural networks autoencoders or extended Kalman filters would represent natural candidates. Further robustness checks would be useful to prove the consistency and stability of the methodology. For example, longer time series and or more granular data: from yearly based to quarterly or monthly based data. Moreover, scaling up the dimension of the analysis, by increasing the number of considered variables, can shed further light on the potentials of the DFM approach.

Lastly, it would be interesting to evaluate the performance in completely different fields like social, cultural or health related ones.

Declarations

The authors declare no competing interests.

References

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

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

[3] Mamatzakis, E.C., Tsionas, M.G.: A bayesian panel stochastic volatility measure of financial stability. International Journal of Finance & Economics (2020)

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

[5] Kabundi, A., Mbelu, A.: Estimating a timevarying financial conditions index for south africa. Working Papers, South African Reserve Bank **8008** (2017)

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

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

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

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

[10] Kliesen, K.L., Smith, D.C.: Measuring financial market stress. Federal Reserve Bank of St. Louis, Economic Synopses **2** (2010)

[11] Oet, M.V., Eiben, R., Bianco, T., Gramlich, D., Ong, S.J.: The financial stress index: Identification of systemic risk conditions. Federal Reserve Bank of Cleveland, Working Paper **11-30** (2011)

[12] Morales, M., Estrada, D.A.: Financial stability index for colombia. Annals of Finance **6**, 555–581 (2010)

[13] Holló, D., Kremer, M., Lo Duca, M.: Ciss – a composite indicator of systemic stress in the financial system. ECB Working Paper **1426** (2012)

[14] Koong, S.S., Law, S.H., Ibrahim, M.H.: Credit expansion and financial stability in malaysia. Journal of Economic Modelling **61**, 339–350 (2017)

[15] Arzamasov, V., Penikas, H.: A financial stability index for israel. Procedia Computer Science **31**, 985–994 (2014)

[16] Sere-Ejembi, A., Ini, S., Salihu, A., Atoi, V., Yaaba, B.: Developing banking system stability index for nigeria. CBN Journal of Applied Statistics **5**(1) (2014)

[17] Jakubik, P., Slacik, T.: How to measure financial (in)stability in emerging europe? Institute of Economic Studies (IES) Working Paper **13**, 1–21 (2013)

[18] Albulescu, C.T.: Financial stability and monetary policy: A reduced-form model for the euro area. Journal for Economic Forecasting **16**(1), 62–81 (2013)

[19] Sales, A.S., Areosa, W.D., Areosa, M.B.M.: Some financial stability indicators for brazil. Banco Central do Brasil Working Paper **287** (2012)

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

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

[22] Grimaldi, M.B.: Detecting and interpreting financial stress in the euro area. ECB Working Paper Series **1214** (2010)

[23] Roye, B.V.: Financial stress and economic activity in germany and the euro area. Kiel Working Paper **1743** (2011)

[24] Louzis, D.P., Vouldis, A.T.: 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 (2011)

[25] Holmes, E.E., Ward, E.J., Scheuerell, M.D.: Analysis of multivariate time-series using the marss package (2018)

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

[27] Khan, S.A., Ammad-ud-din, M.: tensorbf: an r package for bayesian tensor (2016)

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

Figure 2: 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'.