Identifying and interpreting the factors in factor models via sparsity : Different approaches

Thomas Despois, Catherine Doz

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2022 – 12

Indentifying and interpreting the factors in factor models via sparsity : Different approaches

Thomas Despois
Catherine Doz

JEL Codes: C32, C38, C55.
Keywords: Factor model, Identification, Factor interpretation, Sparsity, Sparse PCA, Factor rotation.

agence nationale de la recherche
AU SERVICE DE LA SCIENCE
INVESTIR L'AVENIR

Identifying and interpreting the factors in factor models via sparsity: Different approaches*

Thomas Despois[1,2] and Catherine Doz[1,2]

[1]Paris School of Economics
[2]Université Paris 1 Panthéon-Sorbonne

This version: March 2, 2022.
First version: July 31, 2019.

Abstract

With the usual estimation methods of factor models, the estimated factors are notoriously difficult to interpret, unless their interpretation is imposed via restrictions. This paper considers different approaches for identifying the factor structure and interpreting the factors without imposing their interpretation: sparse PCA and factor rotations. We establish a new consistency result for the factors estimated by sparse PCA. Monte Carlo simulations show that our exploratory methods accurately estimate the factor structure, even in small samples. We also apply them to two standard large datasets about international business cycles and the US economy: for each empirical application, they identify the same factor structure, offering a clear economic interpretation of the estimated factors. These exploratory methods can justify or complement approaches which impose the factor structure a priori, and can also be useful for applications in which factor interpretation is usually overlooked.

JEL classification: C32, C38, C55.
Keywords: factor model, identification, factor interpretation, sparsity, sparse PCA, factor rotation.

*Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France.
E-mail addresses: thomas.despois@psemail.eu and catherine.doz@univ-paris1.fr.
We are grateful to the editor Michael McCracken and three anonymous referees for comments that substantially improved the paper. We would also like to thank for helpful discussions Marie Bessec, Domenico Giannone, Sylvia Kaufmann, Siem Jan Koopman, Eric Renault, Mirco Rubin, Takashi Yamagata, as well as the participants at several PSE and Paris 1 internal seminars, the International Macroeconomics Workshop of the Banque de France - PSE Chair, the Tinbergen Institute Seminar, the 68th Annual Meeting of the French Economic Association, the 2019 Annual Conference of the International Association for Applied Econometrics, the 2019 European Summer Meeting of the Econometric Society, and the 2020 $(EC)^2$ Conference on High Dimensional Modeling in Time Series. This work was supported by the French National Research Agency with the EUR grant ANR-17-EURE-0001. Earlier versions of this paper circulated under the title "Another look into the factor model black box: factor interpretation and structural (in)stability."

1 Introduction

Factor models have been successfully used in various applications and have become increasingly popular in economics and finance. However, they suffer from a weakness: with the usual estimation methods, e.g. principal component analysis (PCA), the estimated factors are difficult, if not impossible, to interpret. Identifying the factor structure and interpreting the factors is essential in many economic and financial applications, in particular to enable a structural analysis (e.g. to study international business cycles).

An individual interpretation of the factors can be attempted through the factor loadings. They represent the link between each factor and each variable. A variable with a large loading (in absolute value) on a factor is highly correlated (in absolute value) with this factor. If a large number of loadings of a given factor are small (or even zero), it means that this factor is mainly related to a small number of variables, which generally enables an economic interpretation. Nevertheless, with the usual estimation methods, the estimated loading matrix tends to have a complex structure with many large loadings for each factor, making the interpretation of the estimated factors difficult, if not impossible.

Most of the literature dealing with the issue of identifying the factor structure and interpreting the factors places restrictions on the loadings to impose a specific interpretation to each factor, whether via hierarchical factor models, or factor rotations with identifying restrictions. As its name suggests, a hierarchical factor model (Kose, Otrok, and Whiteman, 2003) creates a hierarchy within the factors, e.g. a world factor, several regional factors, and many country-specific factors. Factor rotations with identifying restrictions were introduced in economics by Gürkaynak, Sack, and Swanson (2005). The goal was to impose the following interpretation for the two factors of their monetary policy application: surprise changes in the current federal funds rate target, and moves in interest rate expectations over the coming year that are not driven by changes in the current funds rate. Their approach was subsequently used by several other papers in monetary policy (e.g. Miranda-Agrippino and Rey, 2020). This methodology was then extended by Bai and Ng (2013), who proposed two rotations with identifying restrictions such that the first variable is affected by the first factor only, the second variable is affected by the first two factors only (resp. by the second factor only), the third variable is affected by the first three factors only (resp. by the third factor only), and so on.

An alternative and more exploratory approach for identifying the factor structure and interpreting the factors consists of using sparse factor models, introducing sparsity in the factor loadings via purely statistical methods with few or no assumptions about the factor structure. Sparse factor models have been used for some time in statistics, especially biostatistics (see e.g. West, 2003, among many others), but they have only very recently been introduced in economics and finance. They turned to be of help with factor interpretation in these fields too, whether implemented via penalization methods (Croux and Exterkate, 2011; Andreou, Gagliardini, Ghysels, and Rubin, 2019; Pelger and Xiong, 2020; Uematsu and Yamagata, 2021a,b), Bayesian endogenous clustering (Francis, Owyang, and Savascin, 2017), or Bayesian methods using a mixture prior inducing sparsity in the loadings (Kaufmann and Schumacher, 2019; Beyeler and Kaufmann, 2021). On a related issue, sparse factor models can also be used to identify the irrelevant variables for the estimation of the factor model (Kaufmann and Schumacher, 2017). In addition, they are able to slightly improve macroeconomic nowcasting and forecasting performances, for both developed and emerging economies (Croux and Exterkate, 2011; Kristensen, 2017; Kim and Swanson, 2018; Cepni, Güney, and Swanson, 2019; Uematsu and Yamagata, 2021a).

The main goal of this paper is to compare the performance of conceptually very different methods for implementing sparse factor models which enable to obtain interpretable factors while remaining deliberately agnostic with respect to the factor structure. We consider two approaches: sparse PCA and factor rotations. Sparse PCA is a machine learning technique which consists of penalizing or constraining the PCA problem to induce sparsity in the loadings. We use the SPCA algorithm (Zou, Hastie, and Tibshirani, 2006), which adds an elastic net penalty to the PCA problem. We also consider factor rotations that optimize a simple structure criterion concerning the loading matrix (instead of placing identifying restrictions). These long-established methods, which originate from statistics and psychometrics, essentially tend to eliminate the medium loadings (in absolute value), so that each estimated factor tends to have only very large and very small loadings (in absolute value). We use two standard factor rotations: varimax (Kaiser, 1958), which maximizes the sum of variances of squared loadings across factors, and quartimin (Carroll, 1953; Jennrich and Sampson, 1966), which minimizes the covariance of squared loadings between factors. In addition, we compare the results of these approaches to the findings in Francis, Owyang, and Savascin (2017) and Beyeler and Kaufmann (2021), who use Bayesian methods. Our approaches are even more agnostic (they do not require a prior or covariate series), are easier to implement, and have a lower computational cost (even sparse PCA). However, the posterior of the Bayesian methods enables to accurately measure the uncertainty in the estimates, e.g. the probability of each variable belonging to each cluster.

The contributions of the paper are both theoretical and empirical. We establish a new consistency result for sparse PCA. Under a standard set of assumptions of approximate dynamic factor models, we prove that sparse PCA consistently estimates the factor space if the ℓ_1 penalty parameter is $O(1/\sqrt{N})$, where N is the number of series. Monte Carlo simulations show that sparse PCA and factor rotations accurately recover the factor structure, even in small samples. They vastly outperform standard PCA when the true structure is sparse. We also study whether these techniques, which are designed to recover potential forms of sparse structures in the loading matrix, would not lead to an incorrect structure when the true structure is not sparse. It turns out they do not, and even tend to slightly outperform PCA, especially when the factors are correlated. The quartimin rotation and sparse PCA (with the SPCA algorithm) appear to be the preferable methods for recovering the correct factor structure (at least among the approaches considered here), but the varimax rotation yields very similar results in many cases. We apply our methods to two standard large datasets about international business cycles and the US economy. For each empirical application, they identify the same factor structure, offering a clear economic interpretation of the estimated factors. The four factors of the application to international business cycles can be interpreted as measures of regional business cycles respectively related to Europe, Latin America, Northern America, and Developed Asia. Francis, Owyang, and Savascin (2017) obtain very comparable results with their Bayesian sparse factor model with endogenous clustering. Our eight US macroeconomic factors are respectively related to output, prices, spreads, interest rates, housing, labor, the stock market, and money and credit. Similarly, Beyeler and Kaufmann (2021) obtain very comparable results with their Bayesian sparse FAVAR. This robustness supports the validity of our respective approaches.

It is quite remarkable that such long-established statistical methods as factor rotations, and state-of-the-art machine learning and Bayesian methods, lead to extremely similar results for two different applications. This suggests that standard estimators like PCA, while computationally convenient, usually recover a quite unnatural factor structure. Whenever any exploratory method allowing for a sparse structure in the loadings is used, it yields the same factor structure. This robustness could also constitute an interesting addition to the results of Giannone, Lenza, and Primiceri (2021): there is an illusion of sparsity in many predictive models, but not necessarily in factor models. These exploratory methods provide new lens to study different issues with factor models. In particular, they can justify or complement approaches which impose the factor structure a priori. For example, most of the literature studying international business cycles with factor models imposes the interpretation of each factor, and that each country is influenced by a single regional business cycle, corresponding to its geographic region. Our methods relax these restrictions, which proves to be relevant. Several economies appear to be mostly associated with the regional business cycle of another geographic region than their own (e.g. Japan is mostly associated with the European business cycle), or to be greatly influenced by several regional business cycles (e.g. the United Kingdom is greatly influenced by both the Northern American and European business cycles). These methods can also be useful for applications in which factor interpretation is usually overlooked, like forecasting and FAVAR models.

The rest of the paper is structured as follows. Section 2 introduces the econometric framework. Section 3 presents the results of Monte Carlo simulations. Section 4 provides two empirical applications. Section 5 concludes. The online Appendix collects the proof of consistency of the factors estimated by sparse PCA, the detailed presentation of our two datasets, and additional empirical results.

Throughout the paper, the ℓ_1 norm of a vector $v \in \mathbb{R}^n$ is denoted by $\|v\|_1 = \sum_{i=1}^n |v_i|$, and its ℓ_2 norm by $\|v\|_2 = \sqrt{v'v} = \sqrt{\sum_{i=1}^n v_i^2}$. For a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, we denote the maximum and minimum eigenvalues of $A'A$ by $\sigma_{max}(A'A)$ and $\sigma_{min}(A'A)$, the Frobenius norm of A by $\|A\|_F = \sqrt{\text{tr}(A'A)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$, and the induced Euclidean norm of A by $\|A\| = \sqrt{\sigma_{max}(A'A)}$. The all-ones vector of length n is denoted 1_n.

2 Econometric framework

2.1 Dynamic factor model

We observe N stationary time series over T periods. These observed variables are assumed to be driven by a fixed number r of unobserved common factors, with $r << N$. Let X be the $T \times N$ matrix of standardized observed variables, F the $T \times r$ matrix of factors, Λ the $N \times r$ matrix of factor loadings, and e the $T \times N$ matrix of idiosyncratic components. Denoting i the cross-section unit ($i = 1, \dots, N$), t the time unit ($t = 1, \dots, T$), x_{it} the observation of the ith variable at time t, $F_t = (f_{1t}, \dots, f_{rt})'$, and $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir})'$, the model is written as

$$x_{it} = \lambda_i' F_t + e_{it}, \text{ for } i = 1, \dots, N, \; t = 1, \dots, T. \tag{1}$$

In matrix notations, with $X_t = (x_{1t}, \dots, x_{Nt})'$ and $e_t = (e_{1t}, \dots, e_{Nt})'$, the model can also be written as

$$X_t = \Lambda F_t + e_t, \text{ for } t = 1, \dots, T, \tag{2}$$

or

$$X = F\Lambda' + e. \tag{3}$$

Factor models are characterized by the fact that (F_t) and (e_t) are unobserved stochastic processes with zero mean, which are assumed to be uncorrelated at all leads and lags. It should be noted that even though the relation between the observed variables X_t and the common factors F_t is static (no lag of F_t enters the model), this is a dynamic factor model, since (F_t) and (e_t) can be serially correlated processes. We assume that it is an approximate factor model, which means that the idiosyncratic components may be weakly cross-correlated and that the factors drive the bulk of the comovements of the observed variables. This is obtained through a standard set of assumptions, including in particular that the r eigenvalues of $\Lambda'\Lambda$ diverge at rate N whereas the covariance matrix of the idiosyncratic components stays bounded when N goes to infinity (see e.g. Stock and Watson, 2002a, Bai and Ng, 2002, and Doz, Giannone, and Reichlin, 2011, for some standard sets of assumptions).

It is well known that even though the factor space and the common component ΛF_t are uniquely identified, F and Λ are identified only up to an invertible matrix. Indeed, for any $r \times r$ invertible matrix R, the factor model is observationally equivalent to $X = \widetilde{F}\widetilde{\Lambda}' + e$, where $\widetilde{F} = F(R')^{-1}$ and $\widetilde{\Lambda} = \Lambda R$. This issue of indeterminacy will be the core of our paper. We aim to estimate the loading matrix Λ, which possibly has sparsity properties, in which case the factors should be interpretable. Indeed, the factor loading λ_{ij} measures the influence of the jth factor on the ith variable, so the jth column of $\widehat{\Lambda}$ can be used to get an interpretation of the jth estimated factor. With the usual estimation methods (e.g. PCA), the estimated loading matrix $\widehat{\Lambda}$ tends to have a complex structure with many large loadings for each factor, making the interpretation of the estimated factors difficult, if not impossible. In this paper, we build on the PCA estimator of the model, and propose several approaches which enable a sparse structure in $\widehat{\Lambda}$ to be recovered.

2.2 Principal component analysis

Stock and Watson (2002a) and Bai and Ng (2002) proved that the factor space of model (1) is consistently estimated by principal components under general sets of assumptions associated with approximate dynamic factor models. PCA can be used to estimate model (1) by solving the following nonlinear least squares problem:

$$\min_{F,\Lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \lambda_i' F_t)^2 = \min_{F,\Lambda} \|X - F\Lambda'\|_F^2. \tag{4}$$

It is well known that the factors obtained as solutions of (4) are the r first principal components of the data covariance matrix (up to scale), and that the factor loadings obtained as solutions of (4) are the corresponding eigenvectors (up to scale). This minimization problem is also well known to be associated to the same indeterminacy issue as the theoretical model: for any $r \times r$ invertible matrix R, if $(\widehat{F}, \widehat{\Lambda})$ is a solution, then $(\widetilde{F}, \widetilde{\Lambda})$ is also a solution if $\widetilde{\Lambda} = \widehat{\Lambda}R$ and $\widetilde{F} = \widehat{F}(R')^{-1}$. Normalization constraints thus have to be added to obtain a unique solution for the estimates of F and Λ (still up to column permutations and column sign changes). The two standard normalizations are $\widehat{F}'\widehat{F}/T = I_r$ with $\widehat{\Lambda}'\widehat{\Lambda}$ diagonal, and $\widehat{\Lambda}'\widehat{\Lambda}/N = I_r$ with $\widehat{F}'\widehat{F}$ diagonal. The first normalization sets the scale of the estimated factors, while the second sets the scale of the estimated loadings. Since most factor rotations require the unrotated estimated factors to be such that $\widehat{F}'\widehat{F}/T = I_r$, we adopt the first normalization. Under this normalization, the estimates of the loadings and factors are obtained with

$$\widehat{\Lambda}_{PCA} = \widehat{P}\widehat{D}^{1/2} \tag{5}$$

and

$$\widehat{F}_{PCA} = X\widehat{P}\widehat{D}^{-1/2} \tag{6}$$

where $\widehat{D} = \text{diag}(\widehat{d}_1, \ldots, \widehat{d}_r)$ is the $r \times r$ diagonal matrix whose entries correspond to the r largest eigenvalues of the data covariance matrix in decreasing order of magnitude, and $\widehat{P} = (\widehat{p}_1, \ldots, \widehat{p}_r)$ is the $N \times r$ matrix containing the corresponding normalized eigenvectors.[1]

[1] They are also such that $\widehat{d}_j = \|X\widehat{p}_j\|_2^2/T$ for $j = 1, \ldots, r$.

2.3 Factor rotations

The original method for obtaining interpretable factors consists of using factor rotations.[2] The idea is to take advantage of the indeterminacy of the estimated loadings and factors. The initially estimated loading matrix $\widehat{\Lambda}$ ($\widehat{\Lambda}_{PCA}$ here), leading to uninterpretable factors, is multiplied by an $r \times r$ invertible matrix R, chosen so that the transformed loading matrix optimizes a simple structure criterion:

$$\widehat{\Lambda}_{rotated} = \widehat{\Lambda} R. \tag{7}$$

The initially estimated factors are then transformed accordingly:

$$\widehat{F}_{rotated} = \widehat{F}(R')^{-1}. \tag{8}$$

The literature offers numerous rotation methods, which may be either *orthogonal* or *oblique*. Orthogonal rotations constrain the factors to stay orthogonal, whereas oblique rotations do not.[3] Even though orthogonal rotations are routinely used, it is highly recommended to apply oblique rotations, at the very least as a starting point for analysis (see Fabrigar, Wegener, MacCallum, and Strahan, 1999, and Costello and Osborne, 2005, among others). Indeed, many authors (e.g. Cattell, 1978, p. 128) argue that perfectly uncorrelated factors are unrealistic in most applications, so that oblique rotations are almost always necessary to recover the correct factor structure. Moreover, oblique rotations do not require the factors to be correlated, but simply allow for this. If the factors are truly uncorrelated, orthogonal and oblique rotations produce nearly identical results, with the correlations between the obliquely rotated factors being close to zero. We chose to study both orthogonal and oblique rotations to ensure the robustness of our results.

Among orthogonal rotations, varimax (Kaiser, 1958) is generally regarded as the best (see e.g. Fabrigar, Wegener, MacCallum, and Strahan, 1999). This method consists of finding the

[2] Factor rotations were originally designed to be applied to the maximum likelihood estimates of exact factor models. When N is large, under the assumptions of approximate factor models, it can be shown that this estimator is asymptotically equivalent to the PCA estimator (for more on this topic, see Chamberlain and Rothschild, 1983, and Schneeweiss and Mathes, 1995). Applying a factor rotation after a PCA is also valid, and commonplace in many disciplines (see Jolliffe, 2002, chapter 11).

[3] Although standard in the literature, the term "oblique rotation" is a misnomer: a rotation should preserve angles. Since the factors are orthogonal prior to a rotation, they should remain so afterwards. "Oblique rotations" are invertible transformations which allow the angles between the transformed factors to be different from 90 degrees, hence their name.

rotation matrix R maximizing the sum of variances of squared loadings across factors:[4]

$$R_{varimax} = \arg\max_R \left\{ \sum_{j=1}^{r} \left[\frac{1}{N} \sum_{i=1}^{N} (\widehat{\Lambda} R)_{ij}^4 - \left(\frac{1}{N} \sum_{i=1}^{N} (\widehat{\Lambda} R)_{ij}^2 \right)^2 \right] \right\} \quad \text{s.t. } R'R = I_r. \tag{9}$$

We compute $R_{varimax}$ with the gradient projection algorithm for orthogonal rotations proposed by Jennrich (2001).

Among oblique rotations, the oblimin family is the most popular because of its great flexibility. Within this family, the quartimin rotation (Carroll, 1953; Jennrich and Sampson, 1966) tends to be recommended because of its simplicity and results (see e.g. Costello and Osborne, 2005). It searches for the matrix R minimizing the covariance of squared loadings between factors:[5]

$$R_{quartimin} = \arg\min_R \left\{ \sum_{j \neq k} \sum_{i=1}^{N} (\widehat{\Lambda} R)_{ij}^2 (\widehat{\Lambda} R)_{ik}^2 \right\} \quad \text{s.t. } \operatorname{diag}((R'R)^{-1}) = 1_r. \tag{10}$$

We compute $R_{quartimin}$ with the gradient projection algorithm for oblique rotations proposed by Jennrich (2002).

The covariance matrix of the rotated estimated factors is

$$\begin{aligned} \frac{1}{T} \widehat{F}'_{rotated} \widehat{F}_{rotated} &= \frac{1}{T} R^{-1} \widehat{F}' \widehat{F} (R')^{-1} \\ &= R^{-1} (R')^{-1} \quad \text{for } \frac{1}{T} \widehat{F}' \widehat{F} = I_r \\ &= (R'R)^{-1}. \end{aligned} \tag{11}$$

For quartimin, the constraint $\operatorname{diag}((R'R)^{-1}) = 1_r$ maintains the scale of the estimated factors prior to the rotation. For varimax, the stronger constraint $R'R = I_r$ also maintains the orthogonality of the estimated factors.

Some points need to be clarified about the effects of a rotation. First, a factor rotation neither deteriorates nor improves the fit of the factor model: the estimated factor space, the total variance explained by the estimated factors, and all the commonalities[6] remain the same. In particular, since PCA consistently estimates the factor space, this result is not lost after a

[4] Hence the name: VARIance MAXimization.

[5] Hence the names: the oblimin rotations all consist of OBLIque rotations via the MINimization of a simplicity criterion, and the quartimin rotation more specifically involves the minimization of a criterion including fourth degree terms.

[6] The commonality of a variable refers to the part of its variance explained by the estimated factors.

factor rotation. Secondly, factor rotations like varimax or quartimin achieve near-sparsity rather than sparsity: they do not generate any zero loadings, but instead ensure that each estimated factor tends to have only very large and very small loadings in absolute value. Thirdly, when the estimated factors are included in a diffusion index or another model, using the unrotated or rotated estimated factors yields exactly the same predicted values $\widehat{y}$. This is because a rotation does not change the estimated factor space, and hence does not change the orthogonal projection of the predicted variable y onto this space either.

We tried several other orthogonal and oblique rotations, leading to very similar results. For the sake of brevity, we only consider the standard varimax and quartimin rotations in the following.

2.4 Sparse principal component analysis

With sparse PCA, the machine learning literature offers an alternative to factor rotations in order to get interpretable factors, starting with Jolliffe, Trendafilov, and Uddin (2003). The general idea is to penalize or constrain the PCA problem to induce sparsity in the estimated loadings. The resulting estimated factors are easier to interpret, but at the cost of a slightly lower explained variance. This constitutes a trade-off between interpretability and explained variance. However, in many applications, the loss of explained variance implied by the sparse factors is considered to be "small and relatively benign" (d'Aspremont, Bach, and El Ghaoui, 2008, among others).

Sparse PCA has been successfully used in various fields, especially biostatistics and finance. Indeed, this method offers two distinctive features with respect to factor rotations: it yields zero loadings (instead of very small loadings in absolute value),[7] and the output can here be governed by one or several hyperparameters, allowing for more flexibility and finer control. In addition, contrary to the PCA estimates, the factors estimated by sparse PCA can be correlated, as with oblique rotations.

There has been extensive research in machine learning about sparse PCA, leading to dozens of competing algorithms (see the survey by Trendafilov, 2014). We chose to work with the SPCA algorithm (Zou, Hastie, and Tibshirani, 2006), one of the two algorithms recommended by Trendafilov (2014). SPCA consists of formulating PCA as a regression-type problem, and

[7] This is crucial in fields like biostatistics (to focus on only a few genes), or finance (to limit the number of assets in the portfolio, thus reducing the transaction costs).

adding an elastic net penalty in order to produce *modified* principal components, which may be sparse.

The elastic net penalty (Zou and Hastie, 2005) is a convex combination of the ℓ_1 (lasso) and the ℓ_2 (ridge) penalties associating their comparative advantages. The lasso penalty enforces sparsity in the estimated coefficients by performing variable selection. Nevertheless, it comes up with limitations, especially when the variables are highly correlated. Zou and Hastie (2005) show that when $T > N$, if there are high correlations between the variables, the prediction performance of the lasso tends to be dominated by ridge regression, and thus also by the elastic net. Moreover, if there is a group of variables with very high pairwise correlations, then the lasso tends to select only one variable from this group and ignore the others. Using a strictly convex penalty instead of the lasso (e.g. by adding a ridge penalty) overcomes this problem by encouraging what is called the *grouping effect* in machine learning: the regression coefficients of a group of highly correlated variables tend to be equal (up to a change of sign if negatively correlated). These issues are crucial with economic and financial datasets since they usually contain groups of highly correlated variables (economic activity series, prices, interest rates, etc.). SPCA is one of the very few sparse PCA algorithms offering the grouping effect.

The SPCA problem can be written as:[8]

$$\min_{A,B} \left\{ \frac{1}{T}\|X - XBA'\|_F^2 + \sum_{j=1}^{r} \kappa_{1j}\|b_j\|_1 + \kappa_2 \sum_{j=1}^{r} \|b_j\|_2^2 \right\} \quad \text{s.t.} \quad A'A = I_r, B = (b_1, \dots, b_r). \tag{12}$$

The jth column of $\widehat{B}$, denoted $\widehat{b}_j$, is used to construct the jth modified principal component, κ_{1j} is the ℓ_1 penalty parameter controlling the level of sparsity in $\widehat{b}_j$, and κ_2 is the ℓ_2 penalty parameter. If desired, the κ_{1j} hyperparameters can be distinct so as to impose a substantially different level of sparsity for each modified principal component.

It is worth emphasizing the links between this minimization problem and PCA. First, elementary calculations show that the solution of the problem

$$\min_{P} \frac{1}{T}\|X - XPP'\|_F^2 \quad \text{s.t.} \quad P'P = I_r \tag{13}$$

[8] In Zou, Hastie, and Tibshirani (2006), the first term of the optimization problem is not premultiplied by $1/T$. However, it is clear that they normalized the data by a $1/\sqrt{T}$ term, since they define the data covariance matrix as $X'X$. Here, we keep the standard notation and define the data covariance matrix as $X'X/T$, hence the premultiplication by $1/T$.

is $\widehat{P}$, the matrix of the r normalized eigenvectors associated to the r largest eigenvalues of $X'X/T$. Thus, equation (12) is a generalization of (13) where PP' has been replaced by BA', with the addition of an elastic net penalty. Secondly, if the penalty terms are removed in (12), i.e. if $\kappa_{1j} = 0$ for $j = 1, \ldots, r$ and $\kappa_2 = 0$, then the solution of (12) is $A = B = \widehat{P}$.[9] Thus, although it is formulated in a different way (due to the presence of the two matrices A and B, which facilitates the construction of the estimation algorithm), (12) is indeed a penalized version of the PCA minimization problem (13).

SPCA algorithm and estimators. For the sake of completeness, we present the main steps of the SPCA algorithm, for given values of the hyperparameters κ (whose tuning is explained below). We refer the reader to Zou, Hastie, and Tibshirani (2006) for a complete presentation. The algorithm is initialized by setting $A^{(1)} = \widehat{P}$, where $\widehat{P}$ is the $N \times r$ matrix whose columns are the normalized eigenvectors associated to the r largest eigenvalues of the data covariance matrix, in decreasing order. Denoting $A^{(k)}$ the current value of the matrix A, the kth iteration of the algorithm then consists of the following two steps.

- **Step 1 (elastic net regression)**
  Given $A^{(k)} = \left(a_1^{(k)}, \ldots, a_r^{(k)}\right)$, compute $B^{(k)} = \left(b_1^{(k)}, \ldots, b_r^{(k)}\right)$ such that

$$B^{(k)} = \arg\min_B \left\{ \frac{1}{T} \|X - XBA^{(k)\prime}\|_F^2 + \sum_{j=1}^{r} \kappa_{1j} \|b_j\|_1 + \kappa_2 \sum_{j=1}^{r} \|b_j\|_2^2 \right\}. \tag{14}$$

  It can be shown (see Zou, Hastie, and Tibshirani, 2006, for details) that this amounts to solve separately for $j = 1, \ldots, r$ the problem

$$b_j^{(k)} = \arg\min_b \left\{ \frac{1}{T} \|X a_j^{(k)} - Xb\|_2^2 + \kappa_{1j} \|b\|_1 + \kappa_2 \|b\|_2^2 \right\}. \tag{15}$$

- **Step 2 (reduced rank Procustes rotation)**
  Given $B^{(k)}$, $A^{(k+1)}$ is defined by

$$A^{(k+1)} = \arg\min_A \frac{1}{T} \|X - XB^{(k)}A'\|_F^2 \quad \text{s.t.} \quad A'A = I_r. \tag{16}$$

[9]This can be proven using the following steps: first, for a given A, the derivation of the objective function with respect to B shows that it is minimum for $B = A$; then the minimization with respect to A, with the condition $A'A = I_r$, gives $A = \widehat{P}$.

It can be shown (see Theorem 4 in Zou, Hastie, and Tibshirani, 2006) that $A^{(k+1)}$ can be obtained in the following way: if the SVD of $\frac{X'X}{T}B^{(k)}$ is given by $\frac{X'X}{T}B^{(k)} = U_k \Delta_k V_k'$, then $A^{(k+1)} = U_k V_k'$.

These two steps are iterated until convergence.[10] In Zou, Hastie, and Tibshirani (2006), denoting $\widehat{B}$ the matrix derived from the algorithm, the loadings are obtained by the normalization of each column of $\widehat{B}$ to unit length, i.e. $\widehat{b}_j / \|\widehat{b}_j\|_2$. We choose to adopt a different normalization in order to be consistent with our PCA normalization, and scale the estimated factors such that $\text{diag}(\widehat{F}'_{SPCA}\widehat{F}_{SPCA}/T) = 1_r$. The estimated loadings and factors are then obtained with $\widehat{\Lambda}_{SPCA} = \widehat{B}\widehat{D}^{1/2}$ and $\widehat{F}_{SPCA} = X\widehat{B}\widehat{D}^{-1/2}$, where $\widehat{d}_j = \|X\widehat{b}_j\|_2^2/T$ for $j = 1, \ldots, r$, with $\widehat{B}$ and $\widehat{\Lambda}$ sparse.

Tuning the SPCA hyperparameters κ. First of all, we set all the ℓ_1 penalty parameters to be equal: $\kappa_{1j} = \kappa_1$ for $j = 1, \ldots, r$. This leaves us with the tuning of only two hyperparameters (κ_1 and κ_2) instead of $r + 1$, drastically reducing the computational cost of the tuning (which can still be high), and also the risk of overfitting.[11] Tuning hyperparameters is always a challenge, especially when the variables are time series. This can be done by cross-validation (CV), or by minimizing an information criterion. The standard CV techniques like k-fold CV or leave-one-out CV, which assume that the subsamples are independent and identically distributed, are invalidated when using time series due to the inherent serial correlation.[12] A time series CV (simulated out-of-sample exercise) would be relevant, but it seems preferable to use an information criterion instead. First, SPCA is not only used for forecasting applications. Secondly, the computational time can be much lower using an information criterion. Thirdly, Smeekes and Wijler (2018) show that tuning the penalty parameters with the BIC criterion rather than with time series CV leads to a better forecasting performance in the context of macroeconomic forecasting using penalized regression methods. Thus, we choose to tune the hyperparameters by minimizing a BIC-type criterion, as Kristensen (2017) does with the unique

[10] It is clear that, by construction, the objective function decreases at each iteration so that the algorithm does converge.

[11] The introduction of this constraint has almost no impact on the results of our two empirical applications.

[12] Except for the very specific case of purely autoregressive models with uncorrelated errors (Bergmeir, Hyndman, and Koo, 2018).

tuning parameter of another sparse PCA algorithm (sPCA-rSVD, Shen and Huang, 2008):

$$(\widehat{\kappa}_1, \widehat{\kappa}_2) = \arg\min_{\kappa_1,\kappa_2} \left\{ \log\left(\frac{1}{NT} \|X - \widehat{F}_{SPCA}(\kappa_1,\kappa_2)\widehat{\Lambda}'_{SPCA}(\kappa_1,\kappa_2)\|_F^2 \right) + m(\kappa_1,\kappa_2)\frac{\log(NT)}{NT} \right\} \tag{17}$$

where m is the number of nonzero loadings in $\widehat{\Lambda}_{SPCA}$. Obviously, $\widehat{\Lambda}_{SPCA}$, $\widehat{F}_{SPCA}$, and m depend on (κ_1, κ_2). We solve (17) via a grid search, by minimizing the BIC-type criterion over the following grid:

$$\kappa_1 \in \{0, 0.1, \dots, 1\}, \tag{18}$$

$$\kappa_2 \in \{0, 0.1, \dots, 1\}. \tag{19}$$

Consistency of the estimated factors. To the best of our knowledge, the issue of the consistency of the factors estimated by sparse PCA has only been addressed by Kristensen (2017), who proved that sparse PCA consistently estimates the factor space in the case of an approximate dynamic factor model.[13] However, it is worth noting that in Kristensen (2017) the ℓ_1 penalty parameter is supposed to tend to zero quickly, since it vanishes asymptotically at rate 1/T. Consequently, the resulting estimates may not really be sparse. Here, we prove that, in the approximate dynamic factor model framework, under a set of assumptions ensuring the consistency of the PCA estimator, sparse PCA consistently estimates the factor space if $\kappa_1 = O(1/\sqrt{N})$. This condition is far less restrictive than the one in Kristensen (2017) since in most economic applications N is of the same order of magnitude as T or even smaller. Different sets of assumptions have been considered in the literature to define approximate dynamic factor models (see e.g. Stock and Watson, 2002a, Bai and Ng, 2002, Forni, Hallin, Lippi, and Reichlin, 2004, or Doz, Giannone, and Reichlin, 2011). Here we use the Doz, Giannone, and Reichlin (2011) set of assumptions, but we could have obtained similar results with, for example, the Stock and Watson (2002a) or Bai and Ng (2002) sets of assumptions. In particular, we use the following assumptions from Doz, Giannone, and Reichlin (2011).

Assumption CR1. $\liminf_{n\to\infty} \frac{1}{n}\sigma_{min}(\Lambda'\Lambda) > 0$.

Assumption CR2. $\limsup_{n\to\infty} \frac{1}{n}\sigma_{max}(\Lambda'\Lambda)$ is finite.

Assumption CR3. $\limsup_{n\to\infty} \sum_{h\in\mathbb{Z}} \|E(e_t e'_{t-h})\|$ is finite, which implies in particular that

[13] Kristensen (2017) considers the case where there is no ridge penalty, i.e. when $\kappa_2 = 0$.

$\limsup_{n\to\infty} \|\Sigma_e\|$ is finite.

Assumption CR4. $\inf_n \sigma_{min}(\Sigma_e) > 0$.

We then obtain the following theorem.

Theorem 1. *Let $\widehat{F}_t$ denote the PCA estimator of F_t, and $\widetilde{F}_t$ the SPCA estimator of F_t. Under Assumptions CR1-CR4, if $\kappa_1 = O(1/\sqrt{N})$, then there exists an invertible matrix M, with $M = O_P(1)$ and $M^{-1} = O_P(1)$, such that*

$$\frac{1}{T}\sum_{t=1}^{T} \|\widetilde{F}_t - M\widehat{F}_t\|_2^2 = O_P\left(\frac{1}{N}\right). \tag{20}$$

Proof. See online Appendix A. □

Thus, we see that the factors estimated by SPCA asymptotically span the same space as the factors estimated by PCA. Since the factors estimated by PCA asymptotically span the true factor space, the same applies for the factors estimated by SPCA. Our current result is a mean result: consistency for any t is left for future research.

3 Monte Carlo simulations

3.1 Design

To assess the performance of the different estimators in recovering the factor structure, we conduct Monte Carlo simulations. The data generating process (DGP) is detailed below.

$$x_{it} = \lambda_i' F_t + \sqrt{\theta} e_{it} = \sum_{j=1}^{r} \lambda_{ij} f_{jt} + \sqrt{\theta} e_{it} \tag{21}$$

$$F_t = \Phi_F F_{t-1} + u_t,\ \Phi_F = \phi_f I_r,\ u_t \overset{iid}{\sim} \mathcal{N}(0_r, \Sigma_u) \tag{22}$$

$$e_t = \Phi_e e_{t-1} + v_t,\ \Phi_e = \phi_e I_r,\ v_t \overset{iid}{\sim} \mathcal{N}(0_r, \Sigma_v) \tag{23}$$

This DGP can generate serially and cross-sectionally correlated factors, as well as serially and cross-sectionally correlated idiosyncratic components. All of the experiments are repeated 1,000 times. We set $N = 60$ and $T = 100$, $r = 3$, $\phi_f = 0.5$, and $\phi_e \in \{0, 0.5\}$.

The covariance (and correlation) matrix of the factors is defined as

$$\Sigma_F = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}. \tag{24}$$

We set $\rho = (\rho_1, \rho_2, \rho_3) = \{(0,0,0), (-0.4, 0.2, 0)\}$, and deduce Σ_u with $\Sigma_u = \Sigma_F - \Phi_F \Sigma_F \Phi_F' = (1 - \phi_f^2)\Sigma_F$.

The covariance (and correlation) matrix of the idiosyncratic components is a Toeplitz matrix defined as $\Sigma_e = (\tau^{|i-j|})_{ij}$. The parameter τ controls the cross-correlation in the idiosyncratic components. For $\tau = 0$, there is no cross-correlation in the idiosyncratic components (exact factor model). We set $\tau \in \{0, 0.5\}$, and deduce Σ_v with $\Sigma_v = \Sigma_e - \Phi_e \Sigma_e \Phi_e' = (1 - \phi_e^2)\Sigma_e$.

Concerning the loadings, we consider a dense structure and a block structure.

1. **Dense structure in the loadings.** The loadings are simply generated according to

$$\lambda_{ij} \overset{iid}{\sim} \mathcal{N}(0, 1). \tag{25}$$

2. **Block structure in the loadings.** Each of the three columns of Λ is partitioned into subvectors of large or small (possibly zero) loadings:

$$\Lambda = \begin{pmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 \end{pmatrix} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{pmatrix}, \tag{26}$$

where Λ_{ij} are vectors of length 20 and

$$\begin{aligned} \Lambda_{ij} &\overset{iid}{\sim} \mathcal{N}(\mu_l . 1_{20}, \sigma_l^2 . I_{20}) \quad \text{if} \quad i = j, \\ \Lambda_{ij} &\overset{iid}{\sim} \mathcal{N}(\mu_s . 1_{20}, \sigma_s^2 . I_{20}) \quad \text{if} \quad i \neq j. \end{aligned} \tag{27}$$

We set $\mu_s = 0$, and $(\sigma_l^2, \sigma_s^2) = \{(0.2, 0), (0.2, 0.2), (0.5, 0.2)\}$. When $\sigma_s^2 = 0$, the small loadings are zero, and the sparsity level in Λ is $2/3$. To be able to compare the results between the dense structure and the block structure, the expected norm (e.g. ℓ_2 norm) of the loadings should be the same in both cases. To achieve this, μ_l is deduced from μ_s, σ_l^2, and σ_s^2 such that

$E(\Lambda_j'\Lambda_j/N) = 1$ like with the dense structure. Finally, for both the dense structure and the block structure cases, θ is set such that the common and idiosyncratic components have the same variance when neither the factors nor the idiosyncratic components are cross-sectionally correlated.

Before computing the metrics assessing their accuracy, the estimated loadings and factors need to undergo a three-step post-processing procedure.

1. The columns of $\widehat{\Lambda}$ are rescaled so that their ℓ_2 norm is equal to the expected ℓ_2 norm of the simulated loadings:

$$\check{\Lambda}_j = \sqrt{N}\frac{\widehat{\Lambda}_j}{\|\widehat{\Lambda}_j\|_2} \text{ for } j = 1, \ldots, r. \tag{28}$$

   This yields $\check{\Lambda}_j'\check{\Lambda}_j/N = 1$, which is equal to $E(\Lambda_j'\Lambda_j/N)$. The metrics we use in the following to measure the estimation accuracy are invariant to the scale of the estimated factors, so we do not rescale them.

2. Since F and Λ are identified up to column permutations, when necessary, the columns of $\widehat{F}$ and $\check{\Lambda}$ are reordered to correspond to the ordering in F and Λ. The columns of Λ and $\check{\Lambda}$ (taken in absolute value) with the smallest distance according to the ℓ_2 norm are iteratively associated. The order of the estimated factors is then switched accordingly.

3. Since F and Λ are also identified up to column sign changes, when necessary, column sign changes are operated in the matrix containing the reordered estimated factors so that each simulated factor is positively correlated with its estimate. Column sign changes in the matrix containing the rescaled and reordered loadings are then operated accordingly. The resulting loading matrix is denoted $\widetilde{\Lambda}$.

For the loadings, the estimation accuracy is measured with the root mean squared error (RMSE) and the mean absolute error (MAE).

$$RMSE = \sqrt{\frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{r}(\lambda_{ij} - \widetilde{\lambda}_{ij})^2} \tag{29}$$

$$MAE = \frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{r}|\lambda_{ij} - \widetilde{\lambda}_{ij}| \tag{30}$$

A smaller RMSE or MAE indicates a more accurate estimation of the loadings.

For the factors, the estimation accuracy is measured with the correlations among the post-processed estimated factors (denoted $\widehat{\rho}$), and the R^2 of the multivariate regression of the true factors onto the estimated factors

$$R^2_{F,\widehat{F}} = \frac{\|P_{\widehat{F}} F\|_F^2}{\|F\|_F^2}, \tag{31}$$

where $P_{\widehat{F}} = \widehat{F}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'$ is the projection matrix onto the space spanned by the estimated factors. An R^2 closer to one implies a more accurate estimation of the space spanned by the true factors. This metric is invariant to the chosen rotation and the post-processing procedure.

3.2 Results

Tables 1 and 2 summarize the results of the simulations with a dense structure in the loadings. The four estimation methods lead to very comparable results. PCA varimax, PCA quartimin, and SPCA are designed to recover potential sparse or near-sparse structures in the loading matrix, but they do not yield an incorrect structure when the true structure is dense. They even tend to slightly outperform standard PCA concerning the estimation accuracy of the loadings (as measured by the RMSE and MAE), especially when the factors are correlated. The factor space is accurately estimated: the R^2s are close to one. With PCA quartimin and SPCA, the correlations among the factors are very accurately estimated when they are uncorrelated, but underestimated in absolute value when they are correlated.

INSERT TABLES 1-2 ABOUT HERE.

Tables 3 and 4 summarize the results of the simulations with a block structure in the loadings. PCA is now vastly outperformed by the three other estimation methods concerning the estimation accuracy of the loadings. As with the dense structure case, PCA varimax and PCA quartimin lead to extremely similar results when the factors are uncorrelated, but PCA quartimin now dominates PCA varimax slightly more clearly when the factors are correlated. Rotated PCA outperforms SPCA when the structure is near-sparse, but the opposite is the case when the structure is sparse. However, the magnitude of the performance improvement with SPCA when the structure is sparse should be interpreted in the light of the high level of sparsity in the model with this DGP. Concerning the factors, the four estimation methods still lead to similar R^2s, close to one. When the factors are uncorrelated, their estimated correlations are approximately equal to zero with PCA quartimin and SPCA, whereas when they are correlated, their correlations are slightly underestimated in absolute value with SPCA, and a little more substantially with PCA quartimin (especially when the structure is near-sparse instead of sparse) but the results are still satisfactory.

INSERT TABLES 3-4 ABOUT HERE.

For each DGP and estimation method, the introduction of serial correlation or cross-correlation in the idiosyncratic components only has a marginal influence on the estimation accuracy (except when adding cross-correlation in the idiosyncratic components to a DGP with a block structure in the loadings, in which case the R^2s are then affected slightly more, but not the other metrics). None of the competing estimation methods seem to be more affected than the others. For a given DGP, the introduction of correlation among the factors improves the general performance of PCA quartimin and SPCA relative to standard PCA and PCA varimax, which require the estimated factors to be uncorrelated.

In short, rotated PCA and SPCA slightly outperform PCA in recovering the correct factor structure if that structure is dense, and vastly if that structure is sparse or near-sparse. The quartimin rotation and SPCA appear to be the preferable methods for recovering the correct factor structure (at least among the approaches considered here).

In our simulations, PCA quartimin always performs as well or better than PCA varimax (and there is usually no reason to impose the estimated factors to be uncorrelated as is the case with orthogonal rotations), and outperforms SPCA when the factor structure is near-sparse. SPCA dominates PCA quartimin when the factor structure is sparse, but this comes at several costs. First, the computational time of SPCA can be considerable, which is a serious inconvenience in some applications, especially forecasting. In our Monte Carlo study, the 1,000 repetitions of the experiment for a given DGP require less than one minute for PCA quartimin, as compared to several days for SPCA. Another problem for forecasting applications is that SPCA cannot handle missing observations per se, and its computational time tends to discourage the use of an EM algorithm to treat them as with PCA (as proposed by Stock and Watson, 2002b). A possible solution could be to first treat the missing observations, e.g. with PCA and an EM algorithm, and then apply SPCA on the resulting balanced panel. Besides, SPCA implies a small loss of explained variance with respect to PCA (rotated or not). In view of this, the quartimin rotation can still represent a good compromise over SPCA when the factor structure is sparse. Ultimately, the choice between PCA quartimin and SPCA also depends on the importance of having sparsity instead of near-sparsity in the estimated loadings (e.g. in financial applications, to limit the number of assets in the portfolio, thus reducing the transaction costs), and the loss of explained variance it implies.

4 Empirical applications

4.1 International business cycles

This application follows the wealth of literature studying international business cycles with factor models. For comparability purposes, we consider a dataset very similar to other papers from this literature, especially Kose, Otrok, and Whiteman (2003), Francis, Owyang, and Savascin (2017), and Kaufmann and Schumacher (2017). Our measure of business cycle conditions is the annual growth rate of the real GDP at constant national prices. We include 60 countries from 7 regions: North America, Latin America, Europe, Africa, Emerging Asia, Developed Asia, and Oceania (see online Appendix B.1 for further details). The series are retrieved from the version 9.1 of the Penn World Tables (Feenstra, Inklaar, and Timmer, 2015). We use the maximum time span possible with this sample (1961-2017), and standardize the series.

The most standard methods for selecting the number of factors are the information criteria proposed by Bai and Ng (2002). Their six main information criteria select between one and four factors for this dataset. We choose to include four factors in the model, like Francis, Owyang, and Savascin (2017) and Kaufmann and Schumacher (2017). For $r = 4$, the grid search of the SPCA hyperparameters yields $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.6, 0.8)$.

As is routinely reported with PCA, the estimated factors are very difficult to interpret (the estimated factor loadings are presented in online Appendix B.2). Concerning the first factor, the countries with the largest loadings are North American, European, and Developed Asian countries. This factor could thus be interpreted as specific to developed countries, but such an interpretation is imperfect. Several emerging or developing countries also have large loadings (Brazil, Colombia, Costa Rica, El Salvador, Guatemala, Mexico, and South Africa), while several developed countries only have medium loadings (Australia, Iceland, Korea, Malaysia, Singapore, and Thailand), or even small loadings (Ireland and New Zealand). Similarly, the second factor tends to be associated with Latin America, but many countries outside this region also have quite large loadings in absolute value. The remaining two estimated factors are impossible to interpret.

Conversely, the loadings estimated by rotated PCA or SPCA systematically offer a clear economic interpretation of the factors (Figure 1). The interpretation tends to be slightly more straightforward with SPCA due to the sparsity in the estimated loadings, but the three methods lead to a very similar pattern of estimated loadings. The largest loadings of the first factor correspond to European countries so the first factor can be interpreted as a measure of the European business cycle. Japan appears to be strongly associated with it. The second factor is the Latin American business cycle: the largest loadings are associated with many Latin American countries and Mexico. South Africa also has a quite large loading. A possible explanation for the presence of South Africa in this group is that most of these countries are also highly dependent on oil prices (whether as exporters or importers). The third factor can be interpreted as the Northern American business cycle: Canada and the USA have large loadings. Several Latin American countries that are closely linked to the economy of the USA, in particular because their economy is or was dollarized, also have large loadings (Costa Rica, El Salvador, Honduras) or medium loadings (Chile, Guatemala). It is worth noting that several other former British Commonwealth countries have medium loadings in absolute value (Australia, Kenya, the United Kingdom, and Zimbabwe). This is especially true with PCA varimax and PCA quartimin. The fourth factor is clearly the Developed Asian business cycle. Indonesia, one of the richest Emerging Asian countries, also has a large loading. Japan has a small loading with PCA varimax and PCA quartimin, and even a zero loading with SPCA. It is not really surprising to find that Japan is more synchronized with the European economies than with the other Asian economies. In particular, Japan started its development earlier than the Asian Tigers and Tiger Cub Economies.

Despite the high level of sparsity in the loadings estimated by SPCA (74.6%), the loss of explained variance is quite limited: the factors estimated by PCA (rotated or not) account for 45.6% of the data variance, versus 44.2% for the factors estimated by SPCA.

INSERT FIGURE 1 ABOUT HERE.

Since PCA varimax, PCA quartimin, and SPCA lead to very similar estimated loadings, they also lead to very similar estimated factors (see online Appendix B.2). The dynamics of the estimated factors are consistent with economic history (see Figure 2 for an illustration with the factors estimated with SPCA). For example, the impact of the Great Recession appears to have been more pronounced on the Northern American and European business cycles. In addition, the Northern American business cycle accurately captures the early 1980s recession, while the Developed Asian business cycle accurately captures the 1997 Asian financial crisis. The estimated factors are strongly correlated with the real GDP growth rates of some representative countries. For example, the correlation between the SPCA estimated Northern American business cycle and the real GDP growth rate of the USA is 85%. These correlations tend to be higher with the SPCA estimates than with the PCA quartimin estimates, and higher with the PCA quartimin estimates than with the PCA varimax estimates. With PCA quartimin and SPCA, which allow the estimated factors to be correlated contrary to PCA and PCA varimax, the pattern of the correlation matrices of the estimated factors is similar (see online Appendix B.2). It is worth noting that the four estimated factors are all positively correlated. However, the correlations tend to be quite substantially higher in absolute value with SPCA. This is essentially in line with the results of our Monte Carlo simulations with a block structure in the loadings and correlated factors. The two most correlated factors are the Northern American and European business cycles (59% with the SPCA estimates, versus 34% with the PCA quartimin estimates).

INSERT FIGURE 2 ABOUT HERE.

These interpretable factors provide new lens to study different issues. For example, like most studies of international business cycles with factor models, we can use the factors in a variance decomposition to measure the relative importance of each regional business cycle in driving national business cycles. Table 5 provides an illustration with the factors estimated by SPCA (see online Appendix B.2 for the results with the factors estimated by PCA varimax or PCA quartimin). Several interesting facts stand out. As commonly reported in this literature, the developing and emerging economies tend to be less driven by international business cycles. Among the European economies, the countries which are part of the European Union tend to be more synchronized with the European business cycle. Most of the literature studying international business cycles with factor models imposes the interpretation of each factor, and that each country is influenced by a single regional business cycle, corresponding to its geographic region. Our methods relax these restrictions, which proves to be relevant. Several economies appear to be mostly associated with the regional business cycle of another geographic region than their own, or to be greatly influenced by several regional business cycles. As seen before with the loadings, Japan is mostly associated with the European business cycle, while several Latin American economies are greatly or even predominantly driven by the Northern American business cycle (especially Costa Rica, El Salvador, and Honduras). Mexico is more influenced by the Latin American business cycle than by the Northern American business cycle, and the United Kingdom is more influenced by the Northern American business cycle than by the European business cycle. It is also worth noting that the North American economies seem to be more affected by the European business cycle than the European economies are affected by the Northern American business cycle.

INSERT TABLE 5 ABOUT HERE.

Interestingly, Francis, Owyang, and Savascin (2017) obtain very similar results concerning the interpretation of their four estimated factors and the variance decomposition, even though they use a different model (a Bayesian sparse factor model with endogenous clustering) and a slightly different dataset (their prior requires some covariate series in addition to the annual real GDP growth rates of the 60 countries considered). More precisely, they construct their model with one global factor and three cluster factors, without imposing the composition of the clusters. Although the first factor is supposed to be global, it is mainly related to the Developed Asian countries (see their Table 4). It explains more than half of the variance of the real GDP growth rate for Indonesia, Korea, Malaysia, Singapore, and Thailand, and far less for the other countries (including Japan). The first cluster includes most of the countries in Europe, and Japan. The second cluster consists of the USA and former British Commonwealth countries (especially Canada and the United Kingdom). The third cluster comprises most of the countries in Latin America, and Mexico. This similarity with our results supports the validity of our respective approaches. Furthermore, we also extend and complement their results. For example, their model constrains each country to be influenced by only a single cluster factor (in addition to the global factor and the idiosyncratic factor), while our methods imply that each country can be influenced by several regional business cycles. Moreover, our techniques are very easy to implement, and have a very low computational time, even SPCA here (about 0.01 seconds for the rotations, and about 1 minute for SPCA, including the tuning of the hyperparameters).

4.2 US economy

The so-called Stock-Watson dataset (Stock and Watson, 2002a,b) is considered the reference when a large dataset of US macroeconomic series is required, in particular to implement factor models. Its composition and time coverage have been updated several times. The latest modification is the publicly available Federal Reserve Economic Data Monthly Database (FRED-MD) designed by McCracken and Ng (2016), which is now the classic benchmark dataset for macroeconomic factor models. The monthly updates and revisions are taken care of by the data specialists at the St. Louis Fed[14] using mostly the time series from the Federal Reserve Economic Data (FRED), the St. Louis Fed's main publicly available economic database. The resulting vintage databases are all available on its website.[15] The variables are classified in eight groups: 1) output and income, 2) labor market, 3) housing, 4) consumption, orders and inventories, 5) money and credit, 6) interest rates, spreads, and exchange rates, 7) prices, and 8) stock market. We use the 2018:9 vintage of FRED-MD: 128 variables, covering the 1959:1-2018:7 time span. To obtain a balanced panel, the analysis is restricted to 1960:1-2018:4, with 123 variables (see online Appendix C.1 for further details). The series are stationarized using the transformations recommended by McCracken and Ng (2016), then standardized.

The six main Bai and Ng (2002) information criteria select between eight and ten factors. We choose to include eight factors in the model, like McCracken and Ng (2016) and Kristensen (2017), among others. For $r = 8$, the grid search of the SPCA hyperparameters yields $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.6)$.

With PCA, the estimated factors are difficult to interpret (the estimated factor loadings are presented in online Appendix C.2). Only four can be given an economic interpretation. Factor 1 can be interpreted as a real economic activity factor (although the interest rates also have a quite large loading). Factor 4 can be interpreted as an interest rates factor (although the housing, forex, and stock market variables also seem quite important). Factor 7 is obviously a stock market factor. Factor 8 is a money and credit factor. Some variables outside the money and credit group have quite large loadings in absolute value, but they actually correspond to determinants of the demand for money and credit: average hourly earnings in the goods-producing, construction, and manufacturing sectors (the three largest loadings in absolute value

[14] https://files.stlouisfed.org/files/htdocs/fred-databases/fredmdchanges.pdf

[15] https://research.stlouisfed.org/econ/mccracken/fred-databases/

in the labor market group), real personal consumption expenditures, retail and food services sales (the two largest loadings in absolute value in the consumption, orders and inventories group). The remaining four estimated factors are impossible to interpret.

Conversely, the loadings estimated by rotated PCA or SPCA systematically offer a straightforward economic interpretation of the factors: output, prices, spreads, interest rates, housing, labor, stock market, and money and credit. More precisely, the three methods lead to the same pattern of estimated loadings (Figure 3). We also notice that the loadings of the interest rates, stock market, and money and credit factors estimated with these methods are very close to the loadings of their PCA estimated counterparts (especially for the stock market, and money and credit factors). It should be stressed that the eight interpretable factors do not simply consist of recovering the eight groups of variables of the dataset, or some subgroups of them. For example, there is neither a consumption, orders and inventories factor, nor a forex factor. Besides, the prices, labor, and money and credit factors appear to have very heterogeneous influences on the variables from the prices, labor, and money and credit groups. Moreover, the money and credit factor involves several variables outside the money and credit group.

Despite the high level of sparsity in the loadings estimated by SPCA (70.3%), the loss of explained variance is absolutely marginal: the factors estimated by PCA (rotated or not) account for 48.0% of the data variance, versus 47.7% for the factors estimated by SPCA.

INSERT FIGURE 3 ABOUT HERE.

Since PCA varimax, PCA quartimin, and SPCA lead to almost identical estimated loadings, they also lead to almost identical estimated factors (see online Appendix C.2). The dynamics of the estimated factors are consistent with economic history. For example, the housing factor recovers the slow-building of the last housing bubble, its progressive burst starting in early 2006, and the slow recovery of the housing sector starting in early 2009. The estimated factors are strongly correlated with some representative observed variables. For example, the correlation between the SPCA estimated output factor and the growth rate of the industrial production is 96%. These correlations tend to be higher with the SPCA estimates than with the PCA quartimin estimates, and higher with the PCA quartimin estimates than with the PCA varimax estimates. The correlation matrices of the estimated factors are extremely similar between PCA quartimin and SPCA, more than in the previous empirical application (see online Appendix C.2). The correlations tend to be only slightly higher in absolute value with

SPCA. The most correlated factors are the three real economic activity factors, namely output, labor, and housing. For example the output factor and the labor factor have a correlation of 61% with the SPCA estimates, versus 41% with the PCA quartimin estimates.

Beyeler and Kaufmann (2021) obtain very similar interpretations for the eight US macroeconomic factors of their baseline model, even though they use a different model (a Bayesian sparse FAVAR), and another database (FRED-QD, a quarterly frequency companion to the monthly frequency FRED-MD, with a different composition). This supports the validity of our respective approaches.

5 Conclusions

With the usual estimation methods of factor models, the estimated factors are notoriously difficult to interpret, unless their interpretation is imposed via restrictions. This paper considers different approaches for identifying the factor structure and interpreting the factors without imposing their interpretation: sparse PCA (a machine learning technique) and factor rotations (originating from statistics and psychometrics).

We prove that sparse PCA consistently estimates the factor space if the ℓ_1 penalty parameter is $O(1/\sqrt{N})$, where N is the number of series. Monte Carlo simulations show that our exploratory methods accurately estimate the factor structure, even in small samples. They slightly outperform PCA in recovering the correct factor structure if that structure is dense, and vastly if that structure is sparse or near-sparse. The quartimin rotation and sparse PCA (with the SPCA algorithm) appear to be the preferable methods for recovering the correct factor structure (at least among the approaches considered here), but the varimax rotation yields very similar results in many cases. We apply our methods to two standard large datasets about international business cycles and the US economy. For each empirical application, they identify the same factor structure, offering a clear economic interpretation of the estimated factors. These interpretations are very close to the results of Francis, Owyang, and Savascin (2017) and Beyeler and Kaufmann (2021), who use Bayesian methods.

It is quite remarkable that such long-established statistical methods as factor rotations, and state-of-the-art machine learning and Bayesian methods, lead to extremely similar results for two different applications. This suggests that standard estimators like PCA, while computationally convenient, usually recover a quite unnatural factor structure. Whenever any exploratory method allowing for a sparse structure in the loadings is used, it yields the same factor structure. These exploratory methods provide new lens to study different issues with factor models. In particular, they can justify or complement approaches which impose the factor structure a priori, and can also be useful for applications in which factor interpretation is usually overlooked.

References

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019). Inference in group factor models with an application to mixed-frequency data. *Econometrica 87*(4), 1267–1305.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*(1), 18–29.

Bergmeir, C., R. J. Hyndman, and B. Koo (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis 120*, 70–83.

Beyeler, S. and S. Kaufmann (2021). Reduced-form factor augmented VAR—Exploiting sparsity to include meaningful factors. *Journal of Applied Econometrics 36*(7), 989–1012.

Carroll, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika 18*(1), 23–38.

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.

Cepni, O., I. E. Güney, and N. R. Swanson (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting 35*(2), 555–572.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1305–1324.

Costello, A. B. and J. W. Osborne (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation 10*(7), 1–9.

Croux, C. and P. Exterkate (2011). Sparse and robust factor modelling. Tinbergen Institute Discussion Papers 11-122/4, Tinbergen Institute.

d'Aspremont, A., F. Bach, and L. El Ghaoui (2008). Optimal solutions for sparse principal component analysis. *Journal of Machine Learning Research 9*, 1269–1294.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods 4*(3), 272.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–82.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The generalized dynamic factor model consistency and rates. *Journal of Econometrics 119*(2), 231–255.

Francis, N., M. T. Owyang, and O. Savascin (2017). An endogenously clustered factor approach to international business cycles. *Journal of Applied Econometrics 32*(7), 1261–1276.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Gürkaynak, R. S., B. Sack, and E. Swanson (2005). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *International Journal of Central Banking 1*(1), 55–93.

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika 66*(2), 289–306.

Jennrich, R. I. (2002). A simple general method for oblique rotation. *Psychometrika 67*(1), 7–19.

Jennrich, R. I. and P. Sampson (1966). Rotation for simple loadings. *Psychometrika 31*(3), 313–323.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer, New York.

Jolliffe, I. T., N. T. Trendafilov, and M. Uddin (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics 12*(3), 531–547.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika 23*(3), 187–200.

Kaufmann, S. and C. Schumacher (2017). Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics 32*(6), 1123–1144.

Kaufmann, S. and C. Schumacher (2019). Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification. *Journal of Econometrics 210*(1), 116–134.

Kim, H. H. and N. R. Swanson (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting 34*(2), 339–354.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country-specific factors. *American Economic Review 93*(4), 1216–1239.

Kristensen, J. T. (2017). Diffusion indexes with sparse loadings. *Journal of Business & Economic Statistics 35*(3), 434–451.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Miranda-Agrippino, S. and H. Rey (2020). US monetary policy and the Global Financial Cycle. *Review of Economic Studies 87*(6), 2754–2776.

Pelger, M. and R. Xiong (2020). Interpretable sparse proximate factors for large dimensions.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

Schneeweiss, H. and H. Mathes (1995). Factor analysis and principal components. *Journal of Multivariate Analysis 55*(1), 105–124.

Shen, H. and J. Z. Huang (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis 99*(6), 1015–1034.

Smeekes, S. and E. Wijler (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting 34*(3), 408–430.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Trendafilov, N. T. (2014). From simple structure to sparse components: a review. *Computational Statistics 29*(3-4), 431–454.

Uematsu, Y. and T. Yamagata (2021a). Estimation of sparsity-induced weak factor models.

Uematsu, Y. and T. Yamagata (2021b). Inference in sparsity-induced weak factor models.

West, M. (2003). Bayesian factor regression models in the "large p, small n" paradigm. In *Bayesian Statistics 7*, pp. 723–732. Oxford University Press.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(2), 301–320.

Zou, H., T. Hastie, and R. Tibshirani (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics 15*(2), 265–286.

A Variance decomposition method

Let $R^2_{A,B}$ denote the R^2 of the multivariate regression of the N_A variables in the $T \times N_A$ matrix A onto the N_B variables in the $T \times N_B$ matrix B:

$$R^2_{A,B} = \frac{\|P_B A\|_F^2}{\|A\|_F^2}, \tag{32}$$

where $P_B = B(B'B)^{-1}B'$ is the projection matrix onto the space spanned by the columns of B.

- The percentage of data variance explained by the r estimated factors is $R^2_{X,\widehat{F}}$. For PCA, several other equivalent computations can be used, for example $\sum_{j=1}^{r} \widehat{d}_j / \sum_{j=1}^{N} \widehat{d}_j$, where $\widehat{d}_j$ is the jth largest eigenvalue of the data covariance matrix $X'X/T$.

- The percentage of the variance of the ith variable explained by the r estimated factors (the *commonality* of the ith variable) is $R^2_{X_i,\widehat{F}}$, where X_i denotes the ith column of X.

- The percentage of the variance of the ith variable explained by the jth estimated factor is $R^2_{X_i,\widehat{F}_j}$, where $\widehat{F}_j$ denotes the jth column of $\widehat{F}$. It also corresponds to the squared correlation between the ith variable and the jth estimated factor. Of course, when all the estimated factors are uncorrelated (as with PCA and PCA varimax), $R^2_{X_i,\widehat{F}} = \sum_{j=1}^{r} R^2_{X_i,\widehat{F}_j}$.

- When the estimated factors are correlated (as allowed by PCA quartimin and SPCA), it can be useful to adjust the above measure. The percentage of the variance of the ith variable explained by the jth estimated factor, controlling for the influence of the other estimated factors on the jth estimated factor, is $R^2_{X_i,\widehat{F}} - R^2_{X_i,\widehat{F}_{-j}}$, where $\widehat{F}_{-j}$ denotes $\widehat{F}$ except its jth column. This can be represented by a Venn diagram. It also corresponds to the squared semi-partial correlation between the ith variable and the jth estimated factor, controlling for the influence of the other estimated factors on the jth estimated factor. Of course, when all the estimated factors are uncorrelated (as with PCA and PCA varimax), this is simply the percentage of the variance of the ith variable explained by the jth estimated factor.

It is worth noting that $R^2_{X,\widehat{F}}$ and $R^2_{X_i,\widehat{F}}$ are invariant to a factor rotation (orthogonal or oblique). Indeed, given that $\widehat{F}_{rotated} = \widehat{F}(R')^{-1}$, it is trivial to show that $P_{\widehat{F}_{rotated}} = P_{\widehat{F}}$.

B Figures and tables

Table 1: Simulation results with a dense structure in the loadings and uncorrelated factors.

Estimation method	ϕ_e	τ	RMSE	MAE	$\widehat{\rho}$	$R^2_{F,\widehat{F}}$
PCA	0	0	0.65	0.52	(0,0,0)	0.90
PCA varimax	0	0	0.63	0.51	(0,0,0)	0.90
PCA quartimin	0	0	0.63	0.51	(0,0,0)	0.90
SPCA	0	0	0.66	0.53	(0,-0.01,0.01)	0.89
PCA	0.5	0	0.66	0.53	(0,0,0)	0.91
PCA varimax	0.5	0	0.65	0.52	(0,0,0)	0.91
PCA quartimin	0.5	0	0.65	0.52	(0,0,0)	0.91
SPCA	0.5	0	0.66	0.53	(0,0,0)	0.91
PCA	0	0.5	0.66	0.53	(0,0,0)	0.88
PCA varimax	0	0.5	0.65	0.52	(0,0,0)	0.88
PCA quartimin	0	0.5	0.65	0.52	(0,0,0)	0.88
SPCA	0	0.5	0.67	0.54	(0,0,0)	0.88
PCA	0.5	0.5	0.67	0.54	(0,0,0)	0.90
PCA varimax	0.5	0.5	0.65	0.53	(0,0,0)	0.90
PCA quartimin	0.5	0.5	0.65	0.53	(0,0,0)	0.90
SPCA	0.5	0.5	0.67	0.54	(0,0,0)	0.90

Note. The parameters ϕ_e and τ control the serial correlation and cross-correlation in the idiosyncratic components, respectively. With these simulations, the factors are uncorrelated, so $\rho = (0, 0, 0)$.

Table 2: Simulation results with a dense structure in the loadings and correlated factors.

Estimation method	ϕ_e	τ	RMSE	MAE	$\widehat{\rho}$	$R^2_{F,\widehat{F}}$
PCA	0	0	0.72	0.58	(0,0,0)	0.89
PCA varimax	0	0	0.67	0.54	(0,0,0)	0.89
PCA quartimin	0	0	0.65	0.52	(-0.09,0.04,0.01)	0.89
SPCA	0	0	0.70	0.56	(-0.13,0.07,0.02)	0.89
PCA	0.5	0	0.73	0.59	(0,0,0)	0.91
PCA varimax	0.5	0	0.68	0.55	(0,0,0)	0.91
PCA quartimin	0.5	0	0.67	0.54	(-0.09,0.04,0.01)	0.91
SPCA	0.5	0	0.71	0.57	(-0.09,0.05,0.01)	0.91
PCA	0	0.5	0.73	0.59	(0,0,0)	0.88
PCA varimax	0	0.5	0.68	0.55	(0,0,0)	0.88
PCA quartimin	0	0.5	0.67	0.54	(-0.09,0.04,0.01)	0.88
SPCA	0	0.5	0.71	0.57	(-0.12,0.05,0.01)	0.88
PCA	0.5	0.5	0.74	0.59	(0,0,0)	0.90
PCA varimax	0.5	0.5	0.69	0.56	(0,0,0)	0.90
PCA quartimin	0.5	0.5	0.68	0.55	(-0.08,0.04,0.01)	0.90
SPCA	0.5	0.5	0.72	0.58	(-0.08,0.04,0.01)	0.90

Note. The parameters ϕ_e and τ control the serial correlation and cross-correlation in the idiosyncratic components, respectively. With these simulations, the correlations among the factors are such that $\rho = (-0.4, 0.2, 0)$.

Table 3: Simulation results with a block structure in the loadings and uncorrelated factors.

Estimation method	σ_l^2	σ_s^2	ϕ_e	τ	RMSE	MAE	$\widehat{\rho}$	$R^2_{F,\widehat{F}}$
PCA	0.2	0	0	0	0.67	0.57	(0,0,0)	0.92
PCA varimax	0.2	0	0	0	0.23	0.18	(0,0,0)	0.92
PCA quartimin	0.2	0	0	0	0.22	0.17	(0,0,0)	0.92
SPCA	0.2	0	0	0	0.15	0.07	(0,0,0)	0.93
PCA	0.2	0.2	0	0	0.65	0.55	(0,0,0)	0.92
PCA varimax	0.2	0.2	0	0	0.24	0.20	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0	0	0.24	0.19	(0,0,0)	0.92
SPCA	0.2	0.2	0	0	0.30	0.24	(0,0,0)	0.92
PCA	0.5	0.2	0	0	0.66	0.54	(0,0,0)	0.91
PCA varimax	0.5	0.2	0	0	0.29	0.23	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0	0	0.28	0.22	(0,0,0)	0.91
SPCA	0.5	0.2	0	0	0.32	0.26	(0,0,0)	0.91
PCA	0.2	0	0.5	0	0.69	0.58	(0,0,0)	0.94
PCA varimax	0.2	0	0.5	0	0.25	0.20	(0,0,0)	0.94
PCA quartimin	0.2	0	0.5	0	0.24	0.18	(0,0,0)	0.94
SPCA	0.2	0	0.5	0	0.16	0.08	(0,0,0)	0.94
PCA	0.2	0.2	0.5	0	0.67	0.56	(0,0,0)	0.93
PCA varimax	0.2	0.2	0.5	0	0.26	0.21	(0,0,0)	0.93
PCA quartimin	0.2	0.2	0.5	0	0.25	0.20	(0,0,0)	0.93
SPCA	0.2	0.2	0.5	0	0.32	0.26	(0,0,0)	0.93
PCA	0.5	0.2	0.5	0	0.67	0.56	(0,0,0)	0.93
PCA varimax	0.5	0.2	0.5	0	0.31	0.24	(0,0,0)	0.93
PCA quartimin	0.5	0.2	0.5	0	0.31	0.24	(0,0,0)	0.93
SPCA	0.5	0.2	0.5	0	0.35	0.28	(0,0,0)	0.93
PCA	0.2	0	0	0.5	0.67	0.58	(0,0,0)	0.85
PCA varimax	0.2	0	0	0.5	0.24	0.19	(0,0,0)	0.85
PCA quartimin	0.2	0	0	0.5	0.22	0.17	(0.01,0,0)	0.85
SPCA	0.2	0	0	0.5	0.15	0.07	(0.01,0,0)	0.85
PCA	0.2	0.2	0	0.5	0.66	0.55	(0,0,0)	0.85
PCA varimax	0.2	0.2	0	0.5	0.25	0.20	(0,0,0)	0.85
PCA quartimin	0.2	0.2	0	0.5	0.24	0.19	(0,0,0)	0.85
SPCA	0.2	0.2	0	0.5	0.30	0.24	(0,0,0)	0.85
PCA	0.5	0.2	0	0.5	0.67	0.55	(0,0,0)	0.84
PCA varimax	0.5	0.2	0	0.5	0.30	0.23	(0,0,0)	0.84
PCA quartimin	0.5	0.2	0	0.5	0.29	0.23	(0,0,0)	0.84
SPCA	0.5	0.2	0	0.5	0.33	0.26	(0,0,0.01)	0.84
PCA	0.2	0	0.5	0.5	0.69	0.58	(0,0,0)	0.88
PCA varimax	0.2	0	0.5	0.5	0.25	0.20	(0,0,0)	0.88
PCA quartimin	0.2	0	0.5	0.5	0.24	0.18	(0.01,0,0)	0.88
SPCA	0.2	0	0.5	0.5	0.17	0.08	(0.01,0,0)	0.88
PCA	0.2	0.2	0.5	0.5	0.67	0.56	(0,0,0)	0.88
PCA varimax	0.2	0.2	0.5	0.5	0.26	0.21	(0,0,0)	0.88
PCA quartimin	0.2	0.2	0.5	0.5	0.25	0.20	(0,0,0)	0.88
SPCA	0.2	0.2	0.5	0.5	0.33	0.26	(0,0,0)	0.88
PCA	0.5	0.2	0.5	0.5	0.68	0.56	(0,0,0)	0.87
PCA varimax	0.5	0.2	0.5	0.5	0.32	0.25	(0,0,0)	0.87
PCA quartimin	0.5	0.2	0.5	0.5	0.31	0.24	(0,0,0)	0.87
SPCA	0.5	0.2	0.5	0.5	0.35	0.28	(0,0,0)	0.87

Note. The variances of the large and small loadings are denoted σ_l^2 and σ_s^2, respectively. The parameters ϕ_e and τ control the serial correlation and cross-correlation in the idiosyncratic components, respectively. With these simulations, the factors are uncorrelated, so $\rho = (0, 0, 0)$.

Table 4: Simulation results with a block structure in the loadings and correlated factors.

Estimation method	σ_l^2	σ_s^2	ϕ_e	τ	RMSE	MAE	$\widehat{\rho}$	$R^2_{F,\widehat{F}}$
PCA	0.2	0	0	0	0.72	0.61	(0,0,0)	0.92
PCA varimax	0.2	0	0	0	0.29	0.23	(0,0,0)	0.92
PCA quartimin	0.2	0	0	0	0.23	0.18	(-0.32,0.16,0.01)	0.92
SPCA	0.2	0	0	0	0.16	0.07	(-0.38,0.18,0)	0.93
PCA	0.2	0.2	0	0	0.71	0.60	(0,0,0)	0.92
PCA varimax	0.2	0.2	0	0	0.30	0.24	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0	0	0.26	0.21	(-0.23,0.11,0.01)	0.92
SPCA	0.2	0.2	0	0	0.32	0.26	(-0.35,0.18,0.01)	0.92
PCA	0.5	0.2	0	0	0.72	0.60	(0,0,0)	0.91
PCA varimax	0.5	0.2	0	0	0.34	0.27	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0	0	0.30	0.24	(-0.23,0.11,0.01)	0.91
SPCA	0.5	0.2	0	0	0.34	0.27	(-0.36,0.18,0.01)	0.91
PCA	0.2	0	0.5	0	0.73	0.62	(0,0,0)	0.94
PCA varimax	0.2	0	0.5	0	0.31	0.25	(0,0,0)	0.94
PCA quartimin	0.2	0	0.5	0	0.25	0.20	(-0.32,0.16,0.01)	0.94
SPCA	0.2	0	0.5	0	0.18	0.09	(-0.38,0.19,0)	0.94
PCA	0.2	0.2	0.5	0	0.72	0.60	(0,0,0)	0.93
PCA varimax	0.2	0.2	0.5	0	0.31	0.25	(0,0,0)	0.93
PCA quartimin	0.2	0.2	0.5	0	0.28	0.22	(-0.22,0.11,0.01)	0.93
SPCA	0.2	0.2	0.5	0	0.36	0.29	(-0.31,0.16,0.02)	0.93
PCA	0.5	0.2	0.5	0	0.73	0.61	(0,0,0)	0.93
PCA varimax	0.5	0.2	0.5	0	0.36	0.28	(0,0,0)	0.93
PCA quartimin	0.5	0.2	0.5	0	0.33	0.26	(-0.22,0.11,0.01)	0.93
SPCA	0.5	0.2	0.5	0	0.38	0.30	(-0.32,0.16,0.02)	0.93
PCA	0.2	0	0	0.5	0.72	0.61	(0,0,0)	0.85
PCA varimax	0.2	0	0	0.5	0.28	0.23	(0,0,0)	0.85
PCA quartimin	0.2	0	0	0.5	0.24	0.18	(-0.30,0.15,0.01)	0.85
SPCA	0.2	0	0	0.5	0.16	0.08	(-0.34,0.17,0.01)	0.86
PCA	0.2	0.2	0	0.5	0.72	0.60	(0,0,0)	0.85
PCA varimax	0.2	0.2	0	0.5	0.29	0.23	(0,0,0)	0.85
PCA quartimin	0.2	0.2	0	0.5	0.26	0.20	(-0.21,0.11,0.01)	0.85
SPCA	0.2	0.2	0	0.5	0.32	0.26	(-0.33,0.16,0.02)	0.85
PCA	0.5	0.2	0	0.5	0.73	0.60	(0,0,0)	0.85
PCA varimax	0.5	0.2	0	0.5	0.34	0.27	(0,0,0)	0.85
PCA quartimin	0.5	0.2	0	0.5	0.31	0.24	(-0.22,0.11,0.01)	0.85
SPCA	0.5	0.2	0	0.5	0.35	0.28	(-0.33,0.17,0.01)	0.84
PCA	0.2	0	0.5	0.5	0.73	0.62	(0,0,0)	0.88
PCA varimax	0.2	0	0.5	0.5	0.30	0.24	(0,0,0)	0.88
PCA quartimin	0.2	0	0.5	0.5	0.25	0.19	(-0.30,0.15,0.01)	0.88
SPCA	0.2	0	0.5	0.5	0.18	0.09	(-0.35,0.18,0)	0.88
PCA	0.2	0.2	0.5	0.5	0.72	0.61	(0,0,0)	0.88
PCA varimax	0.2	0.2	0.5	0.5	0.30	0.24	(0,0,0)	0.88
PCA quartimin	0.2	0.2	0.5	0.5	0.27	0.22	(-0.21,0.11,0.01)	0.88
SPCA	0.2	0.2	0.5	0.5	0.36	0.29	(-0.29,0.15,0.02)	0.88
PCA	0.5	0.2	0.5	0.5	0.74	0.61	(0,0,0)	0.88
PCA varimax	0.5	0.2	0.5	0.5	0.36	0.28	(0,0,0)	0.88
PCA quartimin	0.5	0.2	0.5	0.5	0.33	0.26	(-0.21,0.11,0.01)	0.88
SPCA	0.5	0.2	0.5	0.5	0.39	0.31	(-0.29,0.16,0.02)	0.87

Note. The variances of the large and small loadings are denoted σ_l^2 and σ_s^2, respectively. The parameters ϕ_e and τ control the serial correlation and cross-correlation in the idiosyncratic components, respectively. With these simulations, the correlations among the factors are such that $\rho = (-0.4, 0.2, 0)$.

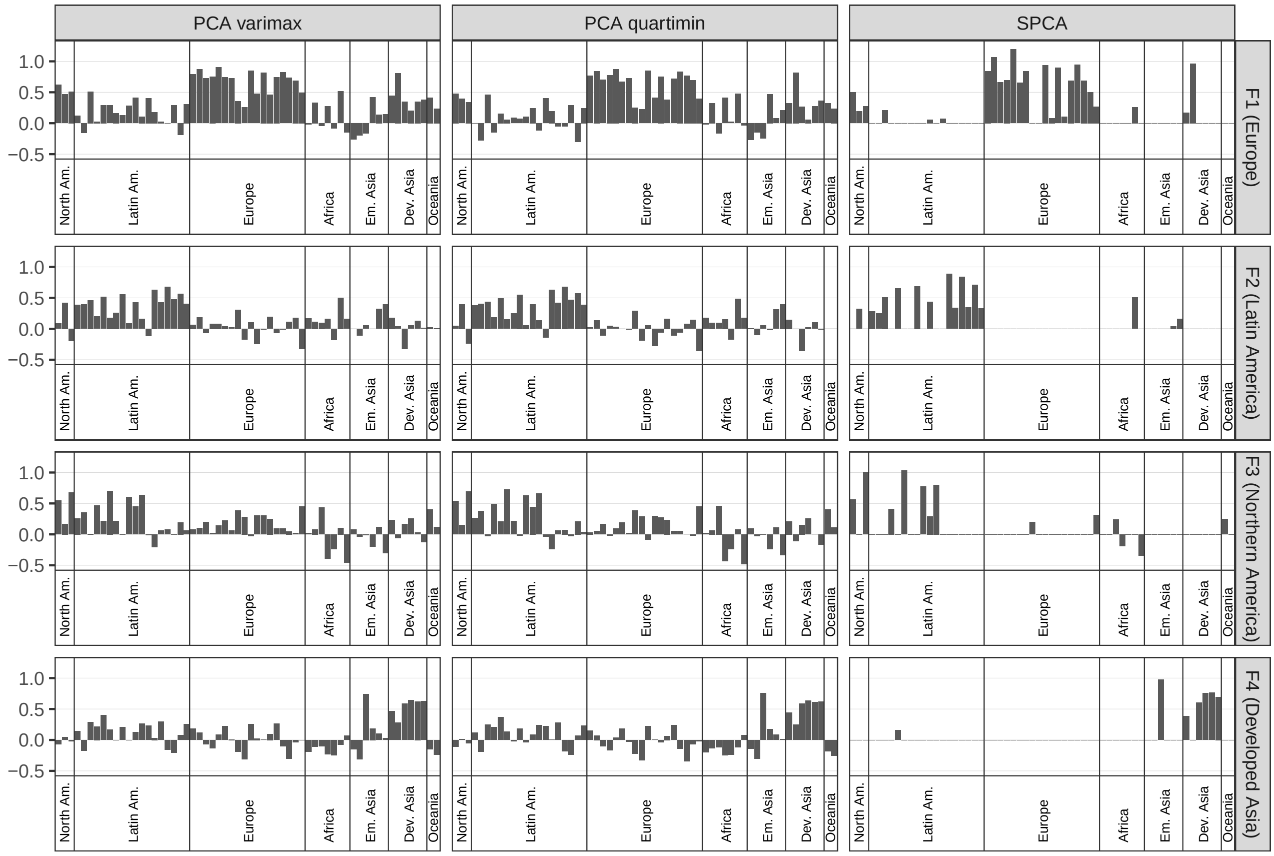


Figure 1: Factor loadings estimated with PCA varimax, PCA quartimin, and SPCA (international business cycles).

Note. The factors are labeled with their economic interpretation. Since F and Λ are identified up to a column sign change, for each factor, we impose the largest loading in absolute value to be positive. The proportions of nonzero loadings for the factors estimated with SPCA are respectively 0.42, 0.27, 0.22, 0.12.

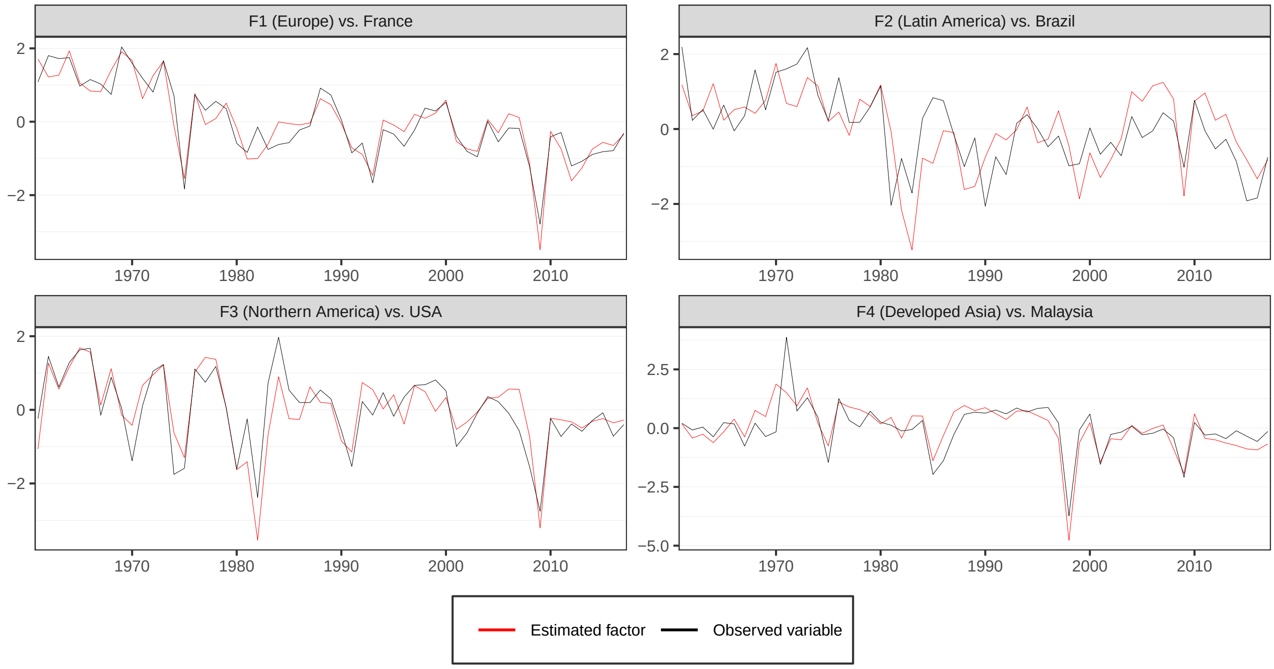


Figure 2: Factors estimated with SPCA, compared with the real GDP growth rates of some representative countries (international business cycles).

Note. The correlations between the factors estimated with SPCA and the real GDP growth rates of the representative countries are respectively 0.94, 0.66, 0.85, 0.82.

Table 5: Variance decomposition with the factors estimated by SPCA (international business cycles).

Country	F1 (Europe)		F2 (Latin Am.)		F3 (Northern Am.)		F4 (Dev. Asia)		Common.
	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	
Canada	49.9	9.3	12.4	0.0	61.1	20.5	4.6	0.6	71.1
Mexico	26.8	5.5	32.4	13.9	14.9	0.2	4.9	0.1	41.8
USA	35.1	3.7	0.7	7.8	72.1	42.8	4.0	0.0	81.6
Argentina	4.5	0.7	24.4	16.8	8.9	2.3	3.1	0.1	26.8
Bolivia	0.4	6.0	7.2	9.4	4.0	6.2	1.5	2.7	19.7
Brazil	32.1	7.2	44.2	20.4	10.5	0.7	15.4	1.7	56.2
Chile	2.3	4.6	6.9	1.7	20.9	17.9	5.5	2.2	27.6
Colombia	15.5	0.1	49.2	27.7	14.9	0.9	23.0	7.5	58.9
Costa Rica	19.0	1.2	15.4	1.4	71.6	49.5	5.2	0.0	73.7
Dominican Republic	4.5	0.0	8.3	3.7	4.9	0.7	3.1	0.5	10.5
Ecuador	3.9	0.1	39.5	35.9	1.3	0.8	3.4	0.2	41.2
El Salvador	14.5	0.1	8.0	0.3	49.5	33.8	1.4	0.4	50.3
Guatemala	27.6	0.6	35.3	12.2	38.8	11.0	8.0	0.1	55.3
Honduras	6.5	6.1	10.1	1.3	49.8	43.1	5.9	1.1	56.6
Jamaica	14.3	7.0	0.7	1.2	4.6	0.0	9.7	3.8	18.9
Panama	3.6	0.3	39.4	40.8	0.0	4.4	0.2	1.1	45.9
Paraguay	1.3	1.4	18.6	15.1	1.3	0.0	8.1	4.2	23.4
Peru	0.3	2.7	43.5	51.4	0.8	0.1	0.1	2.2	52.5
Trinidad and Tobago	8.9	3.1	24.4	19.1	2.0	0.9	0.0	3.7	30.0
Uruguay	0.8	12.0	28.3	36.9	1.5	1.5	0.0	0.1	42.5
Venezuela	12.9	0.9	33.0	20.2	6.8	0.0	6.2	0.3	34.9
Austria	68.0	41.8	12.3	0.0	16.9	0.9	15.0	0.7	69.6
Belgium	83.6	47.3	23.5	1.1	22.2	1.0	16.1	0.2	85.9
Denmark	54.8	34.0	5.5	0.8	22.5	0.4	3.1	1.1	57.2
Finland	55.6	43.7	8.7	0.0	10.5	1.9	3.7	0.8	58.4
France	87.8	52.8	14.2	0.1	28.1	0.0	11.8	0.0	87.9
Germany	65.4	30.2	12.4	0.0	27.5	0.3	16.6	1.2	67.0
Greece	50.9	32.7	6.9	0.2	16.8	0.0	4.2	0.4	51.6
Iceland	17.9	2.6	15.0	4.4	20.4	5.1	0.4	2.2	30.1
Ireland	6.1	4.4	0.5	4.0	6.2	2.7	0.3	1.9	14.4
Italy	75.8	45.3	16.0	0.1	17.0	1.8	21.2	2.0	79.6
Luxembourg	25.4	15.2	0.0	5.5	13.2	1.3	3.8	0.1	31.5
Netherlands	75.5	39.1	10.7	0.4	37.7	1.9	6.0	0.8	78.4
Norway	26.8	6.0	14.8	2.3	19.3	2.0	6.4	0.1	32.1
Portugal	61.8	38.3	6.4	1.0	17.8	0.2	14.1	1.0	63.9
Spain	68.1	52.4	5.7	1.1	18.4	0.2	3.0	1.7	71.8
Sweden	50.7	40.6	10.8	0.5	11.4	1.0	0.3	5.7	57.6
Switzerland	50.1	34.5	13.1	0.9	11.6	1.0	3.0	1.2	52.8
United Kingdom	26.8	7.3	0.0	9.9	36.5	17.0	3.6	0.0	50.5
Cameroon	0.2	0.7	0.5	1.0	0.1	0.3	0.3	0.3	1.9
Côte d'Ivoire	10.8	5.4	5.0	1.1	5.0	0.1	0.0	1.6	13.3
Kenya	0.3	1.4	0.5	0.0	8.3	9.9	0.0	0.3	10.7
Morocco	3.8	17.4	1.3	1.7	5.6	19.0	2.6	6.1	29.7
Senegal	2.7	0.6	5.0	1.3	7.1	3.3	6.8	3.5	12.6
South Africa	30.8	8.8	37.5	17.4	12.2	0.1	4.7	0.3	48.3
Zimbabwe	3.7	0.2	0.0	1.9	22.7	21.7	0.2	0.2	26.1
Bangladesh	6.9	6.6	1.0	0.0	0.2	1.7	1.8	0.2	8.8
India	6.1	0.7	4.8	1.0	3.7	0.1	5.5	1.8	9.6
Indonesia	0.9	9.6	1.5	0.6	0.6	0.6	54.5	67.4	71.9
Pakistan	14.2	11.7	1.4	0.2	0.8	2.6	6.9	2.0	18.9
Philippines	4.0	0.1	10.0	5.2	1.7	0.1	7.7	3.4	14.0
Sri Lanka	1.2	0.6	14.4	15.9	1.1	7.0	0.8	0.0	21.8
Hong Kong SAR	27.2	2.7	13.9	0.7	18.2	1.2	38.6	18.9	50.4
Japan	66.2	41.2	11.4	0.0	13.7	1.9	20.7	2.8	70.8
Korea	14.3	2.6	0.2	10.8	7.2	0.5	51.1	42.7	63.3
Malaysia	8.8	0.3	4.7	0.0	8.0	0.9	67.0	56.7	68.0
Singapore	16.0	1.0	9.3	0.4	5.4	0.2	60.1	43.7	62.0
Thailand	15.0	2.7	4.1	0.1	3.2	0.7	48.1	35.2	50.8
Australia	20.1	3.5	4.2	0.0	32.0	14.5	0.0	3.3	37.5
New Zealand	4.4	2.9	0.3	0.1	1.9	0.1	0.1	0.2	4.8

Note. This table shows for each country the percentage of the variance of the real GDP growth rate explained by each estimated factor, and the commonality. Since the factors estimated by SPCA are correlated, we also report an adjusted measure of the percentage of the variance explained by each estimated factor, controlling for the influence of the other estimated factors. The variance decomposition method is detailed in Appendix A.

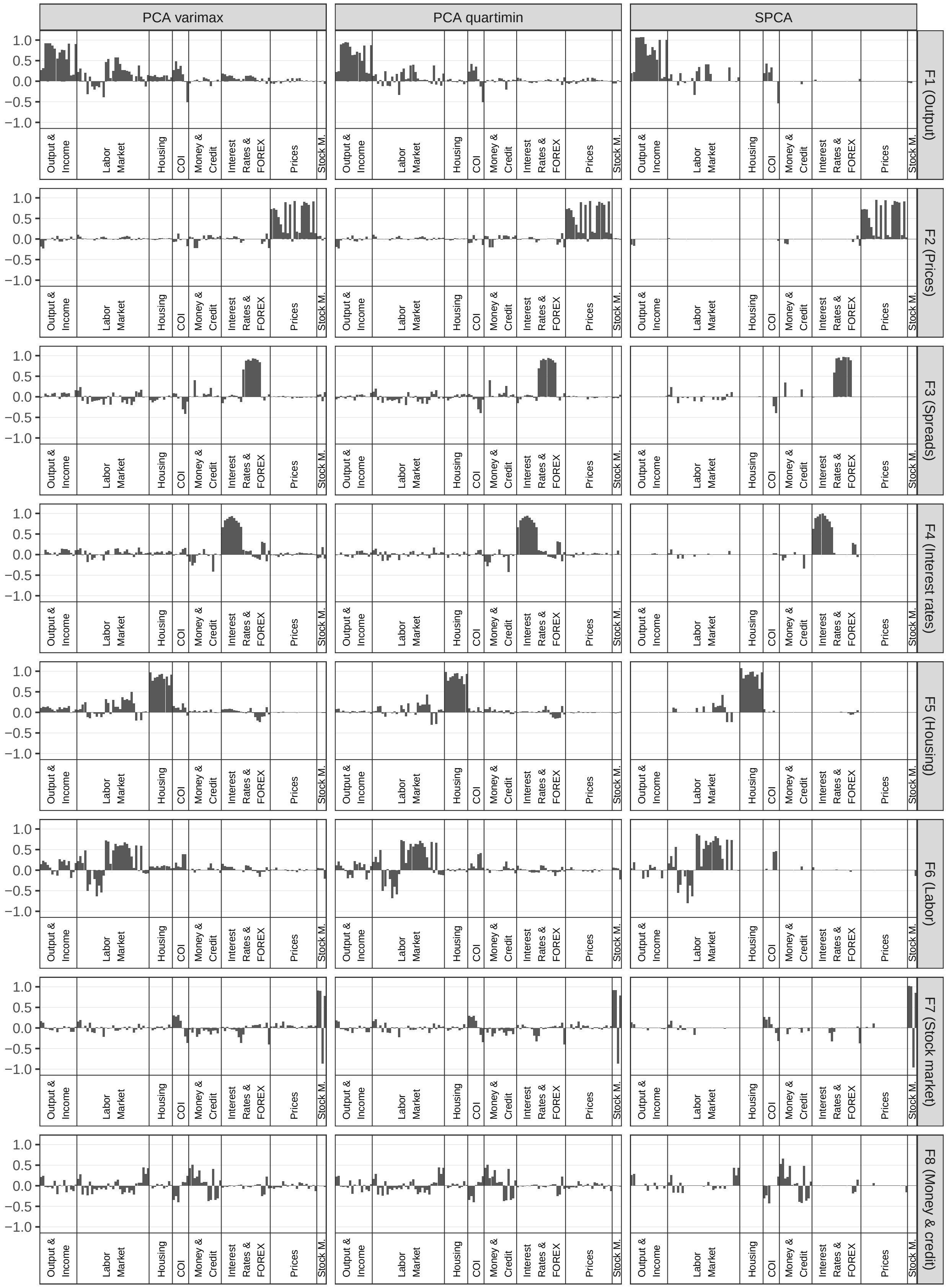


Figure 3: Factor loadings estimated with PCA varimax, PCA quartimin, and SPCA (FRED-MD).

Note. The factors are labeled with their economic interpretation. Since F and Λ are identified up to a column sign change, for each factor, we impose the largest loading in absolute value to be positive. The proportions of nonzero loadings for the factors estimated with SPCA are respectively 0.36, 0.23, 0.25, 0.27, 0.26, 0.35, 0.29, 0.37.

IDENTIFYING AND INTERPRETING THE FACTORS IN FACTOR MODELS VIA SPARSITY: DIFFERENT APPROACHES

Online Appendix

Thomas Despois and Catherine Doz

A Proof of Theorem 1

Preliminary results and notations

We use the same set of assumptions as in Doz, Giannone, and Reichlin (2011). We suppose that the normalization condition is the same as in Doz, Giannone, and Reichlin (2011), which is innocuous, since F_t is defined up to a rotation matrix. In particular, the true value of the loading matrix Λ, satisfies $\Lambda'\Lambda = D$, where D is a diagonal $r \times r$ matrix whose terms go to infinity linearly with N.[1]

PCA notations

We denote:

- $\widehat{d}_1 \geqslant \cdots \geqslant \widehat{d}_n$ the ordered eigenvalues of $\frac{X'X}{T}$
- $\widehat{p}_1, \ldots, \widehat{p}_n$ an associated orthonormal set of eigenvectors
- $\widehat{P} = (\widehat{p}_1, \ldots, \widehat{p}_r)$
- $\widehat{D} = \text{diag}(\widehat{d}_1, \ldots, \widehat{d}_r)$

Other notations

- We use the same notations as in the description of the SPCA algorithm.
- For any matrix A, we take $\|A\| = \sqrt{\sigma_{max}(A'A)}$, where $\sigma_{max}(A'A)$ is the maximum eigenvalue of $A'A$. For any symmetric positive matrix A, we then have $\|A\| = \sigma_{max}(A)$.
- For any matrix A, we say that $A = O_P(f(N))$ if $\|A\| = O_P(f(N))$.
- For any symmetric matrix A, we denote by $\sigma_{min}(A)$ the smallest eigenvalue of A.

[1] Here we drop the index 0 which Doz, Giannone, and Reichlin (2011) use for the true values of the parameters.

To prove our theorem, we use properties which have been obtained by Doz, Giannone, and Reichlin (2011) under their set of assumptions or which can be easily derived from properties which they have stated. In particular, we use the following properties.

Preliminary properties (PCA properties)

i) $\widehat{D} = O_P(N)$

ii) $\widehat{D}^{-1} = O_P\left(\frac{1}{N}\right)$

iii) $\left(\frac{X'X}{T}\right)^{-1} = O_P(1)$

Proof.

i) Doz, Giannone, and Reichlin (2011) have shown that (see Lemma 2 in their Appendix):

$$\frac{1}{N}\left(\widehat{D} - D\right) = O\left(\frac{1}{N}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

We can thus write: $\widehat{D} = D + O(1) + O_P\left(\frac{N}{\sqrt{T}}\right)$.
As all the terms of D tend to infinity linearly with N, we thus obtain $\widehat{D} = O_P(N)$.

ii) Doz, Giannone, and Reichlin (2011) have also shown in the same lemma that:

$$N\left(\widehat{D}^{-1} - D^{-1}\right) = O\left(\frac{1}{N}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$$

We can thus write: $\widehat{D}^{-1} = D^{-1} + O\left(\frac{1}{N^2}\right) + O_P\left(\frac{1}{N\sqrt{T}}\right)$.
As all the terms of D tend to infinity linearly with N, we thus obtain $\widehat{D}^{-1} = O_P\left(\frac{1}{N}\right)$.

iii) If we denote $\widehat{\Sigma}_e = \frac{X'X}{T} - \widehat{\Lambda}\widehat{\Lambda}'$, Doz, Giannone, and Reichlin (2011) have shown that:[2]

$$\widehat{\sigma}_{ij,e} - \sigma_{ij,e} = O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \quad \text{(A1)}$$

and that the result is uniform w.r.t. (i, j).

By Assumption (CR4) we know that $\sigma_{min}\left(\Sigma_e\right) = c > 0$.

By the Weyl theorem, we know that:

$$\sigma_{min}\left(\widehat{\Sigma}_e\right) \geqslant \sigma_{min}\left(\widehat{\Sigma}_e - \Sigma_e\right) + \sigma_{min}\left(\Sigma_e\right) \quad \text{(A2)}$$

[2] Σ_e is denoted Ψ_0 in Doz, Giannone, and Reichlin (2011) and $\widehat{\Sigma}_e$ is denoted $\widehat{\Psi}$.

Further:

$$\begin{aligned}
\sigma_{min}\left(\widehat{\Sigma}_e - \Sigma_e\right) &= \min_{x'x=1} x'\left(\widehat{\Sigma}_e - \Sigma_e\right)x \\
&= \min_{x'x=1} \sum_{i,j} x_i x_j \left(\widehat{\sigma}_{ij,e} - \sigma_{ij,e}\right) \\
&\leqslant \min_{x'x=1} \sum_{i,j} |x_i||x_j| \left|\widehat{\sigma}_{ij,e} - \sigma_{ij,e}\right| \\
&\leqslant \max_{i,j} \left|\widehat{\sigma}_{ij,e} - \sigma_{ij,e}\right| \min_{x'x=1} \sum_{i,j} |x_i||x_j| \\
&\leqslant \max_{i,j} \left|\widehat{\sigma}_{ij,e} - \sigma_{ij,e}\right| \text{ by the Cauchy-Schwarz inequality} \\
&= O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

Using Equation (A2), we thus have: $\sigma_{min}\left(\widehat{\Sigma}_e\right) \geqslant c + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$.

As $\sigma_{min}\left(\frac{X'X}{T}\right) \geqslant \sigma_{min}\left(\widehat{\Sigma}_e\right)$, we also get $\sigma_{min}\left(\frac{X'X}{T}\right) \geqslant c + O_P\left(\frac{1}{\sqrt{N}}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$ so that $\sigma_{max}\left(\frac{X'X}{T}^{-1}\right) \leqslant \frac{1}{c} + O_P\left(\frac{1}{\sqrt{T}}\right)$ and $\left(\frac{X'X}{T}\right)^{-1} = O_P(1)$. □

Before proving our theorem, we now provide a few intermediary results concerning the algorithm, which are stated in the following properties.

Proposition 1. *At iteration k, $B^{(k)}$ can be written as*

$$B^{(k)} = \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{X'X}{T}A^{(k)} - \frac{\kappa_1}{2}Z^{(k)}\right) \tag{A3}$$

where $Z^{(k)}$ is a $N \times r$ matrix whose general term satisfies $|z_{ij}^{(k)}| \leqslant 1$.

Proof. We know from Zou, Hastie, and Tibshirani (2006) (see the result recalled in our presentation of step 1 of the algorithm) that $B^{(k)}$ is obtained as $B^{(k)} = \left(b_1^{(k)}, \ldots, b_r^{(k)}\right)$ where

$$b_j^{(k)} = \arg\min_b \,(a_j^{(k)} - b)'\frac{X'X}{T}(a_j^{(k)} - b) + \kappa_1\|b\|_1 + \kappa_2\|b\|_2^2, \text{ for } j = 1, \ldots, r. \tag{A4}$$

For a given k, let us denote $f_j(b) = (a_j^{(k)} - b)'\frac{X'X}{T}(a_j^{(k)} - b) + \kappa_1\|b\|_1 + \kappa_2\|b\|_2^2$ and $b_j^{(k)} = \arg\min_b f_j(b)$. Standard convex analysis results (see Rockafellar, 1970, Part VI, Section 27) imply that 0 belongs to the subdifferential of f_j in $b_j^{(k)}$, denoted as $\partial f_j(b_j^{(k)})$.

Let us then calculate $\partial f_j(b)$ the subdifferential of f_j in b, for any b.

If we denote $g_j(b) = (a_j^{(k)} - b)'\frac{X'X}{T}(a_j^{(k)} - b) + \kappa_2\|b\|_2^2$, g_j is differentiable, with gradient

$$\nabla g_j(b) = -2\frac{X'X}{T}a_j^{(k)} + 2\frac{X'X}{T}b + 2\kappa_2 b.$$

For $b = (b_1, \ldots, b_n)'$ let us denote $h(b) = \|b\|_1 = \sum_{i=1}^n |b_i|$. If $z = (z_1, \ldots, z_n)'$, we know (see Rockafellar, 1970, Part V, Section 23) that z belongs to $\partial h(b)$, the subdifferential of h in b, if and only if:

$$\begin{cases} z_i = 1 & \text{if } b_i > 0 \\ z_i = -1 & \text{if } b_i < 0 \\ z_i \in [-1, 1] & \text{if } b_i = 0 \end{cases} \tag{A5}$$

Thus, the elements of $\partial f_j(b)$ can be written as:

$$-2\frac{X'X}{T}a_j^{(k)} + 2\frac{X'X}{T}b + 2\kappa_2 b + \kappa_1 z \tag{A6}$$

where z satisfies Equations (A5).

We then obtain that $b_j^{(k)} = \arg\min_b f_j(b)$ if and only if there exists a z_j such that:

$$-2\frac{X'X}{T}a_j^{(k)} + 2\frac{X'X}{T}b_j^{(k)} + 2\kappa_2 b_j^{(k)} + \kappa_1 z_j = 0 \tag{A7}$$

and

$$\begin{cases} z_{ij} = 1 & \text{if } b_{ij}^{(k)} > 0 \\ z_{ij} = -1 & \text{if } b_{ij}^{(k)} < 0 \\ z_{ij} \in [-1, 1] & \text{if } b_{ij}^{(k)} = 0 \end{cases} \tag{A8}$$

Solving Equation (A7) for $b_j^{(k)}$, we then get:

$$b_j^{(k)} = \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{X'X}{T}a_j^{(k)} - \frac{\kappa_1}{2}z_j\right) \tag{A9}$$

If we denote $B^{(k)} = \left(b_1^{(k)}, \dots, b_r^{(k)}\right)$, $A^{(k)} = \left(a_1^{(k)}, \dots, a_r^{(k)}\right)$, and $Z^{(k)} = (z_1, \dots, z_r)$, the result of Proposition 1 follows. □

Proposition 1 gives the general form of $B^{(k)}$ as a function of $A^{(k)}$ at step 1 of iteration k. We now study the general form of $A^{(k)}$, and state the following proposition.

Proposition 2. *Under standard Assumptions A1-A3 and Assumptions CR1-CR4 in Doz, Giannone, and Reichlin (2011), if $\kappa_1 = O\left(\frac{1}{\sqrt{N}}\right)$, the following result holds: for any k, $A^{(k)}$ can be written as $A^{(k)} = \widehat{P}M_k + O_P\left(\frac{1}{N}\right)$ where M_k is a $r \times r$ matrix which satisfies $M_k'M_k = I_r + O_P\left(\frac{1}{N}\right)$.*

The proof of Proposition 2 will be done by mathematical induction, and relies on the following lemma.

Lemma 1. *If we suppose that $A^{(k)} = \widehat{P}M_k + O_P\left(\frac{1}{N}\right)$ where M_k is a $r \times r$ matrix which satisfies $M_k'M_k = I_r + O_P\left(\frac{1}{N}\right)$, then:*

(i) $B^{(k)} = \widehat{P}M_k + O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}$

(ii) $\frac{X'X}{T}B^{(k)} = \widehat{P}\widehat{D}M_k + O_P(1)$

Proof of Lemma 1.

i) Using Proposition 1, we know from Equation (A3) that

$$B^{(k)} = \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \left(\frac{X'X}{T} A^{(k)} - \frac{\kappa_1}{2} Z^{(k)}\right)$$

with $Z^{(k)}$ is a $N \times r$ matrix whose general term satisfies $|z_{ij}^{(k)}| \leqslant 1$.

If $A^{(k)} = \widehat{P} M_k + O_P\left(\frac{1}{N}\right)$, this can be written as

$$B^{(k)} = \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \left(\widehat{P} M_k + O_P\left(\frac{1}{N}\right)\right) - \frac{\kappa_1}{2} \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)} \tag{A10}$$

or equivalently

$$B^{(k)} = \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \widehat{P} M_k + \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \times O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2} \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)} \tag{A11}$$

Let us now study the first two terms of (A11).

We know that the columns of $\widehat{P}$ are eigenvectors of $\frac{X'X}{T}$, associated to the eigenvalues contained in the diagonal matrix $\widehat{D}$. It follows that the columns of $\widehat{P}$ are also eigenvectors of $\frac{X'X}{T} + \kappa_2 I_N$ associated to the eigenvalues contained in the matrix $\widehat{D} + \kappa_2 I_r$, and eigenvectors of $(\frac{X'X}{T} + \kappa_2 I_N)^{-1}$ associated to the eigenvalues contained in the matrix $(\widehat{D} + \kappa_2 I_r)^{-1}$.

We thus get:

$$\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \widehat{P} M_k = \widehat{P}(\widehat{D} + \kappa_2 I_r)^{-1} \widehat{D} M_k \tag{A12}$$

Further, the diagonal terms of $(\widehat{D} + \kappa_2 I_r)^{-1} \widehat{D}$ are $\frac{\widehat{d}_i}{\widehat{d}_i + \kappa_2}$ for $i = 1, \ldots, r$, with

$$\frac{\widehat{d}_i}{\widehat{d}_i + \kappa_2} = \frac{1}{1 + \frac{\kappa_2}{\widehat{d}_i}} \sim 1 - \frac{\kappa_2}{\widehat{d}_i} = 1 + O_P\left(\frac{1}{N}\right) \tag{A13}$$

so that $(\widehat{D} + \kappa_2 I_r)^{-1} \widehat{D} = I_r + O_P\left(\frac{1}{N}\right)$.

As $\widehat{P}'\widehat{P} = I_r$, we have $\widehat{P} = O_P(1)$. As $M_k' M_k = I_r + O_P\left(\frac{1}{N}\right)$, we also have $M_k = O_P(1)$. We can then write:

$$\widehat{P}(\widehat{D} + \kappa_2 I_r)^{-1} \widehat{D} M_k = \widehat{P}(I_r + O_P\left(\frac{1}{N}\right)) M_k = \widehat{P} M_k + O_P\left(\frac{1}{N}\right) \tag{A14}$$

Using (A12), we thus get:

$$\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \widehat{P} M_k = \widehat{P} M_k + O_P\left(\frac{1}{N}\right). \tag{A15}$$

Turning to the second term of (A11), we notice that the eigenvalues of $\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T}$ are $\frac{\widehat{d}_i}{\widehat{d}_i + \kappa_2}, i = 1, \ldots, n$, which are all smaller than 1.

We thus get: $\left\| \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \frac{X'X}{T} \right\| \leqslant 1$, so that this second term is $O_P\left(\frac{1}{N}\right)$.

Part (i) of the lemma then follows.

ii) It follows from (i) that:

$$\frac{X'X}{T}B^{(k)} = \frac{X'X}{T}\left(\widehat{P}M_k + O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\right) \tag{A16}$$

$$= \widehat{P}\widehat{D}M_k + \frac{X'X}{T} \times O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2}\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)} \tag{A17}$$

As $\frac{X'X}{T} = O_P(N)$, we thus get:

$$\frac{X'X}{T}B^{(k)} = \widehat{P}\widehat{D}M_k + O_P(1) - \frac{\kappa_1}{2}\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)} \tag{A18}$$

Further:

- $\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}$ has the same eigenvalues than $\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\frac{X'X}{T}$ so that

$$\left\|\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\right\| = \left\|\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\frac{X'X}{T}\right\| \leqslant 1$$

- $\left\|Z^{(k)}\right\|^2 \leqslant \text{tr}\left(Z^{(k)\prime}Z^{(k)}\right) = \sum_{j=1}^{r}\sum_{i=1}^{N}\left(z_{ij}^{(k)}\right)^2 \leqslant rN$ so that $\left\|Z^{(k)}\right\| = O_P(\sqrt{N})$.

It then follows that:

$$\left\|\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\right\| \leqslant \left\|\frac{X'X}{T}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\right\| \left\|Z^{(k)}\right\| = O_p(\sqrt{N})$$

As $\kappa_1 = O\left(\frac{1}{\sqrt{N}}\right)$, part (ii) of the lemma follows. □

Proof of Proposition 2.

We prove Proposition 2 by mathematical induction.

For $k = 1$, we have $A^{(1)} = \widehat{P}$ and the result is true with $M_1 = I_r$.

If we suppose that the result is true for k, then Lemma 1 applies.

Besides, we know that $A^{(k+1)}$ is obtained from step 2 of iteration k in the following way:

if the SVD of $\frac{X'X}{T}B^{(k)}$ is written as $\frac{X'X}{T}B^{(k)} = U_k\Delta_k V_k'$, then $A^{(k+1)} = U_k V_k'$.

Further, the SVD of $\frac{X'X}{T}B$ is obtained in the following way:

- V_k is a $r \times r$ matrix whose columns are normalized eigenvectors associated to the r eigenvalues of $B^{(k)\prime}\left(\frac{X'X}{T}\right)^2 B^{(k)}$, so that $V_k'V_k = I_r$,

- Δ_k^2 is the diagonal $r \times r$ matrix whose diagonal terms are the associated eigenvalues
- U_k is the $N \times r$ matrix defined by $U_k = \frac{X'X}{T} B^{(k)} V_k \Delta_k^{-1}$, so that $U_k' U_k = I_r$ and

$$\frac{X'X}{T} B^{(k)} = U_k \Delta_k V_k'$$

From result (ii) of Lemma 1, we know that:

$$\frac{X'X}{T} B^{(k)} = \widehat{P}\widehat{D} M_k + O_P(1) \tag{A19}$$

Given the fact that $\widehat{P}'\widehat{P} = I_r$ and that $\widehat{D} = O_P(N)$, we first get

$$\begin{aligned} B^{(k)\prime} \left(\frac{X'X}{T}\right)^2 B^{(k)} &= (\widehat{P}\widehat{D} M_k + O_P(1))'(\widehat{P}\widehat{D} M_k + O_P(1)) \\ &= M_k' \widehat{D}^2 M_k + O_P(N) \end{aligned} \tag{A20}$$

so that

$$V_k \Delta_k^2 V_k' = M_k' \widehat{D}^2 M_k + O_P(N)$$

where V_k is a $r \times r$ matrix and $V_k' V_k = I_r$.

Thus Δ_k^2, the diagonal matrix which contains the eigenvalues of $B^{(k)\prime} \left(\frac{X'X}{T}\right)^2 B^{(k)}$, has terms which all diverge linearly with N^2, like the terms of $\widehat{D}^2$, and Δ_k has terms which all diverge linearly with N. We also get:

$$\begin{aligned} \Delta_k^2 &= V_k' M_k' \widehat{D}^2 M_k V_k + O_P(N) \\ \text{so that } I_r &= \Delta_k^{-1} \left(V_k' M_k' \widehat{D}^2 M_k V_k + O_P(N)\right) \Delta_k^{-1} \\ \text{and } I_r &= \Delta_k^{-1} V_k' M_k' \widehat{D}^2 M_k V_k \Delta_k^{-1} + O_P\left(\frac{1}{N}\right) \end{aligned} \tag{A21}$$

Now, we have seen that U_k is defined by:

$$U_k = \frac{X'X}{T} B^{(k)} V_k \Delta_k^{-1} \tag{A22}$$

As we know that $A^{(k+1)} = U_k V_k'$, we have $A^{(k+1)\prime} A^{(k+1)} = I_r$ by construction.

Using (A22), we can write $A^{(k+1)}$ as:

$$A^{(k+1)} = \frac{X'X}{T} B^{(k)} V_k \Delta_k^{-1} V_k'$$

Using Equation (A19), and the fact that $\Delta_k^{-1} = O_P\left(\frac{1}{N}\right)$, we then obtain:

$$\begin{aligned} A^{(k+1)} &= \left(\widehat{P}\widehat{D} M_k + O_P(1)\right) V_k \Delta_k^{-1} V_k' \\ &= \widehat{P}\widehat{D} M_k V_k \Delta_k^{-1} V_k' + O_P\left(\frac{1}{N}\right). \end{aligned}$$

If we denote $M_{k+1} = \widehat{D} M_k V_k \Delta_k^{-1} V_k'$, we have $A^{(k+1)} = \widehat{P} M_{k+1} + O_P\left(\frac{1}{N}\right)$.

As $A^{(k+1)\prime}A^{(k+1)} = I_r$, we get:

$$\left(\widehat{P}M_{k+1} + O_P\left(\frac{1}{N}\right)\right)'\left(\widehat{P}M_{k+1} + O_P\left(\frac{1}{N}\right)\right) = I_r \tag{A23}$$

As $\widehat{P}'\widehat{P} = I_r$ we know that $\widehat{P} = O_P(1)$.

Further, $M_{k+1} = \widehat{D}M_kV_k\Delta_k^{-1}V_k'$, with:

. $\widehat{D} = O_P(N)$

. $\Delta_k^{-1} = O_P\left(\frac{1}{N}\right)$

. $V_k'V_k = I_r$ so that $V_k = O_P(1)$

. $M_k'M_k = I_r + O_P\left(\frac{1}{N}\right)$ by assumption, so that $M_k = O_P(1)$.

We thus get: $M_{k+1} = O_P(1)$.

It then follows from Equation (A23) that

$$M_{k+1}'M_{k+1} + O_P\left(\frac{1}{N}\right) = I_r \tag{A24}$$

which completes the proof of Proposition 2. □

We are now ready to prove our consistency theorem for the estimated factors.

Proof of Theorem 1.

We suppose that the algorithm converges at iteration k and we construct the estimated loadings and factors as mentioned in the description of the algorithm. More precisely, we use the following notations: $\widetilde{B} = B^{(k)}$, $\widetilde{\Delta} = \operatorname{diag}\left(\widetilde{B}'\frac{X'X}{T}\widetilde{B}\right)$, and $\widetilde{F}_t = \widetilde{\Delta}^{-1/2}\widetilde{B}'x_t$.

Using results (i) and (ii) of Lemma 1, and using the fact that $\widehat{D} = O_P(N)$, $\widehat{P} = O_P(1)$, and $M_k = O_P(1)$, we get that:

$$\begin{aligned}
\widetilde{B}'\frac{X'X}{T}\widetilde{B} &= \left(\widehat{P}M_k + O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\right)'\left(\widehat{P}\widehat{D}M_k + O_P(1)\right) \\
&= M_k'\widehat{D}M_k + O_P(1) - \frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\widehat{P}\widehat{D}M_k - \frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \times O_P(1) \\
&= M_k'\widehat{D}M_k + O_P(1) - \frac{\kappa_1}{2}Z^{(k)\prime}\widehat{P}\left(\widehat{D} + \kappa_2 I_N\right)^{-1}\widehat{D}M_k - \frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \times O_P(1)
\end{aligned}$$

We have seen that $Z^{(k)} = O_P(\sqrt{N})$. As we assume that $\kappa_1 = O\left(\frac{1}{\sqrt{N}}\right)$, we get $\kappa_1 Z^{(k)} = O_P(1)$.

We also have seen that $\left(\widehat{D} + \kappa_2 I_N\right)^{-1}\widehat{D} \leqslant I_r$ so that $\left(\widehat{D} + \kappa_2 I_N\right)^{-1}\widehat{D} = O_P(1)$.
Further, $\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \leqslant \left(\frac{X'X}{T}\right)^{-1} = O_P(1)$. We thus get:

$$\widetilde{B}'\frac{X'X}{T}\widetilde{B} = M_k'\widehat{D}M_k + O_P(1)$$

If we denote $M_k = \left(m_1^{(k)}, \ldots, m_r^{(k)}\right)$, the diagonal terms of $M_k'\widehat{D}M_k$ are

$$m_i^{(k)\prime}\widehat{D}m_i^{(k)} = \sum_{j=1}^{r} \widehat{d}_j \left(m_{ij}^{(k)}\right)^2$$

As $M_k'M_k = I_r + O_P\left(\frac{1}{N}\right)$, we have $\sum_{j=1}^{r} \widehat{d}_j \left(m_{ij}^{(k)}\right)^2 \in [\widehat{d}_r + O_P(1), \widehat{d}_1 + O_P(1)]$ for $i = 1, \ldots, r$ so that all the terms of $\widetilde{\Delta}$ diverge linearly with N like the terms of $\widehat{D}$. We thus have:

$$\widetilde{\Delta}^{-1/2} = O_P\left(\frac{1}{\sqrt{N}}\right)$$

As $\widetilde{F}_t = \widetilde{\Delta}^{-1/2}\widetilde{B}'x_t$, we can write, using result (i) of Lemma 1:

$$\widetilde{F}_t = \widetilde{\Delta}^{-1/2}\left[M_k'\widehat{P}' + O_P\left(\frac{1}{N}\right) - \frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\right]x_t$$

and we know that the $O_P\left(\frac{1}{N}\right)$ term comes from Equation (A13) and is the same one for any t.

Using the fact that $\widehat{F}_t = \widehat{D}^{-1/2}\widehat{P}'x_t$, we then get:

$$\widetilde{F}_t = \widetilde{\Delta}^{-1/2}M_k'\widehat{D}^{1/2}\widehat{F}_t + O_P\left(\frac{1}{N}\right) - \widetilde{\Delta}^{-1/2}\frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} x_t \qquad \text{(A25)}$$

If we denote $M = \widetilde{\Delta}^{-1/2}M_k'\widehat{D}^{1/2}$, it is clear that M is invertible and that $M = O_P(1)$.

Let us now show that $\frac{1}{T}\sum_{t=1}^{T} \|\widetilde{F}_t - M\widehat{F}_t\|_2^2 = O_P\left(\frac{1}{N}\right)$.

We can write:

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T} \|\widetilde{F}_t - M\widehat{F}_t\|_2^2 \\ =&\frac{1}{T}\sum_{t=1}^{T}\left(O_P\left(\frac{1}{N}\right) - \widetilde{\Delta}^{-1/2}\frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} x_t\right)'\left(O_P\left(\frac{1}{N}\right) - \widetilde{\Delta}^{-1/2}\frac{\kappa_1}{2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} x_t\right)\end{aligned}$$

We have seen that:

. the $O_P\left(\frac{1}{N}\right)$ term does not depend on t

. $Z^{(k)} = O_P(\sqrt{N})$

. $\widetilde{\Delta}^{-1/2} = O_P\left(\frac{1}{\sqrt{N}}\right)$

. $\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \leqslant \left(\frac{X'X}{T}\right)^{-1} = O_P(1)$

We can decompose $\frac{1}{T}\sum_{t=1}^{T} \|\widetilde{F}_t - M\widehat{F}_t\|_2^2$ into three terms that we study separately.

- The first term has the form $\frac{1}{T}\sum_{t=1}^{T} O_P\left(\frac{1}{N}\right) \times O_P\left(\frac{1}{N}\right)$ and the $O_P\left(\frac{1}{N}\right)$ term does not depend on t, so this term is $O_P\left(\frac{1}{N^2}\right)$.

- The second term is

$$2 \times O_P\left(\frac{1}{N}\right) \times \frac{1}{T}\sum_{t=1}^{T} \widetilde{\Delta}^{-1/2}\frac{\kappa_1}{2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right) x_t \tag{A26}$$

which can be also written as

$$\kappa_1 \times O_P\left(\frac{1}{N}\right) \times \widetilde{\Delta}^{-1/2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)\frac{1}{T}\sum_{t=1}^{T} x_t \tag{A27}$$

Since $Z^{(k)} = O_P(\sqrt{N}), \widetilde{\Delta}^{-1/2} = O_P\left(\frac{1}{\sqrt{N}}\right), \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} = O_P(1)$, we know that:

$$\kappa_1 \times O_P\left(\frac{1}{N}\right) \times \widetilde{\Delta}^{-1/2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right) = O_P\left(\frac{1}{N\sqrt{N}}\right).$$

Now $\frac{1}{T}\sum_{t=1}^{T} x_t = \Lambda \frac{1}{T}\sum_{t=1}^{T} F_t + \frac{1}{T}\sum_{t=1}^{T} e_t$, with:

  - $\Lambda = O(\sqrt{N})$
  - $\frac{1}{T}\sum_{t=1}^{T} F_t = O_P\left(\frac{1}{\sqrt{T}}\right)$ since (F_t) is a stationary process
  - $\frac{1}{T}\sum_{t=1}^{T} e_t = O_P\left(\frac{\sqrt{N}}{\sqrt{T}}\right)$ by Assumption (CR3) since

$$E\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t\right\|^2\right) = E\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\mathbb{Z}} e_t' e_{t-h}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\mathbb{Z}} \text{tr } E\left(e_t e_{t-h}'\right) \leqslant N \sum_{h\in\mathbb{Z}} \|E(e_t e_{t-h}')\| = O(N).$$

We thus get that $\frac{1}{T}\sum_{t=1}^{T} x_t = O_P\left(\frac{\sqrt{N}}{\sqrt{T}}\right)$ and, using (A27), we obtain that (A26) is $O_P\left(\frac{1}{N\sqrt{T}}\right)$.

- The third term is

$$\frac{\kappa_1^2}{4}\frac{1}{T}\sum_{t=1}^{T} x_t'\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\widetilde{\Delta}^{-1} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} x_t \tag{A28}$$

which can also be written as

$$\begin{aligned}
&\frac{\kappa_1^2}{4}\frac{1}{T}\sum_{t=1}^{T} \text{tr}\left[\widetilde{\Delta}^{-1/2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} x_t x_t'\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\widetilde{\Delta}^{-1/2}\right]\\
=&\frac{\kappa_1^2}{4}\text{tr}\left[\widetilde{\Delta}^{-1/2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\widetilde{\Delta}^{-1/2}\right]\\
=&\frac{\kappa_1^2}{4}\text{tr}\left[\widetilde{\Delta}^{-1/2} Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{X'X}{T}\right)\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} Z^{(k)}\widetilde{\Delta}^{-1/2}\right]
\end{aligned} \tag{A29}$$

Now, we can write

$$\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1/2}\left(\frac{X'X}{T}\right)\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1/2} \leqslant I_N$$

so that

$$\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{X'X}{T}\right)\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \leqslant \left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1} \leqslant \left(\frac{X'X}{T}\right)^{-1} = O_P(1)$$

Since $\kappa_1 = O\left(\frac{1}{\sqrt{N}}\right)$, $Z^{(k)\prime}Z^{(k)} = O_P(N)$, and $\widetilde{\Delta}^{-1/2} = O_P\left(\frac{1}{\sqrt{N}}\right)$, we then get

$$\frac{\kappa_1^2}{4}\widetilde{\Delta}^{-1/2}Z^{(k)\prime}\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}\left(\frac{X'X}{T}\right)\left(\frac{X'X}{T} + \kappa_2 I_N\right)^{-1}Z^{(k)}\widetilde{\Delta}^{-1/2} = O_P\left(\frac{1}{N}\right)$$

As this matrix is $r \times r$, its trace is also $O_P\left(\frac{1}{N}\right)$, so that (A29) is $O_P\left(\frac{1}{N}\right)$.

It then follows that the summation of the three terms of our decomposition is also $O_P\left(\frac{1}{N}\right)$ i.e. that

$$\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{F}_t - M\widehat{F}_t\|_2^2 = O_P\left(\frac{1}{N}\right).$$

□

B Supplementary material for the application to international business cycles

B.1 Data

The regional definitions follow Kose, Otrok, and Whiteman (2003). The Developing Asia region is renamed Emerging Asia.

- North America: Canada, Mexico, USA.
- Latin America: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Panama, Paraguay, Peru, Trinidad and Tobago, Uruguay, Venezuela.
- Europe: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom.
- Africa: Cameroon, Côte d'Ivoire, Kenya, Morocco, Senegal, South Africa, Zimbabwe.
- Emerging Asia: Bangladesh, India, Indonesia, Pakistan, Philippines, Sri Lanka.
- Developed Asia: Hong Kong SAR, Japan, Korea, Malaysia, Singapore, Thailand.
- Oceania: Australia, New Zealand.

B.2 Additional results

Table B1: Selected number of factors (international business cycles).

Criterion	IC_{p1}	IC_{p2}	IC_{p3}	PC_{p1}	PC_{p2}	PC_{p3}
r^*	1	1	4	2	1	4

Note. Number of factors selected by each of the six main Bai and Ng (2002) information criteria, allowing for at most four factors.

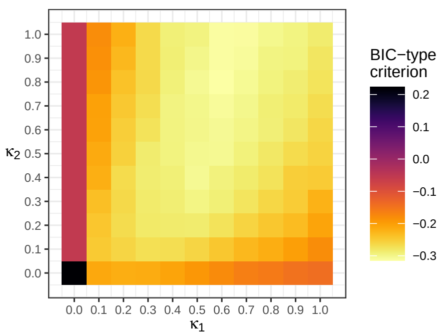


Figure B1: Grid search of the SPCA hyperparameters κ_1 and κ_2 (international business cycles).

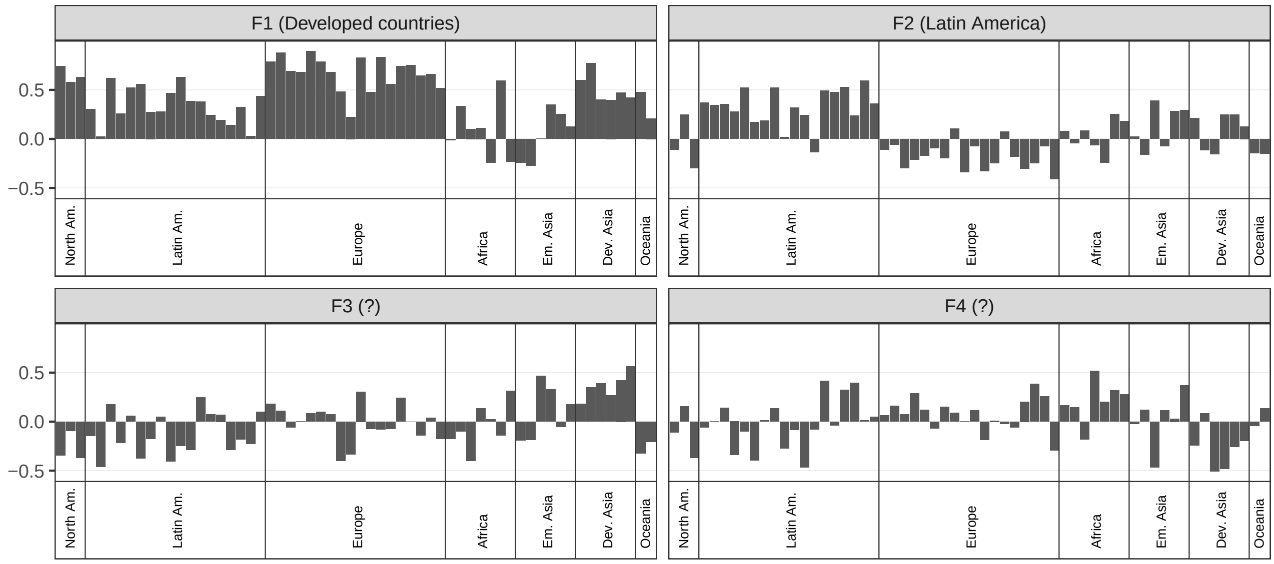


Figure B2: Factor loadings estimated with PCA (international business cycles).

Note. The interpretable factors are labeled with their economic interpretation, and the uninterpretable factors with a question mark. Since F and Λ are identified up to a column sign change, for each factor, we impose the largest loading in absolute value to be positive.

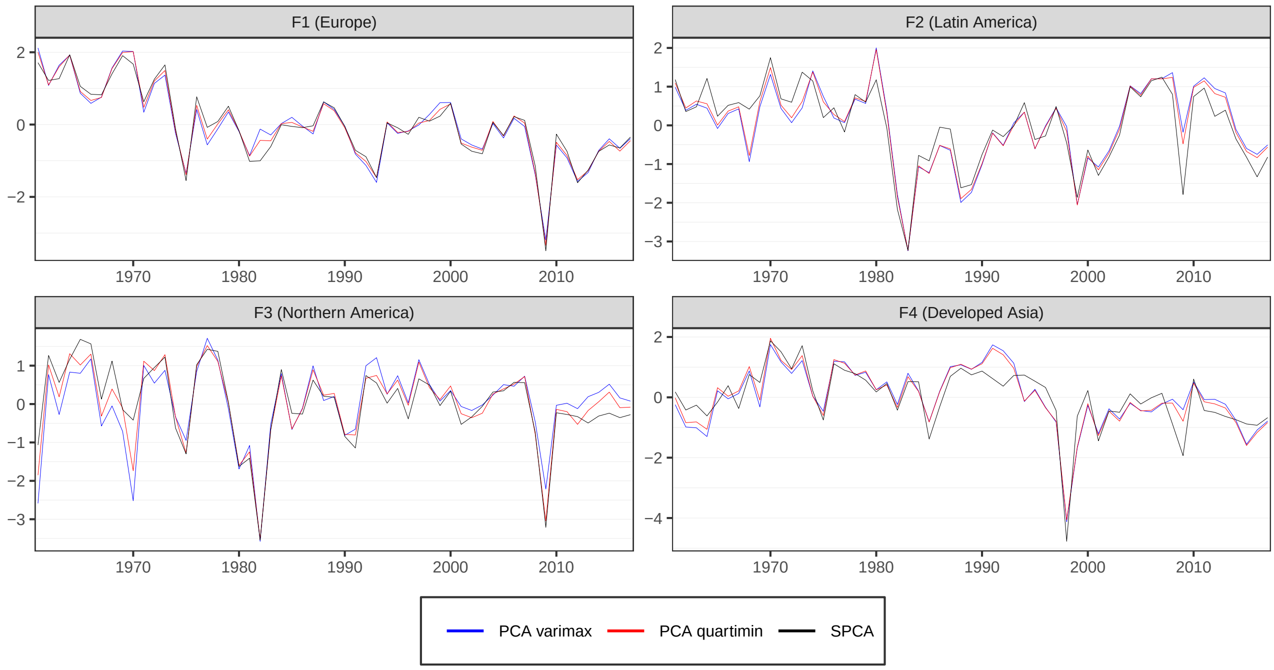


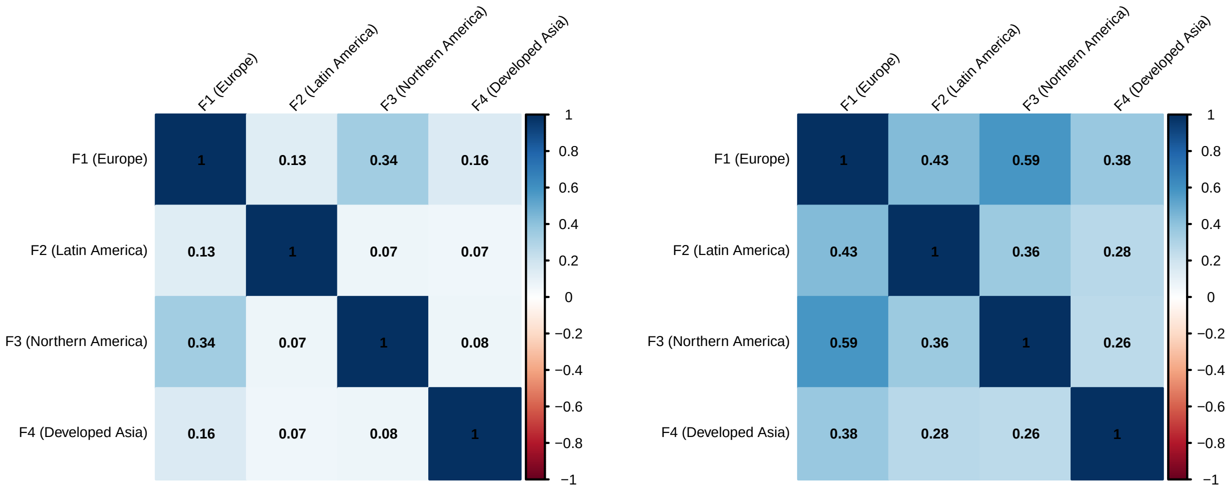
Figure B3: Factors estimated with PCA varimax, PCA quartimin, and SPCA (international business cycles).

Table B2: Correlations between the estimated factors (international business cycles).

Factors	PCA varimax vs. PCA quartimin	PCA varimax vs. SPCA	PCA quartimin vs. SPCA
F1 (Europe)	1.00	0.97	0.99
F2 (Latin America)	1.00	0.90	0.93
F3 (Northern America)	0.96	0.85	0.95
F4 (Developed Asia)	0.99	0.87	0.90

Table B3: Correlations between the estimated factors and the real GDP growth rates of some representative countries (international business cycles).

Pairs	PCA varimax	PCA quartimin	SPCA
F1 (Europe) vs. France	0.91	0.92	0.94
F2 (Latin America) vs. Brazil	0.47	0.52	0.66
F3 (Northern America) vs. USA	0.69	0.79	0.85
F4 (Developed Asia) vs. Malaysia	0.65	0.67	0.82



(a) PCA quartimin

(b) SPCA

Figure B4: Correlation matrices of the estimated factors (international business cycles).

Note. The correlation matrices of the factors estimated with PCA and PCA varimax are not displayed since they are simply the identity matrix, by construction.

Table B4: Variance decomposition with the factors estimated by PCA varimax (international business cycles).

Country	F1 (Europe)	F2 (Latin Am.)	F3 (Northern Am.)	F4 (Dev. Asia)	Common.
Canada	39.3	0.8	30.2	0.5	70.7
Mexico	22.1	18.2	3.0	0.2	43.5
USA	26.5	4.0	46.9	0.0	77.5
Argentina	1.5	15.4	6.9	2.0	25.7
Bolivia	2.6	15.8	12.5	3.0	33.9
Brazil	26.5	21.9	0.0	8.3	56.7
Chile	0.1	4.0	22.3	4.9	31.2
Colombia	8.7	26.8	4.8	16.3	56.6
Costa Rica	8.6	3.3	50.4	2.7	65.0
Dominican Republic	2.8	7.0	4.8	0.0	14.5
Ecuador	1.7	31.7	0.0	4.2	37.6
El Salvador	8.1	0.7	37.4	0.0	46.2
Guatemala	17.1	18.5	20.5	1.7	57.8
Honduras	1.1	2.6	40.9	6.9	51.5
Jamaica	16.7	1.4	0.0	5.5	23.6
Panama	3.3	40.8	4.3	0.1	48.6
Paraguay	0.1	18.3	0.4	8.8	27.6
Peru	0.0	46.4	0.6	2.5	49.5
Trinidad and Tobago	8.5	22.9	0.0	4.4	35.8
Uruguay	3.8	32.7	3.9	0.7	41.0
Venezuela	9.8	16.9	0.4	6.7	33.7
Austria	63.4	0.5	0.6	3.5	67.9
Belgium	77.4	3.5	1.0	1.4	83.3
Denmark	53.4	0.4	4.0	0.5	58.3
Finland	57.7	0.7	0.0	1.7	60.1
France	83.1	0.7	2.0	0.7	86.5
Germany	55.8	0.2	5.0	5.0	66.0
Greece	53.5	0.1	0.4	0.0	53.9
Iceland	13.2	9.8	15.1	3.6	41.7
Ireland	6.7	3.2	8.2	9.9	27.9
Italy	73.0	1.1	0.1	6.9	81.0
Luxembourg	22.8	6.1	9.4	0.1	38.3
Netherlands	68.1	0.0	9.3	0.0	77.4
Norway	21.7	3.7	6.4	1.0	32.8
Portugal	56.8	0.4	0.9	7.2	65.4
Spain	69.2	0.0	0.9	1.1	71.2
Sweden	55.0	1.3	0.2	9.4	66.0
Switzerland	48.5	3.2	0.0	0.1	51.8
United Kingdom	24.9	10.8	20.5	0.0	56.2
Cameroon	0.1	2.9	0.0	3.7	6.7
Côte d'Ivoire	11.4	1.4	0.7	1.2	14.7
Kenya	0.2	1.0	19.1	1.1	21.3
Morocco	7.6	2.5	15.4	5.4	31.0
Senegal	0.8	3.5	5.7	6.2	16.1
South Africa	27.2	25.9	1.1	0.6	54.9
Zimbabwe	2.2	2.5	21.6	0.5	26.8
Bangladesh	6.9	0.0	0.6	2.3	9.8
India	3.8	1.3	0.2	9.9	15.2
Indonesia	2.7	0.3	0.0	56.4	59.5
Pakistan	17.7	0.0	4.0	3.6	25.3
Philippines	2.0	10.5	1.3	1.1	15.0
Sri Lanka	2.1	16.1	9.2	0.1	27.5
Hong Kong SAR	20.0	3.3	5.3	22.0	50.6
Japan	67.0	0.1	0.3	8.1	75.5
Korea	12.1	11.1	2.8	35.0	60.9
Malaysia	4.2	0.3	6.7	42.0	53.1
Singapore	12.3	1.7	0.1	39.6	53.7
Thailand	14.7	0.0	1.7	39.7	56.0
Australia	17.4	0.1	16.2	2.4	36.1
New Zealand	5.8	0.0	1.4	5.8	13.0

Note. This table shows for each country the percentage of the variance of the real GDP growth rate explained by each estimated factor, and the commonality. The variance decomposition method is detailed in Appendix A of the paper.

Table B5: Variance decomposition with the factors estimated by PCA quartimin (international business cycles).

Country	F1 (Europe)		F2 (Latin Am.)		F3 (Northern Am.)		F4 (Dev. Asia)		Common.
	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	Unadj.	Adj.	
Canada	43.2	20.0	2.1	0.2	49.6	26.2	0.0	1.2	70.7
Mexico	25.3	13.5	21.8	15.8	9.9	2.0	1.3	0.0	43.5
USA	29.3	10.1	2.2	5.8	63.0	43.1	0.2	0.2	77.5
Argentina	2.6	0.0	16.8	14.3	9.2	6.2	2.8	1.4	25.7
Bolivia	1.7	6.8	14.6	16.0	9.1	12.9	3.1	3.4	33.9
Brazil	30.4	18.3	26.8	19.1	3.3	0.1	12.4	6.2	56.7
Chile	0.5	2.0	4.7	3.4	22.5	21.7	5.4	4.1	31.2
Colombia	12.4	2.1	31.0	24.0	10.6	3.8	20.1	13.5	56.6
Costa Rica	12.2	0.3	4.9	2.2	60.1	47.5	4.4	1.7	65.0
Dominican Republic	3.7	0.6	7.8	6.3	7.1	4.2	0.1	0.0	14.5
Ecuador	2.8	0.4	33.4	30.5	0.4	0.0	5.3	3.3	37.6
El Salvador	10.4	1.0	1.4	0.3	44.7	35.2	0.1	0.1	46.2
Guatemala	21.8	5.2	22.3	15.7	32.5	17.8	3.6	0.8	57.8
Honduras	2.7	1.2	3.5	1.9	43.2	39.5	8.1	5.7	51.5
Jamaica	16.9	14.1	0.6	2.0	1.2	0.1	7.6	4.8	23.6
Panama	4.1	3.4	42.1	40.0	1.6	5.1	0.4	0.0	48.6
Paraguay	0.4	0.3	19.4	17.6	0.9	0.3	9.5	7.8	27.6
Peru	0.1	0.3	45.4	46.4	0.9	0.5	2.0	3.3	49.5
Trinidad and Tobago	9.3	7.2	24.2	21.8	0.8	0.1	2.7	5.7	35.8
Uruguay	2.2	8.2	31.5	33.1	2.4	4.0	0.5	0.4	41.0
Venezuela	12.1	5.2	19.7	15.1	2.9	0.1	9.1	5.3	33.7
Austria	65.6	51.0	1.9	0.0	9.6	0.1	7.6	2.1	67.9
Belgium	80.5	61.9	6.8	1.9	12.7	0.2	4.9	0.5	83.3
Denmark	53.7	42.7	0.0	1.1	15.4	2.5	0.0	1.1	58.3
Finland	57.1	52.7	1.9	0.2	5.5	0.0	0.2	2.8	60.1
France	85.5	66.6	2.5	0.1	16.0	0.8	3.6	0.2	86.5
Germany	59.2	39.1	1.3	0.0	18.9	3.2	9.5	3.4	66.0
Greece	53.8	45.9	0.7	0.0	7.2	0.0	0.8	0.1	53.9
Iceland	15.3	5.6	11.5	8.4	23.0	13.3	1.9	5.1	41.7
Ireland	6.2	4.5	2.7	3.7	10.9	7.5	8.1	10.6	27.9
Italy	75.2	62.2	3.3	0.3	5.2	0.6	12.7	4.9	81.0
Luxembourg	23.2	14.8	4.1	7.7	17.7	7.9	0.6	0.0	38.3
Netherlands	70.4	49.5	0.3	0.4	27.6	6.6	0.9	0.2	77.4
Norway	24.3	12.4	5.6	2.6	14.8	5.0	2.4	0.4	32.8
Portugal	58.6	44.8	0.0	1.2	9.7	0.3	12.2	5.5	65.4
Spain	68.6	60.5	0.2	0.3	10.5	0.2	0.0	2.0	71.2
Sweden	53.7	51.4	2.6	0.6	6.3	0.0	4.9	11.9	66.0
Switzerland	49.3	42.3	5.4	2.1	4.9	0.0	0.2	0.5	51.8
United Kingdom	25.6	13.9	7.9	13.2	31.8	18.3	0.3	0.0	56.2
Cameroon	0.0	0.0	2.6	3.0	0.0	0.0	3.6	3.9	6.7
Côte d'Ivoire	11.7	9.1	2.0	1.0	3.1	0.4	0.5	1.7	14.7
Kenya	0.0	2.5	1.0	0.9	16.5	19.2	1.0	1.3	21.3
Morocco	6.2	14.8	2.7	2.5	9.0	16.7	4.1	5.9	31.0
Senegal	1.5	0.0	4.2	3.0	7.2	5.3	7.0	5.4	16.1
South Africa	30.1	19.6	29.8	23.1	7.0	0.5	0.0	1.4	54.9
Zimbabwe	2.7	0.1	2.0	3.1	23.0	20.9	0.2	0.6	26.8
Bangladesh	7.0	6.6	0.1	0.0	0.0	0.8	3.1	1.9	9.8
India	4.8	1.9	2.0	0.9	1.2	0.1	11.4	9.1	15.2
Indonesia	1.6	5.4	0.5	0.3	0.1	0.0	53.2	57.1	59.5
Pakistan	17.4	19.2	0.1	0.1	0.4	5.1	5.3	3.0	25.3
Philippines	2.9	0.5	11.6	9.8	2.8	1.1	1.7	0.7	15.0
Sri Lanka	2.2	3.7	16.6	15.9	5.5	10.0	0.2	0.0	27.5
Hong Kong SAR	24.0	9.2	5.6	2.2	13.6	4.0	27.3	19.1	50.6
Japan	68.4	58.6	1.2	0.0	3.5	1.1	13.9	6.1	75.5
Korea	13.4	6.2	7.8	13.2	7.1	2.1	38.5	33.9	60.9
Malaysia	6.3	0.3	1.0	0.1	10.9	5.9	44.9	39.7	53.1
Singapore	15.1	6.5	3.3	1.0	2.4	0.0	44.5	36.9	53.7
Thailand	16.3	11.3	0.4	0.0	0.0	2.4	44.4	37.7	56.0
Australia	18.8	9.1	0.3	0.0	25.0	14.4	1.0	3.3	36.1
New Zealand	5.5	4.8	0.1	0.0	3.0	1.1	4.5	6.5	13.0

Note. This table shows for each country the percentage of the variance of the real GDP growth rate explained by each estimated factor, and the commonality. Since the factors estimated by PCA quartimin are correlated, we also report an adjusted measure of the percentage of the variance explained by each estimated factor, controlling for the influence of the other estimated factors. The variance decomposition method is detailed in Appendix A of the paper.

C Supplementary material for the application to the US economy

C.1 Data

We use the 2018:9 vintage of FRED-MD. The composition of FRED-MD is subject to minor changes over time. The 128 variables of this vintage are listed below. Most of them are retrieved from the Federal Reserve Economic Database (FRED), so their mnemonic is the same as in FRED. Some series require adjustments to the raw data available in FRED (see McCracken and Ng, 2016). The transformation code (TC) indicates how the series was transformed to ensure stationarity: (1) no transformation, (2) Δx_t, (3) $\Delta^2 x_t$, (4) $\log(x_t)$, (5) $\Delta \log(x_t)$, (6) $\Delta^2 \log(x_t)$, (7) $\Delta(x_t/x_{t-1} - 1)$. In order to work with a balanced panel, we drop 5 of 128 variables. The binary entry BP indicates whether that variable is included in the balanced panel 1960:1-2018:4.

Table C6: Output and income.

Mnemonic	Description	TC	BP
RPI	Real Personal Income	5	1
W875RX1	Real Personal Income Excluding Transfer Receipts	5	1
INDPRO	IP Index	5	1
IPFPNSS	IP: Final Products and Nonindustrial Supplies	5	1
IPFINAL	IP: Final Products (Market Group)	5	1
IPCONGD	IP: Consumer Goods	5	1
IPDCONGD	IP: Durable Consumer Goods	5	1
IPNCONGD	IP: Nondurable Consumer Goods	5	1
IPBUSEQ	IP: Business Equipment	5	1
IPMAT	IP: Materials	5	1
IPDMAT	IP: Durable Materials	5	1
IPNMAT	IP: Nondurable Materials	5	1
IPMANSICS	IP: Manufacturing (Standard Industrial Classification)	5	1
IPB51222s	IP: Residential Utilities	5	1
IPFUELS	IP: Fuels	5	1
CUMFNS	Capacity Utilization: Manufacturing	2	1

Table C7: Labor market.

Mnemonic	Description	TC	BP
HWI	Help-Wanted Index for United States	2	1
HWIURATIO	Ratio of Help Wanted to Number of Unemployed	2	1
CLF16OV	Civilian Labor Force	5	1
CE16OV	Civilian Employment	5	1
UNRATE	Civilian Unemployment Rate	2	1
UEMPMEAN	Average Duration of Unemployment (Weeks)	2	1
UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	5	1
UEMP5TO14	Civilians Unemployed for 5-14 Weeks	5	1
UEMP15OV	Civilians Unemployed - 15 Weeks and Over	5	1
UEMP15T26	Civilians Unemployed for 15-26 Weeks	5	1
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	5	1
CLAIMSx	Initial Claims	5	1
PAYEMS	All Employees: Total Nonfarm	5	1
USGOOD	All Employees: Goods-Producing Industries	5	1
CES1021000001	All Employees: Mining and Logging: Mining	5	1
USCONS	All Employees: Construction	5	1
MANEMP	All Employees: Manufacturing	5	1
DMANEMP	All Employees: Durable Goods	5	1
NDMANEMP	All Employees: Nondurable Goods	5	1
SRVPRD	All Employees: Service-Providing Industries	5	1
USTPU	All Employees: Trade, Transportation and Utilities	5	1
USWTRADE	All Employees: Wholesale Trade	5	1
USTRADE	All Employees: Retail Trade	5	1
USFIRE	All Employees: Financial Activities	5	1
USGOVT	All Employees: Government	5	1
CES0600000007	Average Weekly Hours: Goods-Producing	1	1
AWOTMAN	Average Weekly Overtime Hours: Manufacturing	2	1
AWHMAN	Average Weekly Hours: Manufacturing	1	1
CES0600000008	Average Hourly Earnings: Goods-Producing	6	1
CES2000000008	Average Hourly Earnings: Construction	6	1
CES3000000008	Average Hourly Earnings: Manufacturing	6	1

Table C8: Housing.

Mnemonic	Description	TC	BP
HOUST	Housing Starts: Total New Privately Owned	4	1
HOUSTNE	Housing Starts, Northeast	4	1
HOUSTMW	Housing Starts, Midwest	4	1
HOUSTS	Housing Starts, South	4	1
HOUSTW	Housing Starts, West	4	1
PERMIT	New Private Housing Permits (SAAR)	4	1
PERMITNE	New Private Housing Permits, Northeast (SAAR)	4	1
PERMITMW	New Private Housing Permits, Midwest (SAAR)	4	1
PERMITS	New Private Housing Permits, South (SAAR)	4	1
PERMITW	New Private Housing Permits, West (SAAR)	4	1

Table C9: Consumption, orders, and inventories.

Mnemonic	Description	TC	BP
DPCERA3M086SBEA	Real Personal Consumption Expenditures	5	1
CMRMTSPLx	Real Manufacturing and Trade Industries Sales	5	1
RETAILx	Retail and Food Services Sales	5	1
ACOGNO	New Orders for Consumer Goods	5	0
AMDMNOx	New Orders for Durable Goods	5	1
ANDENOx	New Orders for Nondefense Capital Good	5	0
AMDMUOx	Unfilled Orders for Durable Goods	5	1
BUSINVx	Total Business Inventories	5	1
ISRATIOx	Total Business: Inventories to Sales Ratio	2	1
UMCSENTx	Consumer Sentiment Index	2	0

Table C10: Money and credit.

Mnemonic	Description	TC	BP
M1SL	M1 Money Stock	6	1
M2SL	M2 Money Stock	6	1
M2REAL	Real M2 Money Stock	5	1
AMBSL	St. Louis Adjusted Monetary Base	6	1
TOTRESNS	Total Reserves of Depository Institutions	6	1
NONBORRES	Reserves Of Depository Institutions	7	1
BUSLOANS	Commercial and Industrial Loans	6	1
REALLN	Real Estate Loans at All Commercial Banks	6	1
NONREVSL	Total Nonrevolving Credit	6	1
CONSPI	Nonrevolving Consumer Credit to Personal Income	2	1
MZMSL	MZM Money Stock	6	1
DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	6	1
DTCTHFNM	Total Consumer Loans and Leases Outstanding	6	1
INVEST	Securities in Bank Credit at All Commercial Bank	6	1

Table C11: Interest rates, spreads, and exchange rates.

Mnemonic	Description	TC	BP
FEDFUNDS	Effective Federal Funds Rate	2	1
CP3Mx	3-Month AA Financial Commercial Paper Rate	2	1
TB3MS	3-Month Treasury Bill: Secondary Market Rate	2	1
TB6MS	6-Month Treasury Bill: Secondary Market Rate	2	1
GS1	1-Year Treasury Constant Maturity Rate	2	1
GS5	5-Year Treasury Constant Maturity Rate	2	1
GS10	10-Year Treasury Constant Maturity Rate	2	1
AAA	Moody's Seasoned Aaa Corporate Bond Yield	2	1
BAA	Moody's Seasoned Baa Corporate Bond Yield	2	1
COMPAPFFx	CP3Mx - FEDFUNDS	1	1
TB3SMFFM	TB3MS - FEDFUNDS	1	1
TB6SMFFM	TB6MS - FEDFUNDS	1	1
T1YFFM	GS1 - FEDFUNDS	1	1
T5YFFM	GS5 - FEDFUNDS	1	1
T10YFFM	GS10 - FEDFUNDS	1	1
AAAFFM	AAA - FEDFUNDS	1	1
BAAFFM	BAA - FEDFUNDS	1	1
TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies	5	0
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	5	1
EXJPUSx	Japan / U.S. Foreign Exchange Rate	5	1
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	5	1
EXCAUSx	Canada / U.S. Foreign Exchange Rate	5	1

Table C12: Prices.

Mnemonic	Description	TC	BP
WPSFD49207	PPI: Finished Goods	6	1
WPSFD49502	PPI: Finished Consumer Goods	6	1
WPSID61	PPI: Intermediate Materials	6	1
WPSID62	PPI: Crude Materials	6	1
OILPRICEx	Crude Oil, Spliced WTI and Cushing	6	1
PPICMM	PPI: Metals and Metal Products	6	1
CPIAUCSL	CPI: All Items	6	1
CPIAPPSL	CPI: Apparel	6	1
CPITRNSL	CPI: Transportation	6	1
CPIMEDSL	CPI: Medical Care	6	1
CUSR0000SAC	CPI: Commodities	6	1
CUSR0000SAD	CPI: Durables	6	1
CUSR0000SAS	CPI: Services	6	1
CPIULFSL	CPI: All Items Less Food	6	1
CUSR0000SA0L2	CPI: All Items Less Shelter	6	1
CUSR0000SA0L5	CPI: All Items Less Medical Care	6	1
PCEPI	Personal Consumption Expenditures: Chain Index	6	1
DDURRG3M086SBEA	Personal Consumption Expenditures: Durable Goods	6	1
DNDGRG3M086SBEA	Personal Consumption Expenditures: Nondurable Goods	6	1
DSERRG3M086SBEA	Personal Consumption Expenditures: Services	6	1

Table C13: Stock market.

Mnemonic	Description	TC	BP
S&P 500	S&P's Common Stock Price Index: Composite	5	1
S&P: indust	S&P's Common Stock Price Index: Industrials	5	1
S&P div yield	S&P's Composite Common Stock: Dividend Yield	2	1
S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	5	1
VXOCLSx	VXO	1	0

C.2 Additional results

Table C14: Selected number of factors (FRED-MD).

Criterion	IC_{p1}	IC_{p2}	IC_{p3}	PC_{p1}	PC_{p2}	PC_{p3}
r^*	8	8	10	9	9	10

Note. Number of factors selected by each of the six main Bai and Ng (2002) information criteria, allowing for at most ten factors.

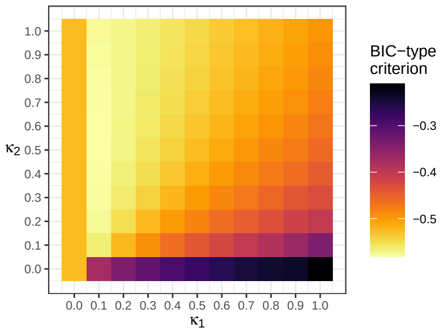


Figure C5: Grid search of the SPCA hyperparameters κ_1 and κ_2 (FRED-MD).

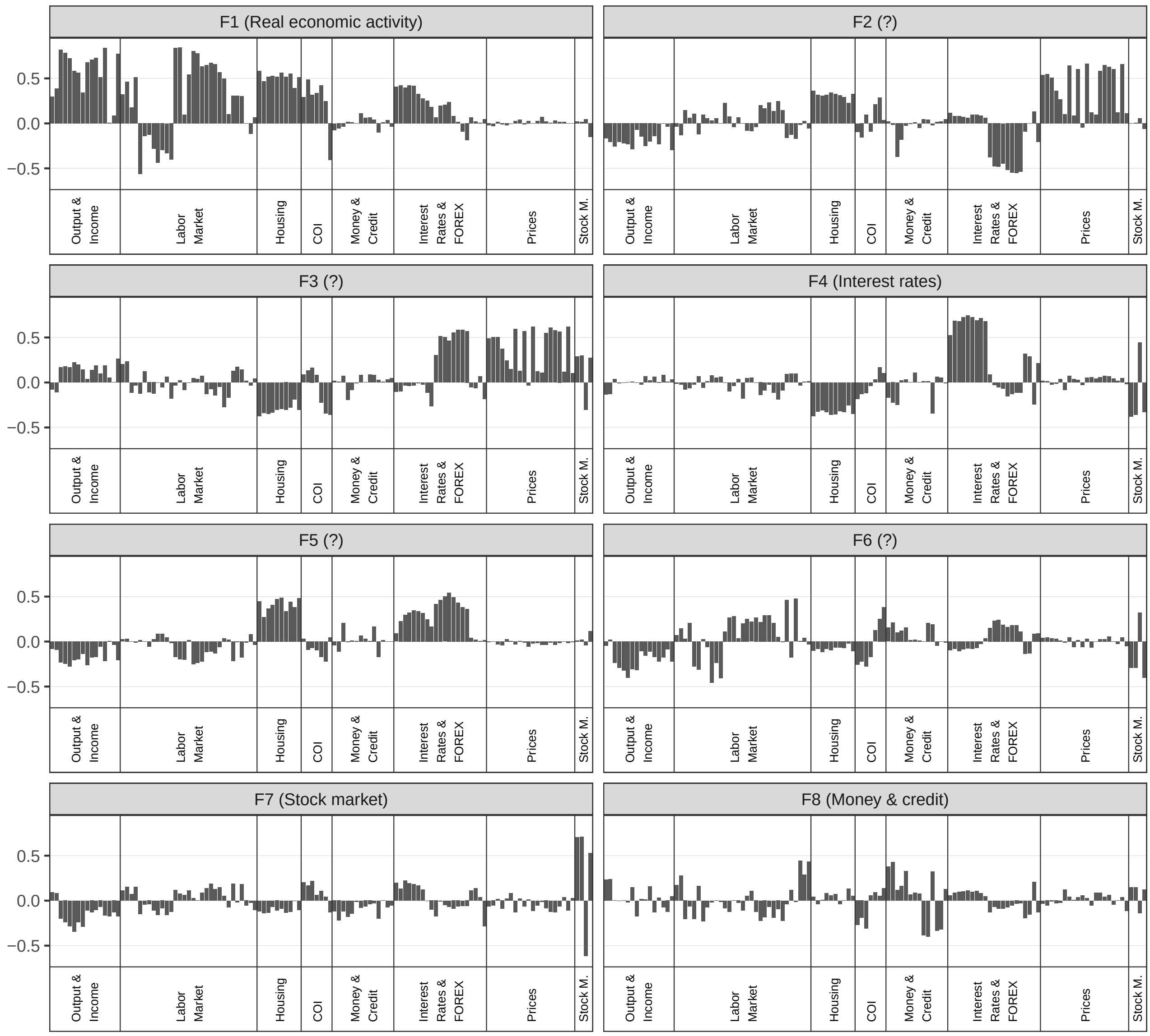


Figure C6: Factor loadings estimated with PCA (FRED-MD).

Note. The interpretable factors are labeled with their economic interpretation, and the uninterpretable factors with a question mark. Since F and Λ are identified up to a column sign change, for each factor, we impose the largest loading in absolute value to be positive.

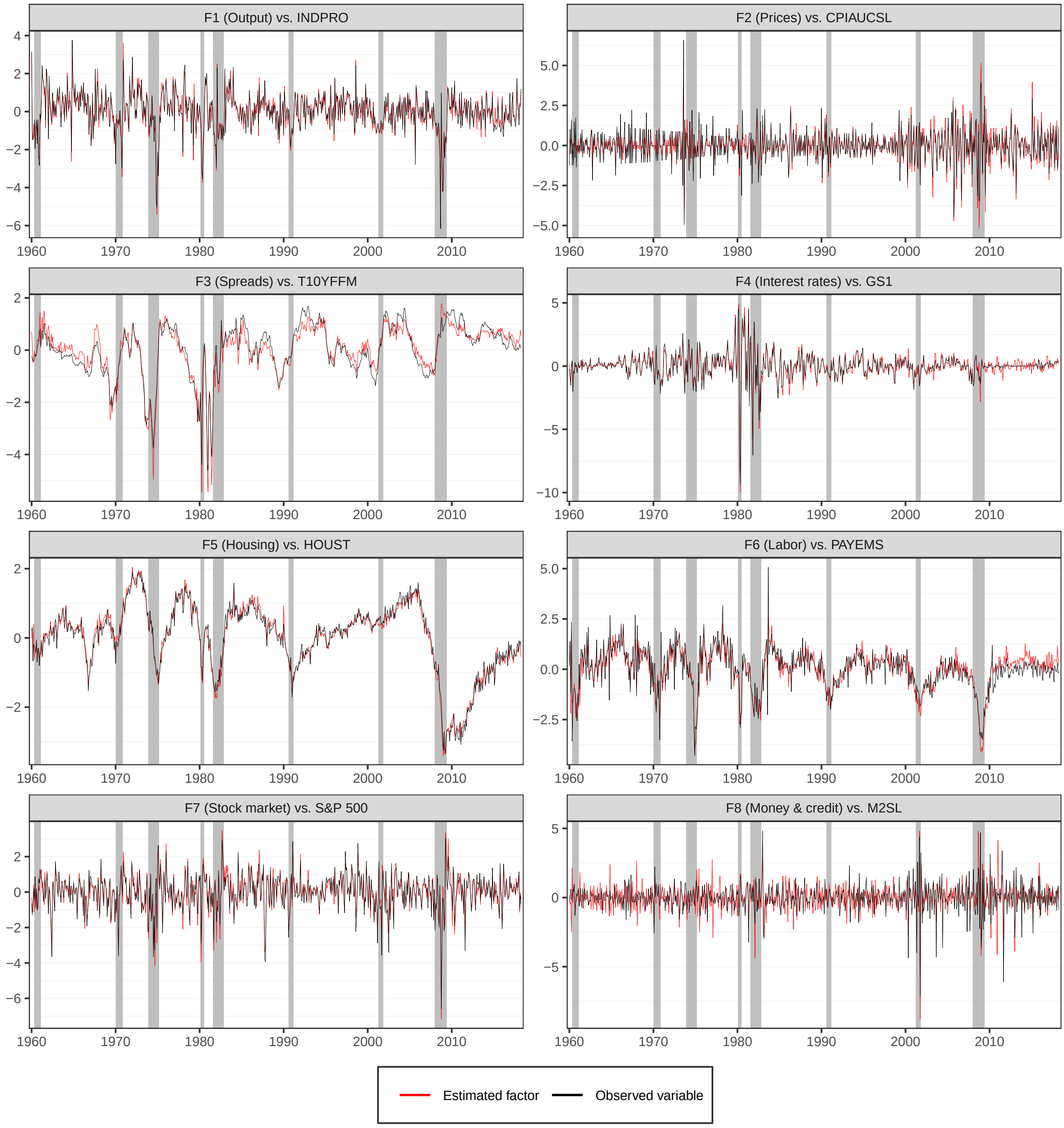


Figure C7: Factors estimated with SPCA, compared with some representative observed variables (FRED-MD).

Note. Shaded areas indicate periods of recession as defined by the National Bureau of Economic Research. The observed variables are stationarized using the transformations recommended by McCracken and Ng (2016). The correlations between the factors estimated with SPCA and the representative observed variables are respectively 0.96, 0.89, 0.95, 0.96, 0.99, 0.91, 0.94, 0.61.

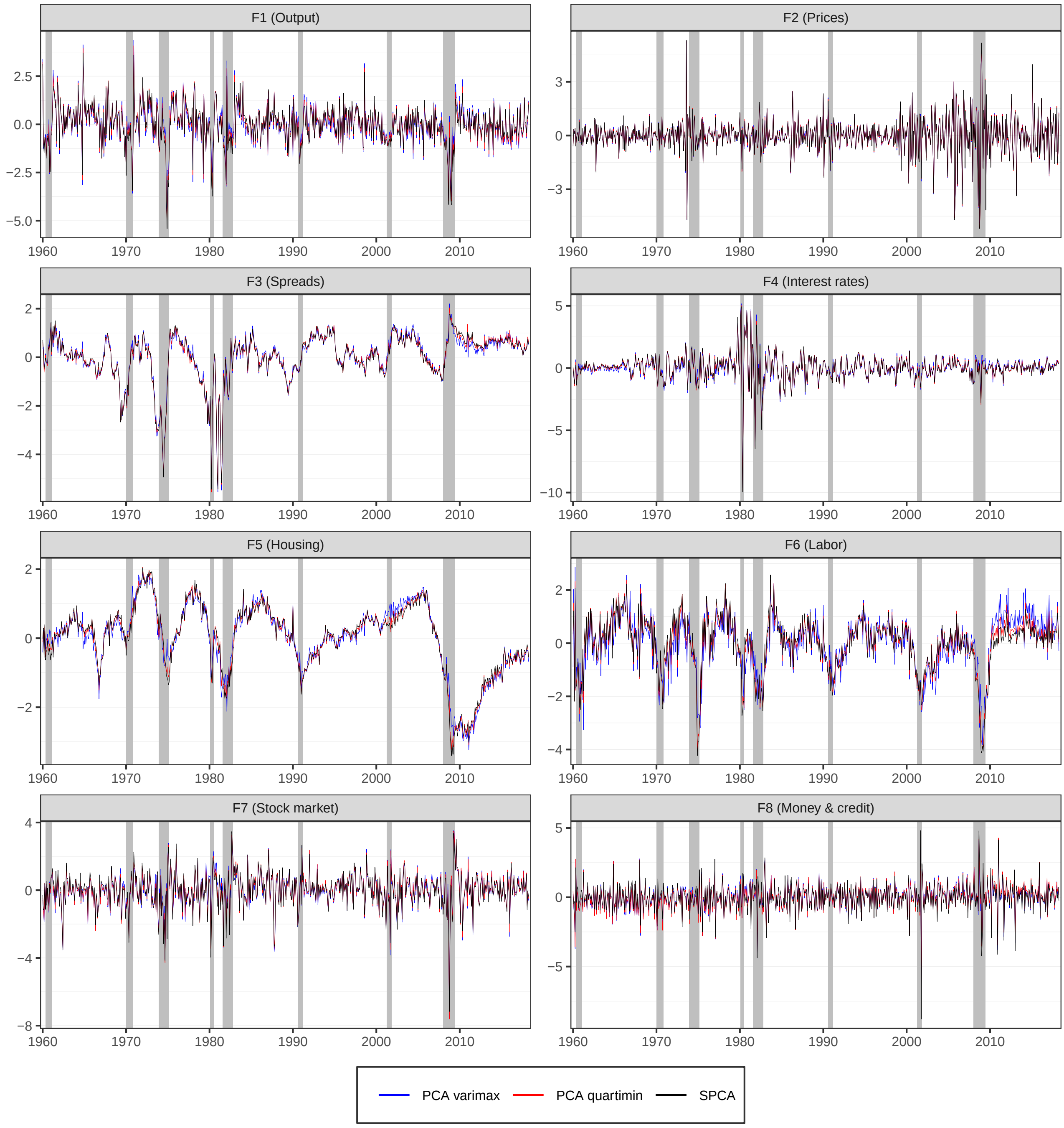


Figure C8: Factors estimated with PCA varimax, PCA quartimin, and SPCA (FRED-MD).

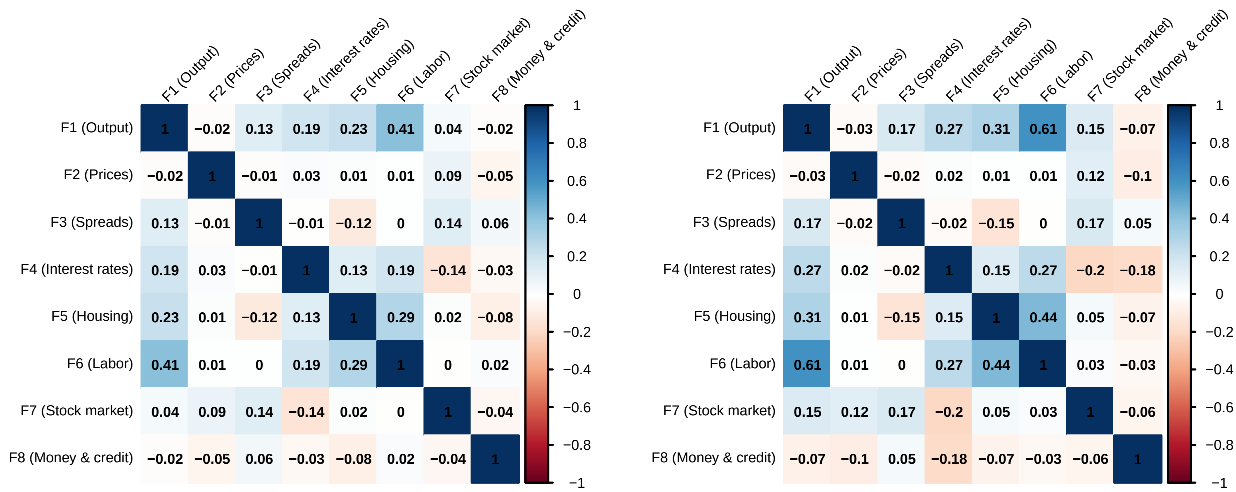
Note. Shaded areas indicate periods of recession as defined by the National Bureau of Economic Research.

Table C15: Correlations between the estimated factors (FRED-MD).

Factors	PCA varimax vs. PCA quartimin	PCA varimax vs. SPCA	PCA quartimin vs. SPCA
F1 (Output)	0.99	0.95	0.99
F2 (Prices)	1.00	1.00	1.00
F3 (Spreads)	0.99	0.99	1.00
F4 (Interest rates)	0.99	0.98	1.00
F5 (Housing)	0.99	0.97	1.00
F6 (Labor)	0.93	0.87	0.99
F7 (Stock market)	0.99	0.96	0.98
F8 (Money & credit)	0.99	0.95	0.96

Table C16: Correlations between the estimated factors and some representative observed variables (FRED-MD).

Pairs	PCA varimax	PCA quartimin	SPCA
F1 (Output) vs. INDPRO	0.92	0.95	0.96
F2 (Prices) vs. CPIAUCSL	0.89	0.89	0.89
F3 (Spreads) vs. T10YFFM	0.93	0.94	0.95
F4 (Interest rates) vs. GS1	0.93	0.95	0.96
F5 (Housing) vs. HOUST	0.97	0.99	0.99
F6 (Labor) vs. PAYEMS	0.72	0.87	0.91
F7 (Stock market) vs. S&P 500	0.90	0.91	0.94
F8 (Money & credit) vs. M2SL	0.51	0.51	0.61



(a) PCA quartimin (b) SPCA

Figure C9: Correlation matrices of the estimated factors (FRED-MD).

Note. The correlation matrices of the factors estimated with PCA and PCA varimax are not displayed since they are simply the identity matrix, by construction.