Another look into the factor model black box: factor interpretation and structural (in)stability

Thomas Despois, Catherine Doz

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2019 – 44

Another look into the factor model black box: factor interpretation and structural (in)stability

Thomas Despois
Catherine Doz

JEL Codes: C32, C33, C38, C55
Keywords: dynamic factor models, factor rotation, sparse PCA, structural break

AGENCE NATIONALE DE LA RECHERCHE
ANR
INVESTIR L'AVENIR

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Another look into the factor model black box: factor interpretation and structural (in)stability*

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

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

This version: February 4, 2020.
First version: July 31, 2019.

Abstract

This paper considers two families of methods allowing to get interpretable factors without imposing their interpretation a priori: factor rotations and sparse PCA. Monte Carlo simulations show their performance in recovering the correct factor structure. In an empirical application with a large U.S. macroeconomic dataset, they recover the same factor structure, offering a clear economic interpretation. This factor representation seems more natural and informative than competitors, offering new lens to disentangle the driving forces within a factor model. In particular, the structural instability appears to be more limited, and of different nature than what was previously found.

JEL classification: C32, C33, C38, C55.
Keywords: dynamic factor models, factor rotation, sparse PCA, structural break.

*Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France.
E-mail addresses: thomas.despois@psemail.eu and catherine.doz@univ-paris1.fr.
We would like to thank the participants at several PSE and Paris 1 internal seminars, the Banque de France-PSE Chair International Macroeconomics Workshop, the Tinbergen Institute Seminar, the 68th Annual Meeting of the French Economic Association, the 2019 Annual Conference of the International Association for Applied Econometrics, and the 2019 European Summer Meeting of the Econometric Society.

1 Introduction

Dynamic factor models (DFMs) have been successfully used in various applications, and have become increasingly popular. Indeed, they present several appealing features. By summarizing a dataset into a small number of latent factors driving the bulk of the comovements of the variables, they provide a very convenient and flexible framework to model together a potentially very large number of time series. Factor models also constitute a rather agnostic approach.[1] The estimated factors can also be used in various ways, especially predictors in a forecasting model (Stock and Watson, 2002a,b), or control variables in a VAR (leading to a Factor-Augmented VAR or FAVAR, originally developed by Bernanke, Boivin and Eliasz, 2005). Moreover, "for certain problems, such as forecasting real economic activity, DFM forecasts are in many cases the best available forecasts" (Stock and Watson, 2016, p. 516). They can also be easily adapted to handle missing observations or mixed frequency data, which is very convenient for nowcasting and forecasting.

Despite all these qualities, DFMs suffer from two weaknesses. First, they constitute black boxes, especially because the factors are very difficult to interpret with the methods generally used to estimate them. Secondly, there is recent and mounting evidence that DFMs can be subject to structural instability, potentially leading to an inconsistent estimation of the factors and thus forecast failures. The central point of this paper is to tackle the uninterpretability issue. In turn, this sheds a whole new light on the structural instability issue, which appear to be closely related.

So far, the factor model black box has been opened from three different angles. First, the stationary (Forni et al., 2009) and non-stationary (Eickmeier, 2009) structural DFM frameworks, transposing to DFMs the methods for identifying shocks in structural VARs, allowed to identify the shocks affecting the factors. Secondly, Giannone, Reichlin and Small (2008) and Bańbura and Rünstler (2011) assessed the contribution of individual series (or groups of series) in the forecasts and in reducing the forecasts uncertainty. Thirdly, an abundant literature has been focusing on getting an economic interpretation of the factors, and thus understanding their role in the model (in particular Kose, Otrok and Whiteman, 2003; Gürkaynak, Sack and Swanson, 2005; Bai and Ng, 2013).

An individual interpretation of the factors can be attempted through the factor loadings. They represent the link between each factor and each variable. If only a small number of loadings associated to a given factor are large (or even non-zero), it means that this factor is mainly linked to a small number of variables, which generally allows to give an economic interpretation to it. Nevertheless, with all the usual estimation methods, like for instance principal component

[1] This was actually the main reason for the introduction of dynamic factor models in economics, as highlighted in the title of the seminal article from Sargent and Sims (1977): *Business Cycle Modeling Without Pretending To Have Too Much A Priori Economic Theory.*

analysis (PCA), the estimated factors are quite difficult to interpret, if not impossible. Most of the literature tried to solve this issue by imposing a specific interpretation to each factor, whether via hierarchical DFMs or via factor rotations with a priori identifying restrictions. As its name suggests, a hierarchical DFM (Kose, Otrok and Whiteman, 2003) enforces a hierarchy in the factors, e.g. a world factor, several regional factors, and many country-specific factors. The factor rotations with a priori 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 has then been used by several other papers in monetary policy (e.g. Miranda-Agrippino and Rey, 2018). This methodology was then extended by Bai and Ng (2013), who proposed two rotations with a priori 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), etc.

An alternative and more agnostic way to get interpretable factors consists in using sparse factor models, introducing sparsity in the factor loadings. Sparse factor models have been used for some time in statistics, especially biostatistics (West, 2003), but they have been introduced in macroeconomics only very recently. They turned out to help the factor interpretation in this field too (Croux and Exterkate, 2011; Kaufmann and Schumacher, 2017, 2019; Beyeler and Kaufmann, 2019), and can also slightly improve the forecasting performance (Croux and Exterkate, 2011; Kristensen, 2017; Kim and Swanson, 2018).

A second weakness of DFMs lies in their potential structural instability and its consequences. The principal components estimator of the factors has been proved to be robust to small (Stock and Watson, 2002a) or even quite substantial (Bates et al., 2013) time variation in the factor loadings. The intuition is that the instabilities will be averaged away if they are moderate enough and sufficiently independent across series. Nevertheless, in case of severe instability, the factor space will not be consistently estimated anymore, potentially leading to forecast failures, e.g. after the occurrence of a large structural break.

However, it is increasingly clear that DFMs are subject to structural instability. Using U.S. macroeconomic data, Stock and Watson (2009) first reported considerable evidence of breaks in 1984, i.e. around the beginning of the Great Moderation. This finding gave birth to a very fast growing literature of tests for structural instability in DFMs, progressively refined so as to allow for an unknown break date (which is then estimated), and then for several breaks.[2] When considering U.S. macroeconomic data, these tests systematically corroborate the presence of instabilities, but yield quite largely different conclusions with respect to the number of breaks,

[2] Breitung and Eickmeier (2011); Chen, Dolado and Gonzalo (2014); Corradi and Swanson (2014); Han and Inoue (2015); Yamamoto and Tanaka (2015); Cheng, Liao and Schorfheide (2016); Baltagi, Kao and Wang (2017); Massacci (2017); Su and Wang (2017); Barigozzi, Cho and Fryzlewicz (2018); Ma and Su (2018); Bai, Han and Shi (forthcoming); etc.

their date, and their form (especially as to whether the breaks are in the loadings and/or in the number of factors). Indeed, Breitung and Eickmeier (2011) confirm the presence of a structural break around 1984,[3] while Chen, Dolado and Gonzalo (2014) also detect a single break, but rather around 1979-1980 (second oil crisis). On the opposite, Ma and Su (2018) identify five breaks: 1979:9 (second oil crisis), 1984:7 (beginning of the Great Moderation), 1990:3, 1995:6 (two labor productivity shocks), and 2002:1 (early 2000s recession). Contrary to the previous studies, Cheng, Liao and Schorfheide (2016) chose to ignore the pre-Great Moderation period but to include the Great Recession, leading to the detection of a single break in the factor loadings, associated with the emergence of a new factor towards the end of 2007. However, studying the 2007–09 recession in the U.S., Stock and Watson (2012) find no evidence for a systematic or widespread break in the factor loadings, or for the emergence of a new factor.

The central point of this paper is to get interpretable factors while remaining deliberately atheoretical. To achieve this, we consider two families of estimation methods: sparse PCA (coming from the machine learning literature), and factor rotations relying on an atheoretical simplicity criterion in the loadings (rather than a priori identifying restrictions).

Monte Carlo simulations show that these methods recover the correct factor structure with accuracy, even in small samples. The improvement over standard PCA can be very substantial. In an empirical application with a large dataset of U.S. macroeconomic and financial variables (the now standard FRED-MD), these methods offer a straightforward economic interpretation for the estimated factors, which appears to be stable over time: output, prices, spreads, interest rates, housing, labor, stock market, money and credit. Moreover, they recover exactly the same factor structure. These results suggest that standard PCA, while computationally convenient, leads to a quite unnatural factor representation. As soon as a method enforcing a simple structure in the loadings is used (no matter the method), it yields the same factor representation.

This factor representation provides new lens to study the structural instability issue. We find very little evidence for the emergence of an additional factor or for a widespread break in the factor loadings at a given time. The loadings exhibit a surprisingly high temporal stability. The structural instability in the model seems to rather consist in breaks in a limited number of loadings. They can be localized, interpreted, and are coherent with economic history.

The rest of the paper is structured as follows. Sections 2 introduces the econometric framework. Sections 3 and 4 respectively presents the results of Monte Carlo simulations and an empirical application on a large U.S. macroeconomic dataset. Section 5 studies the structural instability within the factor model. Section 6 concludes.

[3] They also applied their tests on euro area data, finding evidence for breaks in 1992 (Maastricht treaty) and 1999 (handover of monetary policy from European national central banks to the ECB).

2 Econometric framework

We observe N stationary time series over T periods. It is assumed that the observed variables are driven by a fixed number r of unobserved common factors, with $r << N$. Let X be the $T \times N$ panel data matrix (with all the variables standardized beforehand), F the $T \times r$ matrix including the r factors, Λ the $N \times r$ factor loading matrix, and e the $T \times N$ idiosyncratic errors matrix. Denoting i the cross-section unit ($i = 1, \ldots, N$), t the time unit ($t = 1, \ldots, T$), and x_{it} the observation of x_i at time t, $F_t = (f_{1t}, \ldots, f_{rt})'$, $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{ir})'$, the model is written as:

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

In matrix notations, with $X_t = (x_{1t}, \ldots, x_{Nt})'$ and $e_t = (e_{1t}, \ldots, e_{Nt})'$, it is also written as

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

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

Factor models are characterized by the fact that (F_t) and (e_t) are unobserved stochastic processes with zero mean, which are supposed to be uncorrelated at all leads and lags. Note that although the relation between the observed variables X_t and the common factors F_t is static (no lag of F_t enters equation (2)), the model is a dynamic factor model, since (F_t) and (e_t) can be autocorrelated processes. Further, we suppose that it is an approximate factor model, which means that the (e_{it})'s may be weakly cross-correlated and that the bulk of the comovements of the (x_{it})'s is captured by the factors. This is obtained through a standard set of assumptions, including in particular the fact that the r eigenvalues of $\Lambda'\Lambda$ diverge at rate N whereas the covariance matrix of the idiosyncratic errors e_t stays bounded when N goes to infinity (see e.g. Stock and Watson, 2002a, and Bai and Ng, 2002, for the most often used sets of assumptions). In particular, it is very often assumed that $\Lambda'\Lambda/N \longrightarrow \Sigma_\Lambda$ when N goes to infinity, with Σ_Λ a positive definite matrix.

It is well known that F and Λ are defined only up to an invertible matrix. Indeed, for any $r \times r$ invertible matrix R, equation (2) can also be written as

$$X_t = \widetilde{\Lambda}\widetilde{F}_t + e_t \tag{4}$$

with $\widetilde{\Lambda} = \Lambda R$ and $\widetilde{F}_t = R^{-1}F_t$. Similarly, equation (3) can also be written as

$$X = \widetilde{F}\widetilde{\Lambda}' + e \quad \text{with} \quad \widetilde{F} = F(R')^{-1}. \tag{5}$$

Due to this indeterminacy, only the subspace of $\mathbb{R}^r$ spanned by the factors, and the subspace of

$\mathbb{R}^N$ spanned by the columns of Λ are identifiable and can be estimated.[4] This indeterminacy issue will be the core of our paper. We aim to estimate a loading matrix Λ having sparsity properties in order to be able to interpret the factors. Indeed, as already mentioned in the introduction, the factor loading λ_{ij} is the weight of the jth factor on the variable x_i and measures the link between the jth factor and the ith variable. The jth column of Λ can thus be used to get an interpretation of the jth factor, since it displays which variables are linked to this factor. With all the usual estimation methods (e.g. PCA), the estimated loading matrix $\widehat{\Lambda}$ tends to have a complex structure, making the interpretation of the estimated factors very difficult, if not impossible. In this paper, we build on the PCA estimator of the model, and propose several approaches allowing to get a simple structure in $\widehat{\Lambda}$ and ease the interpretation of the estimated factors.

Notation. For a vector $v \in \mathbb{R}^n$, we denote its ℓ_1 norm as $\|v\|_1 = \sum_{i=1}^n |v_i|$, and its ℓ_2 norm as $\|v\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$. The ℓ_0 norm of v refers to its cardinality, i.e. its number of non-null elements. The Frobenius norm of a matrix $A \in \mathbb{R}^{m\times n}$ is denoted $\|A\|_F = \sqrt{\mathrm{Tr}(A'A)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$.

2.1 Principal component analysis

Stock and Watson (2002a) and Bai and Ng (2002) proved that the factors in model (1) are consistently estimated by principal components under general sets of assumptions associated with approximate DFMs. Further, Stock and Watson (2002a) proved that the result is also valid in case of a small temporal instability in the factor loadings. Bates et al. (2013) later established that the factors are still consistently estimated under a more substantial structural instability, for example if a large discrete break occurs in the factor loadings for a fraction $O(N^{-1/2})$ of the series, or if the factor loadings follow independent random walks (as long as their innovations are relatively small, and independent across series).

PCA is a dimensionality reduction technique which summarizes N variables into r factors. It can be formulated as a nonlinear least squares problem:

$$\min_{F,\Lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \lambda_i' F_t)^2, \tag{6}$$

which can be rewritten as

$$\min_{F,\Lambda} \|X - F\Lambda'\|_F^2. \tag{7}$$

[4]If one wants to uniquely identify the factors and the loadings, it is necessary to impose normalization constraints. For instance, it can be imposed that $E(F_tF_t') = I_r$ and that Σ_Λ is a diagonal matrix or that $E(F_tF_t') = \Sigma_F$ is a diagonal matrix whereas $\Lambda'\Lambda/N \longrightarrow I_r$ when N goes to infinity. Under both these normalization conditions, F and Λ are uniquely identified, up to column permutations and column sign changes (see e.g. Stock and Watson, 2002a, and Bai and Ng, 2002). Similar sets of constraints can be imposed at the estimation stage, in order to get a unique solution, as we will see below for the PCA estimators.

It is clear that the minimization problem (6) and (7) is 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 conditions have thus to be added in order to get a unique solution of the minimization problem. Two standard sets of normalization conditions are used: they are empirical counterparts of the normalization conditions which have been mentioned before for the theoretical model.

Normalization A. $\dfrac{\widehat{F}'\widehat{F}}{T} = I_r$ and $\widehat{\Lambda}'\widehat{\Lambda}$ diagonal.

Normalization B. $\dfrac{\widehat{\Lambda}'\widehat{\Lambda}}{N} = I_r$ and $\widehat{F}'\widehat{F}$ diagonal.

Since most factor rotations rely on the assumption that $\widehat{F}'\widehat{F}/T = I_r$, we adopt normalization A. Under this normalization, the estimation of the loadings and factors can be obtained with:

$$\widehat{\Lambda}_{PCA} = \widehat{V}\widehat{D}^{1/2} \tag{8}$$

$$\widehat{F}_{PCA} = X\widehat{V}\widehat{D}^{-1/2} \tag{9}$$

where $\widehat{D} := \text{diag}(\hat{d}_1, \ldots, \hat{d}_r)$ is the $r \times r$ diagonal matrix whose entries are the r largest eigenvalues of the data covariance matrix $X'X/T$ in decreasing order of magnitude, and $\widehat{V}$ is the $N \times r$ matrix containing the corresponding eigenvectors.

Remark 1. For the following, it should be kept in mind that, by construction, PCA yields to a loading matrix $\widehat{\Lambda}_{PCA}$ whose columns are orthogonal, and to uncorrelated factors which account for successively decreasing amounts of variance.

2.2 Factor rotations

The historical method to get interpretable factors is to use factor rotations.[5] The idea is to take advantage of the indeterminacy of the estimated loadings and factors. The initial PCA estimated loading matrix $\widehat{\Lambda}_{PCA}$, leading to uninterpretable factors, is multiplied by an $r \times r$ invertible matrix R such that the transformed matrix of loadings satisfies a given simple structure criterion. The estimated factors are then transformed accordingly, and should now be interpretable.

$$\widehat{\Lambda}_{RPCA} = \widehat{\Lambda}_{PCA}R \tag{10}$$

$$\widehat{F}_{RPCA} = \widehat{F}_{PCA}(R')^{-1} \tag{11}$$

Note that, as we will see below, the literature classically names as "rotation" any invertible matrix which is used in such transformations. Some of them are true rotations: they satisfy

[5] Factor rotations were originally designed to be applied on the results of maximum likelihood estimation 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 factor rotations after a PCA is also valid, and commonplace in many disciplines (see Jolliffe, 2002, chapter 11).

$R'R = I_r$, they transform the orthogonal PCA estimated factors into a set of orthogonal factors, and are called "*orthogonal* rotations". Some others are not rotations and are classically called "*oblique* rotations": they do not preserve the angles. We will use the same terminology, even if it is not really appropriate, and we will name as "rotation" any of the invertible matrices R which will be used in what follows.

Some points should be clarified about the effects of a rotation. First, it does not yield any zero loading, but instead ensures that each estimated factor tends to have only very large and very small loadings in absolute value. In that respect, rotations achieve near-sparsity rather than sparsity. Secondly, using factor rotations does not deteriorate nor improve the fit of the factor model: the previously estimated factor space remains the same, and so do all the communalities,[6] and the total variance explained by the factors. Thirdly, when it comes to include the estimated factors in a diffusion index or another model, using the unrotated or rotated factors yields exactly the same predicted values $\widehat{y}$. After a rotation, we stay within the same factor space, so the orthogonal projection of the predicted variable y on the space spanned by the factors remains unchanged.

The factor models literature offers numerous rotation methods, which can be either "orthogonal" or "oblique". Orthogonal rotations constrain factors to stay orthogonal, whereas oblique rotations do not. Even if orthogonal rotations are routinely used, it is highly recommended to apply oblique rotations, at the very least as a starting point for the analysis (Fabrigar et al., 1999; Costello and Osborne, 2005). Indeed, perfectly uncorrelated factors are quite unrealistic in most fields, and oblique rotations are almost always necessary to recover the correct structure in the loadings (Cattell, 1978, p. 128). Moreover, oblique rotations do not force the correlation of the factors, but rather simply allow for it. If the factors are truly uncorrelated, orthogonal and oblique rotations produce nearly identical results, with the correlations among obliquely rotated factors being close to zero. We chose to consider both orthogonal and oblique rotations to ensure the robustness of our results.

Among orthogonal rotations, varimax (Kaiser, 1958) is regarded as the best one. It is overwhelmingly the most widely used rotation. This method consists in finding the rotation matrix R maximizing the variance of the squared loadings within each factor.[7]

$$R_{varimax} = \arg\max_{R} \left\{ \frac{1}{N} \sum_{k=1}^{r} \sum_{i=1}^{N} (\widehat{\Lambda}_{PCA} R)^4_{ik} - \sum_{k=1}^{r} \left(\frac{1}{N} \sum_{i=1}^{N} (\widehat{\Lambda}_{PCA} R)^2_{ik} \right)^2 \right\} \text{ s.t. } R'R = I_r \quad (12)$$

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

Concerning oblique rotations, the oblimin family is the most popular due to its great flexibility. Within this family, the quartimin rotation (Carroll, 1953; Jennrich and Sampson, 1966)

[6]The communality of a variable refers to the part of its variance explained by the estimated factors, i.e. the R^2 of the regression of the variable on the estimated factors.

[7]Hence the name: VARIance MAXimization of the squared loadings within each factor.

tends to be recommended because of its simplicity and its results (see Costello and Osborne, 2005, among others). It searches for the matrix R minimizing the covariance of squared loadings[8] between factors, and the solution is thus achieved through the minimization of the following simplicity criterion.[9]

$$R_{quartimin} = \arg\min_{R} \left\{ \sum_{j \neq k} \sum_{i=1}^{N} (\widehat{\Lambda}_{PCA} R)_{ij}^2 (\widehat{\Lambda}_{PCA} R)_{ik}^2 \right\} \text{ s.t. } \operatorname{diag}((R'R)^{-1}) = 1_r \tag{13}$$

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

Remark 2. After a rotation, $\widehat{\Lambda}'_{RPCA}\widehat{\Lambda}_{RPCA}$ is not diagonal, $\widehat{F}'_{RPCA}\widehat{F}_{RPCA}/T \neq I_r$ (unless the rotation is orthogonal), but $\operatorname{diag}(\widehat{F}'_{RPCA}\widehat{F}_{RPCA}/T) = 1_r$.

Remark 3. We have tried several other orthogonal and oblique rotations, yielding very similar results. For the sake of brevity, we only consider the standard varimax and quartimin rotations in the following.

Remark 4. Of course, factors estimated with another method than PCA can also be rotated. As a robustness check, we applied the varimax and quartimin rotations on factors estimated with two other methods: the two-step (Doz, Giannone and Reichlin, 2011) and the quasi-maximum likelihood (Doz, Giannone and Reichlin, 2012) estimators. Contrary to PCA, they explicitly take into account the dynamics of the factors. After a preliminary study, we obtained very similar results (available upon request).

2.3 Sparse principal components analysis

With sparse PCA, the machine learning literature offered 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 with respect to the norms ℓ_0 or ℓ_1 so as to induce sparsity in the loadings.[10] The resulting estimated "sparse factors" are thus linear combinations of only a small number of variables. They are easier to interpret, at the cost of a slightly smaller explained variance. This constitutes a trade-off between interpretability and statistical fidelity (explaining most of the variance in the data). However, it is considered that

[8]An oblique rotation produces two matrices of loadings: the pattern P (the matrix of partial regression weights of variables on factors), and the structure S (the matrix of covariances between factors and variables). Denoting C the factors covariance matrix, we have $S = PC$. For an orthogonal rotation, $C = I_r$ so $S = P$. The simplicity criterion of an oblique rotation is optimized while using the pattern loading matrix. In the following, the obliquely rotated factors are interpreted with the pattern loadings, as it is usually done.

[9]Hence the names: the oblimin rotations all consist in 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.

[10]Sparse factor models can also be implemented with Bayesian methods, using a mixture prior inducing sparsity in the loadings. This approach was proposed by West (2003), and has been adopted in economics by Kaufmann and Schumacher (2017, 2019), and Beyeler and Kaufmann (2019).

in many applications, the decrease in statistical fidelity implied by the sparse factors is "small and relatively benign" (d'Aspremont, Bach and El Ghaoui, 2008, among others).

Sparse PCA has been successfully applied in various fields: image processing, biostatistics, economics, finance, media data (twitter, newspapers, etc.), voting data, etc. Indeed, this method offers two key advantages over factor rotations: it yields exact zero loadings (instead of very small loadings in absolute value),[11] and the output can here be governed by one or several hyperparameters, allowing for more flexibility and a finer control. When estimated with sparse PCA instead of (unrotated) PCA, the factors are not necessarily ranked by decreasing proportion of explained variance (like rotated factors), and can be correlated (like obliquely rotated factors). Moreover, Kristensen (2017) proved that sparse PCA, initially developed for cross-sectional data, still provides consistent estimates of the factors in the case of a DFM (under appropriate conditions on the amount of shrinkage).

There has been extensive research in machine learning about sparse PCA, leading to several 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 in formulating PCA as a regression-type problem, and adding an elastic net penalty (Zou and Hastie, 2005), which 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 automatic variable selection. Nevertheless it comes up with limitations, especially when the variables are highly correlated (which is obviously the case here with macroeconomic and financial variables). When $T > N$, if there are high correlations between the variables, the prediction performance of the lasso tends to be dominated by ridge, and thus also by the elastic net (Zou and Hastie, 2005). 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 does not care which one. 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). We consider the grouping effect as crucial for a macroeconomic factor model, since it precisely considers groups of highly correlated variables (economic activity variables, prices, interest rates, etc.). SPCA is one of the very few sparse PCA algorithms using a strictly convex penalty.

Zou, Hastie and Tibshirani (2006) start by noting that the formulation (7) of PCA amounts to solve the following problem:

$$\min_{V} \|X - XVV'\|_F^2 \quad \text{s.t.} \quad V'V = I_r, \tag{14}$$

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

whose solution $\widehat{V}$ is the $N \times r$ matrix containing the eigenvectors corresponding to the r largest eigenvalues of the data covariance matrix $X'X/T$. They then add an elastic net penalty to (14) in order to produce *modified* principal components, possibly sparse. This leads to the following regression-type optimization problem:

$$\min_{U,V} \|X - XVU'\|_F^2 + \sum_{k=1}^{r} \kappa_{1k}\|V_k\|_1 + \kappa_2 \sum_{k=1}^{r} \|V_k\|_2^2 \;\text{ s.t. }\; U'U = I_r. \tag{15}$$

The kth column of the new $\widehat{V}$, denoted $\widehat{V}_k$, is the kth modified principal component, κ_{1k} the ℓ_1 penalty hyperparameter controlling the level of sparsity in the kth modified principal component, and κ_2 the ℓ_2 penalty hyperparameter. If desired, the κ_{1k} hyperparameters can be distinct so as to impose a substantially different level of sparsity in each modified principal component.

Knowing the hyperparameters κ (whose tuning is explained below), and after initializing U with the first r ordinary principal components, the optimization is performed via a block descent: (15) is alternatively minimized for V given U (with an elastic net regression), and for U given V (with a reduced rank Procrustes rotation), until convergence.[12] As with ordinary PCA under normalization A, the estimated loadings and factors are obtained with $\widehat{\Lambda}_{SPCA} = \widehat{V}\widehat{D}^{1/2}$ and $\widehat{F}_{SPCA} = X\widehat{V}\widehat{D}^{-1/2}$, but now $\widehat{V}$ and $\widehat{\Lambda}$ are sparse, and $\hat{d}_k = \|X\widehat{V}_k\|_2^2/T$, $\forall k \in \{1, \ldots, r\}$.

Remark 5. With SPCA, like after a rotation (see Remark 2), $\widehat{\Lambda}'_{SPCA}\widehat{\Lambda}_{SPCA}$ is not diagonal, $\widehat{F}'_{SPCA}\widehat{F}_{SPCA}/T \neq I_r$, but $\text{diag}(\widehat{F}'_{SPCA}\widehat{F}_{SPCA}/T) = 1_r$.

Tuning the SPCA hyperparameters κ. First of all, we consider there is no theoretical reason to force some factors to be far more sparse than others in a macroeconomic application.[13] Thus, we fix all the ℓ_1 penalization parameters equal: $\kappa_{1k} = \kappa_1$ $\forall k \in \{1, \ldots, r\}$. This leaves us with the tuning of only two hyperparameters (κ_1 and κ_2) instead of $r + 1$, which drastically reduces the computational cost of the tuning, and also the risk of overfitting.

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, assuming that the subsamples are independent and identically distributed, are invalidated when using time series due to the inherent serial correlation.[14] A time series CV (simulated out-of-sample exercise) would be relevant, but it seems preferable to use an information criterion instead. Smeekes and Wijler (2018) showed that tuning the penalization parameters with the BIC criterion rather than with time series CV leads to a better forecasting performance in the context of macroeconomic forecasting using

[12] See Zou, Hastie and Tibshirani (2006) for more details.

[13] The sparse factors (or at least most of them) are expected to have an economically meaningful interpretation: output factor, labor market factor, prices factor, etc. There is no theoretical reason for the labor market factor to be built from far less variables than the prices factor or the output factor for example.

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

penalized regression methods (ridge regression, lasso, adaptive lasso, elastic net, and adaptive elastic net). Besides, the estimation time is much smaller.

We chose to tune the hyperparameters by minimizing a BIC-type criterion, as Kristensen (2017) did with the unique tuning parameter of another sparse PCA algorithm (sPCA-rSVD, Shen and Huang, 2008):

$$(\widehat{\kappa}_1, \widehat{\kappa}_2) = \underset{\kappa_1,\kappa_2}{\operatorname{argmin}} \log\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[x_{it} - \widehat{\lambda}_{SPCA,i}(\kappa_1,\kappa_2)'\widehat{F}_{SPCA,t}(\kappa_1,\kappa_2)\right]^2\right) + m(\kappa_1,\kappa_2)\frac{\log(NT)}{NT} \tag{16}$$

where m is the number of non-zero loadings in $\widehat{\Lambda}_{SPCA}$. Obviously, m, $\widehat{\Lambda}_{SPCA}$ and $\widehat{F}_{SPCA}$ depend on (κ_1, κ_2). We solve (16) via a grid search, by minimizing the BIC-type criterion over the following grid:

$$\kappa_1 \in \{0, 0.1, \ldots, 1\},$$
$$\kappa_2 \in \{0, 0.1, \ldots, 1\}.$$

3 Monte Carlo simulations

3.1 Design

In this section, we run Monte Carlo simulations to assess the performance of the different estimators in recovering the structure of the loadings. In our experiments, we consider two structures: dense and by block. The data-generating process (DGP) for each case is detailed below.

$$x_{it} = \lambda_i' F_t + \sqrt{\theta} e_{it} = \sum_{j=1}^{r} \lambda_{ij} f_{jt} + \sqrt{\theta} e_{it} \tag{17}$$

$$F_t = \Phi_F F_{t-1} + u_t,\ \ \Phi_F = \phi_f I_r,\ \ u_t \overset{iid}{\sim} \mathcal{N}(0_r, \Sigma_u) \tag{18}$$

$$e_{it} = \phi_e e_{i,t-1} + \epsilon_{it},\ \ \epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{19}$$

This DGP can generate serially and cross-sectionally correlated factors, and serially correlated errors. All of the experiments are repeated 1,000 times. We use $N = 60$ and $T = 100$ in our baseline specification (like Kaufmann and Schumacher, 2017, who study a related issue). We set $r = 3$, $\phi_f = 0.5$, $\phi_e \in \{0, 0.5\}$. The covariance (and correlation) matrix of the factors is defined as

$$\Sigma_F = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix} \tag{20}$$

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

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

$$\lambda_{ij} \overset{iid}{\sim} \mathcal{N}(0,1). \tag{21}$$

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

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

   with Λ_{ij} being vectors of length 20 and

$$\begin{aligned} \Lambda_{ij} &\overset{iid}{\sim} \mathcal{N}(\mu_l.1_{20}, \sigma_l^2.I_{20}) \quad \text{if} \quad i = j, \\ \Lambda_{ij} &\overset{iid}{\sim} \mathcal{N}(\mu_s.1_{20}, \sigma_s^2.I_{20}) \quad \text{if} \quad i \neq j. \end{aligned} \tag{23}$$

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 $2/3$. In order to be able to compare the results between the dense structure and the block structure, the expected norm of the loadings should be the same in both cases. To do that, μ_l is deduced from μ_s, σ_l^2, and σ_s^2 such that $E(\Lambda_k'\Lambda_k)/N = 1$ like with the dense structure. Finally, for both the dense structure and the block structure cases, θ is set such that the common component and the idiosyncratic component both explain half of the total variation in X_t when the factors are uncorrelated.

Before computing the metrics assessing the estimation precision, the estimated loadings and factors need to undergo a three-step postprocessing.

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

$$\dot{\widehat{\Lambda}}_k = \sqrt{N} \frac{\widehat{\Lambda}_k}{\|\widehat{\Lambda}_k\|_2} \quad \forall k \in \{1, \ldots, r\} \tag{24}$$

   We have $\dot{\widehat{\Lambda}}_k{}' \dot{\widehat{\Lambda}}_k / N = 1$, which is equal to $E(\Lambda_k'\Lambda_k)/N$. The factors do not need to be rescaled for the metrics we use in the following.

2. Since F and Λ are identified up to column permutations, when necessary, the columns of $\widehat{F}$ and $\dot{\widehat{\Lambda}}$ are reordered to correspond to the ordering in F and Λ. The columns (taken in absolute value) with the smallest distance according to the ℓ_2 norm are iteratively associated. The order of the 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 estimation. 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 precision is measured with the root mean squared error (RMSE) and the mean absolute error (MAE).

$$RMSE(\widetilde{\Lambda}) = \sqrt{\frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{r}(\lambda_{ij}-\widetilde{\lambda}_{ij})^2} \tag{25}$$

$$MAE(\widetilde{\Lambda}) = \frac{1}{Nr}\sum_{i=1}^{N}\sum_{j=1}^{r}|\lambda_{ij}-\widetilde{\lambda}_{ij}| \tag{26}$$

For the factors, this assessment is done with the correlations among the postprocessed estimated factors (denoted $\widehat{\rho}$), and the trace R^2 statistic for the multivariate regression of the estimated factors onto the true factors.

$$R^2_{\widehat{F},F} = \frac{\mathrm{Tr}(F'\widehat{F}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'F)}{\mathrm{Tr}(F'F)} \tag{27}$$

A trace R^2 closer to one implies better performance of the factors estimates. This metric is invariant to the chosen rotation and the estimation postprocessing.

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 comparable results. PCA varimax, PCA quartimin and SPCA do not fallaciously recover a sparse or near-sparse structure in the loadings. They even tend to slightly improve upon standard PCA, especially when the factors are correlated. The factor space is precisely estimated: the trace R^2s are close to one. With PCA quartimin and SPCA, the correlations among factors are very accurately estimated when they are uncorrelated, but are underestimated in absolute value when they are correlated.

Table 1: Dense structure in the loadings with uncorrelated factors.

Estimation method	ϕ_e	RMSE	MAE	$\widehat{\rho}$	$R^2_{\widehat{F},F}$
PCA	0	0.65	0.52	(0,0,0)	0.90
PCA varimax	0	0.63	0.51	(0,0,0)	0.90
PCA quartimin	0	0.63	0.51	(0,0,0)	0.90
SPCA	0	0.66	0.53	(0,-0.01,0.01)	0.89
PCA	0.5	0.67	0.54	(0,0,0)	0.89
PCA varimax	0.5	0.66	0.53	(0,0,0)	0.89
PCA quartimin	0.5	0.66	0.53	(0.01,0,0)	0.89
SPCA	0.5	0.68	0.54	(0,0,0)	0.89

Table 2: Dense structure in the loadings with correlated factors.

Estimation method	ϕ_e	RMSE	MAE	$\widehat{\rho}$	$R^2_{\widehat{F},F}$
PCA	0	0.72	0.58	(0,0,0)	0.89
PCA varimax	0	0.67	0.54	(0,0,0)	0.89
PCA quartimin	0	0.65	0.52	(-0.09,0.04,0.01)	0.89
SPCA	0	0.70	0.56	(-0.13,0.07,0.02)	0.89
PCA	0.5	0.74	0.59	(0,0,0)	0.89
PCA varimax	0.5	0.69	0.56	(0,0,0)	0.89
PCA quartimin	0.5	0.68	0.55	(-0.09,0.04,0.01)	0.89
SPCA	0.5	0.72	0.57	(-0.11,0.05,0.02)	0.89

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 of the structure in the loadings. As with a dense structure, PCA varimax and PCA quartimin lead to extremely similar results when the factors are uncorrelated, but PCA quartimin now dominates PCA varimax more clearly when the factors are correlated. Rotated PCA outperforms SPCA when the structure is near-sparse, but it is the opposite when the structure is sparse. However, the magnitude of the performance improvement with SPCA when the structure is sparse should be read in the light of the high level of sparsity in the model in this case. Concerning the factors, the four estimation methods lead to similar trace R^2s, close to one. The correlations among factors are accurately recovered with SPCA, though slightly underestimated in absolute value. The underestimation is a bit more substantial with PCA quartimin (especially when the structure is not sparse), but the results are still satisfactory.

Table 3: Block structure in the loadings with uncorrelated factors.

Estimation method	σ_l^2	σ_s^2	ϕ_e	RMSE	MAE	$\widehat{\rho}$	$R^2_{\widehat{F},F}$
PCA	0.2	0	0	0.67	0.57	(0,0,0)	0.92
PCA varimax	0.2	0	0	0.23	0.18	(0,0,0)	0.92
PCA quartimin	0.2	0	0	0.22	0.17	(0,0,0)	0.92
SPCA	0.2	0	0	0.15	0.07	(0,0,0)	0.93
PCA	0.2	0.2	0	0.65	0.55	(0,0,0)	0.92
PCA varimax	0.2	0.2	0	0.24	0.20	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0	0.24	0.19	(0,0,0)	0.92
SPCA	0.2	0.2	0	0.30	0.24	(0,0,0)	0.92
PCA	0.5	0.2	0	0.66	0.54	(0,0,0)	0.91
PCA varimax	0.5	0.2	0	0.29	0.23	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0	0.28	0.22	(0,0,0)	0.91
SPCA	0.5	0.2	0	0.32	0.26	(0,0,0)	0.91
PCA	0.2	0	0.5	0.68	0.58	(0,0,0)	0.92
PCA varimax	0.2	0	0.5	0.26	0.21	(0,0,0)	0.92
PCA quartimin	0.2	0	0.5	0.25	0.20	(0,0,0)	0.92
SPCA	0.2	0	0.5	0.17	0.08	(0.01,0,0)	0.93
PCA	0.2	0.2	0.5	0.67	0.56	(0,0,0)	0.92
PCA varimax	0.2	0.2	0.5	0.28	0.22	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0.5	0.27	0.22	(0,0,0)	0.92
SPCA	0.2	0.2	0.5	0.32	0.26	(0,0,0)	0.92
PCA	0.5	0.2	0.5	0.67	0.55	(0,0,0)	0.91
PCA varimax	0.5	0.2	0.5	0.32	0.26	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0.5	0.32	0.25	(0,0,0)	0.91
SPCA	0.5	0.2	0.5	0.35	0.28	(0,0,0)	0.92

Table 4: Block structure in the loadings with correlated factors.

Estimation method	σ_l^2	σ_s^2	ϕ_e	RMSE	MAE	$\widehat{\rho}$	$R^2_{\widehat{F},F}$
PCA	0.2	0	0	0.72	0.61	(0,0,0)	0.92
PCA varimax	0.2	0	0	0.29	0.23	(0,0,0)	0.92
PCA quartimin	0.2	0	0	0.23	0.18	(-0.32,0.16,0.01)	0.92
SPCA	0.2	0	0	0.16	0.07	(-0.38,0.18,0)	0.93
PCA	0.2	0.2	0	0.71	0.60	(0,0,0)	0.92
PCA varimax	0.2	0.2	0	0.30	0.24	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0	0.26	0.21	(-0.23,0.11,0.01)	0.92
SPCA	0.2	0.2	0	0.32	0.26	(-0.35,0.18,0.01)	0.92
PCA	0.5	0.2	0	0.72	0.60	(0,0,0)	0.91
PCA varimax	0.5	0.2	0	0.34	0.27	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0	0.30	0.24	(-0.23,0.11,0.01)	0.91
SPCA	0.5	0.2	0	0.34	0.27	(-0.36,0.18,0.01)	0.91
PCA	0.2	0	0.5	0.73	0.62	(0,0,0)	0.92
PCA varimax	0.2	0	0.5	0.32	0.26	(0,0,0)	0.92
PCA quartimin	0.2	0	0.5	0.27	0.21	(-0.30,0.15,0.01)	0.92
SPCA	0.2	0	0.5	0.19	0.09	(-0.38,0.19,0)	0.93
PCA	0.2	0.2	0.5	0.72	0.60	(0,0,0)	0.92
PCA varimax	0.2	0.2	0.5	0.33	0.26	(0,0,0)	0.92
PCA quartimin	0.2	0.2	0.5	0.30	0.24	(-0.21,0.11,0.01)	0.92
SPCA	0.2	0.2	0.5	0.35	0.28	(-0.33,0.17,0.02)	0.92
PCA	0.5	0.2	0.5	0.74	0.61	(0,0,0)	0.91
PCA varimax	0.5	0.2	0.5	0.37	0.29	(0,0,0)	0.91
PCA quartimin	0.5	0.2	0.5	0.34	0.27	(-0.22,0.11,0.01)	0.91
SPCA	0.5	0.2	0.5	0.38	0.30	(-0.33,0.17,0.01)	0.91

For each DGP and estimation method, the introduction of serial correlation in the idiosyncratic errors only marginally deteriorates the estimation of the loadings and factors. 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 force the estimated factors to be uncorrelated.

In brief, rotated PCA and SPCA slightly improve upon PCA in recovering the correct factor structure if that structure is dense, and vastly if that structure is sparse or near-sparse. We consider the quartimin rotation and SPCA as the preferable approaches to recover the correct factor structure (at least among the methods considered here).

In our simulations, PCA quartimin performs as well or better than PCA varimax, and there is usually no reason to impose the factors to be uncorrelated as it is done with orthogonal rotations. SPCA outperforms PCA quartimin (and PCA varimax) when the factor structure is sparse, but this comes at several expenses. First, the computational cost of SPCA can be considerable, which can be a serious inconvenience in some applications, especially forecasting exercises. In our Monte Carlo study, the 1,000 repetitions of the experiment for a given DGP requires less than one minute for PCA quartimin, versus several days for SPCA. Another problem for forecasting applications is that SPCA can not handle missing observations per se, and its computational cost tends to discourage the use of an EM algorithm to treat them.[15] On the opposite, a factor rotation can be applied after estimation methods which efficiently handle missing observations.[16] Finally, sparse PCA implies a small loss of explained variance with respect to PCA (rotated or not). In view of this, the quartimin rotation can still be a good compromise over SPCA when the structure is sparse.

[15] 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.

[16] PCA, via an EM algorithm (Stock and Watson, 2002b); the two-step estimator, via the Kalman filter (Doz, Giannone and Reichlin, 2011); the quasi-maximum likelihood estimator, via an EM algorithm and the Kalman filter (Doz, Giannone and Reichlin, 2012); etc.

4 Empirical application

4.1 Data

Macroeconomic factor models are usually implemented with a monthly or quarterly database of around 100 macroeconomic and financial variables. For the U.S., the historical reference dataset for macroeconomic factor models is the so-called Stock-Watson dataset (Stock and Watson, 2002a,b). 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). It 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[17] using mostly the time series from the Federal Reserve Economic Data (FRED), the St. Louis Fed's main publicly available economic database. Resulting vintage databases are all available on its website.[18] The variables are classified in 8 groups: 1) output and income, 2) labor market, 3) housing, 4) consumption, orders and inventories, 5) money and credit, 6) interest rates and exchange rates, 7) prices, and 8) stock market.

We proceed with its 2018:9 version: 128 variables, covering the 1959:1-2018:7 time span. In order to get a balanced panel, the analysis is restricted to 1960:1-2018:4, with 123 variables (see Appendix A for further details). The series were stationarized using the transformations recommended by McCracken and Ng (2016), then standardized. We do not treat outliers[19] because this may remove some structural breaks (Breitung and Eickmeier, 2011).

4.2 Tuning the hyperparameters

Number of factors r. The standard method to select the number of (static) factors to include in a factor model was proposed by Bai and Ng (2002). They developed a family of information criteria minimizing the variance of the idiosyncratic component e, subject to a penalty depending on N and T. They showed that for both a large N and a large T their criteria lead to a consistent estimation of the number of factors under a standard set of assumptions, even in case of heteroscedasticity (along time or across variables), autocorrelation, or correlation across the idiosyncratic terms. The selection of r is performed by applying the six main criteria of Bai and Ng (2002) over the full balanced panel 1960:1-2018:4, with $r_{max} = 10$. We choose to select $r = 8$, as McCracken and Ng (2016)[20] and Kristensen (2017) did when considering their full sample.

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

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

[19] In the macroeconomic factor models literature, an observation is usually defined as an outlier if it deviates from the sample median by more than ten interquartile ranges.

[20] McCracken and Ng (2016) base their choice on the PC_{p2} criterion, which finds 8 factors when an outlier adjustment is performed, and 9 factors otherwise (just like here). They prefer to treat the outliers for their

Table 5: Selected number of factors.

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

SPCA hyperparameters κ_1 and κ_2. Considering the full balanced panel and setting $r = 8$, the grid search yields $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.6)$.

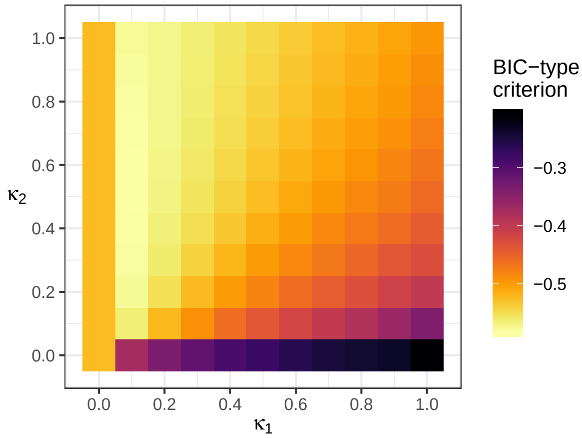


Figure 1: Grid search of the SPCA hyperparameters κ_1 and κ_2.

4.3 Interpreting the factors

As routinely reported, the factors estimated with PCA are very difficult to interpret. When estimating the factors and loadings with the full balanced panel, it appears that only three of the eight factors can be given an economic interpretation (see Figure 2). Factor 4 can be considered as the interest rates factor (even if the housing and stock market variables seem also quite important). Factor 7 is obviously the stock market factor. Factor 8 is the money and credit factor. Some variables outside the money and credit group have a quite strong loading, 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 the labor market group), real personal consumption expenditures, retail and food services sales (the two largest loadings in the consumption, orders and inventories group). All the other estimated factors are impossible to interpret.

empirical analysis, so they select 8 factors.

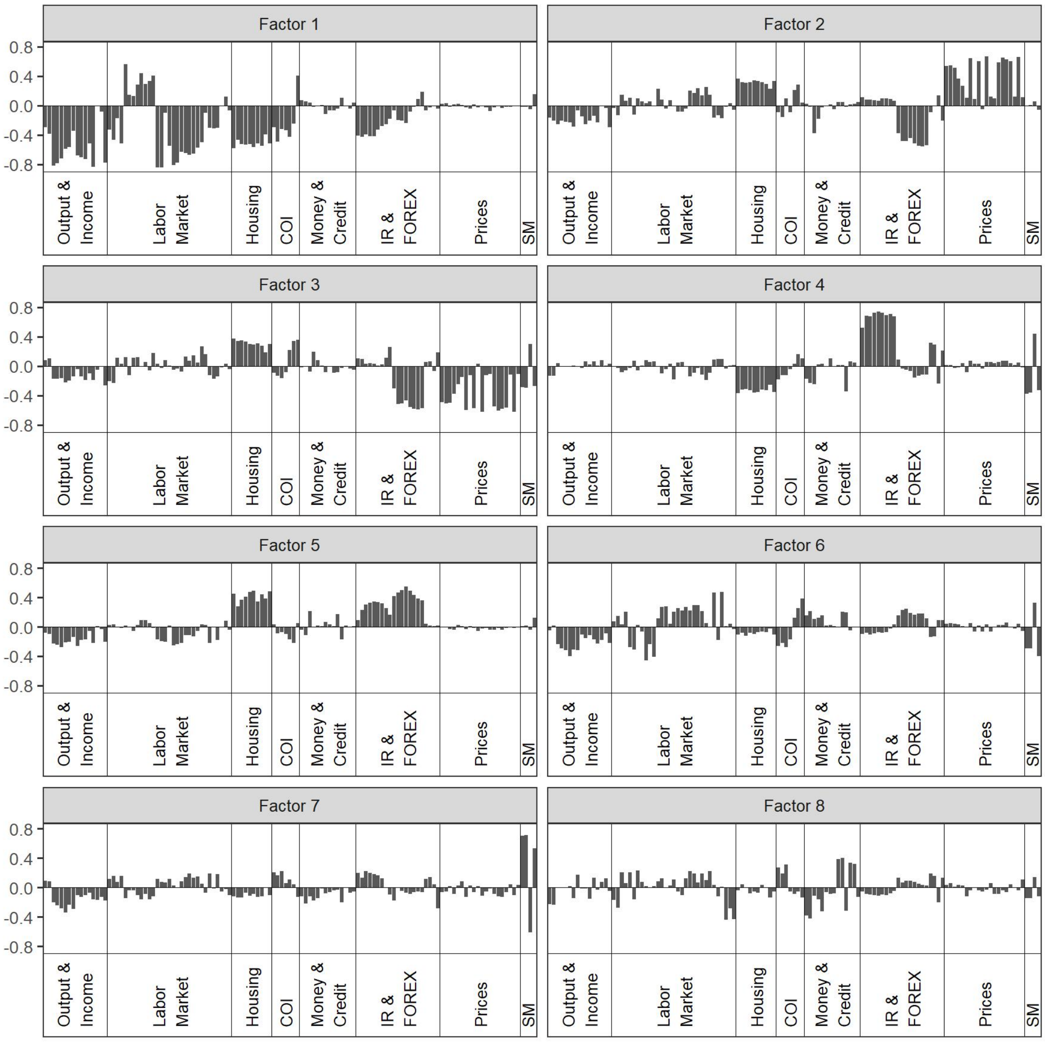


Figure 2: PCA factor loadings.

On the opposite, the factors estimated with rotated PCA or SPCA systematically offer a straightforward economic interpretation: output, prices, spreads, interest rates, housing, labor, stock market, money and credit. More precisely, they recover the same loadings pattern (Figure 3). We can also notice that the loadings of the only three interpretable (unrotated) PCA estimated factors (interest rates, stock market, money and credit) are very close to the loadings of their rotated and sparse factors counterparts (especially for the last two).

Interestingly, Beyeler and Kaufmann (2019) get very similar interpretations for the 8 U.S. macroeconomic factors of their baseline model, even though they use a different estimation method (Bayesian sparse factor model), and another database (FRED-QD, a quarterly frequency companion to FRED-MD, with a different composition). This reassures the validity of our respective approaches.

It should be stressed that the 8 interpretable factors do not simply consist in recovering the 8 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 variables from the prices, labor, and money and credit groups appear to have very heterogeneous weights in the construction of the prices, labor, and money and credit factors. Moreover, the money and credit factor involves several variables outside the money and credit group.

Since quartimin and varimax lead to the same loadings pattern as SPCA, they of course all lead to almost perfectly identical factors (see Figure 4 and Table 6). Their dynamics is 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 correlations among the factors are also extremely similar for PCA quartimin and SPCA (Figure 5).

The rotated and sparse factors display a very comparable explanatory power (Table 7). Together, the rotated factors explain 47.95% of the dataset variance (exactly like the unrotated factors, by construction), versus 42.59% for the sparse factors. With unrotated PCA, the share of variance accounted by the first factor is large (almost 15%) and quickly drops for the following factors, whereas the repartition is more homogeneous with the other methods considered. Overall, that repartition is remarkably similar between PCA quartimin, PCA varimax and SPCA (the only discrepancy being the labor factor, whose explanatory power is significantly smaller with SPCA). The three interpretable (unrotated) PCA estimated factors exhibit very similar explanatory power with their rotated and sparse factors counterparts. The relatively weak part of variance explained by the stock market factor can be explained by the very small number of financial variables in FRED-MD (five, one of which is dropped in order to work with a balanced panel).

These results suggest that the standard principal components estimator of the factors and loadings, i.e. the unrotated solution, while computationally convenient, recovers a quite unnatural factor representation. As soon as a method enforcing a simple structure in the loadings is used (no matter the method), it yields same factor representation, with a simple structure in the loadings offering a clear economic interpretation of the factors. It is quite remarkable that a state-of-the-art machine learning technique and such long-established statistical methods as factor rotations lead to the same results for this kind of data. This factor representation seems more natural and informative than the unrotated solution, and should probably be preferred as reference in DFM applications.

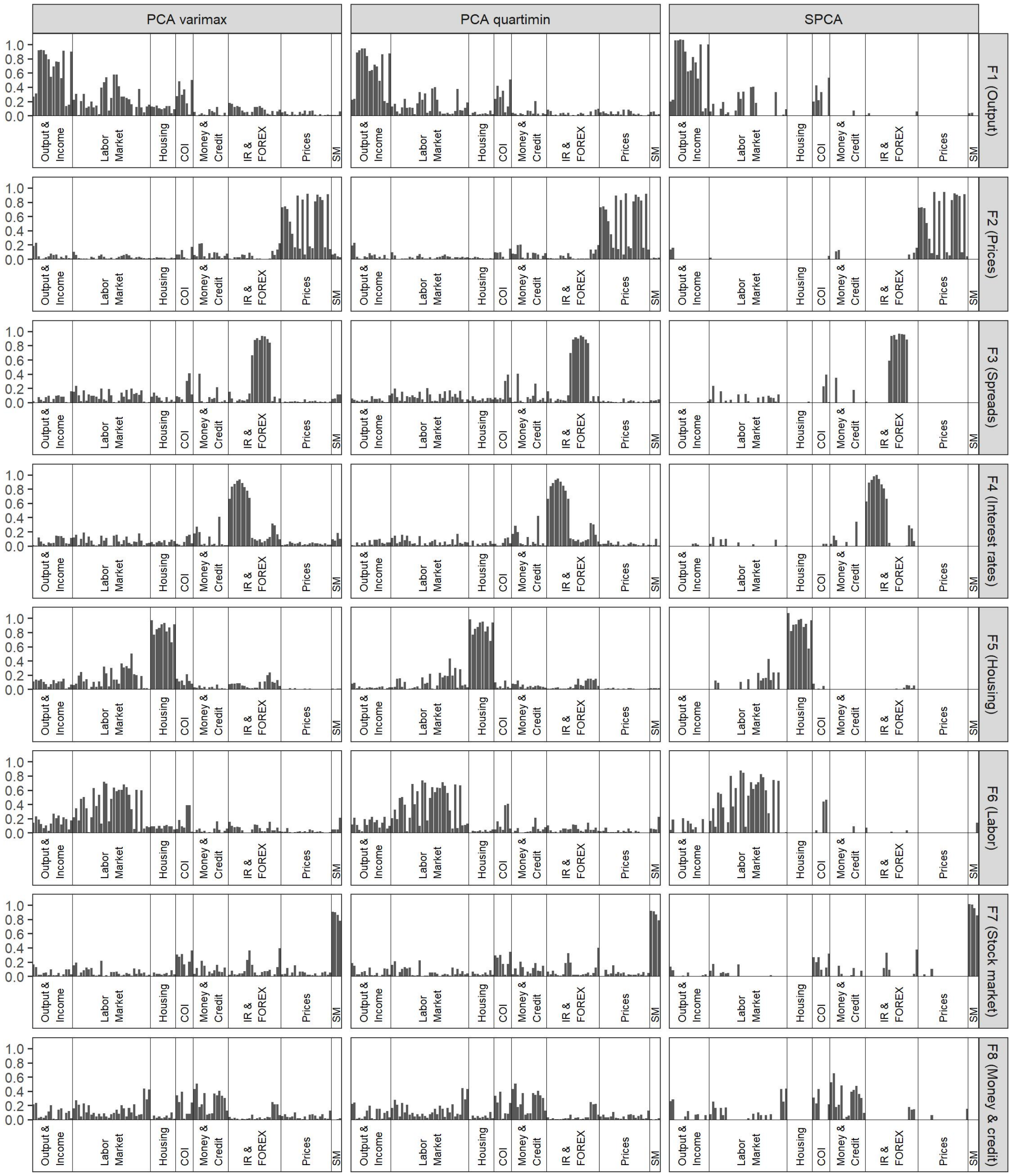


Figure 3: Comparison of the factor loadings of the rotated and sparse factors.

Notes: Remember that the loadings are still identified only up a column sign change. Thus, in order to facilitate the comparisons, the estimated loadings are displayed in absolute value. The proportion of non-zero loadings for the factors estimated with SPCA is, respectively: 0.64, 0.77, 0.75, 0.73, 0.74, 0.65, 0.71, 0.63.

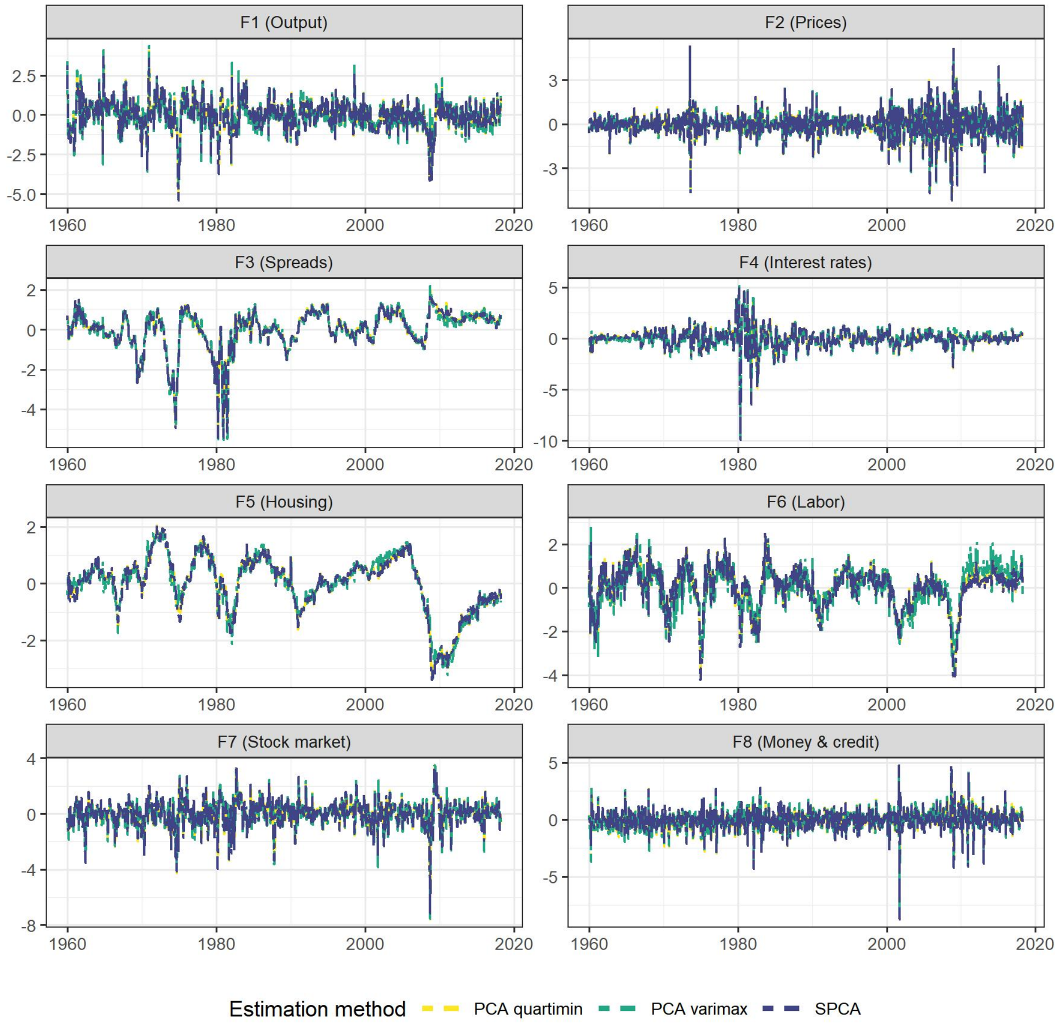


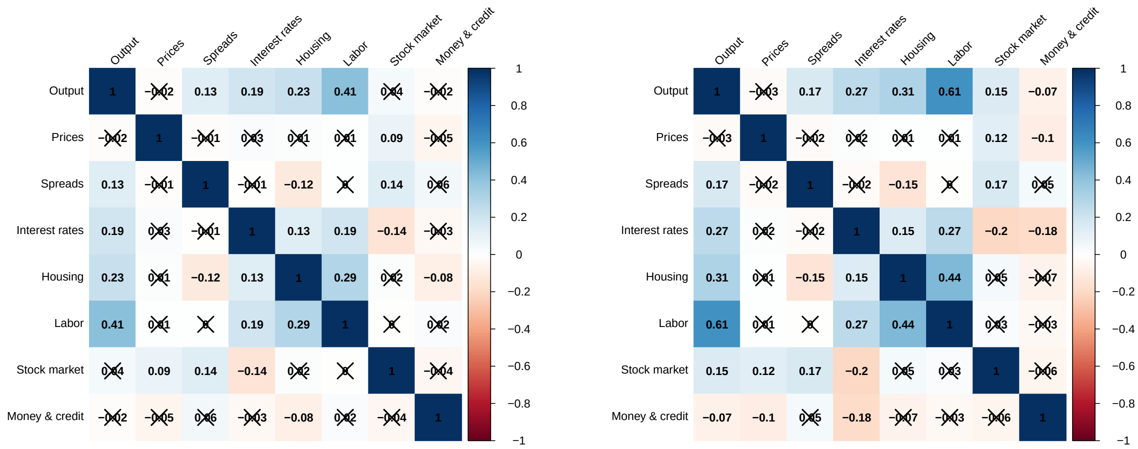
Figure 4: Comparison of the rotated and sparse factors.

Notes: Remember that the factors are identified up to a sign change. We reversed the sign of some factors in order to make their evolution more consistent with their interpretation (e.g. the output and labor factors should collapse in the wake of the 2007-09 financial crisis, not rise). The sign of the following rotated PCA factors is reversed: F1 (Output), F3 (Spreads) and F8 (Money & credit). For the SPCA factors: F1 (Output), F2 (Prices), F4 (Interest rates), F6 (Labor), and F7 (Stock market).

Table 6: Correlations among the estimated factors, depending on the estimation method.

Factors	PCA varimax - PCA quartimin	PCA varimax - SPCA	PCA quartimin - SPCA
Output	0.99	0.95	0.99
Prices	1.00	1.00	1.00
Spreads	0.99	0.99	1.00
Interest rates	0.99	0.98	1.00
Housing	0.99	0.97	1.00
Labor	0.93	0.87	0.99
Stock market	0.99	0.96	0.98
Money & credit	0.99	0.95	0.96

Notes: See Figure 4.



(a) PCA quartimin (b) SPCA

Figure 5: Correlations among the factors.

Notes: See Figure 4. The correlation coefficients which are not significant at the 5% level are crossed out. The correlation matrices of the factors estimated with PCA and PCA varimax are not displayed, since they are simply the identity matrix.

Table 7: Percentage of the dataset total variance explained by the factors, depending on the estimation method.

Factor	PCA	Factor	PCA varimax	PCA quartimin	SPCA
1 (?)	14.77	1 (Output)	9.45	8.49	9.14
2 (?)	7.27	2 (Prices)	7.06	7.05	6.88
3 (?)	7.05	3 (Spreads)	5.85	5.83	5.51
4 (Interest rates)	5.67	4 (Interest rates)	5.92	5.98	5.30
5 (?)	4.31	5 (Housing)	7.44	7.24	6.18
6 (?)	3.44	6 (Labor)	6.11	7.15	4.06
7 (Stock market)	2.96	7 (Stock market)	3.55	3.61	3.15
8 (Money & credit)	2.47	8 (Money & credit)	2.56	2.60	2.37
Total	47.95	Total	47.95	47.95	42.59

Notes: Remember that, by construction, the factors estimated with (unrotated) PCA are ranked by decreasing proportion of explained variance, contrary to the factors estimated with rotated PCA or SPCA.

5 Relation with the structural instability issue

5.1 Recursive estimation exercise

In view of the recent results concerning the structural instability in DFMs, we can wonder about the stability over time of the interpretability (and interpretation) of the factors when using the previous factor representation. To study this issue, we use a recursive estimation of the model (as it is done in most forecasting applications). We start the recursive estimation in 1975:1, i.e. we successively estimate the model with the subsamples 1960:1-1975:1, 1960:1-1975:2,. . . , 1960:1-2018:4. For the sake of clarity, the number of estimated factors is kept fixed in this exercise. The results with PCA varimax, PCA quartimin and SPCA are still extremely similar (see Appendix B), so we only present the results obtained with PCA quartimin in the following.

It appears that the recursively estimated loadings parallel the previous results concerning the factor interpretability (Figure 7). For every period, all the factors have a clear economic interpretation, which is the same as before when considering the full sample. Indeed, the loadings exhibit a surprisingly high temporal stability. This is especially visible with the prices factor, switching between columns 2, 3 and 4 over time.[21]

Working with this factor representation sheds a whole new light on the structural instability issue. We cannot distinguish any widespread break in the loadings at any date. Instead, we witness a global stability, combined with several breaks affecting a limited number of loadings. These breaks can be localized and interpreted. In addition, they are coherent with economic history. We can notice three break dates on Figure 7.

Following the Volcker Shock of 1980, a break occurs in some loadings of the interest rates factor. There is an increase in the weight of the Effective Federal Funds Rate and the Moody's Seasoned Baa Corporate Bond Yield, and to a lesser extent, of the 3-Month AA Financial Commercial Paper rate and the Moody's Seasoned Aaa Corporate Bond Yield.

Similarly, the expansionary monetary policy started in 2001 (as an answer to the burst of the dot-com bubble in 2000, the subsequent recession, and the consequences of the 09/11 terrorist attacks) induces a break in some loadings of the money and credit factor. Before this monetary policy shock, the money and credit factor was rather a factor of *demand for money and credit*. Indeed, it was built from medium loadings for the variables in the money and credit group, and large loadings for variables corresponding to determinants of the demand for money and credit (average hourly earnings in the goods-producing, construction, and manufacturing sectors, retail and food services sales). Following the expansionary monetary policy shock, it turns into a factor of *supply of money*. The loadings of the demand determinants collapse, while the loadings of

[21] Remember that Λ is identified up to column permutations. We did not reorder the columns according to the factor interpretations.

several variables measuring the supply of money erupt (M1SL, M2SL, AMBSL, TOTRESNS, NONBORRES).

The 2007-09 financial crisis and its aftermath generate breaks in the loadings of the money supply factor, and of the housing factor. The former now becomes a more general factor of demand and supply of money and credit, with medium loadings for both the determinants of the demand for money and credit, and the variables measuring the supply of money. Concerning the housing factor, the variables concerning the Northeast and the Midwest now tend to be as important as the variables concerning the South and the West.

In addition, there is no evidence for the emergence of an additional factor. All eight factors are indeed associated with non-zero loadings for all the subsamples, and the part of the dataset total variance explained by the factors remains remarkably stable over time (Figure 6). Since the number of factors has been fixed, if a new factor appeared at a given date, it would end up in the idiosyncratic term and cause a drop in the fraction of explained variance. This is in line with the results of Stock and Watson (2012). Focusing on the 2007–09 recession and its aftermath, they find no evidence for the emergence of an additional factor (but rather for a change in the dynamics of the factors).

In short, the true instability in the DFM seems to consist in breaks in a limited number of loadings, which can be localized and interpreted. These results have important implications. They reassure the robustness of the factor model and the representation it gives of the economy. Here, the dynamics of the variables change over time (and thus so do the dynamics of the factors), but the way the factors are built stays fairly stable.

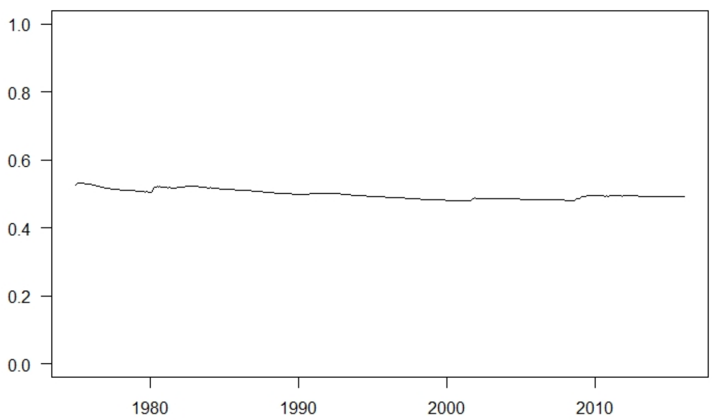


Figure 6: Time evolution of the part of the dataset total variance explained by the PCA estimated factors (rotated or not).

Notes: The dates on the x-axis correspond to the end of the estimation sample. When using the full sample (last point), we simply recover the total in Table 7. Remember that, by construction, the total part of explained variance is identical when considering the factors estimated by rotated or unrotated PCA.

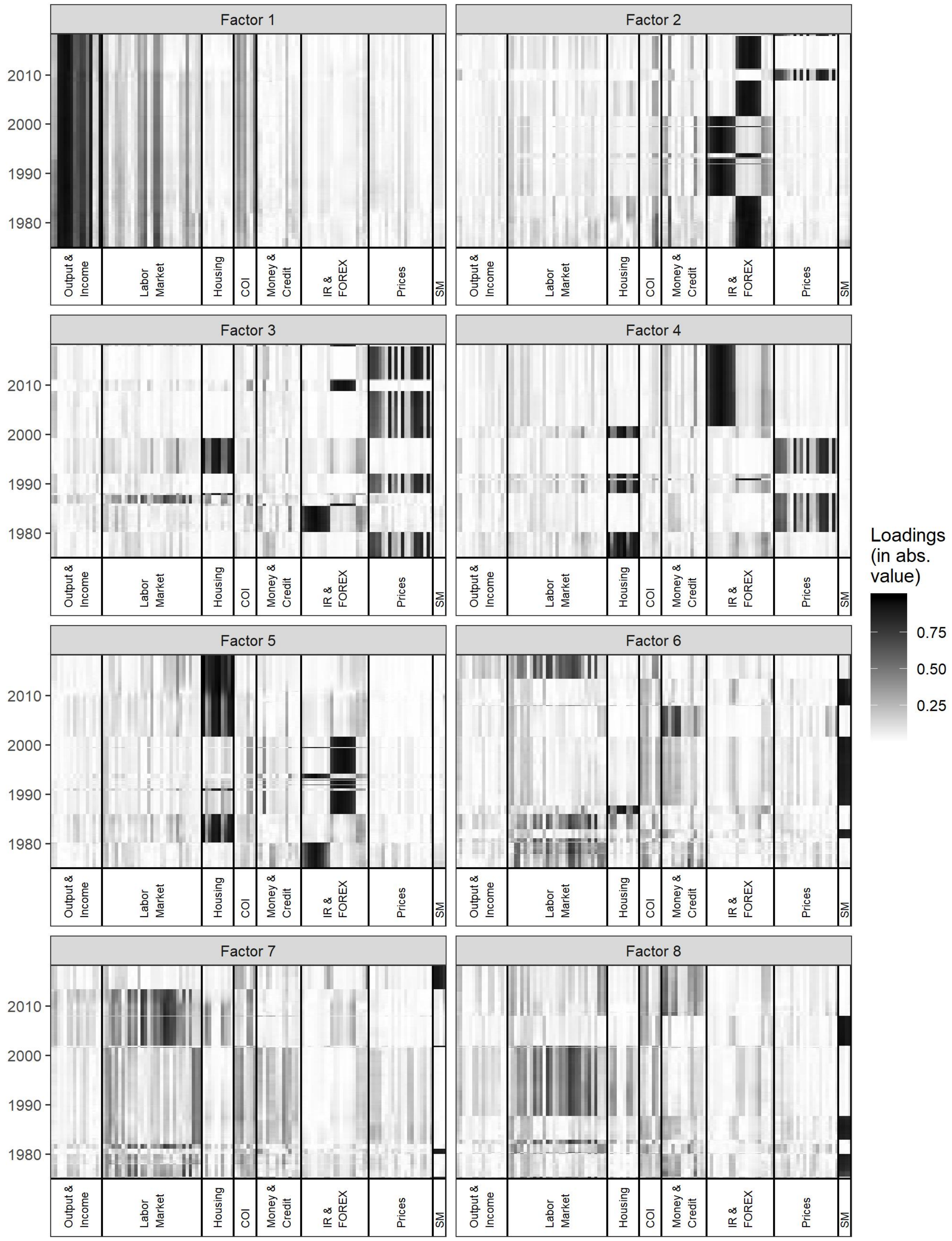


Figure 7: Time evolution of the loadings estimated with PCA quartimin (in absolute value).

Notes: With each factor, for a given date and variable, the larger the loading (in absolute value), the darker the box. The dates on the y-axis correspond to the end of the estimation sample. When using the full sample (top line of each subfigure), we simply recover the loadings from Figure 3. Remember that Λ is identified up to column permutations. We did not reorder the columns according to the factor interpretations.

5.2 Tests

The tests for structural instability in DFMs recently proposed in the literature can offer a second assessment. However, they should be used with caution because at this time very little is known about their power against different forms of instability (Stock and Watson, 2016). Here, we study two main forms of structural instability: global (break in the number of factors and/or widespread break in the loadings), and individual (break in the loadings of a given variable).

Global structural instability. We chose to rely on the test proposed by Chen, Dolado and Gonzalo (2014) due to its performance, its elegant formulation, and its flexibility. Indeed, it can allow for one or several breaks, at possibly unknown dates (and can estimate the break dates when unknown). In the case of a single break with known date (resp. unknown date), it relies on a simple Wald test (resp. a Sup-Wald test, Andrews, 1993). When several breaks with unknown dates are allowed, it relies on a Bai-Perron test (Bai and Perron, 1998, 2003), and the number of breaks is selected by BIC. We chose to allow for several breaks with unknown dates. The results are reported in Table 8. With the factors estimated with (unrotated) PCA, three breaks are detected, corresponding to the two oil crisis and the Great Recession (the estimated break dates are 1971:3, 1979:8, and 2008:3). On the opposite, no break is detected when the PCA factors are rotated using quartimin or varimax, or when the factors are estimated by SPCA. This corroborates the results of the recursive estimation exercise.

Table 8: Results of the Chen-Dolado-Gonzalo test.

Breaks	PCA	PCA varimax	PCA quartimin	SPCA
0	2038.92	**2038.92**	**1867.30**	**1636.69**
1	1936.17	2042.76	1869.64	1646.82
2	1904.61	2064.70	1892.67	1661.46
3	**1904.12**	2078.83	1915.95	1692.21
4	1920.44	2098.52	1949.18	1715.94
5	1953.32	2131.00	1984.66	1753.49

Notes: Bold indicates the smallest BIC for a given estimation method. Three breaks are detected with PCA, but the BICs for two and three breaks are very close. When setting the number of breaks to three (resp. two), the estimated break dates with PCA are 1971:3, 1979:8, and 2008:3 (resp. 1978:4 and 2008:3).

Individual structural instability. We consider the equation

$$x_{it} = \lambda_i' \widehat{F}_t + e_{it}, \tag{28}$$

where $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir})'$, and $\widehat{F}_t = (\widehat{f}_{1t}, \dots, \widehat{f}_{rt})'$ are the factors estimated with PCA quartimin on the full balanced sample. When testing individually the stability of the loadings with a Sup-Wald test (Andrews, 1993), it results that between 17% and 36% of the loadings experience a break (depending on the specification of the test, and at $\alpha = 5\%$). These numbers are in line with what was previously found in the literature. This is not surprising as by construction the factors estimated with rotated or unrotated PCA will yield the same the proportion of rejections for these tests. Nevertheless, the factors are now interpretable, so the sources of the instabilities can be disentangled. It is possible to get a clearer idea regarding which loadings are affected, i.e. which factors and which variables. The most salient example is probably the housing factor in the wake of the subprime crisis (Figure 8). We can notice two main features. Again, the variables concerning the Northeast and the Midwest now tend to be as important as the variables concerning the South and the West. Secondly, the midterm financing conditions in the economy (T5YFFM, T10YFFM), and the financing conditions of the firms (AAAFFM, BAAFFM) gained considerable weight in the construction of the factor summarizing the housing sector. This is of course very consistent with the narratives of the 2007–9 recession.

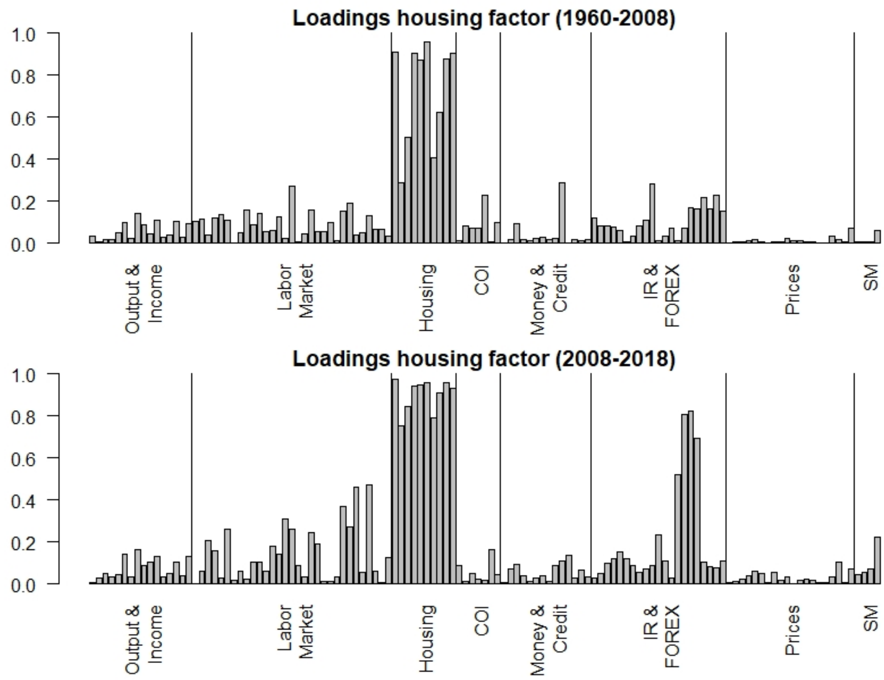


Figure 8: Loadings of the housing factor before and after 2008, estimated by PCA quartimin.

6 Conclusion

When using the principal components estimator of the factors and loadings, the established unrotated solution, while computationally convenient, recovers an uninterpretable factor representation. In this paper, we compare several simple and atheoretical approaches to get an economic interpretation of the factors, leading to almost identical results. As soon as a method enforcing a simple structure in the loadings is used (no matter the method), it yields exactly the same factor representation. This factor representation seems more natural and informative than the unrotated solution. It should probably be preferred as reference when using a factor model. Moreover, the quartimin rotation and sparse PCA (with the SPCA algorithm) seem to be the preferable methods to recover this factor representation.

When working with the interpretable factors, we find no evidence for the emergence of an additional factor or for a widespread break in the factor loadings at a given time. The structural instability in the DFM seems to rather consist in breaks in a limited number of loadings. They can be localized, interpreted, and are coherent with economic history.

We are currently working on two papers using this factor representation to model international business cycles (extending the works of Kose, Otrok and Whiteman, 2003; Del Negro and Otrok, 2008; etc.), and derive DFM forecasts based on interpretable factors.

References

Andrews, D. W. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica* **61**(4), 821–856.

Bai, J., Han, X. and Shi, Y. (forthcoming), 'Estimation and inference of change points in high dimensional factor models', *Journal of Econometrics* .

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

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

Bai, J. and Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**, 47–78.

Bai, J. and Perron, P. (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics* **18**(1), 1–22.

Baltagi, B. H., Kao, C. and Wang, F. (2017), 'Identification and estimation of a large factor model with structural instability', *Journal of Econometrics* **197**(1), 87–100.

Bańbura, M. and Rünstler, G. (2011), 'A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP', *International Journal of Forecasting* **27**(2), 333–346.

Barigozzi, M., Cho, H. and Fryzlewicz, P. (2018), 'Simultaneous multiple change-point and factor analysis for high-dimensional time series', *Journal of Econometrics* **206**(1), 187–225.

Bates, B. J., Plagborg-Møller, M., Stock, J. H. and Watson, M. W. (2013), 'Consistent factor estimation in dynamic factor models with structural instability', *Journal of Econometrics* **177**(2), 289–304.

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

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005), 'Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly Journal of Economics* **120**(1), 387–422.

Beyeler, S. and Kaufmann, S. (2019), 'Reduced-form factor augmented VAR - exploiting sparsity to include meaningful factors'.

Breitung, J. and Eickmeier, S. (2011), 'Testing for structural breaks in dynamic factor models', *Journal of Econometrics* **163**(1), 71–84.

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.

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

Chen, L., Dolado, J. J. and Gonzalo, J. (2014), 'Detecting big structural breaks in large factor models', *Journal of Econometrics* **180**(1), 30–48.

Cheng, X., Liao, Z. and Schorfheide, F. (2016), 'Shrinkage estimation of high-dimensional factor models with structural instabilities', *The Review of Economic Studies* **83**(4), 1511–1543.

Corradi, V. and Swanson, N. R. (2014), 'Testing for structural stability of factor augmented forecasting models', *Journal of Econometrics* **182**(1), 100–118.

Costello, A. B. and Osborne, J. W. (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 Exterkate, P. (2011), Sparse and robust factor modelling, Tinbergen Institute Discussion Papers 11-122/4, Tinbergen Institute.

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

Del Negro, M. and Otrok, C. (2008), 'Dynamic factor models with time-varying parameters: measuring changes in international business cycles', *Federal Reserve Bank of New York Staff Reports No. 326* .

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

Doz, C., Giannone, D. and Reichlin, L. (2012), 'A quasi–maximum likelihood approach for large, approximate dynamic factor models', *Review of Economics and Statistics* **94**(4), 1014–1024.

Eickmeier, S. (2009), 'Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model', *Journal of Applied Econometrics* **24**(6), 933–959.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. and Strahan, E. J. (1999), 'Evaluating the use of exploratory factor analysis in psychological research.', *Psychological Methods* **4**(3), 272.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2009), 'Opening the black box: Structural factor models with large cross sections', *Econometric Theory* **25**(5), 1319–1347.

Giannone, D., Reichlin, L. and Small, D. (2008), 'Nowcasting: The real-time informational content of macroeconomic data', *Journal of Monetary Economics* **55**(4), 665–676.

Gürkaynak, R. S., Sack, B. and Swanson, E. (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.

Han, X. and Inoue, A. (2015), 'Tests for parameter instability in dynamic factor models', *Econometric Theory* **31**(5), 1117–1152.

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 Sampson, P. (1966), 'Rotation for simple loadings', *Psychometrika* **31**(3), 313–323.

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

Jolliffe, I. T., Trendafilov, N. T. and Uddin, M. (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 Schumacher, C. (2017), 'Identifying relevant and irrelevant variables in sparse factor models', *Journal of Applied Econometrics* **32**(6), 1123–1144.

Kaufmann, S. and Schumacher, C. (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 Swanson, N. R. (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., Otrok, C. and Whiteman, C. H. (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* pp. 1–18.

Ma, S. and Su, L. (2018), 'Estimation of large dimensional factor models with an unknown number of breaks', *Journal of Econometrics* **207**(1), 1–29.

Massacci, D. (2017), 'Least squares estimation of large dimensional threshold factor models', *Journal of Econometrics* **197**(1), 101–129.

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

Miranda-Agrippino, S. and Rey, H. (2018), 'US monetary policy and the global financial cycle'.

Sargent, T. J. and Sims, C. A. (1977), 'Business cycle modeling without pretending to have too much a priori economic theory', *New Methods in Business Cycle Research* **1**, 145–168.

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

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

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

Stock, J. H. and Watson, M. (2009), 'Forecasting in dynamic factor models subject to structural instability', *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry* **173**, 205.

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

Stock, J. H. and Watson, M. W. (2012), 'Disentangling the channels of the 2007—09 recession', *Brookings Papers on Economic Activity* **1**, 81–135.

Stock, J. H. and Watson, M. W. (2016), Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, *in* 'Handbook of macroeconomics', Vol. 2, Elsevier, pp. 415–525.

Su, L. and Wang, X. (2017), 'On time-varying factor models: Estimation and testing', *Journal of Econometrics* **198**(1), 84–101.

Trendafilov, N. T. (2014), 'From simple structure to sparse components: a review', *Computational Statistics* **29**(3-4), 431–454.

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

Yamamoto, Y. and Tanaka, S. (2015), 'Testing for factor loading structural change under common breaks', *Journal of Econometrics* **189**(1), 187–206.

Zou, H. and Hastie, T. (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., Hastie, T. and Tibshirani, R. (2006), 'Sparse principal component analysis', *Journal of Computational and Graphical Statistics* **15**(2), 265–286.

A Data

The composition of FRED-MD has slightly changed over time. The 128 variables of its 2018:9 version are listed below. Most of them originate 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 A1: Output and income series.

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 A2: Labor market series.

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 & 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 & 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 A3: Housing series.

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 A4: Consumption, orders, and inventories series.

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 A5: Money and credit series.

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 A6: Interest 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 A7: Prices series.

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 A8: Stock market series.

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

B Recursive estimation exercise: additional results

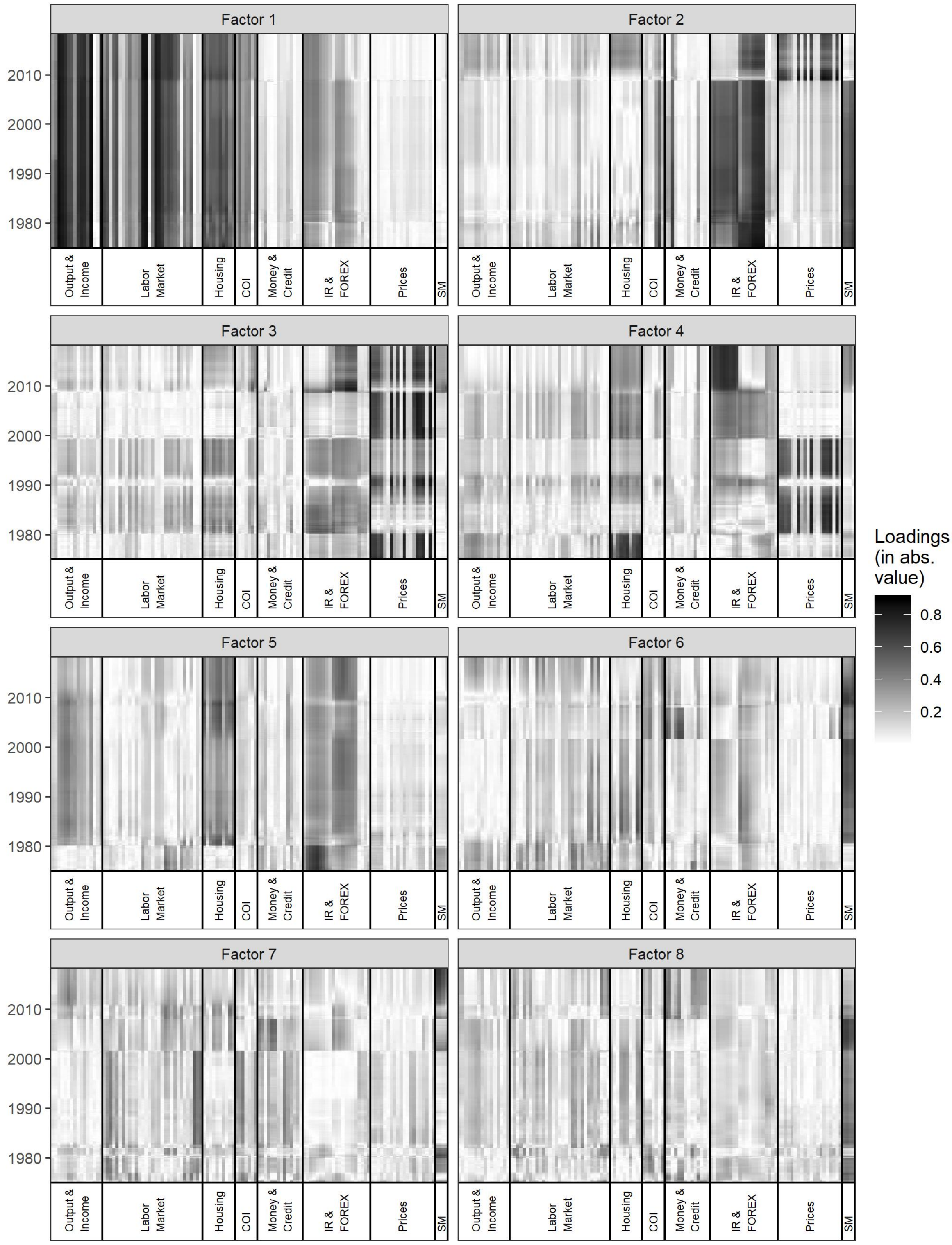


Figure 9: Time evolution of the loadings estimated with PCA (in absolute value).

Notes: See Figure 7.

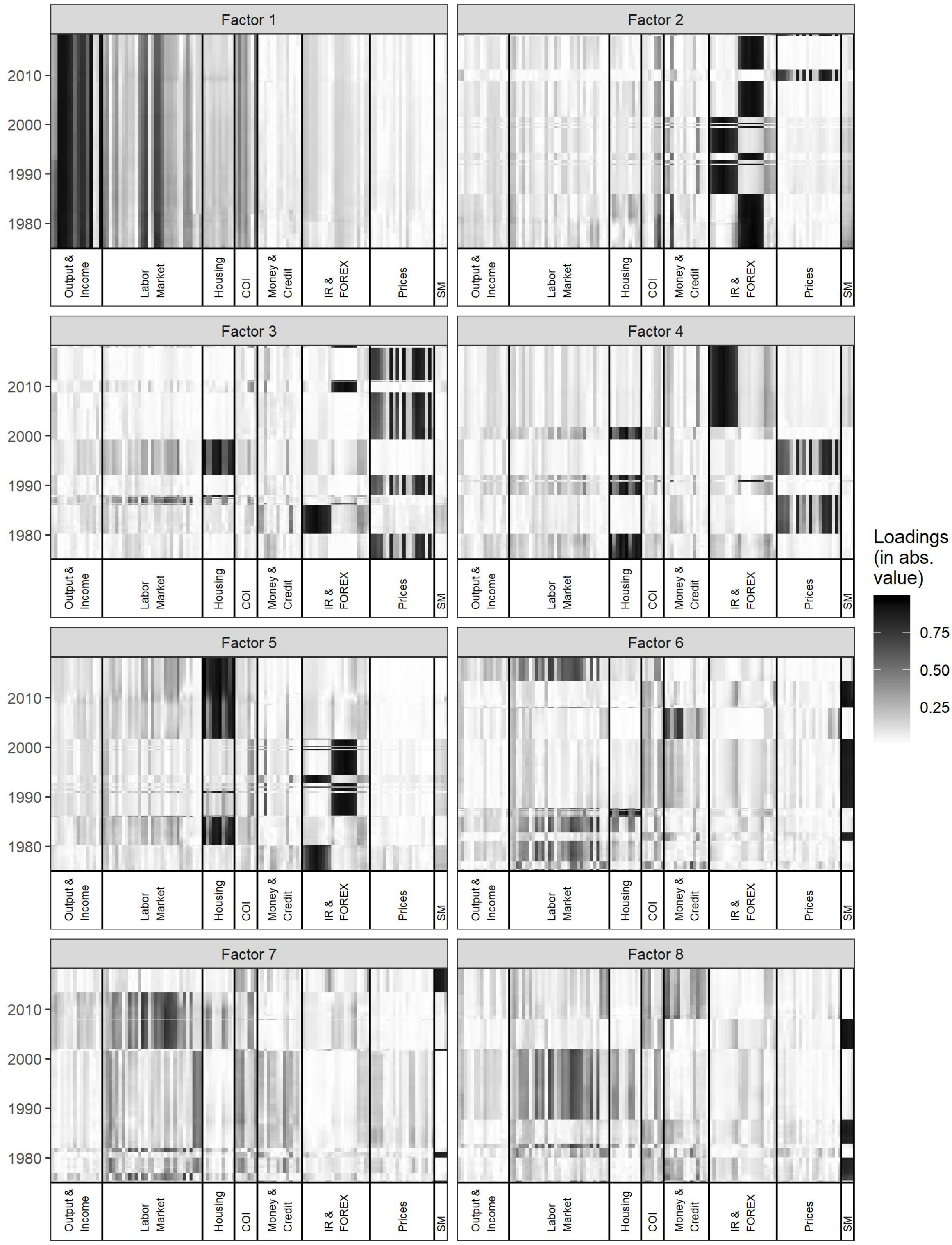


Figure 10: Time evolution of the loadings estimated with PCA varimax (in absolute value).

Notes: See Figure 7.

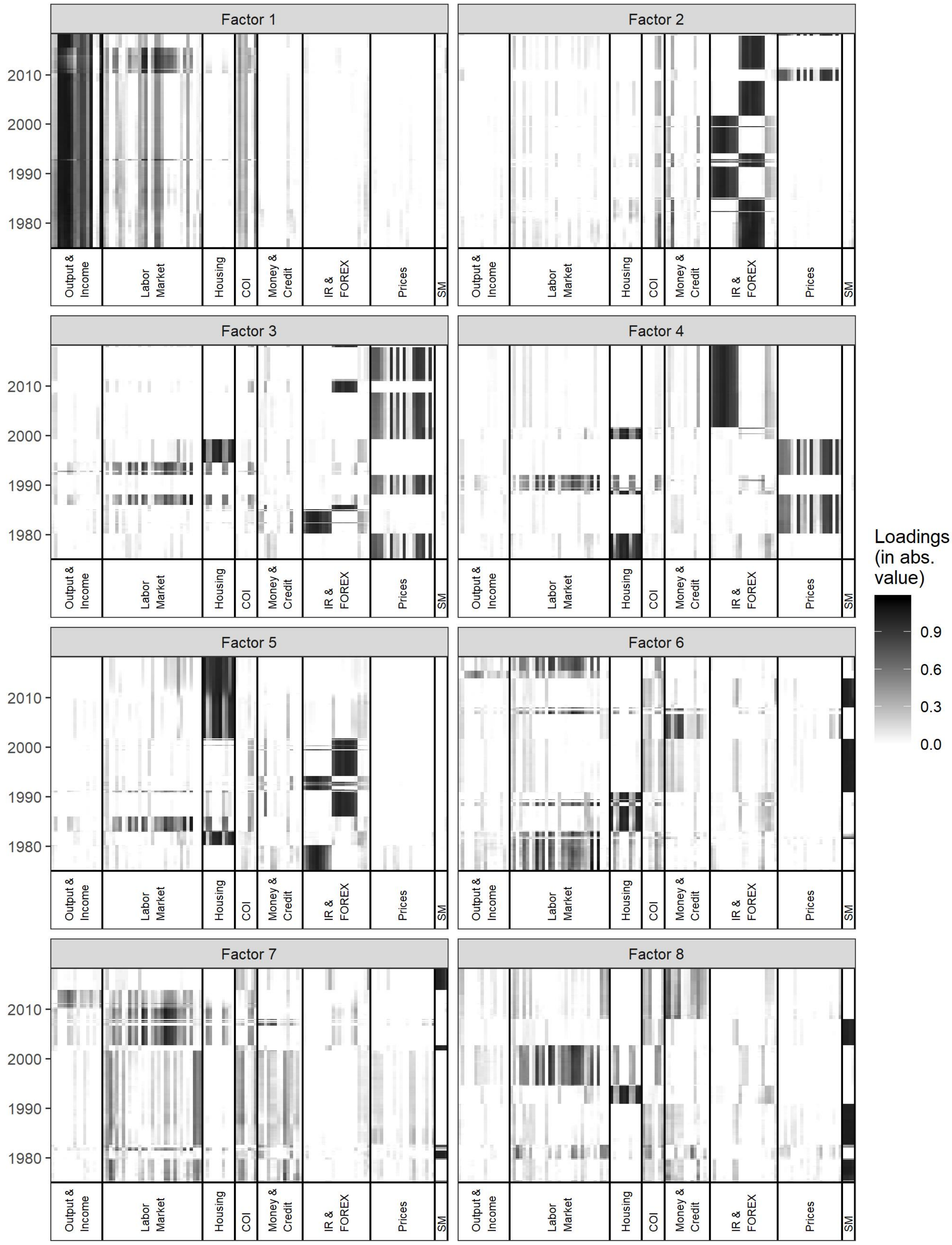


Figure 11: Time evolution of the loadings estimated with SPCA (in absolute value).

Notes: See Figure 7. The SPCA hyperparameters are set to $(\kappa_1, \kappa_2) = (0.1, 0.6)$ throughout this exercise. Indeed, they show little sensitivity to the time window. For example, the tuned SPCA hyperparameters are $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.6)$ with the full sample 1960:1-2018:4, and $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.3)$ with the subsample 1960:1-1989:12. These minor variations leave the estimated loadings almost unaffected. Based on this, we consider more appropriate to keep κ_1 and κ_2 fixed (like the number of factors r), rather than repeating the grid search for each subsample.