



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

Thomas Despois, Catherine Doz

## ► To cite this version:

Thomas Despois, Catherine Doz. Another look into the factor model black box: factors interpretation and structural (in)stability. 2019. halshs-02235543v1

**HAL Id: halshs-02235543**

**<https://shs.hal.science/halshs-02235543v1>**

Preprint submitted on 1 Aug 2019 (v1), last revised 31 Aug 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**WORKING PAPER N° 2019 – 44**

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

**Thomas Despois  
Catherine Doz**

**JEL Codes: C32, C33, C38, C55**

**Keywords : dynamic factor models, factor rotations, sparse PCA, structural break**

# ANOTHER LOOK INTO THE FACTOR MODEL BLACK BOX: FACTORS INTERPRETATION AND STRUCTURAL (IN)STABILITY\*

Thomas Despois<sup>1,2</sup> and Catherine Doz<sup>1,2</sup>

<sup>1</sup>Paris School of Economics

<sup>2</sup>Université Paris 1 Panthéon-Sorbonne

July 31, 2019

COMMENTS WELCOME.

## Abstract

Dynamic factor models (DFMs) have been successfully used in various applications, and have become increasingly popular. Nevertheless, they suffer from two weaknesses. First, the factors are very difficult to interpret with the usual estimation methods. Secondly, there is recent and mounting evidence that they can be subject to structural instability, potentially leading to an inconsistent estimation of the factors. The central point of this paper is to tackle the uninterpretability issue. We consider two families of methods to get interpretable factors: factor rotations and sparse PCA. When using a large dataset of US macroeconomic and financial variables, they recover the same factor representation, with a simple structure in the loadings offering a clear economic interpretation of the factors, which appears to be stable over time. This provides new lens for the DFM applications, and especially to study the structural instability issue. Using this factor representation, we find no evidence for the emergence of a new factor or for a widespread break in the factor loadings at a given time. The structural instability in the model seems to rather consist in the order of the factors switching over time (apparently in relation with the structural breaks in the data), and breaks in a limited number of loadings which can be localized and interpreted.

*JEL classification:* C32, C33, C38, C55.

*Keywords:* dynamic factor models, factor rotations, sparse PCA, structural break.

---

\*E-mail addresses: [thomas.despois@psemail.eu](mailto:thomas.despois@psemail.eu) and [catherine.doz@univ-paris1.fr](mailto:catherine.doz@univ-paris1.fr).

We would like to thank 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, and the 2019 Annual Conference of the International Association for Applied Econometrics for helpful comments.

# 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 accounting for the widespread comovements of the observed 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.<sup>1</sup> 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, Bernanke, Boivin and Elias, 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,<sup>2</sup> 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 to the uninterpretability issue.

So far, the factor model black box has been opened from three different angles. First, the stationary (Forni, Giannone, Lippi and Reichlin, 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 can be attempted through the  $R^2$  of the regression of the variables on the factors, or equivalently through the factor loadings if the factors were estimated with a principal component analysis (PCA). Nevertheless, with all the usual estimation methods, the factors are quite difficult to interpret, if not impossible. Most of the literature tried to solve

---

<sup>1</sup>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*.

<sup>2</sup>Via an EM algorithm (Stock and Watson, 2002b), the Kalman filter (Doz, Giannone and Reichlin, 2011), or both (Doz, Giannone and Reichlin, 2012).

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 technique 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 factors interpretation in this field too (Croux and Exterkate, 2011, Kaufmann and Schumacher, 2017, 2019, Beyeler and Kaufmann, 2018), and can also slightly improve the forecasting performance (Croux and Exterkate, 2011, Kim and Swanson, 2016, Kristensen, 2017).

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, Plagborg-Møller, Stock and Watson, 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.<sup>3</sup> When considering US macroeconomic data, these tests systematically corroborate the presence of instabilities, but yield quite largely different conclusions with respect to the number of breaks, their date, and their form (especially as to whether the breaks are in the loadings and/or in the

---

<sup>3</sup>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), etc.

number of factors). Indeed, [Breitung and Eickmeier \(2011\)](#) confirm the presence of a structural break around 1984,<sup>4</sup> 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: 1979m09 (second oil crisis), 1984m07 (beginning of the Great Moderation), 1990m03, 1995m06 (two labor productivity shocks), and 2002m01 (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.

In this paper, we first try to get interpretable factors while remaining as agnostic as possible. To achieve this, we estimate the factors and loadings by sparse PCA (coming from the machine learning literature), and by factor rotations relying on the minimization of an atheoretical simplicity criterion in the loadings (rather than a priori identifying restrictions). Using a dataset of 123 US macroeconomic and financial variables (the now standard FRED-MD), the rotated and sparse factors offer a straightforward economic interpretation, 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 loadings pattern. 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 simple structure.

This factor representation provides new lens to study the structural instability issue. We find very little evidence for the emergence of a new 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 the order of the factors switching over time (apparently in relation with the structural breaks in the data), and breaks in a limited number of loadings which can be localized and interpreted.

The rest of the paper is structured as follows. Sections 2 and 3 respectively introduces the econometric and empirical frameworks. Section 4 presents the factors interpretation. Section 5 studies the structural instability within the factor model. Section 6 concludes.

## 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  $r \ll N$ . Let  $X$  be the  $T \times N$  panel data matrix (with all the variables standardized beforehand),  $F$  the  $T \times r$  matrix

---

<sup>4</sup>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).

including the  $r$  factors,  $\Lambda$  the  $N \times r$  factor loadings matrix, and  $e$  the  $T \times N$  idiosyncratic errors matrix. Denoting  $i$  the cross-section unit ( $i = 1, \dots, N$ ),  $t$  the time unit ( $t = 1, \dots, T$ ), and  $x_{it}$  the observation of  $x_i$  at time  $t$ ,  $F_t = (F_{1t}, \dots, F_{rt})'$ ,  $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir})'$ , the model is written as:

$$x_{it} = \lambda_i' F_t + e_{it}, \quad \text{for } i = 1, \dots, N, \quad t = 1, \dots, T \quad (1)$$

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

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

$$\text{or } X = F \Lambda' + e \quad (3)$$

Note that 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  $r$  largest eigenvalues of  $\Lambda' \Lambda$  diverge at rate  $N$  whereas the covariance matrix of the idiosyncratic vector  $e_t$  stays bounded when  $N$  goes to infinity (see e.g. [Stock and Watson \(2002a\)](#), [Bai and Ng \(2002\)](#) for the most often used sets of assumptions).

**Notation** For a vector  $v \in \mathbb{R}^n$ , we denote its  $\ell_1$  norm as  $\|v\|_1 = \sum_{i=1}^n |v_i|$ , and its  $\ell_2$  norm as  $\|v\|_2 = \sqrt{\sum_{i=1}^n v_i^2}$ . The  $\ell_0$  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  $\|A\|_F = \sqrt{\text{Tr}(A' A)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$ .

## 2.1 Principal component analysis

[Stock and Watson \(2002a\)](#) proved that the factors in model (1) are consistently estimated by principal components under the general set of assumptions associated with approximate DFMs, and 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:

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

which can be rewritten as

$$\min_{F, \Lambda} \|X - F\Lambda'\|_F^2. \quad (5)$$

It is common knowledge that  $F$  and  $\Lambda$  are not separately identifiable. For any  $r \times r$  non-singular matrix  $R$ ,  $F\Lambda'$  is observationally equivalent to  $\tilde{F}\tilde{\Lambda}'$ , where  $\tilde{\Lambda} = \Lambda R$  and  $\tilde{F} = F(R')^{-1}$ . In order to uniquely identify  $F$  and  $\Lambda$  (still up to a column sign change), we need to normalize them with a set of  $r^2$  restrictions. In approximate factor models, where the eigenvalues of  $\Lambda'\Lambda$  are supposed to diverge at rate  $N$ , the two usual normalizations are  $\Lambda'\Lambda/N = I_r$  with  $F'F$  diagonal, or  $F'F/T = I_r$  with  $\Lambda'\Lambda$  diagonal.

Denoting  $\hat{P}$  the  $N \times r$  matrix containing the eigenvectors corresponding to the  $r$  largest eigenvalues of  $X'X$ , the estimation of the loadings and factors under the normalization  $\Lambda'\Lambda/N = I_r$  and  $F'F$  diagonal, and can be obtained with

$$\hat{\Lambda} = \sqrt{N}\hat{P} \quad (6)$$

$$\hat{F} = X\hat{\Lambda}/N \quad (7)$$

For the following, it should be kept in mind that, by construction, PCA yields orthogonal estimated loadings (i.e. the columns of  $\hat{\Lambda}$  are orthogonal), and uncorrelated estimated factors ranked by decreasing proportion of explained variance.

## 2.2 Factor rotations

The historical method to get interpretable factors is the use of factor rotations.<sup>5</sup> The idea is to take advantage of the rotational indeterminacy of the estimated loadings and factors. The previously estimated matrix of loadings  $\hat{\Lambda}$ , leading to uninterpretable factors, is multiplied by an  $r \times r$  rotation matrix  $R$  such that the rotated matrix of loadings optimizes a certain simple structure criterion. The estimated factors are then rotated accordingly, and should now be interpretable.

$$\hat{\Lambda}_{rotated} = \hat{\Lambda}R \quad (8)$$

$$\hat{F}_{rotated} = \hat{F}(R')^{-1} \quad (9)$$

Some points should be made clear concerning 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

---

<sup>5</sup>Factor rotations were originally designed to be applied on the results of an explanatory factor analysis (EFA), i.e. the exact static factor model. EFA is closely related to PCA, and when  $N$  is large, leads to almost identical results under the assumptions of approximate factor models (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). Nonetheless, following a rotation, the PCA estimated factors do not necessarily explain successively less variance anymore.



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,<sup>6</sup> 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  $\hat{y}$ . 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 uncorrelated, whereas oblique rotations do not.<sup>7</sup> 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 seems quite unrealistic in most fields. Moreover, oblique rotations do not force the factors correlation, but rather simply allow for it. If the transformed factors are found to be only slightly correlated, we can pass to an orthogonal rotation. 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.<sup>8</sup>

$$\hat{R}_{varimax} = \arg \max_R \left\{ \frac{1}{N} \sum_{k=1}^r \sum_{i=1}^N (\hat{\Lambda} R)_{ik}^4 - \sum_{k=1}^r \left( \frac{1}{N} \sum_{i=1}^N (\hat{\Lambda} R)_{ik}^2 \right)^2 \right\} \quad (10)$$

Concerning oblique rotations, the direct oblimin family (Jennrich and Sampson, 1966) is the most popular due to its great flexibility. Within this family, the direct quartimin rotation 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<sup>9</sup> between factors, and the solution is thus achieved through the minimization of the following

---

<sup>6</sup>The communality of a variable refers to the part of its variance explained by the factors, i.e. the  $R^2$  of the regression of the variable on the factors.

<sup>7</sup>The term "oblique rotation" is technically a misnomer. A rotation is an isometry, so it should preserve angles (by property). Since the factors are orthogonal prior to any rotation, they should remain so afterwards. "Oblique rotations" are invertible transformations which allow the angles between the rotated factors to be lower than 90 degrees, hence their name.

<sup>8</sup>Hence the name: VARIance MAXimization of the squared loadings within each factor.

<sup>9</sup>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,  $C$  is obviously diagonal (and if the factors normalization is such that  $F'F/T = I_r$ , then  $C = I_r$  and  $S = P$ ). The simplicity criterion of an oblique rotation is optimized while using the pattern loadings matrix. In the following, the obliquely rotated factors are interpreted with the pattern loadings, as it is usually done.

simplicity criterion:<sup>10</sup>

$$\hat{R}_{quartimin} = \arg \min_R \left\{ \sum_{j \neq k} \sum_{i=1}^N (\hat{\Lambda}R)_{ij}^2 (\hat{\Lambda}R)_{ik}^2 \right\} \quad (11)$$

For the sake of brevity, we only consider the standard varimax and quartimin rotations in the following. Nevertheless, we have tried several other rotations, and they all yield extremely similar results.

### 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  $\ell_0$  or  $\ell_1$  so as to induce sparsity in the loadings.<sup>11</sup> 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 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),<sup>12</sup> 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

<sup>10</sup>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.

<sup>11</sup>Sparse factor models can also be implemented with a Bayesian approach, using a mixture prior inducing sparsity in the loadings (e.g. [West, 2003](#)).

<sup>12</sup>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).

adding an elastic net penalty (Zou and Hastie, 2005), which is a convex combination of the  $\ell_1$  (lasso) and the  $\ell_2$  (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 (5) of the PCA problem amounts to solve

$$\min_P \|X - XPP'\|_F^2 \quad \text{s.t.} \quad P'P = I_r, \quad (12)$$

with  $P$  an  $N \times r$  matrix, and  $\hat{P}$  corresponding to the eigenvectors associated to the  $r$  biggest eigenvalues of  $X'X$ . They then add an elastic net penalty to this regression-type problem, which gives:

$$\min_{P,U} \|X - XPU'\|_F^2 + \sum_{k=1}^r \kappa_{1k} \|P_k\|_1 + \kappa_2 \sum_{k=1}^r \|P_k\|_2^2 \quad \text{s.t.} \quad U'U = I_r \quad (13)$$

where  $P_k$  is the  $k$ th modified principal component (the  $k$ th column of  $P$ ),  $\kappa_{1k}$  the  $\ell_1$  penalty hyperparameter controlling the level of sparsity in the  $k$ th modified principal component, and  $\kappa_2$  the  $\ell_2$  penalty hyperparameter. If desired, the  $\kappa_{1k}$  hyperparameters can be chosen so as to impose a substantially different level of sparsity in each modified principal component.

Knowing the  $r+1$  hyperparameters  $\kappa$  (whose tuning is explained below), and after initializing  $U$  with the first  $r$  ordinary principal components, the optimization is performed via a block descent: (13) is alternatively minimized for  $P$  given  $U$  (with an elastic net regression), and for  $U$  given  $P$  (with a reduced rank Procrustes rotation), until convergence.<sup>13</sup> As with ordinary PCA, the estimated loadings and factors are obtained with  $\hat{\Lambda} = \sqrt{N}\hat{P}$  and  $\hat{F} = X\hat{\Lambda}/N$ , but  $\hat{P}$  and  $\hat{\Lambda}$  are now sparse.

---

<sup>13</sup>See Zou, Hastie and Tibshirani (2006) for more details.

### 3 Empirical framework

#### 3.1 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<sup>14</sup> 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.<sup>15</sup> 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<sup>16</sup> because this may remove some structural breaks (Breitung and Eickmeier, 2011).

#### 3.2 Selecting the number of factors

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, even in case of heteroscedasticity (along time or across variables), autocorrelation, or correlation across the idiosyncratic terms.

For the sake of clarity, the number of factors  $r$  is kept fixed in the next section. The selection of  $r$  is thus performed by applying the six main criteria of Bai and Ng (2002) over the full

---

<sup>14</sup><https://files.stlouisfed.org/files/htdocs/fred-databases/fredmdchanges.pdf>

<sup>15</sup><https://research.stlouisfed.org/econ/mccracken/fred-databases/>

<sup>16</sup>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.

balanced sample, with  $r_{max} = 10$ . We choose to select  $r = 8$ , as [McCracken and Ng \(2016\)](#)<sup>17</sup> and [Kristensen \(2017\)](#) did when considering their full sample.

Table 1: Selected number of factors (full balanced panel).

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

### 3.3 Tuning the SPCA hyperparameters

Before starting any tuning, we consider that there is no theoretical reason to force some factors to be far more sparse than others in a macroeconomic application.<sup>18</sup> Thus, we fix all of the eight  $\ell_1$  penalization parameters equal:  $\kappa_{1k} = \kappa_1 \ \forall k \in \{1, \dots, r\}$ . This leaves us with the tuning of only two hyperparameters ( $\kappa_1$  and  $\kappa_2$ ) instead of nine, which drastically reduces the computational cost of the tuning, and 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.<sup>19</sup> 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 penalized regression methods (ridge regression, lasso, adaptative lasso, elastic net, and adaptative 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](#)):

$$(\widehat{\kappa}_1, \widehat{\kappa}_2) = \underset{\kappa_1, \kappa_2}{\operatorname{argmin}} \log \left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [x_{it} - \widehat{\lambda}_i(\kappa_1, \kappa_2)' \widehat{F}_t(\kappa_1, \kappa_2)]^2 \right) + m(\kappa_1, \kappa_2) \frac{\log(NT)}{NT} \quad (14)$$

where  $m$  is the number of non-null loadings in  $\widehat{\Lambda}$ . They obviously both depend on  $(\kappa_1, \kappa_2)$ . This

<sup>17</sup>[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 empirical analysis, so they select 8 factors.

<sup>18</sup>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 correlated with far less variables than the prices factor or the output factor for example.

<sup>19</sup>Except for the very specific case of purely autoregressive models with uncorrelated errors ([Bergmeir, Hyndman and Koo, 2018](#)).

BIC-type criterion is minimized with a grid search over the following grid:

$$\kappa_1 \in \{0, 0.1, \dots, 0.5\}$$

$$\kappa_2 \in \{0, 0.1, \dots, 1\}$$

Considering the full balanced sample 1960:1-2018:4, we get  $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.6)$ .

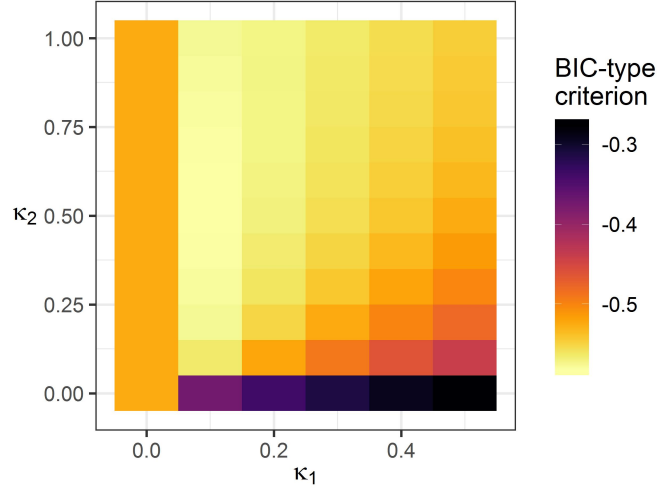


Figure 1: Grid search of the SPCA tuning parameters (full balanced panel).

*Notes:* The grid search takes 4 minutes on R 3.5.1 with an Intel Core i5 3.3 GHz processor and 8 Go RAM.

The tuning of the hyperparameters exhibits little sensitivity to the time window. For example, considering the sample 1960:1-1989:12, the tuned hyperparameters become  $(\widehat{\kappa}_1, \widehat{\kappa}_2) = (0.1, 0.3)$ , letting the estimated loadings almost unchanged. Based on this, we consider that it is acceptable and reasonable to keep  $(\kappa_1, \kappa_2)$  fixed in the following (like the number of factors  $r$ ), rather than repeating the grid search exercise for each time period.

## 4 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 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 strong loadings in the labor market group), real personal consumption expenditures, retail and food services sales (the two strong loadings in the consumption, orders and inventories group). All the other estimated factors are impossible to interpret.

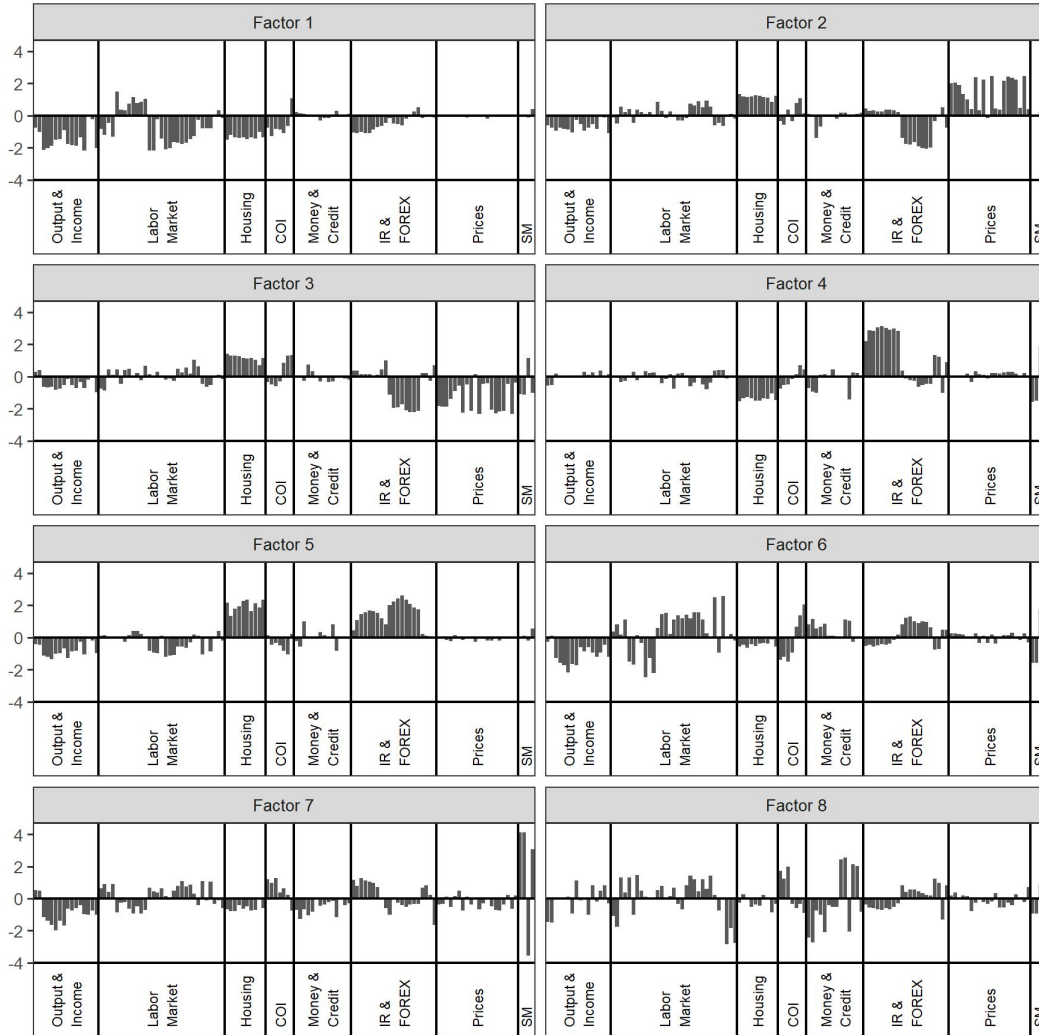


Figure 2: PCA factor loadings (full balanced panel).

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 exactly the same loadings pattern (Figure 3). Interestingly, we can notice that the loadings of the three interpretable (unrotated) PCA estimated factors (interest rates, stock market, money and credit) are very similar to the loadings of their rotated and sparse factors counterparts (especially for the last two).

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 factors involve several variables outside the money and credit group.

Since quartimin and varimax lead to exactly the same loadings pattern as SPCA, they of course all lead to almost perfectly identical factors (Figure 4). 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 rotated and sparse factors also display an extremely similar explanatory power (Table 2). 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.



We consider the quartimin rotation as the preferable approach to recover this factor representation (at least among the methods considered here). Factor rotations and sparse PCA lead to the same results, but with factor rotations, the implementation is easier and the estimation time far smaller, especially because they do not require to tune any hyperparameter.<sup>20</sup> Besides, the factor rotations can be applied on factors estimated with other methods than PCA, like the two-step estimator (Doz, Giannone and Reichlin, 2011) or the quasi-maximum likelihood estimator (Doz, Giannone and Reichlin, 2012). Moreover, there is usually no reason to impose the factors to be uncorrelated as it is done with orthogonal rotations. For example, in our case, we should expect the output factor and the labor factor to be highly correlated (as it is recovered by PCA with quartimin rotation and SPCA).

Table 2: Percentage of the dataset total variance explained by the factors, depending on the estimation method (full balanced panel).

Factor	PCA	Factor	PCA quartimin	PCA varimax	SPCA
1 (?)	14.77	1 (Output)	8.39	8.45	9.14
2 (?)	7.27	2 (Prices)	7.07	7.06	6.88
3 (?)	7.05	3 (Spreads)	5.79	5.78	5.51
4 (Interest rates)	5.67	4 (Interest rates)	5.86	5.86	5.30
5 (?)	4.31	5 (Housing)	7.06	7.03	6.18
6 (?)	3.44	6 (Labor)	7.57	7.58	4.06
7 (Stock market)	2.96	7 (Stock market)	3.64	3.64	3.15
8 (Money & credit)	2.47	8 (Money & credit)	2.57	2.55	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.

<sup>20</sup>Some factor rotations involve an hyperparameter, but that is not the case for the most standard ones like quartimin and varimax.

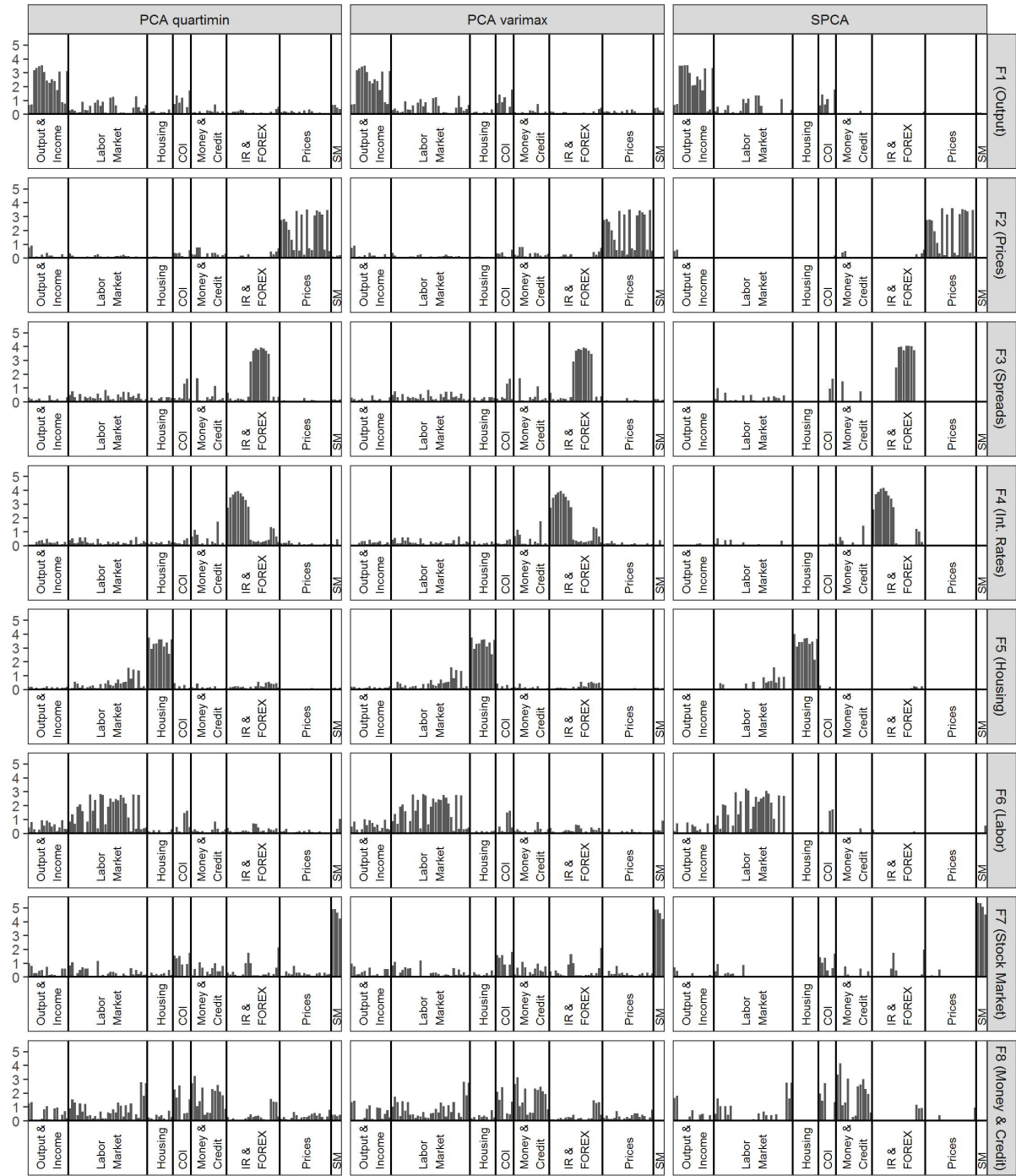


Figure 3: Comparison of the factor loadings (in absolute value) of the rotated and sparse factors (full balanced panel).

*Notes:* The proportion of non-null 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.

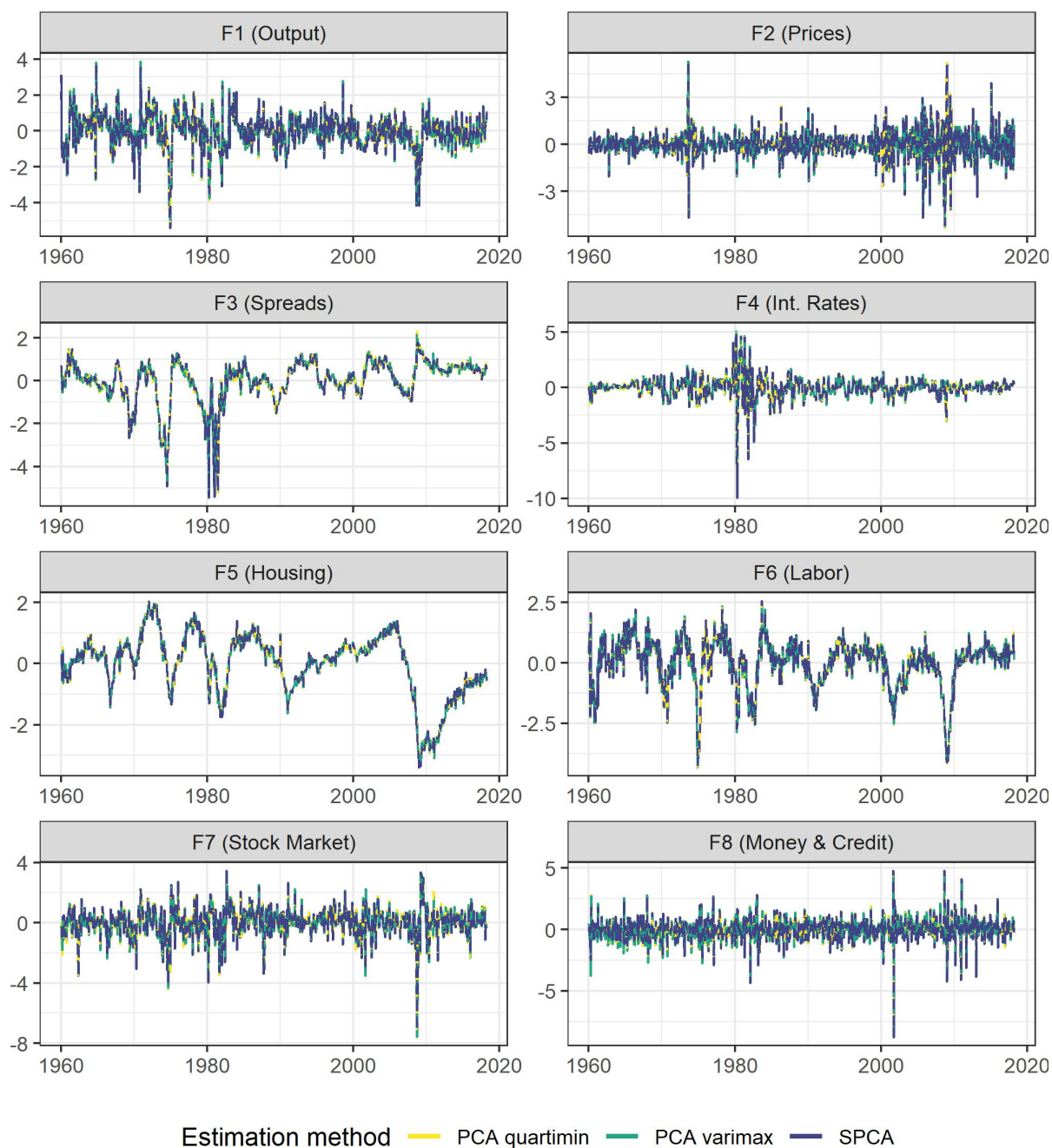


Figure 4: Comparison of the rotated and sparse factors (full balanced panel).

*Notes:* Remember that the factors are identified up to a sign change. On this figure, 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 factors is reversed: F1 (Output), F3 (Spreads) and F8 (Money & Credit). For the sparse factors: F1 (Output), F2 (Prices), F4 (Interest Rates), F6 (Labor), and F7 (Stock Market). In order to facilitate the comparisons, each factor is also standardized.

## 5 Structural instability in the DFM

### 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 brevity, we only present the results obtained by PCA with quartimin rotation. The results with varimax rotation and SPCA are still extremely similar, and we consider PCA with quartimin rotation as the preferable approach to recover this factor representation.

It appears that the recursively estimated loadings recover the previous results concerning the factors interpretability (Figure 5). For every period, all the factors have a clear economic interpretation, stable over time, and which is the same as before when considering the full sample. What is unstable is rather the order of the columns of  $\Lambda$  (and thus also the ordering of the factors). This is especially visible with the prices factor, switching between columns 2, 3 and 4 over time. These switches seem to be in relation with the structural breaks in the data, changing the order of the factors through time. Indeed, the main visible switch dates are consistent with the structural breaks previously found in the literature: second oil crisis, beginning of the Great Moderation, early 2000s recession, beginning of the Great Recession.

Working with this factor representation sheds a whole new light on the structural instability issue. Indeed, the loadings surprisingly exhibit a very high temporal stability. We cannot distinguish any widespread break in the loadings at any date. There is only one very noticeable break, affecting some loadings of the housing factor in the wake of the subprime crisis. Moreover, there is no evidence for the emergence of an additional factor. All of the eight factors are indeed associated with non-zero loadings for all the sub-samples. Moreover, the part of the dataset total variance explained by the factors remains remarkably stable over time (Figure 6). 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 the columns of  $\Lambda$  (and thus also the ordering of the factors) switching over time, and breaks in a limited number of loadings. 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 to build the factors summarizing them stays very stable.

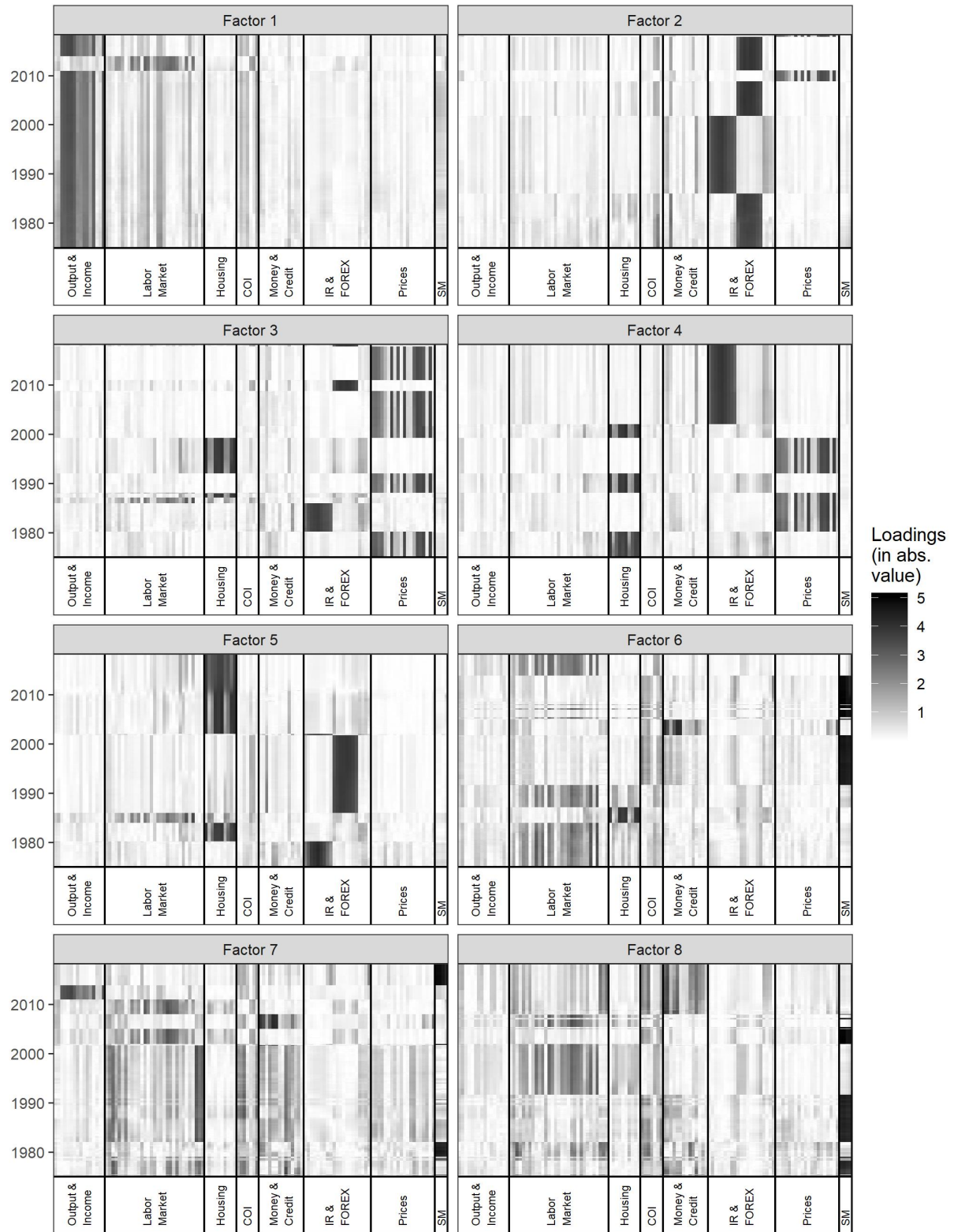


Figure 5: Time evolution of the loadings estimated with PCA quartimin (in absolute value).

*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.

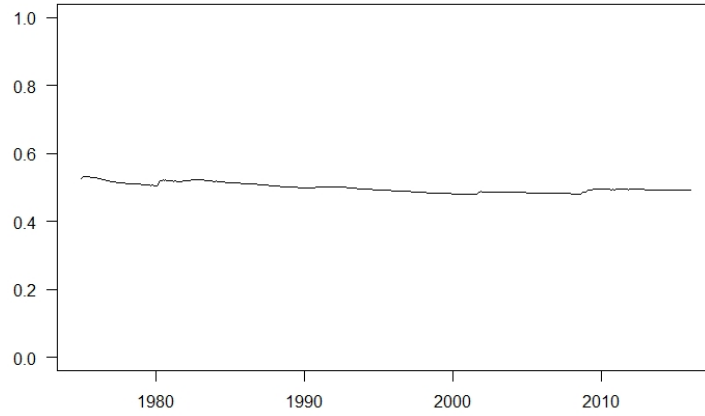


Figure 6: Time evolution of the part of the dataset total variance explained by the PCA estimated factors (rotated or not).

*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 2. Remember that, by construction, the part of explained variance is identical when considering the factors estimated by rotated or unrotated PCA.

## 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).

Concerning the 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 3. 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 1971m03, 1979m08, and 2008m03). 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 3: Results of the Chen-Dolado-Gonzalo test.

Breaks	PCA	PCA quartimin	PCA varimax	SPCA
0	699.21	<b>9.061</b>	<b>7.852</b>	<b>-38.83</b>
1	596.46	13.937	13.023	-28.71
2	564.90	31.829	31.712	-14.07
3	<b>564.40</b>	51.943	53.876	16.69
4	580.73	82.334	83.818	40.42
5	613.61	118.894	120.195	77.98

Concerning the individual structural instability, we consider the equation

$$x_{it} = \lambda_i' \hat{f}_t + e_{it} \quad (15)$$

with  $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir})'$  and  $\hat{f}_t = (\hat{f}_{1t}, \dots, \hat{f}_{rt})'$  the factors estimated with PCA and quartimin on the full balanced sample. When testing the stability of the loadings jointly with a Bai-Perron test (Bai and Perron, 1998, 2003), it results that 42.7% of the 123 variables experience at least one break in their loadings. When testing the stability of the loadings individually with a Sup-Wald test (Andrews, 1993), it results that between 17% and 36% of the loadings experience a break (depending on the specification, 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 7). We can notice two main features. First, the variables concerning the Northeast and the Midwest are now 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 market. This is of course very consistent with the narratives of the 2007–9 recession.

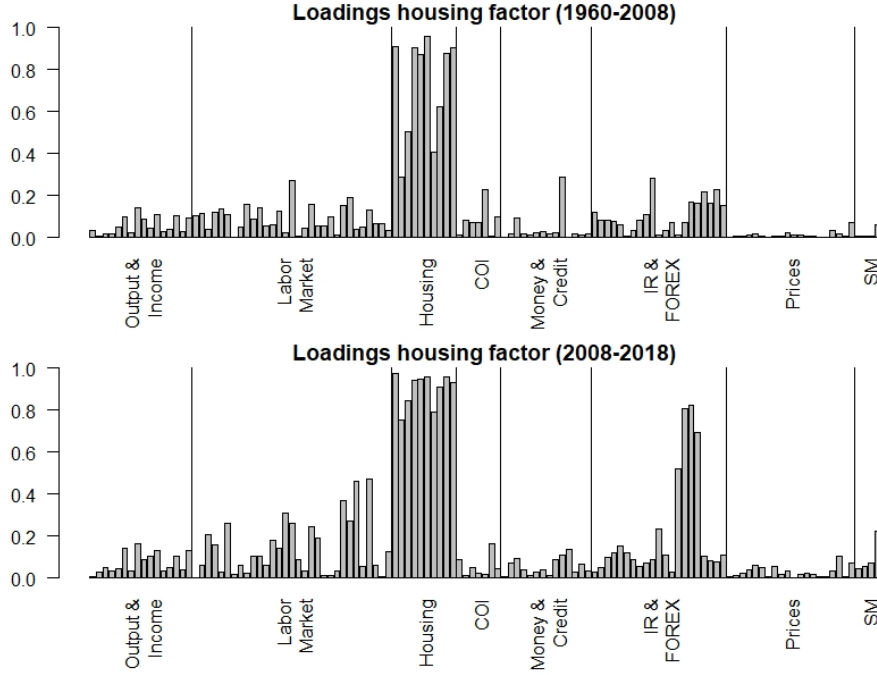


Figure 7: Loadings of the housing factor before and after 2008, estimated by PCA quartimin.

## 6 Conclusion

In this paper, we propose several simple and agnostic approaches to get an economic interpretation of the factors, leading to almost identical results. Our results suggest that when using the principal components estimator of the factors and loadings, the established unrotated solution, while computationally convenient, recovers an uninterpretable factor representation. As soon as a method enforcing a simple structure in the loadings is used (no matter the method), it yields exactly the same simple structure. This factor representation seems more natural and informative than the unrotated solution, and should probably be preferred as reference in other DFM applications, like diffusion index forecasting or FAVAR modelling. Moreover, in our view, the quartimin rotation is the preferable method to recover this factor representation.

When working with the interpretable factors, we find no evidence for the emergence of a new 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 the order of the factors switching over time (apparently in relation with the structural breaks in the data), and breaks in a limited number of loadings which can be localized and interpreted.

We are considering several extensions to this paper. A first development, that we have already started, consists in rotating factors estimated using methods which take the dynamics into account, like in [Doz, Giannone and Reichlin \(2011, 2012\)](#). The results seem to be very consistent. A second development, also initiated, investigates the temporal instability in the diffusion index forecasting equation when using the rotated and sparse factors. Finally, we can



wonder to what extent our results concerning the structural instability are recovered when using other data. For example, modeling international cycles with DFMs yield time-varying factor loadings because the relations between countries largely evolve over time (Del Negro and Otrok, 2008). It would be interesting to apply our approach with the euro area, an emerging country, or financial data.

## References

- Andrews, D. W. (1993), ‘Tests for parameter instability and structural change with unknown change point’, *Econometrica* **61**(4), 821–856.
- 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.
- Bañbura, 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* .
- 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 Elias, 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. (2018), Factor augmented var revisited-a sparse dynamic factor model approach, Technical report, Kiel, Hamburg: ZBW-Leibniz-Informationszentrum Wirtschaft.
- Breitung, J. and Eickmeier, S. (2011), ‘Testing for structural breaks in dynamic factor models’, *Journal of Econometrics* **163**(1), 71–84.
- 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, Technical report, Tinbergen Institute Discussion Paper.
- d’Aspremont, A., Bach, F. and El Ghaoui, L. (2008), ‘Optimal solutions for sparse principal component analysis’, *Journal of Machine Learning Research* **9**(Jul), 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. 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. (2016), ‘Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods’, *International Journal of Forecasting*.
- Kose, M. A., Otrok, C. and Whiteman, C. H. (2003), ‘International business cycles: World, region, and country-specific factors’, *The 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.
- Smeeke, 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.

# Appendix

## A Data

The composition of FRED-MD has slightly changer 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. The transformation code (TC) indicates how the series was transformed to ensure stationarity: (1) no transformation, (2)  $\Delta x_t$ , (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 the five variables listed in Table A9.

Table A1: Output and income series.

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

Table A2: Labor market series.

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

Table A3: Housing series.

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

Table A4: Consumption, orders, and inventories series.

Mnemonic	Description	TC
DPCERA3M086SBEA	Real personal consumption expenditures	5
CMRMTSPLx	Real Manufacturing and Trade Industries Sales	5
RETAILx	Retail and Food Services Sales	5
NAPM	ISM: PMI Composite Index	1
NAPMNOI	ISM: New Orders Index	1
NAPMSDI	ISM: Supplier Deliveries Index	1
NAPMII	ISM: Inventories Index	1
ACOGNO	New Orders for Consumer Goods	5
AMDMNOx	New Orders for Durable Goods	5
ANDENOx	New Orders for Nondefense Capital Good	5
AMDMUOx	Unfilled Orders for Durable Goods	5
BUSINVx	Total Business Inventories	5
ISRATIOx	Total Business: Inventories to Sales Ratio	2
UMCSENTx	Consumer Sentiment Index	2

Table A5: Money and credit series.

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

Table A6: Interest and exchange rates.

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

Table A7: Prices series.

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

Table A8: Stock market series.

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

Table A9: Dropped variables.

Mnemonic	Description	Group
ACOGNO	New Orders for Consumer Goods	Consumption, Orders and Inventories
ANDENOX	New Orders for Nondefense Capital Goods	Consumption, Orders and Inventories
UMCSENTx	Consumer Sentiment Index	Consumption, Orders and Inventories
TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies	Interest Rate and Exchange Rates
VXOCLSx	VXO	Stock Market