Diffusion Indexes with Sparse Loadings

Johannes Tang Kristensen[a,b,*]

[a] *Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*

[b] *Center for Research in Econometric Analysis of Time Series (CREATES), Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark*

This version: March 11, 2014.

Abstract

The use of large-dimensional factor models in forecasting has received much attention in the literature with the consensus being that improvements on forecasts can be achieved when comparing with standard models. However, recent contributions in the literature have demonstrated that care needs to be taken when choosing which variables to include in the model. A number of different approaches to determining these variables have been put forward. These are, however, often based on ad-hoc procedures or abandon the underlying theoretical factor model.

In this paper we will take a different approach to the problem by using the LASSO as a variable selection method to choose between the possible variables and thus obtain sparse loadings from which factors or diffusion indexes can be formed. This allows us to build a more parsimonious factor model which is better suited for forecasting compared to the traditional principal components (PC) approach. We provide an asymptotic analysis of the estimator and illustrate its merits empirically in a forecasting experiment based on US macroeconomic data. Overall we find that compared to PC we obtain improvements in forecasting accuracy and thus find it to be an important alternative to PC.

Keywords: Forecasting, Factors Models, Principal Components Analysis, LASSO

JEL classifications: C38, C53, E27, E37.

*I would like to thank Stefan Holst Milton Bache, Anders Bredahl Kock, Marcelo C. Medeiros, and seminar participants at the Department of Business and Economics, University of Southern Denmark, for their helpful comments and suggestions. Financial support from the The Danish Council for Independent Research, Social Sciences (11-116607/FSE) and support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation is gratefully acknowledged.

Email address: johannes@sam.sdu.dk (Johannes Tang Kristensen)

1. Introduction

Many of the initial attempts at estimating factor models proposed in the literature were quite seriously limited in the amount of data they could handle. Because of this the prevailing methodology used for a number of years now is that of asymptotic principal components (PC). This non-parametric method is capable of handling enormous datasets at almost no computational cost. A recent survey by Stock and Watson (2011) provides a thorough overview of the state of the literature, and is an important addition to previous surveys (e.g. Bai and Ng, 2008b; Stock and Watson, 2006).

One area where the use of factor models has become particularly popular is in macroeconomic forecasting. The literature on using factors estimated from large datasets using PC was initiated by Stock and Watson (2002a,b). Although the estimated factors are often difficult or even impossible to give any economic interpretation, they argued that in the context of macroeconomic forecasting one possible interpretation of the estimated factors is in terms of the diffusion indexes developed by NBER business cycle analysts to measure common movement in macroeconomic variables. Due to this they referred to the estimated factors as diffusion indexes and this has now become the standard terminology when concerned with macroeconomic forecasting using factors estimated by PC (or similar methods).

Although diffusion indexes are conceptually very appealing in their ability to allow the use of very large datasets in a parsimonious manner, they do not necessarily give forecasting performance gains when the number of included variables is increased. Boivin and Ng (2006) investigate the problem. They show using both simulations and real data that forecasting performance is not always improved by including more variables and in fact that in some cases using a smaller dataset of pre-screened variables better forecasting results can be obtained.

Screening of the data is also the topic of Bai and Ng (2008a) where the screening is based on the variable we wish to forecast. The idea is to use various methods to try to determine a subset of the data best suited to forecast the variable thereby obtaining a set of targeted predictors. These are then used to estimated the diffusion indexes. A similar idea is entertained in Dias, Pinheiro, and Rua (2010) where the determination of the targeted predictors is incorporated in the estimation of the diffusion indexes hence the name targeted diffusion indexes. In both cases the authors find improvements in the forecasting performance.

In this paper we will investigate a different way of solving this problem of which variables to include. We will use the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996) to estimate diffusion indexes with sparse loadings, i.e. loadings where only some of the entries of the vectors differ from zero. Like Boivin and Ng (2006) this will produce diffusion indexes where the individual indexes or factors are linear combinations of only a subset of the variables and not all the variables like in the classical PC approach. However, unlike Bai and Ng (2008a) and Dias et al. (2010) our estimated diffusion indexes will not be "targeted" which is in line with the original model in Stock and Watson (2002a) where the factors or diffusion indexes are common to all variables we wish to forecast. However, the sparseness introduced by the LASSO does allow for cases where only some of these common factors are relevant to a variable.

The combination of the LASSO and PC has been explored extensively in the statistics literature and is often referred to as sparse principal components (SPC). Examples of papers considering estimation of SPC include Jolliffe, Trendafilov, and Uddin (2003); Zou, Hastie, and Tibshirani (2006); Shen and Huang (2008); Witten, Tibshirani, and Hastie (2009). However, in macroeconomic forecasting the use of SPC has received very little attention. Croux and Exterkate (2011) is one exception, here the authors consider a robustified version of SPC in a typical macroeconomic forecasting setting. They argue that sparsity in the loadings could help in making the factors more easily interpretable. However, they do not provide any asymptotic justification for their approach. This paper provides both a theoretical and an empirical contribution; first we show that the SPC factor estimator is consistent under assumptions common to the macroeconomic forecasting literature, and that this estimator can be easily computed using the method of Shen and Huang (2008). In addition to this we give a simple method for determining the number of factors using ridge regression. Second, we apply the SPC factor estimator to a typical macroeconomic dataset and show that improvements in forecasting accuracy can be achieved.

The paper is organized as follows. In Section 2 we start by briefly going over the traditional PC approach to estimating factor models. Against this backdrop we then detail how SPC can be used to estimate factor models with sparse loadings. We give an asymptotic analysis of the SPC estimator where we show that the estimated factors will be consistent and give a simple alternative to existing methods for determining the number of factors based on ridge regression. Section 4 provides a number of Monte Carlo simulations that highlight the main differences between the PC and SPC estimators. In Section 5 a pseudo real-time forecasting experiment is conducted in order to judge the forecasting performance of the SPC estimator. Finally, Section 6 concludes.

2. Diffusion index forecasting

The model of Stock and Watson (2002a,b) is based on the idea that we observe a large number of macroeconomic variables, possibly many more than then number of temporal observations. These variables contain information we want to express concisely in a much lower dimension in order to forecast key variables, i.e. we want to extract factors or diffusion indexes from the dataset. More specifically we observe n variables X_t over T periods, and we assume that these variables can be modelled using a factor model. In addition to this we have a scalar time series y_t which is related to the factors and possible other exogenous variables, w_t, and that we wish to forecast. Hence our basic model is:

$$X_t = \Lambda F_t + e_t \tag{1}$$

$$y_{t+h} = \beta_F' F_t + \beta_w' w_t + \varepsilon_{t+h} \tag{2}$$

where Λ is a matrix of loadings associated with the factors and h is the forecast horizon.

A small note on dimensions and notation: X_t is $n \times 1$ with elements x_{it}, when convenient we will used matrix notation and collect these in $X = (X_1, \ldots, X_T)'$. The model has r factors and hence the loadings matrix Λ is $n \times r$, rows of this matrix will be denoted λ_i $(1 \times r)$, and columns $\underline{\lambda}_j$ $(n \times 1)$. F_t is $r \times 1$ and when used in matrix notation will be collected in $F = (F_1, \ldots, F_T)'$ $(T \times r)$. When referring to columns of F these will be denoted $\underline{F}_i$ $(T \times 1)$. w_t is $q \times 1$ and when used in matrix notation will be collected in $W = (w_1, \ldots, w_T)'$ $(T \times q)$. Finally, $\|\cdot\|_2$ is the Euclidean norm, i.e. for $z \in \mathbb{R}^n$ we have $\|z\|_2 = \sqrt{\sum_{i=1}^n z_i^2}$.

2.1. The classical approach

Estimation of the model in (1)–(2) is done using a two-step approach. In the first step the factors (or diffusion indexes) and associated loadings in (1) are estimated by means of principal components (PC). These estimates are then used in the second step estimation of (2) by OLS. The fact that PC can be used to estimate (1) may not be obvious. However, it can easily be motivated by realizing that PC is basically a least squares estimator. Consider the following non-linear least squares objective function

$$V^{\mathrm{LS}}(F, \Lambda; X) = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \lambda_i F_t)^2 \tag{3}$$

which is to be minimized over both F and Λ. Since the parameters are unidentified we need to impose restrictions on the problem, this is done by requiring that the loadings are orthogonal and have a fixed length, e.g. $\Lambda'\Lambda/n = I_r$. A solution to this problem is easily found by concentrating out F and imposing the identifying restrictions to get an equivalent maximization problem $\mathrm{tr}\left[\Lambda' X' X \Lambda\right]$ where $\mathrm{tr}[\cdot]$ denotes the matrix trace. Hence the loadings are estimated as the eigenvectors of $X'X$ corresponding to its r largest eigenvalues, and the factors are given as:

$$\hat{F}^{\mathrm{LS}} = X \hat{\Lambda}^{\mathrm{LS}} / n \tag{4}$$

One of the properties of the PC estimator is that it produces orthogonal loadings and uncorrelated factors (assuming the data have been centered). A consequence of this property is that factors can be estimated sequentially, which is also a very natural way of considering the estimator, i.e. an estimator where subsequent factors explain as much of the residual variance as possible. When considering variations of this standard PC estimation approach and subsequently the accompanying theory, it is often easier to work with such a sequential approach to the estimation. We must, however, keep in mind that what consequences this has on the estimator at hand.

2.2. A LASSO regularized approach

The basic observation underlying the idea of introducing sparsity in the loadings is that in the classical case the PC estimated factors are linear combinations of all the X-variables, see (4). Some of the loadings may be very small but never zero. Hence, even though the estimated factors allow us to be very parsimonious in the forecasting equation, (2), the factors are by no means parsimonious. It would therefore seem interesting to modify the PC estimator such that the estimated loadings will be sparse and hence parsimony will be achieved also in the factors. For this purpose we will employ a LASSO penalized version of the PC estimator, which we will denote a sparse principal components (SPC) estimator. Consider first the problem of estimating a single factor. Simply augmenting the least squares criterion in (3) with a LASSO penalty will give us the following objective function:

$$V^{\mathrm{LASSO}}(\underline{F}, \underline{\lambda}; X, \psi_T) = (nT)^{-1} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \lambda_i F_t)^2 + \psi_T \sum_{i=1}^{n} |\lambda_i| \right] \tag{5}$$

Note that the function is written in terms of $\underline{F}$ and $\underline{\lambda}$ to make it explicit that we are only estimating a single factor. Furthermore, the objective function now also depends on the LASSO tuning parameter ψ_T.

One of the appealing features that this estimation method has, is that just as in the PC case the estimated factor will be a linear combinations of the X-variables:

$$\widehat{\underline{F}}_1^{\text{LASSO}} = X\hat{\underline{\lambda}}_1^{\text{LASSO}}/n \tag{6}$$

However, the crucial difference is that the loadings will now be sparse, in the sense that some of the entries of $\hat{\underline{\lambda}}_1^{\text{LASSO}}$ will be zero. Hence, the factor may depend only on a subset of the X-variables.

Commonly, we are of course interested in more than one factor. Subsequent factors can be estimated in a sequential approach as detailed in the following definition:

Definition 1. *Sparse PC Factor Estimator: The SPC estimates of the first factor and associated loadings are defined as:*

$$(\widehat{\underline{F}}_1, \hat{\underline{\lambda}}_1) = \underset{\underline{F},\underline{\lambda}}{\operatorname{argmin}}\, V^{\text{LASSO}}(\underline{F}, \underline{\lambda}; X, \psi_T) \quad s.t. \quad \underline{\lambda}'\underline{\lambda}/n = 1$$

Let the residuals from the estimation of the k*th factor be defined as* e_k*, then for* $k > 1$ *the subsequent estimates are given as:*

$$(\widehat{\underline{F}}_k, \hat{\underline{\lambda}}_k) = \underset{\underline{F},\underline{\lambda}}{\operatorname{argmin}}\, V^{\text{LASSO}}(\underline{F}, \underline{\lambda}; e_{k-1}, \psi_T) \quad s.t. \quad \underline{\lambda}'\underline{\lambda}/n = 1$$

Hence the SPC factor estimates of r *factors and associated loadings are given as* $\widehat{F} = (\widehat{\underline{F}}_1, \dots, \widehat{\underline{F}}_r)$ *and* $\hat{\Lambda} = (\hat{\underline{\lambda}}_1, \dots, \hat{\underline{\lambda}}_r)$.

The sparsity of the estimator, unfortunately, comes at a cost. One of key features of the PC estimator is lost, orthogonality of the loadings, and hence relation (6) only holds for the first factor. Subsequent factors will also have an additive term relating to previous factors. Note, however, that this is a finite sample feature. Asymptotically the loadings will still be orthogonal as we shall see below.

In practice the estimation of the model is computationally more involved than the usual PC estimation. However, Shen and Huang (2008) provide a simple and fast estimation method for minimizing (5) presented in Algorithm 1 below.

Algorithm 1. *Sparse PCA via regularized SVD (Shen and Huang, 2008, Alg. 1).*

Apply SVD to obtain the a rank 1 approximation of the data $X = usv'$. *Set* $\underline{\lambda}^{(0)} = sv$ *and* $\underline{F}^{(0)} = u$*, hence the latter is the first PC normalized to have length one, and the former is the (non-normalized) loadings. Then step* i *of the algorithm is given as:*

1. *Compute penalized loadings:* $\underline{\lambda}^{(i)} = \operatorname{sgn}\left(X'\underline{F}^{(i-1)}\right)\max\left(\left|X'\underline{F}^{(i-1)}\right| - \psi_T, 0\right)$.
2. *Compute normalized factor:* $\underline{F}^{(i)} = X\underline{\lambda}^{(i)}/\left\|X\underline{\lambda}^{(i)}\right\|_2$.
3. *Check for convergence.*

When convergence is achieved after k *iterations, normalize the factor and loadings to get the final estimates:* $\underline{\lambda} = \underline{\lambda}^{(k)}/\left\|\underline{\lambda}^{(k)}\right\|_2$*, and* $\underline{F} = \underline{F}^{(k)}\left\|\underline{\lambda}^{(k)}\right\|_2$.

The algorithm is based on a simple alternating strategy which is often encountered in bilinear models. In the PC literature, for example, an analog is the NIPALS algorithm (Esbensen, Geladi, and Wold, 1987), and a similar approach is also used in robust factor model estimation, see e.g. Croux and Exterkate (2011) and Kristensen (2013). The basic idea is that for a given F or Λ the problem reduces to a number of linear regressions. In our case with the LASSO penalty we see in step 1 that for a given Λ, F is estimated by the standard LASSO, and in fact in this simple case with only a single regressor the solution has a closed form (Tibshirani, 1996). Now, for a given Λ there is no penalty term, and F is obtained by standard regression. This is step 2. Note, however, as discussed above we need to impose restrictions to identify the parameters. This is usually done by restricting the length of Λ. However, because we use the LASSO to estimate Λ this is not easily done and the algorithm therefore restricts the length of F instead. This of course has no implications for the final estimates as long as we remember to rescale them.

In the next section we show that this approach yields consistent estimates of the factors. However, in spite of this one could still be worried that the penalty will induce a bias in the factors which could affect the performance of the estimator in finite samples. A typical solution to this problem is to only use the LASSO as a variable selection device and rerun the estimation with only the selected variables. Such an approach is often referred to as Post-LASSO, see e.g. Belloni and Chernozhukov (2013). In our case this can easily be accomplished by modifying the first step of Algorithm 1. Here instead of running a LASSO regression to obtain the loadings we run an OLS regression but only for the variables which have been selected, the rest are set equal to zero. This approach is summarized in Algorithm 2 and will be referred to as Post-SPC when used in the context of Definition 1. In the simulations and empirical application below we will include it to assess the severity of the bias introduced by the penalty.

Algorithm 2. *Post-SPC. Let $\widehat{\underline{F}}$ and $\hat{\underline{\lambda}}$ be the estimates from Algorithm 1, and let $\underline{X}_j$ be the jth column of X. Then step i of the algorithm is given as:*

1. *Compute loadings:* $\underline{\lambda}_j^{(i)} = \begin{cases} \left(\underline{F}^{(i-1)\prime}\underline{F}^{(i-1)}\right)^{-1}\underline{F}^{(i-1)\prime}\underline{X}_j & \text{if } \hat{\underline{\lambda}}_j \neq 0 \\ 0 & \text{if } \hat{\underline{\lambda}}_j = 0 \end{cases}$ *for j=1,...,n*
2. *Compute normalized factor:* $\underline{F}^{(i)} = X\underline{\lambda}^{(i)}/\|X\underline{\lambda}^{(i)}\|_2$.
3. *Check for convergence.*

When convergence is achieved after k iterations, normalize the factor and loadings to get the final estimates: $\underline{\lambda} = \underline{\lambda}^{(k)}/\|\underline{\lambda}^{(k)}\|_2$, *and* $\underline{F} = \underline{F}^{(k)}\|\underline{\lambda}^{(k)}\|_2$.

3. Asymptotic properties

For the asymptotic analysis of the SPC estimator we adopt the asymptotic framework of Stock and Watson (2002a) and hence Assumptions 1 and 2 below are identical to their Assumptions F1 and M1. In addition to this we need to make assumptions regarding the LASSO penalty, these are stated in Assumption 3.

Assumption 1. *Factors and factor loadings*

a. $\Lambda'\Lambda/n \to I_r$.

b. $\mathbb{E}\left[F_t F_t'\right] = \Sigma_{FF}$, *where* Σ_{FF} *is a diagonal matrix with elements* $\sigma_{ii} > \sigma_{jj} > 0$ *for* $i < j$.

c. $\left|\lambda_{i,j}\right| \leq \bar{\lambda} < \infty$ *for* $1 \leq i \leq n$, $1 \leq j \leq r$.

d. $T^{-1}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_{FF}$.

Assumption 2. *Moments of the errors*

a. $\mathbb{E}\left[e_t' e_{t+u}/n\right] = \gamma_{n,t}(u)$, *and* $\lim_{n\to\infty}\sup_t \sum_{u=-\infty}^{\infty}\left|\gamma_{n,t}(u)\right| < \infty$.

b. $\mathbb{E}\left[e_{it}e_{jt}\right] = \tau_{ij,t}$, $\lim_{n\to\infty}\sup_t n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\tau_{ij,t}\right| < \infty$.

c. $\lim_{n\to\infty}\sup_{t,s} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\operatorname{cov}(e_{is}e_{it}, e_{js}e_{jt})\right| < \infty$.

Assumption 3. *LASSO penalty*

a. $T^{-1}\psi_T \to 0$.

Notice that we make no explicit assumptions on the sparsity of the loadings. We only require that they conform to Assumption 1a which is a standard assumption for the PC estimator. However, by choosing SPC over PC we implicitly assume that the loadings are sparse or at least that a sparse representation is more suitable for the application at hand, e.g. in a bias/variance trade-off sense. We will show that the penalty introduced in the SPC estimator does not interfere with the asymptotics and that the estimator is therefore still consistent. Hence, the sparsity imposed by the SPC estimator can be seen as a finite sample correction. In order to do so we need only restrict the speed at which the penalty may tend to infinity (Assumption 3). This is similar to the standard results on the LASSO estimator in e.g. Knight and Fu (2000).

Under this set of assumptions we can show that the SPC factor estimator will be consistent as summarized in the following theorem:

Theorem 1. *Let* S_i *denote a variable with values of* ± 1, *let* $n, T \to \infty$, *and suppose that assumptions 1–3 hold. Then* S_i *can be chosen such that*

a. $S_i \widehat{F}_{it} \xrightarrow{p} F_{it}$ *for* $i = 1, 2, \ldots, r$.

b. $T^{-1}\sum_{t=1}^{T}\widehat{F}_{it}^2 \xrightarrow{p} 0$ *for* $i = r+1, \ldots, k$.

The intuition behind the results is quite simple. Since the LASSO penalty is $o(T)$ the penalty term will disappear asymptotically. Hence the proof, which is given in the appendix, shows that the penalized objective function for a single factor converges uniformly to the asymptotic objective function considered in Stock and Watson (2002a). Extending the results to all r factors is then done using a sequential argument where the established consistency of the previous factors is used.

3.1. Forecasting and determining the number of factors

Before the estimated factors can be used in the forecasting equation we need to be able to determine the number of factors r. A number of methods for doing this have been proposed for the PC factor estimator, with the IC_p information criteria of Bai and Ng (2002) being the most commonly used. Alternatively BIC is also often used in spite of it not being consistent (Stock and Watson, 1998). It has, however, shown to give good results empirically. Due to the existence of methods to determine r the results on the forecasting equation in Stock and Watson (2002a) take r to be known. We start by stating a set of assumptions about the forecasting equation:

Assumption 4. *Forecasting equation. Let* $z_t = (F_t', w_t')'$ *and* $\beta = (\beta_F', \beta_w')'$ *then:*

a. $\mathbb{E}\left[z_t z_t'\right] = \Sigma_{zz} = \begin{bmatrix} \Sigma_{FF} & \Sigma_{Fw} \\ \Sigma_{wF} & \Sigma_{ww} \end{bmatrix}$ *is positive definite.*

b. $T^{-1}\sum_t z_t z_t' \xrightarrow{p} \Sigma_{zz}$.

c. $T^{-1}\sum_t z_t \varepsilon_{t+h} \xrightarrow{p} 0$.

d. $T^{-1}\sum_t \varepsilon_{t+h}^2 \xrightarrow{p} \sigma^2$.

e. $|\beta_i| < \infty$ *for* $1 \le i \le r+q$.

Assumption 4 corresponds to Assumption Y1 in Stock and Watson (2002a), and as they argue items a–c are standard conditions that imply consistency of the OLS estimator. The added assumptions are needed because the factors are not observed. Based on this set of assumptions we can now restate Theorem 2 of Stock and Watson (2002a) which also holds for the SPC factor estimator by the results of Theorem 1.

Theorem 2 (Stock and Watson (2002a, Thm. 2)). *Let* S_i *denote a variable with values of* ± 1*, let* $n, T \to \infty$*, and suppose that assumptions 1–4 hold. Then* S_i *can be chosen such that*

a. $(\hat{\beta}_F' \widehat{F}_T + \hat{\beta}_w w_T) - (\beta_F' F_T + \beta_w w_T) \xrightarrow{p} 0$.

b. $\hat{\beta}_w - \beta_w \xrightarrow{p} 0$ *and* $S_i \hat{\beta}_{iF} - \beta_{iF} \xrightarrow{p} 0$ *for* $i = 1, \ldots, r$.

There is, however, a caveat associated with this result. Since we have not verified that any of the existing methods for determining r also apply to the SPC estimator it is problematic to assume r known. The results in Stock and Watson (2002a) do not immediately extend to the case where r is unknown. One reason for this is that in the case where one estimates more than r factors the OLS estimator will be (asymptotically) infeasible due to singularities. A common solution to the problem of a singular design is to replace OLS by e.g. ridge regression. We therefore propose determining the set of relevant factors using ridge regression prior to estimating the forecasting equation. We thereby view the problem of determining r as a variable selection problem as detailed in the following definition:

Definition 2. *Thresholded ridge regression. The method consists of the following three steps:*

1. *Run ridge regression:*

$$\hat{\beta}^{\mathrm{RR}} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T-h} \left(y_{t+h} - \sum_{i=1}^{k} \beta_i \widehat{F}_{it}\right)^2 + \kappa_T \sum_{i=1}^{k} \beta_i^2$$

2. *Select factors for which* $\left|\hat{\beta}_i^{\mathrm{RR}}\right| > \beta_{\mathrm{thr}}$.

3. *Make forecasts based on OLS estimates obtained from the forecasting equation including only selected factors:*

$$y_{t+h} = \sum_{\{i:|\hat{\beta}_i^{\mathrm{RR}}|>\beta_{\mathrm{thr}}\}} \beta_{Fi} \widehat{F}_{it} + \beta_w' w_t + \varepsilon_{t+h} \tag{7}$$

In order to show that this method yields consistent results we will make the following set of assumptions:

Assumption 5. *Thresholded ridge regression*

a. $\Sigma_{Fw} = 0$.

b. $T^{-1}\kappa_T \to \kappa$ *where* $0 < \kappa < \infty$.

c. $0 < \beta_{\text{thr}} < \min_{\{i:\beta_{Fi} \neq 0\}}((\sigma_{ii} + \kappa)^{-1}\sigma_{ii}|\beta_{Fi}|$.

The first item of the assumption is clearly the most debatable, we need to have that (asymptotically) there is no covariation between the factors and the other observable variables. This might not always be the case. The second item requires the penalty parameter to converge to a positive finite number. This ensures that the problem is well-defined asymptotically. The estimator will of course be inconsistent, however, since we only want to select factors this is not a problem. Finally the last item requires us to have knowledge of a lower bound of the relevant parameters. In practice, however, we simply set this threshold to a low value. As we shall see later using a fraction of the smallest OLS estimate as the threshold appears to yield good results. Based on these assumptions we can state the following theorem:

Theorem 3. *Let S_i denote a variable with values of* ± 1, *let* $n, T \to \infty$, *and suppose that assumptions 1–5 hold. Then S_i can be chosen such that*

a. $S_i \hat{\beta}_{Fi} \xrightarrow{p} \beta_{Fi}$ *for* $i \leq r$.

b. $P(\hat{\beta}_{Fi} = 0) \to 1$ *for* $i > r$.

c. $\hat{\beta}_w \xrightarrow{p} \beta_w$.

Hence, just as it was the case in Theorem 2 we achieve consistency of the coefficients associated with the factors and the other observable variables. However, the crucial difference is that with probability tending to one we will only include the true factors, i.e. the coefficients associated with the superfluous factors will be exactly equal to zero. Another diffence is the implicit assumption often made that the variable being forecast is related to all r factors. In the way Assumption 5 is defined this need not be the case for Theorem 3 to hold. Finally, we should note that Theorem 3 holds for both the PC and SPC factor estimators.

4. Monte Carlo evidence

For our Monte Carlo analysis we will use a simplified version of the setup considered in Stock and Watson (2002a) The data-generating process will be:

$$x_{it} = \lambda_i F_t + e_{it}$$
$$(1 - aL)e_{it} = (1 + b^2)v_{it} + bv_{i+1,t} + bv_{i-1,t}$$

Hence, we include the possibility that the error term e_{it} is correlated i.e. it will be serially correlated with an AR(1) coefficient of a and cross-series correlated with a (spatial) MA(1) coefficient b. The error is driven by the random variable v_{it} which will be standard normal. Finally, both the factors F_t and loadings λ_i will be generated as independent standard normal variables. We will impose sparsity on the generated loadings by setting a fraction τ of them equal to zero. For the SPC estimator we need to select the LASSO tuning-parameter ψ_T. This will be done using a BIC-type criterion:

$$\psi_T = \underset{\psi_T}{\operatorname{argmin}} \log\left(nT^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[x_{it} - \hat{\lambda}_i(\psi_T)'\widehat{F}_t(\psi_T)\right]^2\right) + m\frac{\log(nT)}{nT}$$

where $\hat{\lambda}_i(\psi_T)$ and $\widehat{F}_t(\psi_T)$ are the SPC estimates for a given ψ_T, and m is the number of non-zero entries in $\hat{\Lambda}$. In the following we determine ψ_T by a simple grid-search. Three different goals will be considered when judging the performance of the estimators; i) ability to correctly estimate the number of factors, ii) precision of the estimated factors, iii) ability to correctly estimate the loadings as being sparse.

In order to determine the number of factors we will employ three different methods, namely the IC_p information criteria of Bai and Ng (2002), the BIC as used in Stock and Watson (2002b), and the thresholded ridge regression proposed above (RR). As the latter two are defined in terms of the forecasting relationship we need to have a variable to forecast, hence we also generate a uni-variate time series to which these methods are applied:

$$y_{t+1} = \iota' F_t + \epsilon_{t+1}$$

where ι is a vector of ones and ϵ_{t+1} is an independent standard normal error term. For the RR method we must select the penalty parameter and the parameter threshold. Let $\hat{\beta}^{\text{OLS}}$ be the OLS estimates of the forecasting relationship, i.e. an unpenalized version of the RR regression, then we set: $\beta_{\text{thr}} = 0.5\min_i\left|\hat{\beta}_i^{\text{OLS}}\right|$. We further select κ_T by applying BIC to the forecasting equation only including the selected factors, i.e. equation (7) in Definition 2.

Assessing the precision of the factor estimates is done, as is common in the literature, by computing the trace R^2 of a multivariate regression of the factor estimates on the true factors

$$R^2 = \operatorname{tr}\left[F'\widehat{F}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'F\right]/\operatorname{tr}\left[F'F\right]$$

and averaging this across Monte Carlo replications. Hence we obtain a statistic that measure how well the estimated factors span the space of the true factors, with values close to 1 being the desired goal.

In Table 1 we give results for three scenarios with a moderate number of zero-entries in the loadings matrix, i.e. $\tau = 0.4$. The first scenario is the very simple case of only a single factor and i.i.d. error terms. In terms of determining the number of factors all three IC_p criteria do very well. IC_3 often has a tendency to overestimate the number of factors, however, this is only seen for the smallest sample size in the PC case. Both the BIC and RR overestimate the number of factors slightly. These results translate directly into the precision of the factors. However, at the largest sample size all methods for determining the number of factors perform comparably and the estimated factors are very close to the true ones and on par with the benchmark case where the number of factors are taken to be known, i.e. $k = r$. In Table 2 the sparsity of the loadings for the SPC estimator is illustrated. We see that the estimator has a tendency to set too many loadings equal to zero. The best case is for $n = 50$, $T = 200$ where the fraction is 0.43. As n increases from this point keeping T fixed we

Table 1. Simulation results for three different scenarios where $\tau = 0.4$. Estimated number of factors and precision of the estimates.

	Data-generating process						Estimated number of factors					Factor R^2 for various choices of k					
Est.	n	T	r	a	b	τ	IC_1	IC_2	IC_3	BIC	RR	IC_1	IC_2	IC_3	BIC	RR	$k=r$
PC	25	50	1	0	0	0.4	1.00	1.00	1.30	1.12	1.47	0.91	0.91	0.91	0.91	0.91	0.91
PC	25	100	1	0	0	0.4	1.00	1.00	1.00	1.06	1.31	0.92	0.92	0.92	0.92	0.92	0.92
PC	50	100	1	0	0	0.4	1.00	1.00	1.00	1.06	1.30	0.96	0.96	0.96	0.96	0.96	0.96
PC	50	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.16	0.96	0.96	0.96	0.96	0.96	0.96
PC	100	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.17	0.98	0.98	0.98	0.98	0.98	0.98
PC	150	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.19	0.98	0.98	0.98	0.98	0.98	0.98
SPC	25	50	1	0	0	0.4	1.00	1.00	1.00	1.12	1.41	0.90	0.90	0.90	0.90	0.91	0.90
SPC	25	100	1	0	0	0.4	1.00	1.00	1.00	1.06	1.25	0.92	0.92	0.92	0.92	0.92	0.92
SPC	50	100	1	0	0	0.4	1.00	1.00	1.00	1.06	1.22	0.96	0.96	0.96	0.96	0.96	0.96
SPC	50	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.17	0.96	0.96	0.96	0.96	0.96	0.96
SPC	100	200	1	0	0	0.4	1.00	1.00	1.00	1.01	1.16	0.98	0.98	0.98	0.98	0.98	0.98
SPC	150	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.15	0.98	0.98	0.98	0.98	0.98	0.98
Post-SPC	25	50	1	0	0	0.4	1.00	1.00	1.00	1.11	1.41	0.91	0.91	0.91	0.91	0.91	0.91
Post-SPC	25	100	1	0	0	0.4	1.00	1.00	1.00	1.07	1.28	0.92	0.92	0.92	0.92	0.92	0.92
Post-SPC	50	100	1	0	0	0.4	1.00	1.00	1.00	1.06	1.25	0.96	0.96	0.96	0.96	0.96	0.96
Post-SPC	50	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.16	0.96	0.96	0.96	0.96	0.96	0.96
Post-SPC	100	200	1	0	0	0.4	1.00	1.00	1.00	1.02	1.16	0.98	0.98	0.98	0.98	0.98	0.98
Post-SPC	150	200	1	0	0	0.4	1.00	1.00	1.00	1.03	1.17	0.98	0.98	0.98	0.98	0.98	0.98
PC	25	50	4	0	0	0.4	3.99	3.98	6.44	3.81	3.49	0.91	0.90	0.91	0.85	0.74	0.91
PC	25	100	4	0	0	0.4	4.00	4.00	4.00	3.87	3.58	0.91	0.91	0.91	0.87	0.78	0.91
PC	50	100	4	0	0	0.4	4.00	4.00	4.00	3.85	3.57	0.95	0.95	0.95	0.92	0.82	0.95
PC	50	200	4	0	0	0.4	4.00	4.00	4.00	3.89	3.68	0.96	0.96	0.96	0.93	0.86	0.96
PC	100	200	4	0	0	0.4	4.00	4.00	4.00	3.92	3.71	0.98	0.98	0.98	0.95	0.88	0.98
PC	150	200	4	0	0	0.4	4.00	4.00	4.00	3.91	3.68	0.98	0.98	0.98	0.96	0.88	0.98
SPC	25	50	4	0	0	0.4	3.93	3.86	3.99	3.85	3.60	0.89	0.88	0.90	0.85	0.76	0.90
SPC	25	100	4	0	0	0.4	3.99	3.99	4.00	3.90	3.63	0.91	0.91	0.91	0.88	0.80	0.91
SPC	50	100	4	0	0	0.4	4.00	4.00	4.00	3.89	3.64	0.95	0.95	0.95	0.92	0.84	0.95
SPC	50	200	4	0	0	0.4	4.00	4.00	4.00	3.90	3.68	0.96	0.96	0.96	0.93	0.87	0.96
SPC	100	200	4	0	0	0.4	4.00	4.00	4.00	3.94	3.76	0.98	0.98	0.98	0.96	0.90	0.98
SPC	150	200	4	0	0	0.4	4.00	4.00	4.00	3.94	3.75	0.98	0.98	0.98	0.96	0.90	0.98
Post-SPC	25	50	4	0	0	0.4	3.95	3.90	4.01	3.83	3.55	0.90	0.89	0.90	0.85	0.76	0.90
Post-SPC	25	100	4	0	0	0.4	3.99	3.99	4.00	3.88	3.60	0.91	0.91	0.91	0.88	0.79	0.91
Post-SPC	50	100	4	0	0	0.4	4.00	4.00	4.00	3.86	3.61	0.95	0.95	0.95	0.92	0.83	0.95
Post-SPC	50	200	4	0	0	0.4	4.00	4.00	4.00	3.89	3.66	0.96	0.96	0.96	0.93	0.86	0.96
Post-SPC	100	200	4	0	0	0.4	4.00	4.00	4.00	3.92	3.73	0.98	0.98	0.98	0.95	0.89	0.98
Post-SPC	150	200	4	0	0	0.4	4.00	4.00	4.00	3.92	3.71	0.98	0.98	0.98	0.96	0.89	0.98
PC	25	50	4	0.5	1	0.4	8.00	7.30	8.00	4.17	3.58	0.60	0.56	0.60	0.39	0.33	0.39
PC	25	100	4	0.5	1	0.4	8.00	7.96	8.00	4.95	4.01	0.53	0.52	0.53	0.40	0.33	0.36
PC	50	100	4	0.5	1	0.4	6.09	2.19	8.00	5.45	4.50	0.59	0.36	0.67	0.58	0.47	0.52
PC	50	200	4	0.5	1	0.4	4.38	2.75	8.00	5.85	4.86	0.55	0.45	0.66	0.61	0.52	0.56
PC	100	200	4	0.5	1	0.4	3.99	3.63	8.00	5.01	4.66	0.78	0.74	0.83	0.78	0.71	0.79
PC	150	200	4	0.5	1	0.4	4.01	3.93	8.00	4.53	4.36	0.87	0.86	0.88	0.84	0.77	0.87
SPC	25	50	4	0.5	1	0.4	2.69	1.41	7.31	3.71	3.24	0.27	0.18	0.53	0.36	0.32	0.38
SPC	25	100	4	0.5	1	0.4	5.51	3.40	7.89	4.59	3.74	0.40	0.29	0.50	0.38	0.32	0.35
SPC	50	100	4	0.5	1	0.4	1.50	1.15	4.72	4.85	4.11	0.28	0.22	0.52	0.53	0.46	0.50
SPC	50	200	4	0.5	1	0.4	2.20	1.77	4.60	5.30	4.61	0.39	0.33	0.56	0.59	0.52	0.55
SPC	100	200	4	0.5	1	0.4	3.11	2.50	4.11	4.58	4.47	0.66	0.55	0.79	0.78	0.74	0.79
SPC	150	200	4	0.5	1	0.4	3.66	3.13	4.11	4.23	4.32	0.81	0.71	0.87	0.86	0.82	0.87
Post-SPC	25	50	4	0.5	1	0.4	3.56	1.82	7.37	3.73	3.29	0.33	0.21	0.54	0.36	0.32	0.38
Post-SPC	25	100	4	0.5	1	0.4	6.16	4.35	7.83	4.73	3.84	0.44	0.34	0.51	0.39	0.33	0.36
Post-SPC	50	100	4	0.5	1	0.4	1.92	1.32	5.92	4.95	4.18	0.33	0.25	0.57	0.54	0.46	0.51
Post-SPC	50	200	4	0.5	1	0.4	2.68	2.09	5.81	5.53	4.72	0.44	0.38	0.60	0.60	0.52	0.56
Post-SPC	100	200	4	0.5	1	0.4	3.52	3.04	4.67	4.63	4.50	0.73	0.64	0.80	0.78	0.74	0.79
Post-SPC	150	200	4	0.5	1	0.4	3.88	3.59	4.59	4.27	4.33	0.85	0.80	0.87	0.85	0.82	0.87

Note: The results are based on 1,000 Monte Carlo replications.

Table 2. Simulation results for three different scenarios where $\tau = 0.4$. Sparsity of the loadings for the SPC estimator.

Data-generating process							Fraction of zero-entries							
n	T	r	a	b	τ	ψ_T	$\hat{\underline{\lambda}}_1$	$\hat{\underline{\lambda}}_2$	$\hat{\underline{\lambda}}_3$	$\hat{\underline{\lambda}}_4$	$\hat{\underline{\lambda}}_5$	$\hat{\underline{\lambda}}_6$	$\hat{\underline{\lambda}}_7$	$\hat{\underline{\lambda}}_8$
25	50	1	0	0	0.4	1.67	0.49	0.83	0.86	0.88	0.88	0.89	0.90	0.91
25	100	1	0	0	0.4	1.62	0.44	0.82	0.84	0.86	0.87	0.88	0.88	0.89
50	100	1	0	0	0.4	1.86	0.47	0.89	0.90	0.91	0.91	0.92	0.92	0.93
50	200	1	0	0	0.4	1.74	0.43	0.86	0.87	0.88	0.89	0.90	0.90	0.91
100	200	1	0	0	0.4	1.99	0.45	0.91	0.92	0.93	0.93	0.93	0.94	0.94
150	200	1	0	0	0.4	2.12	0.46	0.94	0.94	0.95	0.95	0.95	0.95	0.95
25	50	4	0	0	0.4	1.72	0.33	0.37	0.41	0.47	0.82	0.85	0.87	0.87
25	100	4	0	0	0.4	1.48	0.24	0.25	0.28	0.31	0.77	0.80	0.81	0.82
50	100	4	0	0	0.4	1.77	0.27	0.29	0.31	0.33	0.86	0.88	0.89	0.90
50	200	4	0	0	0.4	1.61	0.20	0.21	0.22	0.23	0.82	0.84	0.85	0.86
100	200	4	0	0	0.4	1.83	0.23	0.23	0.24	0.24	0.88	0.90	0.90	0.91
150	200	4	0	0	0.4	1.96	0.24	0.24	0.25	0.26	0.91	0.92	0.92	0.93
25	50	4	0.5	1	0.4	4.68	0.54	0.61	0.67	0.71	0.73	0.76	0.78	0.81
25	100	4	0.5	1	0.4	4.18	0.43	0.49	0.53	0.57	0.60	0.62	0.64	0.66
50	100	4	0.5	1	0.4	5.79	0.57	0.62	0.67	0.72	0.76	0.78	0.80	0.82
50	200	4	0.5	1	0.4	5.10	0.43	0.46	0.51	0.57	0.63	0.67	0.69	0.71
100	200	4	0.5	1	0.4	6.42	0.52	0.55	0.57	0.61	0.78	0.81	0.83	0.84
150	200	4	0.5	1	0.4	7.04	0.56	0.59	0.61	0.63	0.84	0.86	0.87	0.88

Note: The results are based on 1,000 Monte Carlo replications.

moved away from the true value of $\tau = 0.4$, hence it could appear that T needs to be large compared to n for this method to perform well. The estimated loadings of the superfluous factors generally have a very large fraction of zero-entries.

In the second scenario we increase the number of factors to 4. Now, BIC and RR tend to underestimate the number of factors. Further, it again appear that the size of T relative to n is crucial, e.g. moving from $n = 100$, $T = 200$ to $n = 150$, $T = 200$ the performance of BIC and RR declines. The SPC estimator now tends to set too few loadings equal to zero. However, there is still a clear distinction between the true factors and the superfluous factors in terms of the fraction of zero-entries. In the third scenario we introduce correlation in the error terms. This clearly makes the model more difficult to estimate. Note, however, that part of this is most likely due to the fact that the unconditional variance of the error term is larger compared to the two previous scenarios and hence the signal-to-noise ratio is lower. One interesting observation in this scenario is that IC_3 completely breaks down for the PC estimator but does very well for the SPC estimator. BIC and RR tend to overestimate the number of factors for large samples. However, especially for small samples they appear to give more reliable results than the IC_p criteria. For the SPC estimator we again see that there is a tendency to set too many loadings equal to zero.

When comparing the use of SPC and Post-SPC there seems to be little difference. Especially for the precision of the factor estimates we only see minute improvements for the smallest sample size. Hence, it does not appear that SPC introduces any noteworthy bias in the factors, a somewhat surprising result. It could, however, still be that the loadings are biased and thus that the use of Post-SPC could correct this. As the loadings are not of direct interest to us, we have, however, not investigated this possibility.

Tables A.3 and A.1 in the appendix provides a similar set of results for the case of $\tau = 0.8$. One of the most striking results is in the last scenario. Here the three IC_p criteria break down completely for both PC and SPC, whereas both BIC and RR perform quite well with the latter having a slight edge. One could perhaps also argue that SPC does a slightly better job of estimating the fraction of zero-entries compared to Table 1. However, we do still see both cases of under- and overestimation. For reference we have also considered the case of $\tau = 0$ to see how much is lost by using SPC when the true model is not sparse. These results are reported in Tables A.4 and A.2 in the appendix. In general PC is preferred when the true model is not sparse. However, the SPC estimator is not lagging far behind, and in fact we see in the case of correlated error terms that the IC_p criteria become more reliable when using the SPC estimator. This is likely due to the fact that loadings of superfluous factors are heavily penalized and hence the information criteria are less inclined to include them.

In general it appears that we can capture the sparsity of the models using the SPC estimator to some degree. However, it is not clear that we necessarily approach the true fraction of zero-entries as the sample size increases, and hence it could appear that our methodology has some outstanding issues. On possible culprit could be the BIC-type information criterion we use to select ψ_T. Another, possible problem is that we use the same ψ_T for all factors. One could imagine using different penalties for each factor. However, since we select ψ_T by a grid search the computational burden would quickly increase and we have therefore not pursued this possibility. Clearly, determining ψ_T is an area with much potential for future research.

Regarding the number of factors, it does appear that the thresholded ridge regression does have acceptable performance. In the simple cases the IC_p criteria are always preferred, but it could appear that RR could have merits in more complex cases. And even if we do not believe the model is sparse the combination of SPC and IC_p could be a more robust method of selecting the number of factors as seen in the case of $\tau = 0$. The RR method does, however, have an additional feature we have not considered in these simulation results, namely that it does not necessarily select consecutive factors. Hence if the variable being forecast is only related to say, the first and third factors it could potentially pick this up whereas the other methods would also include the second factor. We would therefore expect RR to perform comparatively better if this were the case.

5. Forecasting

To illustrate the merits of the SPC estimator we will perform a simulated real-time out-of-sample forecast experiment as in Stock and Watson (2002b). Note, however, that it is not true real-time forecasting as the dataset only contains the final vintages of the variables. The dataset used is from Ludvigson and Ng (2010) and consists of 131 monthly US macroeconomic variables spanning the period 1964:1 to 2007:12, and is an updated version of the dataset used in Stock and Watson (2005).

We will forecast variables similar to Stock and Watson (2002b): the consumer price index, all items (punew); the personal consumption expenditure implicit price deflator (gmdc); the consumer price index less food (puxf); and the producer price index for finished goods (pwfsa). Total industrial production (ips10); real personal income less transfers (a0m051); real manufacturing and trade sales (mtq); and number of employees on nonagricultural payrolls (lhnag). The variables will be forecast for the period 1975:1–2007.12 using an expanding window of data.

For the estimation of the factors the entire dataset is used after being transformed to stationarity and standardized. The forecasts will be obtained as direct forecasts by fitting the forecasting equation

$$y_{t+h}^h = \alpha_h + \beta_h' \widehat{F}_t + \sum_{j=1}^{p} \theta_{h,j} y_{t-j+1} + \epsilon_{t+h} \tag{8}$$

where y_{t+h}^h is defined appropriately according to the assumed integration order of the underlying variable, see Stock and Watson (2002b) for details. In this common form we need to specify the number of factors, k, and the AR lag length, p. In the results presented here we select p using BIC with a maximum value of 6, and determine k either by BIC, RR, or the IC_p criterion of Bai and Ng (2002), or we simply fix it. The maximum number of factors is set to 8. All forecasting results will be reported as mean squared forecast errors (MSFE) relative to the MSFE of an AR(p) forecast where p is again selected using BIC with a maximum value of 6.

A natural alternative to the SPC approach would be to use the LASSO directly in the forecasting equation to select which X-variables should be used for forecasting. Hence in the results we also include this approach, i.e. we estimate

$$y_{t+h}^h = \alpha_h + \beta_h' X_t + \sum_{j=1}^{p} \theta_{h,j} y_{t-j+1} + \epsilon_{t+h}$$

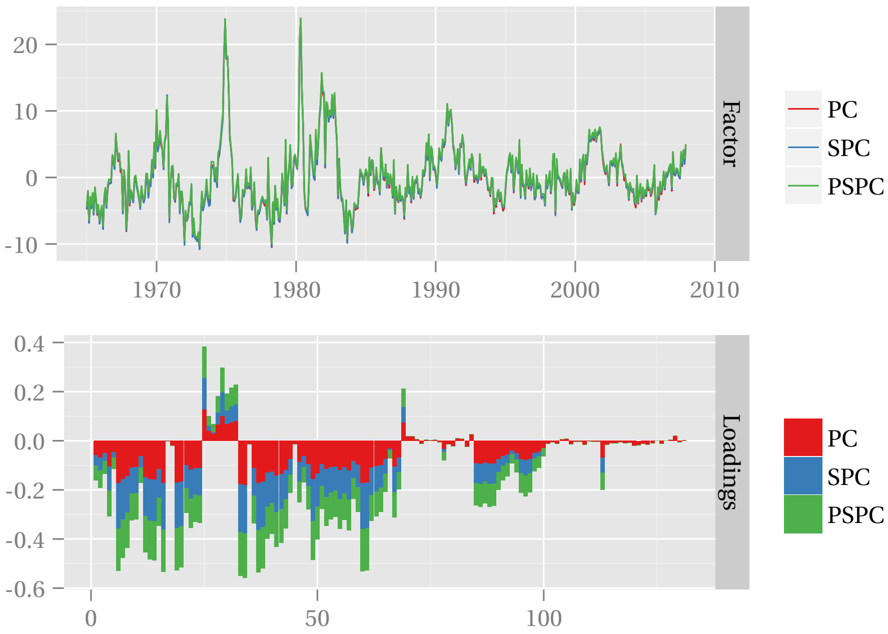
using the LASSO where only β_h is penalized and the tuning parameter is selected using BIC. This will be denoted “LASSO” in the results.[1] Note that in the one factor case this can be seen as an unrestricted version of the SPC approach. Recall that if we estimate a single factor using SPC it will be given as $\widehat{F}_{1,t}^{\mathrm{LASSO}} = \underline{\hat{\lambda}}_1^{\mathrm{LASSO}\prime} X_t / n$. Hence in (8) we have $\beta_h' \widehat{F}_t = \beta_h' \underline{\hat{\lambda}}_1^{\mathrm{LASSO}\prime} X_t / n \equiv \tilde{\beta}_h' X_t$ where $\tilde{\beta}_h = \beta_h' \underline{\hat{\lambda}}_1^{\mathrm{LASSO}} / n$. Sparsity of the loadings will then imply sparsity of $\tilde{\beta}_h$.

Before turning to the actual forecasting results we start by a visual inspection of the consequences for the estimated factors (and loadings) when the LASSO penalty is present. In Figures 1 and 2 the first two estimated factors and associated loadings based on the entire dataset are plotted. In the first plot we see that the estimated factor looks quite similar in all three cases. This is interesting because in the SPC cases roughly 40% of the variables are deemed irrelevant. Hence in appears that the SPC estimator produces a comparable but more parsimonious estimate of the factor. In the second plot we see that for the second factor the effect of the penalty is larger leaving only 37% non-zero entries in the loadings. The plot of the estimated factor also differs more now. Similar plots are provided for the remaining six factors in Figures A.1–A.6 in the appendix.

The forecasting results are presented in Table 3 for the 12-month horizon and in Table 4 for the 6- and 24-month horizons. In a majority of cases either SPC or Post-SPC outperforms PC and only in a single case is the AR model preferred. Even though the performance is quite close in many cases, there are, however, cases where the difference is substantial. For example in the case of industrial production at the six month horizon the best PC forecast has a relative MSFE of 0.9468, whereas it is

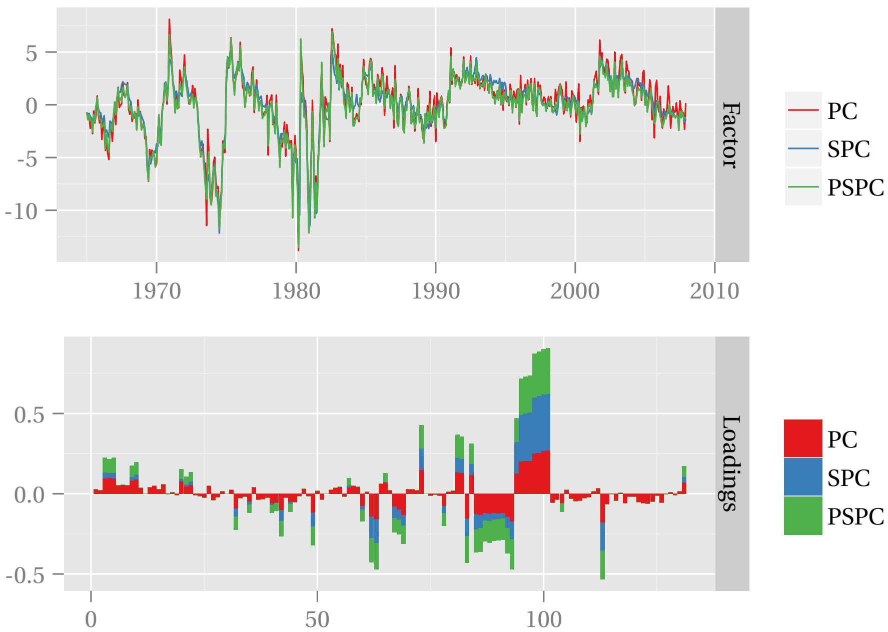
[1] The estimation is carried out using the `glmnet` package of Friedman, Hastie, and Tibshirani (2010).

Figure 1. Estimates of the first factor and associated loadings



Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.634.

Figure 2. Estimates of the second factor and associated loadings



Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.374.

Table 3. Forecasting results for the 12-month horizon.

Est.	k	h	IP	PI	M&TS	Emp.	CPI	C.defl.	CPI exc.	PPI
PC	8	12	0.8686	0.9700	0.8296	0.8353	0.7103	0.8554	0.7949	0.9526
PC	BIC	12	0.8826	1.0004	0.8369	**0.8255**	0.6961	0.8591	0.7921	0.9359
PC	IC_1	12	0.8634	0.9793	0.8309	0.8397	0.7070	0.8564	0.7916	0.9494
PC	IC_2	12	0.8998	1.0009	0.8471	0.8538	0.7008	0.8419	0.7745	0.9420
PC	IC_3	12	0.8686	0.9700	0.8296	0.8353	0.7103	0.8554	0.7949	0.9526
PC	RR	12	0.8886	0.9752	0.8265	0.8371	0.7211	0.8792	0.8093	0.9686
SPC	8	12	0.8889	0.9919	0.8099	0.8583	0.7379	0.8954	0.8178	0.9804
SPC	BIC	12	0.8406	**0.9436**	**0.7701**	0.8386	0.7124	0.8403	0.7918	0.9454
SPC	IC_1	12	0.8537	0.9683	0.7961	0.8292	0.7229	0.8532	0.7706	0.9283
SPC	IC_2	12	0.8749	0.9544	0.8178	0.8633	0.7690	0.9200	0.7887	0.9764
SPC	IC_3	12	0.8817	0.9888	0.8174	0.8541	0.7358	0.8836	0.8102	0.9629
SPC	RR	12	0.8631	0.9650	0.8042	0.8456	0.7015	0.8497	0.7483	**0.9066**
Post-SPC	8	12	0.8649	0.9822	0.7922	0.8444	0.7301	0.8889	0.8211	0.9771
Post-SPC	BIC	12	0.8582	0.9688	0.7850	0.8335	**0.6958**	0.8452	0.8051	0.9549
Post-SPC	IC_1	12	0.8378	0.9653	0.7939	0.8257	0.7068	0.8532	0.7638	0.9412
Post-SPC	IC_2	12	0.8605	0.9493	0.8123	0.8305	0.7296	0.8603	0.7790	0.9638
Post-SPC	IC_3	12	**0.8348**	0.9570	0.8059	0.8590	0.7095	0.8536	0.7894	0.9578
Post-SPC	RR	12	0.8560	0.9845	0.8009	0.8363	0.7034	0.8260	0.7670	0.9261
LASSO		12	1.1426	1.0334	0.8586	0.8928	0.7071	**0.7826**	**0.7131**	0.9549
RMSFE(AR)		12	0.0364	0.0231	0.0334	0.0141	0.0015	0.0011	0.0018	0.0026

Notes: The results are reported as MSFEs relative to AR(p) forecasts with $0 \le 0 \le 6$ choosen by BIC. Bold indicates lowest value in a column, underlined indicates lowest value in a block, i.e. between to horizontal lines. All models include p AR lags where p is chosen by BIC. The last row gives the RMSFE is the benchmark AR model.

0.8329 for the best SPC model. Interestingly, the simple LASSO model does quite well in a number of cases and it is not clear whether LASSO or SPC would be the preferred approach, it all comes down to which variable is being forecast. It is also not clear whether SPC or Post-SPC should be preferred. The performance of the two approaches is very close and it appears quite arbitrary which has the lowest MSFE. This is in line with the Monte Carlo results where their performance was practically identical. As to the problem of determining the number of factors, there is no clear answer to which method to use. But we do see that the RR approach does perform well in a number of cases. This might indicate that for some variables it is beneficial to not include all r factors but only a subset of these.

6. Concluding remarks

In this paper we have investigated the possibility of using sparse principal components to estimate diffusion indexes or factors with sparse loadings. We showed that consistency of the factors, in the sense of Stock and Watson (2002a), was maintained, and proposed a simple alternative to existing methods for determining the number of factors. The methodology still has its shortcomings, and especially the problem of selecting the penalty parameter is an area the requires more research. However, based on both the simulation study and forecasting experiment we are confident that it will prove an important alternative to traditional PC factor estimation in the area of macroeconomic forecasting.

Table 4. Forecasting results for the 6- and 24-month horizons.

Est.	k	h	IP	PI	M&TS	Emp.	CPI	C.defl.	CPI exc.	PPI
PC	8	6	0.9468	0.8508	0.8988	0.7788	0.8323	0.9769	0.8653	1.0500
PC	BIC	6	0.9948	0.8772	0.9432	0.7859	0.8351	0.9721	0.8339	0.9986
PC	IC_1	6	0.9468	0.8632	0.9096	0.7721	0.8370	0.9795	0.8679	1.0536
PC	IC_2	6	0.9594	0.8810	0.8859	0.7530	0.8172	0.9583	0.8486	1.0364
PC	IC_3	6	0.9468	0.8508	0.8988	0.7788	0.8323	0.9769	0.8653	1.0500
PC	RR	6	0.9668	0.8763	0.9251	0.7535	0.8743	0.9219	0.8384	1.0262
SPC	8	6	0.9522	0.8817	0.9465	0.8028	0.8524	1.0157	0.8859	1.0617
SPC	BIC	6	**0.8329**	0.8732	0.8381	0.8088	0.8450	0.9354	0.8461	0.9849
SPC	IC_1	6	0.9202	0.8662	0.8463	0.7577	0.8173	0.9475	0.8484	1.0154
SPC	IC_2	6	0.9075	0.8505	0.8561	0.7662	0.8212	0.9626	0.8362	1.0320
SPC	IC_3	6	0.9332	0.8688	0.8668	0.7981	0.8543	1.0004	0.8808	1.0537
SPC	RR	6	0.9124	0.8844	0.9135	0.7991	0.8260	**0.9159**	0.8334	**0.9823**
Post-SPC	8	6	0.9361	0.8699	0.9149	0.7922	0.8481	1.0224	0.8886	1.0651
Post-SPC	BIC	6	0.8779	0.8667	0.8626	0.8063	0.8336	0.9502	0.8457	0.9826
Post-SPC	IC_1	6	0.9055	0.8632	0.8423	0.7505	**0.8099**	0.9488	0.8389	1.0232
Post-SPC	IC_2	6	0.8600	**0.8262**	**0.8335**	0.7469	0.8133	0.9524	0.8448	1.0225
Post-SPC	IC_3	6	0.9125	0.8521	0.8942	0.7801	0.8315	0.9925	0.8654	1.0477
Post-SPC	RR	6	0.8908	0.8437	0.8663	0.7902	0.8219	0.9213	**0.8215**	0.9868
LASSO		6	0.9893	0.8863	0.8957	**0.7390**	0.8515	0.9268	0.8451	1.0198
RMSFE(AR)		6	0.0219	0.0154	0.0218	0.0080	0.0015	0.0011	0.0018	0.0027
Est.	k	h	IP	PI	M&TS	Emp.	CPI	C.defl.	CPI exc.	PPI
PC	8	24	0.8125	1.1434	0.7980	0.9165	0.6356	0.7488	0.6902	0.8558
PC	BIC	24	0.7830	1.1522	0.7927	0.8953	0.6363	0.7392	0.6640	0.8789
PC	IC_1	24	0.8194	1.1560	0.8222	0.9225	0.6383	0.7512	0.6875	0.8571
PC	IC_2	24	0.8527	1.1721	0.8624	0.9654	0.6425	0.7556	0.6612	0.8629
PC	IC_3	24	0.8125	1.1434	0.7980	0.9165	0.6356	0.7488	0.6902	0.8558
PC	RR	24	0.7942	1.1489	0.8293	0.9079	0.6422	0.7851	0.6521	0.8708
SPC	8	24	0.8333	1.1652	0.7811	0.9176	0.6694	0.8161	0.7165	0.8881
SPC	BIC	24	0.8227	1.1389	0.7979	0.8936	0.6533	0.7717	0.6759	0.8901
SPC	IC_1	24	**0.7787**	1.0786	0.7411	0.8898	0.6633	0.7849	0.6718	0.8607
SPC	IC_2	24	0.7816	**1.0346**	**0.7158**	0.8780	0.7395	0.8749	0.7263	0.9270
SPC	IC_3	24	0.8263	1.1585	0.7865	0.9138	0.6719	0.8128	0.7170	0.8806
SPC	RR	24	0.8131	1.1366	0.8021	0.9024	0.6089	0.7253	0.6657	0.8461
Post-SPC	8	24	0.8354	1.1729	0.8012	0.9248	0.6450	0.7841	0.7023	0.8713
Post-SPC	BIC	24	0.8258	1.1307	0.7801	0.9139	0.6383	0.7623	0.6550	0.8872
Post-SPC	IC_1	24	0.7974	1.1023	0.7707	0.9014	0.6449	0.7747	0.6625	0.8644
Post-SPC	IC_2	24	0.8012	1.0454	0.7512	0.8888	0.6724	0.7957	0.6840	0.8975
Post-SPC	IC_3	24	0.8355	1.1540	0.8291	0.9596	0.6528	0.7845	0.6768	0.8857
Post-SPC	RR	24	0.8037	1.1295	0.7798	0.9233	**0.5844**	0.7074	**0.6332**	0.8462
LASSO		24	0.8877	1.2851	0.7394	**0.8723**	0.7067	**0.6928**	0.7559	**0.8195**
RMSFE(AR)		24	0.0567	0.0365	0.0518	0.0246	0.0018	0.0012	0.0020	0.0027

Notes: The results are reported as MSFEs relative to AR(p) forecasts with $0 \leq 0 \leq 6$ choosen by BIC. Bold indicates lowest value in a column, underlined indicates lowest value in a block, i.e. between to horizontal lines. All models include p AR lags where p is chosen by BIC. The last row in each of the two parts gives the RMSFE is the benchmark AR model.

Appendix: Proofs

We start by introducing some notation. When possible the same, or at least similar, notation will be used as that of Stock and Watson (2002a). The proof of Theorem 1 is sequential in the same manner as Definition 1, i.e. we start by showing uniform convergence of each objective function of the definition to a corresponding asymptotic objective function. Recall the model:

$$X_t = \Lambda F_t + e_t$$

Rewriting the objective function we get

$$V^{\text{LASSO}}(\underline{F}, \underline{\lambda}; X, \psi_T) = (nT)^{-1} \text{tr}\left[(X - \underline{F}\,\underline{\lambda}')'(X - \underline{F}\,\underline{\lambda}')\right] + S(\underline{\lambda})$$

where $S(\underline{\lambda})$ is the LASSO penalty. The expression for the factor estimate is given by the first-order condition:

$$\frac{\partial V^{\text{LASSO}}(\underline{F}, \underline{\lambda}; X, \psi_T)}{\partial \underline{F}} = (nT)^{-1} 2\left[\underline{F}\,\underline{\lambda}'\underline{\lambda} - X\underline{\lambda}\right] = 0 \Leftrightarrow \underline{F} = X\underline{\lambda}/n$$

Hence the concentrated objective function becomes

$$V^{\text{LASSO}}(\underline{\lambda}; X, \psi_T) = (nT)^{-1}\left[\text{tr}(X'X) - \text{tr}(n^{-1}\underline{\lambda}\,\underline{\lambda}'X'X)\right] + S(\underline{\lambda})$$

The first term is independent of $\underline{\lambda}$ and can thus be discarded, changing signs we are therefore left with the equivalent maximization problem:

$$n^{-2}T^{-1}\text{tr}\left[\underline{\lambda}'X'X\underline{\lambda}\right] - S(\underline{\lambda}) = R_1(\underline{\lambda}) - S(\underline{\lambda})$$

Therefore the first step gives us the following estimates: $\hat{\underline{\lambda}}_1 = \text{argmax}_{\underline{\lambda}}\, R_1(\underline{\lambda}) - S(\underline{\lambda})$ and $\widehat{F}_{1t} = \hat{\underline{\lambda}}_1' X_t/n$. The residuals from the jth estimation are $\hat{e}_t^{(j)} = X_t - \sum_{i=1}^{j}\hat{\underline{\lambda}}_i \widehat{F}_{it}$ and hence the estimates of the $(j+1)$th step are: $\hat{\underline{\lambda}}_{j+1} = \text{argmax}_{\underline{\lambda}}\, R_{j+1}(\underline{\lambda}) - S(\underline{\lambda})$ and $\widehat{F}_{j+1,t} = \hat{\underline{\lambda}}_{j+1}'\hat{e}_t^{(j)}/n$. Thus, all objective functions under consideration are constructed from the following quantities:

$$\begin{aligned}
R_1(\gamma) &= n^{-2}T^{-1}\gamma'X'X\gamma = n^{-2}T^{-1}\gamma'\sum_t X_t X_t'\gamma \\
R_l(\gamma) &= n^{-2}T^{-1}\gamma'\sum_t \hat{e}_t^{(l-1)}\hat{e}_t^{(l-1)\prime}\gamma \\
S(\gamma) &= n^{-1}T^{-1}\psi_T\sum_i |\gamma_i|
\end{aligned}$$

and the corresponding asymptotic objective functions will be formed using:

$$\begin{aligned}
R_1^*(\gamma) &= n^{-2}T^{-1}\gamma'\Lambda F'F\Lambda'\gamma = n^{-2}T^{-1}\gamma'\Lambda\sum_t F_t F_t'\Lambda'\gamma \\
&= n^{-2}T^{-1}\gamma'\sum_{i=1}^{r}\underline{\lambda}_i\sum_t F_{it}\sum_{j=1}^{r}F_{jt}\underline{\lambda}_j'\gamma \\
R_l^*(\gamma) &= n^{-2}T^{-1}\gamma'\sum_{i=l}^{r}\underline{\lambda}_i\sum_t F_{it}\sum_{j=l}^{r}F_{jt}\underline{\lambda}_j'\gamma
\end{aligned}$$

where $\sum_t = \sum_{t=1}^T$ and $\sum_i = \sum_{i=1}^n$. In all maximizations we need to impose a length restriction on the loadings, and hence we will make use of the following set: $\Gamma = \{\gamma \mid \gamma'\gamma/n = 1\}$. Finally, due to the sequential nature of the problem we need a few definitions pertaining to the steps of the sequential estimation:

$$\begin{aligned}
\delta_t^{(l)} &= \sum_{i=1}^{l-1} [\underline{\lambda}_i F_{it} - \underline{\hat{\lambda}}_i \widehat{F}_{it}] \\
F^{(l)} &= (\underline{F}_l, \dots, \underline{F}_r) \\
\Lambda^{(l)} &= (\underline{\lambda}_l, \dots, \underline{\lambda}_r) \\
T^{-1} F^{(l)\prime} F^{(l)} &\xrightarrow{p} \Sigma_{FF}^{(l)}
\end{aligned}$$

hence $\Sigma_{FF}^{(l)}$ is a submatrix of Σ_{FF} as defined in Assumption 1.

The proofs are built up from a number of smaller results which we present first, the proofs of the theorems then follow at the end. The proof of Theorem 1 is split in two parts. The first part proves consistency of the first factor, and the second part the subsequent factors. The reason for this is that the first part is almost identical to the proof of Stock and Watson (2002a, Thm. 1b), whereas the second part requires more modifications. All results named **(R#)** are taken from Stock and Watson (2002a), in the cases where small changes to their results are needed a prime has been added to the name and any needed changes in the proofs are detailed. Additional results are labelled **(T#)**.

(R2) $\sup_{\gamma\in\Gamma}(n^2T)^{-1}\gamma' e' e \gamma \xrightarrow{p} 0$

(R3) Let q_t denote a sequence of random variables with $T^{-1}\sum_t q_t^2 \sim O_p(1)$. Then

$$\sup_{\gamma\in\Gamma}\left|T^{-1}\sum_t q_t (N^{-1}\sum_i \gamma_i e_{it})\right| \xrightarrow{p} 0$$

(R4) $\sup_{\gamma\in\Gamma}\left|T^{-1}\sum_t F_{jt}(n^{-1}\sum_i \gamma_i e_{it})\right| \xrightarrow{p} 0$ for $j = 1, 2, \dots, r$

(R5) $\sup_{\gamma\in\Gamma}(n^2T)^{-1}\left|\gamma'\Lambda F' e\gamma\right| \xrightarrow{p} 0$

(R6′) $\sup_{\gamma\in\Gamma}\left|R_1(\gamma) - S(\gamma) - R_1^*(\gamma)\right| \xrightarrow{p} 0.$

Proof.

$$R_1(\gamma) - S(\gamma) - R_1^*(\gamma) = (n^2T)^{-1}\gamma' e' e\gamma + 2(n^2T)^{-1}\gamma'\Lambda F'\gamma - (nT)^{-1}\psi_T \sum_i |\gamma_i|$$

Hence

$$\begin{aligned}
\sup_{\gamma\in\Gamma}\left|R_1(\gamma) - S(\gamma) - R_1^*(\gamma)\right| \le{} & \sup_{\gamma\in\Gamma}(n^2T)^{-1}\left|\gamma' e' e\gamma\right| \\
& + \sup_{\gamma\in\Gamma} 2(n^2T)^{-1}\left|\gamma'\Lambda F' e\gamma\right| + \sup_{\gamma\in\Gamma}(nT)^{-1}\psi_T\sum_i|\gamma_i|
\end{aligned}$$

The first term on the right-hand side converges to 0 by (R2), and the second term converges to 0 by (R5). Consider now the third term:

$$(nT)^{-1}\psi_T\sum_i|\gamma_i| \le T^{-1}\psi_T \to 0$$

The inequality uses the fact that for any $\gamma \in \Gamma$ we have that $n^{-1}\sum_i \gamma_i^2 = 1$ which implies $n^{-1}\sum_i|\gamma_i| \le 1$. The convergence then follows by Assumption 3.a. □

(R7′) $\left|\sup_{\gamma\in\Gamma}[R_1(\gamma)-S(\gamma)]-\sup_{\gamma\in\Gamma}R_1^*(\gamma)\right| \xrightarrow{p} 0.$

Proof. Stock and Watson (2002a, (R7)) using (R6′) instead of (R6). □

(R8) $\sup_{\gamma\in\Gamma}R_1^*(\gamma)\xrightarrow{p}\sigma_{11}$

(R9′) $\sup_{\gamma\in\Gamma}R_1(\gamma)-S(\gamma)\xrightarrow{p}\sigma_{11}$

Proof. Stock and Watson (2002a, (R9)) using (R7′) instead of (R7). □

(R10′) Let $\hat{\underline{\lambda}}_1 = \text{argsup}_{\gamma\in\Gamma}R_1(\gamma)-S(\gamma)$; then $\sup_{\gamma\in\Gamma}R_1^*(\hat{\underline{\lambda}}_1)\xrightarrow{p}\sigma_{11}$

Proof. Stock and Watson (2002a, (R10)) using (R6′) and (R9′) instead of (R6) and (R9), respectively. □

(R11′) $(S_1\hat{\underline{\lambda}}_1'\Lambda/n)\xrightarrow{p}(1,0,0,\ldots,0)'$

Proof. Stock and Watson (2002a, (R11)) using (R10′) instead of (R10). □

(R15′) $S_1\widehat{F}_{1t}-F_{1t}\xrightarrow{p}0$

Proof.

$$\begin{aligned}
S_1\widehat{F}_{1t}-F_{1t} &= S_1\hat{\underline{\lambda}}_1'X_t/n-F_{1t}\\
&= S_1\hat{\underline{\lambda}}_1'\big(e_t+\sum_{i=1}^{r}\underline{\lambda}_iF_{it}\big)/n-F_{1t}\\
&= S_1\hat{\underline{\lambda}}_1'e_t/n+\big(S_1\hat{\underline{\lambda}}_1'\lambda_1/n-1\big)F_{1t}+S_1\hat{\underline{\lambda}}_1'\big(\sum_{i=2}^{r}\underline{\lambda}_iF_{it}\big)/n
\end{aligned}$$

The first term is $o_p(1)$, this is shown in the proof of Stock and Watson (2002a, (R15)). Since $|F_T|$ is $O_p(1)$ by Assumption 1 if follows from (R11′) that the second and third terms are $o_p(1)$ using the same argument as in the proof of (R15). □

(T1) $n^{-1}S_i\hat{\underline{\lambda}}_i'S_j\hat{\underline{\lambda}}_j\xrightarrow{p}0\quad$ for $i\neq j,\quad i,j\leq r$

Proof. Let the loadings be represented as:

$$\hat{\underline{\lambda}}_j=\Lambda^{(j)}(\Lambda^{(j)\prime}\Lambda^{(j)}/n)^{-1/2}\hat{\alpha}_j+\widehat{V}_j$$

where $\widehat{V}_j'\Lambda^{(j)}=0$. Then, by the argument given in (T11) below, it follows that $\widehat{V}_j'\widehat{V}_j/n\xrightarrow{p}0$, $\hat{\alpha}_{j1}^2\xrightarrow{p}1$ and $\hat{\alpha}_{jk}^2\xrightarrow{p}0$ for $k>1$. Assume $i>j$ such that when $\hat{\alpha}_j$ is length $q_j=r-j+1$ and $\hat{\alpha}_i$ is length $q_i=r-i+1$ we have that $q_i<q_j$. Then

$$\begin{aligned}
n^{-1}\hat{\underline{\lambda}}_i'\hat{\underline{\lambda}}_j &= \hat{\alpha}_i'(\Lambda^{(i)\prime}\Lambda^{(i)}/n)^{-1/2\prime}(\Lambda^{(i)\prime}\Lambda^{(j)}/n)(\Lambda^{(j)\prime}\Lambda^{(j)}/n)^{-1/2}\hat{\alpha}_j\\
&+\hat{\alpha}_i'(\Lambda^{(i)\prime}\Lambda^{(i)}/n)^{-1/2\prime}(\Lambda^{(i)\prime}\widehat{V}_j/n)\\
&+(\widehat{V}_i'\Lambda^{(j)}/n)(\Lambda^{(j)\prime}\Lambda^{(j)}/n)^{-1/2}\hat{\alpha}_j\\
&+\widehat{V}_i'\widehat{V}_j/n
\end{aligned}$$

By Assumption 1, $\Lambda^{(i)\prime}\Lambda^{(i)}/n\to I_{q_i}$, $\Lambda^{(j)\prime}\Lambda^{(j)}/n\to I_{q_j}$, and $\Lambda^{(i)\prime}\Lambda^{(j)}/n\to\big(0_{q_i\times(i-j)},I_{q_i}\big)$. Therefore, since $\hat{\alpha}$ converges to zero except for the first term which is bounded the first term of the expression is $o_p(1)$. Considering the second and third terms we have that for any column of Λ, $|\underline{\lambda}'\widehat{V}/n|\leq(\underline{\lambda}'\underline{\lambda}/n)^{1/2}(\widehat{V}'\widehat{V}/n)^{1/2}\xrightarrow{p}0$ since $\widehat{V}'\widehat{V}/n\xrightarrow{p}0$, hence the two terms are $o_p(1)$. Finally, for the last term we have that $|\widehat{V}_i'\widehat{V}_j/n|\leq(\widehat{V}_i'\widehat{V}_i/n)^{1/2}(\widehat{V}_j'\widehat{V}_j/n)^{1/2}\xrightarrow{p}0$, thus the result follows from that fact that $|S_iS_j|=1$. □

(T2) $n^{-1}\sum_i \delta_{it}^{(l)}\delta_{is}^{(l)} \xrightarrow{p} 0, \quad t=1,\dots,T, \quad s=1,\dots,T, \quad l=2,\dots,r+1$

Proof.

$$\begin{aligned}
n^{-1}\sum_i \delta_{it}^{(l)}\delta_{is}^{(l)} &= n^{-1}\sum_i \sum_{k=1}^{l-1}\sum_{h=1}^{l-1}(\lambda_{ki}F_{kt}-\hat{\lambda}_{ki}\widehat{F}_{kt})(\lambda_{hi}F_{hs}-\hat{\lambda}_{hi}\widehat{F}_{hs}) \\
&= \sum_{k=1}^{l-1}\sum_{h=1}^{l-1}\big[n^{-1}\sum_i \lambda_{ki}\lambda_{hi}F_{kt}F_{hs} - n^{-1}\sum_i \lambda_{ki}S_h\hat{\lambda}_{hi}F_{kt}S_h\widehat{F}_{hs} \\
&\qquad + n^{-1}\sum_i S_k\hat{\lambda}_{ki}S_h\hat{\lambda}_{hi}S_k\widehat{F}_{kt}S_h\widehat{F}_{hs} - n^{-1}\sum_i S_k\hat{\lambda}_{ki}\lambda_{hi}S_k\widehat{F}_{kt}F_{hs}\big]
\end{aligned}$$

where the fact that $SS=1$ is used. Regarding the first term in the square bracket we have that $n^{-1}\sum_i \lambda_{ki}\lambda_{hi} \to 1$ for $k=h$ and 0 otherwise by Assumption 1. Considering the second term we have that $n^{-1}\sum_i \lambda_{ki}S_h\hat{\lambda}_{hi} \xrightarrow{p} 1$ for $k=h$ and 0 otherwise by (R11′) and (T11), and by (R15′) and (T12) that $S_h\widehat{F}_{ht} \xrightarrow{p} F_{ht}$. For the third term we have by (R15′) and (T12) that $S_k\widehat{F}_{kt} \xrightarrow{p} F_{kt}$ and $S_h\widehat{F}_{ht} \xrightarrow{p} F_{ht}$. Furthermore, $n^{-1}\sum_i S_k\hat{\lambda}_{ki}S_h\hat{\lambda}_{hi}=1$ for $k=h$ and converges to 0 in probability by (T1) otherwise. Similarly for the fourth term. We therefore have that the two parts of the square bracket are $o_p(1)$ and the result follows. □

(T3) $\sup_{\gamma\in\Gamma}\big|T^{-1}\sum_t\big(n^{-1}\sum_i\gamma_i\delta_{it}^{(l)}\big)^2\big| \xrightarrow{p} 0, \quad \text{for } l=2,\dots,r+1$

Proof.

$$\begin{aligned}
T^{-1}\sum_t\big(n^{-1}\sum_i\gamma_i\delta_{it}^{(l)}\big)^2 &= n^{-2}T^{-1}\sum_t\sum_i\sum_j\gamma_i\gamma_j\delta_{it}^{(l)}\delta_{jt}^{(l)} \\
&= n^{-2}\sum_i\sum_j\gamma_i\gamma_j\big(T^{-1}\sum_t\delta_{it}^{(l)}\delta_{jt}^{(l)}\big) \\
&\le \Big(n^{-2}\sum_i\sum_j\gamma_i^2\gamma_j^2\Big)^{1/2}\Big(n^{-2}\sum_i\sum_j\big(T^{-1}\sum_t\delta_{it}^{(l)}\delta_{jt}^{(l)}\big)^2\Big)^{1/2}
\end{aligned}$$

Now since $\gamma'\gamma/n=1$ we have that the first term equals 1, hence

$$\begin{aligned}
\sup_{\gamma\in\Gamma}\big|T^{-1}\sum_t\big(n^{-1}\sum_i\gamma_i\delta_{it}^{(l)}\big)^2\big| &\le \Big(n^{-2}\sum_i\sum_j\big(T^{-1}\sum_t\delta_{it}^{(l)}\delta_{jt}^{(l)}\big)^2\Big)^{1/2} \\
&= \Big(T^{-2}\sum_t\sum_s\big[n^{-1}\sum_i\delta_{it}^{(l)}\delta_{is}^{(l)}\big]\big[n^{-1}\sum_j\delta_{jt}^{(l)}\delta_{js}^{(l)}\big]\Big)^{1/2}
\end{aligned}$$

where the terms in the two square brackets converge to 0 in probability by (T2) thus completing the proof. □

(T4) $\sup_{\gamma\in\Gamma}(n^2T)^{-1}\big|\gamma'\sum_{i=l}^r\underline{\lambda}_i\sum_t F_{it}\delta_t^{(l)\prime}\gamma\big| \xrightarrow{p} 0, \quad \text{for } l=2,\dots,r$

Proof.

$$(n^2T)^{-1}\big|\gamma'\sum_{i=l}^r\underline{\lambda}_i\sum_t F_{it}\delta_t^{(l)\prime}\gamma\big| \le \big|n^{-1}\gamma'\sum_{i=l}^r\underline{\lambda}_i\big|\big|T^{-1}\sum_t F_{it}\big(n^{-1}\sum_j\delta_{jt}^{(l)\prime}\gamma_j\big)\big|$$

Hence:

$$\sup_{\gamma\in\Gamma}(n^2T)^{-1}\big|\gamma'\sum_{i=l}^r\underline{\lambda}_i\sum_t F_{it}\delta_t^{(l)\prime}\gamma\big| \le \big(\max_{l\le i\le r}\sup_{\gamma\in\Gamma}\big|\gamma'\underline{\lambda}_i/n\big|\big)$$

$$
\begin{aligned}
&\times \sum_{i=l}^{r} \sup_{\gamma\in\Gamma} \big| T^{-1} \sum_t F_{it} \big(n^{-1} \sum_j \delta_{jt}^{(l)} \gamma_j\big) \big| \\
&\le \big(\sup_{\gamma\in\Gamma} (\gamma'\gamma/n)^{1/2}\big) \big(\max_{l\le i\le r} (\underline{\lambda}_i'\underline{\lambda}_i/n)^{1/2}\big) \\
&\times \sum_{i=l}^{r} \sup_{\gamma\in\Gamma} \big| T^{-1} \sum_t F_{it} \big(n^{-1} \sum_j \delta_{jt}^{(l)} \gamma_j\big) \big|
\end{aligned}
$$

The first term is 1 by definition, the second term converges to 1 by Assumption 1, hence we need to show that the third term converges to 0 in probability to arrive at the result. Consider the following:

$$
\sup_{\gamma\in\Gamma} \big| T^{-1} \sum_t F_{it} \big(n^{-1} \sum_j \delta_{jt}^{(l)} \gamma_j\big) \big| \le \big(T^{-1} \sum_t F_{it}^2\big)^{1/2} \big(\sup_{\gamma\in\Gamma} T^{-1} \sum_t \big(n^{-1} \sum_j \delta_{jt}^{(l)} \gamma_j\big)^2\big)^{1/2}
$$

The first term is $O_p(1)$ by Assumption 1 and the second term is $o_p(1)$ by (T3), thus the desired result follows. □

(T5) $\sup_{\gamma\in\Gamma} (n^2 T)^{-1} \big|\gamma' \sum_t \delta_t^{(l)} e_t' \gamma\big| \xrightarrow{p} 0, \quad \text{for } l = 2, \ldots, r+1$

Proof.

$$
\begin{aligned}
(n^2 T)^{-1} \big|\gamma' \sum_t \delta_t^{(l)} e_t{}' \gamma\big| &= \big| T^{-1} \sum_t (n^{-1} \sum_i \gamma_i e_{it})(n^{-1} \sum_j \gamma_j \delta_{jt}^{(l)}) \big| \\
&\le \Big[T^{-1} \sum_t \big(n^{-1} \sum_i \gamma_i e_{it}\big)^2 \Big]^{1/2} \Big[T^{-1} \sum_t \big(n^{-1} \sum_j \gamma_j \delta_{jt}^{(l)}\big)^2 \Big]^{1/2}
\end{aligned}
$$

where the first term is $o_p(1)$ by (R2) and the second term is $o_p(1)$ by (T3). □

(T6) $\sup_{\gamma\in\Gamma} \big| R_l(\gamma) - S(\gamma) - R_l^*(\gamma) \big| \xrightarrow{p} 0, \quad \text{for } l = 2, \ldots, r.$

Proof.

$$
\begin{aligned}
R_l(\gamma) - S(\gamma) - R_l^*(\gamma) &= (n^2 T)^{-1} \gamma' \sum_t e_t e_t' \gamma \\
&+ (n^2 T)^{-1} \gamma' \sum_t \delta_t^{(l)} \delta_t^{(l)\prime} \gamma \\
&+ 2 (n^2 T)^{-1} \gamma' \sum_{i=l}^{r} \underline{\lambda}_i \sum_t F_{it} \delta_t^{(l)\prime} \gamma \\
&+ 2 (n^2 T)^{-1} \gamma' \sum_{i=l}^{r} \underline{\lambda}_i \sum_t F_{it} e_t' \gamma \\
&+ 2 (n^2 T)^{-1} \gamma' \sum_t \delta_t^{(l)} e_t' \gamma \\
&- (nT)^{-1} \psi_T \sum_i |\gamma_i|
\end{aligned}
$$

Hence

$$
\begin{aligned}
\sup_{\gamma\in\Gamma} \big| R_l(\gamma) - S(\gamma) - R_l^*(\gamma) \big| &\le \sup_{\gamma\in\Gamma} (n^2 T)^{-1} \big| \gamma' \sum_t e_t e_t' \gamma \big| \\
&+ \sup_{\gamma\in\Gamma} (n^2 T)^{-1} \big| \gamma' \sum_t \delta_t^{(l)} \delta_t^{(l)\prime} \gamma \big|
\end{aligned}
$$

$$
\begin{aligned}
&+\sup_{\gamma\in\Gamma} 2(n^2T)^{-1}\left|\gamma'\sum_{i=l}^{r}\underline{\lambda}_i\sum_t F_{it}\delta_t^{(l)\prime}\gamma\right|\\
&+\sup_{\gamma\in\Gamma} 2(n^2T)^{-1}\left|\gamma'\sum_{i=l}^{r}\underline{\lambda}_i\sum_t F_{it}e_t'\gamma\right|\\
&+\sup_{\gamma\in\Gamma} 2(n^2T)^{-1}\left|\gamma'\sum_t \delta_t^{(l)}e_t'\gamma\right|\\
&+\sup_{\gamma\in\Gamma}(nT)^{-1}\psi_T\sum_i\left|\gamma_i\right|
\end{aligned}
$$

The first term on the right-hand side converges to 0 by (R2), the second term converges to 0 by (T3), the third term converges to 0 by (T4), the fourth term converges to 0 by (R5) (note that (R5) is only shown for $l=1$, however, it can easily be shown to also hold for other values of l), the fifth term converges to 0 by (T5), and the sixth term converges to 0 by the argument used in (R6′). □

(T7) $\left|\sup_{\gamma\in\Gamma}[R(\gamma)-S(\gamma)]-\sup_{\gamma\in\Gamma}R^*(\gamma)\right| \xrightarrow{p} 0.$

Proof. $\left|\sup_{\gamma\in\Gamma}[R(\gamma)-S(\gamma)]-\sup_{\gamma\in\Gamma}R^*(\gamma)\right| \le \sup_{\gamma\in\Gamma}\left|R(\gamma)-S(\gamma)-R^*(\gamma)\right| \xrightarrow{p} 0$ where the first inequality follows by the definition of the sup and the convergence follows from (T6). □

(T8) $\sup_{\gamma\in\Gamma}R_l^*(\gamma)\xrightarrow{p}\sigma_{ll}$ for $l=2,\dots,r$.

Proof. Let γ be represented as $\gamma=\Lambda^{(l)}(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{-1/2}\alpha+V$ where $V'\Lambda^{(l)}=0$, hence since $\gamma\in\Gamma$ we have that $\alpha'\alpha\le 1$. We can therefore write

$$
\begin{aligned}
\sup_{\gamma\in\Gamma}R_l^*(\gamma) &= \sup_{\alpha,\alpha'\alpha\le 1}\alpha'(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{1/2\prime}(F^{(l)\prime}F^{(l)}/T)(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{1/2}\alpha\\
&= \sup_{\alpha,\alpha'\alpha\le 1}\alpha' C_{nT}^{(l)}\alpha=\hat{\sigma}_{ll}
\end{aligned}
$$

where $\hat{\sigma}_{ll}$ is the largest eigenvalue of $C_{nT}^{(l)}$. Now $(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{1/2}\to I$ and $F^{(l)\prime}F^{(l)}/T\xrightarrow{p}\Sigma_{FF}^{(l)}$ by Assumption 1, hence $C_{nT}^{(l)}\xrightarrow{p}\Sigma_{FF}^{(l)}$ and $\hat{\sigma}_{ll}\xrightarrow{p}\sigma_{ll}$. □

(T9) $\sup_{\gamma\in\Gamma}R_l(\gamma)-S(\gamma)\xrightarrow{p}\sigma_{ll}$

Proof. This follows from (T7) and (T8). □

(T10) Let $\hat{\underline{\lambda}}_l=\operatorname{argsup}_{\gamma\in\Gamma}R_l(\gamma)-S(\gamma)$; then $R_l^*(\hat{\underline{\lambda}}_l)\xrightarrow{p}\sigma_{ll}$.

Proof. This follows from (T6) and (T9). □

(T11) $(S_l\hat{\underline{\lambda}}_l'\Lambda/n)\xrightarrow{p}\ell_l'$ for $l=2,\dots,r$ where $\ell_{l,i}=1$ for $i=l$ and 0 otherwise.

Proof. Let $\hat{\underline{\lambda}}_l=\Lambda^{(l)}(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{-1/2}\hat{\alpha}+\widehat{V}$ and recall that $R_l^*(\hat{\underline{\lambda}}_l)=\hat{\alpha}C_{nT}^{(l)\prime}\hat{\alpha}$ where $C_{nT}^{(l)}$ is defined in (T8). Then

$$
\begin{aligned}
R_l^*(\hat{\underline{\lambda}}_l)-\sigma_{ll} &= \hat{\alpha}'(C_{nT}^{(l)}-\Sigma_{FF}^{(l)})\hat{\alpha}+\hat{\alpha}'\Sigma_{FF}^{(l)}\hat{\alpha}-\sigma_{ll}\\
&= \hat{\alpha}'(C_{nT}^{(l)}-\Sigma_{FF}^{(l)})\hat{\alpha}+(\hat{\alpha}_1^2-1)\sigma_{ll}+\sum_{l<i\le r}\hat{\alpha}_{i-l+1}^2\sigma_{ii}
\end{aligned}
$$

We have by (T10) that the left-hand-side is $o_p(1)$, and the first term on the right-hand-side is $o_p(1)$ since $C_{nT}^{(l)} \xrightarrow{p} \Sigma_{FF}^{(l)}$ and $\hat{\alpha}$ is bounded. Hence $(\hat{\alpha}_1^2 - 1)\sigma_{ll} + \sum_{l<i\leq r} \hat{\alpha}_{i-l+1}^2 \sigma_{ii}$ $\xrightarrow{p} 0$. By Assumption 1 we have that $\sigma_{ii} > 0$ for $i = 1, \ldots, r$, hence we must have that $\hat{\alpha}_1^2 \xrightarrow{p} 1$ and $\hat{\alpha}_i^2 \xrightarrow{p} 0$ for $i > 1$.

We have that $\underline{\hat{\lambda}}_l' \underline{\hat{\lambda}}/n = 1$ but from the definition above we also have that $\underline{\hat{\lambda}}_l' \underline{\hat{\lambda}}/n = \hat{\alpha}'(\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{-1/2\prime}\hat{\alpha} + \widehat{V}'\widehat{V}/n$. By Assumption 1, $\Lambda^{(l)\prime}\Lambda^{(l)}/n \to I$ and as shown above $\hat{\alpha} \xrightarrow{p} (1, 0, \ldots, 0)$ therefore we must have that $\widehat{V}'\widehat{V}/n$ is $o_p(1)$.

Consider the expression

$$
\begin{aligned}
S_l \underline{\hat{\lambda}}_l' \Lambda/n &= [S_l \Lambda^{(l)} (\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{-1/2}\hat{\alpha} + S_l \widehat{V}]' \Lambda/n \\
&= S_l \hat{\alpha}' (\Lambda^{(l)\prime}\Lambda^{(l)}/n)^{-1/2\prime} \Lambda^{(l)\prime}\Lambda/n + S_l \widehat{V}'\Lambda/n \qquad \text{(A.1)}
\end{aligned}
$$

Now, $\Lambda^{(l)\prime}\Lambda/n \to \left(0_{(r-l+1)\times(l-1)}, I_{r-l+1}\right)$ by Assumption 1 therefore the first term converges to ℓ_l'. For the second term we have for each column of Λ that

$$
\left|\widehat{V}'\underline{\lambda}_j/n\right| \leq (\widehat{V}'\widehat{V}/n)^{1/2} (\underline{\lambda}_j'\underline{\lambda}_j/n)^{1/2} \xrightarrow{p} 0
$$

since the first term is $o_p(1)$ and the second is bounded. Therefore the second term of (A.1) converges to a vector of zeros and the result follows. □

(T12) $S_l \widehat{F}_{lt} - F_{lt} \xrightarrow{p} 0$ for $l = 2, \ldots, r$.

Proof.

$$
\begin{aligned}
S_l \widehat{F}_{lt} - F_{lt} &= S_l \underline{\hat{\lambda}}_l' \hat{e}_t^{(l)}/n - F_{lt} \\
&= S_l \underline{\hat{\lambda}}_l' \Big(e_t + \delta_t^{(l)} + \sum_{i=l}^{r} \underline{\lambda}_i F_{it}\Big)/n - F_{lt} \\
&= S_l \underline{\hat{\lambda}}_l' e_t/n + S_l \underline{\hat{\lambda}}_l' \delta_t^{(l)}/n + \left(S_l \underline{\hat{\lambda}}_l' \lambda_l/n - 1\right) F_{lt} + S_l \underline{\hat{\lambda}}_l' \Big(\sum_{i=l+1}^{r} \underline{\lambda}_i F_{it}\Big)/n
\end{aligned}
$$

The first term is $o_p(1)$, this is shown in the proof of Stock and Watson (2002a, (R15)). Since $|F_T|$ is $O_p(1)$ by Assumption 1 we have that the third and fourth terms are $o_p(1)$ by (T11). Considering the second term we have that

$$
\left|S_l \underline{\hat{\lambda}}_l' \delta_t^{(l)}/n\right| \leq \left(\underline{\hat{\lambda}}_l' \underline{\hat{\lambda}}_l/n\right)^{1/2} \left(\delta_t^{(l)\prime}\delta_t^{(l)}/n\right)^{1/2}
$$

where the first term equals 1 and the second term is $o_p(1)$ by (T2). Hence the result follows. □

(T13) $T^{-1}\sum_t \widehat{F}_{lt}^2 \xrightarrow{p} 0$, for $l > r$

Proof.

$$
T^{-1}\sum_t \widehat{F}_{lt}^2 = n^{-2} T^{-1} \underline{\hat{\lambda}}_l' \sum_t \hat{e}_t^{(l-1)} \hat{e}_t^{(l-1)\prime} \underline{\hat{\lambda}}_l = R_l(\underline{\hat{\lambda}}_l)
$$

where $\underline{\hat{\lambda}}_l = \operatorname{argsup}_{\gamma\in\Gamma} R_l(\gamma) - S(\gamma)$. Now

$$
\sup_{\gamma\in\Gamma} \left|R_l(\gamma)\right| \leq \sup_{\gamma\in\Gamma} n^{-2} T^{-1} \left|\gamma' \sum_t e_t e_t' \gamma\right|
$$

$$
\begin{aligned}
&+\sup_{\gamma\in\Gamma} n^{-2}T^{-1}\big|\gamma'\sum_t \delta_t^{(r+1)}\delta_t^{(r+1)\prime}\gamma\big| \\
&+\sup_{\gamma\in\Gamma} n^{-2}T^{-1}\big|\sum_{r<k<l}\gamma\underline{\hat{\lambda}}_k\sum_t \widehat{F}_{kt}\widehat{F}_{kt}\underline{\hat{\lambda}}_k'\gamma\big| \qquad\qquad \text{(A.2)} \\
&+\sup_{\gamma\in\Gamma} 2n^{-2}T^{-1}\big|\gamma'\sum_t e_t\delta_t^{(r+1)\prime}\gamma\big| \\
&+\sup_{\gamma\in\Gamma} 2n^{-2}T^{-1}\big|\gamma'\sum_t e_t\sum_{r<k<l}\widehat{F}_{kt}\underline{\hat{\lambda}}_k'\gamma\big| \\
&+\sup_{\gamma\in\Gamma} 2n^{-2}T^{-1}\big|\gamma'\sum_t \delta_t^{(r+1)}\sum_{r<k<l}\widehat{F}_{kt}\underline{\hat{\lambda}}_k'\gamma\big|
\end{aligned}
$$

Now for $l = r+1$ we have that the third, fifth and sixth terms disappear. Further, by (R2) the first term is $o_p(1)$, by (T3) the second term is $o_p(1)$, and by (T5) the fourth term is $o_p(1)$.

In the case $l > r+1$ we must check the third, fifth and sixth terms. Consider the third term:

$$
\begin{aligned}
\sup_{\gamma\in\Gamma}(n^2T)^{-1}\big|\sum_{r<k<l}\gamma\underline{\hat{\lambda}}_k\sum_t \widehat{F}_{kt}\widehat{F}_{kt}\underline{\hat{\lambda}}_k'\gamma\big| &= \sup_{\gamma\in\Gamma}\big|(n^2T)^{-1}\sum_{r<k<l}\sum_t\sum_i\sum_j \gamma_i\gamma_j\underline{\hat{\lambda}}_{ki}\underline{\hat{\lambda}}_{kj}\widehat{F}_{kt}^2\big| \\
&\le \max_{r<k<l}\sup_{\gamma\in\Gamma}\big|n^{-2}\sum_i\sum_j\gamma_i\gamma_j\underline{\hat{\lambda}}_{ki}\underline{\hat{\lambda}}_{kj}\big|\sum_{r<h<l}T^{-1}\sum_t\widehat{F}_{ht}^2
\end{aligned}
$$

For the first term we have that

$$
\big|n^{-2}\sum_i\sum_j\gamma_i\gamma_j\underline{\hat{\lambda}}_{ki}\underline{\hat{\lambda}}_{kj}\big| \le \big(n^{-2}\sum_i\sum_j\gamma_i^2\gamma_j^2\big)^{1/2}\big(n^{-2}\sum_i\sum_j\underline{\hat{\lambda}}_{ki}^2\underline{\hat{\lambda}}_{kj}^2\big)^{1/2} = 1
$$

and the second term is $o_p(1)$ as argued above.

Consider the fifth term:

$$
\begin{aligned}
n^{-2}T^{-1}\big|\gamma'\sum_t e_t\sum_{r<k<l}\widehat{F}_{kt}\underline{\hat{\lambda}}_k'\gamma\big| &= \big|\sum_{r<k<l}T^{-1}\sum_t\big(n^{-1}\sum_i\gamma_t e_{it}\big)\big(n^{-1}\sum_j\gamma_j\underline{\hat{\lambda}}_{kj}\widehat{F}_{kt}\big)\big| \\
&\le \big[(l-r-1)T^{-1}\sum_t\big(n^{-1}\sum_i\gamma_t e_{it}\big)^2\big]^{1/2} \\
&\quad\times\big[\sum_{r<k<l}T^{-1}\sum_t\big(n^{-1}\sum_j\gamma_j\underline{\hat{\lambda}}_{kj}\widehat{F}_{kt}\big)^2\big]^{1/2}
\end{aligned}
$$

The first term is $o_p(1)$ by (R2), and the second term is the square root of (A.2) and hence $o_p(1)$ as argued above. Lastly, the same argument holds for the sixth term if we rely on (T3) instead of (R2). $\square$

(R17) $T^{-1}\sum_t S\widehat{F}_tF_t' \xrightarrow{p} \Sigma_{FF}$ where $\widehat{F}_t$ is $r\times 1$.

(R18) $T^{-1}\sum_t \widehat{F}_t\widehat{F}_t' \xrightarrow{p} \Sigma_{FF}$ where $\widehat{F}_t$ is $r\times 1$.

(R20) $T^{-1}\sum_t S\widehat{F}_tw_t' \xrightarrow{p} \Sigma_{Fw}$ where $\widehat{F}_t$ is $r\times 1$.

(R21) $T^{-1}\sum_t S\widehat{F}_t\varepsilon_{t+h} \xrightarrow{p} 0$ where $\widehat{F}_t$ is $r\times 1$.

(T14) Let $\widehat{F}_t$ be the $k\ge r$ estimated factors then:

$$
T^{-1}\sum_t\widehat{F}_t\widehat{F}_t' \xrightarrow{p} \begin{bmatrix}\Sigma_{FF} & 0_{r\times k-r}\\ 0_{k-r\times r} & 0_{k-r\times k-r}\end{bmatrix}
$$

Proof. The first diagonal element follows from (R18). For the remaining elements consider $T^{-1}\sum_t \widehat{F}_{kt}^2$. For $k \le r$ this will be $O_p(1)$ by (R18) and Assumption 1, and for $k > r$ it will be $o_p(1)$ by (T13). Since

$$T^{-1}\Big|\sum_t \widehat{F}_{kt}\widehat{F}_{lt}\Big| \le \Big[T^{-1}\sum_t \widehat{F}_{kt}^2\Big]^{1/2}\Big[T^{-1}\sum_t \widehat{F}_{lt}^2\Big]^{1/2}$$

the results then follows. □

(T15) Let $\widehat{F}_t$ be the $k \ge r$ estimated factors then:

$$T^{-1}\sum_t S\widehat{F}_t F_t' \xrightarrow{p} \begin{bmatrix} \Sigma_{FF} \\ 0_{k-r\times r} \end{bmatrix}$$

Proof. The proof is analogous to that of (T14) only using (R17) instead of (R18). □

(T16) Let $\widehat{F}_t$ be the $k \ge r$ estimated factors then:

$$T^{-1}\sum_t S\widehat{F}_t w_t' \xrightarrow{p} \begin{bmatrix} \Sigma_{Fw} \\ 0_{k-r\times r} \end{bmatrix}$$

Proof. For the first element on the right-hand side the result follows by (R20), for the second element we have that:

$$T^{-1}\Big|\sum_t \widehat{F}_{lt} w_{it}\Big| \le \Big[T^{-1}\sum_t \widehat{F}_{lt}^2\Big]^{1/2}\Big[T^{-1}\sum_t w_{it}^2\Big]^{1/2}$$

Since the first term is $o_p(1)$ by (T13) and the second is $O_p(1)$ by Assumption 4 the result follows. □

(T17) Let $\widehat{F}_t$ be the $k \ge r$ estimated factors then:

$$T^{-1}\sum_t S\widehat{F}_t \varepsilon_{t+h} \xrightarrow{p} 0$$

Proof. For the first r factors the result follows by (R21) and for the remaining factors we have that

$$T^{-1}\Big|\sum_t \widehat{F}_{lt}\varepsilon_{t+h}\Big| \le \Big[T^{-1}\sum_t \widehat{F}_{lt}^2\Big]^{1/2}\Big[T^{-1}\sum_t \varepsilon_{t+h}^2\Big]^{1/2}$$

Since the first term is $o_p(1)$ by (T13) and the second is $O_p(1)$ by Assumption 4 the result follows. □

(T18) $S_i\hat{\beta}_{Fi}^{\text{RR}} \xrightarrow{p} (\sigma_{ii}+\kappa)^{-1}\sigma_{ii}\beta_{Fi}$ for $i = 1,\ldots,r$, and $\hat{\beta}_{Fi}^{\text{RR}} \xrightarrow{p} 0$ for $i > r$.

Proof. In matrix from the forecasting model is $Y = F\beta_F + W\beta_w + \varepsilon$, hence the ridge regression estimate of β_F is:

$$S\hat{\beta}_F^{\text{RR}} = \left(T^{-1}\widehat{F}'\widehat{F} + T^{-1}\kappa_T I\right)^{-1} T^{-1}S\widehat{F}'(F\beta_F + W\beta_w + \varepsilon)$$

Now for the first term we have by (T14) that

$$\left(T^{-1}\widehat{F}'\widehat{F} + T^{-1}\kappa_T I\right)^{-1} \xrightarrow{p} \begin{bmatrix} \left(\Sigma_{FF}+\kappa I\right)^{-1} & 0 \\ 0 & \kappa^{-1}I \end{bmatrix}$$

which is a diagonal matrix. Combining this with (T15), (T16), (T17) and Assumption 5 the desired result follows. □

(T19) $P(\hat{\beta}_{Fi} = 0) \to 1$ for $i > r$.

Proof.

$$P(\hat{\beta}_{Fi} = 0) = P(|\hat{\beta}_{Fi}^{\mathrm{RR}}| < \beta_{\mathrm{thr}})$$

Since we choose $\beta_{\mathrm{thr}} < \min_{\{i:\beta_{Fi}\neq 0\}}((\sigma_{ii} + \kappa)^{-1}\sigma_{ii}|\beta_{Fi}|$ and for $i \le r$ we have by (T18) that $S_i\hat{\beta}_{Fi} \xrightarrow{p} (\sigma_{ii} + \kappa)^{-1}\sigma_{ii}\beta_{Fi}$ the result follows. □

Proof of Theorem 1. For $i = 1$ the result follows by (R15′), and for $i = 2, \ldots, r$ it follows by (T12). □

Proof of Theorem 3. For part b the result follows by (T19). Parts a and c then follow by Theorem 2 since by part b only the true factors are included with probability tending to one. □

Appendix: Additional results

Table A.1. Simulation results for three different scenarios where $\tau = 0.8$. Sparsity of the loadings for the SPC estimator.

Data-generating process							Fraction of zero-entries							
n	T	r	a	b	τ	ψ_T	$\underline{\hat{\lambda}}_1$	$\underline{\hat{\lambda}}_2$	$\underline{\hat{\lambda}}_3$	$\underline{\hat{\lambda}}_4$	$\underline{\hat{\lambda}}_5$	$\underline{\hat{\lambda}}_6$	$\underline{\hat{\lambda}}_7$	$\underline{\hat{\lambda}}_8$
25	50	1	0	0	0.8	1.75	0.79	0.86	0.88	0.89	0.89	0.90	0.91	0.92
25	100	1	0	0	0.8	1.66	0.76	0.83	0.85	0.87	0.88	0.89	0.89	0.90
50	100	1	0	0	0.8	1.92	0.79	0.90	0.91	0.92	0.92	0.93	0.93	0.93
50	200	1	0	0	0.8	1.81	0.77	0.87	0.89	0.90	0.90	0.91	0.91	0.92
100	200	1	0	0	0.8	2.06	0.79	0.92	0.93	0.94	0.94	0.94	0.94	0.95
150	200	1	0	0	0.8	2.20	0.81	0.95	0.95	0.95	0.95	0.96	0.96	0.96
25	50	4	0	0	0.8	2.29	0.77	0.80	0.83	0.86	0.90	0.91	0.92	0.93
25	100	4	0	0	0.8	2.09	0.73	0.75	0.77	0.81	0.88	0.89	0.90	0.91
50	100	4	0	0	0.8	2.35	0.75	0.77	0.78	0.80	0.94	0.94	0.95	0.95
50	200	4	0	0	0.8	2.18	0.70	0.71	0.72	0.73	0.92	0.93	0.93	0.93
100	200	4	0	0	0.8	2.52	0.73	0.75	0.75	0.76	0.96	0.96	0.96	0.96
150	200	4	0	0	0.8	2.72	0.75	0.77	0.77	0.78	0.97	0.97	0.98	0.98
25	50	4	0.5	1	0.8	4.89	0.62	0.69	0.74	0.77	0.79	0.81	0.83	0.84
25	100	4	0.5	1	0.8	4.91	0.59	0.65	0.69	0.71	0.73	0.74	0.75	0.76
50	100	4	0.5	1	0.8	6.02	0.73	0.77	0.79	0.82	0.83	0.84	0.86	0.87
50	200	4	0.5	1	0.8	5.91	0.69	0.72	0.75	0.77	0.78	0.80	0.81	0.82
100	200	4	0.5	1	0.8	6.73	0.78	0.79	0.81	0.83	0.84	0.86	0.87	0.88
150	200	4	0.5	1	0.8	7.36	0.83	0.84	0.85	0.86	0.88	0.89	0.90	0.90

Note: The results are based on 1,000 Monte Carlo replications.

Table A.2. Simulation results for three different scenarios where $\tau = 0$. Sparsity of the loadings for the SPC estimator.

Data-generating process							Fraction of zero-entries							
n	T	r	a	b	τ	ψ_T	$\underline{\hat{\lambda}}_1$	$\underline{\hat{\lambda}}_2$	$\underline{\hat{\lambda}}_3$	$\underline{\hat{\lambda}}_4$	$\underline{\hat{\lambda}}_5$	$\underline{\hat{\lambda}}_6$	$\underline{\hat{\lambda}}_7$	$\underline{\hat{\lambda}}_8$
25	50	1	0	0	0	1.61	0.20	0.82	0.85	0.87	0.87	0.88	0.89	0.91
25	100	1	0	0	0	1.53	0.13	0.79	0.82	0.84	0.85	0.86	0.87	0.87
50	100	1	0	0	0	1.80	0.15	0.87	0.89	0.90	0.90	0.91	0.91	0.92
50	200	1	0	0	0	1.68	0.10	0.84	0.86	0.87	0.88	0.88	0.89	0.90
100	200	1	0	0	0	1.92	0.11	0.90	0.92	0.92	0.92	0.93	0.93	0.93
150	200	1	0	0	0	2.04	0.12	0.93	0.93	0.94	0.94	0.94	0.94	0.95
25	50	4	0	0	0	1.42	0.15	0.18	0.22	0.26	0.77	0.79	0.82	0.84
25	100	4	0	0	0	1.32	0.10	0.12	0.15	0.17	0.72	0.75	0.77	0.78
50	100	4	0	0	0	1.57	0.12	0.14	0.16	0.17	0.82	0.84	0.85	0.87
50	200	4	0	0	0	1.46	0.08	0.10	0.10	0.11	0.78	0.80	0.82	0.83
100	200	4	0	0	0	1.65	0.09	0.10	0.11	0.11	0.85	0.86	0.87	0.88
150	200	4	0	0	0	1.74	0.09	0.10	0.11	0.12	0.88	0.89	0.89	0.90
25	50	4	0.5	1	0	4.24	0.41	0.49	0.56	0.61	0.66	0.69	0.72	0.76
25	100	4	0.5	1	0	3.56	0.27	0.34	0.40	0.45	0.49	0.52	0.55	0.57
50	100	4	0.5	1	0	5.30	0.38	0.45	0.51	0.58	0.68	0.72	0.75	0.78
50	200	4	0.5	1	0	4.62	0.26	0.30	0.33	0.39	0.55	0.60	0.63	0.65
100	200	4	0.5	1	0	5.76	0.31	0.34	0.37	0.40	0.73	0.76	0.78	0.80
150	200	4	0.5	1	0	6.27	0.33	0.36	0.38	0.41	0.79	0.81	0.83	0.84

Note: The results are based on 1,000 Monte Carlo replications.

Table A.3. Simulation results for three different scenarios where $\tau = 0.8$. Estimated number of factors and precision of the estimates.

	Data-generating process						Estimated number of factors					Factor R^2 for various choices of k					
Est.	n	T	r	a	b	τ	IC_1	IC_2	IC_3	BIC	RR	IC_1	IC_2	IC_3	BIC	RR	$k = r$
PC	25	50	1	0	0	0.8	1.00	1.00	1.32	1.20	1.63	0.72	0.72	0.72	0.73	0.74	0.72
PC	25	100	1	0	0	0.8	1.00	1.00	1.00	1.12	1.41	0.75	0.75	0.75	0.75	0.76	0.75
PC	50	100	1	0	0	0.8	1.00	1.00	1.00	1.07	1.37	0.87	0.87	0.87	0.88	0.88	0.87
PC	50	200	1	0	0	0.8	1.00	1.00	1.00	1.05	1.20	0.88	0.88	0.88	0.88	0.88	0.88
PC	100	200	1	0	0	0.8	1.00	1.00	1.00	1.05	1.19	0.94	0.94	0.94	0.94	0.94	0.94
PC	150	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.21	0.96	0.96	0.96	0.96	0.96	0.96
SPC	25	50	1	0	0	0.8	1.00	1.00	1.00	1.17	1.47	0.74	0.74	0.74	0.75	0.75	0.74
SPC	25	100	1	0	0	0.8	1.00	1.00	1.00	1.09	1.27	0.76	0.76	0.76	0.77	0.77	0.76
SPC	50	100	1	0	0	0.8	1.00	1.00	1.00	1.05	1.23	0.88	0.88	0.88	0.88	0.88	0.88
SPC	50	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.18	0.89	0.89	0.89	0.89	0.89	0.89
SPC	100	200	1	0	0	0.8	1.00	1.00	1.00	1.04	1.18	0.94	0.94	0.94	0.94	0.94	0.94
SPC	150	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.14	0.96	0.96	0.96	0.96	0.96	0.96
Post-SPC	25	50	1	0	0	0.8	1.00	1.00	1.00	1.17	1.47	0.74	0.74	0.74	0.74	0.75	0.74
Post-SPC	25	100	1	0	0	0.8	1.00	1.00	1.00	1.08	1.30	0.76	0.76	0.76	0.76	0.76	0.76
Post-SPC	50	100	1	0	0	0.8	1.00	1.00	1.00	1.06	1.24	0.88	0.88	0.88	0.88	0.88	0.88
Post-SPC	50	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.18	0.89	0.89	0.89	0.89	0.89	0.89
Post-SPC	100	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.18	0.94	0.94	0.94	0.94	0.94	0.94
Post-SPC	150	200	1	0	0	0.8	1.00	1.00	1.00	1.03	1.19	0.96	0.96	0.96	0.96	0.96	0.96
PC	25	50	4	0	0	0.8	2.77	2.30	5.82	3.74	3.46	0.60	0.52	0.73	0.67	0.58	0.72
PC	25	100	4	0	0	0.8	2.92	2.66	3.35	3.90	3.63	0.61	0.57	0.68	0.71	0.63	0.74
PC	50	100	4	0	0	0.8	3.73	3.50	3.96	3.92	3.68	0.83	0.79	0.87	0.84	0.76	0.87
PC	50	200	4	0	0	0.8	3.85	3.77	3.94	3.94	3.75	0.85	0.84	0.87	0.86	0.80	0.88
PC	100	200	4	0	0	0.8	4.00	4.00	4.00	3.95	3.78	0.94	0.94	0.94	0.92	0.87	0.94
PC	150	200	4	0	0	0.8	4.00	4.00	4.00	3.95	3.79	0.96	0.96	0.96	0.94	0.88	0.96
SPC	25	50	4	0	0	0.8	2.12	1.73	2.84	3.87	3.78	0.48	0.41	0.61	0.69	0.65	0.72
SPC	25	100	4	0	0	0.8	2.58	2.37	3.00	3.96	3.86	0.56	0.52	0.63	0.72	0.68	0.74
SPC	50	100	4	0	0	0.8	3.40	3.14	3.85	4.01	3.96	0.77	0.72	0.85	0.86	0.82	0.87
SPC	50	200	4	0	0	0.8	3.74	3.65	3.88	3.97	3.89	0.84	0.82	0.86	0.87	0.83	0.88
SPC	100	200	4	0	0	0.8	3.99	3.98	4.00	4.01	3.97	0.94	0.94	0.94	0.93	0.91	0.94
SPC	150	200	4	0	0	0.8	4.00	4.00	4.00	4.00	3.98	0.96	0.96	0.96	0.95	0.93	0.96
Post-SPC	25	50	4	0	0	0.8	2.21	1.81	2.91	3.84	3.73	0.50	0.43	0.62	0.69	0.63	0.72
Post-SPC	25	100	4	0	0	0.8	2.63	2.42	3.06	3.93	3.79	0.57	0.53	0.64	0.72	0.67	0.74
Post-SPC	50	100	4	0	0	0.8	3.48	3.19	3.87	3.97	3.85	0.79	0.73	0.85	0.86	0.81	0.87
Post-SPC	50	200	4	0	0	0.8	3.76	3.69	3.90	3.95	3.82	0.84	0.83	0.86	0.86	0.82	0.88
Post-SPC	100	200	4	0	0	0.8	4.00	3.99	4.00	4.00	3.93	0.94	0.94	0.94	0.93	0.90	0.94
Post-SPC	150	200	4	0	0	0.8	4.00	4.00	4.00	3.99	3.95	0.96	0.96	0.96	0.95	0.92	0.96
PC	25	50	4	0.5	1	0.8	8.00	7.53	8.00	2.18	2.53	0.34	0.33	0.34	0.12	0.15	0.18
PC	25	100	4	0.5	1	0.8	8.00	7.99	8.00	2.36	2.74	0.25	0.25	0.25	0.10	0.11	0.14
PC	50	100	4	0.5	1	0.8	3.48	1.10	8.00	3.00	2.97	0.15	0.06	0.30	0.15	0.15	0.17
PC	50	200	4	0.5	1	0.8	1.30	1.03	8.00	3.90	3.47	0.07	0.06	0.26	0.17	0.15	0.17
PC	100	200	4	0.5	1	0.8	1.04	1.00	8.00	5.28	4.25	0.11	0.11	0.40	0.32	0.26	0.27
PC	150	200	4	0.5	1	0.8	1.07	1.00	8.00	5.75	4.61	0.16	0.15	0.50	0.44	0.36	0.37
SPC	25	50	4	0.5	1	0.8	2.36	1.25	7.64	1.89	2.39	0.11	0.07	0.32	0.11	0.15	0.18
SPC	25	100	4	0.5	1	0.8	5.13	2.52	7.97	2.07	2.60	0.16	0.09	0.24	0.10	0.12	0.14
SPC	50	100	4	0.5	1	0.8	1.01	1.00	2.78	2.37	2.68	0.07	0.07	0.13	0.14	0.16	0.18
SPC	50	200	4	0.5	1	0.8	1.00	1.00	1.40	3.03	3.08	0.07	0.07	0.09	0.16	0.16	0.17
SPC	100	200	4	0.5	1	0.8	1.00	1.00	1.24	4.11	3.70	0.15	0.15	0.17	0.34	0.33	0.32
SPC	150	200	4	0.5	1	0.8	1.00	1.00	1.63	4.59	4.15	0.19	0.19	0.30	0.53	0.52	0.48
Post-SPC	25	50	4	0.5	1	0.8	3.01	1.50	7.50	1.92	2.39	0.14	0.08	0.31	0.11	0.14	0.18
Post-SPC	25	100	4	0.5	1	0.8	5.45	3.00	7.89	2.07	2.56	0.17	0.10	0.24	0.09	0.11	0.14
Post-SPC	50	100	4	0.5	1	0.8	1.04	1.00	4.18	2.40	2.73	0.07	0.07	0.18	0.14	0.15	0.18
Post-SPC	50	200	4	0.5	1	0.8	1.01	1.00	2.10	3.22	3.19	0.07	0.07	0.11	0.16	0.16	0.17
Post-SPC	100	200	4	0.5	1	0.8	1.00	1.00	1.77	4.40	3.80	0.14	0.14	0.21	0.34	0.32	0.32
Post-SPC	150	200	4	0.5	1	0.8	1.00	1.00	2.50	4.74	4.24	0.19	0.19	0.38	0.52	0.51	0.47

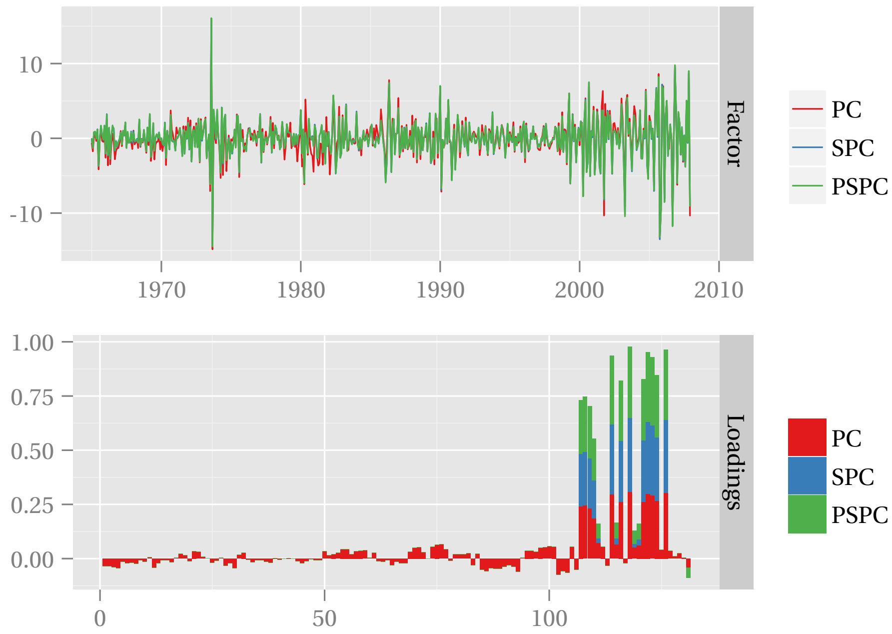
Note: The results are based on 1,000 Monte Carlo replications.

Table A.4. Simulation results for three different scenarios where $\tau = 0$. Estimated number of factors and precision of the estimates.

	Data-generating process						Estimated number of factors					Factor R^2 for various choices of k					
Est.	n	T	r	a	b	τ	IC_1	IC_2	IC_3	BIC	RR	IC_1	IC_2	IC_3	BIC	RR	$k=r$
PC	25	50	1	0	0	0	1.00	1.00	1.41	1.13	1.46	0.94	0.94	0.94	0.94	0.94	0.94
PC	25	100	1	0	0	0	1.00	1.00	1.00	1.06	1.32	0.95	0.95	0.95	0.95	0.95	0.95
PC	50	100	1	0	0	0	1.00	1.00	1.00	1.05	1.25	0.97	0.97	0.97	0.97	0.97	0.97
PC	50	200	1	0	0	0	1.00	1.00	1.00	1.04	1.16	0.97	0.97	0.97	0.97	0.97	0.97
PC	100	200	1	0	0	0	1.00	1.00	1.00	1.03	1.17	0.98	0.98	0.98	0.98	0.98	0.98
PC	150	200	1	0	0	0	1.00	1.00	1.00	1.02	1.14	0.99	0.99	0.99	0.99	0.99	0.99
SPC	25	50	1	0	0	0	1.00	1.00	1.00	1.12	1.40	0.94	0.94	0.94	0.94	0.94	0.94
SPC	25	100	1	0	0	0	1.00	1.00	1.00	1.05	1.25	0.95	0.95	0.95	0.95	0.95	0.95
SPC	50	100	1	0	0	0	1.00	1.00	1.00	1.06	1.25	0.97	0.97	0.97	0.97	0.97	0.97
SPC	50	200	1	0	0	0	1.00	1.00	1.00	1.04	1.14	0.97	0.97	0.97	0.97	0.97	0.97
SPC	100	200	1	0	0	0	1.00	1.00	1.00	1.04	1.15	0.98	0.98	0.98	0.98	0.98	0.98
SPC	150	200	1	0	0	0	1.00	1.00	1.00	1.03	1.16	0.99	0.99	0.99	0.99	0.99	0.99
Post-SPC	25	50	1	0	0	0	1.00	1.00	1.00	1.12	1.43	0.94	0.94	0.94	0.94	0.94	0.94
Post-SPC	25	100	1	0	0	0	1.00	1.00	1.00	1.05	1.28	0.95	0.95	0.95	0.95	0.95	0.95
Post-SPC	50	100	1	0	0	0	1.00	1.00	1.00	1.06	1.25	0.97	0.97	0.97	0.97	0.97	0.97
Post-SPC	50	200	1	0	0	0	1.00	1.00	1.00	1.03	1.15	0.97	0.97	0.97	0.97	0.97	0.97
Post-SPC	100	200	1	0	0	0	1.00	1.00	1.00	1.04	1.16	0.98	0.98	0.98	0.98	0.98	0.98
Post-SPC	150	200	1	0	0	0	1.00	1.00	1.00	1.03	1.15	0.99	0.99	0.99	0.99	0.99	0.99
PC	25	50	4	0	0	0	4.00	4.00	6.37	3.85	3.54	0.94	0.94	0.94	0.89	0.78	0.94
PC	25	100	4	0	0	0	4.00	4.00	4.00	3.87	3.54	0.94	0.94	0.94	0.90	0.81	0.94
PC	50	100	4	0	0	0	4.00	4.00	4.00	3.87	3.61	0.97	0.97	0.97	0.93	0.84	0.97
PC	50	200	4	0	0	0	4.00	4.00	4.00	3.90	3.67	0.97	0.97	0.97	0.95	0.87	0.97
PC	100	200	4	0	0	0	4.00	4.00	4.00	3.90	3.64	0.98	0.98	0.98	0.96	0.88	0.98
PC	150	200	4	0	0	0	4.00	4.00	4.00	3.91	3.66	0.99	0.99	0.99	0.96	0.89	0.99
SPC	25	50	4	0	0	0	4.00	4.00	4.01	3.86	3.53	0.93	0.93	0.93	0.88	0.77	0.93
SPC	25	100	4	0	0	0	4.00	4.00	4.00	3.87	3.59	0.94	0.94	0.94	0.91	0.81	0.94
SPC	50	100	4	0	0	0	4.00	4.00	4.00	3.89	3.62	0.97	0.97	0.97	0.93	0.85	0.97
SPC	50	200	4	0	0	0	4.00	4.00	4.00	3.91	3.70	0.97	0.97	0.97	0.95	0.87	0.97
SPC	100	200	4	0	0	0	4.00	4.00	4.00	3.90	3.63	0.98	0.98	0.98	0.96	0.88	0.98
SPC	150	200	4	0	0	0	4.00	4.00	4.00	3.89	3.66	0.99	0.99	0.99	0.96	0.89	0.99
Post-SPC	25	50	4	0	0	0	4.00	4.00	4.06	3.87	3.56	0.94	0.94	0.94	0.89	0.78	0.94
Post-SPC	25	100	4	0	0	0	4.00	4.00	4.00	3.88	3.57	0.94	0.94	0.94	0.91	0.81	0.94
Post-SPC	50	100	4	0	0	0	4.00	4.00	4.00	3.88	3.62	0.97	0.97	0.97	0.93	0.85	0.97
Post-SPC	50	200	4	0	0	0	4.00	4.00	4.00	3.91	3.68	0.97	0.97	0.97	0.95	0.88	0.97
Post-SPC	100	200	4	0	0	0	4.00	4.00	4.00	3.89	3.65	0.98	0.98	0.98	0.96	0.88	0.98
Post-SPC	150	200	4	0	0	0	4.00	4.00	4.00	3.89	3.63	0.99	0.99	0.99	0.96	0.88	0.99
PC	25	50	4	0.5	1	0	8.00	7.71	8.00	5.02	4.12	0.73	0.71	0.73	0.59	0.47	0.54
PC	25	100	4	0.5	1	0	8.00	7.97	8.00	5.44	4.38	0.68	0.68	0.68	0.59	0.47	0.53
PC	50	100	4	0.5	1	0	7.14	4.26	8.00	5.21	4.59	0.80	0.72	0.81	0.74	0.64	0.73
PC	50	200	4	0.5	1	0	5.61	4.24	8.00	5.15	4.71	0.79	0.77	0.81	0.77	0.68	0.77
PC	100	200	4	0.5	1	0	4.13	4.00	8.00	4.31	4.19	0.89	0.89	0.91	0.87	0.79	0.89
PC	150	200	4	0.5	1	0	4.03	4.00	8.00	4.15	4.04	0.93	0.93	0.94	0.90	0.83	0.93
SPC	25	50	4	0.5	1	0	4.15	2.26	7.52	4.69	3.87	0.50	0.35	0.69	0.55	0.45	0.52
SPC	25	100	4	0.5	1	0	7.04	5.70	7.97	5.30	4.38	0.63	0.57	0.67	0.58	0.47	0.53
SPC	50	100	4	0.5	1	0	3.27	2.59	6.21	4.97	4.36	0.63	0.54	0.75	0.71	0.61	0.70
SPC	50	200	4	0.5	1	0	3.94	3.71	5.84	5.00	4.55	0.74	0.72	0.78	0.75	0.66	0.76
SPC	100	200	4	0.5	1	0	4.00	3.98	4.43	4.22	4.08	0.89	0.88	0.89	0.86	0.78	0.89
SPC	150	200	4	0.5	1	0	4.00	4.00	4.35	4.07	3.99	0.92	0.92	0.92	0.90	0.82	0.92
Post-SPC	25	50	4	0.5	1	0	5.08	3.08	7.58	4.68	3.88	0.57	0.42	0.70	0.56	0.45	0.53
Post-SPC	25	100	4	0.5	1	0	7.51	6.53	7.96	5.38	4.47	0.66	0.62	0.68	0.59	0.48	0.53
Post-SPC	50	100	4	0.5	1	0	3.81	3.05	7.02	5.05	4.37	0.68	0.61	0.78	0.72	0.61	0.71
Post-SPC	50	200	4	0.5	1	0	4.21	3.90	6.96	5.09	4.62	0.76	0.75	0.80	0.76	0.68	0.76
Post-SPC	100	200	4	0.5	1	0	4.00	4.00	5.44	4.22	4.15	0.89	0.89	0.90	0.86	0.79	0.89
Post-SPC	150	200	4	0.5	1	0	4.00	4.00	5.40	4.04	3.94	0.93	0.93	0.93	0.90	0.82	0.93

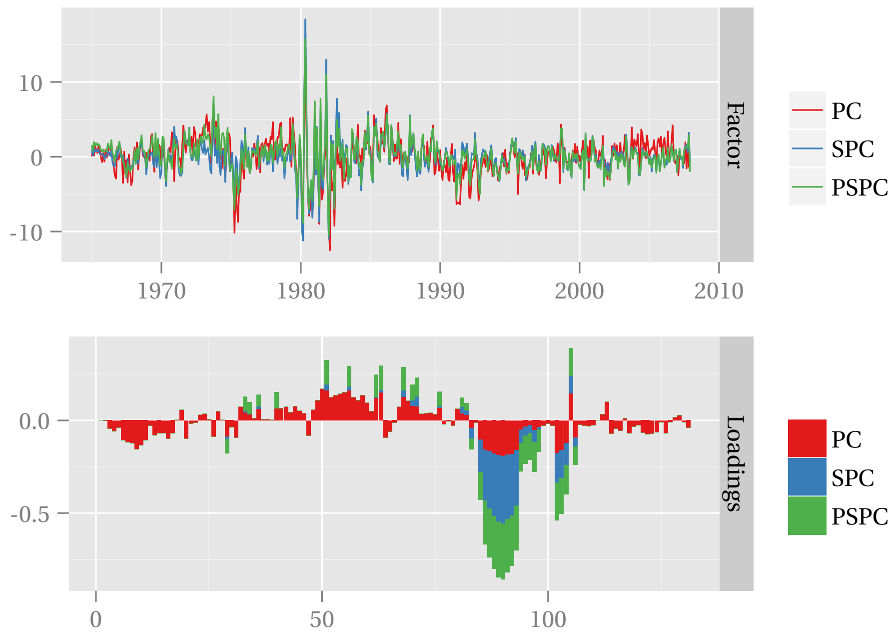
Note: The results are based on 1,000 Monte Carlo replications.

Figure A.1. Estimates of the third factor and associated loadings



Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.130.

Figure A.2. Estimates of the fourth factor and associated loadings

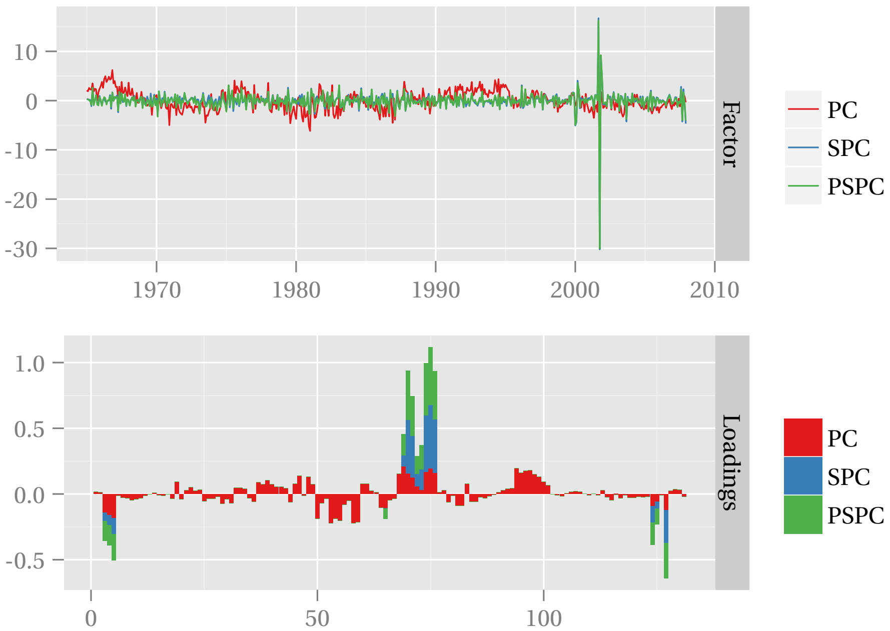


Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.267.

Figure A.3. Estimates of the fifth factor and associated loadings

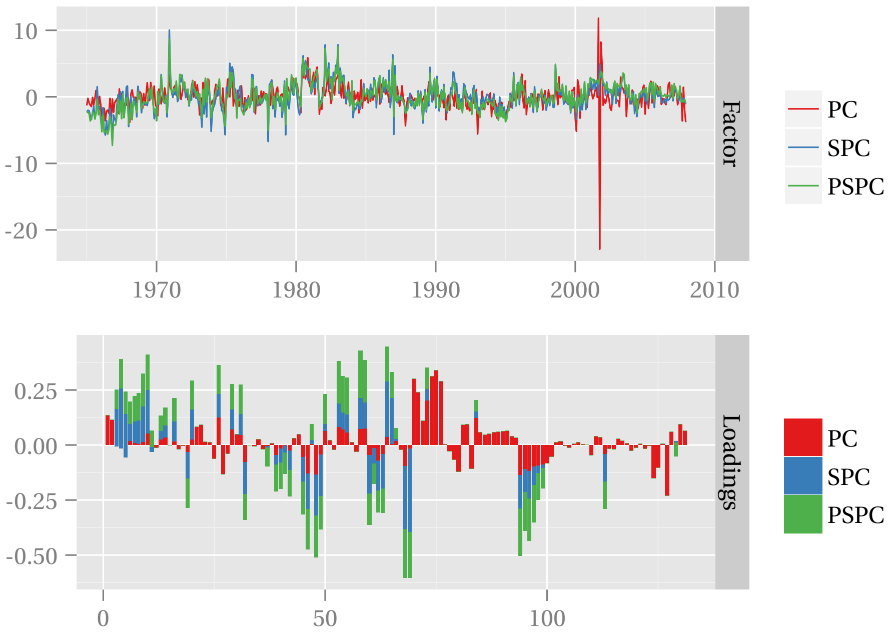
Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.328.

Figure A.4. Estimates of the sixth factor and associated loadings



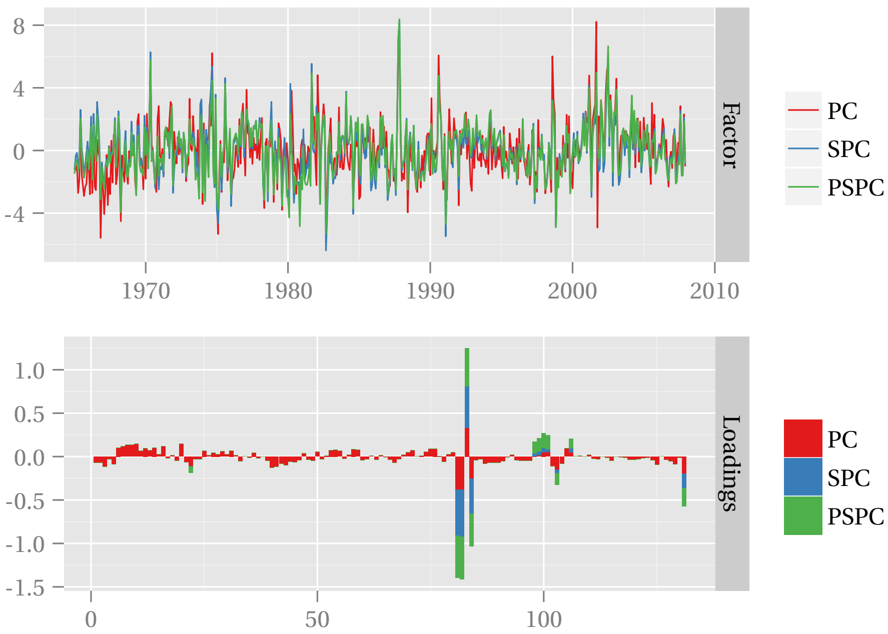
Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.115.

Figure A.5. Estimates of the seventh factor and associated loadings



Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.405.

Figure A.6. Estimates of the eighth factor and associated loadings



Notes: PSPC refers to Post-SPC. Fraction of non-zero loadings for (Post-)SPC: 0.092.

Appendix: Data Description

The dataset used is from Ludvigson and Ng (2010) and can be downloaded from their homepages. The full list of variables along with descriptions from Ludvigson and Ng (2010) has been reproduced below. The majority of the variables are from the Global Insights Basic Economics Database. The remaining variables are either from The Conference Boards Indicators Database (TCB) or calculated by the authors using Global Insights or TCB data (AC). Transforming the variables to be stationary is done according to the transformation codes (TC): 1, no transformation; 2, first difference; 4, logarithms; 5, first difference of logarithms; 6, second difference of logarithms. In addition to this the following abbreviations are used: SA, seasonally adjusted; NSA, not seasonally adjusted; AR, annual rate; SAAR, seasonally adjusted at an annual rate.

Table A.5. Data description

No.	Short name	Mnemonic	TC	Description
1	PI	ypr	5	Personal Income (AR, Bil. Chain 2000 $) (TCB)
2	PI less transfers	a0m051	5	Personal Income Less Transfer Payments (AR, Bil. Chain 2000 $) (TCB)
3	Consumption	cons_r	5	Real Personal Consumption Expenditures (AC) (Bil. $) pi031 / gmdc
4	M&T sales	mtq	5	Manufacturing And Trade Sales (Mil. Chain 1996 $) (TCB)
5	Retail sales	a0m059	5	Sales Of Retail Stores (Mil. Chain 2000 $) (TCB)
6	IP: total	ips10	5	Industrial Production Index - Total Index
7	IP: products	ips11	5	Industrial Production Index - Products, Total
8	IP: final prod	ips299	5	Industrial Production Index - Final Products
9	IP: cons gds	ips12	5	Industrial Production Index - Consumer Goods
10	IP: cons dble	ips13	5	Industrial Production Index - Durable Consumer Goods
11	IP: cons nondble	ips18	5	Industrial Production Index - Nondurable Consumer Goods
12	IP: bus eqpt	ips25	5	Industrial Production Index - Business Equipment
13	IP: matls	ips32	5	Industrial Production Index - Materials
14	IP: dble matls	ips34	5	Industrial Production Index - Durable Goods Materials
15	IP: nondble matls	ips38	5	Industrial Production Index - Nondurable Goods Materials
16	IP: mfg	ips43	5	Industrial Production Index - Manufacturing (Sic)
17	IP: res util	ips307	5	Industrial Production Index - Residential Utilities
18	IP: fuels	ips306	5	Industrial Production Index - Fuels
19	NAPM prodn	pmp	1	Napm Production Index (Percent)
20	Cap util	utl11	2	Capacity Utilization (SIC-Mfg) (TCB)
21	Help wanted indx	lhel	2	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)
22	Help wanted/emp	lhelx	2	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf
23	Emp CPS total	lhem	5	Civilian Labor Force: Employed, Total (Thous.,Sa)
24	Emp CPS nonag	lhnag	5	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)

Table A.5. Data description (continued)

No.	Short name	Mnemonic	TC	Description
25	U: all	lhur	2	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)
26	U: mean duration	lhu680	2	Unemploy. By Duration: Average (Mean) Duration In Weeks (Sa)
27	U < 5 wks	lhu5	5	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)
28	U 5–14 wks	lhu14	5	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)
29	U 15+ wks	lhu15	5	Unemploy.By Duration: Persons Unempl.15 Wks+ (Thous.,Sa)
30	U 15–26 wks	lhu26	5	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)
31	U 27+ wks	lhu27	5	Unemploy.By Duration: Persons Unempl.27 Wks+ (Thous,Sa)
32	UI claims	claimuii	5	Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)
33	Emp: total	ces002	5	Employees On Nonfarm Payrolls: Total Private
34	Emp: gds prod	ces003	5	Employees On Nonfarm Payrolls - Goods-Producing
35	Emp: mining	ces006	5	Employees On Nonfarm Payrolls - Mining
36	Emp: const	ces011	5	Employees On Nonfarm Payrolls - Construction
37	Emp: mfg	ces015	5	Employees On Nonfarm Payrolls - Manufacturing
38	Emp: dble gds	ces017	5	Employees On Nonfarm Payrolls - Durable Goods
39	Emp: nondbles	ces033	5	Employees On Nonfarm Payrolls - Non-durable Goods
40	Emp: services	ces046	5	Employees On Nonfarm Payrolls - Service-Providing
41	Emp: TTU	ces048	5	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities
42	Emp: wholesale	ces049	5	Employees On Nonfarm Payrolls - Wholesale Trade.
43	Emp: retail	ces053	5	Employees On Nonfarm Payrolls - Retail Trade
44	Emp: FIRE	ces088	5	Employees On Nonfarm Payrolls - Financial Activities
45	Emp: Govt	ces140	5	Employees On Nonfarm Payrolls - Government
46	Avg hrs	ces151	1	Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Goods-Producing
47	Overtime: mfg	ces155	2	Avg Weekly Hrs of Prod or Nonsup Workers On Private Nonfarm Payrolls - Mfg Overtime Hours
48	Avg hrs: mfg	a0m001	1	Average Weekly Hours, Mfg. (Hours) (TCB)
49	NAPM empl	pmemp	1	Napm Employment Index (Percent)
50	Starts: nonfarm	hsfr	4	Housing Starts:Nonfarm(1947-58);Total Farm & Nonfarm(1959-)(Thous.,Saar)
51	Starts: NE	hsne	4	Housing Starts:Northeast (Thous.U.)S.A.
52	Starts: MW	hsmw	4	Housing Starts:Midwest(Thous.U.)S.A.
53	Starts: South	hssou	4	Housing Starts:South (Thous.U.)S.A.
54	Starts: West	hswst	4	Housing Starts:West (Thous.U.)S.A.
55	BP: total	hsbr	4	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
56	BP: NE	hsbne	4	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A

Table A.5. Data description (continued)

No.	Short name	Mnemonic	TC	Description
57	BP: MW	hsbmw	4	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.
58	BP: South	hsbsou	4	Houses Authorized By Build. Permits:South(Thou.U.)S.A.
59	BP: West	hsbwst	4	Houses Authorized By Build. Permits:West(Thou.U.)S.A.
60	PMI	pmi	1	Purchasing Managers' Index (Sa)
61	NAPM new ordrs	pmno	1	Napm New Orders Index (Percent)
62	NAPM vendor del	pmdel	1	Napm Vendor Deliveries Index (Percent)
63	NAPM Invent	pmnv	1	Napm Inventories Index (Percent)
64	Orders: cons gds	a1m008	5	Mfrs' New Orders, Consumer Goods And Materials (Mil. $) (TCB)
65	Orders: dble gds	a0m007	5	Mfrs' New Orders, Durable Goods Industries (Bil. Chain 2000 $) (TCB)
66	Orders: cap gds	a0m027	5	Mfrs' New Orders, Nondefense Capital Goods (Mil. Chain 1982 $) (TCB)
67	Unf orders: dble	a1m092	5	Mfrs' Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $) (TCB)
68	M&T invent	a0m070	5	Manufacturing And Trade Inventories (Bil. Chain 2000 $) (TCB)
69	M&T invent/sales	a0m077	2	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $) (TCB)
70	M1	fm1	6	Money Stock: M1 (Curr, Trav.Cks, Dem Dep, Other Ck'able Dep) (Bil. $,Sa)
71	M2	fm2	6	Money Stock: M2 (M1+O'nite Rps, Euro$, G/P&B/D & Mmmfs&Sav& Sm Time Dep (Bil. $, Sa)
72	Currency	fmscu	6	Money Stock: Currency held by the public (Bil. $,Sa)
73	M2 (real)	fm2_r	5	Money Supply: Real M2, fm2 / gmdc (AC)
74	MB	fmfba	6	Monetary Base, Adj For Reserve Requirement Changes (Mil. $,Sa)
75	Reserves tot	fmrra	6	Depository Inst Reserves:Total, Adj For Reserve Req Chgs(Mil. $,Sa)
76	Reserves nonbor	fmrnba	6	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil. $,Sa)
77	C&I loans	fclnbw	6	Commercial & Industrial Loans Outstanding + NonFin Comm. Paper(Mil. $,Sa)
78	C&I loans	fclbmc	1	Wkly Rp Lg Com'l Banks:Net Change Com'l & Indus Loans(Bil. $,Saar)
79	Cons credit	ccinrv	6	Consumer Credit Outstanding - Nonrevolving(G19)
80	Inst cred/PI	ccipy	2	Ratio, Consumer Installment Credit To Personal Income (Pct.) (TCB)
81	S&P 500	fspcom	5	S&P's Common Stock Price Index: Composite (1941-43=10)
82	S&P: indust	fspin	5	S&P's Common Stock Price Index: Industrials (1941-43=10)
83	S&P div yield	fsdxp	2	S&P's Composite Common Stock: Dividend Yield (% Per Annum)
84	S&P PE ratio	fspxe	5	S&P's Composite Common Stock: Price-Earnings Ratio (%,Nsa)
85	Fed Funds	fyff	2	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)
86	Comm paper	cp90	2	Commercial Paper Rate
87	3 mo T-bill	fygm3	2	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)
88	6 mo T-bill	fygm6	2	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)

Table A.5. Data description (continued)

No.	Short name	Mnemonic	TC	Description
89	1 yr T-bond	fygt1	2	Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)
90	5 yr T-bond	fygt5	2	Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)
91	10 yr T-bond	fygt10	2	Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)
92	Aaa bond	fyaaac	2	Bond Yield: Moody's Aaa Corporate (% Per Annum)
93	Baa bond	fybaac	2	Bond Yield: Moody's Baa Corporate (% Per Annum)
94	CP-FF spread	scp90	1	cp90-fyff (AC)
95	3 mo-FF spread	sfygm3	1	fygm3-fyff (AC)
96	6 mo-FF spread	sfygm6	1	fygm6-fyff (AC)
97	1 yr-FF spread	sfygt1	1	fygt1-fyff (AC)
98	5 yr-FF spread	sfygt5	1	fygt5-fyff (AC)
99	10 yr-FF spread	sfygt10	1	fygt10-fyff (AC)
100	Aaa-FF spread	sfyaaac	1	fyaaac-fyff (AC)
101	Baa-FF spread	sfybaac	1	fybaac-fyff (AC)
102	Ex rate: avg	exrus	5	United States;Effective Exchange Rate(Merm)(Index No.)
103	Ex rate: Switz	exrsw	5	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)
104	Ex rate: Japan	exrjan	5	Foreign Exchange Rate: Japan (Yen Per U.S.$)
105	Ex rate: UK	exruk	5	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
106	Ex rate: Canada	exrcan	5	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)
107	PPI: fin gds	pwfsa	6	Producer Price Index: Finished Goods (82=100,Sa)
108	PPI: cons gds	pwfcsa	6	Producer Price Index: Finished Consumer Goods (82=100,Sa)
109	PPI: int materials	pwimsa	6	Producer Price Index: Intermed Mat.Supplies & Components(82=100,Sa)
110	PPI: crude matls	pwcmsa	6	Producer Price Index: Crude Materials (82=100,Sa)
111	Spot market price	psccom	6	Spot market price index: bls & crb: all commodities(1967=100)
112	PPI: nonferrous matls	pw102	6	Producer Price Index: Nonferrous Materials (1982=100, Nsa)
113	NAPM com price	pmcp	1	Napm Commodity Prices Index (Percent)
114	CPI-U: all	punew	6	Cpi-U: All Items (82-84=100,Sa)
115	CPI-U: apparel	pu83	6	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
116	CPI-U: transp	pu84	6	Cpi-U: Transportation (82-84=100,Sa)
117	CPI-U: medical	pu85	6	Cpi-U: Medical Care (82-84=100,Sa)
118	CPI-U: comm.	puc	6	Cpi-U: Commodities (82-84=100,Sa)
119	CPI-U: dbles	pucd	6	Cpi-U: Durables (82-84=100,Sa)
120	CPI-U: services	pus	6	Cpi-U: Services (82-84=100,Sa)
121	CPI-U: ex food	puxf	6	Cpi-U: All Items Less Food (82-84=100,Sa)
122	CPI-U: ex shelter	puxhs	6	Cpi-U: All Items Less Shelter (82-84=100,Sa)
123	CPI-U: ex med	puxm	6	Cpi-U: All Items Less Midical Care (82-84=100,Sa)
124	PCE defl	gmdc	6	Pce, Impl Pr Defl:Pce (2000=100) (AC) (BEA)
125	PCE defl: dlbes	gmdcd	6	Pce, Impl Pr Defl:Pce; Durables (2000=100) (AC) (BEA)
126	PCE defl: nondble	gmdcn	6	Pce, Impl Pr Defl:Pce; Nondurables (2000=100) (AC) (BEA)
127	PCE defl: service	gmdcs	6	Pce, Impl Pr Defl:Pce; Services (2000=100) (AC) (BEA)

Table A.5. Data description (continued)

No.	Short name	Mnemonic	TC	Description
128	AHE: goods	ces275	6	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm – Goods-Producing
129	AHE: const	ces277	6	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm – Construction
130	AHE: mfg	ces278	6	Avg Hourly Earnings of Prod or Nonsup Workers On Private Nonfarm – Manufacturing
131	Consumer expect	hhsntn	2	U. Of Mich. Index Of Consumer Expectations (Bcd-83)

References

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

Bai, J., Ng, S., 2008a. Forecasting economic time series using targeted predictors. Journal of Econometrics 146 (2), 304–317.

Bai, J., Ng, S., 2008b. Large dimensional factor analysis. Foundations and Trends in Econometrics 3 (2), 89–163.

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19 (2), 521–547.

Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? Journal of Econometrics 132 (1), 169–194.

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

Dias, F., Pinheiro, M., Rua, A., 2010. Forecasting using targeted diffusion indexes. Journal of Forecasting 29 (3), 341–352.

Esbensen, K., Geladi, P., Wold, S., 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 2, 37–52.

Friedman, J. H., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33 (1), 1–22.

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

Knight, K., Fu, W., 2000. Asymptotics for LASSO-type estimators. Annals of Statistics, 1356–1378.

Kristensen, J. T., 2013. Factor-based forecasting in the presence of outliers: Are factors better selected and estimated by the median than by the mean? Studies in Nonlinear Dynamics & Econometrics, to appear.

Ludvigson, S., Ng, S., 2010. A factor analysis of bond risk premia. In: Ulah, A., Giles, D. (Eds.), Handbook of Empirical Economics and Finance. Statistics: A Series of Textbooks and Monographs. Chapman and Hall, pp. 313–372.

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

Stock, J., Watson, M., 1998. Diffusion indexes. NBER Working Paper 6702, National Bureau of Economic Research.

Stock, J., Watson, M., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

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

Stock, J., Watson, M., 2005. Implications of dynamic factor models for VAR analysis. NBER Working Paper 11467, National Bureau of Economic Research.

Stock, J., Watson, M., 2006. Forecasting with many predictors. In: Elliott, G., Granger, C. W., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 1. Elsevier, pp. 515–554.

Stock, J., Watson, M., 2011. Dynamic factor models. In: Clements, M. P., Hendry, D. F. (Eds.), Oxford Handbook of Economic Forecasting. Oxford University Press, USA, pp. 35–59.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.

Witten, D., Tibshirani, R., Hastie, T., 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10 (3), 515–534.

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