Journal of Econometrics 176 (2013) 18–29

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Principal components estimation and identification of static factors

Jushan Bai[a,*], Serena Ng[b]

[a] Department of Economics, Columbia University, IAB 1019, 420 West 118th Street, New York, NY 10027, United States
[b] Department of Economics, Columbia University, IAB 1117, 420 West 118th Street, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 13 December 2012
Received in revised form
13 December 2012
Accepted 25 March 2013
Available online 1 April 2013

JEL classification:
C30
C33

Keywords:
Diffusion indices
FAVAR
Rotation
Factor space
Skew-symmetric matrices

ABSTRACT

It is known that the principal component estimates of the factors and the loadings are rotations of the underlying latent factors and loadings. We study conditions under which the latent factors can be estimated asymptotically without rotation. We derive the limiting distributions for the estimated factors and factor loadings when N and T are large and make precise how identification of the factors affects inference based on factor augmented regressions. We also consider factor models with additive individual and time effects. The asymptotic analysis can be modified to analyze identification schemes not considered in this analysis.

1. Introduction

Large dimensional factor analysis has been found to be useful in an increasingly large number of applications, and the theoretical properties of the principal components estimates are quite well understood. The method of principal components estimates the space spanned by the latent factors instead of the factors themselves. Thus, if F_t is the $r \times 1$ vector of latent factors, and $\tilde{F}_t$ is the vector of factor estimates, there exists an $r \times r$ invertible matrix H such that $\tilde{F}_t$ estimates $H'F_t$. Asymptotic results are stated in terms of $\tilde{F}_t - H'F_t$. Similarly, if λ_i is the vector of factor loadings and $\tilde{\lambda}_i$ is the corresponding estimate, asymptotic results are known for $\tilde{\lambda}_i - H^{-1}\lambda_i$.

In some instances, the object of interest is the conditional mean, and interpretation of the parameters that determine the conditional mean is not necessary. For example, in diffusion index forecasting analysis of Stock and Watson (2002), the primary interest is the predicted value of the dependent variable. In factor augmented regressions, the factors are merely present to control for latent common effects. In problems with errors-in-variables or endogeneity such as considered in Bai and Ng (2010), one only needs the factors to be strongly correlated with the endogenous regressor to validate the factors as instruments. In all these cases, we are not interested in the coefficients on the factors per se and being able to estimate a rotation of F_t suffices.

There are, however, cases when the parameters of interest are the coefficients associated with the factors, or even the factors themselves. For example, in arbitrage pricing theory, restrictions on the factor loadings would be necessary to pin down the sensitivity to risk factors such as inflation, real activity, and financial markets. In factor augmented regressions of the form $y_t = \alpha'\tilde{F}_t + W_t'\beta + \varepsilon_t$, one might be interested in testing a hypothesis concerning α. Since the asymptotic theory is only available for $\sqrt{T}(\hat{\alpha} - H^{-1}\alpha)$, the test is uninformative except when α is zero. If H is known, $\hat{\alpha}$ can be given economic interpretation.

We study three sets of restrictions such that F and Λ are exactly identified. If the underlying F and Λ that generate the data satisfy those restrictions then H is asymptotically an identity matrix. This is useful because $\tilde{F}_t$ can be treated as though they were the latent F_t and α can be learnt from $\hat{\alpha}$. We derive the asymptotic distributions for the estimated factors and the loadings based on these restrictions. In case there exist no F and Λ that satisfy any of the identification conditions considered here, the rotation matrix H will not be an identity matrix asymptotically and we will be estimating rotations of the underlying F and Λ. Other identification conditions may be considered; the method developed in this paper should be useful to derive the corresponding limiting distributions.

* Corresponding author.
E-mail addresses: jb3064@columbia.edu, jushan.bai@columbia.edu (J. Bai), Serena.Ng@columbia.edu (S. Ng).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.03.007

Our analysis is extended to allow for (i) additive individual effects, (ii) common time effects, and (iii) heterogeneous time trends in the panel of data.

2. Factor models and identification

Let T and N denote the sample size in the time series and cross-section dimensions, respectively. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the observation X_{it} has a factor structure represented as

$$X_{it} = \lambda_i' F_t + e_{it}.$$

As written, there are no deterministic terms. Individual fixed effects and time trends will be considered subsequently. Let X and e be $T \times N$ matrices. The factor model in matrix form is

$$X = F\Lambda' + e$$

where $F = (F_1, F_2, \ldots, F_T)'$ is the $T \times r$ matrix of factors and $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$ is the $N \times r$ matrix of factor loadings. Our objective is to estimate both F and Λ. We make the following assumptions:

Assumption A. There exists an $M < \infty$, not depending on N and T, such that

(a) $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F > 0$ is a $r \times r$ non-random matrix.
(b) λ_i is either deterministic such that $\|\lambda_i\| \leq M$, or it is stochastic such that $E\|\lambda_i\|^4 \leq M$. In either case, $N^{-1}\Lambda'\Lambda \xrightarrow{p} \Sigma_\Lambda > 0$ is a $r \times r$ non-random matrix as $N \to \infty$.
(c.i) $E(e_{it}) = 0$ and $E|e_{it}|^8 \leq M$.
(c.ii) $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t, s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i, j). Furthermore, $\sum_{i=1}^{N} \bar{\sigma}_{ij} \leq M$ for each j, $\sum_{t=1}^{T} \tau_{ts} \leq M$ for each s, and $\frac{1}{NT}\sum_{i,j,t,s=1} |\sigma_{ij,ts}| \leq M$.
(c.iii) For every (t, s), $E|N^{-1/2}\sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M$.
(d) $\{\lambda_i\}$, $\{F_t\}$, and $\{e_{it}\}$, are three mutually independent groups.
(e) (i) $N^{-1/2}\sum_{i=1}^{N} \lambda_i e_{it} \xrightarrow{d} N(0, \Gamma_t)$; (ii) $T^{-1/2}\sum_{t=1}^{T} F_t e_{it} \xrightarrow{d} N(0, \Phi_i)$.

Assumptions A(a) and (b) imply the existence of r factors. The idiosyncratic errors e_{it} are allowed to be cross-sectionally and serially correlated, but only weakly as stated under condition A(c). If e_{it} are iid, then A(c.ii) and A(c.iii) are satisfied. Assumption A(d) allows within group dependence, meaning that F_t can be serially correlated, λ_i can be correlated over i, and e_{it} can have serial and cross-sectional correlations that are not too strong so that A(a)–(c) hold. We assume no dependence between the factor loadings and the factors, or between the factors and the idiosyncratic errors, which is the meaning of mutual independence between groups. Part (e) of Assumption A defines the limiting covariance of the factors.

The method of principal components minimizes the objective function $\text{tr}[(X - F\Lambda')'(X - F\Lambda')]$ by choosing the normalizations that $F'F/T = I_r$ and $\Lambda'\Lambda$ is diagonal. The estimator for F, denoted $\tilde{F} = (\tilde{F}_1, \ldots, \tilde{F}_T)'$, is a $T \times r$ matrix consisting of r unitary eigenvectors (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix $XX'/(TN)$ in decreasing order. Then $\tilde{\Lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N)' = X'\tilde{F}/T$ is a $N \times r$ matrix of estimated factor loadings. The estimators $\tilde{F}$ and $\tilde{\Lambda}$ satisfy the normalization restrictions since $\tilde{F}'\tilde{F}/T = I_r$ holds by construction and $\tilde{\Lambda}'\tilde{\Lambda}/N = \tilde{V}$ where $\tilde{V}$ is a $r \times r$ diagonal matrix consisting of the r largest eigenvalues of $XX'/(TN)$.

While the restrictions used by the principal components estimator identify the space spanned by the columns of F and the space spanned by the columns of Λ, they do not necessarily identify the individual columns of F or of Λ. To be precise, let H be an $r \times r$ matrix whose transpose is

$$H' = \tilde{V}^{-1}(\tilde{F}'F/T)(\Lambda'\Lambda/N). \tag{1}$$

Under Assumption A, Stock and Watson (2002) and Bai and Ng (2002) showed that H is invertible, $\tilde{F}$ estimates FH (a rotation of F), and $\tilde{\Lambda}$ estimates $\Lambda H'^{-1}$ (a rotation of Λ), though the product $\tilde{F}\tilde{\Lambda}'$ estimates $F\Lambda'$.

We are specifically interested in conditions under which we can identity the columns of F and the columns of Λ from the product $F\Lambda'$. Notice that $F\Lambda' = FRR^{-1}\Lambda'$ for any $r \times r$ invertible matrix R, and R has r^2 free parameters. Thus we need at least r^2 restrictions in order to identity F and Λ, see Lawley and Maxwell (1971). While more than r^2 restrictions can be imposed as in Heaton and Solo (2004) and Reis and Watson (2010), the method of principal components is not suitable for imposing over-identifying restrictions. We consider three sets of restrictions that will lead to exact identification. We then show that if the true F and true Λ satisfy these restrictions, then the corresponding rotation matrix is asymptotically an identity matrix.[1]

Identifying restrictions:		
	Restrictions on F	Restrictions on Λ
(2.1): PC1	$\frac{1}{T}F'F = I_r$	$\Lambda'\Lambda$ is a diagonal matrix with distinct entries
(2.2): PC2	$\frac{1}{T}F'F = I_r$	$\Lambda = \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix}$, $\Lambda_1 = \begin{pmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} \end{pmatrix}$, $\lambda_{ii} \neq 0$, $i = 1, \ldots, r$
(2.3): PC3	Unrestricted	$\Lambda = \begin{pmatrix} I_r \\ \Lambda_2 \end{pmatrix}$

2.1. PC1

PC1 requires that the diagonal elements of $\Lambda'\Lambda$ are distinct and positive and are arranged in decreasing order. The standard method of principal components implicitly invokes the first restriction in PC1 but does not require the diagonal matrix $\Lambda'\Lambda$ to have distinct elements. Without this restriction, the principal components estimator cannot identity the individual columns of F and those of Λ, and there will be rotational indeterminacy. Under PC1, the normalization on F gives $r(r + 1)/2$ restrictions since a symmetric matrix contains $r(r+1)/2$ free parameters. The diagonality of $\Lambda'\Lambda$ gives $r(r - 1)/2$ restrictions. Together, the two normalizations lead to exactly r^2 restrictions. We show in the Appendix that if the restrictions defined by PC1 also hold for the underlying F and Λ that generate the data, then

$$H = I_r + O_p(\delta_{NT}^{-2}), \tag{2}$$

where δ_{NT} denotes $\min\left[\sqrt{N}, \sqrt{T}\right]$ throughout this paper.

PC1 is a statistical restriction and is often used in the maximum likelihood estimation, see Lawley and Maxwell (1971). This identification condition is less restrictive than it appears. A block diagonal matrix of factor loadings also satisfies PC1.[2] For example, with

[1] By symmetry, three different sets of identification restrictions can be obtained by switching F and Λ. For example, $\frac{1}{T}F'F$ is diagonal and $\frac{1}{N}\Lambda'\Lambda = I_r$. Since the asymptotic results still hold by switching the role of F and Λ, we only consider the three sets of restrictions given above.

[2] An extension of this model is the inclusion of a global factor, see for example, Moench and Ng (2011), Hallin and Liska (2008) and Wang (2008). However, the factor loading matrix does not necessarily satisfy PC1; it will satisfy PC2 if there is a cross-section unit which is affected by the global factor only.

$r = 3$, the following loading matrix will satisfy PC1:

$$\Lambda = \begin{bmatrix} \pi_1 & 0 & 0 \\ 0 & \pi_2 & 0 \\ 0 & 0 & \pi_3 \end{bmatrix}$$

where π_i is a vector of $N_i \times 1$ with $N_1 + N_2 + N_3 = N$. The loading matrix implies that the first factor affects the first N_1 individuals, the second factor affects the next N_2 individuals, and so on. This case is potentially useful for economic applications. PC1 still holds under an arbitrary permutation of the cross sections. Thus in the block diagonal case, it is not required to know which cross section units belong to the first group (affected by the first factor) and which belong to the second group, and so forth.

The next two sets of restrictions, PC2 and PC3, involve ordering the data. Both of which are frequently used in empirical work.

2.2. PC2

PC2 restricts Λ_1 to be an invertible lower triangular matrix. It thus requires knowledge of which variable is affected by the first factor only, which variable is affected by the first two factors only, and so on.[3] PC2 is analogous to a triangular system of simultaneous equations. The choice of the first r variables of X_t and their ordering provide the auxiliary information for identification.

Given the unrestricted estimates $\tilde{F}$ and $\tilde{\Lambda}$, it is easy to obtain estimators satisfying PC2. Let $\tilde{\Lambda}_1$ be the first $r \times r$ block of $\tilde{\Lambda}$ and let $\hat{F}$ and $\hat{\Lambda}$ denote the estimators that satisfy PC2, i.e., $\hat{F}'\hat{F}/T = I_r$ and $\hat{\Lambda}_1$ is lower triangular. Then $\hat{F}$ and $\hat{\Lambda}$ can be obtained as follows.

- Step 1: obtain a QR decomposition of $\tilde{\Lambda}_1'$ to yield $\tilde{\Lambda}_1' = Q \cdot R$, where R is an upper triangular matrix with positive diagonal elements, and Q is an $r \times r$ orthogonal matrix such that $Q'Q = I_r$. This decomposition is unique for any invertible $\tilde{\Lambda}_1$.
- Step 2: define

$$\hat{F} = \tilde{F} \cdot Q, \qquad \hat{\Lambda} = \tilde{\Lambda} \cdot Q = \begin{bmatrix} R' \\ \hat{\Lambda}_2 \end{bmatrix}.$$

By construction, $\hat{F}'\hat{F}/T = Q'(\tilde{F}'\tilde{F}/T)Q = Q'Q = I_r$. The new rotation matrix is $H^* = HQ$.

Since $\hat{F}$ and $\hat{\Lambda}$ are rotations of the principal component estimates $\tilde{F}$ and $\tilde{\Lambda}$, they are equivalent in a certain sense. However, their asymptotic distributions will be different. We show in the Appendix that H^* is asymptotically an identity matrix, but $\sqrt{T}(H^* - I_r)$ is asymptotically non-negligible unless $r = 1$. More specifically, if the true F and Λ satisfy PC2, then

$$\begin{cases} H^* - I_r = O_p(\delta_{NT}^{-2}), & r = 1 \\ H^* - I_r = O_p(T^{-1/2}), & r > 1. \end{cases}$$

This implies that $\Xi = \sqrt{T}(H^* - I_r) = o_p(1)$ for $r = 1$. In fact, when $r = 1$, PC1 and PC2 are identical and (2) is in agreement with $\Xi = o_p(1)$. However, $\Xi = \sqrt{T}(H^* - I_r) = O_p(1)$ when $r > 1$. In fact, the limit of Ξ is a skew-symmetric random matrix.[4] In consequence, the limiting distributions of $\hat{F}_t$ and $\hat{\lambda}_i$ will be affected by Ξ.

2.3. PC3

The third set of identification restrictions specify the first $r \times r$ block of Λ (denoted Λ_1) to be an identity matrix and leaves the factor process F completely unrestricted other than requiring $F'F/T$ to be invertible so that r factors exist. Unlike PC1 and PC2, all r^2 restrictions are imposed on Λ under PC3. The restrictions imply that the first variable X_{1t} is affected by the first factor only, the second variable X_{2t} is affected by the second factor only, etc. The resulting structure resembles the classical 'errors-in-variables' model in which $X_{it} = F_{ti} + e_{it}$ for $i = 1, \ldots, r$, as in Pantula and Fuller (1986), and Wansbeek and Meijer (2000, pp. 148–150). While PC3 requires the choice of the first r variables, the estimators for Λ and F are easy to obtain. Given the principal components estimates $\tilde{\Lambda}$ and $\tilde{F}$, let

$$\hat{\Lambda} = \tilde{\Lambda}\tilde{\Lambda}_1^{-1}, \qquad \hat{F} = \tilde{F}\tilde{\Lambda}_1'.$$

The rotation matrix in this case is $H^\dagger = H\tilde{\Lambda}_1'$ because $\hat{F} = \tilde{F}\tilde{\Lambda}_1' = FH\tilde{\Lambda}_1' + o_p(1)$. If the F and Λ underlying the data satisfy PC3, then $H^\dagger$ will converge in probability to I_r. It follows that $\hat{F}$ estimates F and $\hat{\Lambda}$ estimates Λ without rotation. We show in the Appendix that

$$\sqrt{T}(H^\dagger - I_r) = \xi_T + o_p(1) \tag{3}$$

where ξ_T is defined in (11) below. The fact that $\sqrt{T}(H^\dagger - I_r)$ is not negligible for all $r \geq 1$ will affect the limiting distributions of $\hat{\lambda}_i$ and $\hat{F}_t$.

Remark 1 (*Local vs. Global Identification*)**.** Conditions for global and local identification of factor models are discussed, for example, in Bekker (1986) and Algina (1980). Both PC1 and PC2 identity F and Λ up to a column sign change. Changing the sign of any column of F and the sign of the corresponding column of Λ will leave the product $F\Lambda'$ unchanged.

The resulting new F and new Λ still satisfy PC1, and hence observationally equivalent to the original F and Λ. Thus PC1 and PC2 are local identification conditions. However, once we fix the column signs of Λ (or F), PC1 and PC2 become global identification conditions. There will be no other F and Λ with the given column signs and the given product $F\Lambda'$.

To understand how global identification is achieved, consider PC2. Once $F\Lambda'$ is given then $\Lambda(F'F/T)\Lambda' = \Lambda\Lambda'$ is known, since $F'F/T = I_r$. Let $C = \Lambda\Lambda'$. From $\Lambda' = (\Lambda_1', \Lambda_2')$, we have

$$\Lambda\Lambda' = \begin{bmatrix} \Lambda_1\Lambda_1' & \Lambda_1\Lambda_2' \\ \Lambda_2\Lambda_1' & \Lambda_2\Lambda_2' \end{bmatrix}, \qquad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where we also partition matrix C correspondingly. Suppose for concreteness that $r = 3$. Knowing C_{11} is equivalent to knowing the elements of

$$\Lambda_1\Lambda_1' = \begin{bmatrix} \lambda_{11}^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{31} \\ - & \lambda_{21}^2 + \lambda_{22}^2 & \lambda_{21}\lambda_{31} + \lambda_{22}\lambda_{32} \\ - & - & \lambda_{31}^2 + \lambda_{32}^2 + \lambda_{33}^2 \end{bmatrix}.$$

If the sign of λ_{11} is known, then λ_{11} is identified from λ_{11}^2. Since $\lambda_{11} \neq 0$, λ_{21} and λ_{31} can be identified, which further implies the identification of λ_{22}^2. If the sign of λ_{22} is known, then λ_{22} is also identified. Since $\lambda_{22} \neq 0$, this implies the identification of λ_{32}. The same reasoning implies the identification of λ_{33}, given its sign. In summary, we can identify Λ_1 provided that Λ_1 is invertible and the signs of λ_{ii} $(i = 1, 2, 3)$ are known.[5] Next, from $C_{21} = \Lambda_2\Lambda_1'$, we identify Λ_2 from $\Lambda_2 = C_{21}(\Lambda_1')^{-1}$. Thus PC2 together with the column signs of Λ (or F) imply global identification in the restricted parameter space that ensures invertibility of Λ_1.

[3] The structure of Λ is similar to Stock and Watson (2005), though they are interested in identification of shocks to the factors rather than the factors. A variation to PC2 is to normalize the diagonal elements λ_{ii} $(i = 1, 2, \ldots, r)$ to be 1, with $F'F/T$ being diagonal (instead of an identity matrix).

[4] A matrix C is skew-symmetric (also known as anti-symmetric) if $C + C' = 0$. So the diagonal elements of a skew-symmetric matrix are zero, and $C_{ij} = -C_{ji}$.

[5] Identification of Λ_1 alone does not require $\lambda_{33} \neq 0$, but further identification of Λ_2 does need $\lambda_{33} \neq 0$ so that Λ_1 is invertible.

PC3 also implies global identification, but sign restrictions are not necessary. To see this, let $C = \Lambda(F'F/T)\Lambda'$ be given. Under PC3,

$$\Lambda(F'F/T)\Lambda' = \begin{bmatrix} (F'F/T) & (F'F/T)\Lambda_2' \\ \Lambda_2(F'F/T) & \Lambda_2\Lambda_2' \end{bmatrix}.$$

Knowing C_{11} is equivalent to knowing $F'F/T$. Thus we identity Λ_2 from $\Lambda_2 = C_{21}(F'F/T)^{-1} = C_{21}C_{11}^{-1}$.

3. Asymptotic theory

We are interested in the implications of using the factor estimates identified using PC1, PC2, or PC3 for inference. To this end, let

$$Z_{Ti} = (F'F/T)^{-1}T^{-1/2}\sum_{t=1}^{T} F_t e_{it}.$$

By Assumption A(e), $Z_{Ti} \xrightarrow{d} Z_i$ where Z_{Ti} is a zero mean normal vector as $T \to \infty$. To derive the limiting distribution for $\hat{F}_t$ and $\hat{\lambda}_i$, we use the asymptotic representations for $\tilde{F}_t$ and $\tilde{\lambda}_i$, given in Theorems 1 and 2 of Bai (2003). Specifically, if $\sqrt{N}/T \to 0$, then

$$\sqrt{N}(\tilde{F}_t - H'F_t) = \tilde{V}^{-1}\left(\frac{\tilde{F}'F}{T}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1) \tag{4}$$

and if $\sqrt{T}/N \to 0$,

$$\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i) = H'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it} + o_p(1). \tag{5}$$

A useful and alternative expression for (4) is

$$\sqrt{N}(\tilde{F}_t - H'F_t) = H'\left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1) \tag{6}$$

because (1) implies $\tilde{V}^{-1}\left(\frac{\tilde{F}'F}{T}\right) = \tilde{V}^{-1}\left(\frac{\tilde{F}'F}{T}\right)(\Lambda'\Lambda/N)(\Lambda'\Lambda/N)^{-1} = H'(\Lambda'\Lambda/N)^{-1}$.

3.1. PC1

Under PC1, $H' = I_r + O_p(\delta_{NT}^{-2})$. It follows that $\sqrt{N}(\tilde{F}_t - F_t) = \sqrt{N}(\tilde{F}_t - H'F_t) + \sqrt{N}(\tilde{H}' - I_r)F_t = \sqrt{N}(\tilde{F}_t - H'F_t) + o_p(1)$, provided that $\sqrt{N}/\delta_{NT}^2 = o(1)$, or equivalently, $\sqrt{N}/T \to 0$. Thus under PC1, we can rewrite (6) as

$$\sqrt{N}(\tilde{F}_t - F_t) = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1). \tag{7}$$

This result says that $\tilde{F}_t$ is asymptotically equivalent to the least squares estimator for F_t in a cross-section regression with Λ as the regressor, as if Λ were observable. Similarly, if $\sqrt{T}/N \to 0$ and $H^{-1} = I_r + O_p(\delta_{NT}^{-2})$, then

$$\sqrt{T}(\tilde{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it} + o_p(1) \tag{8}$$

because $F'F/T = I_r$ and $\sqrt{T}(H^{-1} - I_r) = o_p(1)$ if $\sqrt{T}/N \to 0$. In view of (8), we can now interpret $\tilde{\lambda}_i$ as the least squares estimator for λ_i in a time series regression with F as regressor, as though it were observed. These representations and the required relative rate between N and T are the same as in (4) and (5), except that we replace H by an identity matrix in view of the identification restrictions.

The fact that H is an r dimensional identity matrix asymptotically simplifies the limiting distributions for $\tilde{F}_t$ and $\tilde{\lambda}_i$ because the right hand sides of (7) and (8) do not depend on any estimated quantities.

Theorem 1. *Suppose that Assumption A and PC1 hold. Let $\tilde{F}_t$ and $\tilde{\lambda}_i$ be obtained by the method of principal components. Then as $N, T \to \infty$ with $\sqrt{N}/T \to 0$, we have*

$$\sqrt{N}(\tilde{F}_t - F_t) \xrightarrow{d} N(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}). \tag{9}$$

Furthermore, if $\sqrt{T}/N \to 0$,

$$\sqrt{T}(\tilde{\lambda}_i - \lambda_i) \xrightarrow{d} N(0, \Phi_i). \tag{10}$$

A formal proof is given in the Appendix. In essence, $\tilde{F}'F/T = I_r + O_p(\delta_{NT}^{-1})$, and $\tilde{V} = \Lambda'\Lambda/N + O_p(\delta_{NT}^{-2})$ under PC1. Thus the limit of $\tilde{F}'F/T$ is I_r and the limit of $\tilde{V}$ is Σ_Λ. Since $\Lambda'\Lambda/N \to \Sigma_\Lambda$ by Assumption A(b), and (9) follows from (7). Furthermore, (8) together with $F'F/T = I_r$ implies (10). Theorem 1 sheds light on the role of identification assumptions on the principal components estimator. As H and Q are now identity matrices, the identification assumptions affect not just where we center the limiting distribution of the factor estimates, but also their asymptotic variances.

Using the limiting result in (10) we can test if λ_i or some components of λ_i are zero. Consider testing the null hypothesis that $R\lambda_i = \bar{\lambda}_i$, where R is a $(q \times r)$ known restriction matrix ($q \le r$) and $\bar{\lambda}_i$ is $q \times 1$, a known vector. Under the null hypothesis,

$$T(R\tilde{\lambda}_i - \bar{\lambda}_i)'(R\hat{\Phi}_i R')^{-1}(R\tilde{\lambda}_i - \bar{\lambda}_i) \xrightarrow{d} \chi_q^2.$$

We can also test restrictions between λ_i and $\lambda_j (i \neq j)$. Put $\delta = (\lambda_i', \lambda_j')'$ and $\hat{\delta} = (\hat{\lambda}_i', \hat{\lambda}_j')'$. Consider the hypothesis $R\delta = \bar{\delta}$, where R is $q \times 2r$ and $\bar{\delta}$ is $q \times 1$. By the asymptotic representation of (8), if $E(e_{it}e_{jt}) = 0$ for $i \neq j$, then $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are asymptotically independent. So let $\hat{\Phi} = \mathrm{diag}(\hat{\Phi}_i, \hat{\Phi}_j)$ (a block-diagonal matrix), then

$$T(R\hat{\delta} - \bar{\delta})'(R\hat{\Phi}R')^{-1}(R\hat{\delta} - \bar{\delta}) \xrightarrow{d} \chi_q^2.$$

If $E(e_{it}e_{jt}) \neq 0$, then Φ will not be a block diagonal matrix, but it is straightforward to estimate the joint asymptotic covariance matrix. Statistics for testing hypotheses concerning the factors F can be similarly constructed.

3.2. PC2

To derive the asymptotic distributions of $\hat{F}_t$ and $\hat{\lambda}_i$ for PC2, and PC3, we need the following:

Assumption B. $(Z_{Ti}', Z_{T1}', \ldots, Z_{Tr}')' \xrightarrow{d} (Z_i', Z_1' \ldots, Z_r')'$.

The random variables Z_{Ti} are defined earlier. Assumption B strengthens A(e) to require the joint convergence of Z_{Ti} and $(Z_{T1}, \ldots, Z_{Tr})$ to the joint limit of Z_i and $(Z_1, \ldots, Z_r)$. Hereafter, we let ξ_T be an $r \times r$ matrix defined by

$$\begin{aligned}\xi_T &= \left(\frac{F'F}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(F_t e_{1t}, \ldots, F_t e_{rt}) \\ &= (Z_{T1}, \ldots, Z_{Tr}).\end{aligned} \tag{11}$$

The limiting distributions of the factor estimates under PC2 depend on whether $r = 1$ or $r > 1$. If $r = 1$, PC1 and PC2 are identical, so the limiting distributions $\hat{F}_t$ and $\hat{\lambda}_i$ are given in Theorem 1. When

$r > 1$, the representations for $\hat{F}_t$ and $\hat{\lambda}_i$ each has an extra term because $\sqrt{T}(H^* - I_r)$ is non-negligible. More specifically, for $i > r$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it} - \sqrt{T}(H^* - I_r)\lambda_i + o_p(1) \quad (12)$$

and for each t,

$$\sqrt{N}(\hat{F}_t - F_t) = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} - (N/T)^{1/2}\sqrt{T}(H^* - I_r)F_t + o_p(1). \quad (13)$$

Let $\Xi = \sqrt{T}(H^* - I_r)$ and let Ξ_{kh} denote the (k, h)th element of Ξ $(1 \le k, h \le r)$. We show in the Appendix that

$$\Xi_{kh} = \begin{cases} (\xi_T \Lambda_1^{'-1})_{kh} + o_p(1), & k > h \\ o_p(1) & k = h \\ -\Xi_{hk} + o_p(1), & k < h \end{cases} \quad (14)$$

where $o_p(1)$ holds if $\sqrt{T}/N \to 0$. The limit of the off-diagonal elements of Ξ are determined by the limit of the off-diagonal elements of $\xi_T(\Lambda_1')^{-1}$, where ξ_T is defined in (11).

It turns out that (12) also holds for $i = 1, 2, \ldots, r$, not just for $i > r$. For $1 \le i \le r$, the last $r - i$ components of $\hat{\lambda}_i$ and of λ_i are zero. Using the asymptotic representation of $\sqrt{T}(H^* - I_r)$ in (14), it can be shown that the last $r - i$ components of the right hand side of (12) indeed have zero limits.

Recall that $\Sigma_F = I_r$ under PC2, and Z_i is the limiting distribution of $\left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it}$. Let $\text{veck}(A)$ denote the column vector that stacks the lower triangular elements of A (excluding the diagonal elements). Note that $\text{veck}(\cdot)$ is different from $\text{vech}(\cdot)$. For any skew-symmetric matrix A, there is a duplication matrix D such that $\text{vec}(A) = D\,\text{veck}(A)$. Eq. (14) implies $\text{veck}(\Xi) = \text{veck}(\xi_T \Lambda_1^{'-(1)}) + o_p(1)$. From $\xi_T(\Lambda_1')^{-1} \xrightarrow{d} (Z_1, Z_2, \ldots, Z_r)(\Lambda_1')^{-1}$ we have $\text{veck}(\Xi) \xrightarrow{d} \eta$, where η is defined as $\eta = \text{veck}((Z_1, Z_2, \ldots, Z_r)(\Lambda_1')^{-1})$. Then

$$\begin{aligned} \sqrt{T}(H^* - I_r)\lambda_i &= \Xi\lambda_i = (\lambda_i' \otimes I_r)\text{vec}(\Xi) \\ &= (\lambda_i' \otimes I_r)D\,\text{veck}(\Xi) \\ &\xrightarrow{d} (\lambda_i' \otimes I_r)D\eta. \end{aligned}$$

Let Z_i be the limit of the first term on the right hand side of (12). We have

Theorem 2. *Suppose that Assumptions A, B, and* PC2 *hold. Let $\hat{F}_t$ and $\hat{\lambda}_i$ denote the estimates with the restrictions of* PC2.

(i) *Let $Z_i =^d N(0, \Phi_i)$. Then for each i and as $N, T \to \infty$ with $\sqrt{T}/N \to 0$,*

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} Z_i - (\lambda_i' \otimes I_r)D\eta$$

where $\eta = \text{veck}[(Z_1, Z_2, \ldots, Z_r)\Lambda_1^{'-1}]$ and D is a duplication matrix linking vec$(\cdot)$ and $\text{veck}(\cdot)$.

(ii) *Let $G_t =^d N(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1})$ and is independent of η. If $N/T \to c$ with $0 \le c < \infty$,*

$$\sqrt{N}(\hat{F}_t - F_t) \xrightarrow{d} G_t + \sqrt{c}(F_t' \otimes I_r)D\eta,$$

In part (i) of Theorem 2, $(\lambda_i' \otimes I_r)D\eta$ is the limit of $\sqrt{T}(H^* - I_r)\lambda_i$, which is also normal since η is normal. Similarly, for part (ii) of the theorem, G_t is the limit of the first term on the right hand side of (13), and $\sqrt{c}(F_t' \otimes I_r)D\eta$ is the limit of the second term of (13). Hypothesis testing can be performed as in Section 3.1.

3.3. PC3

Similar to PC2, the representations for $\hat{F}_t$ and $\hat{\lambda}_i$ each has an extra term because $\sqrt{T}(H^\dagger - I_r)$ is non-negligible. As λ_i is known for $i \le r$, we only need to consider $i \ge r + 1$. We show in the Appendix that

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it} - \sqrt{T}(H^\dagger - I_r)\lambda_i + o_p(1) \quad (15)$$

and for each t,

$$\sqrt{N}(\hat{F}_t - F_t) = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} + (N/T)^{1/2}\sqrt{T}(H^\dagger - I_r)'F_t + o_p(1) \quad (16)$$

where $\sqrt{T}(H^\dagger - I_r)$ is given in (3).

Theorem 3. *Suppose that Assumptions A, B, and* PC3 *hold. Let $\hat{F}_t$ and $\hat{\lambda}_i$ denote the estimates with the restrictions of* PC3.

(i) *Let $Z_i =^d N(0, \Sigma_F^{-1}\Phi_i\Sigma_F^{-1})$. Then for $i \ge r + 1$, as $N, T \to \infty$ with $\sqrt{T}/N \to 0$,*

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} Z_i - (Z_1, \ldots, Z_r)\lambda_i.$$

(ii) *Let $G_t =^d N(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1})$ and is independent of $(Z_1, \ldots, Z_r)$. If $N/T \to c$ with $0 \le c < \infty$,*

$$\sqrt{N}(\hat{F}_t - F_t) \xrightarrow{d} G_t + \sqrt{c}(Z_1, \ldots, Z_r)'F_t.$$

To understand part (i) of Theorem 3, note that Z_i is the limit of the first term on the right hand side of (15). Under Assumptions A, B, and PC3, the second term in (15) satisfies

$$\sqrt{T}(H^\dagger - I_r) \xrightarrow{d} (Z_1, Z_2, \ldots, Z_r),$$

which is an $r \times r$ matrix of random variables.[6] Although $F'F/T$ (whose limit is Σ_F) is not required to be an identity matrix under PC3, Z_i is normally distributed. As a consequence, $(Z_1, \ldots, Z_r)\lambda_i$ is also normally distributed if λ_i is non-random. It follows that $\hat{\lambda}_i$ is still normally distributed. Similarly, part (ii) of Theorem 3 comes from the fact that G_t is the limiting random variable for the first term on the right hand side of (16). Again, hypothesis testing can be performed similarly as in Section 3.1.

4. Implications for factor-augmented regressions

Consider the infeasible regression model

$$y_t = F_t'\alpha + W_t'\beta + \varepsilon_t$$

where F_t is not observable and is replaced by $\hat{F}_t$ estimated under one of the three identification assumptions. Let $\hat{\delta} = (\hat{\alpha}', \hat{\beta}')'$ denote the least squares estimator of the "factor augmented regression"

$$y_t = \hat{F}_t'\alpha + W_t'\beta + v_t = \hat{z}_t'\delta + v_t \quad (17)$$

where $v_t = \varepsilon_t + (F_t - \hat{F}_t)'\alpha$, $\hat{z}_t = (\hat{F}_t', W_t')'$, and $\delta = (\alpha', \beta')'$. To state the asymptotic behavior of $\hat{\delta}$, we also need the following:

[6] The matrix convergence in distribution implicitly refers to the convergence with vectorization. In any event, $\sqrt{T}(H^\dagger - I_r)\lambda_i$ is already a vector, so its convergence to the vector $(Z_1, \ldots, Z_r)\lambda_i$ is well defined.

Assumption C. For $z_t = (F_t', W_t')'$, $E\|z_t\|^4 \le M < \infty$; $E(\varepsilon_t | z_{t-1}, z_{t-2}, \ldots) = 0$; z_t and ε_t are independent of the idiosyncratic errors e_{is} for all i and s. Furthermore, $\frac{1}{T}\sum_{t=1}^T z_t z_t' \xrightarrow{p} \Sigma_{zz} > 0$ and $T^{-1/2}\sum_{t=1}^T z_t\varepsilon_t \xrightarrow{d} N(0, \Sigma_{zz,\varepsilon})$, where $\Sigma_{zz,\varepsilon} = \text{plim}\frac{1}{T}\sum_{t=1}^T \varepsilon_t^2 z_t z_t' > 0$.

If F_t was observed, then under Assumption C, the asymptotic variance of $\hat{\delta}$ would be given by $\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}$. As shown in Bai and Ng (2006), $\hat{\alpha}$ is an estimate of $H^{-1}\alpha$ (and not α) when $\tilde{F}_t$ is used in place of F_t. The following theorem studies the properties of $\hat{\delta}$ when $\hat{F}_t$ is used in place of F_t.

Theorem 4. *Suppose $\sqrt{T}/N \to 0$ and Assumptions A–C hold. Define $\Sigma_\delta = \Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}$. Let $\delta' = (\alpha', \beta')$ and let $\hat{\delta}$ be obtained by the least squares estimation of factor augmented regression (17), where $\hat{F}_t$ is obtained under the restrictions defined by* PC1, PC2, *or* PC3. *Then*

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \text{Avar}(\hat{\delta}))$$

where $\text{Avar}(\hat{\delta}) = \Sigma_\delta$ *under* PC1, $\text{Avar}(\hat{\delta}) = \Sigma_\delta + \text{diag}[(\alpha' \otimes I_r) D\text{var}(\eta)D'(\alpha \otimes I_r), 0]$ *under* PC2, *and* $\text{Avar}(\hat{\delta}) = \Sigma_\delta + \text{diag}(\text{var}[(Z_1, \ldots, Z_r)\alpha], 0)$ *under* PC3. *Furthermore, η and D are defined in Section 3.2, and $(Z_1, \ldots, Z_r)$ is defined in Section 3.3;* $\text{diag}(A, B)$ *refers to the block diagonal matrix with blocks A and B.*

Theorem 4 states that under PC1, $\hat{\delta}$ has properties as though the latent factors F_t were available as regressors. Although the distribution of $\hat{\beta}$ is invariant to identification assumptions used, the distribution of $\hat{\alpha}$ does depend on whether PC1, PC2, or PC3 is used.

To understand Theorem 4, note that under PC1,

$$\sqrt{T}(\hat{\alpha} - \alpha) = \sqrt{T}(\hat{\alpha} - H^{-1}\alpha) - \sqrt{T}(H - I)H^{-1}\alpha.$$

The first term on the right is analyzed by Bai and Ng (2006). Under PC1, $\sqrt{T}(H - I_r) = o_p(1)$ provided $\sqrt{T}/N \to 0$ since $H - I_r = O_p(\delta_{NT}^{-2})$. As H is asymptotically an identity matrix, $\hat{\alpha}$ now directly estimates α. Thus, the limiting distribution for $\sqrt{T}(\hat{\alpha} - H^{-1}\alpha)$ stated in Bai and Ng (2006) simplifies to the case of standard least squares as if F_t were observed. Under PC1, the asymptotic variance of $\Sigma_{\hat{\delta}}$ can be consistently estimated by

$$\widehat{\Sigma_{\hat{\delta}}} = \left(\frac{1}{T}\sum_{t=1}^T \hat{z}_t\hat{z}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{z}_t\hat{z}_t'\hat{\upsilon}_t^2\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{z}_t\hat{z}_t'\right)^{-1}$$

which is White's heteroskedasticity robust covariance estimator using $\hat{z}_t$ as regressors.

Under PC2 and PC3, $\sqrt{T}(H^* - I_r)$ and $\sqrt{T}(H^\dagger - I_r)$ are not asymptotically negligible when $r > 1$. The asymptotic variance of $\hat{\alpha}$ under PC2 has an extra term given by the variance of $(\alpha \otimes I_r)D\eta$. Under PC3, the extra term in the asymptotic variance of $\hat{\alpha}$ is due to $\text{var}[(Z_1, \ldots, Z_r)\alpha]$. Details on estimation of the asymptotic variances are given in Appendix A. It is however useful to note that if e_{jt} are independent for $j = 1, 2, \ldots, r$, then the normal vectors Z_j are also independent. In such a case, $\text{var}[(Z_1, \ldots, Z_r)\alpha] = \sum_{k=1}^r \Phi_k\alpha_k$ can be consistently estimated by $\sum_{k=1}^r \hat{\Phi}_k\hat{\alpha}_k$.

It is useful to remark that when $\hat{F}_t$ estimates F_t instead of a rotation of F_t, we can give economic interpretation to the coefficients on the regressors $\hat{F}_t$. For example, in factor augmented autoregressions (FAVAR) or for the factor models considered in this paper we can obtain the impulse responses of each observable X_{it} in the panel to the common shocks that drive F_t.[7] Suppose that $F_t = A_1F_{t-1} + \cdots + A_pF_{t-p} + A_0u_t$, where u_t is a vector of structural shocks, and A_0 is a $r \times r$ matrix linking the structural shocks u_t to the reduced form shocks υ_t such that $\upsilon_t = A_0u_t$.[8] Observing F_t (with economic interpretations for each component) allows us to use standard structural VAR analysis to identity A_0 and compute the impulse responses $\frac{\partial F_{t+k}}{\partial u_t}$. It follows that we can compute the impulse responses for the observable variables $\frac{\partial X_{i,t+k}}{\partial u_t} = \lambda_i'\frac{\partial F_{t+k}}{\partial u_t}$ for each i and for all $k \ge 0$.

5. Factor models with deterministic terms

In practice, the data are demeaned and trends are removed before the factors are estimated. Factor models with deterministic terms are of the form

$$X_{it} = \mu_i + \delta_i(t) + \lambda_i'F_t + e_{it}$$

where μ_i is an individual fixed effect and $\delta_i(t)$ is a time effect. When $\delta_i(t) = \delta_t$, the time effects are common. When $\delta_i(t) = \delta_i \cdot t$, we have individual specific linear trends. These treatments of deterministic terms will be analyzed in the next three subsections.

5.1. Individual fixed effects

We first assume that the time effect is absent. The model in vector form is written as

$$X_t = \mu + \Lambda F_t + e_t.$$

The model is observationally equivalent to the following model $X_t = \mu^* + \Lambda F_t^* + e_t$ where $\mu^* = \mu + \Lambda\bar{F}$, and $F_t^* = F_t - \bar{F}$. We impose the restriction $\bar{F} = \frac{1}{T}\sum_{t=1}^T F_t = 0$. Equivalently, with $\iota_T = (1, 1, \ldots, 1)'$, a $T \times 1$ vector, the restriction is

$$\iota_T'F = \sum_{t=1}^T F_t = 0. \tag{FE1}$$

In the absence of fixed effects, the principal components estimator is based on the $T \times T$ data matrix $X'X$, where $X = [X_1, X_2, \ldots, X_T]$. To account for the fixed effects, we need to demean the data. Equivalently, we can estimate μ by $\bar{X} = \frac{1}{T}\sum_{t=1}^T X_t$ and use the residuals to estimate Λ and F. The demeaned data matrix is

$$Z = [X_1 - \bar{X}, \ldots, X_T - \bar{X}] = X - \bar{X}\iota_T'.$$

The principal components of F, denoted $\tilde{F}$, corresponds to the eigenvectors $\left(\text{multiplied by } \sqrt{T}\right)$ of the r largest eigenvalues of the data matrix $Z'Z$. That is,

$$(NT)^{-1}Z'Z\tilde{F} = \tilde{F}\tilde{V} \tag{18}$$

where $\tilde{V}$ is $r \times r$ diagonal matrix consisting of the first r largest eigenvalues, arranged in decreasing order. The factor loading estimator is $\tilde{\Lambda} = Z\tilde{F}/T$. By construction, F and Λ already satisfy PC1, namely, that $\tilde{F}'\tilde{F}/T = I_r$ and $\tilde{\Lambda}'\tilde{\Lambda} =$ diagonal. We now want to show that (i) these estimates also satisfy the constraint (FE1) and (ii) that $\tilde{\lambda}_i$ has the same expression with or without demeaning.

To see (i), first note that $\iota_T'Z' = \iota_T'X' - (\iota_T'\iota_T)\bar{X}' = \iota_T'X' - T\bar{X}'$ which equals zero by the definition of $\bar{X}$. Multiply ι_T' on each side of (18), we have

$$0 = \iota_T'Z'Z = \iota_T'\tilde{F}\tilde{V}.$$

[7] Similar issues have been considered by Stock and Watson (2005) and Forni et al. (2009).

[8] The model is still static even though F_t is dynamic.

Since $\tilde{V}$ is an invertible (diagonal) matrix of eigenvalues, it follows that $\iota_T'\tilde{F} = \sum_{t=1}^T \tilde{F}_t = 0$, which is (FE1). The principal components estimator for Λ can now be rewritten as

$$\tilde{\Lambda} = Z\tilde{F}/T = (X - \bar{X}\iota_T')\tilde{F}/T = X\tilde{F}/T$$

where the last equality makes use of the result $\iota_T'\tilde{F} = 0$. Therefore, the expression for $\tilde{\lambda}_i$ has the same form with or without demeaning the data.

To show (ii) that the limiting distribution for $\tilde{\lambda}$ is of the same form with or without fixed effects note that since $F_t = F_t - \bar{F}$ and $\bar{F} = 0$ by assumption, the model in demeaned data is

$$X_{it} - \bar{X}_i = \lambda_i' F_t + e_{it} - \bar{e}_i.$$

Replacing e_{it} with $e_{it} - \bar{e}_i$ in (8) and since $\sum_{t=1}^T F_t \bar{e}_i = \left(\sum_{t=1}^T F_t\right)\bar{e}_i = 0$,

$$\begin{aligned}\sqrt{N}(\tilde{\lambda}_i - \lambda_i) &= \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t(e_{it} - \bar{e}_i) + o_p(1) \\ &= \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t e_{it} + o_p(1).\end{aligned}$$

This representation coincides with (8). Thus under Assumptions A, B, PC1 and (FE1), the limit is again $\sqrt{N}(\tilde{\lambda}_i - \lambda_i) \sim N(0, \Phi_i)$, which is (10). The limiting distribution for $\tilde{F}_t$ also has the same form with or without demeaning. Replacing e_{it} with $e_{it} - \bar{e}_i$ in (7), we have

$$\begin{aligned}\sqrt{N}(\tilde{F}_t - F_t) &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i(e_{it} - \bar{e}_i) + o_p(1) \\ &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it} - T^{-1/2}\left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \lambda_i e_{it} + o_p(1) \\ &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it} + o_p(1).\end{aligned}$$

The second term on the right hand side is $O_p(T^{-1/2})$ because $(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \lambda_i e_{it} = O_p(1)$. The asymptotic representation for $\tilde{F}_t$ is thus the same as when fixed effects are absent. This implies that the limiting distribution has the same form.

The estimators under identification restrictions PC2 and PC3 are constructed exactly the same way as when fixed effects are absent, but using the newly defined principal components estimators $\tilde{F}$ and $\tilde{\Lambda}$. Thus when (FE1) holds, the expressions for $\tilde{\lambda}_i$ and $\tilde{F}_t$ are the same with or without demeaning.

5.2. Common time effects

We now allow for common time effects.

$$X_{it} = \mu_i + \delta_t + \lambda_i' F_t + e_{it}.$$

For identification, we now need the additional restriction[9]

$$\frac{1}{N}\sum_{i=1}^N \lambda_i = 0. \tag{FE2}$$

To estimate the model, we first remove the cross-section mean and time series mean from the data. Let $\dot{X}_{it} = X_{it} - \bar{X}_{i.} - \bar{X}_{.t} + \bar{X}_{..}$, where $\bar{X}_{i.}$ is time series mean for each i, $\bar{X}_{.t}$ is the cross-section mean for period t, and $\bar{X}_{..}$ is the overall mean of X_{it}. The variable $\dot{X}_{it}$ is the usual within group transformation of X_{it}. By similarly defining $\dot{e}_{it}$, the demeaned model is

$$\dot{X}_{it} = \lambda_i' F_t + \dot{e}_{it}.$$

This is now in the form of a pure factor model without individual and time effects. We can again estimate the model using the data $\dot{X}_{it}$, with any of the three sets of identification restrictions, PC1, PC2, and PC3. There is no need to directly impose the fixed effects restrictions (FE1) and (FE2). When (within-group) transformed data are used, these restrictions are automatically satisfied.

The limiting distributions can again be derived using representation (8) with e_{it} replaced by $\dot{e}_{it} = e_{it} - \bar{e}_{i.} - \bar{e}_{.t} + \bar{e}_{..}$. Specifically,

$$\begin{aligned}&T^{-1/2}\sum_{t=1}^T F_t(e_{it} - \bar{e}_{i.} - \bar{e}_{.t} + \bar{e}_{..}) \\ &= T^{-1/2}\sum_{t=1}^T F_t e_{it} - T^{-1/2}\sum_{t=1}^T F_t \bar{e}_{.t} \\ &= T^{-1/2}\sum_{t=1}^T F_t e_{it} - T^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T F_t e_{it} \\ &= T^{-1/2}\sum_{t=1}^T F_t e_{it} - O_p(T^{-1/2})\end{aligned}$$

where the first equality follows from $\left(\sum_{t=1}^T F_t\right)\bar{e}_{i.} = 0$ and $\left(\sum_{t=1}^T F_t\right)\bar{e}_{..} = 0$ since $\sum_{t=1}^T F_t = 0$. Thus the limiting distribution is still determined by the limit of $(F'F/T)^{-1}T^{-1/2}\sum_{t=1}^T F_t e_{it}$. Similarly,

$$N^{-1/2}\sum_{i=1}^N \lambda_i \dot{e}_{it} = N^{-1/2}\sum_{i=1}^N \lambda_i e_{it} + O_p(N^{-1/2}).$$

It follows that the limiting distribution for the factor loadings is of the same form as when fixed effects are absent. The values of the limiting variances will, however, be general different. If there are no fixed effects in the true model but demeaned data are used in estimation, the resulting estimates for the factors and their loadings will, in general, have larger variances than those without demeaning the data.

To see this, recall that under PC1 or PC2, the estimated factor loadings in the fixed effects model are represented by

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t e_{it} + o_p(1)$$

whether or not the fixed effects are estimated. If $e_{it} \sim (0, \sigma^2)$, F_t is a stationary vector, then the limiting distribution is

$$\sqrt{N}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} N(0, \sigma^2[E(F_tF_t')]^{-1}).$$

Now estimation of the fixed effects will also remove the mean from F_t.[10] Although the representation looks the same, the limiting variance of $\hat{\lambda}_i$ is then $\sigma^2[\text{var}(F_t)]^{-1}$. As F_t can have non-zero mean, the second moment $E(F_tF_t')$ is in general larger than the variance of F_t. As $E(F_tF_t') \geq \text{var}(F_t)$ implies $[E(F_tF_t')]^{-1} \leq [\text{var}(F_t)]^{-1}$, the limiting variance of $\hat{\lambda}_i$ is smaller when fixed effects are known to be absent and are not estimated.

---

9 The restriction may be replaced by $E(\lambda_i) = 0$ if each λ_i is considered to be a vector of random variables.

10 Our assumption that $\bar{F} = 0$ is asymptotically equivalent to $E(F_t) = 0$.

5.3. Heterogeneous trends

Instead of common time effects, consider a model with heterogeneous coefficients on the linear trends:

$$X_{it} = \mu_i + \delta_i t + \lambda_i' F_t + e_{it}.$$

We now assume that F_t is a zero mean process that does not contain a linear trend because in the presence of $\mu_i + \delta_i t$, we cannot separately identify the heterogeneous trends and the factor process. For example, suppose that $F_t = c + dt + \eta_t$, where η_t is a zero mean process, we can rewrite the model as $X_{it} = \mu_i^* + \delta_i^* t + \lambda_i' \eta_t + e_{it}$ with $\mu_i^* = \mu_i + \lambda_i' c$ and $\delta_i^* = \delta_i + \lambda_i' d$. We can only identify η_t.

We focus on the identification restriction PC1, i.e., $F'F/T = I_r$ and $\Lambda'\Lambda$ is diagonal. Let X_{it}^τ denote the residuals from the least squares detrending for each series i. We have

$$X_{it}^\tau = \lambda_i' F_t^\tau + e_{it}^\tau,$$

where F_t^τ and e_{it}^τ are also the residuals from the least squares detrending (no actual detrending is performed on them since they are unobservable). Let a_F and b_F be the OLS coefficients when F_t is regressed on $[1, t]$, and $a_{i,e}$ and $b_{i,e}$ are similarly defined, we have

$$F_t^\tau = F_t - a_F - b_F t$$
$$e_{it}^\tau = e_{it} - a_{i,e} - b_{i,e} t.$$

While F_t^τ is not equal to F_t, one can easily show that $F_t^\tau = F_t + O_p(T^{-1/2})$. Note that $F'F/T = I_r$ implies that $F^{\tau\prime} F^\tau / T = I_r + O_p(1/T)$ because F_t is a zero mean sequence by assumption in this section. Together with diagonality of $\Lambda'\Lambda$ under PC1, we can use earlier arguments to show that

$$\begin{aligned}\sqrt{N}(\tilde{\lambda}_i - \lambda_i) &= \left(\frac{F^{\tau\prime} F^\tau}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t^\tau e_{it}^\tau + o_p(1) \\ &= \left(\frac{F^{\tau\prime} F^\tau}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t^\tau e_{it} + o_p(1).\end{aligned}$$

Note that we can replace e_{it}^τ by e_{it} because $\{F_t^\tau\}$ is orthogonal to the sequence $\{1, t\}$. Similarly,

$$\begin{aligned}\sqrt{N}(\tilde{F}_t - F_t^\tau) &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it}^\tau + o_p(1) \\ &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it} - \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i (a_{i,e} + b_{i,e} t) + o_p(1) \\ &= \left(\frac{\Lambda'\Lambda}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it} + o_p(1).\end{aligned}$$

The last equality follows from the fact that $a_{i,e} + b_{i,e} t$ is a linear combination of $\frac{1}{T}\sum_{s=1}^T e_{is}$ and $\left(\frac{1}{T}\sum_{s=1}^T \frac{s}{T} e_{is}\right)\frac{t}{T}$, each of which is $O_p(T^{-1/2})$. Using the assumption that e_{it} has weak cross-sectional correlation, we can show that $N^{-1/2}\sum_{i=1}^N \lambda_i(a_{i,e} + b_{i,e} t) = O_p(T^{-1/2})$. Asymptotic normality for $\sqrt{T}(\tilde{\lambda}_i - \lambda_i)$ and for $\sqrt{N}(\tilde{F}_t - F_t^\tau)$ follows from the fact that $T^{-1/2}\sum_{t=1}^T F_t^\tau e_{it}$ and $N^{-1/2}\sum_{i=1}^N \lambda_i e_{it}$ are asymptotically normal. Once the data are demeaned and detrended, the estimation procedure is identical to the case with or without linear trends. In addition, the asymptotic variances for $\hat{\lambda}_i$ and $\hat{F}_t$ are estimated as if there were no deterministic terms. Analogous arguments can be used to establish that the limiting distributions under PC2 and PC3 also have the same form as the case without deterministic intercepts or trends. Details are omitted.

Table 1
Marginal R^2: $\hat{F}_t$ rotated under PC2.

	Series	Factor 1	2	3	4	5	6	7	8
1	ces002	0.789	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	ips10	0.564	0.349	0.000	0.000	0.000	0.000	0.000	0.000
3	sfygt1	0.034	0.043	0.794	0.000	0.000	0.000	0.000	0.000
4	puxhs	0.002	0.000	0.000	0.769	0.000	0.000	0.000	0.000
5	fygt1	0.068	0.016	0.007	0.004	0.797	0.000	0.000	0.000
6	hsbr	0.154	0.005	0.006	0.000	0.019	0.739	0.000	0.000
7	fmrra	0.000	0.001	0.000	0.009	0.000	0.001	0.648	0.000
8	fspcom	0.003	0.029	0.020	0.001	0.050	0.000	0.001	0.602

6. An application

Stock and Watson (2005) analyzed 132 series over the sample 1959:1 to 2003:12. The predictors include series in 14 categories: real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories; orders; stock prices; exchange rates; interest rates and spreads; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous. The series are transformed by taking logarithms and/or differencing so that the transformed series are approximately stationary. The IC_1 and IC_2 criteria developed in Bai and Ng (2002) find 7 static factors explaining over 40 percent of the variation in the data.

Stock and Watson (2005) performed variance decompositions and reported that the first factor explains much of the variation in production and employment related series, while the second factor explains movements in interest rates, consumption, and stock prices. Variation in inflation is mainly explained by the second and third factor. Factor four is highly correlated with interest rate movements, factor five with employment, factor six with exchange rates, stock returns, and hourly earnings.

We use the Stock–Watson data extended to 2007:12 and used in Ludvigson and Ng (2011). After deleting a series that is no longer published, the new dataset has 131 series. We first transform the data to be stationary. The demeaned and standardized data are then used to estimate the factors. The first 7 factors still explain 45% of the variation in the data, though the IC_2 criterion now finds the optimal number of factors to be 8.

An important aspect of PC2 is that it uses the ordering of the variables to identify the factors. We reorder the data such that the first eight series are (1) ces002, total employees on non-far payroll; (2) ips10, industrial production total index; (3) sfygt1, spread between one-year T-bill rate (fygt1) and fed funds rate; (4) puxhs, CPI excluding shelter; (5) fygt1, one year T-bill rate; (6) hsbr, housing units authorized; (7) fmrra, total reserves; (8) fspcom, S&P 500 index. Under PC2, employment responds to the first factor only while industrial production responds to the first two factors. The interest rate spread responds to factors one to three, while inflation responds to factors one to four, and so on. This in turn implies that shocks to $\hat{F}_1$ are shocks to employment, while shocks to $\hat{F}_2$ are industrial production shocks orthogonal to employment, and so forth.

Table 1 reports the marginal explanatory power of the j-th factor. The (i, j)th entry of the table is computed as follows. Let $R^2(j)$ be the R^2 in a regression of the series in question on the first j rotated factors. We first regress the ith series on the first j rotated factors to get $R^2(j)$, and then regress the same series on the first $j - 1$ rotated factors to get $R^2(j-1)$. The (i, j)th entry equals the difference between $R^2(j)$ and $R^2(j-1)$. The results conform that under PC2, the first two factors are real activity factors while factor four is inflation. Factors three and five are related to interest rates, while factor seven is a monetary factor. Factor six is a housing factor, and factor 8 is that of the stock market.

It is useful to compare the marginal R^2s obtained by regressing these same series on the standard principal component estimates,

Table 2
Marginal R^2: $\tilde{F}_t$.

	Series	Factor 1	2	3	4	5	6	7	8
1	ces002	0.695	0.005	0.000	0.017	0.050	0.004	0.001	0.016
2	ips10	0.662	0.032	0.002	0.076	0.092	0.001	0.008	0.041
3	sfygt1	0.113	0.385	0.005	0.025	0.162	0.139	0.038	0.004
4	puxhs	0.003	0.028	0.701	0.035	0.000	0.001	0.002	0.000
5	fygt1	0.196	0.144	0.018	0.257	0.242	0.003	0.011	0.022
6	hsbr	0.288	0.005	0.010	0.173	0.188	0.218	0.024	0.017
7	fmrra	0.000	0.001	0.028	0.007	0.001	0.142	0.477	0.003
8	fspcom	0.002	0.170	0.004	0.009	0.027	0.064	0.003	0.426

$\tilde{F}_t$. This is reported in Table 2. The results are in line with what was reported in Stock and Watson (2005) that the first two factors highly correlated with output and employment data. However, the remaining factors load on a variety of other variables.

Using the PC2 rotation, the eight factors are much more concentrated on the variations in eight series which facilitates the interpretation of these factors. This is useful in subsequent factor augmented regressions in which economic interpretation of the coefficients on $\hat{F}$ is warranted.

7. Conclusion

This paper considers principal-components-based estimation of factors and factor loadings. In general, the method does not separately identify the factors and factor loadings but only their rotations. This paper considers identification restrictions under which the latent factors and the loadings are identified so that the estimates are not rotated. Three sets of restrictions are considered. We show that if the underlying factors and factor loadings satisfy the restrictions used in the estimation, then the rotation matrix is asymptotically an identity matrix. Limiting distributions are derived, and the asymptotic covariance matrices are obtained for each case separately. Other restrictions may also be considered and the asymptotic properties of the corresponding estimators may be derived based on similar arguments.

Acknowledgments

We thank Silvia Goncalves, Benoit Perron, and Dalibor Stevanovic for helpful comments. The authors acknowledge financial support from the NSF (SES-0962410, SES-0962431).

Appendix A

This appendix shows how to consistently estimate the asymptotic covariances under PC1–PC3.

PC1. This is straightforward. We estimate Σ_Λ by $\hat{\Sigma}_\Lambda = \tilde{\Lambda}'\tilde{\Lambda}/N$. To estimate Φ_i and Γ_t, we can use one of the three methods given in Bai and Ng (2006). Let $\hat{\Phi}_i$ and $\hat{\Gamma}_t$ denote these estimates. Then $\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}$ is estimated by $\hat{\Sigma}_\Lambda^{-1}\hat{\Gamma}_t\hat{\Sigma}_\Lambda^{-1}$.

PC2. To estimate the asymptotic variance of $\hat{\lambda}_i$, first consider the case when e_{it} are cross-sectionally independent, so that Z_i are independent over i. This implies that Z_i $(i > r)$ is independent of η (the latter depends on $(Z_1, \ldots, Z_r)$). Noting that $(F'F/T) = I_r$ under PC2,

$$\mathrm{Avar}(\hat{\lambda}_i) = \Phi_i + (\lambda_i' \otimes I_r)D\,\mathrm{var}(\eta)\,D'\,(\lambda_i \otimes I_r)$$

which is the sum of the variances of Z_i and of $(\lambda_i' \otimes I_r)D\eta$. To estimate the variance of η, we let $\zeta_t = \mathrm{veck}[F_t(e_{1t}, \ldots, e_{rt})\Lambda_1'^{-1}]$. Then η is the limit of $T^{-1/2}\sum_{t=1}^T \zeta_t$. In the absence of serial correlation in e_{jt} $(j = 1, 2, \ldots, r)$, the variance of η is equal to the probability limit of $\frac{1}{T}\sum_{t=1}^T \zeta_t\zeta_t'$, and is estimated by $\widehat{\mathrm{var}}(\eta) = \frac{1}{T}\sum_{t=1}^T \hat{\zeta}_t\hat{\zeta}_t'$ with $\hat{\zeta}_t = \mathrm{veck}[\hat{F}_t(\hat{e}_{1t}, \ldots, \hat{e}_{rt})\hat{\Lambda}_1'^{-1}]$. With serial correlation in e_{jt}, the variance of η is the limit of $\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T E(\zeta_t\zeta_s')$, and it is estimated by the Newey–West method using the series $\hat{\zeta}_t (t = 1, 2, \ldots, T)$. Given $\widehat{\mathrm{var}}(\eta)$, we estimate $\mathrm{Avar}(\hat{\lambda}_i)$ by

$$\widehat{\mathrm{Avar}}(\hat{\lambda}_i) = \hat{\Phi}_i + (\hat{\lambda}_i' \otimes I_r)D\,\widehat{\mathrm{var}}(\eta)\,D'(\hat{\lambda}_i \otimes I_r)$$

where $\hat{\Phi}_i = \frac{1}{T}\sum_{t=1}^T \hat{F}_t\hat{F}_t'\hat{e}_{it}^2$ in the absence of serial correlation in e_{it}, and $\hat{\Phi}_i$ is constructed by the Newey–West method based on the series $\hat{F}_t\hat{e}_{it}$ in the presence of serial correlation.

If the e_{it}s are cross-sectionally correlated, Z_i can be correlated with η. Especially for the case of $i \le r$, Z_i is correlated with η. To account for this correlation, we let τ_t be the vector that stacks F_te_{it} and ζ_t so τ_t is an $r + r(r-1)/2$ dimensional vector. Then $\sqrt{T}(\hat{\lambda}_i - \lambda_i) = [I_r, -(\lambda_i' \otimes I_r)D]T^{-1/2}\sum_{t=1}^T \tau_t + o_p(1)$. In the absence of serial correlation in e_{it}, we estimate the variance of $T^{-1/2}\sum_{t=1}^T \tau_t$ by $\hat{V}_\tau = \frac{1}{T}\sum_{t=1}^T \hat{\tau}_t\hat{\tau}_t'$; in the presence of serial correlation, $\hat{V}_\tau$ is the Newey–West estimator using the series $\hat{\tau}_t$. Finally,

$$\widehat{\mathrm{Avar}}(\hat{\lambda}_i) = [I_r, -(\hat{\lambda}_i' \otimes I_r)D]\hat{V}_\tau[I_r, -(\hat{\lambda}_i' \otimes I_r)D]'.$$

Consider now estimating the asymptotic variance of $\hat{F}_t$. Whether or not e_{it} are cross sectionally correlated, G_t is independent of η since G_t is obtained by the CLT with the entire cross sections, and η only depends on e_{it} for $i \le r$. Thus

$$\mathrm{Avar}(\hat{F}_t) = \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1} + c(F_t' \otimes I_r)D\,\mathrm{var}(\eta)\,D'(F_t \otimes I_r).$$

It is estimated by

$$\mathrm{Avar}(\hat{F}_t) = \hat{\Sigma}_\Lambda^{-1}\hat{\Gamma}_t\hat{\Sigma}_\Lambda^{-1} + (N/T)(\hat{F}_t' \otimes I_r)D\,\widehat{\mathrm{var}}(\eta)\,D'(\hat{F}_t \otimes I_r)$$

where $\hat{\Sigma}_\Lambda = (\hat{\Lambda}'\hat{\Lambda}/N)$, and $\hat{\Gamma}_t$ is given by any one of the three methods in Bai and Ng (2006) using the series $\hat{\lambda}_i\hat{e}_{it}$ $(i = 1, 2, \ldots, N)$. Furthermore, Our earlier discussion on estimating $\mathrm{var}(\eta)$ does not assume $e_{1t}, \ldots, e_{rt}$ to be uncorrelated, so $\widehat{\mathrm{var}}(\eta)$ given earlier is valid whether or not e_{it} are cross-sectionally correlated.

PC3. We separately discuss whether or not e_{it} is cross-sectionally independent.

Case i: If e_{it} are cross-sectionally independent, then Z_i are independent over i and

$$\mathrm{Avar}(\hat{\lambda}_i) = \Sigma_F^{-1}\left(\Phi_i + \sum_{k=1}^r \Phi_k\lambda_{ik}^2\right)\Sigma_F^{-1}$$

which is the sum of variance of Z_i and that of $(Z_1, \ldots, Z_r)\lambda_i$. Furthermore, as G_t is the limit from the central limit theorem applied to all the cross section units, G_t is independent of $Z_1, \ldots, Z_r$. Thus

$$\mathrm{Avar}(\hat{F}_t) = \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1} + c^2\sum_{r=1}^k \Phi_kF_{tk}^2. \quad (19)$$

An estimate of $\mathrm{Avar}(\hat{F}_t)$ is given by $\hat{\Sigma}_\Lambda^{-1}\hat{\Gamma}_t\hat{\Sigma}_\Lambda^{-1} + (N/T)\sum_{k=1}^r \hat{\Phi}_k\hat{F}_{tk}^2$, and an estimate of $\mathrm{Avar}(\hat{\lambda}_i)$ is $\hat{\Sigma}_F^{-1}(\hat{\Phi}_i + \sum_{l=1}^r \hat{\Phi}_k\hat{\lambda}_{ik}^2)\hat{\Sigma}_F^{-1}$, where $\hat{\Sigma}_F = \hat{F}'\hat{F}/T$, $\hat{\Sigma}_\Lambda = (\hat{\Lambda}'\hat{\Lambda}/N)$, and $\hat{\Gamma}_t$ and $\hat{\Phi}_i$ have the same form as under PC1 and PC2 but using the new $\hat{F}$ and $\hat{\Lambda}$.

Case ii: If e_{it} is cross-sectionally correlated, then combining (15) and (3), we have

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1}(I_r, -I_r)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t \otimes b_{it}\right) + o_p(1)$$

where b_{it} is a 2 by 1 vector with e_{it} as the first element and $(e_{1t}, \ldots, e_{rt})\lambda_i = \sum_{k=1}^r e_{kt}\lambda_{ik}$ as the second element. Thus the limiting covariance is given by

$$\mathrm{Avar}(\hat{\lambda}_i) = \Sigma_F^{-1}(I_r, -I_r)\Psi_i(I_r, -I_r)'\Sigma_F^{-1}$$

where $\Psi_i = \lim \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T E(F_tF_s' \otimes b_{it}b_{is}')$, which specializes to $\Psi_i = \text{plim}\frac{1}{T}\left(\sum_{t=1}^T(F_tF_t' \otimes b_{it}b_{it}')\right)$ in the absence of time series correlation. To estimate Ψ_i, apply the Newey–West estimator to the sequence $\hat{F}_t \otimes \hat{b}_{it}$. The asymptotic variance is estimated by $\widehat{\text{Avar}}(\hat{\lambda}_i) = \hat{\Sigma}_F^{-1}(I_r, -I_r)\hat{\Psi}_i(I_r, -I_r)'\hat{\Sigma}_F^{-1}$.

Although G_t is still independent of $Z_1, \ldots, Z_r$ (because G_t is obtained from averaging the entire cross sections), $Z_1, \ldots, Z_r$ are dependent among themselves. Under PC3 and cross section dependence,

$$\sqrt{N}(H^\dagger - I_r)'F_t = (F_t' \otimes I_r)\text{vec}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^T F_s' \otimes a_s\right) + o_p(1)$$
$$\xrightarrow{d} (F_t' \otimes I_r)\text{vec}[(Z_1, \ldots, Z_r)']$$

where $a_t = (e_{1t}, \ldots, e_{rt})'$. Let

$$\Upsilon = \text{Avar}(\text{vec}[(Z_1, \ldots, Z_r)'])$$
$$= \frac{1}{T}\sum_{s=1}^T\sum_{t=1}^T E[\text{vec}(F_s' \otimes a_s)\text{vec}(F_t \otimes a_t')]$$

which simplifies to $\Upsilon = \frac{1}{T}\sum_{s=1}^T E[\text{vec}(F_s' \otimes a_s)\text{vec}(F_s \otimes a_s')]$ in the absence of time series correlations. Let c be the limit of N/T. The limiting variance of $\sqrt{N}(\hat{F}_t - F_t)$ becomes

$$\text{Avar}(\hat{F}_t) = \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1} + c^2(F_t' \otimes I_r)\Upsilon(F_t \otimes I_r).$$

To estimate Υ, apply the Newey–West estimator to the sequence $\hat{F}_s' \otimes \hat{a}_s$. The asymptotic variance of $\hat{F}_t$ is estimated by $\widehat{\text{Avar}}(\hat{F}_t) = \hat{\Sigma}_\Lambda^{-1}\hat{\Gamma}_t\hat{\Sigma}_\Lambda^{-1} + (N/T)^2(\hat{F}_t' \otimes I_r)\hat{\Upsilon}(\hat{F}_t \otimes I_r)$.

Appendix B

Proof of (2). Rewrite

$$\tilde{F}'F/T = (\tilde{F} - FH)'F/T + H'F'F/T$$
$$= H'F'F/T + O_p(\delta_{NT}^{-2}) \tag{20}$$

because $(\tilde{F} - FH)'F/T = O_p(\delta_{NT}^{-2})$, see Lemma B.2 of Bai (2003). Right multiply H to both sides,

$$\tilde{F}'FH/T = H'(F'F/T)H + O_p(\delta_{NT}^{-2}).$$

Rewrite the left hand side of above as

$$\tilde{F}'FH/T = \tilde{F}'(FH - \tilde{F} + \tilde{F})/T = O_p(\delta_{NT}^{-2}) + I_r$$

because $\tilde{F}'(FH - \tilde{F})/T = O_p(\delta_{NT}^{-2})$ and $\tilde{F}'\tilde{F}/T = I_r$, see Lemma B.3 of Bai (2003). Equating the above two equations we obtain

$$I_r = H'(F'F/T)H + O_p(\delta_{NT}^{-2}). \tag{21}$$

Thus if $(F'F/T) = I_r$, we have

$$I_r = H'H + O_p(\delta_{NT}^{-2}). \tag{22}$$

Ignore the $O_p(\delta_{NT}^{-2})$ term, the above shows that H is an orthogonal matrix so that its eigenvalues are either 1 or -1. We need to show that H is a diagonal matrix. From the definition of H

$$H' = \tilde{V}^{-1}(\tilde{F}'F/T)(\Lambda'\Lambda/N) = \tilde{V}^{-1}H'(\Lambda'\Lambda/N) + O_p(\delta_{NT}^{-2})$$

where we use the fact that $\tilde{F}'F/T = H' + O_p(\delta_{NT}^{-2})$ under $F'F/T = I_r$, see (20). Multiplying $\tilde{V}$ on both sides and taking the transpose

$$(\Lambda'\Lambda/N)H = H\tilde{V} + O_p(\delta_{NT}^{-2}). \tag{23}$$

This equation implies that H (up to a negligible term) is a matrix consisting of eigenvectors of $(\Lambda'\Lambda/N)$. The latter matrix is diagonal and has distinct eigenvalues by assumption. Thus, each eigenvalue is associated with a unique unitary eigenvector (up to a sign change) and each eigenvector has a single non-zero element. This implies that H is a diagonal matrix up to an $O_p(\delta_{NT}^{-2})$ order. It is already known that the eigenvalues of H are 1 or -1, H is a diagonal matrix with elements of 1 or -1 as its elements. Without loss of generality, we can assume all elements are 1 (otherwise multiply the corresponding columns of $\tilde{F}$ and $\tilde{\Lambda}$ by -1). This implies $H = I_r + O_p(\delta_{NT}^{-2})$. Moreover, from (23) we obtain

$$(\Lambda'\Lambda/N) = \tilde{V} + O_p(\delta_{NT}^{-2}). \quad \square$$

Proof of Theorem 1. Result (2) leads to representations (7) and (8). The theorem is a direct consequence of these representations and Assumption A. $\square$

Proof of (14). Note $H^* = HQ$ is the rotation matrix under PC2. Under PC2, $F'F/T = I_r$, thus (22) holds. This implies that H is an orthogonal matrix, up to a negligible term, and so is HQ since Q is also orthogonal. Furthermore, left multiply (22) by Q' and right multiply it by Q, and use $Q'Q = I_r$, we have

$$I_r = Q'H'HQ + O_p(\delta_{NT}^{-2}). \tag{24}$$

We next show HQ is a diagonal matrix, up to an $O_p(T^{-1/2})$ term. By (5), for each i, $\tilde{\lambda}_i - H^{-1}\lambda_i = O_p(T^{-1/2})$, we have

$$\tilde{\Lambda}_1' = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_r) = H^{-1}(\lambda_1, \ldots, \lambda_r) + O_p(T^{-1/2}).$$

That is, $\tilde{\Lambda}_1' = H^{-1}\Lambda_1' + O_p(T^{-1/2})$. By the QR decomposition, we have $QR = \tilde{\Lambda}_1' = H^{-1}\Lambda_1' + O_p(T^{-1/2})$. Since Λ_1' is also an upper triangular matrix (an assumption of PC2) and H^{-1} is an orthogonal matrix up to a negligible term, by the uniqueness of the QR decomposition, we have $Q = H^{-1} + O_p(T^{-1/2})$. Right multiply H on each side we have $HQ = I_r + O_p(T^{-1/2})$. When $r = 1$, HQ is a scalar, and combined with (24), we strengthen the rate to $HQ = I_r + O_p(\delta_{NT}^{-2})$. For general $r > 1$, the rate cannot be improved. Let $\Delta = HQ - I_r = O_p(T^{-1/2})$. Eq. (24) implies $(\Delta + I_r)'(\Delta + I_r) = O_p(\delta_{NT}^{-2})$. That is, $\Delta'\Delta + \Delta' + \Delta = O_p(\delta_{NT}^{-2})$. But $\Delta'\Delta = O_p(1/T)$, so $\Delta' + \Delta = O_p(\delta_{NT}^{-2})$. This implies that the diagonal elements of Δ are all $O_p(\delta_{NT}^{-2})$ and Δ is skew-symmetric up to an $O_p(\delta_{NT}^{-2})$ term (and especially for $r = 1$, $\Delta = O_p(\delta_{NT}^{-2})$).

We next derive the asymptotic representation for Δ. Using (5), we can write

$$\tilde{\Lambda}_1' - H^{-1}\Lambda_1' = H'\frac{1}{T}\sum_{t=1}^T F_t(e_{1t}, \ldots, e_{rt}) + o_p(T^{-1/2}).$$

Left multiplying H and using $HH' = I_r + O_p(\delta_{NT}^{-2}) = (F'F/T)^{-1} + O_p(\delta_{NT}^{-2})$ [see (22), which still holds under PC2], we have

$$H\tilde{\Lambda}_1' - \Lambda_1' = \left(\frac{F'F}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^T F_t(e_{1t}, \ldots, e_{rt}) + o_p(T^{-1/2}).$$

The first term on the right hand side is $T^{-1/2}\xi_T$, where ξ_T given in (11), so that

$$H\tilde{\Lambda}_1' - \Lambda_1' = T^{-1/2}\xi_T + o_p(T^{-1/2}).$$

By the QR decomposition of $\tilde{\Lambda}_1'$, $H\tilde{\Lambda}_1' = HQR = (HQ - I)R + R = \Delta R + R$. Thus $H\tilde{\Lambda}_1' - \Lambda_1' = \Delta R + (R - \Lambda_1')$. It follows that

$$\Delta = -(R - \Lambda_1')R^{-1} + T^{-1/2}\xi_T R^{-1} + o_p(T^{-1/2}).$$

Since both R and Λ_1' are upper triangular matrices, the below diagonal elements of Δ are equal to the corresponding elements of $T^{-1/2}\xi_T R^{-1} + o_p(T^{-1/2})$. Since Δ is skew-symmetric up to an $O_p(\delta_{NT}^{-2})$ order, the elements of Δ above the diagonal are also given.

That is, $\Delta_{ij} = T^{-1/2}(\xi_T R^{-1})_{ij} + o_p(T^{-1/2})$ for $i > j$, and $\Delta_{ij} = -\Delta_{ji} + O_p(\delta_{NT}^{-2})$ for $i < j$, and $\Delta_{ii} = O_p(\delta_{NT}^{-2})$ $(i, j = 1, 2, \ldots, r)$. Furthermore, we can replace R by Λ_1'. To see this, by the uniqueness of QR decomposition, $R = \Lambda_1' + o_p(1)$. So $T^{-1/2}\xi_T R^{-1} = T^{-1/2}\xi_T(\Lambda_1')^{-1} + T^{-1/2}\xi_T o_p(1) = T^{-1/2}\xi_T(\Lambda_1')^{-1} + o_p(T^{-1/2})$. Finally, (14) is obtained by noting $\Xi = \sqrt{T}\Delta$. □

Proof of (12). Using $\hat{\lambda}_i = Q'\tilde{\lambda}_i$,

$$\hat{\lambda}_i - \lambda_i = Q'\tilde{\lambda}_i - \lambda_i = Q'(\tilde{\lambda}_i - H^{-1}\lambda_i) + Q'H^{-1}(I - HQ)\lambda_i.$$

Multiplying $\sqrt{T}$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = Q'\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i) - Q'H^{-1}\sqrt{T}(HQ - I_r)\lambda_i.$$

Since $Q'H^{-1} = I_r + o_p(1)$, the second term on the right hand side is $-\sqrt{T}(H^* - I_r)\lambda_i + o_p(1)$. Using (5), the first term on the right hand side is $Q'H'T^{-1/2}\sum_{t=1}^T F_t e_{it} + o_p(1)$. But $Q'H' = I_r + o_p(1) = (F'F/T)^{-1} + o_p(1)$ under PC2. Combining the results yield (12). This argument holds for all $i = 1, 2, \ldots, N$. □

Proof of (13). Using $\hat{F}_t = Q'\tilde{F}_t$,

$$\hat{F}_t - F_t = Q'\tilde{F}_t - F_t = Q'(\tilde{F}_t - H'F_t) + (Q'H' - I_r)F_t.$$

Multiplying $\sqrt{N}$,

$$\sqrt{N}(\hat{F}_t - F_t) = Q'\sqrt{N}(\tilde{F}_t - H'F_t) + (N/T)^{1/2}\sqrt{T}(Q'H' - I_r)F_t.$$

From (6), the first term on the right is $Q'H'(\Lambda'\Lambda/N)^{-1}\sum_{i=1}^N \lambda_i e_{it} + o_p(1)$; but $Q'H' = I_r + o_p(1)$. For the second term on the right, $\sqrt{T}(Q'H' - I_r)F_t = -\sqrt{T}(HQ - I_r)F_t + o_p(1)$ because $\sqrt{T}(HQ - I_r)$ is skew-symmetric up to an $o_p(1)$ term when $\sqrt{T}/N \to 0$. Combining results we obtain (13). □

Proof of Theorem 2. This is a direct consequence of (14), (12), (13), Assumptions A and B. □

Proof of (3). Note $H^\dagger = H\tilde{\Lambda}_1'$ is the rotation matrix under PC3. Since the principal components estimator satisfies $\tilde{\lambda}_i - H^{-1}\lambda_i = O_p(T^{-1/2})$, we have

$$\tilde{\Lambda}_1' = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_r) = H^{-1}(\lambda_1, \ldots, \lambda_r) + O_p(T^{-1/2}).$$

Left multiply H to obtain $H\tilde{\Lambda}_1' = I_r + O_p(T^{-1/2})$ because $(\lambda_1, \ldots, \lambda_r) = I_r$ under PC3. That is, $H^\dagger = I_r + O_p(T^{-1/2})$ so $H^\dagger \xrightarrow{p} I_r$. Using representation (5), we have

$$\sqrt{T}(H^\dagger - I_r) = HH'\frac{1}{\sqrt{T}}\sum_{t=1}^T(F_t e_{1t}, \ldots, F_t e_{rt}) + o_p(1).$$

However, (21) implies $HH' = (F'F/T)^{-1} + O_p(\delta_{NT}^{-2})$. This proves (3). □

Proof of (15). Recall that

$$\begin{aligned}\hat{\lambda}_i - \lambda_i &= \tilde{\Lambda}_1'^{-1}\tilde{\lambda}_i - \lambda_i = \tilde{\Lambda}_1'^{-1}(\tilde{\lambda}_i - H^{-1}\lambda_i)\\ &\quad + (\tilde{\Lambda}_1'^{-1}H^{-1} - I_r)\lambda_i.\end{aligned}$$

Multiply $\sqrt{T}$ on each side

$$\begin{aligned}\sqrt{T}(\hat{\lambda}_i - \lambda_i) &= \tilde{\Lambda}_1'^{-1}\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i)\\ &\quad + \tilde{\Lambda}_1'^{-1}H^{-1}\sqrt{T}(I_r - H^\dagger)\lambda_i.\end{aligned}$$

For the first term on the right hand side, using (5),

$$\tilde{\Lambda}_1^{-1}\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i) = (\tilde{\Lambda}_1'^{-1}H^{-1})(HH')\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t e_{it} + o_p(1).$$

Since $H\tilde{\Lambda}' = I_r + o_p(1)$ its inverse is also $I_r + o_p(1)$. Furthermore, as argued earlier, $HH' = (F'F/T)^{-1} + O_p(\delta_{NT}^{-2})$. Thus

$$\tilde{\Lambda}_1^{-1}\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i) = \left(\frac{F'F}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t e_{it} + o_p(1).$$

The second term on the right hand side equals $\sqrt{T}(I_r - H^\dagger)\lambda_i + o_p(1)$. This proves (15). □

Proof of (16). First note that $\hat{F}_t - F_t = \tilde{\Lambda}_1\tilde{F}_t - F_t = \tilde{\Lambda}_1(\tilde{F}_t - H'F_t) + \tilde{\Lambda}_1 H'F_t - F_t = \tilde{\Lambda}_1(\tilde{F}_t - H'F_t) + (H'^\dagger - I_r)F_t$. It follows that

$$\sqrt{N}(\hat{F}_t - F_t) = \tilde{\Lambda}_1\sqrt{N}(\tilde{F}_t - H'F_t) + (N/T)^{1/2}\sqrt{T}(H^\dagger - I_r)'F_t.$$

From (6), and using $\tilde{\Lambda}_1 H' = I_r + o_p(1)$, the first term on the right hand side is

$$\tilde{\Lambda}_1\sqrt{N}(\tilde{F}_t - H'F_t) = \left(\frac{\Lambda'\Lambda}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{it} + o_p(1).$$

Combining the two equations leads to (16). □

Proof of Theorem 3. This follows from (3), (15), (16), and Assumptions A and B. □

Proof of Theorem 4. We first consider the case of identification under PC1 so that we use $\tilde{F}_t$ in place of F_t in the regression model. We can rewrite the model as in Bai and Ng (2006)

$$\begin{aligned}y_t &= \begin{pmatrix}\tilde{F}_t' & W_t'\end{pmatrix}\begin{pmatrix}H^{-1}\alpha\\ \beta\end{pmatrix} + \varepsilon_t + (F_t'H - \tilde{F}')H^{-1}\alpha\\ &= \hat{z}_t'\delta^* + \varepsilon_t + a_t\end{aligned}$$

where $\hat{z}_t' = (\tilde{F}_t', W_t')'$, $\delta^* = (\alpha'H^{-'}, \beta')'$, and a_t represents the last term on the right hand side. When $\sqrt{T}/N \to 0$, Bai and Ng (2006) shows that the error a_t is negligible, and the least squares estimator $\hat{\delta}$ has the standard limiting distribution as if $\tilde{F}_t$ contains no estimation error (as if $H'F_t$ were observable). More specifically,

$$\sqrt{T}(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \Phi_0^{-'}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}\Phi_0^{-1})$$

where $\Phi_0 = \text{diag}(V^{-1}Q\Sigma_\Lambda, I)$ and $V^{-1}Q\Sigma_\Lambda$ is the probability limit of H, where Q represents the probability limit of $\tilde{F}'F/T$. In our case, the limit of H is an identity matrix (also follows from $Q = I_r$ and $V = \Sigma_\Lambda$ in the present case) so that Φ_0 is an identity matrix. This implies that

$$\sqrt{T}(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \Sigma_\delta)$$

where $\Sigma_\delta = \Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}$. Furthermore,

$$\sqrt{T}(\hat{\delta} - \delta) = \sqrt{T}(\hat{\delta} - \delta^*) + \sqrt{T}[(\alpha - H^{-1}\alpha)', 0']'.$$

But $\sqrt{T}(\alpha - H^{-1}\alpha) = \sqrt{T}(H - I_r)H^{-1}\alpha = o_p(1)$ provided that $\sqrt{T}/N \to 0$ because $H - I_r = O_p(\delta_{NT}^{-2})$. It follows that under $\sqrt{T}/N \to 0$, $\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta)$.

We next consider PC3. We use $\hat{F}_t$ in place of F_t, where $\hat{F}_t$ is defined in the main text. Since $\hat{F}_t$ is an estimate of $H^{\dagger\prime}F_t$, we define $\delta^\dagger = [(H^{\dagger-1}\alpha)', \beta']'$. Then $y_t = \hat{z}_t'\delta^\dagger + \varepsilon_t + a_t^\dagger$, here $a_t^\dagger = (F_t'H^\dagger - \hat{F}')H^{\dagger-1}\alpha$. The same argument in Bai and Ng (2006) leads to

$$\sqrt{T}(\hat{\delta}-\delta^{\dagger}) \xrightarrow{d} N(0, \Phi_0^{-\prime}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}\Phi_0^{-1})$$

where $\Phi_0 = \text{diag}(\text{plim}H^{\dagger}, I)$. Under PC3, $\text{plim}H^{\dagger} = I_r$. Thus, $\sqrt{T}(\hat{\delta}-\delta^{\dagger}) \xrightarrow{d} N(0, \Sigma_\delta)$, where Σ_δ is defined earlier. Next,

$$\sqrt{T}(\hat{\delta}-\delta) = \sqrt{T}(\hat{\delta}-\delta^{\dagger}) + \sqrt{T}[(\alpha - H^{\dagger-1}\alpha)', 0']'.$$

But the term

$$\sqrt{T}(\alpha - H^{\dagger-1}\alpha) = \sqrt{T}(H^{\dagger} - I_r)H^{\dagger-1}\alpha$$

is not negligible and $\sqrt{T}(H^{\dagger} - I_r) \xrightarrow{d} (Z_1, \ldots, Z_r)$ and $H^{\dagger-1}\alpha = \alpha + o_p(1)$. It follows that

$$\sqrt{T}(\hat{\delta}-\delta) \xrightarrow{d} N(0, \Sigma_\delta) + \begin{bmatrix} (Z_1, \ldots, Z_r)\alpha \\ 0 \end{bmatrix}.$$

Since the normal random variable $N(0, \Sigma_\delta)$ is derived from the central limit theorem (CLT) involving $\{\varepsilon_t\}$, while $(Z_1, \ldots, Z_r)$ are derived from the CLT involving $\{e_{it}\}$, these normal variables are independent of each other under Assumption C. Therefore, the asymptotic variance of $\hat{\delta}$ is equal to $\Sigma_\delta + \text{diag}(\text{var}[(Z_1, \ldots, Z_r)\alpha], 0)$, where diag means block-diagonal. Under the assumption that e_{jt} are independent over $j = 1, 2, \ldots, r$, then $Z_1, \ldots, Z_r$ are also independent so that $\text{var}[(Z_1, \ldots, Z_r)\alpha] = \sum_{k=1}^{r}\Phi_k\alpha_k$. For dependent e_{jt} over j, $(Z_1, \ldots, Z_r)\alpha = (\alpha' \otimes I_r)\text{vec}(Z_1, \ldots, Z_r)$. Consistent estimation of $\text{var}(\text{vec}(Z_1, \ldots, Z_r))$ is discussed in Appendix A.

Finally consider PC2. Define $\delta^* = [(H^{*-1}\alpha)', \beta']'$. The same analysis as in PC3 gives

$$\sqrt{T}(\hat{\delta}-\delta) = \sqrt{T}(\hat{\delta}-\delta^*) + \sqrt{T}[(\alpha - H^{*-1}\alpha)', 0']'$$

with $\sqrt{T}(\hat{\delta}-\delta^*) \xrightarrow{d} N(0, \Sigma_\delta)$. Furthermore, $\sqrt{T}(\alpha - H^{*-1}\alpha) = \sqrt{T}(H^* - I_r)H^{*-1}\alpha = \sqrt{T}(H^* - I_r)\alpha + o_p(1) = (\alpha' \otimes I_r)D\ \text{veck}(\xi_T \Lambda_1^{\prime-1}) + o_p(1)$, which converges in distribution to $(\alpha' \otimes I_r)D\eta$. Thus the asymptotic variance of $\hat{\delta}$ is equal to $\Sigma_\delta + \text{diag}[(\alpha' \otimes I_r)D\ \text{var}(\eta)D'(\alpha \otimes I_r), 0]$, where diag means block-diagonal. □

References

Algina, J., 1980. A note on identification in the oblique and orthogonal factor analysis models. Psychometrika 45 (3), 393–395.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–172.

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

Bai, J., Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions. Econometrica 74 (4), 1133–1150.

Bai, J., Ng, S., 2010. Instrumental variables estimation in a data rich environment. Econometric Theory 26 (6), 1607–1637.

Bekker, P., 1986. A note on the identification of restricted factor loading matrices. Psychometrika 51, 607–611.

Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: identifying shocks and propagation mechanisms in VAR and factor models. Econometric Theory 25, 1319–1347.

Hallin, M., Liska, R., 2008. Dynamic factors in the presence of block structure. European University Institute WP 2008/22.

Heaton, C., Solo, V., 2004. Identification of causal factor models of stationary time series. Econometrics Journal 7, 618–627.

Lawley, D.N., Maxwell, A.E., 1971. Factor Analysis in a Statistical Method. Butterworth, London.

Ludvigson, S., Ng, S., 2011. A factor analysis of bond risk premia. In: Gilles, D., Ullah, A. (Eds.), Handbook of Empirical Economics and Finance. Chapman and Hall, pp. 313–372.

Moench, E., Ng, S., 2011. A factor analysis of housing market dynamics in the US and the regions. Econometrics Journal 14, C1–C24.

Pantula, S., Fuller, W., 1986. Computational algorithms for the factor model. Communications in Statistics. Simulation and Computation 15, 227–259.

Reis, R., Watson, M., 2010. Relative goods' prices, pure inflation, and the Phillips correlation. American Economic Journal: Macroeconomics 2 (3), 128–157.

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

Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. NBER WP 11467.

Wang, P., 2008. Large dimensional factor models with a multi-level factor structure: identification, estimation and inference. Unpublished manuscript, Department of Economics, New York University.

Wansbeek, T., Meijer, E., 2000. Measurement Error and Latent Variables in Econometrics. Elsevier, Amsterdam.