Principal component 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

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.

© 2013 Elsevier B.V. All rights reserved.

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 \( \mathbf{F}_t \) is the \( r \times 1 \) vector of latent factors, and \( \hat{\mathbf{F}}_t \) is the vector of factor estimates, there exists an \( r \times r \) invertible matrix \( H \) such that \( \hat{\mathbf{F}}_t \) estimates \( H^t \mathbf{F}_t \). Asymptotic results are stated in terms of \( \mathbf{F}_t - H^t \mathbf{F}_t \). Similarly, if \( \lambda_i \) is the vector of factor loadings and \( \hat{\lambda}_i \) is the corresponding estimate, asymptotic results are known for \( \hat{\lambda}_i - H^{-1} \lambda_i \).

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 \( \mathbf{F}_t \) 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 \( \mathbf{y}_t = \mathbf{\alpha} \hat{\mathbf{F}}_t + \mathbf{W}_t \mathbf{\beta} + \mathbf{\epsilon}_t \), one might be interested in testing a hypothesis concerning \( \mathbf{\alpha} \). Since the asymptotic theory is only available for \( \sqrt{T}(\hat{\alpha} - H^{-1} \alpha) \), the test is uninformative except when \( \alpha \) is zero. If \( H \) is known, \( \hat{\alpha} \) can be given economic interpretation.

We study three sets of restrictions such that \( F \) and \( \Lambda \) are exactly identified. If the underlying \( F \) and \( \Lambda \) that generate the data satisfy those restrictions then \( H \) is asymptotically an identity matrix. This is useful because \( \hat{\mathbf{F}}_t \) can be treated as though they were the latent \( \mathbf{F}_t \) and \( \alpha \) 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 \( \Lambda \) 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 \( \Lambda \). Other identification conditions may be considered; the method developed in this paper should be useful to derive the corresponding limiting distributions.
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, ..., N and t = 1, ..., T, the observation $X_{it}$ has a factor structure represented as $X_{it} = \Lambda F_t + \epsilon_{it}$.

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 $X = F' + e$

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

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

(a) $E|F|^2 \leq M$ and $\frac{1}{T} \sum_{t=1}^T F_tF_t' \frac{d}{\Sigma_e} > 0$ is a $r \times r$ non-random matrix.

(b) $\lambda_i$ is either deterministic such that $|\lambda_i| \leq M$, or it is stochastic such that $E|\lambda_i|^2 \leq M$. In either case, $N^{-1} \Lambda' \Lambda \frac{d}{\Sigma_e} \Lambda' > 0$ is a $r \times r$ non-random matrix as $N \rightarrow \infty$.

(c) $E(e_{it}^2) = 0$ and $E(e_{it})^T \leq M$.

(d) $E(e_{it}e_{js}) = \delta_{ij} \tau_t$ for all $(t, s)$ and $|\tau_t| \leq \tau_0$ for all $t$.

(e) For each $t$, $E[N^{-1/2} \sum_{i=1}^N |e_{it} - \delta_{ij}|^4] \leq M$.

(f) $[\Lambda', [F_t]],$ and $[e_t]$, are three mutually independent groups.

(g) $N^{-1/2} \sum_{i=1}^N \lambda_i e_{it} \xrightarrow{d} N(0, \delta_0^2)$. 

Assumptions A(a) and (b) imply the existence of $r$ factors. The idiosyncratic errors $\epsilon_{it}$ can be allowed to be cross-sectionally and serially correlated, but only weakly as stated under condition A(c). If $\epsilon_{it}$ 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, $\lambda_i$ can be correlated over $i$, and $\epsilon_{it}$ 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 $\text{tr}(X' - \Lambda' \Lambda)'(X - \Lambda' \Lambda)$ by choosing the normalizations that $F'F/T = I_r$ and $\Lambda' \Lambda$ is diagonal. The estimator for $F$, denoted $\hat{F} = (\hat{F}_1, \ldots, \hat{F}_T)'$, 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 $\Lambda = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)' = XX'F'F/T$ is a $N \times r$ matrix of estimated factor loadings. The estimators $\hat{F}$ and $\hat{\Lambda}$ satisfy the normalization restrictions since $\hat{F}'F/T = I_r$ holds by construction and $\hat{\Lambda}'\Lambda/N = V$ where $V$ 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 $\Lambda$, they do not necessarily identify the individual columns of $F$ or of $\Lambda$. To be precise, let $H$ be an $r \times r$ matrix whose transpose is $H' = \hat{V}^{-1}(\hat{V}'\hat{F}/T)(\Lambda' \Lambda/N)$.

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

We are specifically interested in conditions under which we can identify the columns of $F$ and the columns of $\Lambda$ from the product $F^* \Lambda^*$. Notice that $F^* \Lambda^* = FRR^{-1}A^*$ 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 identify $F$ and $\Lambda$, 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 $\Lambda$ satisfy these restrictions, then the corresponding rotation matrix is asymptotically an identity matrix.\footnote{By symmetry, three different sets of identification restrictions can be obtained by switching $F$ and $\Lambda$. For example, $HF$ is diagonal and $\hat{F}^* \hat{\Lambda}^*$ = $I_r$. Since the asymptotic results still hold by switching the role of $F$ and $\Lambda$, we only consider the three sets of restrictions given above.}

<table>
<thead>
<tr>
<th>Identifying restrictions:</th>
<th>Restrictions on $F$</th>
<th>Restrictions on $\Lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.1): PC1</td>
<td>$\frac{1}{T}F'F = I_r$</td>
<td>$\Lambda' \Lambda$ is a diagonal matrix with distinct entries</td>
</tr>
<tr>
<td>(2.2): PC2</td>
<td>$\frac{1}{T}F'F = I_r$</td>
<td>$\Lambda = (A_{11}, \ldots, A_{1r})$, $A_{ij} = 0$ for $i = 1, \ldots, r$</td>
</tr>
<tr>
<td>(2.3): PC3</td>
<td>Unrestricted</td>
<td>$\Lambda = (A_{11}, \ldots, A_{1r})$, $A_{ij} = 0$ for $i = 1, \ldots, r$</td>
</tr>
</tbody>
</table>

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 identify the individual columns of $F$ and those of $\Lambda$, and there will be rotational indeterminacy. Under PC1, the normalization on $F$ gives $r(t + 1)/2$ restrictions since a symmetric matrix contains $r(t + 1)/2$ free parameters. The diagonality of $\Lambda' \Lambda$ gives $r(t + 1)/2$ 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 $\Lambda$ that generate the data, then

$$H = I_r + O_p(\delta_N^{-2}),$$

where $\delta_N$ denotes $\lim \sqrt{N}, \sqrt{T}$ 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.\footnote{An extension of this model is the inclusion of a global factor, see for example, Moonch 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.} For example, with
$r = 3$, 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 $\pi_i$ is a vector of $N_1 \times 1$ 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 $\Lambda_1$ 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 $X$ and their ordering provide the auxiliary information for identification.

Given the unrestricted estimates $\hat{F}$ and $\hat{\Lambda}$, it is easy to obtain estimators satisfying PC2. Let $\hat{F}_1$ be the first $r \times r$ block of $\hat{F}$ and let $\hat{\Lambda}$ denote the estimators that satisfy PC2, i.e., $\hat{\Lambda} \hat{F}/T = I$ 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 $\hat{\Lambda}_1$, denoted $\hat{\Lambda}_1 = Q \cdot R$, where $R$ is an upper triangular matrix with positive diagonal elements, and $Q$ is an $r \times r$ orthogonal matrix such that $Q \cdot Q^T = I$. This decomposition is unique for any invertible $\Lambda_1$.

- **Step 2:** Define

$$
\hat{F} = \hat{F} \cdot Q, \quad \hat{\Lambda} = \hat{\Lambda} \cdot Q = \begin{bmatrix} R^T & \hat{\Lambda}_2 \end{bmatrix}.
$$

By construction, $\hat{F} \hat{F}^T / Q^T (\hat{F} \hat{F}^T / Q) Q = Q \cdot Q^T = I$. The new rotation matrix is $H^* = HQ$.

Since $\hat{F}$ and $\hat{\Lambda}$ are rotations of the principal component estimates $\hat{F}$ and $\hat{\Lambda}$, 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 $\sqrt{\hat{T}}(H^* - I)$ is asymptotically non-negligible unless $r = 1$. More specifically, if the true $F$ and $\Lambda$ satisfy PC2, then

$$
\text{Hessian} = O_p(T^{-1/2}), \quad r = 1
$$

$$
\text{Hessian} = O_p(T^{-1}), \quad r > 1.
$$

This implies that $\sqrt{\hat{T}}(H^* - I) = O_p(1)$ for $r = 1$. In fact, when $r = 1$, PC1 and PC2 are identical and (2) is in agreement with $\text{Hessian} = O_p(1)$. However, $\sqrt{\hat{T}}(H^* - I) = O_p(1)$ when $r > 1$. In fact, the limit of $\text{Hessian}$ is a skew-symmetric random matrix.\(^4\) In consequence, the limiting distributions of $\hat{F}_1$ and $\hat{\Lambda}_1$ will be affected by $\text{Hessian}$.

2.3. PC3

The third set of identification restrictions specify the first $r \times r$ block of $\Lambda$ (denoted $\Lambda_1$) to be an identity matrix and leaves the factor process $F$ completely unrestricted other than requiring $F^T F$ to be invertible so that $r$ factors exist. Unlike PC1 and PC2, all $r^2$ restrictions are imposed on $\Lambda$ under PC3. The restrictions imply that the first variable $X_1$ is affected by the first factor only, the second variable $X_2$ is affected by the second factor only, etc. The resulting structure resembles the classical ‘errors-in-variables’ model in which $X_k = F_k + e_k$ 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 $\Lambda$ and $F$ are easy to obtain. Given the principal components estimates $\hat{\Lambda}$ and $\hat{F}$, let

$$
\hat{\Lambda} = \hat{\Lambda} \hat{\Lambda}^{-1}, \quad \hat{F} = \hat{F} \hat{\Lambda}^{-1}.
$$

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

$$
\sqrt{\hat{T}}(H^* - I) = \xi_t + o_p(1)
$$

where $\xi_t$ is defined in (11) below. The fact that $\sqrt{\hat{T}}(H^* - I)$ is negligible for all $r \geq 1$ will affect the limiting distributions of $\hat{\Lambda}_1$ and $\hat{F}_1$.

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 identify $F$ and $\Lambda$ up to a column sign change. Changing the sign of any column of $F$ and the sign of the corresponding column of $\Lambda$ will leave the product $FA'$ unchanged.

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

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

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

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

$$
\begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ - & \lambda_{22} & \lambda_{23} \\ - & - & \lambda_{33} \end{bmatrix}.
$$

If the sign of $\lambda_{11}$ is known, then $\lambda_{12}$ is identified from $\lambda_{22}$. Since $\lambda_{11} \neq 0$, $\lambda_{22}$ and $\lambda_{33}$ can be identified, which further implies the identification of $\lambda_{33}$. If the sign of $\lambda_{22}$ is known, then $\lambda_{23}$ is also identified. Since $\lambda_{22} \neq 0$, this implies the identification of $\lambda_{33}$. The same reasoning implies the identification of $\lambda_{33}$, given its sign. In summary, we can identify $\Lambda_1$ provided that $\Lambda_1$ is invertible and the signs of $\lambda_{ii}$ ($i = 1, 2, 3$) are known.\(^5\) Next, from $C_{11} = \Lambda_2 \Lambda_1'$, we identify $\Lambda_2$ from $\Lambda_2 = C_{12}(\Lambda_1')^{-1}$. Thus PC2 together with the column signs of $\Lambda$ (or $F$) imply global identification in the restricted parameter space that ensures invertibility of $\Lambda_1$.

---

\(^3\) The structure of $\Lambda$ 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 $\lambda_i$ ($i = 1, 2, \ldots, r$) to be 1, with $F^T F$ being diagonal (instead of an identity matrix).

\(^4\) A matrix $C$ is skew-symmetric (also known as anti-symmetric) if $C + C^T = 0$. So the diagonal elements of a skew-symmetric matrix are zero, and $C_i = -C_{ii}$.

\(^5\) Identification of $\Lambda_1$ alone does not require $\lambda_{13} \neq 0$, but further identification of $\Lambda_2$ does need $\lambda_{13} \neq 0$ so that $\Lambda_1$ is invertible.
PC3 also implies global identification, but sign restrictions are not necessary. To see this, let $C = A(F'F/T)A'$ be given. Under PC3,

$$A(F'F/T)A' = \begin{bmatrix} (F'F/T) & (F'F/T)A_2 \\ A_2(F'F/T) & A_2A_2' \end{bmatrix}. $$

Knowing $C_{11}$ is equivalent to knowing $F'F/T$. Thus we identity $A_2$ from $A_2 = C_{21}(F'F/T)^{-1} = C_{21}C_{11}^{-1}$.

### 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

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

By Assumption A(e), $Z_n \xrightarrow{d} Z_l$ where $Z_l$ is a zero mean normal vector as $T \to \infty$. To derive the limiting distribution for $\tilde{F}_t$ and $\tilde{\lambda}_i$, 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

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

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

$$\sqrt{T}(\tilde{\lambda}_i - H^{-1}\lambda_i) = H' \sqrt{T} \sum_{i=1}^{T} F_i e_{it} + o_p(1).$$

A useful and alternative expression for (4) is

$$\sqrt{N}(\tilde{F}_t - H'F_t) = H'(A'A/N)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} + o_p(1)$$

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

#### 3.1. PC1

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

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

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

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

because $F'F/T = I_r$ and $\sqrt{T}(H' - I_r) = o_p(1)$ if $\sqrt{T}/N \to 0$. In view of (7), we can now interpret $\tilde{\lambda}_i$ as the least squares estimator for $\lambda_i$ 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

$$\sqrt{N}(\tilde{F}_t - F_t) \xrightarrow{d} N(0, \Sigma^{-1}F_t \Sigma^{-1}).$$

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

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

A formal proof is given in the Appendix. In essence, $\tilde{F}'F/T = I_r + O_p(\delta_{\theta}^{-1})$, and $V = A'A/N + O_p(\delta_{\theta}^{-2})$ under PC1. Thus the limit of $F'F/T$ is $I_r$ and the limit of $V$ is $\Sigma_{\theta}$. Since $A'A/N \to \Sigma_{\theta}$ 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 $\lambda_i$ or some components of $\lambda_i$ are zero. Consider the null hypothesis that $R\lambda_i = \tilde{\lambda}_i$, where $R$ is a $(q \times r)$ known restriction matrix $(q \leq r)$ and $\tilde{\lambda}_i$ is $q \times 1$, a known vector. Under the null hypothesis,

$$T(R\tilde{\lambda}_i - \tilde{\lambda}_i)(R\tilde{\Phi}_i R')^{-1}(R\tilde{\lambda}_i - \tilde{\lambda}_i) \xrightarrow{d} \chi^2_q.$$
r > 1, the representations for \( \hat{F}_t \) and \( \hat{\lambda}_t \) each has an extra term because \( \sqrt{T}(H^* - I) \) is non-negligible. More specifically, for \( i > r \),

\[
\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left( \frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_t \lambda_i + o_p(1) \tag{12}
\]

and for each \( t \),

\[
\sqrt{T}(\hat{F}_t - F_t) = \left( \frac{A' A}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_t \lambda_i \alpha_i + o_p(1) \tag{13}
\]

Let \( \Xi = \sqrt{T}(H^* - I) \) and let \( \Xi_{kh} \) denote the \((k, h)\) element of \( \Xi \) (\( 1 \leq k, h \leq r \)). We show in the Appendix that

\[
\Xi_{kh} = \left\{ \begin{array}{ll}
\omega_p(1), & k > h \\
\omega_{kh} + \omega_p(1), & k = h \\
\omega_{hk} + \omega_p(1), & k < h
\end{array} \right.	ag{14}
\]

where \( \omega_p(1) \) holds if \( \sqrt{T}/N \to 0 \). The limit of the off-diagonal elements of \( \Xi \) are determined by the limit of the off-diagonal elements of \( \xi_t(\Lambda_t)^{-1} \), where \( \xi_t \) is defined as (11).

It turns out that (12) also holds for \( i = 1, 2, \ldots, r \), not just for \( i > r \). For \( 1 \leq i \leq r \), the last \( r - i \) components of \( \hat{\lambda}_i \) and \( \lambda_i \) are zero. Using the asymptotic representation of \( \sqrt{T}(H^* - I) \) 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 \( \Sigma_\lambda = I_r \) under PC2, and \( Z_t \) is the limiting distribution of \( \left( \frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_t \). Let vec(\( A \)) denote the column vector that stacks the lower triangular elements of \( A \) (excluding the diagonal elements). Note that vec(\( A \)) is different from vec(\( A \)). For any skew-symmetric matrix \( A \), there is a duplication matrix \( D \) such that vec(\( D A \)) = vec(\( \xi_t(\Lambda_t)^{-1} \)) + o_p(1). From \( \xi_t(\Lambda_t)^{-1} \to (Z_1, Z_2, \ldots, Z_r)(\Lambda_t)^{-1} \) we have vec(\( \Xi \)) \( \to \eta \), where \( \eta \) is defined as \( \eta = \text{vec}(Z_1, Z_2, \ldots, Z_r)(\Lambda_1)^{-1} \). Then

\[
\sqrt{T}(H^* - I) \hat{\lambda}_t = \Xi_{\lambda_t} = (\lambda_t \odot I_r) \text{vec}(\Xi) = (\lambda_t \odot I_r) D \text{vec}(\Xi) \xrightarrow{d} (\lambda_t \odot I_r) \eta.
\]

Let \( Z_t \) 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 \( \hat{\lambda}_t \) denote the estimates with the restrictions of PC2.

(i) Let \( Z_t = N(0, \Sigma_\xi) \). Then for each \( i \) and as \( N, T \to \infty \) with \( \sqrt{T}/N \to 0 \),

\[
\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} Z_i - (\lambda_i \odot I_r) \eta
\]

where \( \eta = \text{vec}((Z_1, Z_2, \ldots, Z_r)(\Lambda_t)^{-1}) \) and \( D \) is a duplication matrix linking vec(\( \lambda \)) and vec(\( \Lambda_t \)).

(ii) Let \( G_t = N(0, \Sigma_\Lambda) \) and is independent of \( \eta \). If \( N/T \to c \) with \( 0 \leq c < \infty \),

\[
\sqrt{T}(\hat{F}_t - F_t) \xrightarrow{d} G_t + \sqrt{c}(F_t \odot I_r) \eta.
\]

In part (i) of Theorem 2, \( (\lambda_t \odot I_r) \eta \) is the limit of \( \sqrt{T}(H^* - I) \lambda_i \), which is also normal since \( \eta \) 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 \( \sqrt{c}(F_t \odot I_r) \eta \) 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 \( \hat{\lambda}_t \) each has an extra term because \( \sqrt{T}(H^* - I) \) is non-negligible. As \( \lambda_i \) is known for \( i \leq r \), we only need to consider \( i \geq r + 1 \). We show in the Appendix that

\[
\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \left( \frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_t \lambda_i + o_p(1) \tag{15}
\]

and for each \( t \),

\[
\sqrt{T}(\hat{F}_t - F_t) = \left( \frac{A' A}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_t
\]

\[
+ (N/T)^{1/2} \sqrt{T}(H^* - I) \hat{F}_t + o_p(1) \tag{16}
\]

where \( \sqrt{T}(H^* - I) \) is given in (3).

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

(i) Let \( Z_t = N(0, \Sigma_\xi) \). Then for \( i \geq r + 1 \) as \( N, T \to \infty \) with \( \sqrt{T}/N \to 0 \),

\[
\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} Z_i - (Z_1, Z_2, \ldots, Z_r) \lambda_i
\]

(ii) Let \( G_t = N(0, \Sigma_\Lambda) \) and is independent of \( (Z_1, Z_2, \ldots, Z_r) \). If \( N/T \to c \) with \( 0 \leq c < \infty \),

\[
\sqrt{T}(\hat{F}_t - F_t) \xrightarrow{d} G_t + \sqrt{c}(Z_1, Z_2, \ldots, Z_r) F_t.
\]

To understand part (i) of Theorem 3, note that \( Z_t \) 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

\[
\sqrt{T}(H^* - I) \to (Z_1, Z_2, \ldots, Z_r),
\]

which is an \( r \times r \) matrix of random variables.\(^6\) Although \( F'F/T \) (whose limit is \( \Sigma_\xi \)) is not required to be an identity matrix under PC3, it is normally distributed. As a consequence, \( (Z_1, \ldots, Z_r) \lambda_i \) is also normally distributed if \( \lambda_i \) is non-random. It follows that \( \lambda_i \) 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

\[
y_t = F_t' \alpha + W_t' \beta + \epsilon_t
\]

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”

\[
y_t = \hat{F}_t' \alpha + W'_t \beta + v_t = \hat{\delta}' \delta + v_t \tag{17}
\]

where \( v_t = \epsilon_t + (F_t - \hat{F}_t)' \alpha, \hat{\delta} = (\hat{F}_t', W'_t)' \), 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, \( \sqrt{T}(H^* - I) \lambda_i \) is already a vector, so its convergence to the vector \( (Z_1, \ldots, Z_r) \lambda_i \) is well defined.
Assumption C. For $z_t = (F_t', W_t' y_t, E|z_t|^2 \leq M < \infty; E|e_{t1}| z_{t-1}, z_{t-2}, \ldots) = 0$; $z_t$ and $e_t$ are independent of the idiosyncratic errors $e_{ti}$ for all $i$ and $s$. Furthermore, $\frac{1}{T} \sum_{t=1}^{T} z_{t} z_{t}^{'} \to P \Sigma_{zz} > 0$ and $\frac{T^{-1/2}}{T} \sum_{t=1}^{T} z_{t} e_{t} \to N(0, \Sigma_{zze})$, where $\Sigma_{zz} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} e_{t} z_{t} e_{t}^{'} > 0$.

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

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

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 $\hat{\delta}$ is invariant to identification assumptions used, the distribution of $\hat{\delta}$ does depend on whether PC1, PC2, or PC3 is used.

To understand Theorem 4, note that under PC1, $\sqrt{T}(\hat{\alpha} - \alpha) = \sqrt{T}(\hat{\alpha} - H^{-1} \alpha) - \sqrt{T}(H^{-1} \alpha)$. The first term on the right is analyzed by Bai and Ng (2006). Under PC1, $\sqrt{T}(H^{-1} \alpha) = o_p(1)$ provided $\sqrt{T} N \to 0$ since $H - I = O_p(\delta_{N2}^2)$. As $H$ is asymptotically an identity matrix, $\hat{\alpha}$ now directly estimates $\alpha$. 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

$$\Sigma_{\hat{\delta}} = \left( \frac{1}{T} \sum_{t=1}^{T} z_{t} z_{t}^{'} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} z_{t} e_{t} e_{t}^{'} \right) \left( \frac{1}{T} \sum_{t=1}^{T} e_{t} e_{t}^{'} \right)^{-1}$$

which is White’s heteroskedasticity robust covariance estimator using $z_t$ as regressors.

Under PC2 and PC3, $\sqrt{T}(H^{-1} \alpha)$ and $\sqrt{T}(H^{-1} \alpha)$ are asymptotically negligible when $r > 1$. The asymptotic variance of $\hat{\alpha}$ under PC2 has an extra term given by the variance of the asymptotic variance of $\hat{\alpha}$ is due to $\text{var}(\hat{\alpha}, \ldots, \hat{\alpha})$. Details on estimation of the asymptotic variances are given in Appendix A. It is however useful to note that if $e_{t1}$ are independent for $j = 1, 2, \ldots, r$, then the normal vectors $Z_t$ are also independent. In such a case, $\text{var}(\hat{\alpha}, \ldots, \hat{\alpha}) = \sum_{t=1}^{T} \phi_{t1} \alpha_t$ can be consistently estimated by $\sum_{t=1}^{T} \phi_{t1} \hat{\alpha}_t$.

It is useful to remark that when $F_t$ estimates $F_t$ instead of a rotation of $F_t$, we can give economic interpretation to the coefficients on the regressors $F_t$. For example, in factor augmented autoregressions (FARVAR) or for the factor models considered in this paper we can obtain the impulse responses of each observable $X_{ti}$ in the panel to the common shocks that drive $F_t$. Suppose that $F_t = A_1 F_{t-1} + \cdots + A_p F_{t-p} + A_0 u_t$, 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 $v_t$ such that $v_t = A_0 u_t$. Observing $F_t$ (with economic interpretations for each component) allows us to use standard structural VAR analysis to identify $A_0$ and compute the impulse responses $\frac{\partial z_{t+k}}{\partial a_{it}}$. It follows that we can compute the impulse responses for the observable variables $\frac{\partial z_{t+k}}{\partial a_{it}} = \lambda_i \frac{\partial z_{t+k}}{\partial a_{it}}$ for each $i$ and for all $k \geq 0$.

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 $X_t = \mu_i + \delta_i(t) + \lambda_i F_t + e_{it}$, where $\mu_i$ is an individual fixed effect and $\delta_i(t)$ is a time effect. When $\delta_i(t) = \delta_i$, the time effects are common. When $\delta_i(t) = \delta_i - t$, 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 $X_t = \mu + A F_t + e_t$.

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

(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 $\mu$ by $\hat{X} = \frac{1}{T} \sum_{t=1}^{T} X_t$ and use the residuals to estimate $\Lambda$ and $F$. The demeaned data matrix is $Z = [X_t - \bar{X}, \ldots, X_{T} - \bar{X}] = X - \bar{X} \iota_t$.

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

$$(NT)^{-1/2} Z' Z = \tilde{F} \tilde{V}$$

(18)

where $\tilde{V}$ is a $r \times r$ diagonal matrix consisting of the first $r$ largest eigenvalues, arranged in decreasing order. The factor loading estimators is $\hat{\Lambda} = Z \tilde{V} / T$. By construction, $F$ and $\Lambda$ already satisfy PC1, namely, that $F \tilde{F} / T = I$, and $\Lambda \tilde{\Lambda} = \text{diagonal}$. We now want to show that (i) these estimates also satisfy the constraint (Fe1) and (ii) that $\hat{\lambda}_i$ has the same expression with or without demeaning.

To see (i), first note that $\iota_t Z' = \iota_t X' - (\iota_t I) X' = \iota_t X' - TX'$ which equals zero by definition of $\bar{X}$. Multiply $\iota_t$ 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 Foroni et al. (2009).

8 The model is still static even though $F_t$ is dynamic.
Since $\bar{V}$ is an invertible (diagonal) matrix of eigenvalues, it follows that $\mathbf{e}' \bar{F} = \sum_{i=1}^{T} \bar{F} \mathbf{e} = 0$, which is (FE1). The principal components estimator for $\lambda$ can now be rewritten as

$$\bar{\lambda} = \bar{Z}'/T = (X - \bar{X})' \bar{F}'/T = \bar{X}' \bar{F}'/T$$

where the last equality makes use of the result $\mathbf{e}' \bar{F} = 0$. 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 $\tilde{\lambda}_i$ is of the same form with or without fixed effects note that since $\bar{F}_t = F_t - \bar{F}$ and $\bar{F} = 0$ by assumption, the model in demeaned data is $X_t - \bar{X}_t = \lambda_i F_t + e_{it} - \tilde{e}_t$.

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

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

$$= \left(\bar{F}' \bar{F} / T\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it} + o_p(1).$$

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 $e_{it} - \tilde{e}_t$ in (7), we have

$$\sqrt{N} (\tilde{F}_t - F_t) = \left(\frac{\bar{A}'}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i (e_{it} - \tilde{e}_t) + o_p(1)$$

$$= \left(\frac{\bar{A}'}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} - T^{-1/2} \left(\frac{\bar{A}'}{N}\right)^{-1}$$

$$\times \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \lambda_i \bar{e}_t + o_p(1)$$

$$= \left(\frac{\bar{A}'}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} + o_p(1).$$

The second term on the right hand side is $O_p(T^{-1/2})$ because $(NT)^{-1/2} \sum_{t=1}^{T} \lambda_i \bar{e}_t = o_p(1)$. 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 $\bar{F}$ and $\bar{A}$. 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

$$\frac{1}{N} \sum_{i=1}^{N} \lambda_i = 0.$$  \hspace{1cm} (FE2)

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

$$\tilde{X}_t = \lambda_i F_t + \tilde{e}_t.$$  

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 $\tilde{X}_t$, 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 $\tilde{e}_t - \bar{X}_t$. Specifically,

$$T^{-1/2} \sum_{t=1}^{T} F_t (e_{it} - \tilde{e}_t) = T^{-1/2} \sum_{t=1}^{T} F_t \tilde{e}_t + O_p(T^{-1/2})$$

$$= T^{-1/2} \sum_{t=1}^{T} F_t \tilde{e}_t + \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} F_t e_{it}$$

$$= T^{-1/2} \sum_{t=1}^{T} F_t \tilde{e}_t + O_p(T^{-1/2})$$

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

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

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

$$\sqrt{N} (\tilde{\lambda}_i - \lambda_i) = \left(\frac{F' F}{T}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} F_i e_{it} + o_p(1)$$

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

$$\sqrt{N} (\tilde{\lambda}_i - \lambda_i) \overset{d}{\rightarrow} N(0, \sigma^2 \left[ E(F_i F_i') \right]^{-1}).$$

Now estimation of the fixed effects will also remove the mean from $F_i$ although the representation looks the same, the limiting variance of $\tilde{\lambda}_i$ is then $\sigma^2 \left[ var(F_i) \right]^{-1}$. As $F_i$ can have non-zero mean, the second moment $E(F_i F_i')$ is in general larger than the variance of $F_i$. As $E(F_i F_i') \geq var(F_i)$ implies $\left[ E(F_i F_i') \right]^{-1} \leq \left[ var(F_i) \right]^{-1}$, the limiting variance of $\tilde{\lambda}_i$ 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 $\lambda_i$ is considered to be a vector of random variables.

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

5.3. Heterogeneous trends

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

\[ X_{nt} = \mu_i + \delta_t + \lambda_i \epsilon_{it} + e_{it}. \]

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

We focus on the identification restriction PC1, i.e., \( F'F/T = I \) and \( A'A \) is diagonal. Let \( X_{nt}' \) denote the residuals from the least squares detrending for each series \( i \). We have

\[ X_{nt}' = \lambda_i F_t + e_{it}', \]

where \( F_t \) and \( e_{it}' \) are also the residuals from the least squares detrending (no actual detrending is performed on them since they are unobservable). Let \( a_{i,t} \) and \( b_{i,t} \) be the OLS coefficients when \( F_t \) is regressed on \([1, t]\), and \( a_{i,t} \) and \( b_{i,t} \) are similarly defined, we have

\[ F_t = F_t - a_{i,t} - b_{i,t} t \]

\[ e_{it}' = e_{it} - a_{i,t} - b_{i,t} t. \]

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

\[ \sqrt{N}(\lambda_i - \lambda_i) = \left( F'F/T \right)^{-1} \sum_{t=1}^T F_t e_{it}' + o_p(1) \]

\[ = \left( F'F/T \right)^{-1} \sum_{t=1}^T F_t e_{it} + o_p(1). \]

Note that we can replace \( e_{it}' \) by \( e_{it} \) because \( \{F_t\} \) is orthogonal to the sequence \([1, t]\). Similarly,

\[ \sqrt{N}(\bar{\lambda}_i - \lambda_i) = \left( \frac{A'A}{N} \right)^{-1} \sum_{i=1}^N \lambda_i e_{it} + o_p(1) \]

\[ = \left( \frac{A'A}{N} \right)^{-1} \sum_{i=1}^N \lambda_i e_{it} - \left( \frac{A'A}{N} \right)^{-1} \times \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i (a_{i,t} + b_{i,t} t) + o_p(1) \]

\[ = \left( \frac{A'A}{N} \right)^{-1} \sum_{i=1}^N \lambda_i e_{it} + o_p(1). \]

The last equality follows from the fact that \( a_{i,t} + b_{i,t} t \) is a linear combination of \( \frac{1}{T} \sum_{t=1}^T e_{it} \) and \( \frac{1}{T} \sum_{t=1}^T e_{it}^{1/2} \), each of which is \( O_p(T^{-1/2}) \). Using the assumption that \( e_{it} \) has weak cross-sectional correlation, we can show that \( N^{-1/2} \sum_{i=1}^N \lambda_i (a_{i,t} + b_{i,t} t) = O_p(T^{-1/2}) \). Asymptotic normality for \( \sqrt{N}(\bar{\lambda}_i - \lambda_i) \) and for \( \sqrt{N}(\bar{\lambda}_i - \lambda_i) \) follows from the fact that \( T^{-1/2} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \) and \( N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \) 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 \( \lambda_i \) and \( \bar{\lambda}_i \) 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 R²: \( \bar{F}_i \) rotated under PC2. |
|-------------------|-------|-------|-------|-------|-------|-------|
| Series            | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
| ces002            | 0.789 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| ips10             | 0.564 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| syfg1t            | 0.034 | 0.043 | 0.794 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| pushx             | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| fyg1t             | 0.068 | 0.016 | 0.007 | 0.004 | 0.797 | 0.000 | 0.000 | 0.000 |
| hsgpr             | 0.154 | 0.005 | 0.086 | 0.000 | 0.019 | 0.739 | 0.000 | 0.000 |
| fnmra             | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.648 | 0.000 |
| fscom             | 0.003 | 0.029 | 0.020 | 0.001 | 0.050 | 0.000 | 0.000 | 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 IC1 and IC2 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 database 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 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) syfg1t, spread between one-year T-bill rate (fyg1t) and fed funds rate; (4) pushx, CPI excluding shelter; (5) fyg1t, one year T-bill rate; (6) hsgpr, housing units authorized; (7) fnmra, total reserves; (8) fscom, 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 \( \bar{F}_i \) are shocks to employment, while shocks to \( \bar{F}_j \) are industrial production shocks orthogonal to employment, and so forth.

### Table 1 reports the marginal explanatory power of the \( j \)-th factor.

The \((i, j)\) 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 \( i \)-th 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 \((i, j)\) entry equals the difference between \( R^2(j) \) and \( R^2(j - 1) \). The results confirm 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 \( R^2 \)’s obtained by regressing these same series on the standard principal component estimates.
Table 2
Marginal $R^2$: $\hat{F}_i$.

<table>
<thead>
<tr>
<th>Series</th>
<th>Factor 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>cspi02</td>
<td>0.695</td>
<td>0.005</td>
<td>0.000</td>
<td>0.017</td>
<td>0.050</td>
<td>0.004</td>
<td>0.001</td>
<td>0.016</td>
</tr>
<tr>
<td>2</td>
<td>0.662</td>
<td>0.032</td>
<td>0.002</td>
<td>0.076</td>
<td>0.092</td>
<td>0.001</td>
<td>0.008</td>
<td>0.041</td>
</tr>
<tr>
<td>3</td>
<td>0.113</td>
<td>0.385</td>
<td>0.005</td>
<td>0.025</td>
<td>0.162</td>
<td>0.139</td>
<td>0.038</td>
<td>0.004</td>
</tr>
<tr>
<td>4</td>
<td>0.003</td>
<td>0.028</td>
<td>0.701</td>
<td>0.035</td>
<td>0.050</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.196</td>
<td>0.144</td>
<td>0.018</td>
<td>0.257</td>
<td>0.242</td>
<td>0.003</td>
<td>0.011</td>
<td>0.022</td>
</tr>
<tr>
<td>6</td>
<td>0.288</td>
<td>0.005</td>
<td>0.010</td>
<td>0.173</td>
<td>0.188</td>
<td>0.218</td>
<td>0.024</td>
<td>0.017</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.001</td>
<td>0.028</td>
<td>0.007</td>
<td>0.001</td>
<td>0.014</td>
<td>0.477</td>
<td>0.003</td>
</tr>
<tr>
<td>8</td>
<td>0.002</td>
<td>0.170</td>
<td>0.004</td>
<td>0.009</td>
<td>0.027</td>
<td>0.064</td>
<td>0.003</td>
<td>0.426</td>
</tr>
</tbody>
</table>

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 Silva 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 $\Sigma_A$ by $\hat{\Sigma}_A = \hat{\Lambda}' \hat{\Lambda} / N$. To estimate $\Phi_i$ and $\Gamma_i$, we can use one of the three methods given in Bai and Ng (2006). Let $\hat{\Phi}_i$ and $\hat{\Gamma}_i$ denote these estimates. Then $\Sigma_A^{-1} \Gamma_i \Sigma_A^{-1}$ is estimated by $\hat{\Sigma}_A^{-1} \hat{\Gamma}_i \hat{\Sigma}_A^{-1}$.

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

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

where $\text{var}(\eta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$, with $\tilde{\xi}_t = \text{vec}(\hat{F}_t (e_{t1}, \ldots, e_{tr}) \Lambda_{r-1}^{-1})$. With serial correlation in $\eta$, the variance of $\eta$ is the limit of $\frac{1}{T} \sum_{t=1}^{T} \sum_{l=1}^{T} E(\xi_l' \xi_t)$, and it is estimated by the Newey–West method using the series $\hat{\xi}_t (t = 1, 2, \ldots, T)$. Given $\text{var}(\eta)$, we estimate $\text{var}(\eta)$ by $\text{Avar}(\hat{\lambda}_i)$.

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

PC3. We separately discuss whether or not $e_s$ is cross-sectionally independent.

Case i: If $e_s$ is cross-sectionally independent, then $Z_i$ are independent over $i$ and

$$\text{Avar}(\hat{\lambda}_i) = \Sigma_A^{-1} \Gamma_i \Sigma_A^{-1} + C^2 \sum_{s=1}^{k} \Phi_s F_s^2,$$

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

$$\text{Avar}(\hat{\lambda}_i) = \Sigma_A^{-1} \Gamma_i \Sigma_A^{-1} + C^2 \sum_{s=1}^{k} \Phi_s F_s^2,$$

where $\text{var}(\eta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$,

which is estimated by $\text{var}(\eta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$. Then $\eta$ is the limit of $\frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$, in the absence of serial correlation in $\eta$, the variance of $\eta$ is equal to the probability limit of $\frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$, and it is estimated by $\text{var}(\eta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{\xi}_t \tilde{\xi}_t'$, with $\tilde{\xi}_t = \text{vec}(\hat{F}_t (e_{t1}, \ldots, e_{tr}) \Lambda_{r-1}^{-1})$. With serial correlation in $\eta$, the variance of $\eta$ is the limit of $\frac{1}{T} \sum_{t=1}^{T} \sum_{l=1}^{T} E(\xi_l' \xi_t)$, and it is estimated by the Newey–West method using the series $\hat{\xi}_t (t = 1, 2, \ldots, T)$. Given $\text{var}(\eta)$, we estimate $\text{var}(\eta)$ by $\text{Avar}(\hat{\lambda}_i)$.

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

Case ii: If $e_s$ is cross-sectionally correlated, then combining (15) and (3), we have

$$\sqrt{T} (\hat{\lambda}_i - \lambda_i) = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{F}_t \otimes b_t \right) + \sigma_{e_s}(1),$$

where $b_t$ is a 2 by 1 vector with $e_s$ as the first element and $(e_{1t}, \ldots, e_{rt}) \lambda_i = \sum_{s=1}^{k} \hat{F}_s \otimes b_t$ as the second element. Thus the limiting covariance is given by

$$\text{Avar}(\hat{\lambda}_i) = \Sigma_F^{-1} (\hat{F}_i - \hat{F}_i) \Psi (I_r - \hat{F}_i) \Sigma_F^{-1}.$$
where $\Psi_t = \lim \frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T} E(F_t F_t' \otimes b_s b_s')$, which specializes to $\Psi_t = \lim \frac{1}{T} \sum_{t=1}^{T} (F_t F_t' \otimes b_s b_s')$ in the absence of time series correlation. To estimate $\Psi_t$, apply the Newey–West estimator to the sequence $\tilde{F}_t \otimes \tilde{b}_t$. The asymptotic variance is estimated by $\text{Var}(\tilde{\lambda}_t) = \Sigma_{\lambda}^{-1} E(\lambda_t, -\lambda_t) \Sigma_{\lambda}^{-1}$.

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,

$$\sqrt{N}(H^f - I_r)F_t = (F_t' \otimes I_r) \text{vec} \left( \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s' \otimes a_s \right) + O_p(1)$$

where $a_s = (e_{t1}, \ldots, e_{tr})'$. Let $Y = \text{Var}(\text{vec}(Z_1, \ldots, Z_T))$

$$Y = \frac{1}{T} \sum_{s=1}^{T} E[\text{vec}(F_s' \otimes a_s) \text{vec}(F_s' \otimes a_s') ]$$

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

$$\text{Var}(F_t) = \Sigma_{\lambda}^{-1} F_t \Sigma_{\lambda}^{-1} + c^2 (F_t' \otimes I_r) Y (F_t \otimes I_r).$$

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

### Appendix B

#### Proof of (2).

Rewrite

$$\tilde{F}' T = (\tilde{F} - HF') T + H' F' T$$

because $\tilde{F} - HF' T = O_p(\delta_{NT}^2)$, see Lemma B.2 of Bai (2003).

Right multiply $H$ to both sides,

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

Rewrite the left hand side of above as

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

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

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

Thus if $(F' F') = I_r$, we have

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

(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 $H$

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

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

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

This equation implies that $H$ (up to a negligible term) is a matrix consisting of eigenvectors of $(\tilde{A}' \Lambda / N)$. 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 $\Lambda$ by $-1$). This implies $H = I_r + O_p(\delta_{NT}^2)$. Moreover, from (23) we obtain

$$(\tilde{A}' \Lambda / N)/\tilde{V} + O_p(\delta_{NT}^2).$$

#### Proof of Theorem 1.

Result (2) leads to representations (7) and (8). The theorem is a direct consequence of these representations and Assumption A.

#### 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

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

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

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

That is, $\tilde{\lambda}_1 = H^{-1} \lambda_1' + O_p(T^{-1/2})$. By the QR decomposition, we have $Q R = \tilde{Q} R = H^{-1} A_1' + Q R'$, Since $\tilde{A}_1$ 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 $Q H = I_r + O_p(T^{-1/2})$. 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, $\Delta' + \Delta + \Delta = O_p(\delta_{NT}^2)$. But $\Delta' + \Delta = O_p(1/T)$, so $\Delta' + \Delta = O_p(\delta_{NT}^2)$. This implies that the diagonal elements of $\Delta$ are all $O_p(\delta_{NT}^2)$ and $\Delta$ 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 $A$. Using (5), we can write

$$(\tilde{\lambda}_1' - H^{-1} \lambda_1') = H^{-1} \frac{1}{T} \sum_{t=1}^{T} F_t(e_{t1}, \ldots, e_{tr}) + O_p(T^{-1/2}).$$

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

$$H(\tilde{\lambda}_1' - \lambda_1') = (F' F' T)^{-1} \frac{1}{T} \sum_{t=1}^{T} F_t(e_{t1}, \ldots, e_{tr}) + O_p(T^{-1/2}).$$

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

$$H(\tilde{\lambda}_1' - \lambda_1') = T^{-1/2} \xi r_t + O_p(T^{-1/2}).$$

By the QR decomposition of $\tilde{A}_1$, $H A_1 = H Q R = (H Q - I_r) + R = \Delta R + R$. Thus $H(\tilde{\lambda}_1' - \lambda_1') = \Delta R + (R - \lambda_1')$. It follows that $\Delta = -(R - \lambda_1')R^{-1} + T^{-1/2} \xi r_t R^{-1} + O_p(T^{-1/2})$.

Since both $R$ and $\lambda_1'$ are upper triangular matrices, the below diagonal elements of $\Delta$ are equal to the corresponding elements of $T^{-1/2} \xi r_t R^{-1} + O_p(T^{-1/2})$. Since $\Delta$ is skew-symmetric up to an $O_p(\delta_{NT}^2)$ order, the elements of $\Delta$ above the diagonal are also given.
Proof of (12). Using \( \hat{\lambda}_i = Q' \hat{\lambda}_i \),
\[
\hat{\lambda}_i - \lambda_i = Q'(\hat{\lambda}_i - H^{-1} \lambda_i) + Q' H^{-1} (I - HQ) \lambda_i.
\]

Multiplying \( \sqrt{T} \),
\[
\sqrt{T} (\hat{\lambda}_i - \lambda_i) = Q' \sqrt{T} (\hat{\lambda}_i - H^{-1} \lambda_i) - Q' H^{-1} \sqrt{T} (HQ - I) \lambda_i.
\]

Since \( Q' H^{-1} = I + o_p(1) \), the second term on the right hand side
is \( - \sqrt{T} (HQ - I) \lambda_i + o_p(1) \). Using (5), the first term on the right hand side
is \( Q'H' (I - I) \lambda_i + o_p(1) \), but \( Q'H' = I + o_p(1) \) \( (F'F/T)^{-1} + o_p(1) \) under PC2. Combining the results yield (12). This argument holds for all \( i = 1, 2, \ldots, N \). \( \square \)

Proof of (13). Using \( \hat{F}_i = Q' \hat{F}_i \),
\[
\hat{F}_i - F_i = Q' (\hat{F}_i - F_i) + (Q' H' - I) F_i.
\]

Multiplying \( \sqrt{N} \),
\[
\sqrt{N} (\hat{F}_i - F_i) = Q' \sqrt{N} (\hat{F}_i - F_i) + (N/T)^{1/2} \sqrt{T} (Q'H' - I) F_i.
\]

From (6), the first term on the right is \( Q'H' (\Lambda'A/N)^{-1} \sum_{i=1}^{N} \lambda_i^\beta_i + o_p(1) \); but \( Q'H' = I + o_p(1) \). For the second term on the right,
\[
\sqrt{T} (Q'H' - I) F_i = - \sqrt{T} (HQ - I) F_i + o_p(1) \text{ because } \sqrt{T} (HQ - I) \text{ is skew-symmetric up to an } o_p(1) \text{ term when } \sqrt{T}/N \to 0.
\]

Combining the results we obtain (13). \( \square \)

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

Proof of (3). Note \( H^T = H \hat{\Lambda}_1' \) is the rotation matrix under PC3. Since the principal components estimator satisfies \( \hat{\lambda}_i - H^{-1} \lambda_i = O_p(T^{-1/2}) \), we have
\[
\hat{\Lambda}_1' = (\hat{\lambda}_1, \ldots, \hat{\lambda}_r) = H^{-1} (\lambda_1, \ldots, \lambda_r) + O_p(T^{-1/2}).
\]

Left multiply \( H \) to obtain \( H \hat{\Lambda}_1' = I + O_p(T^{-1/2}) \) because \( (\lambda_1, \ldots, \lambda_r) = I \) under PC3. That is, \( H^T = I + O_p(T^{-1/2}) \) so \( H^T \to I \). Using representation (5), we have
\[
\sqrt{T} (H^T - I) = H^{-1} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} (F_i e_{it}, \ldots, F_i e_{it}) + o_p(1).
\]

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

Proof of (15). Recall that
\[
\hat{\lambda}_i - \lambda_i = \hat{\lambda}_i^{-1} \hat{\lambda}_i - \lambda_i = \hat{\lambda}_i^{-1} (\hat{\lambda}_i - H^{-1} \lambda_i) + (\hat{\lambda}_i^{-1} H^{-1} - I) \lambda_i.
\]

Multiply \( \sqrt{T} \) on each side
\[
\sqrt{T} (\hat{\lambda}_i - \lambda_i) = \hat{\lambda}_i^{-1} \sqrt{T} (\hat{\lambda}_i - H^{-1} \lambda_i) + \hat{\lambda}_i^{-1} (H^{-1} - I) \lambda_i.
\]

For the first term on the right hand side, using (5),
\[
\hat{\lambda}_i^{-1} \sqrt{T} (\hat{\lambda}_i - H^{-1} \lambda_i) = (\hat{\lambda}_i^{-1} H^{-1}) (H H^T)^{-1} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} F_i e_{it} + o_p(1).
\]

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

Thus
\[
\hat{\lambda}_i^{-1} \sqrt{T} (\hat{\lambda}_i - H^{-1} \lambda_i) = \left( \frac{F'F}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} F_i e_{it} + o_p(1).
\]

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

Proof of (16). First note that \( \hat{F}_i - F_i = \hat{\Lambda}_1 (\hat{F}_i - H F_i) = \hat{\Lambda}_1 (\hat{F}_i - H F_i) + (H^T - I) F_i \). It follows that
\[
\sqrt{N} (\hat{F}_i - F_i) = \hat{\Lambda}_1 \sqrt{N} (\hat{F}_i - H F_i) + (N/T)^{1/2} \sqrt{T} (H^T - I) F_i.
\]

From (6), and using \( \hat{\Lambda}_1 H^T = I + o_p(1) \), the first term on the right hand side is
\[
\hat{\Lambda}_1 \sqrt{N} (\hat{F}_i - H F_i) = \left( \frac{\Lambda'A}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} + o_p(1).
\]

Combining the two equations leads to (16). \( \square \)

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

Proof of Theorem 4. We first consider the case of identification under PC1 so that we use \( F_i \) in place of \( \hat{F}_i \) in the regression model. We can rewrite the model as in Bai and Ng (2006)
\[
y_t = (\hat{F}_i, W_i') \left( H^{-1} \alpha \beta \right) + e_t + (F_i H - \hat{F}_i) H^{-1} \alpha
\]

where \( \hat{F}_i = (\hat{F}_i', W_i') \), \( \delta^* = (\alpha'H^T, \beta') \), and \( \alpha_t \) represents the last term on the right hand side. When \( \sqrt{T}/N \to 0 \), Bai and Ng (2006) shows that the error \( \alpha_t \) is negligible, and the least squares estimator \( \hat{\delta} \) has the standard limiting distribution as if \( \hat{F}_i \) contains no estimation error (as if \( H^T F_i \) were observable). More specifically,
\[
\sqrt{T} (\delta - \delta^*) \to N(0, \Phi_0^{-1} \Sigma_{zz,zz} \Sigma_{zz} \Phi_0^{-1})
\]

where \( \Phi_0 = \text{diag}(V^{-1} Q \Sigma_A, I) \) and \( V^{-1} Q \Sigma_A \) is the probability limit of \( H \), where \( Q \) represents the probability limit of \( F_i F_i' \). In our case, the limit of \( H \) is an identity matrix (also follows from \( Q = I \) and \( V = \Sigma_A \) in the present case) so that \( \Phi_0 \) is an identity matrix. This implies that
\[
\sqrt{T} (\delta - \delta^*) \to N(0, \Sigma_A)
\]

where \( \Sigma_A = \Sigma_{zz,zz} \Sigma_{zz}^{-1} - 1 \). Furthermore,
\[
\sqrt{T} (\delta - \delta^*) \to \sqrt{T} (\delta - \delta^*) + \sqrt{T} [ (\alpha - H^{-1} \alpha'), 0'].
\]

But \( \sqrt{T} (\alpha - H^{-1} \alpha) = \sqrt{T} (H - I) H^{-1} \alpha = o_p(1) \) provided that
\[
\sqrt{T}/N \to 0 \text{ because } H - I = O_p(\delta_{NT}^{-2}). \text{ It follows that under } \sqrt{T}/N \to 0, \sqrt{T} (\delta - \delta^*) \to N(0, \Sigma_A). \]

We next consider PC3. We use \( \hat{F}_i \) in place of \( F_i \), where \( \hat{F}_i \) is defined in the main text. Since \( \hat{F}_i \) is an estimate of \( H^T F_i \), we define \( \delta^* = (\alpha', \beta') \). Then \( y_t = \hat{\xi}_t^* + e_t + a_t^0 = (F_i^T H - \hat{F}_i) H^{-1} \alpha \). The same argument in Bai and Ng (2006) leads to
\[ \sqrt{T}(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \Phi_0^{-1} \Sigma_0^{-1} \Sigma_{zz} \Sigma_{zz} \Phi_0^{-1}) \]

where \( \Phi_0 = \text{diag}(\text{plim}H_1^T, I) \). Under PC3, \( \text{plim}H_1^T = I_r \). Thus, \( \sqrt{T}(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \Sigma_0) \), where \( \Sigma_0 \) is defined earlier. Next, \( \sqrt{T}(\hat{\delta} - \delta) = \sqrt{T}(\hat{\delta} - \delta^*) + \sqrt{T}(\alpha - H^{-1}\alpha)^T, 0' \).

But the term
\[ \sqrt{T}(\alpha - H^{-1}\alpha) = \sqrt{T}(H^T - I_r)H^{-1} \alpha \]

is not negligible and \( \sqrt{T}(H^T - I_r) \xrightarrow{d} (Z_1, \ldots, Z_r) \) and \( H^{-1}
\alpha = \alpha_0(1) \). It follows that
\[ \sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_0) + \left[(Z_1, \ldots, Z_r)\alpha_0(1)\right] \]  

Since the normal random variable \( N(0, \Sigma_0) \) is derived from the central limit theorem (CLT) involving \( \{e_j\} \), while \( (Z_1, \ldots, Z_r) \) are derived from the CLT involving \( \{e_j\} \), these normal variables are independent of each other under Assumption C. Therefore, the asymptotic variance of \( \hat{\delta} \) is equal to \( \Sigma_0 + \text{diag}(\text{var}(Z_1, \ldots, Z_r) \alpha) \), where \( \alpha \) is defined earlier. For dependent \( e_j \) 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 \( e_j \) 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 \( \delta^* = [(H^{r-1}\alpha)', \beta']' \). The same analysis as in PC3 gives
\[ \sqrt{T}(\hat{\delta} - \delta^*) \xrightarrow{d} N(0, \Sigma_0) \]. Furthermore, \( \sqrt{T}(\alpha - H^{-1}\alpha) = \sqrt{T}(H^T - I_r)H^{r-1}\alpha = \sqrt{T}(H^T - I_r) \alpha + \alpha_0(1) = (\alpha' \otimes I_r)D \text{vec}(\xi T A_1^{-1}) + o_p(1) \), which converges in distribution to \((\alpha' \otimes I_r)D \eta \). Thus the asymptotic variance of \( \hat{\delta} \) is equal to \( \Sigma_0 + \text{diag}(\alpha' \otimes I_r)D \text{var}(\eta)D' (\alpha' \otimes I_r), 0' \), where \( \text{diag} \) means block-diagonal.

References

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


