Columbia University

Department of Economics
Discussion Paper Series

Principal Components Estimation and Identication of the Factors

Jushan Bai
Serena Ng

Discussion Paper No.: 1112-02

Department of Economics
Columbia University
New York, NY 10027

August 2011

Principal Components Estimation and Identification of the Factors

Jushan Bai* Serena Ng †

June 20, 2011

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 factor estimates 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.

Keywords: diffusion indices, FAVAR, rotation, factor space, skew-symmetric matrices.

JEL classification: C30, C33

*Department of Economics, Columbia University, IAB 1019, 420 West 118th Street, New York, NY 10027. Email:jushan.bai@columbia.edu

†Department of Economics, Columbia University, IAB 1012, 420 West 118 Street, New York, NY 10027. Email: Serena.Ng@columbia.edu.

We thank three anonymous referees and Dalibor Stevanovic for helpful comments. The authors acknowledge financial support from the NSF (SES-0962410, SES-0962431).

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 estimates obtained by the method of principal components 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 $\widetilde{F}_t$ is the vector of factor estimates, there exists an $r \times r$ invertible matrix H such that $\widetilde{F}_t$ estimates $H'F_t$. Asymptotic results are stated in terms of $\widetilde{F}_t - H'F_t$. Similarly, if λ_i is the vector of factor loadings and $\widetilde{\lambda}_i$ is the corresponding estimate, asymptotic results are known for $\widetilde{\lambda}_i - H^{-1}\lambda_i$.

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 object of interest is the 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, one might be interested in the sensitivity to different risk factors, and knowing whether the factors are related to real macroeconomic activity, to inflation, or to financial markets is useful. This would involve putting restrictions on the factor loadings. In factor augmented regressions of the form $y_t = \alpha'\widetilde{F}_t + W_t'\beta + \varepsilon_t$, one might be interested in testing hypothesis concerning α. Since the asymptotic theory is only available for $\sqrt{T}(\widehat{\alpha} - H^{-1}\alpha)$, the test is uninformative except when α is zero.

To be able to give meaningful interpretation to the factor estimates within the principal components framework, one would need to know H. In general, H is unrestricted and need not be a diagonal matrix. However, the more assumptions we impose on the data generating process, the more we know about H. In the important case when the assumptions imply that H is an identity matrix, the principal components estimator will directly estimate F, instead of a rotation of it. This case is important because $\widetilde{F}_t$ can be treated as though they were the latent F_t. In factor-augmented regressions, $\widehat{\alpha}$ can be given economic interpretation.

We study three sets of restrictions on F and Λ such that H is an identity matrix asymptotically. We also derive the asymptotic distributions for the estimated factors and the loadings under each set of the restrictions. After presenting the results for factor models with no deterministic components, the 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 $i = 1, \dots N$ and $t = 1, \dots T$, 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 $F = (F_1, F_2, ..., F_T)'$ is the $T \times r$ matrix of factors and $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)'$ 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 $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.

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), $E|N^{-1/2}\sum_{i=1}^N \Big[e_{is}e_{it} - E(e_{is}e_{it})\Big]|^4 \leq M$.

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 (A.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)-(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. Although both F_t and e_{it} are allowed to be dynamic processes, the model is a (generalized) static factor model as opposed to a generalized dynamic factor model. For the latter, the readers are referred to Forni et al. (2000).

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 $\widetilde{F} = (\widetilde{F}_1, ..., \widetilde{F}_T)'$, 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 $\widetilde{\Lambda} = (\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_N)' = X'\widetilde{F}/T$ is a $N \times r$ matrix of estimated factor loadings. The estimators $\widetilde{F}$ and $\widetilde{\Lambda}$ satisfy the normalization restrictions since $\widetilde{F}'\widetilde{F}/T = I_r$ holds by construction. It is also easy to show that $\widetilde{\Lambda}'\widetilde{\Lambda}/N = \widetilde{V}$ is a diagonal matrix, where $\widetilde{V}$ 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

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

Under Assumption A, Stock and Watson (2002) and Bai and Ng (2002) showed that H is invertible and $\widetilde{F}$ estimates FH (a rotation of F) and $\widetilde{\Lambda}$ estimates $\Lambda H'^{-1}$ (a rotation of Λ), though the product $\widetilde{F}\widetilde{\Lambda}'$ 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). We consider restrictions that will lead to exact identification, under which we show that H is asymptotically an identity matrix.[1] The $H = I_r$ case is of special interest because the factor estimates can be treated as though they were the latent factors underlying the data, and not just the space spanned by them.

We consider three sets of restrictions that lead to exact identification. One can also use more than r^2 restrictions. Examples of over-identifying factor models in economics can be found, for example, in Heaton and Solo (2004) and Reis and Watson (2010). However, the principal components method is not suitable for imposing over-identifying restrictions.

[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.

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	$\frac{1}{T}F'F = I_r$	$\Lambda = \begin{pmatrix}\Lambda_1 \\ \Lambda_2\end{pmatrix}, \quad \Lambda_1 = \begin{pmatrix}\lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr}\end{pmatrix}, \lambda_{ii} \neq 0, i = 1, \dots r$
(2.3): PC3	unrestricted	$\Lambda = \begin{pmatrix} I_r \\ \Lambda_2 \end{pmatrix}$

PC1 requires that the diagonal elements of $\Lambda'\Lambda$ are distinct and positive. PC2 assumes that diagonal elements of Λ_1 are nonzero so that Λ_1 is of full rank. PC3 also requires $F'F/T$ to be invertible, which is a requirement of r factors.

2.1 PC1

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. 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. But will these restrictions uniquely identify F and Λ?

The answer turns out to be yes, if the restrictions defined by PC1 also hold for the underlying F and Λ that generate the data. Under this assumption, it is shown in the Appendix that

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

where δ_{NT} denote $\min[\sqrt{N}, \sqrt{T}]$ throughout this paper. That is to say, if the underlying F satisfies $F'F/T = I_r$ and the underlying Λ is such that $\Lambda'\Lambda$ is a diagonal matrix with distinct elements, then H can be taken as an identity matrix with the implication that $\widetilde{F}_t$ estimates F_t without rotation asymptotically. Distinctness of the diagonal entries of $\Lambda'\Lambda$ ensures that each eigenvalue of $\Lambda'\Lambda$ is associated with a unique unitary eigenvector up to a sign change.

The intuition for (2) is the following. The data generating process implies

$$XX' = F\Lambda'\Lambda F' + \sigma^2 I_T + Fe' + eF' + (ee' - \sigma^2 I_T).$$

If we divide both sides by NT, we see that the last three terms are negligible (assuming time series stationarity and weak serial correlation for simplicity). The first r eigenvectors associated with the first r largest eigenvalues of the matrix $XX'/(NT)$ are thus determined by the matrix $T^{-1}F'(\Lambda'\Lambda/N)F'$. For $j = 1, 2, ..., r$, let F^j be the jth column of F. Under PC1, we can write

$$\frac{1}{T}F(\Lambda'\Lambda/N)F' = v_1\frac{1}{T}F^1F^{1\prime} + v_2\frac{1}{T}F^2F^{2\prime} + \cdots v_r\frac{1}{T}F^rF^{r\prime}, \tag{3}$$

where $\Lambda'\Lambda/N = \text{diag}(v_1, v_2, ..., v_r)$, with $v_i > v_{i+1}$, and $F^{i\prime}F^j/T = 1$ for $i = j$ and 0 for $i \neq j$. The j-th unitary eigenvector of the matrix in (3) is either F^j or $-F^j$ (no other unitary vector can be the eigenvector). Thus we can determine the individual columns of F without rotation.

The assumption that $\Lambda'\Lambda/N$ is a diagonal matrix is motivated on statistical ground and may be difficult to interpret. However, consider the following special factor loading matrix[2]:

$$\Lambda = \text{block-diag}(\pi_1, \pi_2, ..., \pi_r) \tag{4}$$

where π_i is $N_i \times 1$ with $N = \sum_{i=1}^r N_i$. If $r = 3$,

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

Under (4), the first factor only affects the first N_1 variables, and second factor only affects the next N_2 variables, and so forth. Knowledge on which variables are affected by the first factor, which variables by the second factor, etc is not required. Row permutation is equivalent to multiplying Λ by an $N \times N$ orthogonal matrix B. So long as $B'B = I_N$, $(B\Lambda)'(B\Lambda) = \Lambda'\Lambda$ is still diagonal. Note that the above restriction is imposed on Λ directly, and is a sufficient condition for $\Lambda'\Lambda$ being diagonal. But if Λ is of the form (4), the principal components estimator does not make use of many of the zero restrictions. However, since the principal components method uses exact identification restrictions, it is possible to test over-identifying restrictions based on the limiting distributions derived in this paper.

2.2 PC2

While PC1 is analytically convenient, such a structure is somewhat unusual. A more plausible set of restrictions is to let the first $r \times r$ block of Λ be a lower triangular matrix, leaving the rest of the elements unconstrained. This leads to PC2. Let Λ_2 be the unrestricted $(N - r) \times r$ submatrix of Λ. The $r(r - 1)/2$ (exclusion) restrictions are now imposed on Λ_1. 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. PC2 assumes that $F'F/T = I_r$, and knowledge of which variable is affected by

[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.

the first factor only, which variable is affected by the first two factors only, and so on. The choice of the first r variables of X_t and their ordering provide the auxiliary information for identification.[3]

Given the unrestricted estimates $\widetilde{F}$ and $\widetilde{\Lambda}$, it is easy to obtain estimators satisfying PC2. Let $\widehat{F}$ and $\widehat{\Lambda}$ denote the estimators that satisfy PC2 so that $\widehat{F}'\widehat{F}/T = I_r$ and $\widehat{\Lambda}_1$ is lower triangular. Also let $\widetilde{\Lambda}_1$ be the first $r \times r$ block of $\widetilde{\Lambda}$, the principal component estimator defined earlier. Then $\widehat{F}$ and $\widehat{\Lambda}$ can be obtained as follows.

- Step 1: obtain a QR decomposition of $\widetilde{\Lambda}_1'$ to yield

$$\widetilde{\Lambda}_1' = Q \cdot R$$

  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 $\widetilde{\Lambda}_1$.

- Step 2: define

$$\widehat{F} = \widetilde{F} \cdot Q, \quad \widehat{\Lambda} = \widetilde{\Lambda} \cdot Q = \begin{bmatrix} R' \\ \widehat{\Lambda}_2 \end{bmatrix}.$$

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

Since $\widehat{F}$ and $\widehat{\Lambda}$ are rotations of the principal component estimates $\widetilde{F}$ and $\widetilde{\Lambda}$, they are equivalent in some sense. However, their asymptotic distributions will be different. We show in the appendix that while H^* is asymptotically an identity matrix and $\Xi = \sqrt{T}(H^* - I_r)$ is skew-symmetric up to an $o_p(1)$ term[4] if F and Λ underlying the data satisfy PC2,

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

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 consistent with $\Xi = o_p(1)$. However, for $r > 1$, the limiting distributions of $\widehat{F}_t$ and $\widehat{\lambda}_i$ will be affected by the limit of $\sqrt{T}(H^* - I_r) = O_p(1)$. Let Ξ_{kh} denote the (k, h)th element of Ξ $(1 \leq k, h \leq r)$. We show in the appendix that

$$\Xi_{kh} = \begin{cases} (\xi_T \Lambda_1'^{-1})_{kh} + o_p(1), & k > h \\ o_p(1) & k = h \\ -\Xi_{hk} + o_p(1), & k < h \end{cases} \tag{5}$$

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 (14).

[3] A variation to PC2 is to normalize the diagonal elements λ_{ii} $(i = 1, 2, ..., r)$ 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 $C_{ij} = -C_{ji}$.

2.3 PC3

The third set of identification restrictions specifies the first $r \times r$ block of Λ (denoted Λ_1) to be an identity matrix and leaves the factor process F completely unrestricted. 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, which resembles classical 'errors-in-variables' models in which $X_{it} = F_{ti} + e_{it}$ for $i = 1, \ldots r$, as in Pantula and Fuller (1986), and Wansbeek and Meijer (2000, p.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 $\widetilde{\Lambda}$ and $\widetilde{F}$, let

$$\widehat{\Lambda} = \widetilde{\Lambda}\widetilde{\Lambda}_1^{-1}, \quad \widehat{F} = \widetilde{F}\widetilde{\Lambda}_1'.$$

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

$$\sqrt{T}(H^\dagger - I_r) = \xi_T + o_p(1). \tag{6}$$

where ξ_T is defined in (14). The fact that $\sqrt{T}(H^\dagger - I_r)$ is not negligible for all $r \geq 1$ will affect the limiting distributions of $\widehat{\lambda}_i$ and $\widehat{F}_t$.

2.4 Local vs. Global Identification

Global and local identifications for factor models are discussed, for example, by 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 only 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. Observing $F\Lambda'$ implies observing $\Lambda(F'F/T)\Lambda' = \Lambda\Lambda'$ since $F'F/T = I_r$. Let $C = \Lambda\Lambda'$. From $\Lambda' = (\Lambda_1', \Lambda_2')$, we have

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

where we also partition the observable matrix C correspondingly. Suppose for concreteness that

$r = 3$. Observing C_{11} is equivalent to observing 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.

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,

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

Observing C_{11} is equivalent to observing $F'F/T$. 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, $Z_{Ti} \xrightarrow{d} Z_i$ for a zero mean normal vector Z_{Ti} as $T \to \infty$. To derive the limiting distribution for $\widehat{F}_t$ and $\widehat{\lambda}_i$, we use the asymptotic representations for $\widetilde{F}_t$ and $\widetilde{\lambda}_i$, given in Theorems 1 and 2 of Bai (2003). Specifically, if $\sqrt{N}/T \to 0$, then

$$\sqrt{N}(\widetilde{F}_t - H'F_t) = \widetilde{V}^{-1}\Big(\frac{\widetilde{F}'F}{T}\Big)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1) \tag{7}$$

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

$$\sqrt{T}(\widetilde{\lambda}_i - H^{-1}\lambda_i) = H'\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it} + o_p(1). \tag{8}$$

A useful and alternative expression for (7) is

$$\sqrt{N}(\widetilde{F}_t - H'F_t) = H'\Big(\frac{\Lambda'\Lambda}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1) \tag{9}$$

[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.

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

3.1 PC1

Under PC1, $H' = I_r + O_p(\delta_{NT}^{-2})$. It follows that $\sqrt{N}(\widetilde{F}_t - F_t) = \sqrt{N}(\widetilde{F}_t - H'F_t) + \sqrt{N}(\widetilde{H}' - I_r) = \sqrt{N}(\widetilde{F}_t - H'F_t) + o_p(1)$, provided that $\sqrt{N}/\delta_{NT}^2 = o(1)$, or equivalently, $\sqrt{N}/T \to 0$. Thus under PC1, we can rewrite (9) as

$$\sqrt{N}(\widetilde{F}_t - F_t) = \Big(\frac{\Lambda'\Lambda}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1). \tag{10}$$

This result says that $\widetilde{F}_t$ 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

$$\sqrt{T}(\widetilde{\lambda}_i - \lambda_i) = \Big(\frac{F'F}{T}\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it} + o_p(1) \tag{11}$$

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 (11), we can now interpret $\widetilde{\lambda}_i$ 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 (7) and (8), 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 $\widetilde{F}_t$ and $\widetilde{\lambda}_i$ because the right hand side of (10) and (11) do not depend on any estimated quantities.

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

$$\sqrt{N}(\widetilde{F}_t - F_t) \quad \xrightarrow{d} \quad N(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}). \tag{12}$$

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

$$\sqrt{T}(\widetilde{\lambda}_i - \lambda_i) \quad \xrightarrow{d} \quad N(0, \Phi_i). \tag{13}$$

A formal proof is given in the Appendix. In essence, $\widetilde{F}'F/T = I_r + O_p(\delta_{NT}^{-1})$, and $\widetilde{V} = \Lambda'\Lambda/N + O_p(\delta_{NT}^{-2})$ under PC1. Thus the limit of $\widetilde{F}'F/T$ is I_r and the limit of $\widetilde{V}$ is Σ_Λ. Since $\Lambda'\Lambda/N \to \Sigma_\Lambda$ by Assumption A.b, and (12) follows from (10). Furthermore, (11) together with $F'F/T = I_r$ imply (13). 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 (13) 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 $(q \leq r)$ and $\bar{\lambda}_i$ is $q \times 1$, a known vector. Under the null hypothesis,

$$T(R\widetilde{\lambda}_i - \bar{\lambda}_i)'(R\widehat{\Phi}_i R')^{-1}(R\widetilde{\lambda}_i - \bar{\lambda}_i)\xrightarrow{d}\chi^2_q.$$

We can also test restrictions between λ_i and λ_j $(i \neq j)$. Put $\delta = (\lambda_i', \lambda_j')'$ and $\widehat{\delta} = (\widehat{\lambda}_i', \widehat{\lambda}_j')'$. 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 (11), if $E(e_{it}e_{jt}) = 0$ for $i \neq j$, then $\widehat{\lambda}_i$ and $\widehat{\lambda}_j$ are asymptotically independent. So let $\widehat{\Phi} = \text{diag}(\widehat{\Phi}_i, \widehat{\Phi}_j)$ (a block-diagonal matrix), then

$$T(R\widehat{\delta} - \bar{\delta})'(R\widehat{\Phi}R')^{-1}(R\widehat{\delta} - \bar{\delta})\xrightarrow{d}\chi^2_q.$$

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 F can be similarly constructed.

3.2 PC2

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

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

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

$$\xi_T = \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (F_t e_{1t}, ..., F_t e_{rt}) = (Z_{T1}, ..., Z_{Tr}). \tag{14}$$

The limiting distributions of the factor estimates under PC2 depend on whether $r > 1$. If $r = 1$, PC1 and PC2 are identical, so the limiting distributions $\widehat{F}_t$ and $\widehat{\lambda}_i$ are given in Theorem 1. It remains to consider $r > 1$. The representations for $\widehat{F}_t$ and $\widehat{\lambda}_i$ each has an extra term because $\sqrt{T}(H^* - I_r)$ is non-negligible. More specifically, for $i > r$,

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

and for each t,

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

It turns out that (15) also holds for $i = 1, 2, ..., r$, not just for $i > r$. For $1 \leq i \leq r$, the last $r - i$ components of $\widehat{\lambda}_i$ and of λ_i are zero. It can be seen from the asymptotic representation of $\sqrt{T}(H^* - I_r)$ that the last $r - i$ components of the right hand side 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 $\text{vec}(A) = D\,\text{veck}(A)$. Equation (5) implies $\text{veck}(\Xi) = \text{veck}(\xi_T \Lambda_1^{\prime -1}) + o_p(1)$. Since $\xi_T(\Lambda_1')^{-1} \xrightarrow{d} (Z_1, Z_2, ..., Z_r)(\Lambda_1')^{-1} \equiv \eta$ and Z_i is the limit of the first term on the right hand side of (15):

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

Theorem 2 *Suppose that Assumptions A, B, and PC2 hold. Let $\widehat{F}_t$ and $\widehat{\lambda}_i$ 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 $\sqrt{T}/N \to 0$,

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

where $\eta = \text{veck}[(Z_1, Z_2, ..., Z_r)\Lambda_1^{\prime -1}]$ 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 $0 \leq c < \infty$,

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

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 (16), and $\sqrt{c}(F_t' \otimes I_r)D\eta$ is the limit of the second term of (16).

Hypothesis testing can be performed similarly as in Section 3.1.

3.3 PC3

Similar to PC2, the representations for $\widehat{F}_t$ and $\widehat{\lambda}_i$ each has an extra term due to the non-negligibility of $\sqrt{T}(H^\dagger - I_r)$. Now λ_i is known for $i \leq r$. Consider $i \geq r + 1$. We show in the Appendix that

$$\sqrt{T}(\widehat{\lambda}_i - \lambda_i) = \left(\frac{F'F}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it} - \sqrt{T}(H^\dagger - I_r)\lambda_i + o_p(1) \qquad (17)$$

and for each t,

$$\sqrt{N}(\widehat{F}_t - F_t) = \Big(\frac{\Lambda'\Lambda}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + (N/T)^{1/2}\sqrt{T}(H^\dagger - I_r)'F_t + o_p(1). \quad (18)$$

where $\sqrt{T}(H^\dagger - I_r)$ is given in (6).

Theorem 3 *Suppose that Assumptions A, B, and PC3 hold. Let $\widehat{F}_t$ and $\widehat{\lambda}_i$ 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 $i \geq r+1$, as $N, T \to \infty$ with $\sqrt{T}/N \to 0$,

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

ii Let $G_t =^d N(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1})$. If $N/T \to c$ with $0 \leq c < \infty$,

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

where G_t is independent of $(Z_1, ..., Z_r)$.

To understand part (i) of Theorem 2, note that Z_i is the limit of the first term on the right hand side of (17). Under Assumptions A, B, and PC3, the second term in (17) satisfies

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

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. So $(Z_1, ..., Z_r)\lambda_i$ is also normally distributed if λ_i is non-random. It follows that $\widehat{\lambda}_i$ is still normally distributed. Similarly, part (ii) of Theorem 2 comes from the fact that G_t is the limiting random variable for the first term on the right hand side of (18).

Again, hypothesis testing can be performed similarly as in Section 3.1.

4 Implications for Factor-Augmented Regressions

Consider the infeasible regression model

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

[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 $(Z_1, ...Z_r)\lambda_i$ is well defined.

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

$$y_t = \widehat{F}_t'\alpha + W_t'\beta + v_t = \widehat{z}_t'\delta + v_t \tag{19}$$

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

Assumption C: For $z_t = (F_t', W_t')'$, $E\|z_t\|^4 \le M < \infty$; $E(\varepsilon_t|z_{t-1}, z_{t-2}, ...) = 0$; 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 is observed, then under Assumption C, the asymptotic variance of $\widehat{\delta}$ is given by $\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}$. In Bai and Ng (2006), we showed that $\widehat{\alpha}$ is an estimate of $H^{-1}\alpha$ (and not α) when $\widetilde{F}_t$ is used in place of F_t. The following theorem studies the properties of $\widehat{\delta}$ when $\widehat{F}_t$ is in place of F_t.

Theorem 4 *Suppose* $\sqrt{T}/N \to 0$ *and Assumptions A, B, and C hold. Define* $\Sigma_\delta = \Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}$. *Let* $\delta' = (\alpha', \beta')$ *and let* $\widehat{\delta}$ *be obtained by the least squares estimation of factor augmented regression (19), where* $\widehat{F}_t$ *is obtained under the restrictions defined by PC1, PC2, or PC3. Then*

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

where $\text{Avar}(\widehat{\delta}) = \Sigma_\delta$ *under PC1,* $\text{Avar}(\widehat{\delta}) = \Sigma_\delta + \text{diag}[(\alpha' \otimes I_r)D\,\text{var}(\eta)D'(\alpha \otimes I_r), 0]$ *under PC2, and* $\text{Avar}(\widehat{\delta}) = \Sigma_\delta + \text{diag}(\text{var}[(Z_1, ..., Z_r)\alpha], 0)$ *under PC3. Furthermore,* η *and* D *are defined in Section 3.2, and* $(Z_1, ..., Z_r)$ *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, $\widehat{\delta}$ has properties as though the latent factors in the data F_t were used as regressors. Although the distribution of $\widehat{\beta}$ is invariant to identification assumptions used, the distribution of $\widehat{\alpha}$ does depend on whether PC1, PC2, or PC3 is used.

To understand Theorem 4, note that under PC1,

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

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, $\widehat{\alpha}$ now directly estimates α. Thus, the limiting distribution for $\sqrt{T}(\widehat{\alpha} - H^{-1}\alpha)$ 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 $\Sigma_{\widehat{\delta}}$ can be consistently estimated by

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

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

Under PC2 and PC3, $\sqrt{T}(H^*-I_r)$ (for $r > 1$) and $\sqrt{T}(H^\dagger-I_r)$ are not asymptotically negligible. The asymptotic variance of $\widehat{\alpha}$ 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 $\widehat{\alpha}$ is due to var$[(Z_1, ..., Z_r)\alpha]$. 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 $j = 1, 2, ..., r$, then the normal vectors Z_j are also independent. In such a case, var$[(Z_1, ..., Z_r)\alpha] = \sum_{k=1}^{r}\Phi_k\alpha_k$ can be consistently estimated by $\sum_{k=1}^{r}\widehat{\Phi}_k\widehat{\alpha}_k$.

It is useful to remark that when $\widehat{F}_t$ estimates F_t instead of a rotation of F_t, we can give economic interpretation to the coefficients on the regressors $\widehat{F}_t$. 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 $F_t = A_1F_{t-1} + \cdots + A_pF_{t-p} + Gu_t$, where u_t is a vector of structural shocks, and G is an $r \times r$ matrix linking the structural shocks u_t to the reduced form shocks v_t such that $v_t = Gu_t$. Observing F_t (with economic interpretations for each component) allows us to use standard structural VAR analysis to identity G and compute $\frac{\partial F_{t+k}}{\partial u_t}$. It follows that $\frac{\partial X_{i,t+k}}{\partial u_t} = \lambda_i'\frac{\partial F_{t+k}}{\partial u_t}$ for each i and for all $k \geq 0$.

5 Factor Models with Deterministic Terms

In practice, the data are demeaned and trends are removed before the factors are estimated. Factor models with determinist 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

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

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

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 $\iota_T = (1, 1, ..., 1)'$, a $T \times 1$ vector, the restriction is[8]

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

In the absence of fixed effects, the principal components estimator is based on the $N \times T$ data matrix $X'X$, where $X = [X_1, X_2, ..., X_T]$. 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

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

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

$$(NT)^{-1} Z'Z\widetilde{F} = \widetilde{F}\widetilde{V}. \tag{20}$$

where $\widetilde{V}$ is $r \times r$ diagonal matrix consisting of the first r largest eigenvalues, arranged in decreasing order. The factor loading estimator is $\widetilde{\Lambda} = Z\widetilde{F}/T$. By construction, F and Λ already satisfy PC1, namely, that $\widetilde{F}'\widetilde{F}/T = I_r$ and $\widetilde{\Lambda}'\widetilde{\Lambda} =$ diagonal. We now want to show that (i) these estimates also satisfy the constraint FE1 and (ii) that $\widetilde{\lambda}_i$ 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 (20), we have

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

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

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

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

[8] This restriction may be replaced by $E(F_t) = 0$ if F_t is a random process without affecting the limiting result.

To show (ii) that the limiting distribution for $\widetilde{\lambda}$ is of the same form with or without fixed effects. 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 (11) and since $\sum_{t=1}^{T} F_t \bar{e}_i = (\sum_{t=1}^{T} F_t)\bar{e}_i = 0$,

$$\sqrt{N}(\widetilde{\lambda}_i - \lambda_i) = \Big(\frac{F'F}{T}\Big)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t(e_{it} - \bar{e}_i) + o_p(1) = \Big(\frac{F'F}{T}\Big)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{it} + o_p(1).$$

This representation coincides with (11). Thus under Assumptions A and B, PC1 and FE1, the limit is again $\sqrt{N}(\widetilde{\lambda}_i - \lambda_i) \sim N(0, \Phi_i)$, which is (13). The limiting distribution for $\widetilde{F}_t$ also has the same form with or without demeaning. Replacing e_{it} with $e_{it} - \bar{e}_i$ in (10), we have

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

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 $\widetilde{F}_t$ 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 $\widetilde{F}$ and $\widetilde{\Lambda}$. Thus when FE1 holds, the expression for $\widetilde{\lambda}_i$ and $\widetilde{F}_t$ 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]

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

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

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 (11) 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 $(\sum_{t=1}^{T} F_t)\bar{e}_{i.} = 0$ and $(\sum_{t=1}^{T} F_t)\bar{e}_{..} = 0$ 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

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

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

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

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 $\widehat{\lambda}_i$ is then $\sigma^2[var(F_t)]^{-1}$. As F_t can have non-zero mean, the second moment $E(F_t F_t')$ is in general larger than the variance of F_t. As $E(F_t F_t') \geq \text{var}(F_t)$ implies $[E(F_t F_t')]^{-1} \leq [\text{var}(F_t)]^{-1}$, the limiting variance of $\widehat{\lambda}_i$ is smaller when fixed effects are known to be absent and are not estimated.

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

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

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

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

section. Together with diagonality of $\Lambda'\Lambda$ under PC1, we can use earlier arguments to show that

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

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

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

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 $(\frac{1}{T}\sum_{s=1}^T \frac{s}{T}e_{is})\frac{t}{T}$, each of which is $O_p(T^{-1/2})$. Since e_{it} is zero-mean and is weakly cross-sectionally correlated, we have $N^{-1/2}\sum_{i=1}^N \lambda_i(a_{i,e} + b_{i,e}t) = O_p(T^{-1/2})$. Asymptotic normality for $\sqrt{T}(\widetilde{\lambda}_i - \lambda_i)$ and for $\sqrt{N}(\widetilde{F}_t - F_t^\tau)$ 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. The asymptotic representation and covariance matrix estimates for $\widehat{\lambda}_i$ and $\widehat{F}_t$ also have the same form as when deterministic terms are absent.

In the preceding discussion, we have focused on the limiting distributions under PC1, showing that the limiting distributions have the same form as the case without deterministic intercepts or trends. The same results hold for PC2 and PC3. The details are omitted.

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 by 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 to use the ordering of the variables in the data in identification of 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 $\widehat{F}_1$ are shocks to employment, while shocks to $\widehat{F}_2$ are industrial production shocks orthogonal to employment, and so forth.

Table 1 reports the marginal R^2, defined as $R^2(k) - R^2(k-1)$, where $R^2(k)$ is the R^2 in a regression of the series in question on k rotated factors. The (i,j)th entry in the table is computed as follows. 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 the two R^2s. 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 stock market.

Table 1: Marginal R^2 : $\widehat{F}_t$ 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

It is useful to compare the marginal R^2s obtained by regressing these same series on the standard principal component estimates, $\widetilde{F}_t$. 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.

Table 2: Margianl R^2: $\widetilde{F}_t$

	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

Using the PC2 rotation, the eight factors are much more concentrated on variation 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 $\widehat{F}$ is warranted.

7 Conclusion

The principal components estimator uses the restrictions that $F'F/T = I_r$ and $\Lambda'\Lambda$ is a diagonal matrix. In general, the method only estimates the space spanned by the factors. This paper considers three sets of restrictions under which the factors and the loadings can be estimated without rotations. Limiting distributions are derived, and the asymptotic covariance matrices are obtained for each case separately. Other restrictions might also imply a rotation matrix H that is an identity matrix. Their asymptotic properties can be derived using the analysis for PC1, PC2, and PC3 as guide.

Appendix A

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

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

PC2. To estimate the asymptotic variance of $\widehat{\lambda}_i$, 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 $(Z_1, ..., Z_r)$). Noting that $(F'F/T) = I_r$ under PC2,

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

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 $\zeta_t = \text{veck}[F_t(e_{1t}, ..., e_{rt})\Lambda_1'^{-1}]$. Then η is the limit of $T^{-1/2}\sum_{t=1}^T \zeta_t$. In the absence of serial correlation in e_{jt} $(j = 1, 2, ..., r)$, 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 $\widehat{\text{var}}(\eta) = \frac{1}{T}\sum_{t=1}^T \widehat{\zeta}_t\widehat{\zeta}_t'$ with $\widehat{\zeta}_t = \text{veck}[\widehat{F}_t(\widehat{e}_{1t}, ..., \widehat{e}_{rt})\widehat{\Lambda}_1'^{-1}]$. 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 $\widehat{\zeta}_t$ $(t = 1, 2, ..., T)$. Given $\widehat{\text{var}}(\eta)$, we estimate $\text{Avar}(\widehat{\lambda}_i)$ by

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

where $\widehat{\Phi}_i = \frac{1}{T}\sum_{t=1}^T \widehat{F}_t\widehat{F}_t'\widehat{e}_{it}^2$ in the absence of serial correlation in e_{it}, and $\widehat{\Phi}_i$ is constructed by the Newey-West method based on the series $\widehat{F}_t\widehat{e}_{it}$ 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 $F_t e_{it}$ and ζ_t so τ_t is an $r + r(r-1)/2$ dimensional vector. Then $\sqrt{T}(\widehat{\lambda}_i - \lambda_i) = [I_r, -(\lambda_i' \otimes I_r)D]T^{-1/2}\sum_{t=1}^T \tau_t + o_p(1)$. In the absence of serial correlation in e_{it}, we estimate the variance of $T^{-1/2}\sum_{t=1}^T \tau_t$ by $\widehat{V}_\tau = \frac{1}{T}\sum_{t=1}^T \widehat{\tau}_t\widehat{\tau}_t'$; in the presence of serial correlation, $\widehat{V}_\tau$ is the Newey-West estimator using the series $\widehat{\tau}_t$. Finally,

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

Consider now estimating the asymptotic variance of $\widehat{F}_t$. 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

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

It is estimated by

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

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

$$\text{Avar}(\widehat{\lambda}_i) = \Sigma_F^{-1}\Big(\Phi_i + \sum_{k=1}^{r}\Phi_k\lambda_{ik}^2\Big)\Sigma_F^{-1}$$

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

$$\text{Avar}(\widehat{F}_t) = \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1} + c^2\sum_{r=1}^{k}\Phi_k F_{tk}^2. \tag{21}$$

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

Case ii: If e_{it} is cross-sectionally correlated, then combining (17) and (6), we have

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

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

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

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

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

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

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

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

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 $\sqrt{N}(\widehat{F}_t - F_t)$ becomes

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

To estimate Υ, apply the Newey-West estimator to the sequence $\widehat{F}_s' \otimes \widehat{a}_s$. The asymptotic variance of $\widehat{F}_t$ is estimated by $\widehat{\text{Avar}}(\widehat{F}_t) = \widehat{\Sigma}_\Lambda^{-1}\widehat{\Gamma}_t\widehat{\Sigma}_\Lambda^{-1} + c^2(\widehat{F}_t' \otimes I_r)\widehat{\Upsilon}(\widehat{F}_t \otimes I_r)$.

Appendix B

Proof of (2). Rewrite

$$\widetilde{F}'F/T = (\widetilde{F} - FH)'F/T + H'F'F/T = H'F'F/T + O_p(\delta_{NT}^{-2}) \tag{22}$$

because $(\widetilde{F} - FH)'F = O_p(\delta_{NT}^{-2})$, see Bai (2003, Lemma B.2). Right multiply H to both sides,

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

Rewrite the left hand side of above as

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

because $\widetilde{F}'(FH - \widetilde{F})/T = O_p(\delta_{NT}^{-2})$ and $\widetilde{F}'\widetilde{F}/T = I_r$. Equate the above two equations we obtain

$$I_r = H'(F'F/T)H + O_p(\delta_{NT}^{-2}). \tag{23}$$

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

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

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

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

where we use the fact that $\widetilde{F}'F/T = H' + O_p(\delta_{NT}^{-2})$ under $F'F/T = I_r$, see (22). Multiplying $\widetilde{V}$ on both sides and taking the transpose

$$(\Lambda'\Lambda/N)H = H\widetilde{V} + O_p(\delta_{NT}^{-2}). \tag{25}$$

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 nonzero 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 $\widetilde{F}$ and $\widetilde{\Lambda}$ by -1). This implies $H = I_r + O_p(\delta_{NT}^{-2})$. Moreover, from (25) we obtain

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

Proof of Theorem 1. Result (2) leads to representations (10) and (11). The theorem is a direct consequence of these representations and Assumptions A and B.

Proof of (5). Note $H^* = HQ$ is the rotation matrix under PC2. Under PC2, $F'F/T = I_r$, thus (24) 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 (24) by Q' and right multiply it by Q, and use $Q'Q = I_r$, we have

$$I_r = Q'H'HQ + O_p(\delta_{NT}^{-2}). \tag{26}$$

We next show HQ is a diagonal matrix, up to an $O_p(T^{-1/2})$ term. By (8), for each i, $\widetilde{\lambda}_i - H^{-1}\lambda_i = O_p(T^{-1/2})$, we have

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

That is, $\widetilde{\Lambda}_1' = H^{-1}\Lambda_1' + O_p(T^{-1/2})$. By the QR decomposition, we have $QR = \widetilde{\Lambda}_1' = H^{-1}\Lambda_1' + O_p(T^{-1/2})$. 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 (26), we strengthen the rate to $HQ = I_r + O_p(\delta_{NT}^{-2})$. For general $r > 1$, the rate cannot be improved. Let $\Delta = HQ - I_r = O_p(T^{-1/2})$. Equation (26) 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 (8), we can write

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

Left multiplying H and using $HH' = I_r + O_p(\delta_{NT}^{-2}) = (F'F/T)^{-1} + O_p(\delta_{NT}^{-2})$ [see (24), which still holds under PC2], we have

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

The first term on the right hand side is $T^{-1/2}\xi_T$, where ξ_T given in (14), so that

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

By the QR decomposition of $\widetilde{\Lambda}_1'$, $H\widetilde{\Lambda}_1' = HQR = (HQ - I)R + R = \Delta R + R$. Thus $H\widetilde{\Lambda}_1' - \Lambda_1' = \Delta R + (R - \Lambda_1')$. 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 $\Delta_{ii} = O_p(\delta_{NT}^{-2})$ $(i, j = 1, 2, ..., r)$. 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, (5) is obtained by noting $\Xi = \sqrt{T}\Delta$.

Proof of (15). Using $\widehat{\lambda}_i = Q'\widetilde{\lambda}_i$,

$$\widehat{\lambda}_i - \lambda_i = Q'\widetilde{\lambda}_i - \lambda_i = Q'(\widetilde{\lambda}_i - H^{-1}\lambda_i) + Q'H^{-1}(I - HQ)\lambda_i$$

Multiplying $\sqrt{T}$,

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

Since $Q'H^{-1} = I_r + o_p(1)$, the second term on the right hand side is $-\sqrt{T}(H^* - I_r)\lambda_i$. Using (8), 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 (15). This argument holds for all $i = 1, 2, ..., N$.

Proof of (16). Using $\widehat{F}_t = Q'\widetilde{F}_t$,

$$\widehat{F}_t - F_t = Q'\widetilde{F}_t - F_t = Q'(\widetilde{F}_t - H'F_t) + (Q'H' - I_r)F_t$$

Multiplying $\sqrt{N}$,

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

From (9), 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. Combining the results yield (16).

Proof of Theorem 2. This is a direct consequence of (5), (15), (16), Assumptions A and B.

Proof of (6). Note $H^\dagger = H\widetilde{\Lambda}_1'$ is the rotation matrix under PC3. Since the principal components estimator satisfies $\widetilde{\lambda}_i - H^{-1}\lambda_i = O_p(T^{-1/2})$, we have

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

Left multiply H to obtain $H\widetilde{\Lambda}_1' = I_r + O_p(T^{-1/2})$ because $(\lambda_1, ..., \lambda_r) = I_r$ under PC3. That is, $H^\dagger = I_r + O_p(T^{-1/2})$ so $H^\dagger \xrightarrow{p} I_r$. Using representation (8), we have

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

However, (23) implies $HH' = (F'F/T)^{-1} + O_p(\delta_{NT}^{-2})$. This proves (6).

Proof of (17). Recall that

$$\widehat{\lambda}_i - \lambda_i = \widetilde{\Lambda}_1'^{-1}\widetilde{\lambda}_i - \lambda_i = \widetilde{\Lambda}_1'^{-1}(\widetilde{\lambda}_i - H^{-1}\lambda_i) + (\widetilde{\Lambda}_1'^{-1}H^{-1} - I_r)\lambda_i.$$

Multiply $\sqrt{T}$ on each side

$$\sqrt{T}(\widehat{\lambda}_i - \lambda_i) = \widetilde{\Lambda}_1'^{-1}\sqrt{T}(\widetilde{\lambda}_i - H^{-1}\lambda_i) + \widetilde{\Lambda}_1'^{-1}H^{-1}\sqrt{T}(I_r - H^{\dagger})\lambda_i.$$

For the first term on the right hand side, using (8),

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

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

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

The second term on the right hand side equals $\sqrt{T}(I_r - H^{\dagger})\lambda_i + o_p(1)$. This proves (17).

Proof of (18). First note that

$$\widehat{F}_t - F_t = \widetilde{\Lambda}_1\widetilde{F}_t - F_t = \widetilde{\Lambda}_1(\widetilde{F}_t - H'F_t) + \widetilde{\Lambda}_1 H'F_t - F_t$$

$$= \widetilde{\Lambda}_1(\widetilde{F}_t - H'F_t) + (H'^{\dagger} - I_r)F_t$$

It follows that

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

From (9), and using $\widetilde{\Lambda}_1 H' = I_r + o_p(1)$, the first term on the right hand side is

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

Combining the two equations leads to (18).

Proof of Theorem 3 . This follows from (6), (17), (18), and Assumptions A and B.

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

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

where $\widehat{z}_t' = (\widetilde{F}_t', W_t')'$, $\delta^* = (\alpha' H^{-\prime}, \beta')'$, 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 $\widehat{\delta}$ has the standard limiting distribution as if $\widetilde{F}_t$ contains no estimation error (as if $H'F_t$ were observable). More specifically,

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

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 $\widetilde{F}'F/T$. 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

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

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

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

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 $\sqrt{T}/N \to 0$, $\sqrt{T}(\widehat{\delta} - \delta) \xrightarrow{d} N(0, \Sigma_\delta)$.

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

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

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

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

But the term

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

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

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

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

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

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

with $\sqrt{T}(\widehat{\delta} - \delta^*)\xrightarrow{d}N(0, \Sigma_\delta)$. Furthermore, $\sqrt{T}(\alpha - H^{*-1}\alpha) = \sqrt{T}(H^* - I_r)H^{*-1}\alpha = \sqrt{T}(H^* - I_r)\alpha + o_p(1) = (\alpha' \otimes I_r)D\,\text{veck}(\xi_T \Lambda_1'^{-1}) + o_p(1)$, which converges in distribution to $(\alpha' \otimes I_r)D\eta$. Thus the asymptotic variance of $\widehat{\delta}$ 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. and Ng, S. 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70:1**, 191–221.

Bai, J. and Ng, S. 2006, Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions, *Econometrica* **74:4**, 1133–1150.

Bai, J. and Ng, S. 2010, Instrumental Variables Estimation in a Data Rich Environment, **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. and Reichlin, L. 2009, Opening the Black Box: Identifying Shocks and Propogation Mechanisms in VAR and Factor Models, *Econometric Theory* **25**, 1319–1347.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2000. The Generalized Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics* **82:4**, 540–554.

Hallin, M, and R. Liska (2008). Dynamic factor models with a block structure, EUI Working Papers, ECO 2008-22.

Heaton, C. and Solo, V. 2004, Identification of Causal Factor Models of Stationary Time Series, *Econometrics Journal* **7**, 618–627.

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

Ludvigson, S. and Ng, S. 2011, A Factor Analysis of Bond Risk Premia, *in* D. Gilles and A. Ullah (eds), *Handbook of Empirical Economics and Finance*, Chapman and Hall, pp. 313–372.

Moench, E. and Ng, S. 2011, A Factor Analysis of Housing Market Dynamics in the US and the Regions, *Econometrics Journal* **14**, C1–C24.

Pantula, S. and Fuller, W. 1986, Computational Algorithoms for the Factor Model, *Communications in Statistics- Simulation and Computation* **15**, 227–259.

Reis, R. and Watson, M. 2010, Relative Goods' Prices, Pure Inflation, and the Phillips Correlation, *American Economic Journal: Macroeconomics* **2:3**, 128–157.

Stock, J. H. and 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. and 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 mansucript, Department of Economics, New York University.

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