INFERENCE ON FACTOR MODELS OF LARGE DIMENSIONS

Jushan Bai*

March 25, 2001

Abstract

This paper develops the inferential theory for factor models of large dimensions. The principal component method is considered because it is easy to implement in practice and it is asymptotically equivalent to the maximum likelihood method (if normality is assumed). We derive the rate of convergence and the limiting distributions of the estimated factors, factor loadings, and common components. The theory is developed under the framework of large cross section (N) and large time dimension (T), a setting to which the classical factor analysis does not apply.

We show that the estimated common components are asymptotically normal with a convergence rate equal to the square root of the minimum of N and T. The estimated factors and their loadings are generally normal, although not always so. Conditions for non-normality are identified. The convergence rate of the estimated factors and factor loadings can be faster than that of the estimated common components. These results are obtained under general conditions that allow for correlations and heteroskedasticities in both dimensions. When the idiosyncratic errors are uncorrelated and homoskedastic, stronger results are obtained. For large N but fixed T, a necessary and sufficient condition for consistency is found.

JEL Classification: C13, C33, C43

Keywords: Approximate factor models, principal components, common components, large model analysis, consistency.

---

*Department of Economics, Boston College. Email: jushan.bai@bc.edu Phone: 617-552-3689

I am grateful to James Stock for his encouragement in pursing this research and to Serena Ng for her constructive comments and help throughout this research. All errors are my own responsibility.

Financial support from the National Science Foundation under grant SBR-9709508 is acknowledged.

1 Introduction

Economists nowadays have the luxury of working with an increasingly large volume of data. For example, the Penn World Tables contain thirty variables for more than one-hundred countries covering the postwar years. The World Bank has data for about two-hundred countries over forty years. State and sectoral level data are also widely available. A useful method for summarizing information in data rich environments is factor analysis. More importantly, many economic problems fall directly into the framework of large-dimension factor models; e.g., the arbitrage pricing theory of Ross (1976); the rank theory of consumer demand systems of Lewbel (1991); diffusion-index forecasting of Stock and Watson (1998, 1999), and disaggregated-business-cycle methodology of Forni and Reichlin (1998). While there is a well developed inferential theory for small-dimension (classical) factor models, the inferential theory for large-dimension factor models is not well understood. The purpose of this paper is to partially fill in this gap.

The classical factor analysis assumes a fixed N with T independent and identically distributed (iid) observations; both the factors and idiosyncratic errors are iid and the idiosyncratic covariance matrix is diagonal; while not essential, normality is often assumed and maximum likelihood estimation is used in estimation. In addition, inferential theory is based on the basic assumption that the sample covariance matrix is root-T consistent and asymptotically normal, see e.g, Anderson (1963, 1984) and Lawley and Maxwell (1971). These assumptions are no longer appropriate for large dimensional models. The iid assumption and diagonality of the idiosyncratic covariance matrix are too strong for economic time series data. The maximum likelihood estimation is not feasible for large-dimension factor models because of the size of the number of parameters to be estimated. Moreover, the consistency of the sample covariance matrix to the population $N \times N$ covariance matrix is not a well defined problem when N and T both approach infinity.

There is a growing literature that recognizes the limitations of the classical factor analysis and proposes new methodologies. Motivated by arbitrage pricing theory, Chamberlain and Rothschild (1983) introduced the notation of "approximate factor model" to allow for non-diagonal covariance matrix. Furthermore, Chamberlain and Rothschild argued that principal component analysis is equivalent to factor analysis (or maximum likelihood method under normality) when N increases to infinity. But they assumed a known $N \times N$ population covariance matrix. Connor and Korajczyk (1986,1988, 1993) studied the unknown covariance matrix and suggested that when N is much larger than T, the N-dimension factor model can be treated as a T-dimension model. They applied the principal component method to the $T \times T$ sample covariance matrix. This approach of computation has been used in most recent literature as in Stock and Watson (1998,1999), Bernanke and Boivin (2000), and others. Forni and Lippi (1997), Forni and Reichlin (1998) considered large dimension dynamic factor models and suggested different methods for estimation. Forni, Hallin, Lippi and Reichlin (2000a) formulated the dynamic principal component method by extending the analysis of Brillinger (1981).

Some preliminary estimation theory of large factor models has been obtained in the literature. Connor and Korajczyk (1986) proved consistency for the estimated factors with T fixed. For inference that requires large T, they used sequential limit argument (N converges to infinity first and then T converges to infinity)[1]. Stock and Watson (1999) studied the uniform consistency of estimated factors and derived certain rate of convergence for large N and large T. The rate of convergence was also studied by Bai and Ng (1999). Forni et al. (2000a,c) established consistency and certain rate of convergence for the estimated common components for dynamic factor models.

However, inferential theory is not well understood for large-dimension factor models. For example, limiting distributions are not available in the literature. In addition, the rates of convergence derived thus far are not the ones that would deliver a (non-degenerate) convergence in distribution. In this paper, we derive the rate of convergence and the limiting distributions for the estimated factors, factor loadings, and common components, estimated by the principal component method. Furthermore, the results are derived under more general setup than classical factor analysis. In addition to large N and large T, we allow for serial and cross-section dependence for the idiosyncratic errors; we also allow for heteroskedasticity in both dimensions. Under classical factor models, say N is fixed, one can only consistently estimate factor loadings but not the factors; see Anderson (1984). In contrast, we demonstrate that both the factors and factor loadings can be consistently estimated (up to a normalization) under the large-model paradigm.

We also consider the case of large N but fixed T. We show that to consistently estimate the factors, a necessary condition is asymptotic independence and asymptotic homoskedasticity in the time dimension. In contrast, under the framework of large N and large T, we establish not only consistency but consistency in the presence of serial correlation and heteroskedasticity.

The rest of the paper is organized as follows. Section 2 sets up the model, introduces notation and assumptions. Section 3 provides the asymptotic theory for the estimated factors, factor loadings, and common components. Section 4 provides additional results in the absence of serial correlation and heteroskedasticity. The case of fixed T is also studied. Section 5 derives consistent estimators for the covariance matrices occurring in the limiting distributions. Concluding remarks are provided in Section 6. All proofs are given in the appendices.

2 Preliminaries

Let X_{it} be the observed data for the i^{th} cross section unit at time t, for $i = 1, \ldots N$, and $t = 1, \ldots T$. Consider the following model

$$X_{it} = \lambda_i' F_t + e_{it} = C_{it} + e_{it}, \tag{1}$$

[1] Connor and Korajczyk (1986) recognized the importance of simultaneous limit theory. They stated that "Ideally we would like to allow N and T to grow simultaneously (possibly with their ratio approaching some limit). We know of no straightforward technique for solving this problem and leave it for future endeavors." Studying simultaneous limit is only a recent endeavor.

where F_t is a $r \times 1$ vector of common factors, λ_i is a $r \times 1$ vector of factor loadings, $C_{it} = \lambda_i' F_t$ is the common component, and e_{it} is the idiosyncratic component of X_{it}. None of the right hand side variables are observable and the only observable variable is X_{it}. In the context of arbitrage pricing theory, X_{it} represents the return of security i at period t; F_t represents the vector of factor returns; λ_i represents security i's exposure to the factors; and e_{it} represents the idiosyncratic component of returns. Although t is labelled as time, it has different meanings with different applications. For example, in the rank theory of consumer demands, t represents the t^{th} household and i represents the i^{th} consumption good.

The above factor models are used in many other applications. For example, the coincident index model of Stock and Watson (1989) (large T but a small N) and diffusion-index forecasting of Stock and Watson (1999) (large N and large T). They showed that an index of economic activity based on $N = 61$ series provides non-trivial improvements to inflation forecasts. Bernanke and Boivin (2000) considered the conduct of monetary policy when there is an abundance of information. In their analysis, the information from over 200 series is pooled with the factor model. Tong (2000) examined the profitability of momentum trading strategies and its linkage with systematic factors, estimated by the principal component method ($N = 311$ and $T = 408$). Forni et al. (2000b) studied business cycles among European countries with large-dimension data sets.

When N is small, the model can be cast under the state space setup and be estimated by maximizing the Gaussian likelihood via the Kalman filter. As N increases, the state space and the number of parameters to be estimated increase very quickly, rendering the estimation problem challenging, if not impossible. But factor models can also be estimated by the method of principal components. As shown by Chamberlain and Rothschild (1983), principal component estimator approaches to the maximum likelihood estimator when N increases (though they did not consider sampling variations). Yet the former is much easier to compute. Thus this paper focuses on the properties of principal component estimators.

Equation (1) can be written as an N-dimension time series

$$X_t = \Lambda F_t + e_t \quad (t = 1, 2, ..., T) \tag{2}$$

where $X_t = (X_{1t}, X_{2t}, ..., X_{Nt})'$, $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_N)'$, and $e_t = (e_{1t}, e_{2t}, ..., e_{Nt})'$. Alternatively, we can rewrite (1) as a T-dimension system with N observations:

$$\underline{X}_i = F\lambda_i + \underline{e}_i \quad (i = 1, 2, ..., N)$$

where $\underline{X}_i = (X_{i1}, X_{i2}, ..., X_{iT})'$, $F = (F_1, F_2, ..., F_T)'$ and $\underline{e}_i = (e_{i1}, e_{i2}, ..., e_{iT})'$. We will also use the matrix notation:

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

where $X = (\underline{X}_1, \underline{X}_2, ..., \underline{X}_N)$ is a $T \times N$ matrix of observed data and $e = (\underline{e}_1, \underline{e}_2, ..., \underline{e}_N)$ is a $T \times N$ matrix of idiosyncratic errors. The matrices Λ $(N \times r)$ and F $(T \times r)$ are both unknown.

Our objective is to derive the large sample properties of the unobserved factors and their loadings estimated with the method of principal components when N and T are both large.

The method of principal components minimizes

$$V(r) = \min_{\Lambda, F} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i' F_t)^2.$$

Concentrating out Λ and using the normalization that $F'F/T = I_r$ (a $r \times r$ identity matrix), the problem is identical to maximizing $tr(F'(XX')F)$. The estimated factor matrix, denoted by $\widetilde{F}$, is $\sqrt{T}$ times eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix XX', and $\widetilde{\Lambda}' = (\widetilde{F}'\widetilde{F})^{-1}\widetilde{F}'X = \widetilde{F}'X/T$ are the corresponding factor loadings. The common component matrix $F\Lambda'$ is estimated by $\widetilde{F}\widetilde{\Lambda}'$.

Because the factors and their loadings are identified only up to scale, the solution to the above minimization problem is not unique. Another solution is given by $(\bar{F}, \bar{\Lambda})$, where $\bar{\Lambda}$ is constructed as $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of the $N \times N$ matrix $X'X$. The normalization that $\bar{\Lambda}'\bar{\Lambda}/N = I_r$ implies $\bar{F} = X\bar{\Lambda}/N$. The common component matrix is estimated by $\bar{F}\bar{\Lambda}'$. The first estimate $(\widetilde{F}, \widetilde{\Lambda})$ is easier to compute when T is less than N and the second estimate $(\bar{F}, \bar{\Lambda})$ is easier to compute when T is larger than N. We introduce a third estimator, which is a normalized version of the second one:

$$\widehat{F} = \bar{F}(\bar{F}'\bar{F}/T)^{1/2}, \quad \text{and} \quad \widehat{\Lambda} = \bar{\Lambda}(\bar{F}'\bar{F}/T)^{-1/2}. \tag{3}$$

It is clear that the three estimates lead to the same estimated common components, i,e., $\widetilde{F}\widetilde{\Lambda}' = \bar{F}\bar{\Lambda}' = \widehat{F}\widehat{\Lambda}'$ and the same optimal objective function $V(r)$. The theoretical analysis will focus on the third estimator, which has been studied by Bai and Ng (1999) to certain extent.

Let $tr(A)$ denote the trace of A and $\|A\| = [tr(A'A)]^{1/2}$ denote its norm. Throughout, we let F_t^0 be the $r \times 1$ vector of true factors and λ_i^0 be the true loadings, with F^0 and Λ^0 being the corresponding matrices. The subsequent analysis assumes the number of factors (r) is known without loss of generality since Bai and Ng (1999) showed that r can be consistently estimated. For consistent estimation of r, the following assumptions are made:

Assumption A: Factors

$E\|F_t^0\|^4 \le M$ with $M < \infty$ and $T^{-1}\sum_{t=1}^{T} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$ for some positive definite matrix Σ_F.

Assumption B: Factor loadings

$\|\lambda_i\| \le \bar{\lambda} < \infty$, and $\|\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda\| \to 0$ for some $r \times r$ positive definite matrix Σ_Λ.

Assumption C: Time and cross-section dependence and heteroskedasticity

There exists a positive constant $M < \infty$ such that for all N and T,

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \le M$;

2. $E(e_s'e_t/N) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \le M$ for all s, and

$$T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)| \le M;$$

3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \le |\tau_{ij}|$ for some τ_{ij} and for all t. In addition,

$$N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \le M;$$

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T|\tau_{ij,ts}| \le M$;

5. For every (t,s), $E\Big|N^{-1/2}\sum_{i=1}^N[e_{is}e_{it} - E(e_{is}e_{it})]\Big|^4 \le M$.

Assumption D: Weak dependence between factors and idiosyncratic errors

$$E\Big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0 e_{it}\|^2\Big) \le M.$$

Assumption A is more general than that of classical factor analysis in which the factors F_t are i.i.d. Here we allow F_t to be dynamic such that $A(L)F_t = \epsilon_t$. However, we do not allow the dynamic enter into X_{it} directly, so that the relationship between X_{it} and F_t is still static. For more general dynamic factor models, readers are referred to Forni et al. (2000a). Assumption B ensures that each factor has a non-trivial contribution to the variance of X_t. We only consider non-random factor loadings for simplicity. Our results still hold when the λ_i's are random, provided they are independent of the factors and idiosyncratic errors, and $E\|\lambda_i\|^4 \le M$. Assumption C allows for limited time series and cross section dependence in the idiosyncratic components. Heteroskedasticities in both the time and cross section dimensions are also allowed. Under stationarity in the time dimension, $\gamma_N(s,t) = \gamma_N(s-t)$, though the condition is not necessary. Given Assumption C1, the remaining assumptions in C are easily satisfied if the e_{it} are independent for all i and t. The allowance for some correlation in the idiosyncratic components sets up a model to have an *approximate factor structure*. It is more general than a *strict factor model* which assumes e_{it} is uncorrelated across i, the framework on which the APT theory of Ross (1976) was based. Thus, the results to be developed will also apply to strict factor models. When the factors and idiosyncratic errors are independent (a standard assumption for conventional factor models), Assumption D is implied by Assumptions A and C. Independence is not required for D to be true. For example, suppose that $e_{it} = \epsilon_{it}\|F_t\|$ with ϵ_{it} being independent of F_t and ϵ_{it} satisfies Assumption C, then Assumption D holds.

Chamberlain and Rothschild (1983) defined an approximate factor model as having bounded eigenvalues for the $N \times N$ covariance matrix $\Omega = E(e_t e_t')$. If we assume e_t to be stationary with $E(e_{it}e_{jt}) = \tau_{ij}$, then from the matrix theory, the largest eigenvalue of Ω is bounded by $\max_i \sum_{j=1}^{N} |\tau_{ij}|$. Thus if we assume $\sum_{j=1}^{N} |\tau_{ij}| \leq M$ for all i and all N, which implies Assumption C3, then (2) will be an approximate factor model in the sense of Chamberlain and Rothschild. Since we also allow for non-stationarity (e.g., heteroskedasticity in the time dimension), our model in certain sense is more general than approximate factor models.

3 Asymptotic Theory

Assumptions $A - D$ are sufficient for consistently estimating the number of factors (r) as well as the factors and factor loadings, see Bai and Ng (1999). However, to derive their limiting distributions, we need additional assumptions.

Assumption E: Weak dependence. For some $M < \infty$ such that for all T and N,

1. For each t, $\sum_{s=1}^{T} |\gamma_N(s,t)| \leq M$.
2. For each i, $\sum_{k=1}^{N} |\tau_{ki}| \leq M$.

This assumption strengthens C2 and C3, respectively, and is still reasonable. For example, in the case of independence over time, $\gamma_N(s,t) = 0$ for $s \neq t$. Then Assumption E1 is equivalent to $\frac{1}{N}\sum_{i=1}^{N} E(e_{it}^2) \leq M$ for all t and N. Under cross-section independence, E2 is equivalent to $E(e_{it})^2 \leq M$, which is implied by Assumption C1.

Assumption F: Moments and central limit theorem. There exists an $M < \infty$ such that for all N and T

1. For each t,

$$E\left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{k=1}^{N} F_s^0 [e_{ks}e_{kt} - E(e_{ks}e_{kt})] \right\|^2 \leq M$$

2. The $r \times r$ matrix satisfies

$$E\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{k=1}^{N} F_t^0 \lambda_k^{0\prime} e_{kt} \right\|^2 \leq M$$

3. For each t, as $N \to \infty$,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 e_{it} \xrightarrow{d} N(0, \Gamma_t)$$

where $\Gamma_t = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i^0 \lambda_j^{0\prime} E(e_{it}e_{jt})$

4. For each i, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^0 e_{it} \xrightarrow{d} N(0, \Phi_i),$$

where $\Phi_i = \text{plim}_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E[F_t^0 F_s^{0\prime} e_{is} e_{it}]$.

Assumption F is not stringent because the sums in F1 and F2 involve zero mean random variables. The last two assumptions are simply central limit theorems, which are satisfied by various mixing processes.

Assumption G: The eigenvalues of the $r \times r$ matrix $(\Sigma_\Lambda \Sigma_F)$ are distinct.

The matrices Σ_F and Σ_Λ are defined in Assumptions A and B, respectively. Assumption G guarantees a unique limit for $(\widetilde{F}' F^0/T)$, which appears in the limiting distributions. Otherwise, its limit can only be determined up to orthogonal transformations. Similar assumption is made in classical factor analysis, see Anderson (1963). Note that this assumption is not needed for determining the number of factors. For example, in Bai and Ng (1999), the number of factors is determined based on the sum of squared residuals $V(r)$, which depends on the projection matrix $P_{\widetilde{F}}$. Projection matrices are invariant to orthogonal transformations. Also, Assumption G is not required for studying the limiting distribution of the estimated common components. The underlying reason is that the common components are identifiable. In the following analysis, we will use the fact that for positive definite matrices A and B, the eigenvalues of AB, BA and $A^{1/2}BA^{1/2}$ etc., all have the same set of eigenvalues.

Proposition 1 *Under Assumptions A-D and G, the following probability limit exists and*

$$\text{plim}_{T,N\to\infty} \frac{\widetilde{F}' F^0}{T} = Q.$$

The matrix Q is invertible and is given by $Q = V^{1/2}\Upsilon' \Sigma_\Lambda^{-1/2}$, where $V = diag(v_1, v_2, ..., v_r)$ and $v_1 > v_2 > \cdots > v_r > 0$ are the eigenvalues of $\Sigma_\Lambda^{1/2} \Sigma_F \Sigma_\Lambda^{1/2}$, and Υ is the corresponding eigenvector matrix such that $\Upsilon'\Upsilon = I_r$.

The proof is provided in the Appendix A. Under assumptions A-G, we shall establish asymptotic normality for the principal component estimators. Asymptotic theory for the principal component estimator exists only in the classical framework, where one of the dimensions is fixed. For example, Anderson (1963) showed asymptotic normality of the estimated principal components for large T and fixed N. Classical factor analysis always starts with the basic assumption that there exists a root-T consistent and asymptotically normal estimator for the underlying $N \times N$ covariance matrix of X_t (assuming N is fixed). The framework for classical factor analysis does not extend to situations when neither dimension of the data is fixed. This is because consistent estimation of the covariance matrix of X_t is not a well defined problem when N and T both approach infinity. Thus, our analysis is necessarily different from the classical approach.

3.1 Limiting Distribution of estimated factors

As noted earlier, F^0 and Λ^0 are not separately identifiable. However, they can be estimated up to an invertible $r \times r$ matrix transformation. As shown in the appendix, for the principal component estimator $\widehat{F}$, there exists an invertible matrix H_1 (whose dependence on N, T will be suppressed for notational simplicity) such that $\widehat{F}$ is an estimator of $F^0 H_1$ and $\widehat{\Lambda}$ is an estimator of $\Lambda^0(H_1')^{-1}$. In addition, $\widehat{F}\widehat{\Lambda}'$ is an estimator of $F^0\Lambda^{0\prime}$, the common components. It is clear that the common components are identifiable. It is also shown in the appendix that $\widetilde{F}$ is estimating $F^0 H_2$ for some $r \times r$ invertible matrix H_2. Likewise, $\widetilde{\Lambda}$ is estimating $\Lambda^0(H_2')^{-1}$. We also note that for many purposes, knowing $F^0 H_1$ is as good as knowing F^0. For example, in regression analysis, using F^0 as the regressor will give the same predicted value as using $F^0 H_1$ as the regressor. Because F^0 and $F^0 H_1$ span the same space, testing the significance of F^0 in a regression model containing F^0 as regressors is the same as testing the significance of $F^0 H_1$. For the same reason, the portfolio-evaluation measurements of Connor and Korajczyk (1986) will give valid results regardless of F^0 or $F^0 H_1$ being used.

Theorem 1 *Suppose that Assumptions A-G hold and that $\widehat{F}$ is defined in equation (3). For each t, as $N, T \to \infty$, we have*
(i) If $\sqrt{N}/T \to 0$, then

$$\begin{aligned}\sqrt{N}(\widehat{F}_t - H_1' F_t^0) &= \Big(\frac{\widetilde{F}'F^0}{T}\Big)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} + o_p(1) \\ &\xrightarrow{d} N(0, Q\Gamma_t Q'),\end{aligned}$$

where Q is defined in Proposition 1 and Γ_t is defined in F3.
(ii) If $\sqrt{N}/T \to \tau > 0$, possibly infinity, then

$$T(\widehat{F}_t - H_1' F_t^0) = O_p(1).$$

Theorem 1 implies the following result for $\widetilde{F}_t$:

Corollary 1 *Under the assumptions of Theorem 1, we have (i): If $\sqrt{N}/T \to 0$, then $\sqrt{N}(\widetilde{F}_t - H_2' F_t^0) \xrightarrow{d} N(0, V^{-1}Q\Gamma_t Q' V^{-1})$, where V is defined in Proposition 1; (ii): If $\sqrt{N}/T \to \tau$, then $T(\widetilde{F}_t - H_2' F_t^0) = O_p(1)$.*

A number of comments are in order.

1. The convergence rate is $\min\{\sqrt{N}, T\}$. When the factor loadings λ_i^0 $(i = 1, 2, ..., N)$ are all known, F_t^0 can be estimated by the cross-section least squares method and rate of convergence will be $\sqrt{N}$. The current rate of convergence reflects the fact that factor loadings are unknown and are also estimated. Under stronger assumptions, however, root-N rate is still achievable (see Section 4). Asymptotic normality is achieved by the central limit theorem as $N \to \infty$. Thus large N is required for this theorem.

2. While restrictions among N and T are needed, the theorem is not a sequential limiting result but a simultaneous one. In addition, the theorem not only holds for a particular relationship between N and T, it holds for many combinations of N and T. The restriction $\sqrt{N}/T \to 0$ is a weak one and is satisfied when $N = cT$ for $c \neq 0$, a constant. The result permits simultaneous inference for large N and large T. For example, the portfolio-performance measurement of Connor and Korajczyk (1986) can be analyzed without recoursing to sequential argument. See footnote 1.

3. The rate of convergence implied by this theorem is useful in regression analysis or in a forecasting equation involving estimated regressors such as

$$Y_{t+1} = \alpha F_t^0 + \beta W_t + u_{t+1}, \quad t = 1, 2, ..., T,$$

where Y_t and W_t are observable, but F_t^0 is not. However, F_t^0 can be replaced by $\widehat{F}_t$. Elementary calculation shows that the estimation effect in $\widehat{F}_t$ can be ignored as long as $\widehat{F}_t = H_1' F_t^0 + o_p(T^{-1/2})$ with H_1 having a full rank. This will be true if $N/T \to \infty$ by Theorem 1 because the convergence rate is either $\sqrt{N}$ or T. Thus if N is sufficiently large relative to T, F_t^0 can be treated as known. Stock and Watson (1998, 1999) considered such a framework of forecasting. Given the rate of convergence for $\widehat{F}_t$, it is easy to show that the forecast $Y_{T+1|T} = \widehat{\alpha}\widehat{F}_T + \widehat{\beta} W_T$ is $\sqrt{T}$-consistent for the conditional mean $E(Y_{T+1}|F_T^0, W_T)$, assuming the condition mean of u_{T+1} is zero. Furthermore, in constructing confidence intervals for the forecast, the estimation effect of $\widehat{F}_T$ can also be ignored provided that N is large relative to T. If $N/T \to \infty$ does not hold, say $N = cT$ $(c \neq 0)$, then the limiting distribution of the forecast will also depend on the limiting distribution of $\widehat{F}_T$ so the confidence interval must reflect this estimation effect. Theorem 1 allows us to account for this effect when constructing confidence intervals.

4. The covariance matrix of the limiting distribution depends on the correlation structure in the cross-section dimension. If e_{it} is independent across i, then

$$\Gamma_t = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_{it}^2 \lambda_i^0 \lambda_i^{0\prime}. \tag{4}$$

If, in addition, $\sigma_{it}^2 = \sigma_{jt}^2 = \sigma_t^2$ for all i, j, we have $\Gamma_t = \sigma_t^2 \Sigma_\Lambda$. In Section 5, we discuss consistent estimation of $Q\Gamma_t Q'$.

Before ending this section, we provide further asymptotic results. Part (ii) of Theorem 1 does not address the limiting distribution of $T(\widehat{F}_t - H_1' F_t^0)$. To derive its limiting distribution, we need an additional assumption.

Assumption H. As $N \to \infty$, $\gamma_N(s,t) \to \gamma(s,t)$ and $\sum_{s=1}^{\infty} |\gamma(s,t)| < \infty$ for each t.

Proposition 2 *Suppose that Assumptions A-H hold.*
(i) If $\sqrt{N}/T \to \infty$, *then*

$$T(\widehat{F}_t - H_1' F_t^0) \xrightarrow{d} (Q')^{-1} \sum_{s=1}^{\infty} F_s^0 \gamma(s,t).$$

(ii) If $\sqrt{N}/T \to \tau$ such that $0 < \tau < \infty$, then

$$T(\widehat{F}_t - H_1' F_t^0) \xrightarrow{d} (Q')^{-1} \sum_{s=1}^{\infty} F_s^0 \gamma(s,t) + \tau^{-1} N(0, Q\Gamma_t Q').$$

Note that the infinite series is well defined because the coefficients $\gamma(s,t)$ are absolutely summable and $E\|F_s\|^2 < M$ for all s by assumption. The two random variables in the limit of (ii) are independent when the idiosyncratic errors e_{it} are independent of factors F_s^0 for all i, t, and s. Usefulness of this result is the rate of convergence as well as the implication that (to be discussed further) serial correlation and heteroskedasticity slow down the rate of convergence.

3.2 Limiting Distribution of estimated factor loadings

The previous section shows that $\widehat{F}$ is estimating $F^0 H_1$. Now we show that $\widehat{\Lambda}$ is estimating $\Lambda^0 (H_1')^{-1}$. That is, $\widehat{\lambda}_i$ is estimating $H_1^{-1}\lambda_i^0$ for every i. Lehmann and Modest (1988) used the estimated loadings to construct various portfolios.

Theorem 2 *Suppose that Assumptions A-G hold and that $\widehat{\Lambda}$ is defined in (3). Then for each i, as $N, T \to \infty$, we have*

(i) If $\sqrt{T}/N \to 0$, then

$$\begin{aligned} \sqrt{T}(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0) &= H_1^{-1}\Big(\frac{F^{0\prime}F^0}{T}\Big)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^0 e_{it} + o_p(1) \\ &\xrightarrow{d} N(0, (Q'V)^{-1}\Phi_i (VQ)^{-1}), \end{aligned}$$

where V and Q are given in Proposition 1 and Φ_i in F4.

(ii) If $\sqrt{T}/N \to c$, where $c > 0$, possibly infinity,

$$N(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0) = O_p(1).$$

Similar to Corollary 1, Theorem 2 implies the limiting distribution for $\widetilde{\Lambda}$.

Corollary 2 *Under the assumptions of Theorem 2, we have (i): If $\sqrt{T}/N \to 0$, then $\sqrt{N}(\widetilde{\lambda}_i - H_2^{-1}\lambda_i^0) \xrightarrow{d} N(0, Q'^{-1}\Phi_i Q^{-1})$; (ii): If $\sqrt{T}/N \to c$, then $N(\widetilde{\lambda}_i - H_2^{-1}\lambda_i^0) = O_p(1)$.*

Some comments are appropriate here. First, the convergence rate is $\min\{\sqrt{T}, N\}$. When the factors F_t^0 $(t = 1, 2, ..., T)$ are all observable, λ_i^0 can be estimated by a time series regression with the i^{th} cross-section unit. In this case the best rate is $\sqrt{T}$. The new rate is due to the fact that F_t^0's are not observable and are also estimated. Second, an analogous result to Proposition 1 also holds for $\widehat{\lambda}_i$. The details are omitted. Third, if T and N are of the same order, then $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$. It follows from Theorems 1 and 2 that both $\widehat{F}_t$ and $\widehat{\lambda}_i$ are asymptotically normal.

3.3 Limiting distribution of estimated common components

The limit theory of estimated common components can be derived from the previous two theorems. Note that $C_{it}^0 = F_t^{0\prime}\lambda_i^0$ and $\widehat{C}_{it} = \widehat{F}_t'\widehat{\lambda}_i = \widetilde{F}_t'\widetilde{\lambda}_i$.

Theorem 3 *Suppose Assumptions A-F hold. As $N, T \to \infty$, we have*

(i) If $N/T \to 0$, then

$$\sqrt{N}(\widehat{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, V_{it})$$

where $V_{it} = \lambda_i^{0\prime}\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\lambda_i^0$, and Σ_Λ and Γ_t are defined earlier.

(ii) If $T/N \to 0$, then

$$\sqrt{T}(\widehat{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, W_{it})$$

where $W_{it} = F_t^{0\prime}\Sigma_F^{-1}\Phi_i\Sigma_F^{-1}F_t^0$, and Σ_F and Φ_i are defined earlier.

(iii) If $T/N \to \pi$, then

$$\sqrt{N}(\widehat{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, V_{it} + \pi W_{it}).$$

Thus the estimated common components are always asymptotically normal. The rate of convergence is $\min\{\sqrt{N}, \sqrt{T}\}$, which is the best rate possible. When F^0 is observable, the best rate for $\widehat{\lambda}_i$ is $\sqrt{T}$. When Λ^0 is observable, the best rate for $\widehat{F}_t$ is $\sqrt{N}$. It follows that when both are estimated, the best rate for $\widehat{\lambda}_i'\widehat{F}_t$ is the minimum of $\sqrt{N}$ and $\sqrt{T}$.

4 Stationary Idiosyncratic Errors

In the previous section, the rate of convergence for $\widehat{F}_t$ is shown to be $\min\{\sqrt{N}, T\}$. If T is fixed, it implies that $\widehat{F}_t$ (or $\widetilde{F}_t$) is not consistent. The result seems to be in conflict with that of Connor and Korajczyk (1986), who showed that the estimator $\widetilde{F}_t$ is consistent under fixed T. This inquiry leads us to the discovery of a necessary and sufficient condition for consistency under fixed T. Connor and Korajczyk imposed the following assumption:

$$\frac{1}{N}\sum_{i=1}^{N} e_{it}e_{is} \to 0, \quad t \neq s, \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N} e_{it}^2 \to \sigma^2 \quad \text{for all } t, \quad \text{as } N \to \infty. \tag{5}$$

We shall call the first condition asymptotic independence (or asymptotic uncorrelation) and the second condition asymptotic homoskedasticity. They established consistency under assumption (5). This assumption appears to be reasonable for assets returns and is commonly used in the finance literature, e.g., Campbell, Lo, and Mackinlay (1997). For many economic variables, violating one of the conditions seems to be easy. We show that assumption (5) is also necessary under fixed T.

Theorem 4 *Assume Assumptions A-G hold. Under a fixed T, a necessary and sufficient condition for consistency is asymptotic independence and asymptotic homoskedasticity.*

The implication is that, for fixed T, consistent estimation is not possible in the presence of serial correlation and heteroskedasticity. In contrast, under large T, we can still obtain consistent estimation. This result highlights the importance of large model framework. In addition, the allowance of serial correlation and heteroskedasticity is not only a more general assumption, but also a significant contribution.

Next, we show that previous results can also be strengthened under uncorrelation and homoskedasticity in the time dimension.

Assumption I: $E(e_{it}e_{is}) = 0$ if $t \neq s$, $Ee_{it}^2 = \sigma_i^2$, and $E(e_{it}e_{jt}) = \tau_{ij}$, for all t, i, and j.

Let $\bar{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^N \sigma_i^2$, which is a bounded sequence by Assumption C2. Let V_{NT} be the diagonal matrix consisting of the first r largest eigenvalues of the matrix $\frac{1}{TN}XX'$, arranged in decreasing order. Lemma A.3 shows $V_{NT} \xrightarrow{p} V$, a positive definite matrix. Define $D_{NT} = (V_{NT} - \frac{1}{T}\bar{\sigma}_N^2)^{-1}V_{NT}$, then $D_{NT} \xrightarrow{p} I_r$, as T and N converge to infinity. Define $H_3 = H_1 D_{NT}$.

Theorem 5 *Under Assumptions A-G and I, as $T, N \to \infty$, we have*

$$\sqrt{N}(\widehat{F}_t - H_3' F_t^0) \xrightarrow{d} N(0, Q\Gamma Q')$$

where $\Gamma = \text{plim}(\Lambda^{0\prime}\Omega\Lambda^0/N)$ *and* $\Omega = E(e_t e_t') = (\tau_{ij})$.

Note that cross-section correlation and heteroskedasticity are still allowed. Thus the result is for approximate factor models. This theorem does not require any restriction between N and T except they both converge to infinity. A similar result holds for $\widetilde{F}_t$. The rate of convergence ($N^{1/2}$) holds even for fixed T, but the limiting distribution is different.

If cross-section independence and cross-section homoskedasticity are assumed, then Theorem 2 part (i) holds without any restriction between N and T. However, cross-section homoskedasticity is unduly restrictive. Assumption I does not improve the result of Theorem 3, which already offers the best rate of convergence. Inspecting Proposition 2 and Theorem 5, we see that serial correlation and heteroskedasticity are responsible for the non-normality results of Proposition 2.

5 Estimating Covariance Matrices

In this section, we devise consistent estimators of the asymptotic variance-covariance matrices appeared in Theorems 1-3.

(1). *Covariance matrix of estimated factors.* This covariance matrix depends on the cross-section correlation of the idiosyncratic errors. Because the order of cross-sectional correlation is unknown, a HAC type estimator (see, Newey and West (1987)) is not feasible.

Thus we will assume cross-section independence for e_{it} $(i = 1, 2, ..., N)$. The asymptotic covariance of $\widehat{F}_t$ is given by $\Pi_t = Q\Gamma_t Q'$, where Γ_t is defined in (4). That is,

$$\Pi_t = \text{plim}\Big(\frac{\widetilde{F}'F^0}{T}\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\sigma_{it}^2\lambda_i^0\lambda_i^{0\prime}\Big)\Big(\frac{F^{0\prime}\widetilde{F}}{T}\Big).$$

This matrix involves the product $F^0\Lambda^{0\prime}$, which can be replaced by an estimate of (F^0, Λ^0). Note that any of the three estimators of (F^0, Λ^0) will give a numerically identical product. We choose $(\widetilde{F}, \widetilde{\Lambda})$. A consistent estimator of the covariance matrix is then given by

$$\widehat{\Pi}_t = \Big(\frac{\widetilde{F}'\widetilde{F}}{T}\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i'\Big)\Big(\frac{\widetilde{F}'\widetilde{F}}{T}\Big) = \Big(\frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i'\Big) \tag{6}$$

where $\widetilde{e}_{it} = X_{it} - \widetilde{\lambda}_i'\widetilde{F}_t$. Note that $\frac{\widetilde{F}'\widetilde{F}}{T} = I$.

(2). *Covariance matrix of estimated factor loadings.* The asymptotic covariance matrix of $\widehat{\lambda}_i$ is given by (see Theorem 2)

$$\Theta_i = (Q'V)^{-1}\Phi_i(VQ)^{-1}.$$

Let $\widetilde{\Phi}_i$ be a HAC estimator, as in Andrews and Monahan (1992), constructed with the series $\{\widetilde{F}_t \cdot \widetilde{e}_{it}\}$ $(t = 1, 2, ..., T)$. Because $\widetilde{F}_t$ estimates $H_2'F_t^0$, the HAC estimator $\widetilde{\Phi}_i$ is not directly estimating Φ_i but a transformation of Φ_i. Define

$$\widehat{\Theta}_i = V_{NT}^{-1}\widetilde{\Phi}_i V_{NT}^{-1}.$$

This estimator is consistent for Θ_i.

(3). *Covariance matrix of estimated common components.* Let

$$\widehat{V}_{it} = \widetilde{\lambda}_i'\Big(\frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i'\Big)\Big(\frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}\Big)^{-1}\widetilde{\lambda}_i,$$

$$\widehat{W}_{it} = \widetilde{F}_t'\Big(\frac{\widetilde{F}'\widetilde{F}}{T}\Big)^{-1}\widetilde{\Phi}_i\Big(\frac{\widetilde{F}'\widetilde{F}}{T}\Big)^{-1}\widetilde{F}_t \equiv \widetilde{F}_t'\widetilde{\Phi}_i\widetilde{F}_t.$$

as estimators of V_{it} and W_{it}, respectively.

Theorem 6 *Assume Assumptions A-G and cross-sectional independence. As, $T, N \to \infty$, $\widehat{\Pi}_t$, $\widehat{\Theta}_i$, $\widehat{V}_{it}$, and $\widehat{W}_{it}$ are consistent for Π_t, Θ_i, V_{it}, and W_{it}, respectively.*

One major point of this theorem is that all limiting covariances are easily estimable.

6 Concluding Remarks

This paper studies factor models under a nonstandard setting: large cross section and large time dimension. Such large-dimension factor models have received an increasing attention in the recent economic literature. This paper considers estimating the model by the principal component method, which is feasible and straightforward to implement. We derive some inferential theory concerning the estimators, including the rate of convergence and limiting distributions. In contrast with the classical factor analysis, we are able to consistently estimate both factors and factors loadings, not just factor loadings. In addition, our results are obtained under very general conditions, allowing for cross-section and time-dimension correlations and heteroskedasticities. We also identify a necessary condition for consistency under fixed T.

Many issues remain to be investigated. It is interesting to investigate the efficiency of the principal component estimator. This estimator is analogous to that of the least squares. A more efficient estimator is perhaps the counterpart of the generalized least squares. The question will then involve properly setting up the generalized least squares method and examining the properties of the estimator. These would include the convergence rate, limiting distribution and asymptotic efficiency. These problems differ from classical regression analysis because the factors and loadings are not observable.

Another broad area of research is empirical applications of the theoretical results derived in this paper. These results have important implications for empirical work. It is not necessary to divide a large sample into small subsamples in order to conform the fixed T requirement, as is done in the existing literature. In fact, a large time dimension delivers consistent estimates even under heteroskedasticity and serial correlation. In contrast, under small T, consistent estimation is not guaranteed. Some interesting applications in economics and finance have already been conducted, e.g., Bernanke and Boivin (2000) and Tong (2000). Many other existing applications that employed factor models can be reexamined with new data sets and a enlarged T dimension. It would be interesting to examine common cycles and comovements in the world economy on the lines of Forni at al. (2000b) and Gregory and Head (1999).

Recently, Granger (2001) and others called for large-model analysis to be one of the forefront items on the econometrics research agenda. It is our view that large-model analysis will become an important modelling paradigm as time progresses. For one thing, data-rich environments will be more pervasive as data collection, storage, and dissemination become more efficient and less costly. For a further example, the increasing interconnectedness in the world economy means that variables across different countries may have tighter linkage than ever before. All these, in combination with modern computing power, make large-model analysis more pertinent and feasible. We hope that the research of this paper will shed light on further development in the analysis of large models.

APPENDIX

Appendix A: Proof of Theorem 1

As in Bai and Ng (1999), we use the identity $\widehat{F} = \frac{1}{TN}XX'\widetilde{F}$. As defined in the main text, let V_{NT} be the $r \times r$ diagonal matrix of eigenvalues of $\frac{1}{TN}XX'$ in decreasing order. By the definition of eigenvectors and eigenvalues, $\frac{1}{TN}XX'\widetilde{F} = \widetilde{F}V_{NT}$. The left hand side is simply $\widehat{F}$. Thus we obtain immediately another identity linking $\widehat{F}$ and $\widetilde{F}$. That is, $\widehat{F} = \widetilde{F}V_{NT}$, or $\widehat{F}_t = V_{NT}\widetilde{F}_t$ $(t = 1, 2, ..., T)$.

Let $H_1 = (\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}\widetilde{F}/T)$ be a $r \times r$ matrix and $C_{NT} = \min(\sqrt{T}, \sqrt{N})$. Theorem 1 is based on the identity, also see Bai and Ng (1999):

$$\widehat{F}_t - H_1'F_t^0 = T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t) + T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} + T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} + T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}, \quad \text{(A.1)}$$

where

$$\begin{aligned} \zeta_{st} &= \frac{e_s'e_t}{N} - \gamma_N(s,t), \\ \eta_{st} &= F_s^{0\prime}\Lambda^{0\prime}e_t/N, \\ \xi_{st} &= F_t^{0\prime}\Lambda^{0\prime}e_s/N. \end{aligned} \quad \text{(A.2)}$$

To analyze each term above, we need the following lemmas.

Lemma A.1 *Under Assumptions A-D,*

$$C_{NT}^2\left(\frac{1}{T}\sum_{t=1}^{T}\|\widehat{F}_t - H_1'F_t^0\|^2\right) = O_p(1)$$

Proof: see Theorem 1 of Bai and Ng (1999).

Let $H_2 = H_1V_{NT}^{-1}$. Note that V_{NT} converges to a positive definite matrix (see Lemma A.3 below). Because $\widehat{F}_t = V_{NT}\widetilde{F}_t$ or $\widetilde{F}_t = V_{NT}^{-1}\widehat{F}_t$, Lemma A.1 leads to the following:

Corollary A.1 *Under Assumptions A-D, we have*

$$C_{NT}^2\left(\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{F}_t - H_2'F_t^0\|^2\right) = O_p(1).$$

From $\widetilde{F}'\widetilde{F}/T = I$ and Assumptions A and B, it follows that $\|H_1\| = O_p(1)$. Because V_{NT} has a positive definite matrix as its limit, we have $\|H_2\| = O_p(1)$. In addition, Lemma A.3 implies both H_1 and H_2 are of full rank.

Lemma A.2 *Under Assumptions A-F, we have*

(a). $T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t) = O_p(\frac{1}{\sqrt{T}C_{NT}})$;

(b). $T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} = O_p(\frac{1}{\sqrt{N}C_{NT}})$;

(c). $T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} = O_p(\frac{1}{\sqrt{N}})$;

(d). $T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st} = O_p(\frac{1}{\sqrt{N}C_{NT}})$.

Proof: Consider part (a). By adding and subtracting terms, we have

$$\begin{aligned}
T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t) &= T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0 + H_2'F_s^0)\gamma_N(s,t) \\
&= T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\gamma_N(s,t) + H_2'T^{-1}\sum_{s=1}^{T}F_s^0\gamma_N(s,t).
\end{aligned}$$

Now $\frac{1}{T}\sum_{s=1}^{T}F_s^0\gamma_N(s,t) = O_p(\frac{1}{T})$ since $E|\sum_{s=1}^{T}F_s^0\gamma_N(s,t)| \le (\max_s E\|F_s^0\|)\sum_{s=1}^{T}|\gamma_N(s,t)| \le M^{1+1/4}$ by Assumptions A and E1. Consider the first term:

$$|T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\gamma_N(s,t)| \le \Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\frac{1}{\sqrt{T}}\Big(\sum_{s=1}^{T}|\gamma_N(s,t)|^2\Big)^{1/2}$$

which is $O_p(\frac{1}{C_{NT}})\frac{1}{\sqrt{T}}O(1) = O_p(\frac{1}{\sqrt{T}C_{NT}})$ by Corollary A.1 and Assumption E1.
Consider part (b).

$$T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} = T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\zeta_{st} + H_2'T^{-1}\sum_{s=1}^{T}F_s^0\zeta_{st}.$$

For the first term,

$$\|T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\zeta_{st}\| \le \Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\zeta_{st}^2\Big)^{1/2}.$$

Furthermore,

$$\begin{aligned}
T^{-1}\sum_{s=1}^{T}\zeta_{st}^2 &= \frac{1}{T}\sum_{s=1}^{T}\Big[\frac{e_s'e_t}{N} - \gamma_N(s,t)\Big]^2 = \frac{1}{T}\sum_{s=1}^{T}\Big[\frac{e_s'e_t}{N} - \frac{E(e_s'e_t)}{N}\Big]^2 \\
&= \frac{1}{N}\frac{1}{T}\sum_{s=1}^{T}\Big[N^{-1/2}\sum_{i=1}^{N}(e_{is}e_{it} - E(e_{is}e_{it}))\Big]^2 = O_p(\frac{1}{N}).
\end{aligned}$$

Thus the first term is $O_p(\frac{1}{C_{NT}})O_p(\frac{1}{\sqrt{N}})$. Next, $T^{-1}\sum_{s=1}^{T}F_s^0\zeta_{st} = \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0(e_{is}e_{it} - E(e_{is}e_{it})) = O_p(\frac{1}{\sqrt{NT}})$ by Assumption F1. Thus

$$T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} = O_p(\frac{1}{C_{NT}}).O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{NT}}) = O_p(\frac{1}{\sqrt{N}C_{NT}}).$$

Consider part (c).

$$T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} = T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\eta_{st} + H_2'T^{-1}\sum_{s=1}^{T}F_s^0\eta_{st}.$$

From $T^{-1}\sum_{s=1}^{T}F_s^0\eta_{st} = (\frac{1}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime})\frac{1}{N}\sum_{k=1}^{N}\lambda_k e_{kt} = O_p(\frac{1}{\sqrt{N}})$. The first term is

$$\|T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\eta_{st}\| \leq \Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\cdot\Big(\frac{1}{T}\sum_{s=1}^{T}\eta_{st}^2\Big)^{1/2}.$$

The first expression is $O_p(1/C_{NT})$ by Corollary A.1. For the second expression,

$$T^{-1}\sum_{s=1}^{T}\eta_{st}^2 = T^{-1}\sum_{s=1}^{T}(F_s^{0\prime}\Lambda^{0\prime}e_t/N)^2 \leq \|\Lambda^{0\prime}e_t/N\|^2T^{-1}\sum_{s=1}^{T}\|F_s^0\|^2 = O_p(\frac{1}{N}),$$

since $\frac{1}{T}\sum_{s=1}^{T}\|F_s^0\|^2 = O_p(1)$, and $\|\Lambda^{0\prime}e_t/\sqrt{N}\|^2 = O_p(1)$ Thus, (c) is $O_p(\frac{1}{\sqrt{N}})$.

Finally for part (d),

$$\begin{aligned}T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st} &= T^{-1}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\Lambda^{0\prime}e_s/N = T^{-1}\sum_{s=1}^{T}(\widetilde{F}_se_s'\Lambda^0/N)F_t^0\\ &= \frac{1}{NT}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)e_s'\Lambda^0F_t^0 + \frac{1}{NT}\sum_{s=1}^{T}H_2'F_s^0e_s'\Lambda^0F_t^0.\end{aligned}$$

Consider the first term

$$\begin{aligned}\|(\frac{1}{NT}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)e_s'\Lambda^0)F_t^0\| &\leq \frac{1}{\sqrt{N}}\Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\Big\|\frac{e_s'\Lambda}{\sqrt{N}}\Big\|^2\Big)^{1/2}\|F_t^0\|\\ &= O_p(\frac{1}{\sqrt{N}})\cdot O_p(\frac{1}{C_{NT}})\cdot O_p(1) = O_p(\frac{1}{\sqrt{N}C_{NT}})\end{aligned}$$

following arguments analogous to those used in part (c). For the second term of (d),

$$\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}F_s^0e_s'\Lambda^0F_t^0 = \frac{1}{\sqrt{NT}}\Big(\sum_{s=1}^{T}\sum_{k=1}^{N}F_s^0\lambda_k^{0\prime}e_{ks}\Big)F_t^0 = O_p(1),$$

by Assumption F2. Thus, $N^{-1}T^{-1}\sum_{s=1}^{T}H_2'F_s^0e_s'\Lambda^0F_t^0 = O_p(\frac{1}{\sqrt{NT}})$ and (d) is $O_p(\frac{1}{\sqrt{N}C_{NT}})$. The proof of Lemma A.2 is complete.

By the definition of eigenvalues and eigenvectors, we have $\frac{1}{TN}XX'\widetilde{F} = \widetilde{F}V_{NT}$. From $\widetilde{F}'\widetilde{F}/T = I$, we further have $T^{-1}\widetilde{F}'\frac{1}{TN}XX'\widetilde{F} = V_{NT}$.

Lemma A.3 *Assume Assumptions A-D hold. As, $T, N \to \infty$,*
(i) $T^{-1}\widetilde{F}'(\frac{1}{TN}XX')\widetilde{F} = V_{NT} \xrightarrow{p} V$
(ii) $\frac{\widetilde{F}'F^0}{T}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\frac{F^{0\prime}\widetilde{F}}{T} \xrightarrow{p} V$
where V is the diagonal matrix consisting of the eigenvalues of $\Sigma_\Lambda\Sigma_F$.

This lemma is implicitly proved by Stock and Watson (1999). The details are omitted.

Because V is positive definite, the lemma says that $\frac{F^{0\prime}\widetilde{F}}{T}$ is of full rank for all large T and N and thus is invertible. We also note that Lemma A.3(ii) shows that a quadratic form of $\frac{F^{0\prime}\widetilde{F}}{T}$ has a unique limit. But this does not guarantee $\frac{F^{0\prime}\widetilde{F}}{T}$ itself has a unique limit unless Assumption G is made. In what follows, an eigenvector matrix of W refers to the matrix whose columns are the eigenvectors of W with unit length and the i^{th} column corresponds to the i^{th} largest eigenvalue.

Proof of Proposition 1: Let V_{NT} be as stated earlier, then $\frac{1}{TN}XX'\widetilde{F} \equiv \widetilde{F}V_{NT}$. Multiplying this identity on both sides by $T^{-1}(\frac{\Lambda'\Lambda^0}{N})^{1/2}F^{0\prime}$, we have

$$\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}T^{-1}F^{0\prime}\left(\frac{XX'}{TN}\right)\widetilde{F} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\widetilde{F}}{T}\right)V_{NT}.$$

Expanding XX' with $X = F^0\Lambda^{0\prime} + e$, we can rewrite above as

$$\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\widetilde{F}}{T}\right) + d_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\widetilde{F}}{T}\right)V_{NT} \qquad \text{(A.3)}$$

where $d_{NT} = \left(\frac{\Lambda 0\prime\Lambda^0}{N}\right)^{1/2}\left[\left(\frac{F^{0\prime}F^0}{T}\right)\Lambda^{0\prime}e'\widetilde{F}/(TN) + \frac{1}{TN}F^{0\prime}e\Lambda^0F^{0\prime}\widetilde{F}/T + \frac{1}{TN}F^{0\prime}ee'\widetilde{F}/T\right] = o_p(1)$. The $o_p(1)$ is implied by Lemma A.2. Let

$$B_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}F^0}{T}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}$$

and

$$R_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\widetilde{F}}{T}\right), \qquad \text{(A.4)}$$

then we can rewrite (A.3) as

$$[B_{NT} + d_{NT}R_{NT}^{-1}]R_{NT} = R_{NT}V_{NT}.$$

Thus each column R_{NT}, though not unit length, is an eigenvector of the matrix $[B_{NT} + d_{NT}R_{NT}^{-1}]$. Let V_{NT}^* be a diagonal matrix consisting of the diagonal elements of $R_{NT}'R_{NT}$. Denote $\Upsilon_{NT} = R_{NT}{V_{NT}^*}^{-1/2}$ so that each column of Υ_{NT} has a unit length, and we have

$$[B_{NT} + d_{NT}R_{NT}^{-1}]\Upsilon_{NT} = \Upsilon_{NT}V_{NT}.$$

Thus Υ_{NT} is the eigenvector matrix of $[B_{NT}+d_{NT}R_{NT}^{-1}]$. Note that $B_{NT}+d_{NT}R_{NT}^{-1}$ converges to $B = \Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ by Assumptions A and B and $d_{NT} = o_p(1)$. Because the eigenvalues of B

are distinct by Assumption G, the eigenvalues of $B_{NT}+d_{NT}R_{NT}^{-1}$ will also be distinct for large N and large T by the continuity of eigenvalues. This implies that the eigenvector matrix of $B_{NT}+d_{NT}R_{NT}^{-1}$ is unique except that each column can be replaced by the negative of itself. In addition, the kth column of R_{NT} (see (A.4)) depends on $\widetilde{F}$ only through the kth column of $\widetilde{F}$ $(k=1,2,...,r)$. Thus the sign of each column in R_{NT} and thus in $\Upsilon_{NT}=R_{NT}V_{NT}^{*\,-1/2}$ is implicitly determined by the sign of each column in $\widetilde{F}$. Thus, given the column sign of $\widetilde{F}$, Υ_{NT} is uniquely determined. By the eigenvector perturbation theory (which requires the distinctness of eigenvalues, see Franklin (1968)), there exists a unique eigenvector matrix Υ of $B=\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ such that $\|\Upsilon_{NT}-\Upsilon\|=o_p(1)$. From $\frac{F^{0\prime}\widetilde{F}}{T}=\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2}\Upsilon_{NT}V_{NT}^{*\,1/2}$, we have $\frac{F^{0\prime}\widetilde{F}}{T}\xrightarrow{p}\Sigma_\Lambda^{-1/2}\Upsilon V^{1/2}$ by Assumption B and by $V_{NT}^*\xrightarrow{p}V$ in view of Lemma A.3(ii).

Proof of Theorem 1:

Case 1: $\sqrt{N}/T\to 0$. By Lemma A.2, we have:

$$\widehat{F}_t-H_1'F_t^0=O_p\Big(\frac{1}{\sqrt{T}C_{NT}}\Big)+O_p\Big(\frac{1}{\sqrt{N}C_{NT}}\Big)+O_p\Big(\frac{1}{\sqrt{N}}\Big)+O_p\Big(\frac{1}{\sqrt{N}C_{NT}}\Big).$$

The limiting distribution is determined by the third term of the right hand side of (A.1) because it is the dominating term. Using the definition of η_{si},

$$\sqrt{N}(\widehat{F}_t-H_1'F_t^0)\;=\;T^{-1}\sum_{s=1}^{T}(\widetilde{F}_sF_s^{0\prime})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0e_{it}+o_p(1).$$

Now $\frac{1}{\sqrt{N}}\sum_{i=1}^N\lambda_i^0e_{it}\xrightarrow{d}N(0,\Gamma_t)$ by Assumption F3. Together with Proposition 1, we have $\sqrt{N}(\widehat{F}_t-H_1'F_t)\xrightarrow{d}N(0,Q\Gamma_tQ')$ as stated.

Case 2: $\sqrt{N}/T\to\tau$. If $\tau=\infty$, then the first term of (A.1) is the dominating term. We have $T(\widehat{F}_t-H_1'F_t^0)=TO_p(\frac{1}{\sqrt{T}C_{NT}})=O_p(1)$ since $C_{NT}=T$. If $0<\tau<\infty$, then the first and the third terms have the same order of magnitude. In this case the third term is also $O_p(1)$ because $T(\widehat{F}_t-H_1'F_t^0)=O_p(1)+O_p(T/\sqrt{N})=O_p(1)$ in view of $T/\sqrt{N}\to 1/\tau$.

Proof of Corollary 1. From $\widetilde{F}_t=V_{NT}^{-1}\widehat{F}_t$ and $H_2'=V_{NT}^{-1}H_1'$, we have $\sqrt{N}(\widetilde{F}_t-H_2'F_t^0)=V_{NT}^{-1}\sqrt{N}(\widehat{F}_t-H_1'F_t^0)$. The corollary follows immediately from Theorem 1 and $V_{NT}\to V$ by Lemma A.3.

To prove Proposition 1, we need the following lemma.

Lemma A.4 *Under Assumptions A-E,*

$$\max_{1\le t\le T}\|\widehat{F}_t-H_1'F_t^0\|=O_p(T^{-1/2})+O_p((T/N)^{1/2}).$$

Proof: We consider each term on the right hand side of (A.1). For the first term,

$$\max_t T^{-1}\|\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t)\|\le T^{-1/2}\Big(T^{-1}\sum_{s=1}^{T}\|\widetilde{F}_s\|^2\Big)^{1/2}\max_t\Big(\sum_{s=1}^{T}\gamma_N(s,t)^2\Big)^{1/2}. \qquad (A.5)$$

The above is $O_p(T^{-1/2})$ follows from $T^{-1}\sum_{s=1}^{T}\|\widetilde{F}_s\|^2 = O_p(1)$ and $\sum_{s=1}^{T}\gamma_N(s,t)^2 \le M_1$ for some $M_1 < \infty$ uniformly in t. The remaining three terms of (A.1) are each $O_p((T/N)^{1/2})$ uniformly in t. To see this, let $\nu_t = T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}$. It suffices to show $\max_t\|\nu_t\|^2 = O_p(T/N)$. But Bai and Ng (1999) proved that $\sum_{t=1}^{T}\|\nu_t\|^2 = O_p(T/N)$ (they used notation b_t instead of $\|\nu_t\|^2$). Bai and Ng also obtained the same result for the third and the fourth terms.

Remark: This lemma gives an upper bound on the maximum deviation of the estimated factors to the true ones (up to a transformation). The bound is not the sharpest possible because we essentially use the argument that $\max_t\|\widehat{F}_t - H_1'F_t^0\| \le \sum_{t=1}^{T}\|\widehat{F}_t - H_1'F_t^0\|$. Note that if $N \ge cT^2$, for some $c > 0$, then the maximum deviation is $O_p(T^{-1/2})$, which is actually a very strong result.

Corollary A.2 *Under Assumptions A-E, we have*

$$\max_{1\le t\le T}\|\widetilde{F}_t - H_2'F_t^0\| = O_p(T^{-1/2}) + O_p((T/N)^{1/2}).$$

Proof of Proposition 2. First we prove that Lemma A.2 (a) can be strengthened to

$$T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t) = T^{-1}H_2'\sum_{s=1}^{T}F_s^0\gamma_N(s,t) + O_p(T^{-3/2}). \tag{A.6}$$

From the proof of Lemma A.2 (a), it is sufficient to show that $\frac{1}{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\gamma_N(s,t) = O_p(\frac{1}{T^{3/2}})$. Corollary A.2 implies $\max_{1\le s\le T}\|\widetilde{F}_s - H_2'F_s^0\| = O_p(T^{-1/2})$ since $\sqrt{N}/T \to \tau$. Now

$$\begin{aligned}\|T^{-1}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\gamma_N(s,t)\| &\le T^{-1}\max_{1\le s\le T}\|\widetilde{F}_s - H_2'F_s^0\|\Big(\sum_{s=1}^{T}|\gamma_N(s,t)|\Big)\\ &= T^{-1}O_p(T^{-1/2})O(1) = O_p(T^{-3/2}).\end{aligned}$$

Case 1: $\sqrt{N}/T \to \infty$. In this case, the first term of (A.1) is the dominating term. We have $T(\widehat{F}_t - H_1'F_t^0) = H_2'\sum_{s=1}^{T}F_s^0\gamma_N(s,t) + o_p(1) \to V^{-1}Q\Sigma_\Lambda\sum_{s=1}^{\infty}F_s^0\gamma(s,t)$, by (A.6), Assumption H, and by $H_2' \to V^{-1}Q\Sigma_\Lambda = Q'^{-1}$. The last equality follows from the definition of Q.

Case 2: $\sqrt{N}/T \to \tau$. The first and the third terms of (A.1) are of the same order of magnitude. We have

$$T(\widehat{F}_t - H_1'F_t^0) = H_2'\sum_{s=1}^{T}F_s^0\gamma_N(s,t) + \frac{T}{\sqrt{N}}\Big(\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_sF_s^{0\prime}\Big)\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it}\Big) + o_p(1).$$

The limiting distribution is obtained by combining results from case 1 and Theorem 1.

Appendix B: Proof of Theorem 2

To prove Theorem 2, we need some preliminary results.

Lemma B.1 *Under Assumptions A-F,* $T^{-1}(\widehat{F}-F^0H_1)'\underline{e}_i = O_p(\frac{1}{C_{NT}^2})$.

Proof: From the identity (A.1), we have

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(\widehat{F}_t - H_1'F_t^0)e_{it} &= T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t)e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}e_{it} + \\ & \quad T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st}e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}e_{it} \\ &= I + II + III + IV. \end{aligned}$$

We begin with I, which can be rewritten as

$$I = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\gamma_N(s,t)e_{it} + T^{-2}H_2'\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0\gamma_N(s,t)e_{it}.$$

The first term is bounded by

$$T^{-1}\Big(\frac{1}{T}\|\sum_{s=1}^{T}\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\Big(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_N(s,t)|^2e_{it}^2\Big)^{1/2} = T^{-1}O_p(C_{NT}^{-1})O_p(1),$$

where the $O_p(1)$ follows from $Ee_{it}^2 \le M$ and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)|^2 \le M$ by Lemma 1(i) of Bai and Ng (1999). The expected value of the second term of I is bounded by (ignore H_2)

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_N(s,t)|(E\|F_s^0\|^2)^{1/2}(Ee_{it}^2)^{1/2} \le MT^{-1}\Big(\cdot T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_N(s,t)|\Big) = O(T^{-1})$$

by Assumption C2. Thus I is $O_p(T^{-1})$. For II, we rewrite it as

$$II \;=\; T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\zeta_{st}e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0\zeta_{st}e_{it}.$$

The second term is $O_p(\frac{1}{\sqrt{NT}})$ by Assumption F1. To see this, the second term can be written as $\frac{1}{\sqrt{NT}}(\frac{1}{T}\sum_{t=1}^{T}z_te_{it})$ with $z_t = \frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{k=1}^{N}F_s^0[e_{ks}e_{kt} - E(e_{ks}e_{kt})]$. By F1, $E\|z_t\|^2 < M$. Thus $E\|z_te_{it}\| \le (E\|Z_t\|^2Ee_{it}^2)^{1/2} \le M$. This implies $\frac{1}{T}\sum_{t=1}^{T}z_te_{it} = O_p(1)$. For the first term, we have

$$\|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\zeta_{st}e_{it}\| \;\le\; \Big(\frac{1}{T}\|\sum_{s=1}^{T}\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\Big(\frac{1}{T}\sum_{t=1}^{T}\zeta_{st}e_{it}\Big)^2\Big)^{1/2}.$$

But

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{st}e_{it} = \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}[e_{ks}e_{kt} - E(e_{ks}e_{kt})]\Big)e_{it} = O_p(N^{-1/2}).$$

So the first term is $O_p(\frac{1}{C_{NT}})\cdot O_p(\frac{1}{\sqrt{N}}) = O_p(\frac{1}{C_{NT}\sqrt{N}})$. Thus $II = O_p(\frac{1}{C_{NT}\sqrt{N}})$.
For III, we rewrite it as

$$III = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)\eta_{st}e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}H_2'F_s^0\eta_{st}e_{it}.$$

The first term is bounded by

$$\Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\cdot\Big(\frac{1}{T}\sum_{s=1}^{T}\Big(\frac{1}{T}\sum_{t=1}^{T}\eta_{st}e_{it}\Big)^2\Big)^{1/2} = O_p(\frac{1}{C_{NT}})O_p(\frac{1}{\sqrt{N}})$$

because $\frac{1}{T}\sum_{t=1}^{T}\eta_{st}e_{it} = \frac{1}{\sqrt{N}}F_s^{0\prime}\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k e_{kt}\Big)e_{it}$, which is $O_p(N^{-1/2})$. The second term of III can be written as

$$H_2'\Big(\frac{1}{T}\sum_{s=1}^{T}F_s^0F_s^{0\prime}\Big)\Big(\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}\lambda_k e_{kt}e_{it}\Big) \tag{B.1}$$

which is $O_p((NT)^{-1/2})$ if cross-section independence holds for the $e's$. Under weak cross-sectional dependence as in Assumption E2, the above is $O_p((NT)^{-1/2}) + O_p(N^{-1})$. This follows from $e_{kt}e_{it} = e_{kt}e_{it} - \tau_{ki,t} + \tau_{ki,t}$, where $\tau_{ki,t} = E(e_{kt}e_{it})$. We have $\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}|\tau_{kit}| \le \frac{1}{N}\sum_{k=1}^{N}\tau_{ki} = O(N^{-1})$ by E2, where $|\tau_{ki,t}| \le \tau_{ki}$. In summary, III is $O_p(\frac{1}{C_{NT}\sqrt{N}}) + O(N^{-1})$. The proof for IV is similar to that of III. Thus $I + II + III + IV = O_p(\frac{1}{T}) + O_p(\frac{1}{C_{NT}\sqrt{N}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N}) = O_p(\frac{1}{C_{NT}^2})$.

Lemma B.2 *Under Assumptions A-F, the $r\times r$ matrix* $T^{-1}(\widehat{F} - F^0H_1)'F^0 = O_p(\frac{1}{C_{NT}^2})$.

Proof: Using the identity (A.1), we have

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}(\widehat{F}_t - H_1'F_t^0)F_t^{0\prime} &= T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\gamma_N(s,t) + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\zeta_{st} + \\
&\quad T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\eta_{st} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\xi_{st} \\
&= I + II + III + IV.
\end{aligned}$$

Term I is $O_p(T^{-1})$. The proof is the same as that of I of Lemma B.1. Next,

$$II = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H_2'F_s^0)F_t^{0\prime}\zeta_{st} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}H_2'F_s^0F_t^{0\prime}\zeta_{st}.$$

The first term is $O_p(\frac{1}{C_{NT}\sqrt{N}})$ following arguments analogous to Lemma B.1. The second term is $O_p(\frac{1}{\sqrt{NT}})$ by Assumption $F1$ and the Cauchy-Schwarz inequality. Thus, $II = O_p(\frac{1}{C_{NT}\sqrt{N}})$. Next, note that $III = T^{-2}\sum_{t=1}^T\sum_{s=1}^T(\widetilde{F}_s - H_2'F_s^0)F_t^{0\prime}\eta_{st} + T^{-2}\sum_{t=1}^T\sum_{s=1}^T H_2'F_s^0F_t^{0\prime}\eta_{st}$. Now $T^{-2}\sum_{t=1}^T\sum_{s=1}^T H_2'F_s^0F_t^{0\prime}\eta_{st} = H_2'(\frac{1}{T}\sum_{s=1}^T F_s^0F_s^{0\prime})\frac{1}{TN}\sum_{t=1}^T\sum_{k=1}^N \lambda_k F_t' e_{kt} = O_p(1)O_p(\frac{1}{\sqrt{NT}})$ by Assumption F2. Consider

$$\|T^{-2}\sum_{t=1}^T\sum_{s=1}^T(\widetilde{F}_s - H_2'F_s^0)F_t^{0\prime}\eta_{st}\| \leq \Big(\frac{1}{T}\sum_{s=1}^T\|\widetilde{F}_s - H_2'F_s^0\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^T\Big\|\frac{1}{T}\sum_{t=1}^T F_t^{0\prime}\eta_{st}\Big\|^2\Big)^{1/2}.$$

The second term can be rewritten as

$$\Big(\frac{1}{T}\sum_{s=1}^T\Big\|\frac{1}{NT}\sum_{t=1}^T F_t^{0\prime}\sum_{k=1}^N F_s^{0\prime}\lambda_k e_{kt}\Big\|^2\Big)^{1/2} = \frac{1}{\sqrt{NT}}\Big(\frac{1}{T}\sum_{s=1}^T\Big\|F_s^{0\prime}\frac{1}{\sqrt{NT}}\sum_{t=1}^T\sum_{k=1}^N\lambda_k F_t^{0\prime}e_{kt}\Big\|^2\Big)^{1/2}$$

which is $O_p(\frac{1}{\sqrt{NT}})$ by F2. Therefore, $T^{-2}\sum_{t=1}^T\sum_{s=1}^T(\widetilde{F}_s - H_2'F_s^0)F_t^{0\prime}\eta_{st} = O_p(\frac{1}{C_{NT}})\cdot O_p(\frac{1}{\sqrt{NT}})$. Thus, $III = O_p(\frac{1}{C_{NT}})\cdot O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{\sqrt{NT}}) = O_p(\frac{1}{\sqrt{NT}})$. The proof for IV is similar to that of III. Thus $I + II + III + IV = O_p(\frac{1}{T}) + O_p(\frac{1}{C_{NT}\sqrt{N}}) + O_p(\frac{1}{\sqrt{NT}}) = O_p(\frac{1}{C_{NT}^2})$.

Lemma B.3 *Under Assumptions A-F,*

$$(\widehat{F}'\widehat{F})^{-1}\widehat{F}'\underline{e}_i = \Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}H_1'\frac{1}{T}\sum_{s=1}^T F_s^0 e_{is} + O_p(\frac{1}{C_{NT}^2}).$$

Proof:

$$\begin{aligned}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'\underline{e}_i &= (\widehat{F}'\widehat{F})^{-1}(\widehat{F} - F^0H_1 + F^0H_1)'\underline{e}_i \\ &= (\widehat{F}'\widehat{F})^{-1}(\widehat{F} - F^0H_1)'\underline{e}_i + (\widehat{F}'\widehat{F})^{-1}(F^0H_1)'\underline{e}_i.\end{aligned}$$

Now $(\frac{\widehat{F}'\widehat{F}}{T}) = O_p(1)$, and $T^{-1}(\widehat{F} - F^0H_1)'\underline{e}_i = O_p(\frac{1}{C_{NT}^2})$ by Lemma B.1. Furthermore,

$$(\widehat{F}'\widehat{F})^{-1}(F^0H_1)'\underline{e}_i = \Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}H_1'\frac{1}{T}\sum_{s=1}^T F_s^0 e_{is}.$$

Lemma B.4 *Under Assumptions A-F,* $(\widehat{F}'\widehat{F})^{-1}\widehat{F}'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0 = O_p(\frac{1}{C_{NT}^2})$.

Proof:

$$\begin{aligned}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0 &= (\widehat{F}'\widehat{F})^{-1}(\widehat{F} - F^0H_1 + F^0H_1)'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0 \\ &= (\widehat{F}'\widehat{F})^{-1}(\widehat{F} - F^0H_1)'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0 \\ &\quad +(\widehat{F}'\widehat{F})^{-1}(F^0H_1)'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0.\end{aligned}$$

Thus

$$\begin{aligned}\|(\widehat{F}'\widehat{F})^{-1}\widehat{F}'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0\| &\leq \|\Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}\| \cdot \Big(T^{-1}\sum_{t=1}^{T}\|\widehat{F}_t - H_1'F_t^0\|^2\Big)\|H_1^{-1}\lambda_i^0\| \\ &+\|\Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}\| \cdot \|H_1\| \cdot \|\frac{F^{0\prime}(F^0H_1 - \widehat{F})}{T}\| \cdot \|H_1^{-1}\lambda_i^0\|.\end{aligned}$$

The first term is $O_p(\frac{1}{C_{NT}^2})$ by Lemma A.1 and the second term is $O_p(\frac{1}{C_{NT}^2})$ by Lemma B.2.

Proof of Theorem 2. The estimator $\widehat{\lambda}_i$ has an alternative expression: $\widehat{\lambda}_i = (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\underline{X}_i$, where $\underline{X}_i = F^0\lambda_i^0 + \underline{e}_i$. That is, $\widehat{\lambda}_i$ is OLS estimator of λ_i^0 using $\widehat{F}$ as the regressor and $\underline{X}_i$ as the dependent variable. Using this representation simplifies the proof significantly. Thus

$$\begin{aligned}\widehat{\lambda}_i &= (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\big[\widehat{F}H_1^{-1}\lambda_i^0 + \underline{e}_i + (F^0 - \widehat{F}H_1^{-1})\lambda_i^0\big] \\ &= H_1^{-1}\lambda_i^0 + (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\underline{e}_i + (\widehat{F}'\widehat{F})^{-1}\widehat{F}'(F^0 - \widehat{F}H_1^{-1})\lambda_i^0.\end{aligned}$$

Thus

$$\widehat{\lambda}_i - H_1^{-1}\lambda_i^0 = (\widehat{F}'\widehat{F})^{-1}\widehat{F}'\underline{e}_i + (\widehat{F}'\widehat{F})^{-1}\widehat{F}'(F^0H_1 - \widehat{F})H_1^{-1}\lambda_i^0.$$

By Lemmas B.3 and B.4,

$$\widehat{\lambda}_i - H_1^{-1}\lambda_i^0 = \Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}H_1'\frac{1}{T}\sum_{s=1}^{T}F_s^0e_{is} + O_p(\frac{1}{C_{NT}^2}). \tag{B.2}$$

Case 1: $\sqrt{T}/N \to 0$. Then $\sqrt{T}/C_{NT}^2 \to 0$ and thus

$$\sqrt{T}(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0) = \Big(\frac{\widehat{F}'\widehat{F}}{T}\Big)^{-1}H_1'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^0e_{is} + o_p(1).$$

Note $(\frac{\widehat{F}'\widehat{F}}{T})^{-1}H_1' = H_1^{-1}(\frac{F^{0\prime}F^0}{T})^{-1} + o_p(1)$ and $H_1^{-1}(\frac{F^{0\prime}F^0}{T})^{-1} = [(\frac{F^{0\prime}F^0}{T})(\frac{\Lambda^{0\prime}\Lambda^0}{N})(\frac{F^{0\prime}\widetilde{F}}{T})]^{-1} \to (\Sigma_F\Sigma_\Lambda Q')^{-1}$, which is equal to $(Q'V)^{-1}$, see Proposition 1. Together with Assumption F4, the desired limiting distribution is obtained.

Case 2 $\sqrt{T}/N \to c$ with $c > 0$. The first term on the right hand side of (B.2) is $O_p(T^{-1/2})$, and the second term is $O_p(N^{-1})$ in view of $N << T$ and thus $C_{NT}^2 = O(N)$. Thus

$$N(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0) = O_p(N/\sqrt{T}) + O_p(1) = O_p(1)$$

because $N/\sqrt{T} \geq 1/c$.

Proof of Corollary 2. The mathematical identity $\widetilde{\Lambda} = \widehat{\Lambda}V_{NT}$ can be shown to hold. That is, $\widetilde{\lambda}_i = V_{NT}\widehat{\lambda}_i$. From $H_2 = H_1V_{NT}^{-1}$, or $H_2^{-1} = V_{NT}H_1^{-1}$, we have $\sqrt{T}(\widetilde{\lambda}_i - H_2^{-1}\lambda_i^0) = V_{NT}\sqrt{T}(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0)$. The corollary follows immediately from Theorem 2 and $V_{NT} \to V$.

Appendix C: Proof of Theorem 3

From $C_{it}^0 = F_t^{0\prime}\lambda_i^0$ and $\widehat{C}_{it} = \widehat{F}_t'\widehat{\lambda}_i$, we have

$$\widehat{C}_{it} - C_{it}^0 = (\widehat{F}_t - H_1'F_t^0)'H_1^{-1}\lambda_i^0 + \widehat{F}_t'(\widehat{\lambda}_i - H_1^{-1}\lambda_i^0).$$

(i): If $\frac{N}{T} \to 0$ then $\frac{\sqrt{N}}{T} \to 0$,

$$\sqrt{N}(\widehat{C}_{it} - C_{it}^0) = \sqrt{N}(\widehat{F}_t - H_1'F_t^0)'H_1^{-1}\lambda_i^0 + O_p((N/T)^{1/2}).$$

By Theorem 1,

$$\begin{aligned}\sqrt{N}(\widehat{F}_t - H_1'F_t^0) &= \Big(\frac{\widetilde{F}'F^0}{T}\Big)\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k e_{kt} + o_p(1)\\ &= \Big(\frac{\widetilde{F}'F^0}{T}\Big)\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k^0 e_{kt} + o_p(1)\\ &= H_1'\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k^0 e_{kt} + o_p(1)\end{aligned}$$

by the definition of H_1. Therefore,

$$\begin{aligned}\lambda_i^{0\prime}(H_1')^{-1}\sqrt{N}(\widehat{F}_t - H_1'F_t^0) &= \lambda_i^{0\prime}\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)^{-1}\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k^0 e_{kt} + o_p(1)\\ &\xrightarrow{d} N(0, \lambda_i^{0\prime}\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\lambda_i^0).\end{aligned}$$

(ii) If $\frac{T}{N} \to 0$ then $\frac{\sqrt{T}}{N} \to 0$. By Theorem 2,

$$\begin{aligned}\sqrt{T}(\widehat{C}_{it} - C_{it}^0) &= o_p(1) + \sqrt{T}\widehat{F}_t'(\widehat{\lambda}_i - H_1^{-1}\lambda_i)\\ &= o_p(1) + F_t^{0\prime}(F^{0\prime}F^0/T)^{-1}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s^0 e_{is}\\ &\xrightarrow{d} N(0, F_t^{0\prime}\Sigma_F^{-1}\Phi_i\Sigma_F^{-1}F_t^0).\end{aligned}$$

(iii) If $N/T \to \pi$,

$$\sqrt{N}(\widehat{C}_{it} - C_{it}^0) = \sqrt{N}(\widehat{F}_t - H_1'F_t^0)'H_1^{-1}\lambda_i^0 + (N/T)^{1/2}\sqrt{T}\widehat{F}_t'(\widehat{\lambda}_i - H_1^{-1}\lambda_i).$$

Assumptions A-F imply the asymptotic independence of $\widehat{F}_t$ and $\widehat{\lambda}_i$. Combining the results from cases 1 and 2 and noting the asymptotic independence, we have

$$\sqrt{N}(\widehat{C}_{it} - C_{it}^0)\xrightarrow{d}N(0, V_{it}) + \sqrt{\pi}N(0, W_{it}) = N(0, V_{it} + \pi W_{it}).$$

Appendix D: Proof of Theorems 4 and 5

To prove Theorem 4, we need the following lemma. In the following, I_k denotes the $k \times k$ identity matrix.

Lemma D.1 *Let A be a $T \times r$ matrix ($T \geq r$) with $rank(A) = r$, and let Ω be a semipositive definite matrix of $T \times T$. If for every A, there exist r eigenvectors of $AA' + \Omega$, denoted by Γ ($T \times r$), such that $\Gamma = AC$ for some $r \times r$ invertible matrix C, then $\Omega = cI_T$ for some $c \geq 0$.*

Note that if $\Omega = cI_T$ for some c, then the r eigenvectors corresponding to the first r largest eigenvalues are of the form AC. This is because AA' and $AA' + cI_T$ have the same set of eigenvectors, and the first r eigenvectors of AA' is of the form AC. Thus $\Omega = cI_T$ is a necessary and sufficient condition for AC to be the eigenvectors of $AA' + \Omega$ for every A.

Proof: We consider two cases, $r = 1$ and $r > 1$.

Consider $r = 1$. Let η_i be the $T \times 1$ vector with the ith element being 1 and 0 elsewhere. For example, $\eta_1 = (1, 0, ..., 0)'$. Let $A = \eta_1$. The lemma's assumption implies that η_1 is an eigenvector of $AA' + \Omega$. That is, for some scalar a (it can be shown that $a > 0$),

$$(\eta_1\eta_1' + \Omega)\eta_1 = \eta_1 a.$$

The above implies $\eta_1 + \Omega_1 = \eta_1 a$, where Ω_1 is the first column of Ω. This in turn implies that all elements, with possible exception for the first element, of Ω_1 are zero. Apply the same reasoning with $A = \eta_i$ ($i = 1, 2, ..., T$), we conclude Ω is a diagonal matrix such that $\Omega = diag(c_1, c_2, ..., c_T)$. Next we argue the constants c_i must be the same. Let $A = \eta_1 + \eta_2 = (1, 1, 0, ..0)'$. The lemma's assumption implies that $\Gamma = (1/\sqrt{2}, 1/\sqrt{2}, 0, ..., 0)'$ is an eigenvector of $AA' + \Omega$. It is easy to verify that $(AA' + \Omega)\Gamma = \Gamma d$ for some scalar d implies that $c_1 = c_2$. Similiar reasoning shows all constants are the same.

Next consider $r > 1$. Let G be an arbitrary $r \times r$ orthogonal matrix. Consider $A = (G', 0)'$. The assumption implies that $\Gamma = ((GC)', 0)'$ is a $T \times r$ matrix of eigenvectors of $AA' + \Omega$. From $\Gamma'\Gamma = I_r$, we see that GC itself is an orthogonal matrix. Partition Ω conformably into a matrix of blocks, with the first row blocks are Ω_{11} and Ω_{12}, the second row blocks are Ω_{21} and Ω_{22}. From $(AA' + \Omega)\Gamma = \Gamma D$, where D is diagonal matrix of $r \times r$, we have

$$\begin{pmatrix} GC \\ 0 \end{pmatrix} + \begin{pmatrix} \Omega_{11}GC \\ \Omega_{21}GC \end{pmatrix} = \begin{pmatrix} GC \\ 0 \end{pmatrix} D \tag{D.1}$$

This implies $\Omega_{21} = 0$ and by symmetry $\Omega_{12} = 0$. Thus Ω is a block diagonal matrix. Furthermore, multiplying Γ' on both sides of (D.1), we get

$$I + C'G'\Omega_{11}GC = D,$$

or equivalently, $C'G'\Omega_{11}GC = D - I$. This implies that Ω_{11} is diagonalizable by an arbitrary orthogonal matrix GC. This is true if and only if $\Omega_{11} = cI_r$. If $T = r$, then the desired result follows. If $T > r$, let $A = (0', G', 0')'$, where the first 0 is a $1 \times r$ vector (G is shifted downward by only one row). Partition Ω into a 3×3 block matrix, and denote the resulting blocks by Ω_{ij}^*. Using the same reasoning, we get $\Omega_{22}^* = c_2 I_r$ and all off-diagonal blocks are zeros. Because $r > 1$, Ω_{11} and Ω_{22}^* have at least one overlapping diagonal element. Thus $c = c_2$. Continue this argument until A is of the form $(0', G')'$ and $\Omega = cI_T$ is proved at the same time.

Proof of Theorem 4. In this proof, all limits are taken as $N \to \infty$. Let $\Psi = \text{plim}_{N\to\infty} N^{-1} ee' = \text{plim} N^{-1} \sum_{i=1}^N e_i e_i'$. That is, the $(t, s)^{th}$ entry of Ψ is the limit of $\frac{1}{N}\sum_{i=1}^N e_{it}e_{is}$. From

$$(TN)^{-1}XX' = T^{-1}F^0(\Lambda^{0\prime}\Lambda^0/N)F^{0\prime} + T^{-1}F^0(\Lambda^{0\prime}e'/N) + T^{-1}(e\Lambda^0/N)F^{0\prime} + T^{-1}(ee'/N),$$

we have $\frac{1}{TN}XX' \xrightarrow{p} B$ with $B = \frac{1}{T}F^0\Sigma_\Lambda F^{0\prime} + \frac{1}{T}\Psi$ because the two middle terms converge to zero. We shall argue that consistency of $\widetilde{F}$ for some transformation of F^0 implies $\Psi = \sigma^2 I_T$. That is, equation (5) holds. Let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_T$ be the eigenvalues of B with $\mu_r > \mu_{r+1}$. Thus, it is assumed that the first r eigenvalues are well seperated with the remaining ones. Without this assumption, it can be shown that consistency is not possible. Let Γ be the $T \times r$ matrix of eigenvectors corresponding to the r largest eigenvalues of B. Because $\frac{1}{TN}XX' \xrightarrow{p} B$, it follows that $\|P_{\widetilde{F}} - P_\Gamma\| = \|\widetilde{F}\widetilde{F}' - \Gamma\Gamma'\| \xrightarrow{p} 0$. This follows from the continuity property of an invariance space when the associated eigenvalues are separated with the rest of eigenvalues, see, e.g., Bhatia (1997). If $\widetilde{F}$ is consistent for some transformation of F^0, that is, $\|\widetilde{F} - F^0D\| \to 0$ with D being a $r \times r$ invertible matrix, then $\|P_{\widetilde{F}} - P_{F^0}\| \to 0$, where $P_{F^0} = F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$. Since the limit of $P_{\widetilde{F}}$ is unique, we have $P_\Gamma = P_{F^0}$. This implies that $\Gamma = F^0C$ for some $r \times r$ invertible matrix C. Since consistency requires this be true for every F^0, not just a particular F^0, we see the existence of r eigenvectors of B in the form of F^0C for all F^0. Apply Lemma D.1 with $A = F^0\Sigma_\Lambda^{1/2}$ and $\Omega = T^{-1}\Psi$, we obtain $\Psi = cI_T$ for some c, which implies condition (5).

Proof of Theorem 5.

By the definition of $\widetilde{F}$ and V_{NT}, we have $\frac{1}{NT}XX'\widetilde{F} = \widetilde{F}V_{NT}$. Since W and $W + cI$ have the same set of eigenvectors for an arbitrary matrix W, we have

$$\Big(\frac{1}{NT}XX' - T^{-1}\bar{\sigma}_N^2 I_T\Big)\widetilde{F} = \widetilde{F}(V_{NT} - T^{-1}\bar{\sigma}_N^2 I_r).$$

Right multiply $D_{NT} = (V_{NT} - T^{-1}\bar{\sigma}_N^2 I_r)^{-1}V_{NT}$ on both sides and use $\widehat{F} = \widetilde{F}V_{NT}$, we obtain

$$\Big(\frac{1}{NT}XX' - T^{-1}\bar{\sigma}_N^2 I_T\Big)\widetilde{F}D_{NT} = \widehat{F}$$

Expanding XX' and note that $\frac{1}{TN}ee' - T^{-1}\bar{\sigma}_N^2 I_T = \frac{1}{TN}[ee' - E(ee')]$, we have

$$\widehat{F}_t - H_3'F_t^0 = D_{NT}T^{-1}\sum_{s=1}^T \widetilde{F}_s\zeta_{st} + D_{NT}T^{-1}\sum_{s=1}^T \widetilde{F}_s\eta_{st} + D_{NT}T^{-1}\sum_{s=1}^T \widetilde{F}_s\xi_{st}, \qquad \text{(D.2)}$$

Equation (D.2) is similar to (A.1). But the first term on the left side of (A.1) disappears and each of the remaining three terms is multiplied by D_{NT}. The middle term of (D.2) is of $O_p(N^{-1/2})$ and each of the other two terms is $O_p(N^{-1/2}C_{NT}^{-1})$ by Lemma A.2. The rest of proof is the same as that of case 1 of Theorem 1, except no restriction between N and T is needed. In addition, $\Gamma_t = \Gamma = \lim \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^0\lambda_j^{0\prime}E(e_{it}e_{jt}) = \lim \frac{1}{N}\Lambda^{0\prime}\Omega\Lambda^0$.

Appendix E: Proof of Theorem 6

First we show $\widehat{\Pi}_t$ is consistent for Π_t. A detailed proof will invole four steps: (i) $\frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' - \frac{1}{N}\sum_{i=1}^{N}e_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' = o_p(1)$; (ii) $\frac{1}{N}\sum_{i=1}^{N}e_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' - H_2^{-1}(\frac{1}{N}\sum_{i=1}^{N}e_{it}^2\lambda_i^0\lambda_i^{0\prime})(H_2')^{-1} = o_p(1)$; (iii) $\frac{1}{N}\sum_{i=1}^{N}e_{it}^2\lambda_i^0\lambda_i^{0\prime} - \frac{1}{N}\sum_{i=1}^{N}\sigma_{it}^2\lambda_i^0\lambda_i^{0\prime} = o_p(1)$; (iv) $H_2^{-1} \to Q$. The first two steps imply that $\widetilde{e}_{it}^2$ can be replaced by e_{it}^2 and $\widetilde{\lambda}_i$ can be replaced by $H_2^{-1}\lambda_i^0$. A rigorous proof for (i) and (ii) can be given, but the details are omitted here (a proof is available from the author). Heuristically, (i) and (ii) follows from $\widetilde{e}_{it} = e_{it}^2 + O_p(C_{NT}^{-1})$ and $\widetilde{\lambda}_i = H_2^{-1}\lambda_i^0 + O_p(C_{NT}^{-1})$, which are the consequences of Theorem 3 and Corollary 2, respectively. The result of (iii) is a special case of White (1980), and (iv) is proved below. Combining these results together with (4), we obtain $\frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' \xrightarrow{p} Q\Gamma_t Q' = \Pi_t$.

Next, we prove $\widehat{\Theta}_i$ is consistent for Θ_i. Because $\widetilde{F}_t$ is estimating $H_2'F_t^0$, the HAC estimator $\widetilde{\Phi}_i$ based on $\widetilde{F}_t\widetilde{e}_{it}$ $(t = 1, 2, ..., T)$ is estimating $H_2^{0\prime}\Phi_i H_2^0$, where H_2^0 is the limit of H_2. The consistency of $\widetilde{\Phi}_i$ for $H_2^{0\prime}\Phi_i H_2^0$ can be proved using the argument of Andrews and Monahan (1992). Now from $H_2' = V_{NT}^{-1}H_1' = V_{NT}^{-1}(\widetilde{F}'F^0/T)(\Lambda^{0\prime}\Lambda/N) \to V^{-1}Q\Sigma_\Lambda$. The latter matrix is equal to Q'^{-1} (see Proposition 1). Thus $\widetilde{\Phi}_i$ is estimating $Q'^{-1}\Phi_i Q^{-1}$. This in turn implies that $V_{NT}^{-1}\widetilde{\Phi}_i V_{NT}^{-1} \to V^{-1}Q'^{-1}\Phi_i Q^{-1}V^{-1} = \Theta_i$ because $V_{NT} \to V$.

Next, we argue $\widehat{V}_{it}$ is consistent for V_{it}. First note that cross-section independence is assumed because V_{it} involves Γ_t. We also note that V_{it} is simply the limit of

$$\lambda_i^{0\prime}\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)^{-1}\Big(\frac{1}{N}\sum_{k=1}^{N}e_{kt}^2\lambda_k^0\lambda_k^{0\prime}\Big)\Big(\frac{\Lambda^{0\prime}\Lambda^0}{N}\Big)^{-1}\lambda_i^0$$

The above expression is scale free in the sense that an identical value will be obtained when replacing λ_j^0 by $A\lambda_j^0$ (for all j), where A is a $r \times r$ invertible matrix. The consistency of $\widehat{V}_{it}$ now follows from the consistency of $\widetilde{\lambda}_j$ for $H_2^{-1}\lambda_j^0$.

Finally, we argue the consistency of $\widehat{W}_{it} = \widetilde{F}_t'\widetilde{\Phi}_i\widetilde{F}_t$ for W_{it}. The consistency of $\widetilde{\Phi}_i$ for $Q'^{-1}\Phi_i Q^{-1}$ is already proved. Now, $\widetilde{F}_t'$ is estimating $F_t^{0\prime}H_2$ but $H_2 \to Q^{-1}$. Thus $\widehat{W}_{it}$ is consistent for $F_t^{0\prime}Q^{-1}Q'^{-1}\Phi_i Q^{-1}Q'^{-1}F_t^0 \equiv W_{it}$ because $Q^{-1}Q'^{-1} = \Sigma_F^{-1}$ (see Proposition 1). This completes the proof of Theorem 6.

References

Anderson, T. W. (1963), Asymptotic Theory for Principal Component Analysis, *Annals of Mathematical Statistics* 34, 122–148.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Andrews, D.W.K. and J.C. Monahan (1992), An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, 953-966.

Bai, J. and Ng, S. (1999), Determining the Number of Factors in Approximate Factor Models. Department of Economics, Boston College. Forthcoming in *Econometrica.*

Bernanke, B. and Boivin, J. (2000), Monetary Policy in a Data Rich Environment. Department of Economics, Princeton University.

Bhatia, R. (1997), *Matrix Analysis.* New York: Springer.

Brillinger, D.R. (1981), *Time Series Data Analysis and Theory*, Hold, Rinehart and Winston Inc.

Campbell, C.J. A.W. Lo, and A.C. Mackinlay (1997). *The Econometrics of Financial Markets*, New Jersey, Princeton University Press.

Chamberlain, G. and Rothschild, M. (1983), Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets, *Econometrica* 51, 1305–1324.

Connor, G. and Korajczyk, R. (1986), Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis, *Journal of Financial Economics* 15, 373–394.

Connor, G. and Korajczyk, R. (1988), Risk and Return in an Equilibrium APT: Application to a New Test Methodology, *Journal of Financial Economics* 21, 255–289.

Connor, G. and Korajczyk, R. (1993), A Test for the Number of Factors in an Approximate Factor Model, *Journal of Finance* XLVIII:4, 1263–1291.

Forni, M. and Lippi, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Oxford University Press, Oxford, U.K.

Forni, M. and Reichlin, L. (1998), Let's Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics, *Review of Economic Studies* 65, 453–473.

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

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000b), Reference Cycles: The NBER Methodology Revisited, CEPR Discussion Paper 2400.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000c), The Generalized Dynamic Factor Model: Consistency and rates. Unpublished manuscript.

Franklin, J.E. (1968), *Matrix Theory*. New Jersey: Prentice-Hall Inc.

Granger, C.W.J. (2001), Macroeconometrics – Past and future. *Journal of Econometrics*, 100, 17-19.

Gregory, A. and Head, A. (1999), Common and Country-Specific Fluctuations in Productivity, Investment, and the Current Account, *Journal of Monetary Economics* 44:3, 423–452.

Lawley, D.N, and A.E. Maxwell (1971). *Factor Analysis as a Statistical Method.* London: Butterworth.

Lehmann, B. N. and Modest, D. (1988), The Empirical Foundations of the Arbitrage Pricing Theory, *Journal of Financial Economics* 21, 213–254.

Lewbel, A. (1991), The Rank of Demand Systems: Theory and Nonparametric Estimation, *Econometrica* **59**, 711–730.

Newey, W.K. and K.D. West (1987), A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance* **13**, 341–360.

Stock, J. H. and Watson, M. (1989), New Indexes of Coincident and Leading Economic Indications, *in* O. J. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual 1989*, M. I. T. Press, Cambridge.

Stock, J. H. and Watson, M. W. (1998), Diffusion Indexes, NBER Working Paper 6702.

Stock, J. H. and Watson, M. W. (1999), Forecasting Inflation, *Journal of Monetary Economics* 44, 293–335.

Tong, Y. (2000), When are momentum profits due to common factor dynamics? Department of Finance, Boston College.

White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica* 48, 817-838.