ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

WILEY

INFERENTIAL THEORY FOR FACTOR MODELS
OF LARGE DIMENSIONS

By Jushan Bai[1]

This paper develops an inferential theory for factor models of large dimensions. The principal components estimator is considered because it is easy to compute and is asymptotically equivalent to the maximum likelihood estimator (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 within the framework of large cross sections (N) and a large time dimension (T), to which classical factor analysis does not apply.

We show that the estimated common components are asymptotically normal with a convergence rate equal to the minimum of the square roots of N and T. The estimated factors and their loadings are generally normal, although not always so. 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. Stronger results are obtained when the idiosyncratic errors are serially uncorrelated and homoskedastic. A necessary and sufficient condition for consistency is derived for large N but fixed T.

KEYWORDS: Approximate factor models, principal components, common components, large model analysis, large data sets, data-rich environment.

1. INTRODUCTION

ECONOMISTS NOW HAVE the luxury of working with very large data sets. 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. Such a data-rich environment is due in part to technological advances in data collection, and in part to the inevitable accumulation of information over time. A useful method for summarizing information in large data sets is factor analysis. More importantly, many economic problems are characterized by factor models; e.g., the Arbitrage Pricing Theory of Ross (1976). While there is a well developed inferential theory for factor models of small dimensions

(classical), the inferential theory for factor models of large dimensions is absent. The purpose of this paper is to partially fill this void.

A factor model has the following representation:

$$(1) \qquad X_{it} = \lambda_i' F_t + e_{it},$$

where X_{it} is the observed datum for the ith cross section at time $t(i = 1, \ldots N;$ $t = 1, \ldots T)$; F_t is a vector $(r \times 1)$ of common factors; λ_i is a vector $(r \times 1)$ of factor loadings; and e_{it} is the idiosyncratic component of X_{it}. The right-hand-side variables are not observable.[2] Readers are referred to Wansbeek and Meijer (2000) for an econometric perspective on factor models. Classical factor analysis assumes a fixed N, while T is allowed to increase. In this paper, we develop an inferential theory for factor models of large dimensions, allowing both N and T to increase.

1.1. Examples of Factor Models of Large Dimensions

(i) *Asset Pricing Models.* A fundamental assumption of the Arbitrage Pricing Theory (APT) of Ross (1976) is that asset returns follow a factor structure. In this case, X_{it} represents asset i's return in period t; F_t is a vector of factor returns; and e_{it} is the idiosyncratic return. This theory leaves the number of factors unspecified, though fixed.

(ii) *Disaggregate Business Cycle Analysis.* Cyclical variations in a country's economy could be driven by global or country-specific shocks, as analyzed by Gregory and Head (1999). Similarly, variations at the industry level could be driven by country-wide or industry-specific shocks, as analyzed by Forni and Reichlin (1998). Factor models allow for the identification of common and specific shocks, where X_{it} is the output of country (industry) i in period t; F_t is the common shock at t; and λ_i is the exposure of country (industry) i to the common shocks. This factor approach of analyzing business cycles is gaining prominence.

(iii) *Monitoring, Forecasting, and Diffusion Indices.* The state of the economy is often described by means of a small number of variables. In recent years, Forni, Hallin, Lippi, and Reichlin (2000b), Reichlin (2000), and Stock and Watson (1998), among others, stressed that the factor model provides an effective way of monitoring economic activity. This is because business cycles are defined as co-movements of economic variables, and common factors provide a natural representation of these co-movements. Stock and Watson (1998) further demonstrated that the estimated common factors (diffusion indices) can be used to improve forecasting accuracy.

(iv) *Consumer Theory.* Suppose X_{ih} represents the budget share of good i for household h. The rank of a demand system is the smallest integer r such that $X_{ih} = \lambda_{i1} G_1(e_h) + \cdots \lambda_{ir} G_r(e_h)$, where e_h is household h's total expenditure, and $G_j(\cdot)$ are unknown functions. By defining the r factors as

---

[2] It is noted that N also represents the number of variables and T represents the number of observations, an interpretation used by classical factor analysis.

$F_h = [G_1(e_h) \ldots G_r(e_h)]'$, the rank of the demand system is equal to the number of factors. An important result due to Gorman (1981) is that demand systems consistent with standard axioms of consumer theory cannot have a rank more than three, meaning that households' allocation of resources if they are utility maximizing should be described by at most three factors. This theory is further generalized by Lewbel (1991).

Applications of large-dimensional factor models are rapidly increasing. Bernanke and Boivin (2000) showed that using the estimated factors makes it possible to incorporate large data sets into the study of the Fed's monetary policy. Favero and Marcellino (2001) examined a related problem for European countries. Cristadoro, Forni, Reichlin, and Veronese (2001) estimated a core inflation index from a large number of inflation indicators. As is well known, factor models have been important in finance and are used for performance evaluations and risk measurement; see, e.g., Campbell, Lo, and Mackinlay (1997, Chapters 5 and 6). More recently, Tong (2000) studied the profitability of momentum trading strategies using factor models.

In all the above and in future applications, it is of interest to know when the estimated factors can be treated as known. That is, under what conditions is the estimation error negligible? This question arises whenever the estimated factors are used as regressors and the significance of the factors is tested, as for example in Bernanke and Boivin (2000), and also in the diffusion index forecasting of Stock and Watson (1998). If the estimation error is not negligible, then the distributional properties of the estimated factors are needed. In addition, it is always useful to construct confidence intervals for the estimates, especially for applications in which the estimates represent economic indices. This paper offers answers to these and related theoretical questions.

1.2. *Limitations of Classical Factor Analysis*

Let $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$ and $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$ be $N \times 1$ vectors, and let $\Sigma = \mathrm{cov}(X_t)$ be the covariance matrix of X_t. Classical factor analysis assumes that N is fixed and much smaller than T, and further that the e_{it} are independent and identically distributed (iid) over time and are also independent across i. Thus the variance-covariance matrix of e_t, $\Omega = E(e_t e_t')$, is a diagonal matrix. The factors F_t are also assumed to be iid and are independent of e_{it}. Although not essential, normality of e_{it} 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

$$\widehat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (X_t - \overline{X})(X_t - \overline{X})'$$

is root-T consistent for Σ and asymptotically normal; see Anderson (1984) and Lawley and Maxwell (1971).

These assumptions are restrictive for economic problems. First, the number of cross sections (N) is often larger than the number of time periods (T) in economic data sets. Potentially important information is lost when a small number of variables is chosen (to meet the small-N, requirement). Second, the assumption that $\sqrt{T}(\hat{\Sigma} - \Sigma)$ is asymptotically normal may be appropriate under a fixed N, but it is no longer appropriate when N also tends to infinity. For example, the rank of $\hat{\Sigma}$ does not exceed $\min\{N, T\}$, whereas the rank of Σ can always be N. Third, the iid assumption and diagonality of the idiosyncratic covariance matrix Ω, which rules out cross-section correlation, are too strong for economic time series data. Fourth, maximum likelihood estimation is not feasible for large-dimensional factor models because the number of parameters to be estimated is large. Fifth, classical factor analysis can consistently estimate the factor loadings (λ_i's) but not the common factors (F_t's). In economics, it is often the common factors (representing the factor returns, common shocks, diffusion indices, etc.) that are of direct interest.

1.3. *Recent Developments and the Main Results of this Paper*

There is a growing literature that recognizes the limitations of classical factor analysis and proposes new methodologies. Chamberlain and Rothschild (1983) introduced the notation of an "approximate factor model" to allow for a non-diagonal covariance matrix. Furthermore, Chamberlain and Rothschild showed that the principal component method is equivalent to factor analysis (or maximum likelihood under normality of the e_{it}) when N increases to infinity. But they assumed a known $N \times N$ population covariance matrix. Connor and Korajczyk (1986, 1988, 1993) studied the case of an unknown covariance matrix and suggested that when N is much larger than T, the factor model can be estimated by applying the principal components method to the $T \times T$ covariance matrix. Forni and Lippi (1997) and Forni and Reichlin (1998) considered large dimensional dynamic factor models and suggested different methods for estimation. Forni, Hallin, Lippi, and Reichlin (2000a) formulated the dynamic principal components 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 a sequential limit argument (N goes to infinity first and then T goes to infinity).[3] Stock and Watson (1999) studied the uniform consistency of estimated factors and derived some rates of convergence for large N and large T. The rate of convergence was also studied by Bai and Ng (2002). Forni et al. (2000a,c) established consistency and some rates of convergence for the estimated common components ($\lambda_i' F_t$) for dynamic factor models.

---

[3] 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 limits is only a recent endeavor.

However, inferential theory is not well understood for large-dimensional 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 (nondegenerate) 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 components method. Furthermore, the results are derived under more general assumptions 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, with a fixed N, one can 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) for large-dimensional factor models.

We also consider the case of large N but fixed T. We show that to estimate the factors consistently, a necessary condition is asymptotic orthogonality and asymptotic homoskedasticity (defined below). In contrast, under the framework of large N and large T, we establish consistency in the presence of serial correlation and heteroskedasticity. That is, the necessary condition under fixed T is no longer necessary when both N and T are large.

The rest of the paper is organized as follows. Section 2 sets up the model. 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. Section 6 reports the simulation results. Concluding remarks are provided in Section 7. All proofs are given in the appendices.

2. ESTIMATION AND ASSUMPTIONS

Recall that a factor model is represented by

$$(2) \qquad X_{it} = \lambda_i' F_t + e_{it} = C_{it} + e_{it},$$

where $C_{it} = \lambda_i' F_t$ is the common component, and all other variables are introduced previously. 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), the principal components estimator converges 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 the principal components estimator.

Equation (2) can be written as an N-dimension time series with T observations:

$$(3) \qquad X_t = \Lambda F_t + e_t \qquad\qquad\qquad\qquad (t = 1, 2, \ldots, T)$$

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

$$X_i = F\lambda_i + e_i \qquad\qquad\qquad\qquad (i = 1, 2, \ldots, N)$$

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

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

where $X = (X_1, X_2, \ldots, X_N)$ is a $T \times N$ matrix of observed data and $e = (e_1, e_2, \ldots, e_N)$ is a $T \times N$ matrix of idiosyncratic errors. The matrices $\Lambda(N \times r)$ and $F(T \times r)$ are both unknown.

Our objective is to derive the large-sample properties of the estimated factors and their loadings when both N and T are 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 $F'F/T = I_r$ (an $r \times r$ identity matrix), the problem is identical to maximizing $\mathrm{tr}(F'(XX')F)$. 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 $\widetilde{F}\widetilde{\Lambda}'$.

Since both N and T are allowed to grow, there is a need to explain the way in which limits are taken. There are three limit concepts: sequential, pathwise, and simultaneous. A sequential limit takes one of the indices to infinity first and then the other index. Let $g(N, T)$ be a quantity of interest. A sequential limit with N growing first to infinity is written as $\lim_{T\to\infty}\lim_{N\to\infty}g(N, T)$. This limit may not be the same as that of T growing first to infinity. A pathwise limit restricts (N, T) to increase along a particular trajectory of (N, T). This can be written as $\lim_{v\to\infty}g(N_v, T_v)$, where (N_v, T_v) describes a given trajectory as v varies. If v is chosen to be N, then a pathwise limit is written as $\lim_{N\to\infty}g(N, T(N))$. A simultaneous limit allows (N, T) to increase along all possible paths such that $\lim_{v\to\infty}\min\{N_v, T_v\} = \infty$. This limit is denoted by $\lim_{N, T\to\infty}g(N, T)$. The present paper mainly considers simultaneous limit theory. Note that the existence of a simultaneous limit implies the existence of sequential and pathwise limits (and the limits are the same), but the converse is not true. When $g(N, T)$ is a

function of N (or T) alone, its limit with respect to N (or T) is automatically a simultaneous limit. Due to the nature of our problem, this paper also considers simultaneous limits with restrictions. For example, we may need to consider the limit of $g(N, T)$ for N and T satisfying $\sqrt{N}/T \to 0$; that is, $\lim_{v \to \infty} g(N_v, T_v)$ over all paths such that $\lim_{v \to \infty} \min\{N_v, T_v\} = \infty$ and $\lim_{v \to \infty} \sqrt{N_v}/T_v = 0$.

Let $\|A\| = [\text{tr}(A'A)]^{1/2}$ denote the norm of matrix A. 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 following assumptions are used in Bai and Ng (2002) to estimate the number of factors consistently:

ASSUMPTION A—Factors: $E\|F_t^0\|^4 \le M < \infty$ and $T^{-1} \sum_{t=1}^{T} F_t^0 F_t^{0\prime} \xrightarrow{P} \Sigma_F$ for some $r \times r$ positive definite matrix Σ_F.

ASSUMPTION B—Factor Loadings: $\|\lambda_i\| \le \bar{\lambda} < \infty$, and $\|\Lambda^{0\prime} \Lambda^0/N - \Sigma_N\| \to 0$ 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), $E|N^{-1/2} \sum_{i=1}^{N} [e_{is} e_{it} - E(e_{is} e_{it})]|^4 \le M$.

ASSUMPTION D—Weak dependence between factors and idiosyncratic errors:

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

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 to 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 nontrivial contribution to the variance of X_t. We only consider nonrandom 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 satisfied if the e_{it} are independent for all i and t. Correlation in the idiosyncratic components allows the 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 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, if $e_{it} = \epsilon_{it}\|F_t\|$ with ϵ_{it} being independent of F_t and ϵ_{it} satisfying Assumption C, then Assumption D holds.[4]

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 e_t is stationary with $E(e_{it}e_{jt}) = \tau_{ij}$, then from matrix theory, the largest eigenvalue of Ω is bounded by $\max_i \sum_{j=1}^N |\tau_{ij}|$. Thus if we assume $\sum_{j=1}^N |\tau_{ij}| \le M$ for all i and all N, which implies Assumption C3, then (3) will be an approximate factor model in the sense of Chamberlain and Rothschild. Since we also allow for nonstationarity (e.g., heteroskedasticity in the time dimension), our model is more general than approximate factor models.

Throughout this paper, the number of factors (r) is assumed fixed as N and T grow. Many economic problems do suggest this kind of factor structure. For example, the APT theory of Ross (1976) assumes an unknown but fixed r. The Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965) implies one factor in the presence of a risk-free asset and two factors otherwise, irrespective of the number of assets (N). The rank theory of consumer demand systems of Gorman (1981) and Lewbel (1991) implies no more than three factors if consumers are utility maximizing, regardless of the number of consumption goods.

In general, larger data sets may contain more factors. But this may not have any practical consequence since large data sets allow us to estimate more factors. Suppose one is interested in testing the APT theory for Asian and the U.S. financial markets. Because of market segmentation and cultural and political differences, the factors explaining asset returns in Asian markets are different from those in the U.S. markets. Thus, the number of factors is larger if data from the two markets are combined. The total number of factors, although increased, is still fixed (at a new level), regardless of the number of Asian and U.S. assets if APT theory prevails in both markets. Alternatively, two separate tests of APT can be conducted, one for Asian markets, one for the U.S.; each individual data set

---

[4] While these assumptions take a similar format as in classical factor analysis in that assumptions were made on the uobservable quantities such as F_t and e_t (data generating process assumptions), it would be better to make assumptions in terms of the observables X_{it}. Assumptions characterized by observable variables may have the advantage of, at least in principle, having the assumptions verified by the data. We plan to pursue this in future research.

is large enough to perform such a test and the number of factors is fixed in each market. The point is that such an increase in r has no practical consequence, and it is covered by the theory. Mathematically, a portion of the cross-section units may have zero factor loadings, which is permitted by the model's assumptions.

It should be pointed out that r can be a strictly increasing function of N (or T). This case is not covered by our theory and is left for future research.

3. ASYMPTOTIC THEORY

Assumptions A–D are sufficient for consistently estimating the number of factors (r) as well as the factors themselves and their loadings. By analyzing the statistical properties of $V(k)$ as a function of k, Bai and Ng (2002) showed that the number of factors (r) can be estimated consistently by minimizing the following criterion:

$$IC(k) = \log(V(k)) + k\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right).$$

That is, for $\bar{k} > r$, let $\hat{k} = \arg\min_{0 \le k \le \bar{k}} IC(k)$. Then $P(\hat{k} = r) \to 1$, as $T, N \to \infty$. This consistency result does not impose any restriction between N and T, except $\min\{N, T\} \to \infty$. Bai and Ng also showed that AIC (with penalty $2/T$) and BIC (with penalty $\log T/T$) do not yield consistent estimators. In this paper, we shall assume a known r and focus on the limiting distributions of the estimated factors and factor loadings. Their asymptotic distributions are not affected when the number of factors is unknown and is estimated.[5] Additional assumptions are needed to derive their limiting distributions:

ASSUMPTION E—Weak Dependence: *There exists $M < \infty$ such that for all T and N, and for every $t \le T$ and every $i \le N$:*
1. $\sum_{s=1}^{T} |\gamma_N(s, t)| \le M$.
2. $\sum_{k=1}^{N} |\tau_{ki}| \le M$.

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 $s \ne t$. Then Assumption E1 is equivalent to $(1/N)\sum_{i=1}^{N} E(e_{it}^2) \le M$ for all t and N, which is implied by C1. Under cross-section independence, E2 is equivalent to $E(e_{it})^2 \le M$. Thus under time series and cross-section independence, E1 and E2 are equivalent and are implied by C1.

---

[5] This follows because $P(\widetilde{F}_t \le x) = P(\widetilde{F}_t \le x, \hat{k} = r) + P(\widetilde{F}_t \le x, \hat{k} \ne r)$. But $P(\widetilde{F}_t \le x, \hat{k} \ne r) \le P(\hat{k} \ne r) = o(1)$. Thus

$$P(\widetilde{F}_t \le x) = P(\widetilde{F}_t \le x, \hat{k} = r) + o(1)$$
$$= P(\widetilde{F}_t \le x | \hat{k} = r)P(\hat{k} = r) + o(1)$$
$$= P(\widetilde{F}_t \le x | \hat{k} = r) + o(1)$$

because $P(\hat{k} = r) \to 1$. In summary, $P(\widetilde{F}_t \le x) = P(\widetilde{F}_t \le x | \hat{k} = r) + o(1)$.

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

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 $\Phi_i = \operatorname{p\,lim}_{T \to \infty} (1/T) \sum_{s=1}^{T} \sum_{t=1}^{T} E[F_t^0 F_s^{0\prime} e_{is} e_{it}]$.

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 $(\Sigma_\Lambda \cdot \Sigma_F)$ are distinct.*

The matrices Σ_F and Σ_Λ are defined in Assumptions A and B. 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. A 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 (2002), 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 distributions of the estimated common components. The 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 $A^{1/2} B A^{1/2}$ are the same.

PROPOSITION 1: *Under Assumptions A–D and G,*

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

The matrix Q is invertible and is given by $Q = V^{1/2} \Upsilon' \Sigma_{\Lambda}^{-1/2}$, where $V = \mathrm{diag}(v_1, v_2, \ldots, v_r), v_1 > v_2 > \cdots > v_r > 0$ 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 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. 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 considered in this paper. This is because consistent estimation of the covariance matrix of X_t is not a well defined problem when N and T increase simultaneously to 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 components estimator $\widetilde{F}$, there exists an invertible matrix H (whose dependence on N, T will be suppressed for notational simplicity) such that $\widetilde{F}$ is an estimator of $F^0 H$ and $\widetilde{\Lambda}$ is an estimator of $\Lambda^0 (H')^{-1}$. In addition, $\widetilde{F}\widetilde{\Lambda}'$ is an estimator of $F^0 \Lambda^{0\prime}$, the common components. It is clear that the common components are identifiable. Furthermore, knowing $F^0 H$ is as good as knowing F^0 for many purposes. For example, in regression analysis, using F^0 as the regressor will give the same predicted value as using $F^0 H$ as the regressor. Because F^0 and $F^0 H$ 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 $F^0 H$. For the same reason, the portfolio-evaluation measurements of Connor and Korajczyk (1986) give valid results whether F^0 or $F^0 H$ is used.

THEOREM 1: *Under Assumptions A–G, as $N, T \to \infty$, we have:*
(i) *if $\sqrt{N}/T \to 0$, then for each t,*

$$\sqrt{N}\left(\widetilde{F}_t - H' F_t^0\right) = V_{NT}^{-1}\left(\frac{\widetilde{F}' F^0}{T}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i^0 e_{it} + o_p(1)$$

$$\xrightarrow{d} N(0, V^{-1} Q \Gamma_t Q' V^{-1}),$$

where V_{NT} is a diagonal matrix consisting of the first r eigenvalues of $(1/NT)XX'$ in decreasing order, V and Q are defined in Proposition 1, and Γ_t is defined in F3;
(ii) *if $\liminf \sqrt{N}/T \geq \tau > 0$, then*

$$T\left(\widetilde{F}_t - H' F_t^0\right) = O_p(1).$$

The dominant case is part (i); that is, asymptotic normality generally holds. Applied researchers should feel comfortable using the normal approximation for the estimated factors. Part (ii) is useful for theoretical purpose when a convergence rate is needed. The theorem says that the convergence rate is $\min\{\sqrt{N}, T\}$. When the factor loadings λ_i^0 $(i = 1, 2, \ldots, N)$ are all known, F_t^0 can be estimated by the cross-section least squares method and the rate of convergence is $\sqrt{N}$. The rate of convergence $\min\{\sqrt{N}, T\}$ reflects the fact that factor loadings are unknown and are estimated. Under stronger assumptions, however, the root-N convergence rate is still achievable (see Section 4). Asymptotic normality of Theorem 1 is achieved by the central limit theorem as $N \to \infty$. Thus large N is required for this theorem.

Additional comments:

1. Although restrictions between N and T are needed, the theorem is not a sequential limit result but a simultaneous one. In addition, the theorem holds not only for a particular relationship between N and T, but also for many combinations of N and T. The restriction $\sqrt{N}/T \to 0$ is not strong. Thus asymptotic normality is the more prevalent situation for empirical applications. The result permits simultaneous inference for large N and large T. For example, the portfolio-performance measurement of Connor and Korajczyk (1986) can be obtained without recourse to a sequential limit argument. See footnote 3.

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

$$Y_{t+1} = \alpha F_t^0 + \beta W_t + u_{t+1} \qquad\qquad (t = 1, 2, \ldots, T),$$

where Y_t and W_t are observable, but F_t^0 is not. However, F_t^0 can be replaced by $\widetilde{F}_t$. A crude calculation shows that the estimation error in $\widetilde{F}_t$ can be ignored as long as $\widetilde{F}_t = H'F_t^0 + o_p(T^{-1/2})$ with H having a full rank. This will be true if $T/N \to 0$ by Theorem 1 because $\widetilde{F}_t - H'F_t^0 = O_p(1/\min\{\sqrt{N}, T\})$. A more careful but elementary calculation shows that the estimation error is negligible if $\widetilde{F}'(\widetilde{F} - F^0 H)/T = o_p(T^{-1/2})$. Lemma B.3 in the Appendix shows that $\widetilde{F}'(\widetilde{F} - F^0 H)/T = o_p(\delta_{NT}^{-2})$, where $\delta_{NT}^2 = \min\{N, T\}$. This implies that the estimation error in $\widetilde{F}$ is negligible if $\sqrt{T}/N \to 0$. Thus for large N, 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 $\widetilde{F}_t$, it is easy to show that the forecast $Y_{T+1|T} = \hat{\alpha}\widetilde{F}_T + \hat{\beta}W_T$ is $\sqrt{T}$-consistent for the conditional mean $E(Y_{T+1}|F_T^0, W_T)$, assuming the conditional mean of u_{T+1} is zero. Furthermore, in constructing confidence intervals for the forecast, the estimation effect of $\widetilde{F}_T$ can be ignored provided that N is large. If $\sqrt{T}/N \to 0$ does not hold, then the limiting distribution of the forecast will also depend on the limiting distribution of $\widetilde{F}_T$; the confidence interval must reflect this estimation effect. Theorem 1 allows us to account for this effect when constructing confidence intervals for the forecast.

3. 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

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

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 $Q\Gamma_t Q'$.

Next, we present a uniform consistency result for the estimated factors.

PROPOSITION 2: *Under Assumptions A–E,*

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

This lemma gives an upper bound on the maximum deviation of the estimated factors from the true ones (up to a transformation). The bound is not the sharpest possible because the proof essentially uses the argument that $\max_t \|\widetilde{F}_t - H' F_t^0\| \le \sum_{t=1}^{T} \|\widetilde{F}_t - H' F_t^0\|$. Note that if $\liminf N/T^2 \ge c > 0$, then the maximum deviation is $O_p(T^{-1/2})$, which is a very strong result. In addition, if it is assumed that $\max_{1 \le s \le T} \|F_s^0\| = O_p(\alpha_T)$ (e.g., when F_t is strictly stationary and its moment generating function exists, $\alpha_T = \log T$), then it can be shown that, for $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$,

$$\max_{1 \le t \le T} \left\| \widetilde{F}_t - H' F_t^0 \right\| = O_p(T^{-1/2} \delta_{NT}^{-1}) + O_p(\alpha_T T^{-1}) + O_p((T/N)^{1/2}).$$

3.2. *Limiting Distribution of Estimated Factor Loadings*

The previous section shows that $\widetilde{F}$ is an estimate of $F^0 H$. Now we show that $\widetilde{\Lambda}$ is an estimate of $\Lambda^0 (H')^{-1}$. That is, $\widetilde{\lambda}_i$ is an estimate of $H^{-1} \lambda_i^0$ for every i. The estimated loadings are used in Lehmann and Modest (1988) to construct various portfolios.

THEOREM 2: *Under Assumptions A–G, as $N, T \to \infty$:*
(i) *if $\sqrt{T}/N \to 0$, then for each i,*

$$\sqrt{T}(\widetilde{\lambda}_i - H^{-1} \lambda_i^0) = V_{NT}^{-1} \left(\frac{\widetilde{F}' F^0}{T} \right) \left(\frac{\Lambda^{0\prime} \Lambda^0}{N} \right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t^0 e_{it} + o_p(1)$$

$$\xrightarrow{d} N(0, (Q')^{-1} \Phi_i Q^{-1}),$$

where V_{NT} is defined in Theorem 1, Q is given in Proposition 1, and Φ_i in F4;
(ii) *if $\liminf \sqrt{T}/N \ge \tau > 0$, then*

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

The dominant case is part (i), asymptotic normality. Part (ii) is of theoretical interest when a convergence rate is needed. This rate is $\min\{\sqrt{T}, N\}$. When the factors $F_t^0 (t = 1, 2, \ldots, T)$ are all observable, λ_i^0 can be estimated by a time series regression with the ith cross-section unit, and the rate of convergence is $\sqrt{T}$. The new rate $\min\{\sqrt{T}, N\}$ is due to the fact that F_t^0's are not observable and are estimated.

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 $\widetilde{C}_{it} = \widetilde{F}_t' \tilde{\lambda}_i$.

THEOREM 3: *Under Assumptions A–F, as* $N, T \to \infty$, *we have for each i and t,*

$$\left(\frac{1}{N} V_{it} + \frac{1}{T} W_{it} \right)^{-1/2} \left(\widetilde{C}_{it} - C_{it}^0 \right) \xrightarrow{d} N(0, 1)$$

where $V_{it} = \lambda_i^{0\prime} \Sigma_{\Lambda}^{-1} \Gamma_t \Sigma_{\Lambda}^{-1} \lambda_i^0$, $W_{it} = F_t^{0\prime} \Sigma_F^{-1} \Phi_i \Sigma_F^{-1} F_t^0$, *and* Σ_{Λ}, Γ_t, Σ_F, *and* Φ_i *are all defined earlier. Both* V_{it} *and* W_{it} *can be replaced by their consistent estimators.*

A remarkable feature of Theorem 3 is that no restriction on the relationship between N and T is required. The estimated common components are always asymptotically normal. The convergence rate is $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. To see this, Theorem 3 can be rewritten as

$$(5) \qquad \frac{\delta_{NT} \left(\widetilde{C}_{it} - C_0^{it} \right)}{\left(\frac{\delta_{NT}^2}{N} V_{it} + \frac{\delta_{NT}^2}{T} W_{it} \right)^{1/2}} \xrightarrow{d} N(0, 1).$$

The denominator is bounded both above and below. Thus 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 $\tilde{\lambda}_i$ is $\sqrt{T}$. When Λ^0 is observable, the best rate for $\widetilde{F}_t$ is $\sqrt{N}$. It follows that when both are estimated, the best rate for $\tilde{\lambda}_i'\widetilde{F}_t$ is the minimum of $\sqrt{N}$ and $\sqrt{T}$. Theorem 3 has two special cases: (a) if $N/T \to 0$, then $\sqrt{N}(\widetilde{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, V_{it})$; (b) if $T/N \to 0$, then $\sqrt{T}(\widetilde{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, W_{it})$. But Theorem 3 does not require a limit for T/N or N/T.

4. STATIONARY IDIOSYNCRATIC ERRORS

In the previous section, the rate of convergence for $\widetilde{F}_t$ is shown to be $\min\{\sqrt{N}, T\}$. If T is fixed, it implies that $\widetilde{F}_t$ 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:

$$(6) \qquad \frac{1}{N}\sum_{i=1}^{N} e_{it}e_{is} \to 0, \quad t \neq s, \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N} e_{it}^2 \to \sigma^2 \quad \text{for all } t, \quad \text{as } N \to \infty.$$

We shall call the first condition asymptotic orthogonality and the second condition asymptotic homoskedasticity (σ^2 not depending on t). They established consistency under assumption (6). This assumption appears to be reasonable for asset returns and is commonly used in the finance literature, e.g., Campbell, Lo, and Mackinlay (1997). For many economic variables, however, one of the conditions could be easily violated. We show that assumption (6) 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 orthogonality 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 the large-dimensional framework.

Next, we show that our previous results can also be strengthened under homoskedasticity and no serial correlation.

ASSUMPTION H—$E(e_{it}e_{is}) = 0$ *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 $\bar{\sigma}_N^2 = (1/N)\sum_{i=1}^{N}\sigma_i^2$, 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 $(1/TN)XX'$. Lemma A.3 (Appendix A) shows $V_{NT} \xrightarrow{p} V$, a positive definite matrix. Define $D_{NT} = V_{NT}(V_{NT} - \frac{1}{T}\bar{\sigma}_N^2)^{-1}$; then $D_{NT} \xrightarrow{p} I_r$, as T and N go to infinity. Define $\bar{H} = HD_{NT}$.

THEOREM 5: *Under Assumptions A–H, as $T, N \to \infty$, we have*

$$\sqrt{N}(\tilde{F}_t - \bar{H}'F_t^0) \xrightarrow{d} N(0, V^{-1}Q\Gamma Q'V^{-1})$$

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 cross-section heteroskedasticity are still allowed. Thus the result is for approximate factor models. This theorem does not require any restriction on the relationship between N and T except that they both go to infinity. The rate of convergence ($\sqrt{N}$) 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) also holds without any restriction on the relationship between N and T. However, cross-section homoskedasticity is unduly restrictive. Assumption H does not improve the result of Theorem 3, which already offers the best rate of convergence.

5. ESTIMATING COVARIANCE MATRICES

In this section, we derive consistent estimators of the asymptotic variance-covariance matrices that appear in Theorems 1–3.

(a) *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} ($i = 1, 2, \ldots, N$). The asymptotic covariance of $\widetilde{F}_t$ is given by $\Pi_t = V^{-1} Q \Gamma_t Q' V^{-1}$, where Γ_t is defined in (4). That is,

$$\Pi_t = \operatorname{plim} V_{NT}^{-1} \left(\frac{\widetilde{F}' F^0}{T} \right) \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_{it}^2 \lambda_i^0 \lambda_i^{0'} \right) \left(\frac{F^{0'} \widetilde{F}}{T} \right) V_{NT}^{-1}.$$

This matrix involves the product $F^0 \Lambda^{0'}$, which can be replaced by its estimate $\widetilde{F} \widetilde{\Lambda}'$. A consistent estimator of the covariance matrix is then given by

$$(7) \qquad \widehat{\Pi}_t = V_{NT}^{-1} \left(\frac{\widetilde{F}' \widetilde{F}}{T} \right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i' \right) \left(\frac{\widetilde{F}' \widetilde{F}}{T} \right) V_{NT}^{-1} = V_{NT}^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i' \right) V_{NT}^{-1}$$

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

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

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

Let $\widehat{\Theta}_i$ be the HAC estimator of Newey and West (1987), constructed with the series $\{\widetilde{F}_t \cdot \tilde{e}_{it}\}$ ($t = 1, 2, \ldots, T$). That is,

$$\widehat{\Theta}_i = D_{0,i} + \sum_{v=1}^{q} \left(1 - \frac{v}{q+1} \right) (D_{vi} + D_{vi}'),$$

where $D_{vi} = (1/T) \sum_{t=v+1}^{T} \widetilde{F}_t \tilde{e}_{it} \tilde{e}_{it-v} \widetilde{F}_{t-v}'$ and q goes to infinity as T goes to infinity with $q/T^{1/4} \to 0$. One can also use other HAC estimators such as Andrews' (1991) data dependent method with quadratic spectral kernel. While a HAC estimator based on $F_t^0 \cdot e_{it}$ (the true factors and true idiosyncratic errors) estimates Φ_i, a HAC estimator based on $\widetilde{F}_t \cdot \tilde{e}_{it}$ is directly estimating Θ_i (because $\widetilde{F}_t$ estimates $H' F_t^0$). The consistency of $\widehat{\Theta}_i$ is proved in the Appendix.

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

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

$$\widehat{W}_{it} = \tilde{F}_t' \left(\frac{\tilde{F}'\tilde{F}}{T}\right)^{-1} \widehat{\Theta}_i \left(\frac{\tilde{F}'\tilde{F}}{T}\right)^{-1} \tilde{F}_t \equiv \tilde{F}_t'\widehat{\Theta}_i\tilde{F}_t$$

be the 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. MONTE CARLO SIMULATIONS

We use simulations to assess the adequacy of the asymptotic results in approximating the finite sample distributions of $\tilde{F}_t$, $\tilde{\lambda}_i$, and $\tilde{C}_{it}$. To conserve space, we only report the results for $\tilde{F}_t$ and $\tilde{C}_{it}$ because $\tilde{\lambda}_i$ shares similar properties to $\tilde{F}_t$. We consider combinations of $T = 50, 100$, and $N = 25, 50, 100, 1000$. Data are generated according to $X_{it} = \lambda_i^{0'}F_t^0 + e_{it}$, where λ_i^0, F_t^0, and e_{it} are i.i.d. $N(0,1)$ for all i and t. The number of factors, r, is one. This gives the data matrix X $(T \times N)$. The estimated factor $\tilde{F}$ is $\sqrt{T}$ times the eigenvector corresponding to the largest eigenvalue of XX'. Given $\tilde{F}$, we have $\tilde{\Lambda} = X'\tilde{F}/T$ and $\tilde{C} = \tilde{F}\tilde{\Lambda}'$. The reported results are based on 2000 repetitions.

To demonstrate that $\tilde{F}_t$ is estimating a transformation of F_t^0 ($t = 1, 2, \ldots, T$), we compute the correlation coefficient between $\{\tilde{F}_t\}_{t=1}^{T}$ and $\{F_t^0\}_{t=1}^{T}$. Let $\rho(\ell, N, T)$ be the correlation coefficient for the ℓth repetition with given (N, T). Table I reports this coefficient averaged over $L = 2000$ repetitions; that is, $(1/L)\sum_{\ell=1}^{L}\rho(\ell, N, T)$.

The correlation coefficient can be considered as a measure of consistency for all t. It is clear from the table that the estimation precision increases as N grows. With $N = 1000$, the estimated factors can be effectively treated as the true ones.

We next consider asymptotic distributions. The main asymptotic results are

$$(8) \qquad \sqrt{N}(\tilde{F}_t - H'F_t^0) \xrightarrow{d} N(0, \Pi_t)$$

TABLE I

AVERAGE CORRELATION COEFFICIENTS BETWEEN
$\{\tilde{F}_t\}_{t=1}^{T}$ AND $\{F_t^0\}_{t=1}^{T}$

	$N = 25$	$N = 50$	$N = 100$	$N = 1000$
$T = 50$	0.9777	0.9892	0.9947	0.9995
$T = 100$	0.9785	0.9896	0.9948	0.9995

where $\Pi_t = V^{-1}Q\Gamma_t Q'V^{-1}$, and

$$\left(\frac{1}{N}V_{it} + \frac{1}{T}W_{it}\right)^{-1/2}(\widetilde{C}_{it} - C_{it}^0) \xrightarrow{d} N(0,1)$$

where V_{it} and W_{it} are defined in Theorem 3. We define standardized estimates as

$$f_t = \widehat{\Pi}_t^{-1/2}\sqrt{N}(\widetilde{F}_t - H'F_t^0) \quad \text{and} \quad c_{it} = \left(\frac{1}{N}\widehat{V}_{it} + \frac{1}{T}\widehat{W}_{it}\right)^{-1/2}(\widetilde{C}_{it} - C_{it}^0)$$

where $\widehat{\Pi}_t$, $\widehat{V}_{it}$, and $\widehat{W}_{it}$ are the estimated asymptotic variances given in Section 5. That is, the standardization is based on the theoretical mean and theoretical variance rather than the sample mean and sample variance from Monte Carlo repetitions. The standardized estimates should be approximately $N(0,1)$ if the asymptotic theory is adequate.

We next compute the sample mean and sample standard deviation (std) from Monte Carlo repetitions of standardized estimates. For example, for f_t, the sample mean is defined as $\bar{f}_t = (1/L)\sum_{\ell=1}^{L} f_t(\ell)$ and the sample standard deviation is the square root of $(1/L)\sum_{\ell=1}^{L}(f_t(\ell) - \bar{f}_t)^2$, where ℓ refers to the ℓth repetition. The sample mean and sample standard deviation for c_{it} are similarly defined. We only report the statistics for $t = [T/2]$ and $i = [N/2]$, where $[a]$ is the largest integer smaller than or equal to a; other values of (t,i) give similar results and thus are not reported.

The first four rows of Table II are for f_t and the last four rows are for c_{it}. In general, the sample means are close to zero and the standard deviations are close to 1. Larger T generates better results than smaller T. If we use the two-sided normal critical value (1.96) to test the hypothesis that f_t has a zero mean with known variance of 1, then the null hypothesis is rejected (marginally) for just two cases: $(T,N) = (50,1000)$ and $(T,N) = (100, 50)$. But if the estimated standard deviation is used, the two-sided t statistic does not reject the zero-mean hypothesis for all cases. As for the common components, c_{it}, all cases except $N = 25$ strongly point to zero mean and unit variance. This is consistent with the asymptotic theory.

TABLE II

SAMPLE MEAN AND STANDARD DEVIATION OF f_t AND c_{it}

			$N=25$	$N=50$	$N=100$	$N=1000$
f_t	$T=50$	mean	0.0235	−0.0189	0.0021	−0.0447
	$T=100$	mean	0.0231	0.0454	−0.0196	0.0186
f_t	$T=50$	std	1.2942	1.2062	1.1469	1.2524
	$T=100$	std	1.2521	1.1369	1.0831	1.0726
c_{it}	$T=50$	mean	−0.0455	−0.0080	−0.0029	−0.0036
	$T=100$	mean	0.0252	0.0315	0.0052	0.0347
c_{it}	$T=50$	std	1.4079	1.1560	1.0932	1.0671
	$T=100$	std	1.1875	1.0690	1.0529	1.0402

We also present graphically the above standardized estimates f_t and c_{it}. Figure 1 displays the histograms of f_t for $T = 50$ and Figure 2 for $T = 100$. The histogram is scaled to be a density function; the sum of the bar heights times the bar lengths is equal to 1. Scaling makes for a meaningful comparison with the standard normal density. The latter is overlayed on the histogram. Figure 3 and Figure 4 show the histogram of c_{it} for $T = 50$ and $T = 100$, respectively, with the density of $N(0, 1)$ again overlayed.

It appears that asymptotic theory provides a very good approximation to the finite sample distributions. For a given T, the larger is N, the better is the approximation. In addition, $\widetilde{F}_t$ does appear to be $\sqrt{N}$ consistent because the histogram is for the estimates $\sqrt{N}(\widetilde{F}_t - H'F_t^0)$ (divided by its variance). The histograms stay within the same range for the standardized $\widetilde{F}_t$ as N grows from 25 to 1000. Similarly, the convergence rate for $\widetilde{C}_{it}$ is $\min\{\sqrt{N}, \sqrt{T}\}$ for the same reason. In general, the limited Monte Carlo simulations lend support to the theory.

FIGURE 1.— Histogram of estimated factors ($T = 50$). These histograms are for the standardized estimates f_t (i.e., $\sqrt{N}(\widetilde{F}_t - HF_t^0)$ divided by the estimated asymptotic standard deviation). The standard normal density function is superimposed on the histograms.
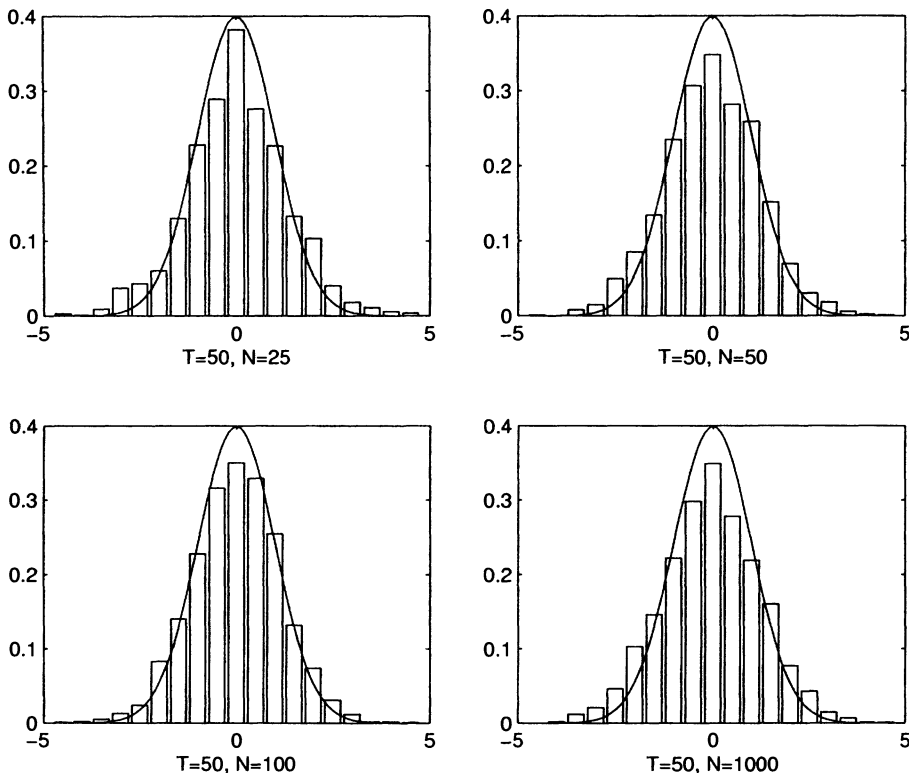
FIGURE 2.— Histogram of estimated factors ($T = 100$). These histograms are for the standard-ized estimates f_t (i.e., $\sqrt{N}(\widetilde{F}_t - HF_t^0)$ divided by the estimated asymptotic standard deviation). The standard normal density function is superimposed on the histograms.
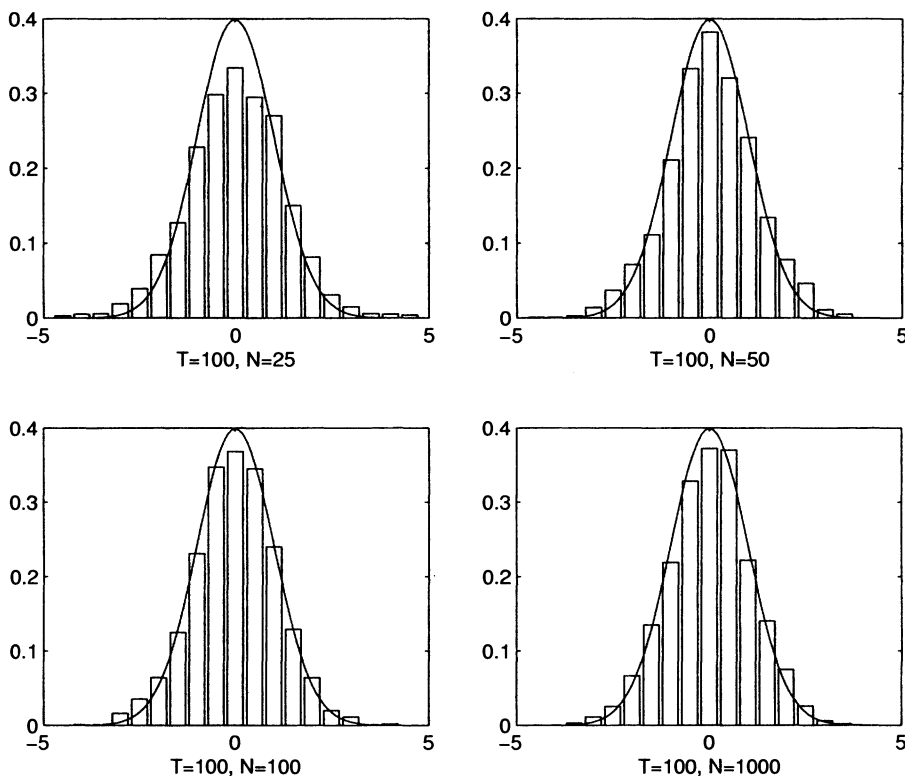
Finally, we construct confidence intervals for the true factor process. This is useful because the factors represent economic indices in various empirical appli-cations. A graphical method is used to illustrate the results. For this purpose, no Monte Carlo repetition is necessary; only one draw (a single data set X ($T \times N$)) is needed. The case of $r = 1$ is considered for simplicity. A small T ($T = 20$) is used to avoid a crowded graphical display (for large T, the resulting graphs are difficult to read because of too many points). The values for N are 25, 50, 100, 1000, respectively. Since $\widetilde{F}_t$ is an estimate of $H'F_t^0$, the 95% confidence interval for $H'F_t^0$, according to (8), is

$$\left(\widetilde{F}_t - 1.96\widehat{\varPi}_t^{1/2}N^{-1/2}, \widetilde{F}_t + 1.96\widehat{\varPi}_t^{1/2}N^{-1/2}\right) \qquad (t = 1, 2, , \ldots, T).$$

Because F^0 is known in a simulated environment, it is better to transform the confidence intervals to those of F_t^0 (rather than $H'F_t^0$) and see if the resulting confidence intervals contain the true F_t^0. We can easily rotate $\widetilde{F}$ toward F^0 by

FIGURE 3.— Histogram of estimated common components ($T = 50$). These histograms are for the standardized estimates c_{it} (i.e., $\min\{\sqrt{N}, \sqrt{T}\}(\widetilde{C}_{it} - C_{it}^0)$ divided by the estimated asymptotic standard deviation). The standard normal density function is superimposed on the histograms.

the regression[6] $F^0 = \widetilde{F}\beta + \text{error}$. Let $\hat{\beta}$ be the least squares estimate of β. Then the 95% confidence interval for F_t^0 is

$$(L_t, U_t) = \left(\hat{\beta}\widetilde{F}_t - 1.96\,\hat{\beta}\widehat{\Pi}_t^{1/2} N^{-1/2}, \hat{\beta}\widetilde{F}_t + 1.96\,\hat{\beta}\widehat{\Pi}_t^{1/2} N^{-1/2}\right)$$

$$(t = 1, 2, \ldots, T).$$

Figure 5 displays the confidence intervals (L_t, U_t) and the true factors F_t^0 ($t = 1, 2, \ldots, T$). The middle curve is F_t^0. It is clear that the larger is N, the narrower are the confidence intervals. In most cases, the confidence intervals contain the true value of F_t^0. For $N = 1000$, the confidence intervals collapse to the true

---

[6] In classical factor analysis, rotating the estimated factors is an important part of the analysis. This particular rotation (regression) is not feasible in practice because F^0 is not observable. The idea here is to show that $\widetilde{F}$ is an estimate of a transformation of F^0 and that the confidence interval before the rotation is for the transformed variable $F^0 H$.
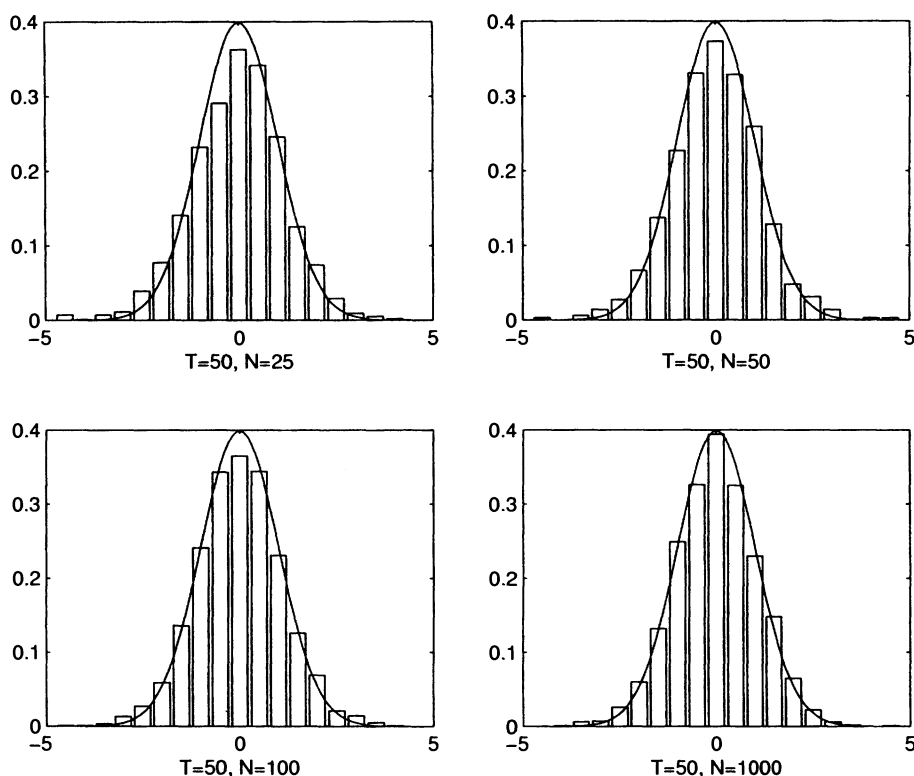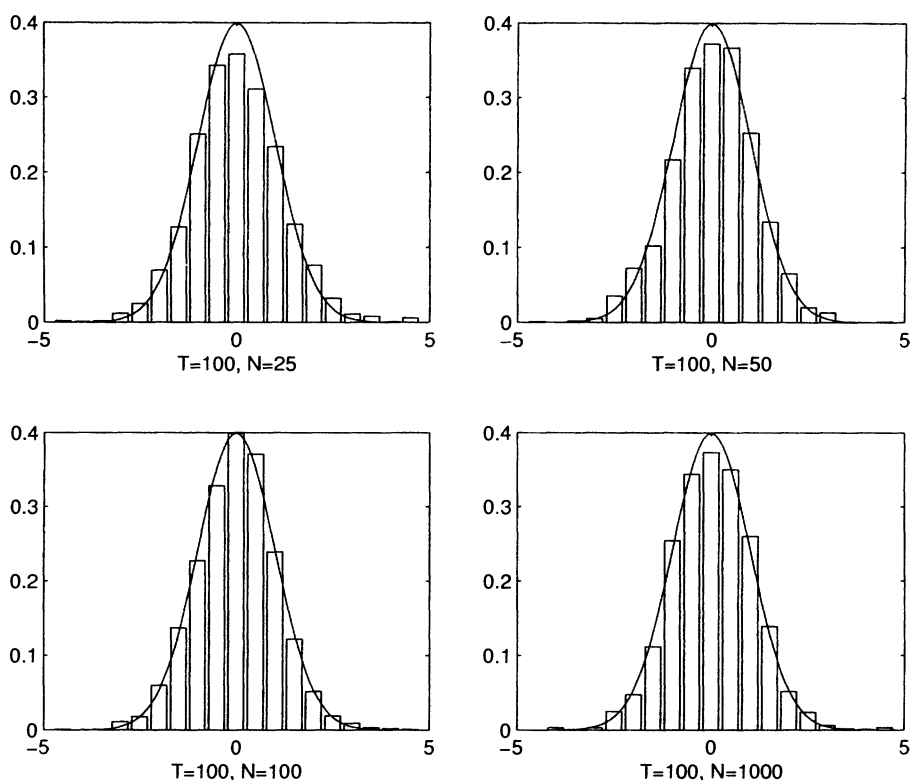
FIGURE 4.— Histogram of estimated common components ($T = 100$). These histograms are for the standardized estimates c_{it} (i.e., $\min\{\sqrt{N}, \sqrt{T}\}(\widetilde{C}_{it} - C_{it}^0)$ divided by the estimated asymptotic standard deviation). The standard normal density function is superimposed on the histograms.

values. This implies that there exists a transformation of $\widetilde{F}$ such that it is almost identical to F^0 when N is large.

7. CONCLUDING REMARKS

This paper studies factor models under a nonstandard setting: large cross sections and a large time dimension. Such large-dimensional factor models have received increasing attention in the recent economic literature. This paper considers estimating the model by the principal components method, which is feasible and straightforward to implement. We derive some inferential theory concerning the estimators, including rates of convergence and limiting distributions. In contrast to classical factor analysis, we are able to estimate consistently both the factors and their loadings, not just the latter. In addition, our results are obtained under very general conditions that allow for cross-sectional and serial
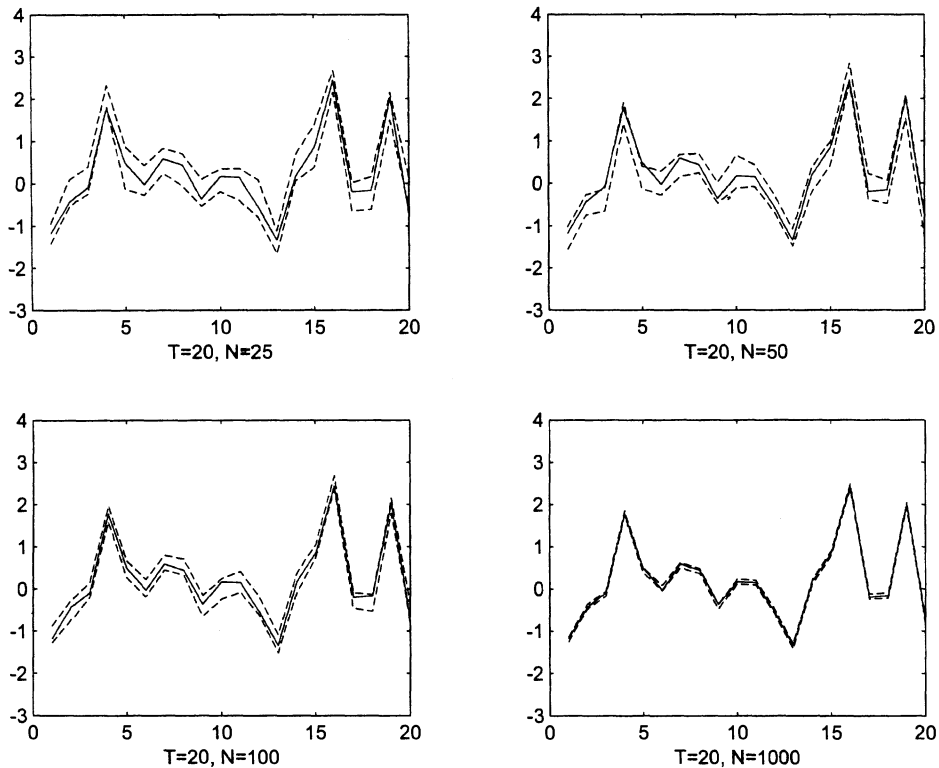
FIGURE 5.— Confidence intervals for $F_t^0(t = 1, \ldots, T)$. These are the 95% confidence intervals (dashed lines) for the true factor process $F_t^0(t = 1, 2, \ldots, 20)$ when the number of cross sections (N) varies from 25 to 1000. The middle curve (solid line) represents the true factor process.

correlations and heteroskedasticities. We also identify a necessary and sufficient condition for consistency under fixed T.

Many issues remain to be investigated. The first is the quality of approximation that the asymptotic distributions provide to the finite sample distributions. Although our small Monte Carlo simulations show adequate approximation, a thorough investigation is still needed to access the asymptotic theory. In particular, it is useful to document under what conditions the asymptotic theory would fail and what are the characteristics of finite sample distributions. The task is then to develop an alternative asymptotic theory that can better capture the known properties of the finite sample distributions. This approach to asymptotic analysis was used by Bekker (1994) to derive a new asymptotic theory for instrumental-variable estimators, where the number of instruments increases with the sample size; also see Hahn and Kursteiner (2000). This framework can be useful for factor analysis because the data matrix has two growing dimensions and some of our asymptotic analysis restricts the way in which N and T can grow. In addition,

Bekker's approach might be useful when the number of factors (r) also increases with N and T. Asymptotic theory for an increasing r is not examined in this paper and remains to be studied.

Another fruitful area of research would be empirical applications of the theoretical results derived in this paper. Our results show that it is not necessary to divide a large sample into small subsamples in order to conform to a fixed T requirement, as is done in the existing literature. A large time dimension delivers consistent estimates even under heteroskedasticity and serial correlation. In contrast, consistent estimation is not guaranteed under fixed T. Many existing applications that employed classical factor models can be reexamined using new data sets with large dimensions. It would also be interesting to examine common cycles and co-movements in the world economy along the lines of Forni et al. (2000b) and Gregory and Head (1999).

Recently, Granger (2001) and others called for large-model analysis to be on the forefront of the econometrics research agenda. We believe that large-model analysis will become increasingly important. Data sets will naturally expand as data collection, storage, and dissemination become more efficient and less costly. Also, the increasing interconnectedness in the world economy means that variables across different countries may have tighter linkages than ever before. These facts, in combination with modern computing power, make large-model analysis more pertinent and feasible. We hope that the research of this paper will encourage further developments in the analysis of large models.

Department of Economics, New York University, 269 Mercer St., 7th Floor, New York, NY 10003 U.S.A.; jushan.bai@nyu.edu

APPENDIX A: PROOF OF THEOREM 1

As defined in the main text, let V_{NT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(1/TN)XX'$ in decreasing order. By the definition of eigenvectors and eigenvalues, we have $(1/TN)XX'\widetilde{F} = \widetilde{F}V_{NT}$ or $(1/NT)XX'\widetilde{F}V_{NT}^{-1} = \widetilde{F}$. Let $H = (\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}\widetilde{F}/T)V_{NT}^{-1}$ be an $r \times r$ matrix and $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Assumptions A and B together with $\widetilde{F}'\widetilde{F}/T = I$ and Lemma A.3 below imply that $\|H\| = O_p(1)$. Theorem 1 is based on the identity (also see Bai and Ng (2002)):

$$(A.1) \qquad \widetilde{F}_t - H'F_t^0 = V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st} \right),$$

where

$$\zeta_{st} = \frac{e_s'e_t}{N} - \gamma_N(s,t),$$

$$(A.2) \qquad \eta_{st} = F_s^{0\prime}\Lambda^{0\prime}e_t/N,$$

$$\xi_{st} = F_t^{0\prime}\Lambda^{0\prime}e_s/N.$$

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

LEMMA A.1: *Under Assumptions A–D,*

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

PROOF: See Theorem 1 of Bai and Ng (2002). Note that they used $\| V_{NT} \widetilde{F}_t - V_{NT} H' F_t^0 \|^2$ as the summand. Because V_{NT} converges to a positive definite matrix (see Lemma A.3 below), it follows that $\| V_{NT} \| = O_p(1)$ and the lemma is implied by Theorem 1 of Bai and Ng. Q.E.D.

LEMMA A.2: *Under Assumptions A–F, we have*
(a) $T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \gamma_N(s, t) = O_p(\frac{1}{\sqrt{T} \delta_{NT}})$;
(b) $T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \zeta_{st} = O_p(\frac{1}{\sqrt{N} \delta_{NT}})$;
(c) $T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \eta_{st} = O_p(\frac{1}{\sqrt{N}})$;
(d) $T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \xi_{st} = O_p(\frac{1}{\sqrt{N} \delta_{NT}})$.

PROOF: Consider part (a). By adding and subtracting terms,

$$T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \gamma_N(s, t) = T^{-1} \sum_{s=1}^{T} (\widetilde{F}_s - H' F_s^0 + H' F_s^0) \gamma_N(s, t)$$

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

Now $(1/T) \sum_{s=1}^{T} F_s^0 \gamma_N(s, t) = O_p(1/T)$ since

$$E \left| \sum_{s=1}^{T} F_s^0 \gamma_N(s, t) \right| \le \left(\max_s E \| F_s^0 \| \right) \sum_{s=1}^{T} |\gamma_N(s, t)| \le M^{1+1/4}$$

by Assumptions A and E1. Consider the first term:

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

which is

$$O_p \left(\frac{1}{\delta_{NT}} \right) \frac{1}{\sqrt{T}} O(1) = O_p \left(\frac{1}{\sqrt{T} \delta_{NT}} \right)$$

by Lemma A.1 and Assumption E1. Consider part (b):

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

For the first term,

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

Furthermore,

$$T^{-1} \sum_{s=1}^{T} \zeta_{st}^2 = \frac{1}{T} \sum_{s=1}^{T} \left[\frac{e_s' e_t}{N} - \gamma_N(s, t) \right]^2 = \frac{1}{T} \sum_{s=1}^{T} \left[\frac{e_s' e_t}{N} - \frac{E(e_s' e_t)}{N} \right]^2$$

$$= \frac{1}{N} \frac{1}{T} \sum_{s=1}^{T} \left[N^{-1/2} \sum_{i=1}^{N} (e_{is} e_{it} - E(e_{is} e_{it})) \right]^2 = O_p \left(\frac{1}{N} \right).$$

Thus the first term is

$$O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\sqrt{N}}\right).$$

Next,

$$T^{-1}\sum_{s=1}^{T}F_s^0\zeta_{st} = \frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s^0(e_{is}e_{it} - E(e_{is}e_{it})) = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

by Assumption F1. Thus

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

Consider part (c):

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

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

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

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

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

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

$$T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st} = T^{-1}\sum_{s=1}^{T}\widetilde{F}_sF_t^{0\prime}\Lambda^{0\prime}e_s/N = T^{-1}\sum_{s=1}^{T}(\widetilde{F}_se_s'\Lambda^0/N)F_t^0$$

$$= \frac{1}{NT}\sum_{s=1}^{T}(\widetilde{F}_s - H'F_s^0)e_s'\Lambda^0 F_t^0 + \frac{1}{NT}\sum_{s=1}^{T}H'F_s^0e_s'\Lambda^0 F_t^0.$$

Consider the first term

$$\left\|\left(\frac{1}{NT}\sum_{s=1}^{T}(\widetilde{F}_s - H'F_s^0)e_s'\Lambda^0\right)F_t^0\right\| \leq \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H'F_s^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{e_s'\Lambda^0}{\sqrt{N}}\right\|^2\right)^{1/2}\|F_t^0\|$$

$$= O_p\left(\frac{1}{\sqrt{N}}\right)\cdot O_p\left(\frac{1}{\delta_{NT}}\right)\cdot O_p(1) = O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)$$

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

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

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

LEMMA A.3: *Assume Assumptions A–D hold. As $T, N \to \infty$:*

(i) $$T^{-1} \widetilde{F}' \left(\frac{1}{TN} XX' \right) \widetilde{F} = V_{NT} \overset{p}{\longrightarrow} V,$$

(ii) $$\frac{\widetilde{F}' F^0}{T} \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right) \frac{F^{0'} \widetilde{F}}{T} \overset{p}{\longrightarrow} V$$

where V is the diagonal matrix consisting of the eigenvalues of $\Sigma_\Lambda \Sigma_F$.

PROOF: From $(1/TN)XX'\widetilde{F} = \widetilde{F} V_{NT}$ and $\widetilde{F}' \widetilde{F}/T = I$, we have $T^{-1} \widetilde{F}' (1/TN) XX' \widetilde{F} = V_{NT}$. The rest is implicitly proved by Stock and Watson (1999). The details are omitted.   Q.E.D.

Because V is positive definite, the lemma says that $F^{0'} \widetilde{F}/T$ 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 $F^{0'} \widetilde{F}/T$ has a limit. But this does not guarantee $F^{0'} \widetilde{F}/T$ itself has a 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 ith column corresponds to the ith largest eigenvalue.

PROOF OF PROPOSITION 1: Multiply the identity $(1/TN)XX'\widetilde{F} \cong \widetilde{F} V_{NT}$ on both sides by $T^{-1} (\Lambda' \Lambda^0/N)^{1/2} F^{0'}$ to obtain:

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

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

(A.3) $$\left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2} \left(\frac{F^{0'} F^0}{T} \right) \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right) \left(\frac{F^{0'} \widetilde{F}}{T} \right) + d_{NT} = \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2} \left(\frac{F^{0'} \widetilde{F}}{T} \right) V_{NT}$$

where

$$d_{NT} = \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2} \left[\left(\frac{F^{0'} F^0}{T} \right) \Lambda^{0'} e' \widetilde{F}/(TN) + \frac{1}{TN} F^{0'} e \Lambda^0 F^{0'} \widetilde{F}/T + \frac{1}{TN} F^{0'} ee' \widetilde{F}/T \right]$$
$$= o_p(1).$$

The $o_p(1)$ is implied by Lemma A.2. Let

$$B_{NT} = \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2} \left(\frac{F^{0'} F^0}{T} \right) \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2}$$

and

(A.4) $$R_{NT} = \left(\frac{\Lambda^{0'} \Lambda^0}{N} \right)^{1/2} \left(\frac{F^{0'} \widetilde{F}}{T} \right);$$

then we can rewrite (A.3) as

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

Thus each column of R_{NT}, though not of length 1, 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 $T_{NT} = R_{NT} V_{NT}^{*-1/2}$ so that each column of T_{NT} has a unit length, and we have

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

Thus T_{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 $\widetilde{F}$ ($k = 1, 2, \ldots, r$). Thus the sign of each column in R_{NT} and thus in $T_{NT} = R_{NT}V_{NT}^{*}{}^{-1/2}$ is implicitly determined by the sign of each column in $\widetilde{F}$. Thus, given the column sign of $\widetilde{F}$, T_{NT} is uniquely determined. By the eigenvector perturbation theory (which requires the distinctness of eigenvalues; see Franklin (1968)), there exists a unique eigenvector matrix T of $B = \Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ such that $\|T_{NT} - T\| = o_p(1)$. From

$$\frac{F^{0\prime}\widetilde{F}}{T} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1/2} T_{NT}V_{NT}^{*}{}^{1/2},$$

we have

$$\frac{F^{0\prime}\widetilde{F}}{T} \xrightarrow{p} \Sigma_\Lambda^{-1/2}TV^{1/2}$$

by Assumption B and by $V_{NT}^* \xrightarrow{p} V$ in view of Lemma A.3(ii). $Q.E.D.$

PROOF OF THEOREM 1: *Case 1:* $\sqrt{N}/T \to 0$. By (A.1) and Lemma A.2, we have

$$(A.5) \qquad \widetilde{F}_t - H'F_t^0 = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right).$$

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

$$(A.6) \qquad \sqrt{N}(\widetilde{F}_t - H'F_t^0) = V_{NT}^{-1}T^{-1}\sum_{s=1}^{T}(\widetilde{F}_sF_s^{0\prime})\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} + o_p(1).$$

Now $(1/\sqrt{N})\sum_{i=1}^{N}\lambda_i^0 e_{it} \xrightarrow{d} N(0, \Gamma_t)$ by Assumption F3. Together with Proposition 1 and Lemma A.3, we have $\sqrt{N}(\widetilde{F}_t - H'F_t) \xrightarrow{d} N(0, V^{-1}Q\Gamma_t Q'V^{-1})$ as stated.

Case 2: If $\liminf \sqrt{N}/T \geq \tau > 0$, then the first and the third terms of (A.1) are the dominant terms. We have

$$T(\widetilde{F}_t - H'F_t^0) = O_p(1) + O_p(T/\sqrt{N}) = O_p(1)$$

in view of $\limsup(T/\sqrt{N}) \leq 1/\tau < \infty$. $Q.E.D.$

PROOF OF PROPOSITION 2: We consider each term on the right-hand side of (A.1). For the first term,

$$(A.7) \qquad \max_t T^{-1}\left\|\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s, t)\right\| \leq T^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\|\widetilde{F}_s\|^2\right)^{1/2}\max_t\left(\sum_{s=1}^{T}\gamma_N(s, t)^2\right)^{1/2}.$$

The above is $O_p(T^{-1/2})$ following from $T^{-1}\sum_{s=1}^{T}\|\widetilde{F}_s\|^2 = O_p(1)$ and $\sum_{s=1}^{T}\gamma_N(s, t)^2 \leq M_1$ 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 (2002) 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. $Q.E.D.$

APPENDIX B: PROOF OF THEOREM 2

To prove Theorem 2, we need some preliminary results.

LEMMA B.1: *Under Assumptions A–F, we have*

$$T^{-1}(\widetilde{F} - F^0 H)' e_i = O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

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

$$T^{-1}\sum_{t=1}^{T}(\widetilde{F}_t - H' F_t^0)e_{it} = V_{NT}^{-1}\left[T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\gamma_N(s,t)e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}e_{it}\right.$$

$$\left. + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st}e_{it} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}e_{it}\right]$$

$$= V_{NT}^{-1}(I + II + III + IV).$$

We begin with I, which can be rewritten as

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

The first term is bounded by

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

where the $O_p(1)$ follows from $Ee_{it}^2 \leq M$ and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)|^2 \leq M$ by Lemma 1(i) of Bai and Ng (2002). The expected value of the second term of I is bounded by (ignore H)

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

by Assumption C2. Thus I is $O_p(T^{-1/2}\delta_{NT}^{-1})$. For II, we rewrite it as

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

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

$$\left\|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H'F_s^0)\zeta_{st}e_{it}\right\|$$

$$\leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - H'F_s^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{st}e_{it}\right)^2\right)^{1/2}.$$

But

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

So the first term is

$$
O_p\!\left(\frac{1}{\delta_{NT}}\right)\cdot O_p\!\left(\frac{1}{\sqrt{N}}\right)=O_p\!\left(\frac{1}{\delta_{NT}\sqrt{N}}\right).
$$

Thus $II = O_p(1/\delta_{NT}\sqrt{N})$.

For III, we rewrite it as

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

The first term is bounded by

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

because

$$
\frac{1}{T}\sum_{t=1}^{T}\eta_{st}e_{it} = \frac{1}{\sqrt{N}}F_s^{0\prime}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\lambda_k e_{kt}\right)e_{it},
$$

which is $O_p(N^{-1/2})$. The second term of III can be written as

$$
\text{(B.1)}\qquad H'\left(\frac{1}{T}\sum_{s=1}^{T}F_s^0 F_s^{0\prime}\right)\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}\lambda_k e_{kt}e_{it}\right)
$$

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 $(1/TN)\sum_{t=1}^{T}\sum_{k=1}^{N}|\tau_{kit}|\le (1/N)\sum_{k=1}^{N}\tau_{ki}=O(N^{-1})$ by E2, where $|\tau_{ki,t}|\le\tau_{ki}$. In summary, III is $O_p(1/\delta_{NT}\sqrt{N})+O(N^{-1})$. The proof of IV is similar to that of III. Thus

$$
I+II+III+IV = O_p\!\left(\frac{1}{\sqrt{T}\delta_{NT}}\right)+O_p\!\left(\frac{1}{\delta_{NT}\sqrt{N}}\right)+O_p\!\left(\frac{1}{\sqrt{NT}}\right)+O_p\!\left(\frac{1}{N}\right)
$$

$$
= O_p\!\left(\frac{1}{\delta_{NT}^2}\right).\qquad\qquad\qquad\qquad\qquad\text{Q.E.D.}
$$

LEMMA B.2: *Under Assumptions A–F, the $r\times r$ matrix $T^{-1}(\widetilde{F}-F^0 H)'F^0 = O_p(\delta_{NT}^{-2})$.*

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

$$
T^{-1}\sum_{t=1}^{T}(\widetilde{F}_t - H'F_t^0)F_t^{0\prime} = V_{NT}^{-1}\Bigg[T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s F_t^{0\prime}\gamma_N(s,t) + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s F_t^{0\prime}\zeta_{st}
$$

$$
+ T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s F_t^{0\prime}\eta_{st} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s F_t^{0\prime}\xi_{st}\Bigg]
$$

$$
= V_{NT}^{-1}(I+II+III+IV).
$$

Term I is $O_p(T^{-1/2}\delta_{NT}^{-1})$. The proof is the same as that of I of Lemma B.1. Next,

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

The first term is $O_p(1/\delta_{NT}\sqrt{N})$ following arguments analogous to Lemma B.1. The second term is $O_p(1/\sqrt{NT})$ by Assumption F1 and the Cauchy-Schwarz inequality. Thus, $II = O_p(1/\delta_{NT}\sqrt{N})$. Next, note that

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

Now

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}H'F_s^0 F_t^{0\prime}\eta_{st} = H'\left(\frac{1}{T}\sum_{s=1}^{T}F_s^0 F_s^{0\prime}\right)\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}\lambda_k F_t' e_{kt} = O_p(1)O_p\left(\frac{1}{\sqrt{NT}}\right)$$

by Assumption F2. Consider

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

The second term can be rewritten as

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

which is $O_p(1/\sqrt{NT})$ by F2. Therefore,

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - H'F_s^0)F_t^{0\prime}\eta_{st} = O_p\left(\frac{1}{\delta_{NT}}\right)\cdot O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Thus,

$$III = O_p\left(\frac{1}{\delta_{NT}}\right)\cdot O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

The proof for IV is similar to that of III. Thus

$$I + II + III + IV = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right) + O_p\left(\frac{1}{\delta_{NT}\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right). \qquad Q.E.I$$

LEMMA B.3: *Under Assumptions A–F, the $r \times r$ matrix $T^{-1}(\widetilde{F} - F^0 H)'\widetilde{F} = O_p(\delta_{NT}^{-2})$.*

PROOF:

$$T^{-1}(\widetilde{F} - F^0 H)'\widetilde{F} = T^{-1}(\widetilde{F} - F^0 H)F^0 H + T^{-1}(\widetilde{F} - F^0 H)'(\widetilde{F} - F^0 H).$$

The lemma follows from Lemma B.2 and Lemma A.1. $\qquad Q.E.D.$

PROOF OF THEOREM 2: From $\widetilde{\Lambda} = \widetilde{F}'X/T$ and $X = F^0\Lambda^{0\prime} + e$, we have $\widetilde{\lambda}_i = T^{-1}\widetilde{F}'F^0\lambda_i^0 + T^{-1}\widetilde{F}'\underline{e}_i$. Writing $F^0 = F^0 - \widetilde{F}H^{-1} + \widetilde{F}H^{-1}$ and using $T^{-1}\widetilde{F}'\widetilde{F} = I$, we obtain

$$\widetilde{\lambda}_i = H^{-1}\lambda_i^0 + T^{-1}H'F^{0\prime}\underline{e}_i + T^{-1}\widetilde{F}'(F^0 - \widetilde{F}H^{-1})\lambda_i^0 + T^{-1}(\widetilde{F} - F^0 H)'\underline{e}_i.$$

Each of the last two terms is $O_p(\delta_{NT}^{-2})$ by Lemmas B.3 and B.1. Thus

$$(B.2) \qquad \widetilde{\lambda}_i - H^{-1}\lambda_i^0 = H'\frac{1}{T}\sum_{s=1}^{T}F_s^0 e_{is} + O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

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

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

But $H' \to V^{-1}Q\Sigma_\Lambda = Q'^{-1}$. The equality follows from the definition of Q; see Proposition 1. Together with Assumption F4, the desired limiting distribution is obtained.

Case 2: $\liminf \sqrt{T}/N \geq c$ 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 $N \ll T$ and thus $\delta_{NT}^2 = O(N)$. Thus

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

because $\limsup(N/\sqrt{T}) \leq 1/c < \infty$.   *Q.E.D.*

APPENDIX C: PROOF OF THEOREM 3

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

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

By adding and subtracting, the second term can be written as

$$F_t^{0\prime}H(\tilde{\lambda}_i - H^{-1}\lambda_i^0) + (\tilde{F}_t - H'F_t^0)'(\tilde{\lambda}_i - H^{-1}\lambda_i^0) = F_t^{0\prime}H(\tilde{\lambda}_i - H^{-1}\lambda_i^0) + O_p(1/\delta_{NT}^2).$$

Thus,

$$\tilde{C}_{it} - C_{it}^0 = \lambda_i^{0\prime}H'^{-1}(\tilde{F}_t - H'F_t^0) + F_t^{0\prime}H(\tilde{\lambda}_i - H^{-1}\lambda_i^0) + O_p(1/\delta_{NT}^2).$$

From (A.5) and (A.6) ,

$$\delta_{NT}(\tilde{F}_t - H'F_t^0) = \frac{\delta_{NT}}{\sqrt{N}} V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right) \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k^0 e_{kt} + O_p(1/\delta_{NT}).$$

From (B.2),

$$\delta_{NT}(\tilde{\lambda}_i - H^{-1}\lambda_i^0) = \frac{\delta_{NT}}{\sqrt{T}} H' \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s^0 e_{is} + O_p\left(\frac{1}{\delta_{NT}}\right).$$

Combining the preceding three equations, we obtain

$$\delta_{NT}(\tilde{C}_{it} - C_{it}^0) = \frac{\delta_{NT}}{\sqrt{N}} \lambda_i^0 H'^{-1} V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T}\right) \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k^0 e_{kt}$$

$$+ \frac{\delta_{NT}}{\sqrt{T}} F_t^{0\prime} HH' \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s^0 e_{is} + O_p\left(\frac{1}{\delta_{NT}}\right).$$

By the definition of H, we have $H'^{-1}V_{NT}^{-1}(\tilde{F}'F^0/T) = (\Lambda^{0\prime}\Lambda^0/N)^{-1}$. In addition, it can be shown that $HH' = (F^{0\prime}F^0/T)^{-1} + O_p(1/\delta_{NT}^2)$. Thus

$$\delta_{NT}(\tilde{C}_{it} - C_{it}^0) = \frac{\delta_{NT}}{\sqrt{N}} \lambda_i^0 \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k^0 e_{kt}$$

$$+ \frac{\delta_{NT}}{\sqrt{T}} F_t^{0\prime} \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s^0 e_{is} + O_p\left(\frac{1}{\delta_{NT}}\right).$$

Let

$$\xi_{NT} = \lambda_i^0 \left(\frac{\Lambda^{0\prime}\Lambda^0}{N} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_k^0 e_{kt}.$$

Then $\xi_{NT} \xrightarrow{d} \xi \stackrel{d}{=} N(0, V_{it})$ by Assumptions B and F3, where V_{it} is defined in Theorem 3. Let

$$\zeta_{NT} = F_t^{0\prime} \left(\frac{F^{0\prime}F^0}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s^0 e_{is}.$$

Then $\zeta_{Nt} \xrightarrow{d} \zeta \stackrel{d}{=} N(0, W_{it})$ by Assumptions A and F4, where W_{it} is defined in Theorem 3. In addition, ξ_{NT} and ζ_{NT} are asymptotically independent because the former is the sum of cross-section random variables and the latter is the sum of a particular time series (ith). Asymptotic independence holds under our general assumptions that allow for weak cross-section and time series dependence for e_{it}. This implies that (ξ_{NT}, ζ_{NT}) converges (jointly) to a bivariate normal distribution. Let $a_{NT} = \delta_{NT}/\sqrt{N}$ and $b_{NT} = \delta_{NT}/\sqrt{T}$; then

$$(C.1) \qquad \delta_{NT}(\widetilde{C}_{it} - C_{it}^0) = a_{NT}\xi_{NT} + b_{NT}\zeta_{NT} + O_p(1/\delta_{NT}).$$

The sequences a_{NT} and b_{NT} are bounded and nonrandom. If they converge to some constants, then asymptotic normality for $\widetilde{C}_{it}$ follows immediately from Slustky's Theorem. However, a_{NT} and b_{NT} are not restricted to be convergent sequences; it requires an extra argument to establish asymptotic normality. We shall use an almost sure representation theory (see Pollard (1984, page 71)). Because $(\xi_{NT}, \zeta_{NT}) \xrightarrow{d} (\xi, \zeta)$, the almost sure representation theory asserts that there exist random vectors $(\xi_{NT}^*, \zeta_{NT}^*)$ and (ξ^*, ζ^*) with the same distributions as (ξ_{NT}, ζ_{NT}) and (ξ, ζ) such that $(\xi_{NT}^*, \zeta_{NT}^*) \longrightarrow (\xi^*, \zeta^*)$ (almost surely). Now

$$a_{NT}\xi_{NT}^* + b_{NT}\zeta_{NT}^* = a_{NT}\xi^* + b_{NT}\zeta^* + a_{NT}(\xi_{NT}^* - \xi^*) + b_{NT}(\zeta_{NT}^* - \zeta^*).$$

The last two terms are each $o_p(1)$ because of the almost sure convergence. Because ξ^* and ζ^* are independent normal random variables with V_{it} and W_{it} as their variances, we have $a_{NT}\xi^* + b_{NT}\zeta^* \stackrel{d}{=} N(0, a_{NT}^2 V_{it} + b_{NT}^2 W_{it})$. That is, $(a_{NT}^2 V_{it} + b_{NT}^2 W_{it})^{-1/2}(a_{NT}\xi^* + b_{NT}\zeta^*) \stackrel{d}{=} N(0, 1)$. This implies that

$$\frac{a_{NT}\xi_{NT}^* + b_{NT}\zeta_{NT}^*}{(a_{NT}^2 V_{it} + b_{NT}^2 W_{it})^{1/2}} \xrightarrow{d} N(0, 1).$$

The above is true with $(\xi_{NT}^*, \zeta_{NT}^*)$ replaced by (ξ_{NT}, ζ_{NT}) because they have the same distribution. This implies that, in view of (C.1),

$$\frac{\delta_{NT}(\widetilde{C}_{it} - C_{it}^0)}{(a_{NT}^2 V_{it} + b_{NT}^2 W_{it})^{1/2}} \xrightarrow{d} N(0, 1).$$

The above is (5), which is equivalent to Theorem 3. *Q.E.D.*

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 $(T > r)$ with $\operatorname{rank}(A) = r$, 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 of $AA' + \Omega$ 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' are 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: Consider the case of $r = 1$. Let η_i be a $T \times 1$ vector with the ith element being 1 and 0 elsewhere. For example, $\eta_1 = (1, 0, \ldots, 0)'$. 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. Applying the same reasoning with $A = \eta_i$ ($i = 1, 2, \ldots, T$), we conclude Ω is a diagonal matrix such that $\Omega = \text{diag}(c_1, c_2, \ldots, c_T)$. Next we argue the constants c_i must be the same. Let $A = \eta_1 + \eta_2 = (1, 1, 0, \ldots 0)'$. The lemma's assumption implies that $\Gamma = (1/\sqrt{2}, 1/\sqrt{2}, 0, \ldots, 0)'$ is an eigenvector of $AA' + \Omega$. It is easy to verify that $(AA' + \Omega)\Gamma = \Gamma d$ for some scalar d implies that $c_1 = c_2$. Similar reasoning shows all the c_j's are the same. This proves the lemma for $r = 1$. The proof for general r is omitted to conserve space, but is available from the author (the proof was also given in an earlier version of this paper).   *Q.E.D.*

PROOF OF THEOREM 4: In this proof, all limits are taken as $N \to \infty$. Let $\Psi = \text{plim}_{N \to \infty} N^{-1}ee' = \text{plim } N^{-1} \sum_{i=1}^{N} \underline{e}_i \underline{e}_i'$. That is, the (t, s)th entry of Ψ is the limit of $(1/N) \sum_{i=1}^{N} e_{it}e_{is}$. 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 $(1/TN)XX' \xrightarrow{p} B$ with $B = (1/T)F^0 \Sigma_\Lambda F^{0\prime} + (1/T)\Psi$ 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 (6) 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 separated from 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 $(1/TN)XX' \xrightarrow{p} B$, 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 from the rest of the eigenvalues; see, e.g., Bhatia (1997). If $\widetilde{F}$ is consistent for some transformation of F^0, that is, $\|\widetilde{F} - F^0 D\| \to 0$ with D being an $r \times r$ 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 $\Gamma = F^0 C$ for some $r \times r$ invertible matrix C. Since consistency requires that this be true for every F^0, not just for a particular F^0, we see the existence of r eigenvectors of B in the form of $F^0 C$ for all F^0. Applying 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 (6).   *Q.E.D.*

PROOF OF THEOREM 5: By the definition of $\widetilde{F}$ and V_{NT}, we have $(1/NT)XX'\widetilde{F} = \widetilde{F}V_{NT}$. Since W and $W + cI$ have the same set of eigenvectors for an arbitrary matrix W, we have

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

Right multiply $J_{NT} = (V_{NT} - T^{-1}\bar{\sigma}_N^2 I_r)^{-1}$ on both sides to obtain

$$\left(\frac{1}{NT}XX' - T^{-1}\bar{\sigma}_N^2 I_T\right)\widetilde{F}J_{NT} = \widetilde{F}.$$

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

$$(D.1) \qquad \widetilde{F}_t - \bar{H}'F_t^0 = J_{NT}T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} + J_{NT}T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} + J_{NT}T^{-1}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st},$$

where $\bar{H} = HV_{NT}J_{NT}$. Equation (D.1) is similar to (A.1). But the first term on the right-hand side of (A.1) disappears and V_{NT}^{-1} is replaced by J_{NT}. The middle term of (D.1) is of $O_p(N^{-1/2})$ and each of the other two terms is $O_p(N^{-1/2}\delta_{NT}^{-1})$ by Lemma A.2. The rest of the proof is the same as that of Theorem 1, except no restriction between N and T is needed. In addition,

$$\Gamma_t = \Gamma = \lim \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^0\lambda_j^{0\prime}E(e_{it}e_{jt}) = \lim \frac{1}{N}\Lambda^{0\prime}\Omega\Lambda^0. \qquad\qquad Q.E.D.$$

APPENDIX E: PROOF OF THEOREM 6

First we show $\widehat{\Pi}_t$ is consistent for Π_t. A detailed proof will involve four steps:

(i) $\quad \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i' - \frac{1}{N}\sum_{i=1}^{N}e_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i' = o_p(1);$

(ii) $\quad \frac{1}{N}\sum_{i=1}^{N}e_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i' - H^{-1}(\frac{1}{N}\sum_{i=1}^{N}e_{it}^2\lambda_i^0\lambda_i^{0\prime})(H')^{-1} = o_p(1);$

(iii) $\quad \frac{1}{N}\sum_{i=1}^{N}e_{it}^2\lambda_i^0\lambda_i^{0\prime} - \frac{1}{N}\sum_{i=1}^{N}\sigma_{it}^2\lambda_i^0\lambda_i^{0\prime} = o_p(1);$

(iv) $\quad H^{-1} \to Q.$

The first two steps imply that $\tilde{e}_{it}^2$ can be replaced by e_{it}^2 and $\tilde{\lambda}_i$ can be replaced by $H^{-1}\lambda_i^0$. 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 $\tilde{e}_{it} = e_{it} + O_p(\delta_{NT}^{-1})$ and $\tilde{\lambda}_i = H^{-1}\lambda_i^0 + O_p(\delta_{NT}^{-1})$, which are the consequences of Theorem 3 and Theorem 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

$$\widehat{\Pi}_t = (V_{NT}^{-1})\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i'(V_{NT}^{-1}) \xrightarrow{p} V^{-1}Q\Gamma_tQ'V^{-1} = \Pi_t.$$

Next, we prove $\widehat{\Theta}_i$ is consistent for Θ_i. Because $\widetilde{F}_t$ is estimating $H'F_t^0$, the HAC estimator $\widehat{\Theta}_i$ based on $\widetilde{F}_t\tilde{e}_{it}$ $(t = 1, 2, \ldots, T)$ is estimating $H^{0\prime}\Phi_iH^0$, where H^0 is the limit of H. The consistency of $\widehat{\Theta}_i$ for $H^{0\prime}\Phi_iH^0$ can be proved using the argument of Newey and West (1987). Now

$$H' = V_{NT}^{-1}(\widetilde{F}'F^0/T)(\Lambda^{0\prime}\Lambda/N) \to V^{-1}Q\Sigma_\Lambda.$$

The latter matrix is equal to Q'^{-1} (see Proposition 1). Thus $\widehat{\Theta}_i$ is consistent for $Q'^{-1}\Phi_iQ^{-1}$.

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

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

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 an $r \times r$ invertible matrix. The consistency of $\widehat{V}_{it}$ now follows from the consistency of $\tilde{\lambda}_j$ for $H^{-1}\lambda_j^0$.

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

REFERENCES

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

———— (1984): *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681.

BERNANKE, B., AND J. BOIVIN (2000): "Monetary Policy in a Data Rich Environment," Unpublished Manuscript, Princeton University.

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

BRILLINGER, D. R. (1981): *Time Series Data Analysis and Theory*. New York: Holt, 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 M. ROTHSCHILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305–1324.

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

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

———— (1993): "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, 48, 1263–1291.

CRISTADORO, R., M. FORNI, L. REICHLIN, AND G. VERONESE (2001): "A Core Inflation Index for the Euro Area," Unpublished Manuscript, CEPR.

FAVERO, C. A., AND M. MARCELLINO (2001): "Large Datasets, Small Models And Monetary Policy in Europe," Unpublished Manuscript, IEP, Bocconi University, and CEPR.

FORNI, M., AND M. LIPPI (1997): *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford, U.K.: Oxford University Press.

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

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000a): "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.

———— (2000b): "Reference Cycles: The NBER Methodology Revisited," CEPR Discussion Paper 2400.

———— (2000c): "The Generalized Dynamic Factor Model: Consistency and Rates," Unpublished Manuscript, CEPR.

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

GORMAN, W. M. (1981): "Some Engel Curves," in *Essays in the Theory and Measurement of Consumer Behavior in Honor of Sir Richard Stone*, ed. by A. Deaton. New York: Cambridge University Press.

GRANGER, C. W. J. (2001): "Macroeconometrics—Past and Future," *Journal of Econometrics*, 100, 17–19.

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

HAHN, J., AND G. KUERSTEINER (2000): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large," Unpublished Manuscript, Department of Economics, MIT.

LAWLEY, D. N., AND A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. London: Butterworth.

LEHMANN, B. N., AND D. MODEST (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.

LINTNER, J. (1965): "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

POLLARD, D. (1984): *Convergence of Stochastic Processes.* New York: Springer-Verlag.

REICHLIN, L. (2000): "Extracting Business Cycle Indexes from Large Data Sets: Aggregation, Estimation, and Identification," Paper prepared for the 2000 World Congress.

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

SHARPE, W. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

STOCK, J. H., AND M. WATSON (1989): "New Indexes of Coincident and Leading Economic Indications," in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard and S. Fischer. Cambridge: MIT Press.

——— (1998): "Diffusion Indexes," NBER Working Paper 6702.

——— (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

TONG, Y. (2000): "When Are Momentum Profits Due to Common Factor Dynamics?" Unpublished Manuscript, Department of Finance, Boston College.

WANSBEEK, T., AND E. MEIJER (2000): *Measurement Error and Latent Variables in Econometrics.* Amsterdam: North-Holland.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," *Econometrica*, 48, 817–838.