Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 122 (2004) 137–183
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Estimating cross-section common stochastic trends in nonstationary panel data

Jushan Bai*

Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA

Accepted 23 October 2003

Abstract

This paper studies large-dimension factor models with nonstationary dynamic factors, also referred to as cross-section common stochastic trends. We consider the problem of estimating the dimension of the common stochastic trends and the stochastic trends themselves. We derive the rates of convergence and the limiting distributions for the estimated common trends and for the estimated loading coefficients. Generalized dynamic factor models with nonstationary factors are also considered. Cointegration among the factors is permitted. The method is applied to the study of employment fluctuations across 60 industries for the U.S. We examine the hypothesis that these fluctuations can be explained by a small number of aggregate factors. We also test whether some observable macroeconomic variables are the underlying factors.

JEL classification: C13; C33; C43

Keywords: Common-stochastic trends; Dynamic factors; Generalized dynamic factor models; Principal components; Nonstationary panel data

1. Introduction

Factor models provide an effective way of synthesizing information contained in large data sets. The latter are increasingly available due to the advancement in data gathering technologies and the natural expansion of data sets over time. In this paper, we examine large-dimension factor models with nonstationary dynamic factors. We consider the problem of estimating the dimension of the factors and deriving the distribution theory of the estimated factors and of the factor loadings.

* Tel.: +212-998-8960.
E-mail address: jushan.bai@nyu.edu (J. Bai).

doi:10.1016/j.jeconom.2003.10.022

Dynamic factor models are useful in at least five areas of economic analysis. The first is index modelling and extraction. Factors are regarded as unobservable economic indices that capture the co-movement of many variables; see Quah and Sargent (1993), Forni and Reichlin (1998) and Stock and Watson (1999a). The second is information synthesizing. Based on large-dimension factor models, Bernanke and Boivin (2000) examine, among other inquiries, the state of an economy when an expert artificial intelligence system instead of the Fed sets the monetary policy. Their analysis demonstrates the usefulness of factor models as a way of aggregating information from thousands of economic indicators. The third is forecasting. Stock and Watson (1999b) and Favero and Marcellino (2001), and Artis et al. (2001) show how dynamic factor models can be used to improve forecasting accuracy. The fourth is modelling cross-section correlations. One major source of cross-section correlation in macroeconomic data is common shocks, e.g., oil price shocks and international financial crises. Cross-section correlation of this nature may well be characterized by common factor models. Consequently, factor models are used in studying world business cycles as in Gregory and Head (1999), and area-wide business cycles as in Forni and Reichlin (1998) and Forni et al. (2000b). Cross-section correlation exists even in micro-level data (e.g., household data) because of herd behavior, fashions or fads. The general state of an economy, such as recessions or booms, also affects household decision making. Factor models allow for heterogeneous responses to common shocks through heterogeneous factor loadings.

Finally, factor models can be used to study cross-section cointegration in nonstationary panel data. In testing the PPP hypothesis, for example, Banerjee et al. (2001) argued that exchange rate series for the European countries share a common trend, which implies strong cross-section correlation. Therefore, the standard assumption of cross-section independence is violated, rendering many panel unit root tests invalid. Hall et al. (1999) considered the problem of determining the number of common trends, but not how to estimate the common factors.

In all the above applications, an often-asked question is how well the factors (or unobservable economic indices) are estimated. Are the estimated factors a transformation of the underlying factors? Can confidence intervals be constructed to assess the accuracy of the estimates? What are the convergence rates of the estimates? Are the estimated factors uniformly consistent for the underlying factors? Under what conditions can the estimated factors be treated as known or the sampling errors be ignored? How important is the common component relative to the idiosyncratic component? This paper provides answers to these questions.

The contributions of this paper include (i) determining the number of nonstationary factors; (ii) deriving the limiting distributions for the estimated factors, factor loadings, and common components; (iii) construction of confidence intervals; (iv) empirical application of the technique. Consistent estimation of the number of factors is important because it enables us to disentangle the common and idiosyncratic components, thereby allowing us to measure the relative importance of each component. The rates of convergence and the limiting distributions allow us to assess the accuracy of the estimates and to construct confidence intervals. The results are useful in testing whether an observable series is one of the underlying factors. Although not a focus of this paper, our results have implications for unit root and cointegration analysis for nonstationary panel data. This topic was recently explored by Bai and Ng (2001) and Moon and Perron (2001) based on factor models. For a recent survey on panel unit roots and cointegration, the readers are referred to Baltagi and Kao (2000). The present paper builds on the work of Bai and Ng (2002) and Bai (2003), where large-dimension factor models of I(0) variables were analyzed. This paper makes a contribution to statistical inference for large-dimension factor models of nonstationary variables. In addition, we derive some inferential theory for generalized dynamic factor models most recently studied by Forni et al. (2000a).

The rest of the paper is organized as follows. Section 2 introduces the model and states the underlying assumptions. We also discuss the estimation procedure and derive some useful properties for the estimated common stochastic trends. Section 3 focuses on determining the number of common trends. Section 4 develops the asymptotic distribution theory for the estimated common trends and the estimated factor loadings. The construction of confidence intervals is also examined. Section 5 analyzes generalized dynamic factor models. Section 6 reports simulation results. In Section 7, we apply our technique to sectoral employment. Section 8 provides concluding remarks. All proofs are provided in the appendix.

2. Model, assumptions, and estimation

We first consider a restricted dynamic factor model in which the factors are dynamic but the relation between the dynamic factors and the observable variable is static. Generalized dynamic models as in Forni et al. (2000a) will be considered later. The analysis of this simpler model gives insight into the general model.

2.1. Model and assumptions

Consider:

$$X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{jt} + e_{it} = \lambda_i' F_t + e_{it} \tag{1}$$

$(i = 1, 2, \ldots, N; t = 1, 2, \ldots, T)$, where e_{it} is an I(0) error process which can be serially correlated for each i, $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{ir})'$, and $F_t = (F_{1t}, \ldots, F_{rt})'$ is a vector of integrated processes such that

$$F_t = F_{t-1} + u_t,$$

and u_t is a vector $(r \times 1)$ of zero-mean I(0) processes (not necessarily i.i.d.) that drive the stochastic trends F_t. For each given i, the process X_{it} is I(1) unless $\lambda_i = 0$. It is clear that X_{it} and F_t are cointegrated for each i. We assume no cointegration among F_t. The case in which F_t is cointegrated is discussed in Section 5.

The right-hand side variables are not observable and only X_{it} is observable. Note that X_{it} and X_{jt} are correlated because they share the same F_t. While having a similar format, model (1) is different from the common trend representation of Stock and Watson (1988) in which the errors driving the common trends and the idiosyncratic errors are the same. The latter property may not be suitable for panel data. In addition, the number of series N is fixed in Stock and Watson (1988). In this paper, we consider the case in which N goes to infinity.

The current set up differs from the classical factor analysis in a number of ways. The size of cross-section units N (or the number of variables) goes to infinity rather than a fixed number; the factors F_t are integrated processes rather than i.i.d. variables; the errors e_{it} are allowed to be correlated in both dimensions rather than i.i.d. in the time dimension and independent in the cross-section dimension. For classical factor analysis, see Lawley and Maxwell (1971) and Anderson (1984).

In classical factor analysis, with N fixed and with T being allowed to grow, an unrestricted variance–covariance matrix $E(e_t e_t')$ will render the factor model completely unidentifiable, where $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$. But this is no longer true if N also tends to infinity; see Chamberlain and Rothschild (1983), who introduced the notion of approximate factor models to allow for cross-section correlation in e_{it}.

In what follows, we use F_t^0, λ_i^0 and r to denote the true common trends, the true factor loading coefficient, and the true number of trends, respectively. At a given t, we have

$$X_t = \Lambda^0 F_t^0 + e_t, \tag{2}$$

where $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$, $\Lambda^0 = (\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0)'$, and $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$.

Let $\underline{X}_i$ be a $T \times 1$ vector of time series observations for the ith cross-section unit. For a given i, we have

$$\underline{X}_i = F^0 \lambda_i^0 + \underline{e}_i, \tag{3}$$

where $\underline{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$, $F^0 = (F_1^0, F_2^0, \ldots, F_T^0)'$, and $\underline{e}_i = (e_{i1}, e_{i2}, \ldots, e_{iT})'$.

Our objective is to estimate r, F^0, and Λ^0. Clearly, F^0 and Λ^0 are not separately identifiable. But they are identifiable up to a transformation. We shall state the assumptions needed for consistent estimation. Throughout, the norm of a matrix A is defined as $\|A\| = [tr(A'A)]^{1/2}$. The notation M stands for a finite positive constant, not depending on N and T; $B(\cdot)$ stands for a Brownian motion process on $[0, 1]$. In addition, $\int BB'$ means $\int_0^1 B(\tau)B(\tau)'\,\mathrm{d}\tau$, etc.

Assumption A (Common stochastic trends). (1) $E\|u_t\|^{4+\delta} \leqslant M$ for some $\delta > 0$ and for all $t \leqslant T$.

(2) As $T \to \infty$, $T^{-2}\sum_{t=1}^T F_t^0 F_t^{0\prime} \xrightarrow{\mathrm{d}} \int B_u B_u'$, where B_u is a vector of Brownian motions with covariance $\Omega_{uu} = \lim_{T\to\infty} 1/T \sum_{s=1}^T \sum_{t=1}^T E(u_t u_s')$; the $r \times r$ matrix Ω_{uu} is positive definite.

(3) (iterated logarithm) $\liminf_{T\to\infty} \log\log(T) T^{-2} \sum_{t=1}^T F_t^0 F_t^{0\prime} = D$, where D is a nonrandom positive definite matrix.

(4) (initial value) $E\|F_0^0\|^4 \leqslant M$.

Assumption B (Heterogeneous factor loadings). The loading λ_i is either deterministic such that $\|\lambda_i^0\| \leqslant M$ or it is stochastic such that $E\|\lambda_i^0\|^4 \leqslant M$. In either case, $\Lambda^{0\prime}\Lambda^0/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$ for some $r \times r$ positive definite non-random matrix Σ_Λ.

Assumption C (Time and cross-section dependence and heteroskedasticity). (1) $E(e_{it}) = 0$, $E|e_{it}|^8 \leqslant M$;

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

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

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

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

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

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

Assumption D. $\{\lambda_i\}, \{u_t\}$, and $\{e_{it}\}$ are three groups of mutually independent stochastic variables.

Sufficient conditions for A2 can be found in Hansen (1992). The assumption that Ω_{uu} is positive definite rules out cointegration among the components of F_t. Cointegration for F_t is equivalent to the presence of both I(1) and I(0) common factors. This case is considered in Section 5. There, it is shown that the number of I(1) and I(0) common factors can be separately identified, and that the cointegrating rank can be consistently estimated.

Assumption A3 is implied by the law of the iterated logarithm; see Lai and Wei (1983). The limiting matrix being positive definite follows from the Cramer-Wold device and the law of the iterated logarithm. Assumption B is standard. Assumption C allows for limited time series and cross-section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross-section dimensions is also allowed. Under stationarity in the time dimension, $\gamma_N(s,t) = \gamma_N(s-t)$. Given Assumption C1, the remaining assumptions in C are easily satisfied if the e_{it} are independent for all i and t. The allowance for weak cross-section correlation in the idiosyncratic components leads to the *approximate factor structure* of Chamberlain and Rothschild (1983). It is more general than a *strict factor model* which assumes e_{it} is uncorrelated across i, a framework underpinning classical factor analysis.

In terms of estimating the dimension of I(1) common factors, Assumption C is more than necessary. For example, we can assume $e_{it} = \tau_i'\eta_t + \varepsilon_{it}$, where η_t are I(0) series and ε_{it} satisfies Assumption C. This implies strong cross-section correlations for e_{it} because η_t are common I(0) factors. Again this is related to cointegration, and the number of I(1) common trends can be consistently determined. In a special case that η_t is equal to u_t, the model implies dynamic factor structure to be explained below. Assumption C is used when deriving the limiting distributions of the estimated common factors and factor loadings.

Assumption D rules out correlation between e_{it} and u_t. Suppose e_{it} and u_t are correlated and consider projecting e_{it} on u_t such that $e_{it} = \tau_i' u_t + \varepsilon_{it}$, with ε_{it} being uncorrelated with u_t. From $u_t = F_t - F_{t-1}$, we can rewrite the factor model as $X_{it} = (\lambda_i + \tau_i)' F_t - \tau_i' F_{t-1} + \varepsilon_{it}$, giving rise to a dynamic factor model. As shown in Section 5, the dimension of F_t can still be consistently estimated.

2.2. Estimating common stochastic trends

In this section we first outline the procedure for estimating common stochastic trends and then derive some preliminary convergence results for the estimated common trends. These results are required for analyzing the properties of the dimension estimator.

Because the true dimension r is unknown, we start with an arbitrary number k ($k < min\{N,T\}$). The superscript in λ_i^k and F_t^k highlights the allowance for k stochastic trends in the estimation. Estimates of Λ^k and F^k are obtained by solving the optimization problem

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} F_t^k)^2 \tag{4}$$

subject to the normalization of either $F^{k\prime}F^k/T^2 = I_k$ or $\Lambda^{k\prime}\Lambda^k/N = I_k$, where T^2 and N correspond to the rates in Assumptions A2 and B, respectively. If we concentrate out Λ^k and use the normalization that $F^{k\prime}F^k/T^2 = I_k$, the optimization problem is identical to maximizing $tr(F^{k\prime}(XX')F^k)$, where $X = (\underline{X}_1, \ldots, \underline{X}_N)$ is $T \times N$. The estimated common-trend matrix, denoted by $\tilde{F}^k$, is T times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix XX'. Given $\tilde{F}^k$, we have $\tilde{\Lambda}^{k\prime} = (\tilde{F}^{k\prime}\tilde{F}^k)^{-1}\tilde{F}^{k\prime}X = \tilde{F}^{k\prime}X/T^2$, which is the corresponding matrix of the estimated factor loadings. Clearly, this is the method of principal components and is used by many researchers, e.g., Connor and Korajczyk (1986) for large N but fixed T and Stock and Watson (1999a) for large N and large T.

The solution to the above minimization problem is not unique, even though the sum of squared residuals $V(k)$ is unique. Another useful solution is $(\bar{F}^k, \bar{\Lambda}^k)$, where $\bar{\Lambda}^k$ is constructed as $\sqrt{N}$ times the eigenvectors corresponding to the k largest eigenvalues of the $N \times N$ matrix $X'X$ and $\bar{F}^k = X\bar{\Lambda}^k/N$. The second solution is easier to compute when $N < T$ and the first is easier when $T < N$.

Define

$$\hat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k/T^2)^{1/2} \quad \text{and} \quad \hat{\Lambda}^k = \bar{\Lambda}^k(\bar{F}^{k\prime}\bar{F}^k/T^2)^{-1/2} \tag{5}$$

a rescaled version of $(\bar{F}^k, \bar{\Lambda}^k)$. The following lemma provides a preliminary but useful property for the estimated common stochastic trends:

Lemma 1. *Assume Assumptions* A–D *hold. For each fixed* $k \geqslant 1$, *there exists an* $(r \times k)$ *matrix* H^k *with* $rank(H^k) = \min\{k, r\}$, *and* $\delta_{NT} = \min\left\{\sqrt{N}, T\right\}$, *such that*

$$\delta_{NT}^2 \left(\frac{1}{T} \sum_{t=1}^{T} \|\hat{F}_t^k - H^{k\prime} F_t^0\|^2 \right) = O_p(1). \tag{6}$$

This lemma is crucial for analyzing the estimated dimension. It is also used when developing the limiting distributions for the various estimators. Because the trends (F^0) can only be identified up to transformation, the principal components method is estimating a rotation of F^0. The lemma says that the time average of the squared deviations between the estimated trends and a transformation of the true trends converges to zero as $N,\ T \to \infty$. The rate of convergence is the minimum of N and T^2.

Stock and Watson (1999a) considered uniform consistency of $\hat{F}_t$ when $N \gg T^2$ and for stationary F_t^0. Using different arguments, stronger uniform result can be obtained:

Proposition 1. *Under assumptions* A–D *and* E1 (*below*),

$$\max_{1 \leqslant t \leqslant T} \|\hat{F}_t^k - H^{k\prime} F_t^0\| = O_p(T^{-1}) + O_p\left(\sqrt{T/N}\right).$$

This bound is not the sharpest possible. The term $O_p\left(\sqrt{T/N}\right)$ can be replaced by $O_p\left(\log T/\sqrt{N}\right)$ if enough moment conditions (e.g., normality assumption) are imposed on the idiosyncratic errors of e_{it}. This lemma implies that when N is sufficiently large relative to T, the estimated common stochastic trends are uniformly consistent. This result is of independent interest.

3. The number of common stochastic trends

3.1. Using data in differences

A useful observation is that the differenced data satisfy all assumptions of Bai and Ng (2002). Thus their information criterion approach is directly applicable to the differenced data. Model (1) under first differencing takes the form

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

Note that the idiosyncratic errors are over-differenced. But over-differencing does not violate any conditions of Bai and Ng. Let

$$V(k) = \min_{\Lambda^k, U^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\Delta X_{it} - \lambda_i^k u_t^k)^2, \tag{7}$$

where $U^k = (u_1^k, u_2^k, \ldots, u_T^k)'$. Consider the criterion of the form:

$$PC(k) = V(k) + kg(N, T),$$

where $g(N,T)$ is a penalty function. Let $kmax$ be a positive integer such that $r < kmax$ and let

$$\hat{k} = \underset{0 \leqslant k \leqslant kmax}{\arg\min}\ PC(k). \tag{8}$$

Denote $C_{NT} = \min\left[\sqrt{N}, \sqrt{T}\right]$. Theorem 2 of Bai and Ng (2002) implies that

Proposition 2. *Under assumptions of* A–D, *if* (i) $g(N,T) \to 0$ *and* (ii) $C_{NT}^2 g(N,T) \to \infty$, *then* $\lim_{N,T\to\infty} P(\hat{k} = r) = 1$.

Let $\hat{\sigma}^2$ be a consistent estimate of $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E(\Delta e_{it})^2$, say $\hat{\sigma}^2 = V(kmax)$. The criteria in (12) (see below) with $\alpha_T = 1$ satisfy the conditions of this proposition.[1] They are derived by Bai and Ng (2002) from the properties of $V(k)$ as a function of k.

3.2. New criteria for data in levels

It is not always desirable to difference the data. For example, for the generalized dynamic factor models of Section 5, differencing the data will not yield consistent estimates for the number of I(1) trends. By introducing new criteria to reflect integrated common trends, we show that the dimension of stochastic trends can be consistently estimated without the need of differencing. This also allows us to simultaneously estimate the stochastic trends themselves and the dimension. Let

$$V(k) = V(k, \hat{F}^k) = \min_{\Lambda^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} \hat{F}_t^k)^2 \tag{9}$$

denote the sum of squared residuals (divided by NT) when k trends are estimated.[2] The objective is to find penalty functions, $g(N,T)$, such that criteria of the form

$$IPC(k) = V(k) + kg(N,T) \tag{10}$$

can consistently estimate r, where the label "IPC" refers to "Integrated Panel Criterion." Again, assume $r < kmax$ and let

$$\hat{k} = \underset{0 \leqslant k \leqslant kmax}{\operatorname{argmin}}\ IPC(k). \tag{11}$$

Theorem 1. *Assume Assumptions* A–D *hold. As* $N, T \to \infty$, *if*

(i) $g(N,T)\frac{\log\log(T)}{T} \to 0$,
(ii) $g(N,T) \to \infty$,

then $\lim_{N,T\to\infty} Prob(\hat{k} = r) = 1$.

[1] For the third criterion, condition (ii) will be violated if N is excessively large relative to T (say $N = \exp(T)$) and vice versa.

[2] This sum of squared residuals does not depend on which estimate of F^0 is used. That is, $V(k) = V(k, \tilde{F}^k) = V(k, \bar{F}^k) = V(k, \hat{F}^k)$, where the three different estimates for F^0 are defined in Section 2.

Note that $g(N,T)$ diverges to infinity, rather than converging to zero as in Proposition 2. While a formal proof of Theorem 1 is provided in the appendix, its rationale is this. When $k < r$, the normalized sum of squared residuals, $V(k)$, is such that $V(k) = \mathrm{O}(T/\log\log(T))$, and when $k \geqslant r$, $V(k) = \mathrm{O_p}(1)$. Thus a penalty function diverging at a slower rate than $T/\log\log(T)$ will pick up at least r trends. But it will not pick up more than r trends because the reduction in sum of squared residuals is only $\mathrm{O_p}(1)$, whereas the cost (penalty) for estimating one more dimension is a larger magnitude. Thus, the criterion function is minimized at r with probability tending to one.

Let $\alpha_T = T/[4\log\log(T)]$, which relates to the law of the iterated logarithm. Let $\hat{\sigma}^2$ be an estimate of $(NT)^{-1}\sum_{i=1}^N \sum_{t=1}^T E(e_{it})^2$. In practice, $\hat{\sigma}^2 = V(kmax)$. Consider the following criteria:

$$IPC_1(k) = V(k) + k\hat{\sigma}^2\alpha_T\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right);$$

$$IPC_2(k) = V(k) + k\hat{\sigma}^2\alpha_T\left(\frac{N+T}{NT}\right)\log C_{NT}^2;$$

$$IPC_3(k) = V(k) + k\hat{\sigma}^2\alpha_T\left(\frac{N+T-k}{NT}\right)\log(NT). \tag{12}$$

The first two criteria satisfy the requirements of Theorem 1. That is, $\frac{\log\log(T)}{T}g(N,T) \to 0$ and $g(N,T) \to \infty$. This is also true for the last criterion unless N is too large relative to T such as $N = \mathrm{O}(e^T)$ (it is still valid when T is large relative to N).[3]

Remark 1. The conditions of Theorem 1 actually imply strong consistency, that is, $P(\lim_{N,T\to\infty}\hat{k} = r) = 1$. Many of the existing results on dimension selection, e.g., Geweke and Meese (1981), although stated in terms of consistency, imply strong consistency. Others such as Hannan and Quinn (1979) are explicit about strong consistency. In our present situation, strong consistency is more desirable than mere consistency. Although it is nontrivial to see why, the underlying reason is related to the limiting behavior of $T^{-2}\sum_{t=1}^T F_t^0 F_t^{0\prime}$. While this matrix converges in distribution to a (stochastic) positive-definite matrix, its infimum limit (i.e., "lim inf") is zero by the law of the iterated logarithm. The implication is that some consistency criteria do not imply strong consistency. For example, replacing condition (i) with $g(N,T)/T \to 0$ implies consistency, but cannot guarantee strong consistency.

Remark 2. A penalty function satisfying conditions of Theorem 1 is able to pick up I(2) nonstationary factors (it can also pick up linear or polynomial trends). However, to separately identify the number of I(1) and the number of I(2) common trends, a heavier penalty function that can pick up I(2) but not I(1) trends is also required. Such a criterion is not considered in this paper.

[3] It is also possible to consider criteria based on log-valued sum of squared residuals such that $\log V(k) + kg(N,T)$. In this case, the scaling factor $\hat{\sigma}^2$ is not necessary. It can be shown that if (i) $g(N,T)/\log(T) \to 0$ and (ii) $g(N,T) \to \infty$, then $\hat{k}$ is consistent for r. The details are omitted.

4. Distribution theory

In this section we investigate the limiting distribution of the estimated common stochastic trends, the estimated loading coefficients, and the estimated common components. The number of common stochastic trends (r) is assumed to be known in this section and k is taken to be r. The limiting distribution is not affected when r is estimated because of consistent estimation and the discreteness of r. We shall simply write the estimated common trends and the estimated loading coefficients as $\hat{F}_t$ and $\hat{\lambda}_i$ without the superscript k. See Section 2.2 for the definition of these estimators. We shall write H^k as H_1, which is a $r \times r$ matrix with full rank. In matrix notation, $\hat{F}$ is estimating $F^0 H_1$. We will also show that $\tilde{F}$ is estimating $F^0 H_2$ for some $r \times r$ matrix H_2 of full rank. In addition, we show $\hat{\Lambda}$ is an estimator for $\Lambda^0 (H_1')^{-1}$ and $\tilde{\Lambda}$ is an estimator for $\Lambda^0 (H_2')^{-1}$. While F^0 and Λ^0 are not separately identifiable, they can be estimated up to a transformation. It is clear that $\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}'$ is an estimator of $F^0 \Lambda^{0\prime}$, the common components. That is, the common components are directly identifiable. For many purposes, knowing $F^0 H_1$ is as good as knowing F^0. For example, in regression analysis, using F^0 as the regressor will give the same predicted value of a left-hand side variable as using $F^0 H_1$ as the regressor. Because F^0 and $F^0 H_1$ span the same space, testing the significance of F^0 in a regression model containing F^0 as regressors is the same as testing the significance of $F^0 H_1$.

Additional assumptions are needed to derive the limiting distributions.

Assumption E (Weak dependence of idiosyncratic errors). For all T and N,

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

where $\gamma_N(s,t)$ and τ_{ki} are defined in Assumption C.

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

Lemma 2. *Under Assumptions* A–E, *we have, for each* t

$$\min\left\{\sqrt{N}, T^{3/2}\right\}(\hat{F}_t - H_1' F_t^0) = O_p(1).$$

The convergence rate is $\min\{\sqrt{N}, T^{3/2}\}$. When the loading coefficients λ_i^0 ($i = 1,2,\ldots,N$) are all known, F_t^0 can be estimated by the least-squares method using the cross-section data at period t, and the rate of convergence will be $\sqrt{N}$. The current rate of convergence applies because the coefficients λ_i^0 are unknown and are also estimated.

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

$$Y_{t+1} = \alpha' F_t^0 + \beta' W_t + \eta_{t+1}, \quad t = 1,2,\ldots,T,$$

where Y_t is scalar, and Y_t and W_t are observable. The integrated process F_t^0 is not observable but can be replaced by $\hat{F}_t$. Let $\hat{\alpha}$ denote the least-squares estimator with $\hat{F}$ as the regressor. The limiting distribution of the estimated coefficients is the same as if F^0H_1 were used as long as N is large relative to T. Note that $\hat{\alpha}$ is estimating $H_1^{-1}\alpha$ rather than α.

Assumption F. The eigenvalues of the $r \times r$ random matrix $\Sigma_\Lambda^{1/2}(\int B_u B_u')\Sigma_\Lambda^{1/2}$ are distinct with probability 1.

The matrices $\int B_u B_u'$ and Σ_Λ are defined in Assumptions A and B. Assumption F guarantees a well-defined limiting random variable for $(\tilde{F}'F^0/T^2)$, which appears in the limiting distributions of $\hat{F}_t$. A similar condition is imposed in classical factor analysis, see Anderson (1963). Note that this assumption is not needed for determining the number of common trends. This is because the number of trends is estimated using the sum of squared residuals $V(r)$, which is well defined (unique) regardless of the distinctness of the said eigenvalues. Also, Assumption F is not required for studying the limiting distribution of the estimated common components. The underlying reason is that the common components are identifiable. In the following analysis, we will use the fact that for positive-definite matrices A and B, the eigenvalues of AB, BA and $A^{1/2}BA^{1/2}$, etc., are the same.

Proposition 3. *Under Assumptions* A–F, *as* $N, T \to \infty$,

$$\frac{\tilde{F}'F^0}{T^2} \xrightarrow{d} Q.$$

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

The proof is provided in the appendix. The random matrix Q appears in the limiting distributions of $\hat{F}_t$ and $\hat{\lambda}_i$.

4.1. Limiting distribution of estimated common trends

Assumptions B and C imply that $N^{-1/2}\sum_{i=1}^N \lambda_i^0 e_{it} = O_p(1)$ for each t. It is now strengthened to have a normal limiting distribution as N tends to infinity.

Assumption G (Central limit theorem). 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 $\Gamma_t = \lim_{N\to\infty}(1/N)\sum_{i=1}^N\sum_{j=1}^N E(\lambda_i^0\lambda_j^{0\prime}e_{it}e_{jt})$.

Theorem 2. *Under Assumptions* A–G, *as* $N, T \to \infty$ *with* $N/T^3 \to 0$, *we have*, *for each* t

$$\sqrt{N}(\hat{F}_t - H_1' F_t^0) = \left(\frac{\tilde{F}' F^0}{T^2}\right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 e_{it} + \mathrm{o_p}(1)$$

$$\xrightarrow{\mathrm{d}} QN(0, \Gamma_t),$$

where Q *is defined in Proposition* 3 *and* Γ_t *is defined in Assumption* G. *In addition*, Q *is independent of* $N(0, \Gamma_t)$.

Corollary 1. *Under Assumptions* A–G *and* $N/T^3 \to 0$, *then* $\sqrt{N}(\tilde{F}_t - H_2' F_t^0) \xrightarrow{\mathrm{d}} V^{-1} QN(0, \Gamma_t)$, *where* H_2 *is a* $r \times r$ *matrix of rank* r, *and* Q *and* V *are defined in Proposition* 3. *In addition*, Q *and* V *are independent of* $N(0, \Gamma_t)$.

While restrictions on N and T are needed, the theorem is not a sequential limiting 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 that $N/T^3 \to 0$ is a weak one and is easily satisfied.

4.2. Limiting distribution of estimated factor loadings

The previous subsection shows that $\hat{F}$ is estimating $F^0 H_1$. Now we show that $\hat{\Lambda}$ is estimating $\Lambda^0 (H_1')^{-1}$. That is, $\hat{\lambda}_i$ is estimating $H_1^{-1} \lambda_i^0$ for every i. First, we state the rate of convergence.

Lemma 3. *Under Assumptions* A–E, *we have*, *for each* i,

$$T(\hat{\lambda}_i - H_1^{-1} \lambda_i^0) = \mathrm{O_p}(1).$$

It is worth noting that the rate of convergence depends only on Assumptions A–E. This rate is the same as if F_t^0 $(t = 1, 2, \ldots, T)$ were known. But if F_t^0 were observable for each t, then we could directly estimate λ_i^0, as opposed to a matrix transformation of λ_i^0.

To obtain the limiting distribution of $\hat{\lambda}_i$, we need an additional assumption.

Assumption H. For each i, as $T \to \infty$,

$$\frac{1}{T} \sum_{t=1}^{T} F_t^0 e_{it} \xrightarrow{\mathrm{d}} \int B_u \, \mathrm{d}B_e^{(i)},$$

where B_u is a $r \times 1$ vector of Brownian motions defined earlier; and $B_e^{(i)}$ is scalar Brownian motion process with variance $\Omega_{ee}^{(i)} = \lim (1/T) \sum_{t=1}^{T} \sum_{s=1}^{T} E(e_{it} e_{is})$. B_u and $B_e^{(i)}$ are independent.

Theorem 3. *Suppose that Assumptions* A–F *and* H *hold. Then for each* i, *as* $N,T\to\infty$, *we have*

$$T(\hat{\lambda}_i - H_1^{-1}\lambda_i^0) = H_1^{-1}\left(\frac{F^{0\prime}F^0}{T^2}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it} + o_p(1)$$

$$\xrightarrow{d} (\Sigma_\Lambda Q')^{-1}\left(\int B_u B_u'\right)^{-1}\int B_u\,\mathrm{d}B_e^{(i)},$$

where Σ_Λ *is given in Assumption* B, *and* Q *is given in Proposition* 3.

The limiting distribution is conditionally normal. Similar to Corollary 1, Theorem 3 implies the following limiting distribution for $\tilde{\Lambda}$:

Corollary 2. *Under the assumptions of Theorem* 3, *we have*

$$T(\tilde{\lambda}_i - H_2^{-1}\lambda_i^0)\xrightarrow{d}V(\Sigma_\Lambda Q')^{-1}\left(\int B_u B_u'\right)^{-1}\int B_u\,\mathrm{d}B_e^{(i)}$$

where V *and* Q *are defined in Proposition* 3.

4.3. Limiting distribution of estimated common components

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

Theorem 4. *Suppose that Assumptions* A–H *hold. As* $N,T\to\infty$, *we have*

(i) *If* $N/T\to 0$, *then for each* (i,t)

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

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

(ii) *If* $T/N\to 0$, *then for each* (i,t) *such that* $t=[T\tau]$,

$$\sqrt{T}(\hat{C}_{it} - C_{it}^0)\xrightarrow{d}B_u(\tau)'\left(\int B_u B_u'\right)^{-1}\int B_u\,\mathrm{d}B_e^{(i)}.$$

(iii) *If* $N/T\to\pi$, *then for each* (i,t) *with* $t=[T\tau]$,

$$\sqrt{N}(\hat{C}_{it} - C_{it}^0)\xrightarrow{d}N(0,V_{it}) + \sqrt{\pi}B_u(\tau)'\left(\int B_u B_u'\right)^{-1}\int B_u\,\mathrm{d}B_e^{(i)}.$$

In practice, (iii) is the most useful case because we can simply replace π by its estimate N/T. The two limiting random variables in (iii) are independent. The rate of convergence for the estimated common components is $\min\{\sqrt{N},\sqrt{T}\}$, which is the best possible rate. When F^0 is observable, the best rate for $\hat{\lambda}_i$ is T. When Λ^0 is observable, the best rate for $\hat{F}_t$ is $\sqrt{N}$. It follows that when both are estimated, the best rate for $\hat{\lambda}_i'\hat{F}_t$ does not exceed the minimum of $\sqrt{N}$ and T. Because the magnitude of $\hat{F}_t$ is large such that $\hat{F}_t = O_p\left(\sqrt{T}\right)$, the best rate for the product is $\min\left\{\sqrt{N}, \sqrt{T}\right\}$.

4.4. Confidence intervals

After obtaining the estimated factors, it is of interest to examine the relationship between some observable economic variables and the estimated statistical factors. The distribution theory derived in the previous section allows us to evaluate whether a given economic series is one of (or a linear combination) the underlying factors.

Suppose R_t $(t = 1, 2 \ldots, T)$ is a scalar observable series and we are to test if R_t is one of the underlying factors (or any linear combination of the underlying factors), i.e., $R_t = \delta' F_t^0$ for all t. Many observable economic variables are indices, and their levels and scales may be arbitrarily set. Therefore, it may be more reasonable to test whether $R_t = \alpha + \delta' F_t^0$ for all t, where α can capture the arbitrariness in levels and δ can capture the arbitrariness in scales. Consider a rotation of $\tilde{F}_t$ toward R_t via the regression

$$R_t = \alpha + \tilde{F}_t'\delta + error. \tag{13}$$

Let $(\hat{\alpha}, \hat{\delta})$ be the least-squares estimator, and define $\hat{R}_t = \hat{\alpha} + \tilde{F}_t'\hat{\delta}$. We have the following:

Proposition 4. *Under the assumptions of Theorem* 2 *and cross-section uncorrelation for the idiosyncratic errors, as* $N, T \to \infty$ *with* $N/T^3 \to 0$, *we have*

$$\frac{\sqrt{N}(\hat{R}_t - \alpha - \delta' F_t^0)}{\left[\hat{\delta}' V_{NT}^{-1} \left(\frac{1}{N}\sum_{i=1}^N \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i'\right) V_{NT}^{-1} \hat{\delta}\right]^{1/2}} \xrightarrow{d} N(0, 1),$$

where $\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i'\tilde{F}_t$ and V_{NT} is a diagonal matrix consisting of the first r largest eigenvalues of $XX'/(T^2N)$. From this, the 95% confidence interval for $R_t = \alpha + \delta' F_t^0$ $(t = 1, 2, \ldots, T)$ is

$$(\hat{R}_t - 1.96 S_t N^{-1/2}, \hat{R}_t + 1.96 S_t N^{-1/2}), \tag{14}$$

where S_t is the denominator expression given in Proposition 4. The test procedure does not require the knowledge of the remaining $(r-1)$ underlying factors. The confidence intervals can be easily plotted along with the true observable series, as illustrated in our simulation and in our empirical application. When the R_t series is a linear combination of the true factors F_t^0, we expect 95% of the T observations on R_t fall into the confidence band.

For the null hypothesis $R_t = \delta' F_t^0$ for all t, the constant regressor in (13) can be suppressed, and Proposition 4 is still valid.

Testing statistics concerning factor loading can be constructed using Theorem 3. A useful test is whether a given cross-section (i) is affected by the common trends. This is equivalent to testing if the factor loading (λ_i) is zero or any sub-vector of λ_i is zero. Because of the conditional normality, the chi-square distribution is applicable for testing $\lambda_i = 0$ and the normal distribution is applicable for testing $\lambda_{ik} = 0$ for component k.

5. Generalized dynamic factor models

Consider the following dynamic factor models:

$$X_{it} = \lambda_i(L)'F_t + e_{it}, \tag{15}$$

$$F_t = F_{t-1} + u_t,$$

where $\lambda_i(L)$ is a vector of polynomials of the lag operator. The relation between X_{it} and F_t is now dynamic, whereas in (1), the relation is static even though the factor F_t itself is dynamic. The above model was considered by Quah and Sargent (1993) and Forni and Reichlin (1998). A dynamic model is more flexible than a static one for empirical applications. For example, it allows shocks to affect different sectors or regions of an economy at different times and also allows for transmission effects. Following Forni et al. (2000a), we assume

$$\lambda_i(L) = \sum_{i=0}^{\infty} a_{ij} L^j$$

with $\sum_{j=0}^{\infty} j|a_{ij}| < \infty$. It is noted that $\lambda_i(L)'F_t$ is well defined because we assume $F_t = 0$ for $t < 0$, which is a standard assumption for integrated processes. Forni and Reichlin (1998) and Forni et al. (2000a) proposed some informal methods to estimate the number of common trends. Here, we show that the dimension of F_t can be consistently estimated for the dynamic models. It is important to note that the differenced data method considered in Section 3.1 does not work for dynamic models. To see this, consider the case in which F_t is a scalar such that $X_{it} = \lambda_{i1}F_t + \lambda_{i2}F_{t-1} + e_{it}$. From a dynamic point of view, there is only one factor. With differencing, we have $\Delta X_{it} = \lambda_{i1}u_t + \lambda_{i2}u_{t-1} + \Delta e_{it}$, which satisfies the conditions of Bai and Ng (2002) with two factors. Thus the differenced data method will identify two factors. With data in levels, however, we can consistently estimate the true number of factors.

Assumption I. Let $\lambda_i = \lambda_i(1) = \sum_{j=0}^{\infty} a_{ij}$, $E\|\lambda_i\|^4 \leqslant M < \infty$ and $(1/N)\sum_{i=1}^{N} \lambda_i\lambda_i' \xrightarrow{p} \Sigma_\Lambda$, a positive-definite matrix. In addition, let $\lambda_i^*(L) = \sum_{j=0}^{\infty} a_{ij}^* L^j$ with $a_{ij}^* = \sum_{k=j+1}^{\infty} a_{ik}$. We assume $E\|\lambda_i^*(L)u_t\|^4 \leqslant M < \infty$ for all i and t.

Theorem 5. *Assume Assumptions* A, C–D *and* I *hold. If* $g(N,T)$ *satisfies conditions* (i) *and* (ii) *of Theorem* 1, *then* $\lim_{N,T\to\infty} P(\hat{k} = r) = 1$.

Thus the dimension of the common factors can be consistently estimated for the generalized dynamic factor models (the proof is provided in Appendix C). Much more challenging is the limiting distribution of the estimated factors for the generalized dynamic factor models. If $\lambda_i(L)$ is a finite order polynomial in L, however, then limiting distributions are not difficult to obtain. Thus for deriving the limiting distribution, we restrict $\lambda_i(L)$ to be a finite order polynomial. To fix ideas, consider

$$X_{it} = \lambda_{i0}' F_t + \lambda_{i1}' F_{t-1} + \cdots + \lambda_{ip}' F_{t-p} + e_{it}. \tag{16}$$

This can be rewritten as

$$X_{it} = \gamma_{i0}' F_t - \gamma_{i1}' \Delta F_{t-1} - \cdots - \gamma_{ip}' \Delta F_{t-p} + e_{it}, \tag{17}$$

where $\gamma_{ik} = \lambda_{ik} + \lambda_{ik+1} + \cdots + \lambda_{ip}$. Denoting

$$\gamma_i' = (\gamma_{i0}', -\gamma_{i1}', \ldots, -\gamma_{ip}') \quad \text{and} \quad \underline{F}_t = (F_t', \Delta F_{t-1}', \ldots, \Delta F_{t-p}')'$$

Eq. (17) can be rewritten as

$$X_{it} = \gamma_i' \underline{F}_t + e_{it} \tag{18}$$

$$= \gamma_{i0}' F_t + \gamma_{i0-}' G_t + e_{it}, \tag{19}$$

where $G_t = (\Delta F_{t-1}', \ldots, \Delta F_{t-p}')'$, and γ_{i0-} is a sub-vector of γ_i other than γ_{i0}. This reparametrization implies that F_t is a vector of I(1) factors and G_t is a vector of I(0) factors.

For model (16), or equivalently (19), to have r of I(1) factors, the limiting matrix of $(1/N)\sum_{i=1}^N \gamma_{i0}\gamma_{i0}'$ must be an $r \times r$ positive-definite matrix. This is, Assumption I holds. The total number of I(1) and I(0) factors is equal to the rank of the limiting matrix of $(1/N)\sum_{i=1}^N \gamma_i\gamma_i'$, where $\gamma_i = (\gamma_{i0}', \gamma_{i0-}')'$. Because the vector γ_i is an invertible transformation of λ_i, where $\lambda_i' = (\lambda_{i0}', \lambda_{i1}', \ldots, \lambda_{ip}')$ (the coefficients of model (16)), the rank of the matrix $(1/N)\sum_{i=1}^N \gamma_i\gamma_i'$ is equal to the rank of $(1/N)\sum_{i=1}^N \lambda_i\lambda_i'$. In (16), some lag of F_t, say F_{t-k} $(k < p)$ may not enter into the equation, i.e., the vector λ_{ik} is zero, or some components of λ_{ik} can be zero. Consider, for example, $X_{it} = \lambda_{i0}F_t + \lambda_{i2}F_{t-2} + e_{it}$, where F_t is a scalar and $\lambda_{i1} = 0$. This model can be rewritten as $X_{it} = (\lambda_{i0} + \lambda_{i2})F_t - \lambda_{i2}(\Delta F_{t-1} + \Delta F_{t-2}) + e_{it}$. In this case, we shall define G_t to be $(\Delta F_{t-1} + \Delta F_{t-2})$. By combining the columns of $(\Delta F_{t-1}, \ldots, \Delta F_{t-p})$ when necessary, we shall assume that the loading matrix of G_t has a full column rank. More specifically, we assume the loading coefficients of model (19) satisfy

Assumption J. *$E\|\gamma_i\|^4 \leqslant M$ and $p\lim_{N\to\infty}(1/N)\sum_{i=1}^N \gamma_i\gamma_i' = \Sigma_\Gamma$*, where Σ_Γ is an $(r+q)\times(r+q)$ positive-definite matrix.

Because there are r of I(1) factors, the above assumption implies the existence of q of I(0) factors. According to Theorem 5, the r I(1) factors can be consistently estimated by the data in levels. In addition, the differenced data approach leads to consistent estimation of the total number of factors $(r+q)$. Thus, q can also be consistently estimated. We state this result in the following corollary.

Corollary 3. *Under Assumptions of* A–D, *and* I–J, *the number of* I(1) *and* I(0) *factors can be consistently estimated.*

General I(0) factors. It is noted that G_t does not have to be the lags of ΔF_t, it can be a vector of general I(0) variables. This is because our proof does not depend on G_t to be the lags of ΔF_t, but only on G_t being I(0). For general G_t, we naturally assume $E\|G_t\|^4 \leqslant M$ for all t, and

$$\begin{pmatrix} T^{-2}\sum_{t=1}^{T} F_tF_t' & T^{-3/2}\sum_{t=1}^{T} F_tG_t' \\ T^{-3/2}\sum_{t=1}^{T} G_t'F_t & T^{-1}\sum_{t=1}^{T} G_tG_t' \end{pmatrix} \xrightarrow{d} \Sigma,$$

where Σ is a $(r+q)\times(r+q)$ positive-definite (random) matrix with probability 1. When G_t is the lags of ΔF_t, this assumption holds, given the assumptions on u_t. It is pointed out that G_t does not have to be mean zero. When a component of G_t is the constant 1 (for all t), the true model has an intercept. The theoretical results of this section holds for general I(0) G_t.

Cointegrated non-stationary factors. The presence of I(0) factors can accommodate cointegration. More specifically, let F_t be a vector of cointegrated I(1) factors. Then there exists an invertible matrix P such that $PF_t=(\xi_t',\eta_t')'$, where ξ_t is a vector of non-cointegrated I(1) factors, and η_t is a vector of I(0) factors that are linear combinations of F_t. With some abuse of notation, denote ξ_t by F_t, η_t by G_t, and γ_iP^{-1} by γ_i so that cointegrated I(1) factors can be expressed as (19). Corollary 3 implies that the dimension of non-cointegrated I(1) factors and the dimension of I(0) factors can be identified.

We next consider the limiting distributions of the estimated factors and of the factor loadings. In deriving the limiting distributions below, both r and q are assumed to be known. Let $\tilde F$ be the r eigenvectors of XX' corresponding to the first r largest eigenvalues normalized such that $\tilde F'\tilde F/T^2=I$ and let $\tilde G$ be the q eigenvectors corresponding to the next q largest eigenvalues, normalized such that $\tilde G'\tilde G/T=I$. Denote

$$\underline{\tilde F}=(\tilde F,\tilde G).$$

Let V_{NT}^r be the diagonal matrix of the first r eigenvalues of the matrix $XX'/(T^2N)$ and V_{NT}^q be the diagonal matrix of the $(r+1)$th to $(r+q)$th largest eigenvalues of the matrix $XX'/(TN)$. Denote $\underline{V}_{NT}=diag(V_{NT}^r,V_{NT}^q)$. We use superscript 0 to represent the true quantities so that $\underline{F}^0=(\underline{F}_1^0,\underline{F}_2^0,\ldots,\underline{F}_T^0)'$ is the $T\times(r+q)$ true factor matrix and $\Gamma^0=(\gamma_1^0,\gamma_2^0,\ldots,\gamma_N^0)'$ is the $N\times(r+q)$ true factor loading matrix. We estimate $\underline{F}^0$ by $\underline{\tilde F}$ and estimate Γ^0 by

$$\tilde\Gamma=X'\underline{\tilde F}\Upsilon_T^{-2},$$

where $\Upsilon_T=diag\left(TI_r,\sqrt{T}I_q\right)$. We have the following results:

Theorem 6. *Under Assumptions* A, C–J, *as* $N,T\to\infty$,

(i) $\Upsilon_T^{-2}\underline{\tilde F}'\underline{F}^0\xrightarrow{d}\underline{Q}$ *for some invertible but random matrix* $\underline{Q}$.

(ii) $\underline{V}_{NT} \xrightarrow{d} \underline{V}$, *where* $\underline{V}$ *is a full-rank diagonal random matrix.*

(iii) *There exists an invertible matrix* $\underline{H}$, *such that if* $N, T \to \infty$ *with* $N/T^2 \to 0$ *and if* $\left(1/\sqrt{N}\right)\sum_{i=1}^{N} \gamma_i^0 e_{it} \xrightarrow{d} N(0, \Omega_t)$, *then*

$$\sqrt{N}(\tilde{\underline{F}}_t - \underline{H}'\underline{F}_t^0) = \underline{V}_{NT}^{-1}(\Upsilon_T^{-2}\tilde{\underline{F}}'\underline{F}^0)\left(\frac{1}{\sqrt{N}}\right)\sum_{i=1}^{N}\gamma_i^0 e_{it} + o_p(1)$$

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

where $\underline{V}$ *and* $\underline{Q}$ *are independent of* $N(0, \Omega_t)$.

(iv) *If* $\Upsilon_T^{-1}\sum_{t=1}^{T} \underline{F}_t^0 e_{it} \xrightarrow{d} W_F$ *(a conditionally normal vector), then*

$$\Upsilon_T(\tilde{\gamma}_i - \underline{H}^{-1}\gamma_i^0) = \underline{V}_{NT}\underline{H}^{-1}(\Upsilon_T^{-1}\underline{F}^{0\prime}\underline{F}^0\Upsilon_T^{-1})^{-1}$$

$$\times \Upsilon_T^{-1}\sum_{t=1}^{T}\underline{F}_t^0 e_{it} + o_p(1) \xrightarrow{d} AW_F,$$

where $A = \underline{V}(\Sigma_\Gamma \underline{Q}')^{-1}\Sigma_F^{-1}$ *and* Σ_F *is a block diagonal matrix with the first block being* $\int B_u B_u'$ *and the second block being* $p\lim (1/T)\sum_{t=1}^{T} G_t G_t'$.

Both the estimated factors and factor loadings are conditionally normal. This implies an easy construction of confidence intervals. Suppose we are interested in testing the hypothesis that an observable sequence R_t is one of (or a linear combination) the underlying factors. The test can be constructed similar to the method given in Section 4.4. Consider rotating the estimated factors toward R_t by running the regression:

$$R_t = \alpha + \delta'\tilde{\underline{F}}_t + error \quad (t = 1, 2, \ldots, T).$$

Let $(\hat{\alpha}, \hat{\delta})$ be the least-squares estimator and define $\hat{R}_t = \hat{\alpha} + \hat{\delta}'\tilde{\underline{F}}_t$.

Corollary 4. *Under the assumptions of Theorem* 6 *and cross-section independence of the idiosyncratic errors*, *as* $N, T \to \infty$ *with* $N/T^2 \to 0$, *we have*

$$\frac{\sqrt{N}(\hat{R}_t - \alpha - \delta'\underline{F}_t^0)}{\left[\hat{\delta}'\underline{V}_{NT}^{-1}\left((1/N)\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\gamma}_i\tilde{\gamma}_i'\right)\underline{V}_{NT}^{-1}\hat{\delta}\right]^{1/2}} \xrightarrow{d} N(0, 1).$$

From this, the 95% confidence interval for R_t $(t = 1, 2, \ldots, T)$ is

$$(\hat{R}_t - 1.96 S_t N^{-1/2}, \hat{R}_t + 1.96 S_t N^{-1/2}), \tag{20}$$

where S_t is the denominator expression of this corollary. If we test $R_t = \delta'\underline{F}_t^0$ (i.e., $\alpha = 0$), the constant in the regression should be suppressed and the corollary continues to hold.

6. Simulation results

The dimension of common trends. We first consider standard dynamic factor models (no lags of F_t entering into X_{it}). Data are generated according to

$$X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{jt} + e_{it}, \tag{21}$$

$$F_{jt} = F_{jt-1} + u_{jt}, \tag{22}$$

$$e_{it} = \rho e_{it-1} + v_{it} + \theta v_{it-1}, \tag{23}$$

where λ_{ij}, u_{jt}, and v_{it} are i.i.d. $N(0,1)$ for all (i,j,t) and are independent of each other. The parameter values are $r=2$, $\rho=0.5$, and $\theta=0.5$. Thirteen combinations of N and T of various sizes are considered. In all cases, $kmax=8$. Both the differenced data and level data methods are used and evaluated. Table 1 reports the average $\hat{k}$ over 1000 simulations. The differenced and level methods are both estimating $r=2$. All criteria perform reasonably well, except the first two criteria for the differenced data with $T=40$ and with $T=N=50$.

We next consider generalized dynamic factor models. Eq. (21) is replaced by

$$X_{it} = \sum_{j=1}^{r} \sum_{k=0}^{p} \lambda_{ijk} F_{jt-k} + e_{it}, \tag{24}$$

where the λ_{ijk} are i.i.d. $N(0,1)$. Eqs. (22) and (23) follow the same specification as in the previous case. Only the configuration of $r=2$ and $p=1$ is reported to conserve space. Given $r=2$, (24) can be rewritten as

$$X_{it} = \lambda_{i10} F_{1t} + \lambda_{i11} F_{1t-1} + \lambda_{i20} F_{2t} + \lambda_{i21} F_{2t-1} + e_{it}. \tag{25}$$

According to Theorem 5, with data in levels, two factors should be identified. By Proposition 2, with data in differences, four factors should be identified. The results are reported in Table 2, with each entry representing the average of $\hat{k}$ over 1000 repetitions. The simulation results are consistent with the theory that the number of factors in a generalized dynamic factor model can be identified.

Estimating common trends. We shall use graphs to present the factor estimates and the true factors. Both the standard and generalized dynamic factor models are considered. For the former, data are generated according to Eqs. (21)–(23) with $r=2$, $\rho=0.5$, $\theta=0.5$. The true factors are denoted by F^0 ($T\times 2$). We fix T at $T=30$. A larger T yields more precise factor estimates but produces a crowded graphic display. We examine the behavior of the factor estimates as N varies from $N=25$ to 50 and then to 100. For each (T,N), we simulate a sample of observations, denoted by X, a $T\times N$ matrix. We use the estimator $\tilde{F}$ ($T\times 2$), which is equal to the eigenvectors of the first two largest eigenvalues of XX' multiplied by T (see Section 2.2). To see that $\tilde{F}$ is estimating a transformation of F^0, we rotate $\tilde{F}$ toward each of the true factor process via the following regression

$$F^0_{kt} = \delta_k' \tilde{F}_t + error \tag{26}$$

Table 1
Estimated number of factors ($\hat{k}$) averaged over 1000 repetitions

N	T	Differenced data			Level data		
		PC_1	PC_2	PC_3	IPC_1	IPC_2	IPC_3
100	40	3.73	2.77	2.00	2.00	2.00	1.92
100	60	2.13	2.00	2.00	2.00	2.00	1.92
200	60	2.00	2.00	2.00	2.00	2.00	1.92
500	60	2.00	2.00	2.00	2.00	2.00	1.93
1000	60	2.00	2.00	2.00	2.00	2.00	1.92
40	100	2.33	2.04	2.00	1.99	1.98	1.84
60	100	2.00	2.00	2.00	1.99	1.99	1.88
60	200	2.00	2.00	2.00	2.00	1.99	1.86
60	500	2.00	2.00	2.00	2.00	2.00	1.87
60	1000	2.00	2.00	2.00	2.00	2.00	1.88
50	50	4.26	2.59	2.00	2.00	1.99	1.91
100	100	2.00	2.00	2.00	2.00	2.00	1.92
200	200	2.00	2.00	2.00	2.00	2.00	1.98

$$\text{DGP}: X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{jt} + e_{it}$$

$$F_{jt} = F_{jt-1} + u_{jt}$$

$$e_{it} = \rho e_{it-1} + v_{it} + \theta v_{it-1}$$

$$r = 2, \;\; \rho = 0.5, \;\; \theta = 0.5$$

Note: The true number of I(1) factors is $r = 2$. When the average of $\hat{k}$ is an integer, the corresponding standard error is exactly zero. In the few cases when the averaged $\hat{k}$ over replications is not an integer, the standard errors are small. In view of the precision of the estimates in the majority of cases, the standard errors in the simulations are not reported.

for $k = 1, 2$. Let $\hat{\delta}_k$ be the least-squares estimate of δ_k. Then $\hat{\delta}_k' \tilde{F}_t$ is the predicted value of F_{kt}^0 using the predictor $\tilde{F}_t$. The precision of the factor estimates increases as N becomes larger. For example, the sample correlation coefficient between the two sequences $\{F_{1t}^0\}$ and $\{\hat{\delta}_1' \tilde{F}_t\}$ is 0.9964 for $N = 25$, 0.9973 for $N = 50$, and 0.9998 for $N = 100$. The sample correlation coefficient between $\{F_{2t}^0\}$ and $\{\hat{\delta}_2' \tilde{F}_t\}$ is 0.9934, 0.9945, and 0.9987, respectively, for $N = 25, 50$ and 100. A plot of $\hat{\delta}_k' \tilde{F}_t$ along with F_{kt}^0 would show that they track each other extremely well, but instead, we have plotted the confidence intervals. Confidence intervals for F_{kt}^0 are computed according to Proposition 4 and Eq. (14). These intervals together with the true factor process are plotted in Fig. 1. The left panels are for the first factor and the right panels for the second factor. The true factor processes are indeed located inside the confidence intervals with the exception of a small number of data points.

Next, we examine the corresponding results for generalized dynamic factor models. The data are generated according to Eqs. (24), (22), and (23) with $r=2$, $p=1$, $\rho=0.5$,

Table 2
Estimated number of factors ($\bar{k}$) averaged over 1000 repititions (Generalized dynamic factor model)

		Differenced data			Level data		
N	T	PC_1	PC_2	PC_3	IPC_1	IPC_2	IPC_3
100	40	4.70	4.17	4.00	2.06	2.02	1.97
100	60	4.01	4.00	4.00	2.00	2.00	1.98
200	60	4.00	4.00	4.00	2.00	2.00	1.98
500	60	4.00	4.00	4.00	2.00	2.00	1.98
1000	60	4.00	4.00	4.00	2.00	2.00	1.98
40	100	4.04	4.00	4.00	2.00	2.00	1.96
60	100	4.00	4.00	4.00	2.00	2.00	1.98
60	200	4.00	4.00	4.00	2.00	2.00	1.98
60	500	4.00	4.00	4.00	2.00	2.00	1.99
60	1000	4.00	4.00	4.00	2.00	2.00	1.98
50	50	5.08	4.08	4.00	2.00	2.00	1.97
100	100	4.00	4.00	4.00	2.00	2.00	1.99
200	200	4.00	4.00	4.00	2.00	2.00	2.00

$$\text{DGP}: X_{it} = \sum_{j=1}^{r}\sum_{k=0}^{p} \lambda_{ijk} F_{jt-k} + e_{it}$$

$$F_{jt} = F_{jt-1} + u_{jt}$$

$$e_{it} = \rho e_{it-1} + v_{it} + \theta v_{it-1}$$

$$r = 2,\ \ p = 1,\ \ \rho = 0.5,\ \ \theta = 0.5$$

Note: The level-data method gives an estimate of r (true value $r = 2$) and the differenced-data method gives an estimate of $r(p+1)$ (true value is 4).

and $\theta = 0.5$. Again T is fixed at 30 and N takes on the values 25, 50, and 100. The same F^0 generated earlier is used here for comparison purpose. But now the lag F^0_{t-1} also enters into the X_{it} equation. In this case, four factors need to be estimated, with two being I(1) and two being I(0). Let $\underline{\tilde{F}}$ be the $T \times 4$ factor estimate described in Section 5. For $k = 1, 2$, we consider the rotation

$$F^0_{kt} = \delta'_k \underline{\tilde{F}}_t + error.$$

Now the sample correlation coefficient between $\{F^0_{1t}\}$ and $\{\hat{\delta}'_1 \underline{\tilde{F}}_t\}$ is 0.9966, 0.9934, and 0.9995, respectively, for $N = 25, 50$ and 100. The correlation coefficient between $\{F^0_{2t}\}$ and $\{\hat{\delta}'_2 \underline{\tilde{F}}_t\}$ is 0.9956, 0.9967, and 0.9989, respectively, as N takes on the three values. Again, $\hat{\delta}'_k \underline{\tilde{F}}_t$ tracks F^0_{kt} extremely well.

Confidence intervals for F^0_{kt} are constructed according to Corollary 4 and Eq. (20). These intervals along with the true factor process $\{F^0_{kt}\}$ are plotted in Fig. 2. The left panels are for factor $\{F^0_{1t}\}$ and the right ones for factor $\{F^0_{2t}\}$. With the exception of a small number of data points, the true factor process lies within the confidence intervals. The distribution theory appears to be adequate.

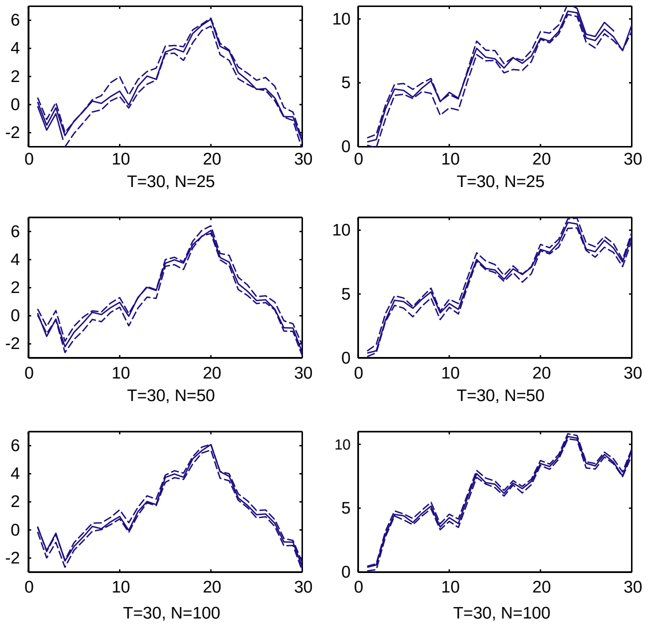

Fig. 1. Confidence intervals for the true factor process. Data are generated according to the model specified in Table 1. The left three panels are the confidence intervals (dashed line) for the first true factor and the intervals are estimated from $N = 25$, 50 and 100 cross-sections, respectively. The true factor process F^0_{1t} (solid line) is also plotted. The right three panels are the confidence intervals for the second factor process along with the true factor process F^0_{2t}.

7. Application: sectoral employment

In this section we study fluctuations in employment across 60 industries for the U.S. We examine the hypothesis that these fluctuations can be explained by a small number of aggregate factors. Quah and Sargent (1993) also studied sectoral employment fluctuations. However, they did not estimate the number of factors. Neither did they conduct hypothesis testings in the absence of an inferential theory. This point was emphasized by Geweke (1993) in his comment on Quah and Sargent's work. With the theory developed in this paper, we are able to address a number of issues raised in Geweke's comments. In addition, interesting insight is gained by examining the relationship between the estimated factors and observable macroeconomic factors. Our empirical analysis also illustrates the techniques proposed in this paper.

The Bureau of Economic Analysis (BEA) reports the number of full-time equivalent (FTE) workers across various industries (NIPA, Tables 6.5b and 6.5c). There are a total of sixty private sector industries. A list of them is provided in Appendix D. The data are annual frequency, ranging from 1948 to 2000. Our analysis is based on the log-valued data. For graphical display, the series are ordered cross-sectionally to

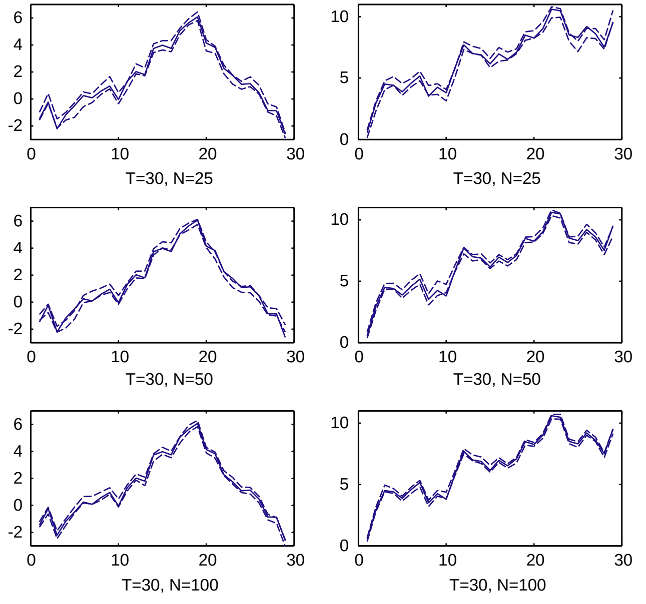

Fig. 2. Confidence intervals for the true factor process. Data are generated according to the model specified in Table 2 (generalized dynamic factor models). The left three panels are the confidence intervals (dashed lines) for the first factor and the intervals are estimated from $N = 25, 50$ and 100 cross-sections, respectively. The true factor process F^0_{1t} is also plotted (solid line). The right three panels are the confidence intervals for the second factor process along with the true factor process F^0_{2t}.

have a better view of the data. Two plots are given according to different methods of ordering. In Fig. 3, we order the cross-sections according to their 1948 values in ascending order. In Fig. 4, the cross-sections are ordered according to their 2000 values also in ascending order. The vertical axis represents the log-valued employment in each sector. The statistical analysis below does not depend on the ordering of cross-sections, and any permutation will give the same results.

The number of factors. For data in levels, we estimate the number of factors using the three criteria in Eq. (12). With $kmax = 6$, the first two criteria suggest four factors and the last criterion gives two factors. If we set $kmax = 4$, all three criteria yield two factors. If we choose *kmax* to 2, all criteria still give two factors. These results provide evidence in support of two nonstationary common factors.

For data in differences, we start with $kmax = 6$, and then set *kmax* at the estimated value in the first around as in the previous paragraph. All criteria indicate three factors. This suggests an additional I(0) factor in the system. The estimated residuals resulting from a three-factor model is plotted in Fig. 5. No discernable pattern is found in the residuals, indicating a reasonable fit.

Macroeconomic factors. Quah and Sargent (1993) examined an observable factor model with total employment and total output (with leads and lags) as potential factors. They found that future GNP has better explaining power than past GNP. In terms of

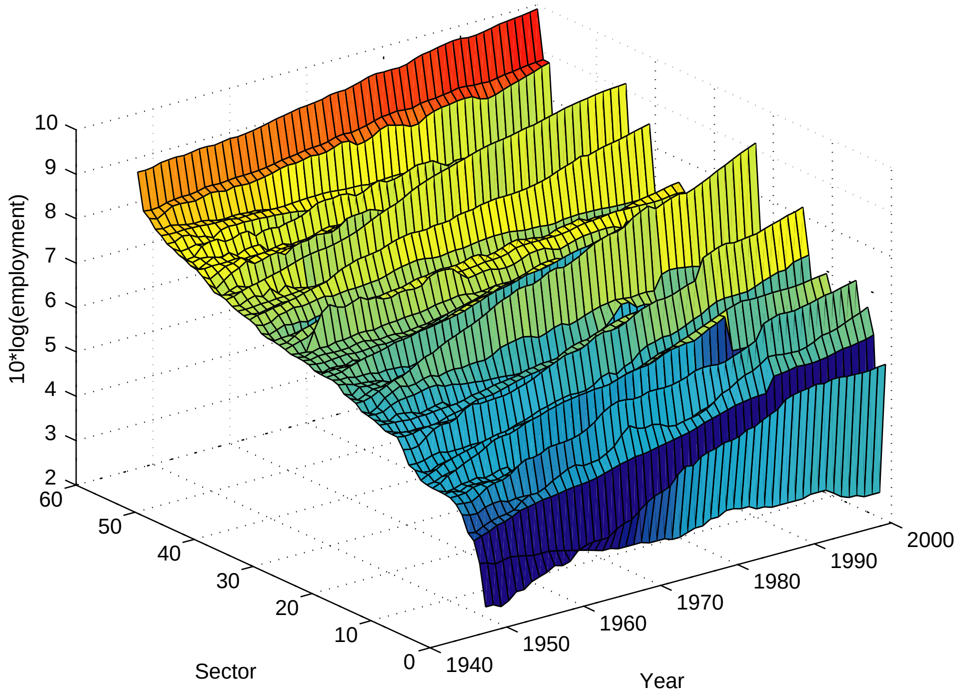

Fig. 3. The number of full-time equivalent employees across 60 sectors. The sectors are arranged in ascending order according to their 1948 values.

the number of factors, our result is consistent with the following empirical model for sectoral employment:

$$S_{it} = \lambda_{i1} E_t + \lambda_{i2} Y_t + \lambda_{i3} Y_{t+1} + e_{it}, \tag{27}$$

where S_{it} is the log-valued employment (multiplied by 10) of sector i; E_t is total employment, defined as the log-valued total employment of the 60 industries; and Y_t is the log-valued real GNP at year t. Eq. (27) can be rewritten as $S_{it} = \lambda_{i1} E_t + (\lambda_{i2} + \lambda_{i3}) Y_t + \lambda_{i3} \Delta Y_{t+1} + e_{it}$, where $\Delta Y_t = Y_t - Y_{t-1}$ is the growth rate of real GNP. This model postulates that disaggregated employment is explained by total employment, total output, and changes in total output. If this specification is correct, we should identify two factors with data in levels, and three factors with data in differences.

Our next inquiry is whether the unobservable factor model is consistent with the empirical model of (27). That is, whether or not total employment and total output are the underlying factors. If correctly specified, model (27) leads to a simpler representation, more efficient estimation, and a direct economic interpretation. But if total employment and GNP are not the underlying factors, model (27) would be misspecified. The advantage of the unobservable factor model is that we can consistently estimate the

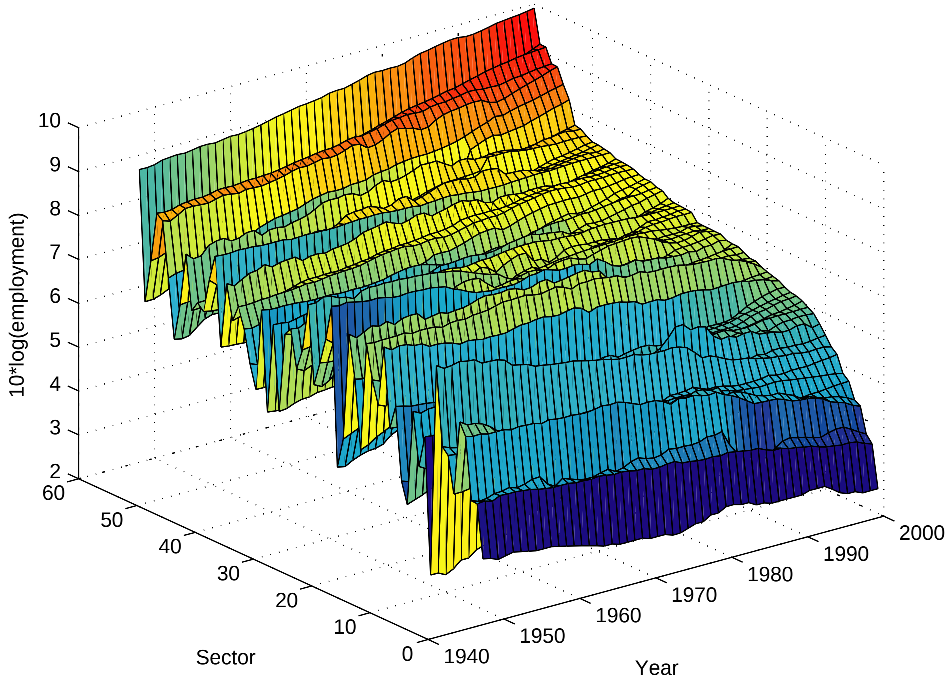

Fig. 4. The number of full-time equivalent employees across 60 sectors. The sectors are arranged in ascending order according to their 2000 values.

underlying factors. Once these factors are estimated, the inferential theory allows us to test if the observable variables are the factors.

To test whether total employment is a true underlying factor, we rotate the three statistical factors toward E_t by running the regression $E_t = \delta' \underline{\tilde{F}}_t + error$, as explained in Section 5 (including an intercept in the regression does not alter the conclusion). We then compute and plot the confidence intervals for the true underlying factor. Also plotted is the observable total employment. It is seen that total employment lies inside the confidence intervals throughout the whole period 1948–2000, see Fig. 6. This suggests that we can accept the hypothesis that total employment is one of the underlying factors.

To test whether GNP is one of the true factors, we rotate the three statistical factors toward Y_t by running the regression $Y_t = \delta' \underline{\tilde{F}}_t + error$. Since there are many periods for which GNP stays outside the confidence intervals, the evidence in supporting GNP as a factor is dubious, see Fig. 7. It remains an open question as to which economic variable constitutes the second nonstationary factor.

Overall, there is strong evidence that employment fluctuations across industries can be well explained by two nonstationary dynamic factors. The evidence is also consistent with the hypothesis that total employment is one of the underlying factors. But the evidence for GNP as a factor is not strong.

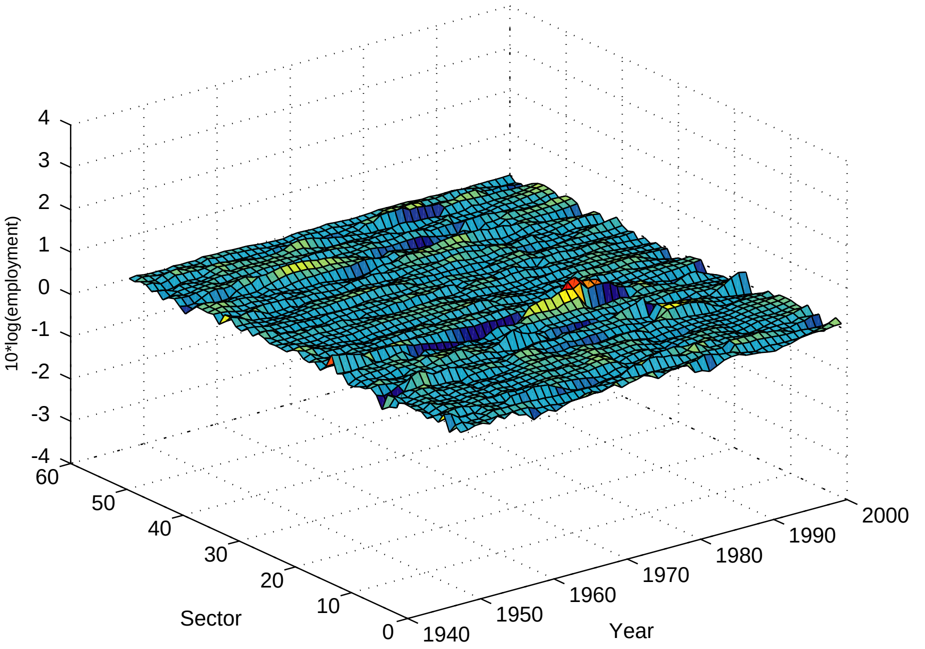

Fig. 5. The number of full-time equivalent employees across 60 sectors. Estimated residuals from a three-factor model.

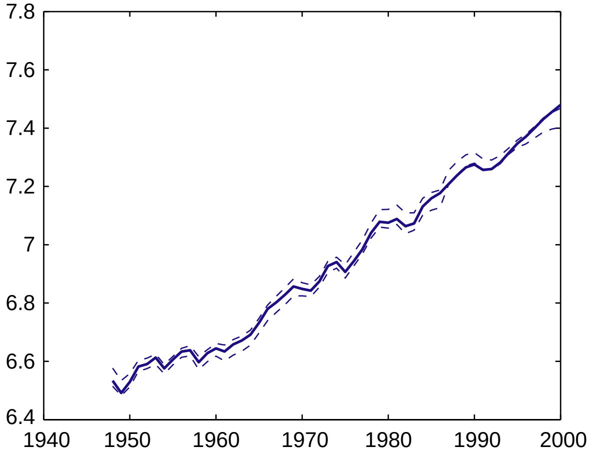

Fig. 6. Confidence intervals for testing total employment as a factor. Confidence intervals—dashed line, log-valued total employment—solid line.

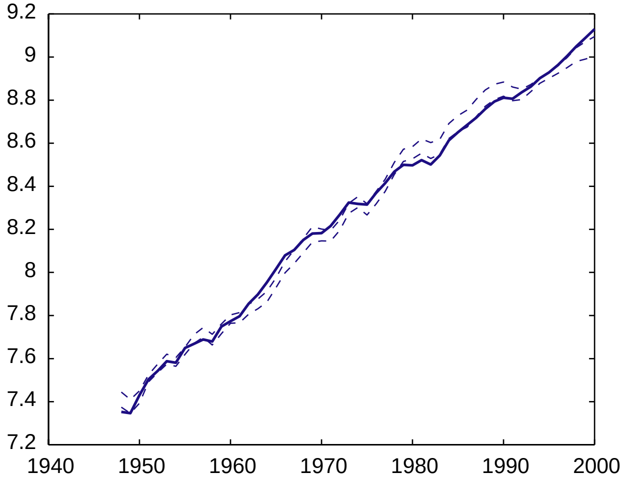

Fig. 7. Confidence intervals for testing GNP as a factor. Confidence intervals—dashed line, log-valued GNP—solid line.

8. Conclusion

This paper studies the estimation and inference of cross-section common stochastic trends in nonstationary panel data. We show that the dimension and the common trends themselves can be consistently estimated. When estimating the dimension, no restriction is imposed on the relationship between N and T. We also derive the rates of convergence and the limiting distributions for the estimated common trends and the estimated loading coefficients. Corresponding results for a class of generalized dynamic factor models are also obtained. The method is applied to the study of employment fluctuations across 60 U.S. industries. It is found that two nonstationary dynamic factors can explain much of the fluctuations in the sectoral employment. The statistical evidence is consistent with the hypothesis that aggregate employment is one of the underlying factors.

High-dimension data analysis that takes into account cross-section correlation and co-movement should become an important framework of econometric analysis, in view of the increasing availability of high-dimensional data sets and the increasing interconnectedness of the world economy. This paper may be considered a step in this direction.

Acknowledgements

I thank Gary Chamberlain, Jinyong Hahn, Peter Hansen, Jerry Hausman, Arthur Lewbel, and seminar participants at Brown University, Ohio State University, Harvard/MIT, NYU, University of Maryland, and Triangle Econometrics Workshop for many useful comments. Serena Ng and Richard Tresch read the manuscript carefully and provided many valuable comments. All remaining errors are my own responsibility.

Financial support from the National Science Foundation under grants SBR-9896329 and SES-0137084 is gratefully acknowledged.

Appendix A. The number of common trends

The proofs of Lemma 1, Proposition 1, and Lemmas A.1 and A.2 below are similar to those of Bai and Ng (2002) and thus are omitted. Detailed proofs are available from the author.

Lemma A.1. *Under Assumptions* A–C, *we have for some* $M<\infty$, *and for all* N *and* T,

$$\text{(i)}\quad E\left(T^{-1}\sum_{t=1}^{T}\|N^{-1/2}e_t'\Lambda^0\|^2\right)=E\left(T^{-1}\sum_{t=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}\lambda_i^0\|^2\right)\leqslant M,$$

$$\text{(ii)}\quad E\left(T^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(N^{-1}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right)\leqslant M,$$

$$\text{(iii)}\quad E\left\|(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\right\|\leqslant M.$$

We use the identity $\hat F^k=N^{-1}X\tilde\Lambda^k$ and $\tilde\Lambda^k=T^{-2}X'\tilde F^k$. That is, $\hat F^k=(1/T^2N)XX'\tilde F^k$. From the normalization $\tilde F^{k\prime}\tilde F^k/T^2=I_k$, we also have $T^{-2}\sum_{t=1}^{T}\|\tilde F_t^k\|^2=O_p(1)$. For $H^k=(\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}\tilde F^k/T^2)$, upon expanding XX', we have

$$\hat F_t^k-H^{k\prime}F_t^0=T^{-2}\sum_{s=1}^{T}\tilde F_s^k\gamma_N(s,t)+T^{-2}\sum_{s=1}^{T}\tilde F_s^k\zeta_{st}+T^{-2}\sum_{s=1}^{T}\tilde F_s^k\eta_{st}$$
$$+T^{-2}\sum_{s=1}^{T}\tilde F_s^k\xi_{st},\qquad\text{(A.1)}$$

where

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

$$\eta_{st}=F_s^{0\prime}\Lambda^{0\prime}e_t/N,$$

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

Lemma A.2. *For each k,* $1 \leqslant k \leqslant r$, *and let* H^k *be the matrix defined earlier, then*

$$V(k,\hat{F}^k) - V(k,F^0H^k) = \mathrm{O_p}(T^{1/2}\delta_{NT}^{-1}),$$

where $V(k,\hat{F}^k)$ *is defined in* (9) *and* $V(k,F^0H^k)$ *is defined by replacing* $\hat{F}^k$ *with* F^0H^k.

Lemma A.3. *For the matrix* H^k *defined in Lemma* 1, *and for each k with* $k < r$, *we have*

$$\liminf_{N,T\to\infty} \frac{\log\log T}{T}(V(k,F^0H^k) - V(r,F^0)) = \tau > 0.$$

Proof. For a matrix A, let $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$. Then

$$\begin{aligned} V(k,F^0H^k) - V(r,F^0) &= N^{-1}T^{-1}\sum_{i}^{N} \underline{X}_i'(P_{F^0} - P_{F^0H^k})\underline{X}_i \\ &= N^{-1}T^{-1}\sum_{i=1}^{N} \lambda_i^{0\prime}F^{0\prime}(P_{F^0} - P_{F^0H^k})F^0\lambda_i^0 \\ &\quad + 2N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i'(P_{F^0} - P_{F^0H^k})F^0\lambda_i^0 \\ &\quad + N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i'(P_{F^0} - P_{F^0H^k})\underline{e}_i = \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

First, note that $P_{F^0} - P_{F^0H^k} \geqslant 0$. Hence, $\mathrm{III} \geqslant 0$. Consider term I:

$$\begin{aligned} \mathrm{I} &= tr\left[T^{-1}F^{0\prime}(P_{F^0} - P_{F^0H^k})F^0N^{-1}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}\right] \\ &= T\,tr\left(\left[\frac{F^{0\prime}F^0}{T^2} - \frac{F^{0\prime}F^0H^k}{T^2}\left(\frac{H^{k\prime}F^{0\prime}F^0H^k}{T^2}\right)^{-1}\frac{H^{k\prime}F^{0\prime}F^0}{T^2}\right]N^{-1}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}\right). \end{aligned}$$

Thus,

$$\frac{\log\log T}{T}I = tr\left(\left[B_T - B_TH^k(H^{k\prime}B_TH^k)^{-1}H^{k\prime}B_T\right]N^{-1}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}\right),$$

where $B_T = (\log\log T)(F^{0\prime}F^0/T^2)$. The matrix in the square brackets is positive semidefinite. Furthermore, because $rank(H^k) = k < r$, the matrix is not identically zero in view of $rank(F^{0\prime}F^0/T^2) = r$. In addition, $N^{-1}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime}$ is positive definite by Assumption B. These imply that the trace is positive and the trace does not converge to zero because $\liminf B_T$ is a positive-definite matrix by Assumption A(iii). This implies that $\liminf_{N,T\to\infty}(\log\log T)\mathrm{I}/T > 0$.

Now $\mathrm{II} = 2N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_{F^0} F^0 \lambda_i^0 - 2N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_{F^0H^k} F^0 \lambda_i^0$. But

$$\left| N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_{F^0} F^0 \lambda_i^0 \right|$$

$$= \left| N^{-1}T^{-1}\sum_{i=1}^N \sum_{t=1}^T e_{it} F_t^{0\prime} \lambda_i^0 \right|$$

$$\leqslant T^{1/2}\left(T^{-2}\sum_{t=1}^T \|F_t^0\|^2\right)^{1/2} \cdot \frac{1}{\sqrt{N}} \left(T^{-1}\sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it}\lambda_i^0 \right\|^2\right)^{1/2}$$

$$= \mathrm{O_p}((T/N)^{1/2}).$$

The last equality follows from Lemma A.1(i). The second term of II is also $\mathrm{O_p}((T/N)^{1/2})$, and hence $(\log\log T)\mathrm{II}/T = \mathrm{O_p}(\log\log T/(TN)^{1/2}) \xrightarrow{\mathrm{p}} 0$. This proves Lemma A.3. □

Lemma A.4. *For each fixed k with $k \geqslant r$, $V(k,\hat{F}^k) - V(r,\hat{F}^r) = \mathrm{O_p}(1)$.*

Proof. This follows easily from $V(k,\hat{F}^k) = \mathrm{O_p}(1)$ for each $k \geqslant r$ because

$$V(k,\hat{F}^k) \leqslant V(r,\hat{F}^r) \leqslant V(r,F^0) \leqslant \frac{1}{TN}\sum_{i=1}^N \sum_{t=1}^T e_{it}^2 = \mathrm{O_p}(1).$$

The first inequality follows from $k \geqslant r$. The second is a defining property of minimization (the sum of squared residuals is no larger than the case of known F^0 and thus not minimized with respect to F^0). The third inequality says that the sum of squared residuals $V(r,F^0)$ is no larger than the sum of squared residuals evaluated at Λ^0 and F^0 (without minimizing with respect to Λ^0 and F^0). The last equality follows from the assumptions on e_{it}. □

Proof of Theorem 1. Consider $k < r$. We have the identity:

$$V(k,\hat{F}^k) - V(r,\hat{F}^r) = [V(k,\hat{F}^k) - V(k,F^0H^k)] + [V(k,F^0H^k) - V(r,F^0H^r)] + [V(r,F^0H^r) - V(r,\hat{F}^r)].$$

Lemma A.2 implies that the first and the third terms are both $\mathrm{O_p}(T^{1/2}\delta_{NT}^{-1})$. Next, consider the second term. Because F^0H^r and F^0 span the same column space, $V(r,F^0H^r) = V(r,F^0)$. Thus the second term can be rewritten as $V(k,F^0H^k) - V(r,F^0)$, which is $\mathrm{O_p}(T/\log\log T)$ [but not $\mathrm{o_p}(T/\log\log T)$ and is positive] by Lemma A.3. Hence, $V(k,\hat{F}^k) - V(r,\hat{F}^r) = \mathrm{O_p}(T/\log\log T)$. Using $PC(k) - PC(r) = V(k,\hat{F}^k) - V(r,\hat{F}^r) - (r-k)g(N,T)$, it follows that if $g(N,T)(\log\log T)/T \to 0$, then $P[PC(k) <$

$PC(r)] \to 0$. Next for $k > r$,

$$P[PC(k) - PC(r) < 0] = P[V(r, \hat{F}^r) - V(k, \hat{F}^k) > (k - r)g(N, T)].$$

By Lemma A.4, $V(r, \hat{F}^r) - V(k, \hat{F}^k) = O_p(1)$. For $k > r$, $(k-r)g(N,T) \geqslant g(N,T) \to \infty$. Thus for $k > r$, $P[PC(k) < PC(r)] \to 0$ as $N, T \to \infty$. □

Appendix B. Distribution theory

Lemma B.1. *Under Assumptions* A–D, *we have for* $k = r$ (*omitting superscript* k)

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

where $H_2 = H_1 V_{NT}^{-1}$ *has full rank and* V_{NT} *is an* $r \times r$ *diagonal matrix consisting of the first* r *largest eigenvalues of* $(1/NT^2)XX'$ *in decreasing order.*

Proof. From the identity $\hat{F} = (1/NT^2)XX'\tilde{F}$ and the definition of eigenvectors and eigenvalues, $(1/T^2N)XX'\tilde{F} = \tilde{F}V_{NT}$. The left-hand side is simply $\hat{F}$. Thus, we obtain immediately another identity linking $\hat{F}$ and $\tilde{F}$. That is, $\hat{F} = \tilde{F}V_{NT}$, or $\hat{F}_t = V_{NT}\tilde{F}_t$ and $\tilde{F}_t = V_{NT}^{-1}\hat{F}_t$, $(t = 1, 2, \ldots, T)$. Thus $\tilde{F}_t - H_2'F_t^0 = V_{NT}^{-1}(\hat{F}_t - H_1'F_t^0)$. The lemma follows from Lemma 1 and $\|V_{NT}^{-1}\| = O_p(1)$ (see Lemma B.3 below). □

Note that both H_1 and H_2 are full rank matrices, and $\|H_1\| = O_p(1)$ and $\|H_2\| = O_p(1)$.

Proof of Lemma 2. By taking $k = r$, we can rewrite (A.1) as

$$\hat{F}_t - H_1'F_t^0 = T^{-2}\sum_{s=1}^{T} \tilde{F}_s \gamma_N(s,t) + T^{-2}\sum_{s=1}^{T} \tilde{F}_s \zeta_{st} + T^{-2}\sum_{s=1}^{T} \tilde{F}_s \eta_{st} + T^{-2}\sum_{s=1}^{T} \tilde{F}_s \xi_{st}. \tag{B.1}$$

The desired result follows from Lemma B.2 below. □

Lemma B.2. *Under Assumptions* A–E, *we have, for each* t,

(a) $T^{-2}\sum_{s=1}^{T} \tilde{F}_s \gamma_N(s,t) = O_p(T^{-3/2})$;
(b) $T^{-2}\sum_{s=1}^{T} \tilde{F}_s \zeta_{st} = O_p\left(\frac{1}{\sqrt{NT}}\right)$;
(c) $T^{-2}\sum_{s=1}^{T} \tilde{F}_s \eta_{st} = O_p\left(\frac{1}{\sqrt{N}}\right)$;
(d) $T^{-2}\sum_{s=1}^{T} \tilde{F}_s \xi_{st} = O_p\left(\frac{1}{\sqrt{NT}}\right)$.

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

$$T^{-2}\sum_{s=1}^{T}\tilde{F}_s\gamma_N(s,t)=T^{-2}\sum_{s=1}^{T}(\tilde{F}_s-H_2'F_s^0)\gamma_N(s,t)+H_2'T^{-2}\sum_{s=1}^{T}F_s^0\gamma_N(s,t). \quad \text{(B.2)}$$

Now $T^{-2}\sum_{s=1}^{T}F_s^0\gamma_N(s,t)=\mathrm{O_p}(T^{-3/2})$ since $E|\sum_{s=1}^{T}F_s^0\gamma_N(s,t)|\leqslant(\max_s E\|F_s^0\|)\sum_{s=1}^{T}|\gamma_N(s,t)|\leqslant T^{1/2}M$ by $E\|F_t^0\|=\mathrm{O}(\sqrt{T})$ and Assumption E(1). Consider the first term:

$$\left\|T^{-2}\sum_{s=1}^{T}(\tilde{F}_s-H_2'F_s^0)\gamma_N(s,t)\right\|\leqslant\frac{1}{T^{3/2}}\left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{F}_s-H_2'F_s^0\|^2\right)^{1/2}\times\left(\sum_{s=1}^{T}|\gamma_N(s,t)|^2\right)^{1/2},$$

which is $T^{-3/2}\mathrm{O_p}(\delta_{NT}^{-1})$ by Lemma B.1 and Assumption E(1). Consider part (b).

$$T^{-2}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}=T^{-2}\sum_{s=1}^{T}(\tilde{F}_s-H_2'F_s^0)\zeta_{st}+H_2'T^{-2}\sum_{s=1}^{T}F_s^0\zeta_{st}=\mathrm{I}+\mathrm{II}.$$

$$\mathrm{I}=\|T^{-2}\sum_{s=1}^{T}(\tilde{F}_s-H_2'F_s^0)\zeta_{st}\|\leqslant\frac{1}{T}\left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{F}_s-H_2'F_s^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\zeta_{st}^2\right)^{1/2}.$$

Furthermore,

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

Thus term I is $\frac{1}{T}\mathrm{O_p}(\frac{1}{\delta_{NT}})\mathrm{O_p}\left(\frac{1}{\sqrt{N}}\right)$. Consider II,

$$\left\|T^{-2}\sum_{s=1}^{T}F_s^0\zeta_{st}\right\|\leqslant\frac{1}{\sqrt{T}}\left(\frac{1}{T^2}\sum_{s=1}^{T}\|F_s^0\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\zeta_{st}^2\right)^{1/2}=\mathrm{O_p}\left(\frac{1}{\sqrt{NT}}\right).$$

Thus (I) is dominated by (II) and hence (b) is $\mathrm{O_p}\left(1/\sqrt{NT}\right)$. Consider part (c).

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

Now $T^{-2}\sum_{s=1}^{T}F_s^0\eta_{st}=((1/T^2)\sum_{s=1}^{T}F_s^0F_s^{0\prime})(1/N)\sum_{k=1}^{N}\lambda_k e_{kt}=\mathrm{O_p}\left(1/\sqrt{N}\right)$. The first term above is

$$\left\|T^{-2}\sum_{s=1}^{T}(\tilde{F}_s-H_2'F_s^0)\eta_{st}\right\|\leqslant T^{-1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{F}_s-H_2'F_s^0\|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{s=1}^{T}\eta_{st}^2\right)^{1/2}.$$

But

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

for $\|\Lambda^{0\prime}e_t/\sqrt{N}\|^2 = \mathrm{O_p}(1)$. Thus, $\|T^{-2}\sum_{s=1}^{T}(\tilde{F}_s - H_2'F_s^0)\eta_{st}\| = \mathrm{O_p}((TN)^{-1/2}\delta_{NT}^{-1})$. Therefore, (c) is $\mathrm{O_p}\left(\frac{1}{\sqrt{N}}\right)$. Finally for part (d),

$$T^{-2}\sum_{s=1}^{T}\tilde{F}_s\xi_{st} = T^{-2}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}\Lambda^{0\prime}e_s/N = T^{-2}\sum_{s=1}^{T}(\tilde{F}_se_s'\Lambda^0/N)F_t^0$$

$$= \frac{1}{NT^2}\sum_{s=1}^{T}(\tilde{F}_s - H_2'F_s^0)e_s'\Lambda^0F_t^0 + \frac{1}{NT^2}\sum_{s=1}^{T}H_2'F_s^0e_s'\Lambda^0F_t^0.$$

Consider the first term

$$\left\|\left(\frac{1}{NT^2}\sum_{s=1}^{T}(\tilde{F}_s - H_2'F_s^0)e_s'\Lambda^0\right)F_t^0\right\| \leqslant \frac{1}{\sqrt{TN}}\left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{F}_s - H_2'F_s^0\|^2\right)^{1/2}$$

$$\times\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{e_s'\Lambda^0}{\sqrt{N}}\right\|^2\right)^{1/2}\frac{\|F_t^0\|}{\sqrt{T}},$$

which is $\mathrm{O_p}\left(1/\sqrt{TN}C_{NT}\right)$. For the second term of (d),

$$\frac{1}{NT^2}\sum_{s=1}^{T}F_s^0e_s'\Lambda^0F_t^0 = \frac{1}{\sqrt{NT}}\left(\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\sum_{k=1}^{N}F_s^0\lambda_k^{0\prime}e_{ks}\right)\left(F_t^0/\sqrt{T}\right) = \mathrm{O_p}((TN)^{-1/2}).$$

Thus, (d) is $\mathrm{O_p}\left(1/\sqrt{NT}\right)$. The proof of Lemma B.2 is complete. □

Lemma B.3. *Assume Assumptions* A–D *hold. Let V_{NT} be the diagonal matrix consisting of the first r largest eigenvalues of $(1/T^2N)XX'$ and let $V_{NT}^{\dagger}$ be counterpart of the matrix $(1/T^2N)F^0(\Lambda^{0\prime}\Lambda^0)F^{0\prime}$. As, $T,N\to\infty$,*

(i) $T^{-2}\tilde{F}'((1/T^2N)XX')\tilde{F} = V_{NT}\xrightarrow{d}V$,
(ii) $V_{NT}^{\dagger} = V_{NT} + \mathrm{o_p}(1)$,
(iii) $\tilde{F}'F^0/T^2(\Lambda^{0\prime}\Lambda^0/N)F^{0\prime}\tilde{F}/T^2\xrightarrow{d}V$,

where V is a diagonal matrix (random) consisting of the eigenvalues of $\Sigma_\Lambda\int B_uB_u'$.

Proof. By the definition of eigenvalues and eigenvectors, we have $(1/T^2N)XX'\tilde{F} = \tilde{F}V_{NT}$. From $\tilde{F}'\tilde{F}/T^2 = I$, we further have $T^{-2}\tilde{F}'(1/T^2N)XX'\tilde{F} = V_{NT}$. Next,

$$\|T^{-4}N^{-1}\tilde{F}'(XX')\tilde{F} - T^{-4}N^{-1}\tilde{F}'F^0(\Lambda^{0\prime}\Lambda^0)F^{0\prime}\tilde{F}\| = \mathrm{o_p}(1) \tag{B.3}$$

is implied by the proof of Lemma 2. Associated with the eigenvalues of $V_{NT}^{\dagger}$ is the eigenvector matrix $\tilde{F}_{\dagger}$ such that $\tilde{F}_{\dagger}'\tilde{F}_{\dagger}/T^2 = \mathrm{I}$. By definition, $T^{-4}N^{-1}\tilde{F}_{\dagger}'F^0(\Lambda^{0\prime}\Lambda^0)$

$F^{0\prime}\tilde{F}_{\dagger} = V_{NT}^{\dagger}$. It can be shown that

$$\|T^{-4}N^{-1}\tilde{F}'F^0(\Lambda^{0\prime}\Lambda^0)F^{0\prime}\tilde{F} - T^{-4}N^{-1}\tilde{F}_{\dagger}'F^0(\Lambda^{0\prime}\Lambda^0)F^{0\prime}\tilde{F}_{\dagger}\| = o_p(1). \tag{B.4}$$

Note that the r largest eigenvalues of the matrix $(1/T^2)F^0(\Lambda^{0\prime}\Lambda^0/N)F^{0\prime}$ are the same as those of the matrix $(\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}F^0/T^2)$. But the latter matrix converges in distribution to $\Sigma_\Lambda \int B_u B_u'$. Thus $V_{NT}^{\dagger}$ converges in distribution to V. Eqs. (B.3) and (B.4) imply $V_{NT} = V_{NT}^{\dagger} + o_p(1)$. Thus $V_{NT} \xrightarrow{d} V$, proving parts (i) and (ii). Part (iii) follows from part (i) and (B.3). □

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 3. Let V_{NT} be as stated earlier, then $(1/T^2N)XX'\tilde{F} = \tilde{F}V_{NT}$. Multiplying this identity on both sides by $T^{-2}(\Lambda^{0\prime}\Lambda^0/N)^{1/2}F^{0\prime}$, we have

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

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

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

$$= \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T^2}\right)V_{NT}, \tag{B.5}$$

where $d_{NT} = (\Lambda^{0\prime}\Lambda^0/N)^{1/2}[(F^{0\prime}F^0/T^2)\Lambda^{0\prime}e'\tilde{F}/(T^2N) + (1/T^2N)F^{0\prime}e\Lambda^0F^{0\prime}(\tilde{F})/T^2 + (1/T^2N)F^{0\prime}ee'\tilde{F}/T^2] = o_p(1)$, where the $o_p(1)$ is implied by Lemma B.2. Let

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

and

$$R_{NT} = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{1/2}\left(\frac{F^{0\prime}\tilde{F}}{T^2}\right), \tag{B.6}$$

then we can rewrite (B.5) as

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

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

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

This equality holds because V_{NT}^* and V_{NT} are diagonal matrices and thus commutable. Therefore Υ_{NT} is the eigenvector matrix of $[B_{NT} + d_{NT}R_{NT}^{-1}]$. Note that $B_{NT} + d_{NT}R_{NT}^{-1}$

converges to $B=\Sigma_\Lambda^{1/2}(\int B_u B_u')\Sigma_\Lambda^{1/2}$ by Assumptions A and B and $d_{NT}=o_p(1)$. Because the eigenvalues of B are distinct by Assumption F, the eigenvalues of $B_{NT}+d_{NT}R_{NT}^{-1}$ will also be distinct for large N and large T. 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 (B.6)) depends on $\tilde F$ only through the kth column of $\tilde F$ ($k=1,2,\ldots,r$). Thus the sign of each column in R_{NT} and thus in $\Upsilon_{NT}=R_{NT}V_{NT^*}^{-1/2}$ is implicitly determined by the sign of each column in $\tilde F$. Thus, given the column sign of $\tilde F$, Υ_{NT} is uniquely determined. By the eigenvector perturbation theory (which requires the distinctness of eigenvalues, see Franklin (1968)), there exists a unique eigenvector matrix Υ of $B=\Sigma_\Lambda^{1/2}\int B_u B_u'\Sigma_\Lambda^{1/2}$ such that $\Upsilon_{NT}\xrightarrow{d}\Upsilon$. Rewrite (B.6) as $F^{0\prime}\tilde F/T^2=(\Lambda^{0\prime}\Lambda^0/N)^{-1/2}R_{NT}=(\Lambda^{0\prime}\Lambda^0/N)^{-1/2}\Upsilon_{NT}(V_{NT}^*)^{1/2}$, we have $F^{0\prime}\tilde F/T^2\xrightarrow{d}\Sigma_\Lambda^{-1/2}\Upsilon V^{1/2}$ by Assumption B and by $V_{NT}^*\xrightarrow{d}V$ in view of Lemma B.3(ii). □

Proof of Theorem 2. By Lemma B.2, we have

$$\hat F_t-H_1'F_t^0=O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{B.7}$$

and

$$\sqrt{N}(\hat F_t-H_1'F_t^0)=O_p(N^{1/2}T^{-3/2})+O_p(T^{-1/2})+O_p(1)+O_p(T^{-1/2}).$$

The limiting distribution is determined by the third term of (B.1). By the definition of η_{st},

$$\sqrt{N}(\hat F_t-H_1'F_t^0)=T^{-2}\sum_{s=1}^T(\tilde F_s F_s^{0\prime})\frac{1}{\sqrt N}\sum_{i=1}^N\lambda_i^0 e_{it}+o_p(1).$$

Now $\left(1/\sqrt{N}\right)\sum_{i=1}^N\lambda_i^0 e_{it}\xrightarrow{d}N(0,\Gamma_t)$ by Assumption G. Together with Proposition 3, we have $\sqrt{N}(\hat F_t-H_1'F_t)\xrightarrow{d}QN(0,\Gamma_t)$ as stated. Note that Q is independent of $N(0,\Gamma_t)$. This is because the limiting behavior of $(\tilde F' F^0/T^2)$ is determined by the common trends only, see (B.4) and the common trends are independent of the idiosyncratic errors. □

Proof of Corollary 1. From $\tilde F_t=V_{NT}^{-1}\hat F_t$ and $H_2'=V_{NT}^{-1}H_1'$, we have $\sqrt{N}(\tilde F_t-H_2'F_t^0)=V_{NT}^{-1}\sqrt{N}(\hat F_t-H_1'F_t^0)$. The corollary follows from Theorem 2 and $V_{NT}\to V$ by Lemma B.3. □

Lemma B.4. *Under Assumptions* A–E, *the* $r\times r$ *matrices satisfy*

(i) $T^{-1}(\hat F-F^0H_1)'F^0=O_p(T^{-1})+O_p(N^{-1/2})$,
(ii) $T^{-1}(\hat F-F^0H_1)'\hat F=O_p(T^{-1})+O_p(N^{-1/2})$,
(iii) $T^{-1}\|\sum_{t=1}^T(\hat F_t-H_1'F_t^0)\|=O_p(T^{-3/2})+O_p((NT)^{-1/2})$.

Furthermore, (i)–(iii) *hold with* $\hat F$ *replaced by* $\tilde F$ *and* H_1 *replaced by* H_2.

Proof. Consider (i). Using the identity (B.1), we have

$$T^{-1}\sum_{t=1}^{T}(\hat{F}_t - H_1'F_t^0)F_t^{0\prime} = T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}\gamma_N(s,t) + T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}\zeta_{st}$$

$$+ T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}\eta_{st} + T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}\xi_{st}$$

$$= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.$$

We shall show $I = O_p(T^{-1})$. We can simply treat $\tilde{F}_s$ as $H_2'F_s^0$ because (after adding and subtracting) the term involving $\tilde{F}_s - H_2'F_s^0$ is a dominated term. Thus it is sufficient to prove $T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0F_t^{0\prime}\gamma_N(s,t) = O_p(T^{-1})$. But this follows from $E\|F_s^0F_t^{0\prime}\gamma_N(s,t)\| \leqslant |\gamma_N(s,t)|\max_{1\leqslant t\leqslant T}E\|F_t^0\|^2 \leqslant MT\gamma_N(s,t)$ and Assumption C2. Consider II. From $\zeta_{st} = e_s'e_t/N - \gamma_N(s,t)$, $\mathrm{II} = (1/T^3N)\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_sF_t^{0\prime}e_s'e_t - \mathrm{I}$. To prove $\mathrm{II} = O_p(T^{-1})$, it suffices to prove (treating $\tilde{F}_s$ as $H_2'F_s^0$, i.e., ignoring $\tilde{F}_s - H_2'F_s^0$)

$$\frac{1}{T^3N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s^0F_t^{0\prime}e_s'e_t = O_p(T^{-1}).$$

The above is bounded by

$$\frac{1}{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}F_s^0e_{is}\right)\left(\frac{1}{T}\sum_{t=1}^{T}F_t^0e_{it}\right)'\right\| \leqslant \frac{1}{T}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t^0e_{it}\right\|^2$$

$$= O_p(T^{-1})$$

because the assumptions imply $E\|\frac{1}{T}\sum_{t=1}^{T}F_t^0e_{it}\|^2 \leqslant M$. Next, rewrite III as

$$\mathrm{III} = \left(\frac{1}{T^2}\sum_{s=1}^{T}\tilde{F}_sF_s^{0\prime}\right)\left(\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}\lambda_k^0F_t^{0\prime}e_{kt}\right).$$

The first expression is $O_p(1)$ and the second is $O_p(N^{-1/2})$. Thus $\mathrm{III} = O_p(N^{-1/2})$. Consider IV. Because η_{st} is a scalar we can use $F_t^{0\prime}\eta_{st} = \eta_{st}F_t^{0\prime} = \eta_{st}'F_t^{0\prime}$ to rewrite IV as

$$\mathrm{IV} = \left(\frac{1}{TN}\sum_{s=1}^{T}\tilde{F}_se_s'\Lambda^0\right)\left(\frac{1}{T^2}\sum_{t=1}^{T}F_t^0F_t^{0\prime}\right).$$

The second expression is $O_p(1)$ and the first can be rewritten as

$$\frac{1}{TN}\sum_{s=1}^{T}(\tilde{F}_s - H_2'F_s^0)\sum_{k=1}^{N}\lambda_k^{0\prime}e_{ks} + H_2'\frac{1}{TN}\sum_{s=1}^{T}\sum_{k=1}^{N}F_s^0\lambda_k^{0\prime}e_{ks}. \tag{B.8}$$

The first term of (B.8) is bounded by $((1/T)\sum_{s=1}^{T}\|\tilde{F}_s - H_2'F_s^0\|^2)^{1/2}((1/T)\sum_{s=1}^{T}\|(1/\sqrt{N})\sum_{k=1}^{N}\lambda_k^0e_{ks}\|^2)^{1/2}N^{-1/2}$, which is $O_p(N^{-1/2}\delta_{NT}^{-1})$. The second term is $O_p(N^{-1/2})$. Combining results, we prove (i).

Part (ii) of this lemma follows easily from (i). Rewrite

$$T^{-1}\hat{F}'(F^0H_1 - \hat{F}) = -T^{-1}(\hat{F} - F^0H_1)'(\hat{F} - F^0H_1). - H_1'T^{-1}F^{0\prime}(\hat{F} - F^0H_1).$$

The norm of the first expression is bounded by $T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_1'F_t^0\|^2 = \mathrm{O_p}(\delta_{NT}^{-2})$ and the second expression is $\mathrm{O_p}(T^{-1} + N^{-1/2})$ by part (i), proving part (ii).

Next consider (iii). By (B.7), each term except the third on the right-hand side is of $\mathrm{O_p}(T^{-3/2})$ or $\mathrm{O_p}((NT)^{-1/2})$, thus the average over t is still of this magnitude. The average over t for the third term is given by $T^{-2}\sum_{s=1}^{T}(\tilde{F}_sF_s^{0\prime})(1/TN)\sum_{i=1}^{N}\lambda_i^0 e_{it}$, which is equal to $\mathrm{O_p}((NT)^{-1/2})$, proving (iii).

The remaining claim of this lemma follows from the relationship between $\hat{F}$ and $\tilde{F}$. □

Proof of Lemma 3 and Theorem 3. The estimator $\hat{\lambda}_i$ has an alternative expression: $\hat{\lambda}_i = (\hat{F}'\hat{F})^{-1}\hat{F}'\underline{X}_i$, where $\underline{X}_i = F^0\lambda_i^0 + \underline{e}_i$. Thus

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

Rewriting $\hat{F}'\underline{e}_i = H_1'F^{0\prime}\underline{e}_i + (\hat{F} - F^0H_1)'\underline{e}_i$, we have

$$\begin{aligned}T(\hat{\lambda}_i - H_1^{-1}\lambda_i^0) &= \left(\frac{\hat{F}'\hat{F}}{T^2}\right)^{-1} H_1'\frac{1}{T}F^{0\prime}\underline{e}_i + \left(\frac{\hat{F}'\hat{F}}{T^2}\right)^{-1}\frac{1}{T}(\hat{F} - F^0H_1)'\underline{e}_i\\ &\quad + \left(\frac{\hat{F}'\hat{F}}{T^2}\right)^{-1}\frac{1}{T}\hat{F}'(F^0H_1 - \hat{F})H_1^{-1}\lambda_i^0. \qquad \text{(B.9)}\end{aligned}$$

Consider each term on the right-hand side above. From $\|(\hat{F}'\hat{F}/T^2)^{-1}\| = \mathrm{O_p}(1)$, $\|H_1\| = \mathrm{O_p}(1)$, and $T^{-1}F^{0\prime}\underline{e}_i = T^{-1}\sum_{t=1}^{T}F_t^0 e_{it} = \mathrm{O_p}(1)$, the first term of (B.9) is $\mathrm{O_p}(1)$. Next,

$$\begin{aligned}T^{-1}\|(\hat{F} - F^0H_1)'\underline{e}_i\| &= T^{-1}\left\|\sum_{t=1}^{T}(\hat{F}_t - H_1'F_t^0)e_{it}\right\|\\ &\leqslant \left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_1'F_t^0\|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}e_{it}^2\right)^{1/2}\\ &= \mathrm{O_p}(\delta_{NT}^{-1})\mathrm{O_p}(1).\end{aligned}$$

So the second expression of (B.9) is $\mathrm{O_p}(\delta_{NT}^{-1})$. Finally, by Lemma B.4(ii), the last expression of (B.9) is $\mathrm{o_p}(1)$. Combining results we obtain Lemma 3. In addition,

$$T(\hat{\lambda}_i - H_1^{-1}\lambda_i^0) = \left(\frac{\hat{F}'\hat{F}}{T^2}\right)^{-1} H_1'\frac{1}{T}\sum_{t=1}^{T}F_t^0 e_{it} + \mathrm{o_p}(1). \qquad \text{(B.10)}$$

Note that $(\hat{F}'\hat{F}/T^2)^{-1}H_1' = H_1^{-1}(F^{0\prime}F^0/T^2)^{-1} + o_p(1)$ and $H_1^{-1}(F^{0\prime}F^0/T^2)^{-1} = [(\Lambda^{0\prime}\Lambda^0/N)(F^{0\prime}\tilde{F}/T^2)]^{-1}(F^{0\prime}F^0/T^2)^{-1} \to (\Sigma_\Lambda Q')^{-1}(\int B_u B_u')^{-1}$, see Proposition 3. This implies Theorem 3 in view of Assumption H. □

Proof of Corollary 2. The mathematical identity $\tilde{\Lambda}=\hat{\Lambda}V_{NT}$ can be shown to hold. That is, $\tilde{\lambda}_i=V_{NT}\hat{\lambda}_i$. From $H_2=H_1V_{NT}^{-1}$, or $H_2^{-1}=V_{NT}H_1^{-1}$, we have $T(\tilde{\lambda}_i-H_2^{-1}\lambda_i^0)=V_{NT}T(\hat{\lambda}_i-H_1^{-1}\lambda_i^0)$. The corollary follows immediately from Theorem 3 and $V_{NT}\to V$. □

Proof of Theorem 4. From $C_{it}^0=F_t^{0\prime}\lambda_i^0$ and $\hat{C}_{it}=\hat{F}_t'\hat{\lambda}_i$, we have

$$\hat{C}_{it}-C_{it}^0=(\hat{F}_t-H_1'F_t^0)'H_1^{-1}\lambda_i^0+\hat{F}_t'(\hat{\lambda}_i-H_1^{-1}\lambda_i^0). \tag{B.11}$$

(i) $N/T\to 0$. Multiplying $\sqrt{N}$ on each side of above,

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

Because $N/T\to 0$ and $\hat{F}_t=O_p(T^{1/2})$, the second term on the right is $o_p(1)$. By Theorem 2,

$$\begin{aligned}\sqrt{N}(\hat{F}_t-H_1'F_t^0)&=\left(\frac{\tilde{F}'F^0}{T^2}\right)\frac{1}{\sqrt{N}}\sum_{j=1}^N\lambda_j^0e_{jt}+o_p(1)\\&=\left(\frac{\tilde{F}'F^0}{T^2}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^N\lambda_j^0e_{jt}+o_p(1)\\&=H_1'\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^N\lambda_j^0e_{jt}+o_p(1)\end{aligned}$$

by the definition of H_1. Therefore,

$$\begin{aligned}\lambda_i^{0\prime}(H_1')^{-1}\sqrt{N}(\hat{F}_t-H_1'F_t^0)&=\lambda_i^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^N\lambda_j^0e_{jt}+o_p(1)\\&\xrightarrow{d}N(0,\lambda^{0\prime}_i\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\lambda_i^0).\end{aligned} \tag{B.12}$$

That is, $\sqrt{N}(\hat{C}_{it}-C_{it}^0)\xrightarrow{d}N(0,\lambda_i^{0\prime}\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\lambda_i^0)$.

(ii) If $T/N\to 0$ then $T^{1/2}(\hat{F}_t-H_1'F_t^0)=o_p(1)$ by (B.7). From (B.11) and (B.10),

$$\begin{aligned}\sqrt{T}(\hat{C}_{it}-C_{it}^0)&=o_p(1)+(T^{-1/2}\hat{F}_t')T(\hat{\lambda}_i-H_1^{-1}\lambda_i)\\&=o_p(1)+(T^{-1/2}F_t^{0\prime})(F^{0\prime}F^0/T^2)^{-1}\frac{1}{T}\sum_{s=1}^T F_s^0e_{is}\\&\xrightarrow{d}B_u(\tau)'\left(\int B_uB_u'\right)^{-1}\int B_u\,\mathrm{d}B_e^{(i)}.\end{aligned} \tag{B.13}$$

(iii) $N/T \to \pi$. Rewrite (B.11) as

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

The desired result follows from cases 1 and 2. In addition, $B_e^{(i)}$ is asymptotically independent of $N^{-1/2}\sum_{j=1}^{N}\lambda_j e_{jt}$ because the latter involves every cross-section unit. Moreover, $B_e^{(i)}$ is independent of B_u by assumption. Thus, the limits of (B.12) and (B.13) are independent. □

Proof of Proposition 4. For notational simplicity, we write H for H_2'. From $\hat{R}_t = \hat{\alpha} + \hat{\delta}'\tilde{F}_t$, $\hat{R}_t - \alpha - \delta' F_t^0 = (\hat{\alpha} - \alpha) + \hat{\delta}'\tilde{F}_t - \delta' F_t^0$. Adding and subtracting terms, and then multiplying $\sqrt{N}$ on each side, we have

$$\sqrt{N}(\hat{R}_t - \alpha - \delta' F_t^0) = \sqrt{N}(\hat{\alpha} - \alpha) + \sqrt{N}(\hat{\delta} - H^{-1\prime}\delta)'\tilde{F}_t + \delta' H^{-1}\sqrt{N}(\tilde{F}_t - HF_t^0).$$

We next show the first two terms on the right-hand side are negligible. Let $\imath = (1,1,\ldots,1)'$ and define $Z = (\imath, \tilde{F})$ to be the regressor matrix of (13). Let $\hat{\beta} = (\hat{\alpha}, \hat{\delta}')'$ be the least-squares estimator. Then $\hat{\beta} = (Z'Z)^{-1}Z'R$, where $R = (R_1, \ldots, R_T)'$. Under the null hypothesis, $R_t = \alpha + \delta' F_t^0$. Adding and subtracting terms, $R_t = \alpha + \delta' H^{-1}\tilde{F}_t + \delta' H^{-1}(HF_t^0 - \tilde{F}_t)$. In matrix notation, $R = Z\gamma - (\tilde{F} - F^0 H')H^{-1\prime}\delta$, where $\gamma = (\alpha, \delta' H^{-1})'$. Thus

$$\hat{\beta} - \gamma = -(Z'Z)^{-1}Z'(\tilde{F} - F^0 H')H^{-1\prime}\delta.$$

Define $D_T = diag\left(\sqrt{T}, T, \ldots, T\right)$, then $D_T^{-1}Z'ZD_T^{-1} = \mathrm{O_p}(1)$. We have

$$D_T(\hat{\beta} - \gamma) = (D_T^{-1}Z'ZD_T^{-1})^{-1}D_T^{-1}Z'(\tilde{F} - F^0 H')H^{-1\prime}\delta.$$

This implies that

$$\hat{\alpha} - \alpha = M_{11}\frac{1}{T}\imath'(\tilde{F} - F^0 H')H^{-1\prime}\delta + M_{12}\frac{1}{T^{3/2}}\tilde{F}'(\tilde{F} - F^0 H')H^{-1}\delta,$$

$$\hat{\delta} - H^{-1\prime}\delta = M_{21}\frac{1}{T^{3/2}}\imath'(\tilde{F} - F^0 H')H^{-1\prime}\delta + M_{22}\frac{1}{T^2} \times \tilde{F}'(\tilde{F} - F^0 H')H^{-1\prime}\delta, \tag{B.14}$$

where M_{ij} is the (i,j)th block entry of $(D_T^{-1}Z'ZD_T^{-1})^{-1}$, partitioned conformably. Each M_{ij} is $\mathrm{O_p}(1)$. By Lemma B.4(iii), $T^{-1}\imath'(\tilde{F} - F^0 H') = \mathrm{O_p}(T^{-3/2}) + \mathrm{O_p}\left(1/\sqrt{NT}\right)$, and by Lemma B.4(ii). $T^{-3/2}\tilde{F}'(\tilde{F} - F^0 H') = \mathrm{O_p}(T^{-3/2}) + \mathrm{O_p}\left(1/\sqrt{NT}\right)$. Thus, $\sqrt{N}(\hat{\alpha} - \alpha) = \mathrm{O_p}(\sqrt{N}T^{-3/2}) + \mathrm{O_p}(T^{-1/2})$ and $\sqrt{N}(\hat{\delta} - H^{-1\prime}\delta) = \mathrm{O_p}(\sqrt{N}T^{-2}) + \mathrm{O_p}(T^{-1})$. Because $\tilde{F}_t = \mathrm{O_p}\left(\sqrt{T}\right)$, we have $\sqrt{N}(\hat{\delta} - H^{-1\prime}\delta)'\tilde{F}_t = \mathrm{O_p}\left(\sqrt{N}T^{-3/2}\right) + \mathrm{O_p}(T^{-1/2})$. Now it is clear that both $\sqrt{N}(\hat{\alpha} - \alpha)$ and $\sqrt{N}(\hat{\delta} - H^{-1\prime}\delta)'\tilde{F}_t$ converge to zero if $N/T^3 \to 0$. Thus,

$$\sqrt{N}(\hat{R}_t - \alpha - \delta' F_t^0) = \mathrm{o_p}(1) + \delta' H^{-1}\sqrt{N}(\tilde{F}_t - HF_t^0).$$

By Theorem 2 and Corollary 1,

$$\sqrt{N}(\tilde{F}_t - HF_t^0) = V_{NT}^{-1}\sqrt{N}(\hat{F}_t - H_1'F_t^0) \tag{B.15}$$

$$= V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T^2}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} + o_p(1). \tag{B.16}$$

Thus,

$$\delta' H^{-1}\sqrt{N}(\tilde{F}_t - HF_t^0) = \delta' H^{-1}V_{NT}^{-1}\left(\frac{\tilde{F}'F^0}{T^2}\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i^0 e_{it} + o_p(1).$$

Because H, V_{NT}^{-1}, and $T^{-2}\tilde{F}'F^0$ are all asymptotically independent of $\left(1/\sqrt{N}\right)\sum_{i=1}^{N}\lambda_i^0 e_{it}$, we have

$$\frac{\sqrt{N}(\hat{R}_t - \alpha - \delta' F_t^0)}{[\delta' H^{-1}V_{NT}^{-1}(\tilde{F}'F^0/T^2)\Gamma_t(\tilde{F}'F^0/T^2)'V_{NT}^{-1}H^{-1\prime}\delta]^{1/2}} \xrightarrow{d} N(0,1),$$

where Γ_t is the asymptotic variance of $\left(1/\sqrt{N}\right)\sum_{i=1}^{N}\lambda_i^0 e_{it}$. That is, $\Gamma_t = p\lim_{N\to\infty}(1/N)\sum_{i=1}^{N}e_{it}^2\lambda_i^0\lambda_i^{0\prime}$ under cross-section uncorrelation for idiosyncratic errors. We can estimate $(\tilde{F}'F^0/T^2)\Gamma_t(\tilde{F}'F^0/T^2)'$ by replacing F^0 with $\tilde{F}$, replacing Λ^0 by $\tilde{\Lambda}$ (equivalently, replacing λ_i^0 with $\tilde{\lambda}_i$ for every i), and by replacing e_{it}^2 with $\tilde{e}_{it}^2$. Noticing that $\tilde{F}'\tilde{F}/T^2 = I$, it is not difficult to show that

$$\left(\frac{\tilde{F}'F^0}{T^2}\right)\Gamma_t\left(\frac{\tilde{F}'F^0}{T^2}\right)' - \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i' = o_p(1).$$

Finally, $H^{-1\prime}\delta$ is estimated by $\hat{\delta}$, see (B.14). This implies that

$$\frac{\sqrt{N}(\hat{R}_t - \alpha - \delta' F_t^0)}{\left[\hat{\delta}' V_{NT}^{-1}\left((1/N)\sum_{i=1}^{N}\hat{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i'\right)V_{NT}^{-1}\hat{\delta}\right]^{1/2}} \xrightarrow{d} N(0,1),$$

proving Proposition 4. □

Appendix C. Generalized dynamic factor models

Lemma C.1. *Under the assumptions of Theorem* 5, *and for* H^k *defined in Lemma* 1, *we have, for each* $k \leqslant r$

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 = O_p(1). \tag{C.1}$$

This lemma provides a rough bound on the behavior of $\hat{F}_t^k$. Although not sharpest possible, the bound is sufficient for our purpose.

Proof. By the Beveridge–Nelson decomposition, we have $\lambda_i(L) = \lambda_i(1) + \lambda_i^*(L)(1-L)$, where $\lambda_i^*(L) = \sum_{j=0}^{\infty} a_{ij}^* L^j$ with $a_{ij}^* = \sum_{k=j+1}^{\infty} a_{ik}$. Thus, $\lambda_i(L)'F_t = \lambda_i(1)F_t + \lambda_i^*(L)u_t$. That is, $X_{it} = \lambda_i(1)'F_t + \lambda_i^*(L)u_t + e_{it}$, which can be rewritten as $X_{it} = \lambda_i' F_t + v_{it}$, where $\lambda_i = \lambda_i(1)$ and $v_{it} = \lambda_i^*(L)u_t + e_{it}$. Note that v_{it} does not have the weak cross-section correlation properties imposed on e_{it}. Thus Lemma C.1 is much weaker than Lemma 1.

Combining the first two terms on the right-hand side of (A.1), we can rewrite (A.1) as

$$\hat{F}_t^k - H^{k\prime}F_t^0 = T^{-2}\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st} + T^{-2}\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st} + T^{-2}\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st} \tag{C.2}$$

where, upon introducing $v_t = (v_{1t}, v_{2t}, \ldots, v_{Nt})'$

$$\zeta_{st} = \frac{v_s' v_t}{N}, \qquad \eta_{st} = F_s^{0\prime}\Lambda^{0\prime}v_t/N, \qquad \xi_{st} = F_t^{0\prime}\Lambda^{0\prime}v_s/N.$$

The rest of proof involving bounding each term on the right-hand side of (C.2). The proof is similar to that of Theorem 1 of Bai and Ng (2002), and thus is omitted. □

Lemma C.2. *Under the assumptions of Theorem* 5, *we have*, *for each* k *with* $1 \leqslant k \leqslant r$,

$$V(k, \hat{F}^k) - V(k, F^0H^k) = O_p(T^{1/2}).$$

Proof. Let $D_k = \hat{F}^{k\prime}\hat{F}^k/T^2$ and $D_0 = H^{k\prime}F^{0\prime}F^0H^k/T^2$. Then adding and subtracting,

$$P_{\hat{F}^k} - P_{F^0H^k} = T^{-2}[(\hat{F}^k - F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k)' + (\hat{F}^k - F^0H^k)D_k^{-1}H^{k\prime}F^{0\prime} + F^0H^kD_k^{-1}(\hat{F}^k - F^0H^k)' + F^0H^k(D_k^{-1} - D_0^{-1})H^{k\prime}F^{0\prime}];$$

$$V(k, F^0H^k) - V(k, \hat{F}^k) = N^{-1}T^{-1}\sum_{i=1}^{N}\underline{X}_i'(P_{\hat{F}^k} - P_{FH^k}^0)\underline{X}_i = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}, \tag{C.3}$$

where the four terms correspond to the decomposition of $P_{\hat{F}^k} - P_{F^0H^k}$ as above. Now

$$I = N^{-1}T^{-3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)'D_k^{-1}(\hat{F}_s^k - H^{k\prime}F_s^0)X_{it}X_{is}$$

$$\leqslant \left(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}[(\hat{F}_t^k - H^{k\prime}F_t^0)'D_k^{-1}(\hat{F}_s^k - H^{k\prime}F_s^0)]^2\right)^{1/2} \times \left[T^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(N^{-1}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right]^{1/2}$$

$$\leqslant \left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)\|D_k^{-1}\|O_p(1) = O_p(1)$$

by Lemma C.1 and Lemma A.1(iii). Note that $\|D_k^{-1}\| = O_p(1)$, which is proved below.

$$\mathrm{II} = N^{-1}T^{-3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)' D_k^{-1} H^{k\prime}F_s^0 X_{it}X_{is}$$

$$\leqslant \left(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 \|H^{k\prime}F_s^0\|^2 \|D_k^{-1}\|^2\right)^{1/2}$$

$$\times \left[T^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(N^{-1}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right]^{1/2}$$

$$\leqslant T^{1/2}\left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)^{1/2} \|D_k^{-1}\| \left(T^{-2}\sum_{s=1}^{T}\|H^{k\prime}F_s^0\|^2\right)^{1/2} O_p(1)$$

$$= T^{1/2}\left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)^{1/2} O_p(1) = O_p(T^{1/2}).$$

It can be verified that III is also $O_p(T^{1/2})$. Next,

$$\mathrm{IV} = N^{-1}T^{-3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t^{0\prime}H^k(D_k^{-1} - D_0^{-1})H^{k\prime}F_s^0 X_{it}X_{is}$$

$$\leqslant T\|D_k^{-1} - D_0^{-1}\| N^{-1}\sum_{i=1}^{N}\left(T^{-2}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\| \cdot |X_{it}|\right)^2 = T\|D_k^{-1} - D_0^{-1}\| O_p(1),$$

where $O_p(1)$ is obtained because the term is bounded (using $|X_{it}| \leqslant \|\lambda_i^0\|\|F_t^0\| + |e_{it}|$) by $2\|H^k\|^2(\frac{1}{T^2}\sum_{t=1}^{T}\|F_t^0\|^2)^2(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i^0\|^2) + 2\|H^k\|^2\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{T^2}\sum_{t=1}^{T}\|F_t^0\||e_{it}|)^2$, which is $O_p(1)$ by Assumptions A and B. Next, we prove that $\|D_k - D_0\| = O_p(T^{-1/2})$. Notice

$$D_k - D_0 = T^{-2}\sum_{t=1}^{T}[\hat{F}_t^k\hat{F}_t^{k\prime} - H^{k\prime}F_t^0F_t^{0\prime}H^k]$$

$$= T^{-2}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)(\hat{F}_t^k - H^{k\prime}F_t^0)'$$

$$+ T^{-2}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)F_t^{0\prime}H^k + T^{-2}\sum_{t=1}^{T}H^{k\prime}F_t^0(\hat{F}_t^k - H^{k\prime}F_t^0)',$$

$$\|D_k - D_0\| \leqslant T^{-2}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2$$
$$+2T^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)^{1/2}\left(T^{-2}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2\right)^{1/2}$$
$$= T^{-1}\mathrm{O_p}(1) + T^{-1/2}\mathrm{O_p}(1)\mathrm{O_p}(1) = T^{-1/2}\mathrm{O_p}(1).$$

Because the limit of $F^{0\prime}F^0/T^2$ is positive definite and $rank(H^k) = k \leqslant r$, D_0 $(k \times k)$ is invertible and $\|D_0^{-1}\| = \mathrm{O_p}(1)$. From $\|D_k - D_0\| = T^{-1/2}\mathrm{O_p}(1) \to 0$, D_k is also invertible and $\|D_k^{-1}\| = \mathrm{O_p}(1)$. From $D_k^{-1} - D_0^{-1} = D_k^{-1}(D_0 - D_k)D_0^{-1}$, we have $\|D_k^{-1} - D_0^{-1}\| = \|D_k - D_0\|\mathrm{O_p}(1) = T^{-1/2}\mathrm{O_p}(1)$. Thus $\mathrm{IV} = T\|D_k - D_0\|\mathrm{O_p}(1) = \mathrm{O_p}(T^{1/2})$. This proves Lemma C.2.

Lemma C.3. *Lemma* A.3 *still holds. That is, for each k with $k < r$, there exists a $\tau > 0$,*

$$\liminf_{N,T\to\infty}\frac{\log\log T}{T}[V(k, F^0H^k) - V(r, F^0)] = \tau > 0.$$

Proof. The proof is the same as that of Lemma A.3 but with $\underline{e}_i$ replaced by $\underline{v}_i$, where $\underline{v}_i = (v_{i1}, v_{i2}, \ldots, v_{iT})'$ with $v_{it} = \lambda_i^*(L)'u_t + e_{it}$. More specifically,

$$V(k, F^0H^k) - V(r, F^0) = N^{-1}T^{-1}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}(P_F^0 - P_{FH}^0)F^0\lambda_i^0$$
$$+ 2N^{-1}T^{-1}\sum_{i=1}^{N}\underline{v}_i'(P_F^0 - P_{FH}^0)F^0\lambda_i^0$$
$$+ N^{-1}T^{-1}\sum_{i=1}^{N}\underline{v}_i'(P_F^0 - P_{FH}^0)\underline{v}_i = \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

The proof for I and III is the same as before. That is, $\mathrm{III} \geqslant 0$ and $\liminf(\log\log T)\,\mathrm{I}/T > 0$. Next, $\mathrm{II} = 2N^{-1}T^{-1}\sum_{i=1}^{N}\underline{v}_i'P_{F^0}F^0\lambda_i^0 - 2N^{-1}T^{-1}\sum_{i=1}^{N}\underline{v}_i'P_{F^0H^k}F^0\lambda_i^0$. But

$$|N^{-1}T^{-1}\sum_{i=1}^{N}\underline{v}_i'P_F^0F^0\lambda_i^0| = |N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}F_t^{0\prime}\lambda_i^0|$$
$$\leqslant T^{1/2}\left(T^{-2}\sum_{t=1}^{T}\|F_t^0\|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}v_{it}\lambda_i^0\right\|^2\right)^{1/2}$$
$$= \mathrm{O_p}(T^{1/2}).$$

The second term of II is also $\mathrm{O_p}(T^{1/2})$, and hence $(\log\log T)\mathrm{II}/T = \mathrm{O_p}(\log\log T/T^{1/2}) \xrightarrow{p} 0$. This proves Lemma C.3. □

Lemma C.4. *For each fixed k with* $k \geqslant r$, $V(k,\hat{F}^k) - V(r,\hat{F}^r) = O_p(1)$.

Proof. This follows easily from $V(k,\hat{F}^k) = O_p(1)$ for each $k \geqslant r$ because

$$V(k,\hat{F}^k) \leqslant V(r,\hat{F}^r) \leqslant V(r,F^0) \leqslant \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 = O_p(1)$$

The first three inequalities are explained in the proof of Lemma A.4. The last equality follows from the assumptions on $\lambda_i(L)$, u_t, and e_{it}. □

Proof of Theorem 5. The proof is identical to that of Theorem 1. Instead of using Lemmas A.2–A.4 we use Lemmas C.2–C.4. □

Proof of Theorem 6 and Corollary 4. The proof of Theorem 6 is similar to that of Theorems 2 and 3; the proof of Corollary 4 is similar to that of Proposition 4. Thus these proofs are omitted. □

Appendix D. List of industries

- Agriculture, forestry, and fishing
  - Farms
  - Agricultural services, forestry, and fishing
- Mining
  - Metal mining
  - Coal mining
  - Oil and gas extraction
  - Nonmetallic minerals, except fuels
- Construction
- Manufacturing
  - Durable goods
    - Lumber and wood products
    - Furniture and fixtures
    - Stone, clay, and glass products
    - Primary metal industries
    - Fabricated metal products
    - Machinery, except electrical
    - Electric and electronic equipment
    - Motor vehicles and equipment
    - Other transportation equipment
    - Instruments and related products
    - Miscellaneous manufacturing industries

Nondurable goods
Food and kindred products
Tobacco manufactures
Textile mill products
Apparel and other textile products
Paper and allied products
Printing and publishing
Chemicals and allied products
Petroleum and coal products
Rubber and miscellaneous plastics products
Leather and leather products

Transportation and public utilities
Transportation
Railroad transportation
Local and interurban passenger transit
Trucking and warehousing
Water transportation
Transportation by air
Pipelines, except natural gas
Transportation services

Communication
Telephone and telegraph
Radio and television

Electric, gas, and sanitary services
Wholesale trade
Retail trade
Finance, insurance, and real estate

Banking
Credit agencies other than banks
Security and commodity brokers
Insurance carriers
Insurance agents, brokers, and service
Real estate
Holding and other investment offices
Services

Hotels and other lodging places
Personal services
Business services
Auto repair, services, and parking
Miscellaneous repair services
Motion pictures

Amusement and recreation services
Health services
Legal services
Educational services
Social services and membership organizations

Social services
Membership organizations
Miscellaneous professional services

References

Anderson, T.W., 1963. Asymptotic theory for principal component analysis. Annals of Mathematical Statistics 34, 122–148.

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Artis, M., Banerjee, A., Marcellino, M., 2001. Factor forecasts for the UK. CEPR Discussion Paper No. 3119.

Bai, J., 2003. Inference on factor models of large dimension. Econometrica 71, 135–171.

Bai, J., Ng, S., 2001. A PANIC attack on unit roots and cointegration. Boston College Working Number 519.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 91–221.

Baltagi, B.H., Kao, C., 2000. Nonstationary panels, cointegration in panels and dynamic panels: a survey. Advances in Econometrics 15, 7–51.

Banerjee, A., Marcellino, M., Osbat, C., 2001. Testing for PPP: should we use panel methods? Department of Economics, European University Institute.

Bernanke, B., Boivin, J., 2000. Monetary policy in a data rich environment, Department of Economics, Princeton University, Princeton, NJ.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51, 1305–1324.

Connor, G., Korajczyk, R., 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. Journal of Financial Economics 15, 373–394.

Favero, C.A., Marcellino, M., 2001. Large datasets, small models and monetary policy in Europe. IEP, Bocconi University, and CEPR.

Forni, M., Reichlin, L., 1998. Let's get real: a factor-analytic approach to disaggregated business cycle dynamics. Review of Economic Studies 65, 453–473.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000a. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000b. Reference cycles: the NBER methodology revisited. CEPR Discussion Paper 2400.

Franklin, J.E., 1968. Matrix Theory. Prentice-Hall, Englewood Cliffs, NJ.

Geweke, J., 1993. Comment on A dynamic index model for large cross sections. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, INdicators, and Forecasting. University of Chicago Press for the NBER, Chicago, pp. 285–306 (Chapter 7).

Geweke, J., Meese, R., 1981. Estimating regression models of finite but unknown order. International Economic Review 23 (1), 55–70.

Gregory, A., Head, A., 1999. Common and country-specific fluctuations in productivity, investment, and the current account. Journal of Monetary Economics 44 (3), 423–452.

Hall, S.G., Lazarova, S., Urga, G., 1999. A principle components analysis of common stochastic trends in heterogeneous panel data: some Monte Carlo evidence. Oxford Bulletin of Economics and Statistics 61, 749–767.

Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society, Series B 41, 190–195.

Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.

Lai, T.L., Wei, C.Z., 1983. Asymptotic properties of general autoregressive models and strong consistency of least squares estimates to their parameters. Journal of Multivariate Analysis 12, 346–370.

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

Moon, R., Perron, B., 2001. Testing for a unit root in panels with dynamic factors. Department of Economics, University of Southern California.

Quah, D., Sargent, T., 1993. A dynamic index model for large cross sections. In: Stock, J.H., Watson, M.W. (Eds.), Business Cycles, INdicators, and Forecasting. University of Chicago Press for the NBER, Chicago, pp. 285–306 (Chapter 7).

Stock, J.H., Watson, M.W., 1988. Testing for common trends. Journal of American Statistical Association 83, 1097–1107.

Stock, J.H., Watson, M.W., 1999a, Diffusion indexes. NBER Working Paper 6702.

Stock, J.H., Watson, M.W., 1999b. Forecasting inflation. Journal of Monetary Economics 44, 293–335.