Econometric Theory, **28**, 2012, 274–308.
doi:10.1017/S0266466611000338

EFFICIENT ESTIMATION OF FACTOR MODELS

IN CHOI
Sogang University

This paper considers the factor model $X_t = \Lambda F_t + e_t$. Assuming a normal distribution for the idiosyncratic error e_t conditional on the factors $\{F_t\}$, conditional maximum likelihood estimators of the factor and factor-loading spaces are derived. These estimators are called generalized principal component estimators (GPCEs) without the normality assumption. This paper derives asymptotic distributions of the GPCEs of the factor and factor-loading spaces. It is shown that variance of the GPCE of the common component is smaller than that of the principal component estimator studied in Bai (2003, *Econometrica* 71, 135–172). The approximate variance of the forecasting error using the GPCE-based factor estimates is derived and shown to be smaller than that based on the principal component estimator. The feasible GPCE (FGPCE) of the factor space is shown to be asymptotically equivalent to the GPCE. The GPCE and FGPCE are shown to be more efficient than the principal component estimator in finite samples.

1. INTRODUCTION

Methods for handling large data sets are becoming increasingly important in economics and finance because such data sets are more readily available these days. Factor models have been popular for large sets of data since they do not require much structural information regarding their data generating processes. They can extract essential information contained in large data sets, and the information can be used for subsequent analyses such as forecasting (e.g., Stock and Watson, 1999, 2002b; Artis, Banerjee, and Marcellino, 2005; Forni, Hallin, Lippi, and Reichlin, 2003; Huang, Lee, and Li, 2006), policy analysis (e.g., Favero, Marcellino, and Neglia, 2005; Bernanke, Boivin, and Eliasz, 2005), predictive regressions (e.g., Ludvigson and Ng, 2007), and instrumental variables estimation (cf. Bai and Ng, 2010). Factor models have also been used in regressions for panel data as a way of modeling cross-sectional correlation. Examples of this approach are Bai and Ng

This paper was presented at the Third Symposium on Econometric Theory and Applications, held in Hong Kong in April 2007. The author thanks the participants, especially Peter Phillips and Tae-Hwy Lee, for constructive comments on the paper. He also thanks Jörg Breitung for sending a copy of his paper related to this one. Guido Kuersteiner and two anonymous referees read this paper very carefully and provided insightful comments, for which the author is grateful. This paper was financially supported by the Research Grant Council of Hong Kong under project number 641207. Address correspondence to In Choi, Department of Economics, Sogang University, #1 Shinsu-dong, Mapo-gu, Seoul, Korea; e-mail: inchoi@google.com, inchoi@sogang.ac.kr.

 © Cambridge University Press 2011

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(2004), Moon and Perron (2004), and Phillips and Sul (2003). Both the static and dynamic factor models have been used for applications. The former is the factor model without leads and lags of the factor, while the latter allows dynamics in the factor. However, even the static factor model allows dynamics in the factor, though it is not explicitly modeled.

A common method of estimating the static factor model is the principal component estimation method proposed by Chamberlain and Rothschild (1983) and Connor and Korajzcyk (1986), and further studied by Stock and Watson (2002a, 2002b), Bai and Ng (2002, 2006), and Bai (2003). Consistent estimation of the factor and factor-loading spaces is established in Stock and Watson (2002a), and asymptotic distributions of the principal component estimators of the factor and factor-loading spaces are derived in Bai (2003). Bai and Ng (2002) study the model selection issue for the static factor model. Besides the principal component method, Forni et al. (2000, 2005) use the dynamic principal component method of Brillinger (1981) to estimate the dynamic factor model. In addition, Kapetanios and Marcellino (2006) use methods for the state-space model to estimate the dynamic factor models with factor lags present.

Though there are now well-established methods for estimating factor models, the efficiency of these methods has not received much study. Exceptions are Jones (2001) and Boivin and Ng (2006). Jones considers time series heteroskedasticity in the idiosyncratic errors and shows through simulation that allowing heteroskedasticity improves the quality of extracted factors. Boivin and Ng provide some simulation results regarding the efficiency of the principal component estimation method.

We focus on the static factor model in this paper and study the issue of estimation efficiency. This issue is obviously important because greater efficiency in estimating the factor and factor-loading spaces will lead to better forecasts and more precise policy analysis. In order to tackle the efficiency issue, we start from the conditional likelihood function of the static factor model under a normality assumption and derive the conditional maximum likelihood estimators (CMLEs) of the factor and factor-loading spaces. These estimators are not identical with the principal component estimators in the literature, indicating that the existing estimators are suboptimal under a normality assumption. The CMLEs derived in this paper will be called generalized principal component estimators (GPCEs) without a normality assumption since they take into account the nonspherical structure of the error vector's variance–covariance matrix as do the generalized least squares (GLS) estimators. In contrast, the existing estimators will be called ordinary principal component estimators (OPCEs). The GPCE is not totally new. Stock and Watson[1] (2006) introduced it using weighted principal component analysis without invoking the likelihood principle. Boivin and Ng (2006) also considered a similar estimator that takes into account cross-sectional heteroskedasticity. However, neither of these articles formally studied the estimator's efficiency. In addition, Forni et al. (2005) introduce an estimator of the factor space, which they call generalized principal components of observed time series. The relationship

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

between this estimator and the GPCE is not obvious and needs to be investigated. More recently, Breitung and Tenhofen (2007) study the GPCE, focusing on serially correlated idiosyncratic errors. However, they do not consider cross-sectional correlation of the idiosyncratic errors which is the focus of this paper.

Asymptotic distributions of the GPCEs of the factor and factor-loading spaces are derived using the techniques of Stock and Watson (2002a), Bai and Ng (2002), and Bai (2003). These will show that the GPCEs enjoy better efficiency than the OPCEs in the limit when the idiosyncratic errors are weakly stationary. Forecasting regressions are also studied using the estimated factor space based on the GPCE, and an efficiency gain in forecasting is reported. The feasible GPCE (FGPCE) of the factor space that uses the estimates of idiosyncratic errors from ordinary principal component estimation is considered. It is shown to be asymptotically equivalent to the GPCE under some additional conditions on the variance-covariance matrix of the idiosyncratic errors, establishing that the FGPCE is as efficient as the GPCE.

This paper proceeds as follows. Section 2 introduces the CMLEs and GPCEs of the factor and factor-loading spaces. Section 3 derives asymptotic distributions of the CMLEs/GPCEs and compares their efficiency with that of the OPCEs. Section 4 studies forecasting regressions using the estimated factors based on the GPCEs. Section 5 establishes asymptotic equivalence of the GPCE and FGPCE. Section 6 reports some simulation results. Section 7 summarizes and concludes. All proofs are relegated to the Appendixes.

The following notation will be used throughout this paper. For a square matrix A and a vector x, $\|A\| = \sqrt{tr(A'A)}$, $ev_{\max}(A)$ and $ev_{\min}(A)$ are the maximum and minimum of the eigenvalues of A, respectively, and $\|A\|_1 = \max_x\{\|Ax\| : \|x\| = 1\}$. Here $\|A\|_1$ is the positive square root of the largest eigenvalue of $A'A$ and is called the spectral norm of the matrix A. It satisfies basic properties of the matrix norm. Note that $\|Aa\| \leq \|A\|_1 \|a\|$ for a square matrix A and a vector a. It follows from this that $\|AB\| \leq \|A\|_1 \|B\|$, where B is any matrix of conformable size. For matrices X and Y, $X \oplus Y = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$. $[X]_{i,j}$ denotes the (i, j)th element of matrix X. All the matrices in this paper are real ones. All the limits are taken as $N, T \to \infty$ unless otherwise stated.

2. CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION OF FACTOR MODELS

Consider the factor model

$$X_t = \Lambda F_t + e_t, \qquad (t = 1, \ldots, T), \tag{1}$$

where X_t is an $N \times 1$ vector of observations, Λ is an $N \times r$ matrix of factor loadings, F_t is an $r \times 1$ vector of latent factors, and e_t is an idiosyncratic component of the model.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Model (1) can be written in scalar notation as

$$X_{it} = \lambda_i' F_t + e_{it}, \qquad (i = 1, \ldots, N;\ t = 1, \ldots, T), \tag{2}$$

where λ_i' is the ith row of matrix Λ. In matrix notation it can also be written as

$$X = F\Lambda' + e, \tag{3}$$

where $X = \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix}$, $F = \begin{bmatrix} F_1' \\ \vdots \\ F_T' \end{bmatrix}$, and $e = \begin{bmatrix} e_1' \\ \vdots \\ e_T' \end{bmatrix}$.

Assume $e_t \mid F \sim iid\ N(0, \Omega)$, where $\Omega > 0$ and Ω is not related to F. Then, the conditional log-likelihood function (multiplied by -2) for model (1) is

$$-2l(\Lambda, F, \Omega) = TN\ln(2\pi) + T\ln|\Omega| + \operatorname{tr}\left\{\Omega^{-1}(X - F\Lambda')'(X - F\Lambda')\right\}. \tag{4}$$

Assume for the moment that the matrix Ω is known. Then, we need to minimize $\operatorname{tr}\left\{\Omega^{-1}(X - F\Lambda')'(X - F\Lambda')\right\}$ with respect to F and Λ for the conditional maximum likelihood estimation of the factor and factor-loading spaces. The standard theory of multivariate regression yields a preliminary CMLE of Λ, $\hat{\Lambda} = X'F(F'F)^{-1}$. Plugging this formula into the log-likelihood function (4) and omitting the first two terms, we obtain $\operatorname{tr}\left\{\Omega^{-1}X'(I - P_F)X\right\}$ with $P_F = F(F'F)^{-1}F'$. With the standardization $F'F = T \times I_r$, the CMLE of the factor space is obtained by maximizing $\operatorname{tr}\left\{\Omega^{-1}X'FF'X\right\} = \operatorname{tr}\left\{F'\left(X\Omega^{-1}X'\right)F\right\}$ with respect to F. Therefore, the CMLE of F, denoted by $\hat{F}$, is $\sqrt{T}$ times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix $X\Omega^{-1}X'$. The CMLE of Λ is given by $\hat{\Lambda} = \frac{1}{T}X'\hat{F}$.

By contrast, the principal component estimator of the factor space (see Connor and Korajzcyk, 1986; Stock and Watson, 2002a; Bai and Ng, 2002) is $\sqrt{T}$ times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix XX', and that of the factor loading is obtained in the same manner as for $\hat{\Lambda}$.

Alternatively, the log-likelihood function (4) can be maximized with respect to F first. Then, the CMLE of Λ, $\ddot{\Lambda}$, is $\sqrt{N}$ times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix $\Omega^{-1}X'X\Omega^{-1}$ when the standardization $\ddot{\Lambda}'\Omega^{-1}\ddot{\Lambda} = N \times I_r$ is used. The CMLE of F is given by $\ddot{F} = \frac{1}{N}X\Omega^{-1}\ddot{\Lambda}$.[2] Notice that the maximized value of the log-likelihood function is the same regardless of which pair of CMLEs is used. This implies that $\hat{F}\hat{\Lambda}' = \ddot{F}\ddot{\Lambda}'$.

The CMLEs introduced so far assume that Ω is known. The CMLE of Ω is defined by $\hat{\Omega} = \frac{1}{N}\left(X - \hat{F}\hat{\Lambda}'\right)'\left(X - \hat{F}\hat{\Lambda}'\right)$. In practice, an initial estimator of Ω is required. Recall that the CMLEs $\hat{F}$ and $\hat{\Lambda}$ are not MLEs. The MLEs require knowing the probabilistic structure of $\{F_t\}$ and use iterative computation methods

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(see Magnus and Neudecker, 1988, and Basilevsky, 1994, for discussions on the MLEs for the factor model with heteroskedastic, cross-sectionally uncorrelated idiosyncratic errors). By contrast, $\hat{F}$ and $\hat{\Lambda}$ do not require distributional assumptions regarding F, though they may be suboptimal compared to the MLEs.

3. ASYMPTOTIC DISTRIBUTIONS OF GENERALIZED PRINCIPAL COMPONENT ESTIMATORS

In this section we drop the normality assumption of Section 2 and study asymptotic distributions of the GPCEs. In particular, we are interested in comparing their properties with those of the OPCEs reported in Bai (2003). For the asymptotic distributions, assume the following.

Assumption 1.

(i) $\mathrm{E}(e_t) = 0$; $\mathrm{E}\left(e_t e_t'\right) = \Omega$ for all t and $M_1 < ev_{\min}(\Omega) < M_2$ for $M_1,\ M_2 > 0$, and $N = 1, 2, \ldots, \infty$; and $\left\| \frac{1}{T}\sum_{t=1}^{T} e_t e_t' - \Omega \right\| = O_p\left(\frac{N}{\sqrt{T}}\right)$.

(ii) Let $\gamma(s,t) = \mathrm{E}\left(\frac{1}{N} e_s' \Omega^{-1} e_t\right)$. Then, $\sum_{s=1}^{T} |\gamma(s,t)|$ is bounded uniformly in t and T.

(iii) $\mathrm{E}\left| \frac{1}{\sqrt{N}} \left[e_s' \Omega^{-1} e_t - \mathrm{E}\left(e_s' \Omega^{-1} e_t\right) \right] \right|^4$ is bounded uniformly in t, s, and N.

Part (i) of this assumption states that $\{e_t\}$ is a zero-mean process with a fixed variance–covariance matrix for every t. The variance-covariance matrix is positive definite for every finite N and even in the limit. Thus, Ω^{-1} exists for $N = 1, 2, \ldots, \infty$.[3] Since $\left\|\Omega^{-1}\right\|_1 = \frac{1}{ev_{\min}(\Omega)}$, it implies that $\left\|\Omega^{-1}\right\|_1$ is positive and bounded above by a constant uniformly in N including $N = \infty$. This fact will be used for proofs in Appendix B. If $\{e_{it} e_{jt}\}$ satisfies the central limit theorem, i.e., $\frac{1}{T}\sum_{t=1}^{T} e_{it} e_{jt} - \omega_{i,j} = O_p\left(\frac{1}{\sqrt{T}}\right)$ with $\omega_{i,j}$ denoting the (i,j)-th element of Ω, the third part of Assumption 1(i) will hold. Part (ii) assumes that $\{e_t\}$ is weakly dependent over time. Because of the assumed serial correlation of idiosyncratic errors, the GPCEs are not CMLEs even under a normality assumption. Part (iii) is a technical assumption required for Lemma B.1 in Appendix B. These assumptions can be proven if more elementary assumptions on $\{e_t\}$ are made.

For the factor-loading matrix and latent factors, the following assumptions are required.

Assumption 2.

(i) $\mathrm{rank}(\Lambda) = r$ and r is known.

(ii) $\frac{\Lambda'\Omega^{-1}\Lambda}{N} \to \Sigma_{\Lambda^*}$, where Σ_{Λ^*} is a positive-definite matrix.

(iii) $\|\lambda_i\| \leq \bar{\lambda} < \infty$ for every i.

(iv) $\frac{F'F}{T} \xrightarrow{p} \Sigma_F$, where Σ_F is a positive-definite matrix.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(v) $\sup_t \mathrm{E}\|F_t\| < \infty$.[4]

(vi) The eigenvalues of the matrix $\Sigma_{\Lambda^*}\Sigma_F$ are distinct.

Parts (i)–(iii) are of standard nature. Part (iv) is the weak law of large numbers. It holds under sufficient moment and other regularity conditions for $\{F_t\}$. Note that it allows for serial correlation in $\{F_t\}$ but no unit root nonstationarity. If $\{F_t\}$ is a vector linear process (cf. Phillips and Solo, 1992) or a mixing process (cf. Davidson, 1994), part (iv) will hold. Part (iv) is required for Lemmas B.2 and B.3 in Appendix B.

In addition, we make the following assumptions.

Assumption 3.

(i) For any t, N, and T, $\mathrm{E}\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T} F_s\left[e_s'\Omega^{-1}e_t - \mathrm{E}\left(e_s'\Omega^{-1}e_t\right)\right]\right\|^2$ is bounded.

(ii) For any N and T, $\mathrm{E}\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\Lambda'\Omega^{-1}e_s F_s'\right\|^2$ is bounded.

(iii) For each t and i, as $T, N \to \infty$,

$$\begin{pmatrix} \frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t \\ \frac{1}{\sqrt{T}}\sum_{s=1}^{T} F_s e_{is} \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \Sigma_{\Lambda^*} & 0 \\ 0 & \Phi_i \end{pmatrix}\right),$$

where $\Phi_i = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{E}\left(F_t F_s' e_{it} e_{is}\right)$.

Let F_{ks} denote the kth element of F_s. Part (i) requires that

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\mathrm{E}(F_{ks}F_{lu}\left[\frac{1}{N}e_s'\Omega^{-1}e_1 - \mathrm{E}\left(\frac{1}{N}e_s'\Omega^{-1}e_1\right)\right] \times \left[\frac{1}{N}e_u'\Omega^{-1}e_1 - \mathrm{E}\left(\frac{1}{N}e_u'\Omega^{-1}e_1\right)\right]$$

be bounded for any k and l. Here we put $t = 1$ without loss of generality. If $\left\{F_{ks}\left[\frac{1}{N}e_s'\Omega^{-1}e_1 - \mathrm{E}\left(\frac{1}{N}e_s'\Omega^{-1}e_1\right)\right]\right\}$ is, for example, a mixing process, this can be obtained under proper conditions (cf. Davidson, 1994, Ch. 19). Part (ii) is the same as requiring that

$$\frac{1}{N}\mathrm{tr}\left\{\Lambda'\Omega^{-1}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}(F_s'F_t e_s e_t')\Omega^{-1}\Lambda\right\}$$

be bounded. This will follow if $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}(F_{ks}F_{lt}e_{is}e_{jt})$ are bounded for every k, s, i, and j. Again, if $\{F_{ks}e_{is}\}$ is a mixing process, this can be obtained under proper conditions. Part (iii) states that joint asymptotic normality holds for $\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t$ and $\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is}$. It is assumed that the two vectors in part (iii) are

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

asymptotically uncorrelated. This holds if, for example, $\{e_t\}$ are uncorrelated over time, $\{F_t\}$ and $\{e_t\}$ are independent, and every element of $\mathrm{E}(F_t)$ is bounded for all t, since

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\mathrm{E}\left(\Lambda'\Omega^{-1}e_t\sum_{s=1}^{T}F_s'e_{is}\right)\\
&\quad=\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\Lambda'\Omega^{-1}\mathrm{E}(e_te_{is})\mathrm{E}\left(F_s'\right)\\
&\quad=\frac{1}{\sqrt{NT}}\Lambda'\Omega^{-1}\mathrm{E}(e_te_{it})\mathrm{E}\left(F_t'\right)\\
&\quad=\frac{1}{\sqrt{NT}}\Lambda'\Omega^{-1}\omega_i\mathrm{E}\left(F_t'\right)\\
&\quad=\frac{1}{\sqrt{NT}}\Lambda'\mathbf{i}\mathrm{E}\left(F_t'\right)\\
&\quad=\frac{1}{\sqrt{NT}}\lambda_i\mathrm{E}\left(F_t'\right)\to 0,
\end{aligned}$$

where $\boldsymbol{\omega}_i$ is the ith column of Ω and $\mathbf{i}$ is an $N\times 1$ vector with the ith element being equal to 1 and the rest zeros.

Asymptotic distributions of the GPCE of the factor space are reported in Theorem 1 below. It shows that $\hat{F}_t$ essentially estimates a rotation of the true, unobservable factors. For many practical applications, however, this is indeed what we need. Moreover, estimating the true factors is not possible without further identifying restrictions.

THEOREM 1. *Suppose that Assumptions 1, 2, and 3 hold.*

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

$$\sqrt{N}\left(\hat{F}_t-J'F_t\right)\xrightarrow{d}N\left(0,W^{-1}\right),$$

where $J=\frac{1}{NT}\left(\Lambda'\Omega^{-1}\Lambda\right)\left(F'\hat{F}\right)W_{NT}^{-1}$, W_{NT} *is an* $r\times r$ *diagonal matrix consisting of the first* r *largest eigenvalues of* $\frac{1}{NT}X\Omega^{-1}X'$ *in descending order, and* $W=diag[w_1,\ldots,w_r],\ldots w_1>>w_r>0$ *are the eigenvalues of* $\Sigma_{\Lambda^*}\Sigma_F$.

(ii) If $\liminf\frac{\sqrt{N}}{T}\geq\tau>0$, $T\left(\hat{F}_t-J'F_t\right)=O_p(1)$.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Some comments are in order for these results. First, $\hat{F}_t$ estimates a rotation of F_t, $J'F_t$, consistently. Note that the rotation matrix is different from that for the OPCE.

Second, the orders of convergence of the GPCE and OPCE of the factor space are the same, as this theorem and Theorem 1 of Bai (2003) show, though their distributions are different.

Third, it is hard to compare efficiency of the GPCE and OPCE because these are estimating different rotations of the factors. However, because $J \xrightarrow{p} R^{-1}$ (R is defined in Theorem 2) and $R'W^{-1}R = \Sigma_{\Lambda^*}^{-1}$, Theorem 1 can be rewritten as

$$\sqrt{N}\left(J'^{-1}\hat{F}_t - F_t\right) \xrightarrow{d} N\left(0, \Sigma_{\Lambda^*}^{-1}\right). \tag{5}$$

In contrast, using Proposition 1 and Theorem 1 of Bai (2003) and letting $\lim_{N\to\infty} \frac{\Lambda'\Lambda}{N} = \Sigma_\Lambda$, we obtain[5] for the OPCE $\hat{F}_{ot}$ and a matrix H defined in Bai (2003),

$$\sqrt{N}\left(H'^{-1}\hat{F}_{ot} - F_t\right) \xrightarrow{d} N\left(0, \Sigma_\Lambda^{-1}\Sigma_{\Lambda^*}\Sigma_\Lambda^{-1}\right), \tag{6}$$

once we assume $\mathrm{E}\left(e_t e_t'\right) = \Omega$ for every t. Interestingly, the asymptotic variances in relations (5) and (6) are similar to those of the GLS and ordinary least squares (OLS) estimators, respectively, in standard regression analysis once we regard Λ as a regressor matrix. This also means that $J'^{-1}\hat{F}_t$ is more efficient than $H'^{-1}\hat{F}_{ot}$, though these are infeasible estimators and F_t is still inestimable.

Fourth, when $r = 1$, the GPCE and OPCE are estimating the same object, and the efficiency gain of GPCE over OPCE can be established heuristically as follows. Since $J \xrightarrow{p} \frac{1}{\sqrt{\Sigma_F}}$ and $w_1 = \Sigma_{\Lambda^*}\Sigma_F$ (use Lemmas B.2 and B.3 in Appendix B for these), part (i) of Theorem 1 can be written as

$$\hat{F}_t \simeq \frac{F_t}{\sqrt{\Sigma_F}} + \frac{1}{\sqrt{N}} N\left(0, \frac{\Sigma_{\Lambda^*}^{-1}}{\Sigma_F}\right). \tag{7}$$

For the OPCE $\hat{F}_{ot}$, we obtain

$$\hat{F}_{ot} \simeq \frac{F_t}{\sqrt{\Sigma_F}} + \frac{1}{\sqrt{N}} N\left(0, \frac{1}{\Sigma_F}\Sigma_\Lambda^{-1}\Sigma_{\Lambda^*}\Sigma_\Lambda^{-1}\right). \tag{8}$$

Since the variance in (8) is larger than that in (7), we can heuristically argue that the GPCE is more efficient than the OPCE in estimating the common object $\frac{F_t}{\sqrt{\Sigma_F}}$. However, F_t still cannot be estimated because Σ_F cannot be identified in practice.

Asymptotic distributions of the GPCE of the factor-loading space are reported in the following theorem. Again, the GPCE estimates a rotation of the true, unobservable factor loadings.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

THEOREM 2. *Let $\hat{\lambda}_i$ be the ith column of $\hat{\Lambda}$. Suppose that Assumptions 1, 2, and 3 hold.*

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

$$\sqrt{T}\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \xrightarrow{d} N\left(0, R'^{-1}\Phi_i R^{-1}\right),$$

where $J = \frac{1}{NT}(\Lambda'\Omega^{-1}\Lambda)\left(F'\hat{F}\right)W_{NT}^{-1}$, W_{NT} *is an* $r \times r$ *diagonal matrix consisting of the first* r *largest eigenvalues of* $\frac{1}{NT}X\Omega^{-1}X'$ *in descending order,* $R = W^{1/2}\Theta'\Sigma_{\Lambda^*}^{-1/2}$, $W = diag[w_1, \ldots, w_r]$, $w_1 > \cdots > w_r > 0$ *are the eigenvalues of* $\Sigma_{\Lambda^*}\Sigma_F$, Θ *is the corresponding eigenvector matrix such that* $\Theta'\Theta = I_r$, *and* Φ_i *is the asymptotic variance-covariance matrix of* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}$.

(ii) If $\liminf \frac{\sqrt{T}}{N} \geq \tau > 0$, $N\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) = O_p(1)$.

This theorem shows that $\hat{\lambda}_i$ is a consistent estimator of a rotation of λ_i, $J^{-1}\lambda_i$ and that asymptotic normality holds if $\frac{\sqrt{T}}{N} \to 0$. The orders of convergence of the GPCE and OPCE of the space of factor loadings are the same.

Again, an efficiency comparison of GPCE and OPCE can be made as before. Using Theorem 2 above and Theorem 2 of Bai (2003), we obtain

$$\sqrt{T}\left(J\hat{\lambda}_i - \lambda_i\right),\ \sqrt{T}\left(H\hat{\lambda}_{oi} - \lambda_i\right) \xrightarrow{d} N(0, \Phi_i),$$

where H is defined in Bai (2003) and $\hat{\lambda}_{oi}$ denotes the OPCE of λ_i. Thus, the GPCE is not expected to have an efficiency advantage over the OPCE.

Indeed, for $r = 1$, a heuristic argument as before gives

$$\hat{\lambda}_i,\ \hat{\lambda}_{oi} \simeq \sqrt{\Sigma_F}\lambda_i + \frac{1}{\sqrt{T}}N\left(0, \frac{\Phi_i}{\Sigma_F}\right).$$

These results come as no surprise once we are reminded that OLS and GLS are equivalent in multivariate linear regression models.[6] The factor loadings are similar to regression coefficients of multivariate regression models, and this explains the asymptotic equivalence of the GPCE and the OPCE.

The serial correlation of $\{e_{it}\}$ affects the variance–covariance matrix in Theorem 2. This indicates that efficiency of the GPCE can be improved if the serial correlation structure is taken into account. The GPCE in this case is not the CMLE even under a normality condition. In related work, Stock and Watson (2005) impose an autoregressive structure on $\{e_{it}\}$ and consider the Cochrane-Orcutt-type transformation of the static factor model. This transformation results in an efficiency gain in the estimation of the factor-loading space.

Using Theorems 1 and 2, the following theorem obtains asymptotic distribution of the GPCE of the common component, $C_{it} = \lambda_i' F_t$. The GPCE of C_{it}, $\hat{C}_{it}$, is defined by $\hat{C}_{it} = \hat{\lambda}_i'\hat{F}_t$.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

THEOREM 3. *Suppose that Assumptions 1, 2, and 3 hold. In addition, assume that* $\{F_t\}$ *and* $\{e_t\}$ *are independent and that* $a_{NT} = \frac{\kappa_{NT}}{\sqrt{N}}$ *and* $b_{NT} = \frac{\kappa_{NT}}{\sqrt{T}}$ *converge to* a *and* b*, respectively, where* $\kappa_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$*. Then*

$$\frac{\hat{C}_{it} - C_{it}}{\sqrt{\frac{1}{N}Y_i + \frac{1}{T}Z_{it}}} \xrightarrow{d} N(0,1), \tag{9}$$

where $Y_i = \lambda_i' \Sigma_{\Lambda^*}^{-1} \lambda_i$ *and* $Z_{it} = F_t' \Sigma_F^{-1} \Phi_i \Sigma_F^{-1} F_t$.

This theorem shows that the approximate variance of $\hat{C}_{it} - C_{it}$, $\frac{1}{N}Y_i + \frac{1}{T}Z_{it}$, is smaller than that of the OPCE, $\frac{1}{N}\lambda_i' \Sigma_\Lambda^{-1} \Sigma_{\Lambda^*} \Sigma_\Lambda^{-1} \lambda_i + \frac{1}{T}Z_{it}$ (cf. Bai, 2003, Thm. 3). This reflects the efficiency gain made by the GPCE. The efficiency gain is being provided by the GPCE of the factor space, since Z_{it} appears commonly in the approximate variances of the GPCE and OPCE. Notice also that the asymptotic variances of the GPCE and OPCE can be reduced if a proper account is made of the serial correlation in the idiosyncratic errors for the estimation of the factor-loading space. But this usually requires assuming a particular serial correlation structure of the idiosyncratic errors.

4. FORECASTING REGRESSIONS

The main purpose of this section is to approximate the variance of the forecasting error based on the GPCE. Forecasting regressions using estimated factors originate from Stock and Watson (2002b) and have been used widely for prediction.

The forecasting equation we consider is

$$y_{t+h} = \alpha' F_t + \beta' Z_t + \varepsilon_{t+h}, \qquad (t = 1, \ldots, T-h), \tag{10}$$

where h is a forecasting horizon, $\{F_t\}$ are factors from model (1) and $\{Z_t\}$ are observable variables. Since $\{F_t\}$ are not observed, we instead use $\{\hat{F}_t\}$ for the purpose of forecasting. In order to forecast inflation rates, for example, Stock and Watson (2002b) use estimated factors based on more than 100 macroeconomic time series, and the lagged inflation rates as observable right-hand-side variables.

The regression equation with the estimated factors is written as

$$y_{t+h} = \alpha' J'^{-1} \hat{F}_t + \beta' Z_t + \varepsilon_{t+h} + \alpha' J'^{-1} \left(J' F_t - \hat{F}_t\right). \tag{11}$$

The OLS estimator of $\left(\alpha' J'^{-1} \ \beta'\right)'$ is obtained by regressing y_{t+h} on $\left(\hat{F}_t \ Z_t\right)$. The estimator from this regression is denoted as $\hat{\delta} = \left(\hat{\alpha}', \hat{\beta}'\right)'$. Notice that the parameter vector α is not identified in this regression equation.

We will study the asymptotic distribution of $\hat{\delta}$ in the following. To do this, we need an additional assumption.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Assumption 4. Let $L_t = (F_t', Z_t')'$.

(i) $\{F_t, Z_t, \varepsilon_t\}_{t=1}^T$ and $\{e_{it}\}_{t=1}^T$ are independent.

(ii) $\mathrm{E}\,\|L_t\|^4$ is bounded for every t.

(iii) $\frac{1}{T}\sum_{t=1}^T L_t L_t' \xrightarrow{p} \Sigma_L = \begin{bmatrix} \Sigma_F & \Sigma_{FZ} \\ \Sigma_{ZF} & \Sigma_Z \end{bmatrix} > 0.$

(iv) $\frac{1}{\sqrt{T}}\sum_{t=1}^T L_t \varepsilon_{t+h} \xrightarrow{d} N(0, \Sigma_{\varepsilon L})$, where $\Sigma_{\varepsilon L} = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \varepsilon_{t+h}^2 L_t L_t'$.

Part (i) implies that the idiosyncratic errors from the factor model and all the random variables in the forecasting regression equation are independent. Part (ii) is used in Theorem 4.[7] Parts (iii) and (iv) are standard assumptions used in regressions.

Asymptotic distributions of the forecasting regression's OLS estimators are reported in the following theorem.

THEOREM 4. *Let* $\delta = \left(\alpha' J^{-1}, \beta'\right)'$ *and suppose that Assumptions 1, 2, 3, and 4 hold. Then, if* $\frac{\sqrt{T}}{N} \to 0$,

$$\sqrt{T}\left(\hat{\delta} - \delta\right) \xrightarrow{d} N\left(0, \Psi'^{-1}\Sigma_L^{-1}\Sigma_{\varepsilon L}\Sigma_L^{-1}\Psi^{-1}\right),$$

where $\Psi = diag\left[R'^{-1}\ I\right]$ *and R is defined in Theorem 2.*

Since $\hat{\alpha}$ estimates a rotation of α, the true parameter α is not identified. This follows because $\hat{F}_t$ estimates a rotation of F_t, not F_t itself. However, β is consistently estimated. Moreover, the GPCE does not bring any efficiency gain over the OPCE for the forecasting regressions since Theorem 4 implies that

$$diag\left[J', I\right]\sqrt{T}\left(\hat{\delta} - \delta\right) \xrightarrow{d} N\left(0, \Sigma_L^{-1}\Sigma_{\varepsilon L}\Sigma_L^{-1}\right),$$

which is the same as the asymptotic distribution of the OLS estimator for the similarly standardized[8] OPCE-based forecasting regressions (cf. Bai and Ng, 2006, Thm. 1). This reflects the fact that α is not identified by either estimator and that there is no common object for efficiency comparison.

However, for the forecasting error, we do obtain an efficiency gain. To analyze this, let $\hat{L}_t = (\hat{F}_t'\ Z_t')'$ and $\hat{y}_{T+h|T} = \hat{\delta}'\hat{L}_T$, and consider the forecasting error

$$\hat{y}_{T+h|T} - y_{T+h} = \left(\hat{\delta} - \delta\right)'\hat{L}_T + \alpha' J'^{-1}\left(\hat{F}_T - J'F_T\right) - \varepsilon_{T+h}.$$

Approximate variance of the forecasting error is reported in the following theorem.

THEOREM 5. *Suppose that Assumptions 1, 2, 3, and 4 hold. In addition, assume* $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ *for all t*; $\mathrm{E}\left(\varepsilon_{t+h}\varepsilon_{T+h}|L_1, ..., L_{T-h}\right) = 0$ *for* $t = 1, \ldots, T-h$;

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

and h is a fixed, positive integer. Then, if $\frac{\sqrt{T}}{N} \to 0$, the approximate variance of the forecasting error, $\hat{y}_{T+h|T} - y_{T+h}$, is

$$\frac{1}{T} L_T' \Sigma_L^{-1} \Sigma_{\varepsilon L} \Sigma_L^{-1} L_T + \frac{1}{N} \alpha' \Sigma_{\Lambda^*}^{-1} \alpha + \sigma_\varepsilon^2. \tag{12}$$

Using the same method[9] as for Theorem 5, we obtain the approximate variance of the OPCE-based forecasting error,

$$\frac{1}{T} L_T' \Sigma_L^{-1} \Sigma_{\varepsilon L} \Sigma_L^{-1} L_T + \frac{1}{N} \alpha' \left[\Sigma_\Lambda^{-1} \Sigma_{\Lambda^*} \Sigma_\Lambda^{-1} \right] \alpha + \sigma_\varepsilon^2. \tag{13}$$

Relations (12) and (13) indicate that the GPCEs offer a smaller forecasting error variance. Again, the efficiency gain originates from the GPCEs' estimation of the factor space.

5. FEASIBLE GENERALIZED PRINCIPAL COMPONENT ESTIMATION

This section studies the FGPCE of the factor space. Because the GPCE of the factor-loading space does not lead to an efficiency gain in the limit as we have learned from Theorem 2, we will focus on the estimation of the factor space only.

Suppose that an estimator of Ω, denoted as $\hat{\Omega}$, is available. Specific restrictions given to the structure of Ω should also apply to $\hat{\Omega}$. The FGPCE of the factor space is obtained by maximizing $\text{tr}\left\{ F' \left(X \hat{\Omega}^{-1} X' \right) F \right\}$ with respect to F. The FGPCE resulting from this operation is denoted as $\hat{F}^f$. The standardization $\hat{F}^{f\prime} \hat{F}^f = T \times I_r$ is employed as before.

An immediate question one may have about this estimator is whether or not it has the same limiting distribution as the GPCE that uses the true variance-covariance matrix. In order to establish asymptotic equivalence of the GPCE and FGPCE, the following assumption is required.

Assumption 5.

(i) $\left\| \hat{\Omega}^{-1} \right\|_1 = O_p(1)$.

(ii) $\left\| \hat{\Omega} - \Omega \right\|_1 \xrightarrow{p} 0$.

(iii) $\frac{1}{\sqrt{N}} \Lambda' \left(\hat{\Omega}^{-1} - \Omega^{-1} \right) e_t \xrightarrow{p} 0$.

(iv) $\sup_t \| F_t \| = O_p(1)$.

Part (i) of this assumption is equivalent to stating that the smallest eigenvalue of $\hat{\Omega}$ is bounded below by a positive constant with probability approaching one as sample sizes increase. Since Ω is positive definite, this will eventually hold if elements of Ω can be estimated consistently in the uniform norm $\|\cdot\|$. Part (ii) states that the largest eigenvalue of $\hat{\Omega} - \Omega$ converges to zero in probability. Part (iii) is required for the asymptotic equivalence of $\frac{1}{\sqrt{N}} \Lambda' \hat{\Omega}^{-1} e_t$ and $\frac{1}{\sqrt{N}} \Lambda' \Omega^{-1} e_t$. Even when all the elements of $\hat{\Omega}^{-1} - \Omega^{-1}$ are $o_p(1)$, this condition is hard to

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

verify unless some parts of the matrix $\hat{\Omega}^{-1} - \Omega^{-1}$ take zero values. It is expected from this that we need some restrictions on the structure of the variance-covariance matrices Ω and $\hat{\Omega}$. Part (iv) is needed to prove Lemma B.6 in Appendix B, which will be used to prove Theorem 6 below. Assumption 5 can be verified when a special structure is given to the variance-covariance matrices Ω and $\hat{\Omega}$ as shown in the examples below.

The following theorem reports asymptotic equivalence of the GPCE and FGPCE.

THEOREM 6. *Suppose that assumptions for Theorem 1 and Assumption 5 hold. Then, the asymptotic distributions of* $\hat{F}_t - J'F_t$ *and* $\hat{F}_t^f - J^{f\prime}F_t$ *are equivalent, where* $J^f = \frac{1}{NT}(\Lambda'\hat{\Omega}^{-1}\Lambda)(F'\hat{F}^f)\hat{W}_{NT}^{-1}$ *with* $\hat{W}_{NT}$ *being an* $r \times r$ *diagonal matrix consisting of the first* r *largest eigenvalues of* $\frac{1}{NT}X\hat{\Omega}^{-1}X'$ *in descending order.*

Because $J - J^f \xrightarrow{p} 0$ (cf. Lemmas B.5 and B.6), this theorem shows that $\hat{F}_t^f$ estimates essentially the same object that $\hat{F}_t$ does and that they have the same limiting distribution. Thus, discussions after Theorem 1 about efficiency comparison between the GPCE and OPCE also apply to the FGPCE.

The following examples illustrate circumstances where asymptotic equivalence of the GPCE and FGPCE holds.

Example 1. Heteroskedasticity only
Let $\Omega = diag\left[\sigma_1^2, \ldots, \sigma_N^2\right]$ and $\hat{\Omega} = diag\left[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2\right]$ where σ_i^2 are positive constants. Assume

(i) $\max_{1\le i\le N}\left\{\left|\hat{\sigma}_i^2 - \sigma_i^2\right|\right\} \xrightarrow{p} 0$,

(ii) $\max_{1\le i\le N}\left\{\left|\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right|\right\} \xrightarrow{p} 0$,

(iii) $\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i\left(\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right)e_{it} \xrightarrow{p} 0$.

Then,

$$\left\|\hat{\Omega} - \Omega\right\|_1 = \max_{1\le i\le N}\left\{\left|\hat{\sigma}_i^2 - \sigma_i^2\right|\right\} \xrightarrow{p} 0$$

and

$$\left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_1 = \max_{1\le i\le N}\left\{\left|\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right|\right\} \xrightarrow{p} 0.$$

This provides $\left|\left\|\hat{\Omega}^{-1}\right\|_1 - \left\|\Omega^{-1}\right\|_1\right| \xrightarrow{p} 0$, from which it is deduced that $\left\|\hat{\Omega}^{-1}\right\|_1 = O_p(1)$. Last, the third part of Assumption 5 is also satisfied by the given assumptions. Thus, we have asymptotic equivalence of the GPCE and FGPCE for the factor space estimation. In practice, we may use the residuals from the ordinary principal component estimation in order to estimate σ_i^2.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Example 2 (Heteroskedasticity and cross-sectional correlation).
Let $N = na$ where $n \to \infty$ and a is a fixed positive constant greater than 1. The variance-covariance matrix Ω has a block-diagonal structure such that $\Omega = \Omega_1 \oplus \Omega_2 \ldots \oplus \Omega_n$. Here each Ω_k is an $a \times a$, positive definite matrix. Assume that there are matrices $\hat{\Omega}_k$ such that

$$\text{(i)}\ \max_{1\le k\le n}\left\{\left\|\hat{\Omega}_k - \Omega_k\right\|_1\right\} \xrightarrow{p} 0,$$

$$\text{(ii)}\ \max_{1\le k\le n}\left\{\left\|\hat{\Omega}_k^{-1} - \Omega_k^{-1}\right\|_1\right\} \xrightarrow{p} 0,$$

$$\text{(iii)}\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Lambda_k\left(\hat{\Omega}_k^{-1} - \Omega_k^{-1}\right)\mathbf{e}_{kt} \xrightarrow{p} 0$$

$$\text{where } \Lambda = \begin{bmatrix}\Lambda_1' \\ \vdots \\ \Lambda_n'\end{bmatrix} \text{ and } e_t = \begin{bmatrix}\mathbf{e}_{1t} \\ \vdots \\ \mathbf{e}_{nt}\end{bmatrix}.$$

Then,

$$\left\|\hat{\Omega} - \Omega\right\|_1 = \max_{1\le k\le n}\left\{\left\|\hat{\Omega}_k - \Omega_k\right\|_1\right\} \xrightarrow{p} 0,$$

and

$$\left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_1 = \max_{1\le k\le n}\left\{\left\|\hat{\Omega}_k^{-1} - \Omega_k^{-1}\right\|_1\right\} \xrightarrow{p} 0.$$

The third part of Assumption 5 is also satisfied. In practice, $\hat{\Omega}_k$ can be constructed as in the previous example by using the residuals from the ordinary principal component estimation.

These examples require prior knowledge on the structure of the variance-covariance matrix Ω. Using a general structure for Ω without any restriction on its elements is not allowed for the following reasons. First, the OPCE requires a restriction on the elements of Ω in order to estimate the idiosyncratic errors consistently (cf. Bai, 2003, Assum. C3). This restriction states that the absolute sum of all the elements of Ω should grow at the rate of N. Because the FGPCE uses the OPCE as an initial estimator, restrictions required for the OPCE are also required for the FGPCE. Second, part (iii) of Assumption 5 is difficult to verify when the absolute sum of all the elements of Ω grows at the rate of N^2. Therefore, some restrictions on Ω, as given in Examples 1 and 2, are inevitable.

Last, we study whether the FGPCE is as efficient as the GPCE in estimating forecasting errors. The following theorem shows that they are asymptotically equivalent in terms of efficiency.

THEOREM 7. *Suppose that assumptions for Theorem 5 and Assumption 5 hold. Then, the approximate variance of the forecasting error that is based on the FGPCE is the same as that in Theorem 5.*

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

6. SIMULATION

6.1. Estimation Efficiency

This section reports simulation results that illustrate the estimation efficiency of the OPCE, GPCE, and FGPCE. We are focused on the estimation of the common component, C_{it}, because a GPCE efficiency gain in the estimation of the common components has theoretical support, as discussed in Section 3. The FGPCE is computed using the OPCE-based estimator of the variance–covariance matrix Ω. More precisely, the following steps were taken.

(i) Calculate the OPCEs and the estimates of the idiosyncratic errors $\hat{e}_{it} = X_{it} - \hat{\lambda}'_{oi}\hat{F}_{ot}$.

(ii) Calculate the estimates of the elements of Ω as $\hat{\omega}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}\hat{e}_{jt}$.

(iii) Use $\hat{\Omega} = (\hat{\omega}_{ij})_{i,j=1,\ldots,N}$ for the calculation of the FGPCE.

In our simulation, particular structures of Ω are assumed, as outlined in the next paragraph, and these are assumed to be known. This structure is also imposed on $\hat{\Omega}$. Note that the sum of the elements of every row of matrix Ω should be bounded for the consistency of the OPCE-based estimates of the common components. While the GPCEs do not require this condition, the FGPCEs do, as they use the OPCEs as initial estimates.

Data were generated by model (1) where $F_t \sim iid\ N(0, I_r)$ and Λ_i is a fixed matrix with entries generated according to the law of standard normal distribution. For sample sizes, we considered $N = 50$ and $T = 50,\ 100$.[10] For $\{e_{it}\}$, we considered three cases: (i) cross-sectional heteroskedasticity with neither serial nor cross-sectional correlation, (ii) cross-sectional heteroskedasticity with serial correlation only, and (iii) cross-sectional heteroskedasticity with weak cross-sectional correlation modeled by a block-diagonal structure of the variance–covariance matrix Ω. The blocks in case (iii) are 5×5 positive-definite matrices. The factors and idiosyncratic errors were generated to be independent.

In Table 1 we report empirical mean squared errors (MSEs) of the OPCE, GPCE, and FGPCE for each C_{it}. We used 2,000 iterations to calculate the empirical MSEs. We report MSEs, rather than variances because biases may also matter in finite samples in estimating the common components. These MSEs take the small-sample biases into account, but the biases are quite negligible relative to the variances in all the cases. In calculating the MSEs, we assume that the structure of the variance–covariance matrix Ω is known.

The results in Table 1 are summarized as follows.

(i) The GPCE and FGPCE significantly improve MSEs relative to the OPCE. In some cases, the MSEs of the GPCE and FGPCE are less than half of those of the OPCE.

(ii) The GPCE and FGPCE show similar performance in most cases, but the GPCE performs better than FGPCE in all cases, as is well expected.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

TABLE 1. Efficiency Comparison for the Estimation of Common Components

		I			II			III		
		OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE
$r=1$	$T=50$	0.0108	0.0060	0.0063	0.0107	0.0065	0.0067	0.0184	0.0101	0.0108
	$T=100$	0.0103	0.0036	0.0042	0.0092	0.0031	0.0036	0.0133	0.0049	0.0054
$r=3$	$T=50$	0.0416	0.0233	0.0254	0.0398	0.0229	0.0252	0.0669	0.0313	0.0357
	$T=100$	0.0319	0.0125	0.0151	0.0336	0.0163	0.0179	0.0571	0.0152	0.0185
$r=5$	$T=50$	0.0755	0.0442	0.0502	0.0486	0.0282	0.0310	0.1034	0.0504	0.0583
	$T=100$	0.0535	0.0194	0.0239	0.0478	0.0207	0.0252	0.0955	0.0265	0.0349
$r=8$	$T=50$	0.1043	0.0657	0.0754	0.1072	0.0738	0.0792	0.1612	0.0831	0.0970
	$T=100$	0.0872	0.0468	0.0580	0.0775	0.0415	0.0479	0.1148	0.0400	0.0513

Notes: Data were generated by $X_{it} = \lambda_i' F_t + e_{it}$, $(i = 1, \ldots, N;\ t = 1, \ldots, T)$, $N = 50$; r denotes the number of factors and is assumed to be known; $\{e_{it}\}$ are serially and cross-sectionally uncorrelated, have common variances across t but have different variances across i; $F_t \sim iid\ N(0, I_r)$; λ_i take fixed values throughout iterations with their values generated according to the law of standard normal distribution. Entries in this table denote empirical mean squared errors of the OPCE, GPCE, and FGPCE of the common components using 2,000 iterations. In estimating the variance-covariance matrix of $\{e_{it}\}$, it is assumed that we know the structure of the variance–covariance matrix Ω. I refers to the case of cross-sectional heteroskedasticity, II to that of cross-sectional heteroskedasticity and serial correlation, and III to that of cross-sectional heteroskedasticity and weak cross-sectional correlation.

(iii) Serial correlation present in part II does not bring much difference to the MSEs.

(iv) Cross-sectional dependence incorporated into part III introduces higher MSEs for all of the estimators.

(v) As the number of factors increases, the performance of all estimators weakens.

(vi) The performance of all estimators improves as T increases.

6.2. Mean Squared Forecasting Errors

This section reports simulation results that compare mean squared forecasting errors based on the OPCE, GPCE, and FGPCE estimates of the factor space. The FGPCEs were computed as in the last section. Data were generated by

$$y_{t+1} = \alpha' F_t + \beta y_t + \varepsilon_{t+1}, \qquad (t = 1, \ldots, T),$$

where $\alpha = [1, \ldots, 1]'$, $\beta = 0.1$,[11] $F_t = 0.1 \times I_r F_{t-1} + v_t$, $v_t \sim iid\ N(0, I_r)$, and $\varepsilon_t \sim iid\ N(0, 1)$. This is a special case of the factor augmented vector autoregressive model popularized by Bernanke et al. (2005). For sample sizes, we considered $N = 25$ and $T = 50,\ 100$.[12] The first $T - 1$ observations are used to forecast the last observation, y_{T+1}. The unobserved factors F_t are contained in the observed data X_{it} generated by $X_{it} = \lambda_i' F_t + e_{it}$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$. In this data generation, λ_i take fixed values throughout the iterations, with their values generated according to the law of standard normal distribution as in the last

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

section. Furthermore, $\{e_{it}\}$, $\{v_t\}$, and $\{\varepsilon_t\}$ were generated to be independent. For $\{e_{it}\}$, we considered the same three cases as in the last section.

Table 2 reports empirical mean squared forecasting errors for y_{T+1} (i.e., the empirical counterpart of $\mathrm{E}(\hat{y}_{T+1|T} - y_{T+1})^2$) using the OPCE, GPCE, and FGPCE estimates of the factor space. The mean squared forecasting errors are the same as the variances of the forecasting errors up to the third decimal point in most cases. We performed 2,000 iterations for the calculations of empirical mean squared forecasting errors. We assume that the structure of the variance–covariance matrix Ω is known in these calculations.

The results in Table 2 are summarized as follows.

(i) Without any exception, the GPCE and FGPCE improve empirical mean squared forecasting errors relative to the OPCE.

(ii) The GPCE and FGPCE show similar performance in most cases, but the GPCE performs better than FGPCE in all the cases.

(iii) Neither serial correlation (part II) nor cross-sectional correlation (part III) brings any noticeable changes in the mean squared forecasting errors.

(iv) The mean squared forecasting errors improve as T increases and r decreases.

TABLE 2. Mean Squared Errors of Forecasting

		I			II			III		
		OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE
$r=1$	$T=50$	1.065	1.065	1.064	1.029	1.025	1.025	1.021	1.020	1.019
	$T=100$	1.014	1.001	1.003	1.065	1.060	1.060	1.065	1.048	1.049
$r=3$	$T=50$	1.083	1.069	1.070	1.166	1.145	1.146	1.118	1.065	1.070
	$T=100$	1.077	1.062	1.064	1.082	1.045	1.053	1.113	1.056	1.060
$r=5$	$T=50$	1.159	1.109	1.123	1.251	1.217	1.223	1.215	1.162	1.168
	$T=100$	1.116	1.077	1.082	1.133	1.095	1.096	1.100	1.020	1.033
$r=8$	$T=50$	1.315	1.238	1.249	1.355	1.254	1.277	1.391	1.224	1.267
	$T=100$	1.151	1.132	1.131	1.203	1.110	1.125	1.143	1.053	1.053

Notes: Data were generated by $y_{t+1} = \alpha' F_t + \beta y_t + \varepsilon_{t+1}$, $\alpha = [1, \ldots, 1]'$, $\beta = 0.1$, $(t = 1, \ldots, T-1)$, $X_{it} = \lambda_i' F_t + e_{it}$, and $F_t = 0.1 \times I_r F_{t-1} + v_t$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$; r denotes the number of factors and is assumed to be known; $\{e_{it}\}$ are serially and cross-sectionally uncorrelated, have common variances across t but have different variances across i; $v_t \sim iidN(0, I_r)$ and $\varepsilon_t \sim iid\ N(0, 1)$; $\{e_{it}\}, \{v_t\}$, and $\{\varepsilon_t\}$ are independent; λ_i take fixed values throughout iterations with their values generated according to the law of standard normal distribution. Entries in this table denote empirical mean squared errors of forecasting for y_{T+1} using the OPCE, GPCE, and FGPCE estimates of the factor spaces. The number of iteration is 2,000. In estimating the variance-covariance matrix of $\{e_{it}\}$, it is assumed that we know the structure of the variance–covariance matrix Ω. I refers to the case of cross-sectional heteroskedasticity, II to that of cross-sectional heteroskedasticity and serial correlation, and III to that of cross-sectional heteroskedasticity and weak cross-sectional correlation.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

6.3. Summary of Simulation Results

Tables 1 and 2 show that the GPCE and FGPCE improve on the OPCE in terms of estimation and forecasting efficiencies in finite samples. These results are consistent with the asymptotic theories in previous sections. However, the GPCE and FGPCE are based on the assumption that the idiosyncratic errors are stationary. When this assumption is violated, the OPCE may perform better in finite samples. Moreover, the known structure of Ω is used to calculate the GPCE and FGPCE. If the structure is not known, the OPCE, which does not require such knowledge, may perform better. In a nutshell, the GPCE and FGPCE bring an efficiency gain over the OPCE under the given assumptions and structures, but the OPCE may be more robust to misspecification. Notably, this is reminiscent of the usual comparison of the GLS and OLS estimators.

7. CONCLUSION AND FURTHER REMARKS

We have derived the asymptotic distributions of the GPCEs of the factor and factor-loading spaces by sending N and T to infinity simultaneously. It was shown that the variance of the GPCE of the common component was smaller than that of the OPCE. Approximate variance of the forecasting error using the factor estimates based on the GPCE was also derived and shown to be smaller than that based on the OPCE. The feasible GPCE of the factor space was shown to enjoy the same level of efficiency as the GPCE under some additional conditions. Simulation results confirmed efficiency gains of the GPCE and FGPCE over the OPCE in finite samples.

NOTES

1. See equation (17) of their paper.
2. Once the transformations $X \to X\Omega^{-1/2}$ and $\Lambda \to \Omega^{-1/2}\Lambda$ are made, the CMLE of Λ and F follow from a standard theory of principal component estimation.
3. One of the referees pointed out that the limit of a sequence of invertible matrices can be noninvertible. This assumption precludes such cases.
4. This assumption is required to prove Lemma A.2(a) of Bai (2003), which is used for the proof of Theorem 1 of this paper.
5. Note that $H \xrightarrow{p} Q^{-1}$, where H and Q are as defined in Bai (2003).
6. In the multivariate linear regression model $Y_t = AX_t + U_t$ $(t = 1, \ldots, T)$ with $U_t \sim iid(0, \Omega)$, the OLS estimator of A is $\left(\sum_{t=1}^{T} Y_t X_t\right)\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}$. For the GLS estimation, we transform the model such that $\Omega^{-1/2}Y_t = \Omega^{-1/2}AX_t + \Omega^{-1/2}U_t$. The GLS estimator is the OLS estimator using this transformed model. However, the GLS estimator is the same as the OLS estimator from the original model. This should be distinguished from the equivalence of GLS and OLS studied by Kruskal (1968), Zyskind (1967), and others.
7. To be more precise, it is required for Lemma A.1 of Bai and Ng (2002), which we will adapt for our use.
8. Using $diag\left[H', I\right]$ for the standardization.
9. Replace J with H and use Theorem 1 of Bai (2003).
10. Unreported results for $N = 25, 100$ are similar to those for $N = 50$.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

11. Different values of β were also tried, but gave no noticeable differences.
12. Unreported results for $N = 25, 100$ are similar to those for $N = 50$.

REFERENCES

Artis, M., A. Banerjee, & M. Marcellino (2005) Factor forecasts for the U.K. *Journal of Forecasting* 24, 279–298.

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–172.

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

Bai, J. & S. Ng (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127–1177.

Bai, J. & S. Ng (2006) Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74, 1133–1150.

Bai, J. & S. Ng (2010) Instrumental variable estimation in a data rich environment. *Econometric Theory* 26, 1577–1606.

Basilevsky, A. (1994) *Statistical Factor Analysis and Related Methods.* Wiley.

Bernanke, B., J. Boivin, & P. Eliasz (2005) Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.

Boivin, J. & S. Ng (2006) Are more data always better for factor analysis? *Journal of Econometrics* 132, 169–194.

Breitung, J. & J. Tenhofen (2007) GLS Estimation of Dynamic Factor Models. Mimeo, University of Bonn.

Brillinger, D.R. (1981) *Time Series. Data Analysis and Theory*. Holden Day.

Chamberlain, G. & M. Rothschild (1983) Arbitrage, factor structure, and mean variance analysis on large asset markets. *Econometrica* 51, 1281–1304.

Connor, G. & R. Korajzcyk (1986) Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15, 373–394.

Davidson, J. (1994) *Stochastic Limit Theory.* Cambridge University Press.

Favero, C., M. Marcellino, & F. Neglia (2005) Principal components at work: The empirical analysis of monetary policy with large data sets. *Journal of Applied Econometrics* 20, 603–620.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000) The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics* 80, 540–554.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2003) Do financial variables help in forecasting inflation and real activity in the euro area? *Journal of Monetary Economics* 50, 1243–1255.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2005) The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830–840.

Huang, H., T.-H. Lee, & C. Li (2006) Forecasting Output Growth and Inflation: How to Use Information in the Yield Curve. Mimeo, University of California, Riverside.

Jones, C.S. (2001) Extracting factors from heteroskedastic asset returns. *Journal of Financial Economics* 62, 293–325.

Kapetanios, G. & M. Marcellino (2006) A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions. Mimeo, Queen Mary, University of London.

Kruskal, W. (1968) When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics* 39, 70–75.

Ludvigson, S.C. & S. Ng (2007) The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics* 83, 171–222.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus.* Wiley.

Moon, H.R. & B. Perron (2004) Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122, 81–126.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Phillips, P.C.B. & D. Sul (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6, 217–259.

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

Stock, J.H. & M.W. Watson (2002a) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

Stock, J.H. & M.W. Watson (2002b) Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20, 147–162.

Stock, J.H. & M.W. Watson (2005) Implications of Dynamic Factor Model for VAR Analysis. NBER Working paper no. W11467.

Stock, J.H. & M.W. Watson (2006) Forecasting with many predictors. In G. Elliott, C.W.J. Granger, & A. Timmerman (eds.), *Handbook of Economic Forecasting*, pp. 515–554, Elsevier.

Zyskind, G. (1967) On canonical forms, non-negative covariance matrices and best and simple least squares estimators in linear models. *Annals of Mathematical Statistics* 38, 1092–1109.

APPENDIX A: Proofs of Main Results

Proof of Theorem 1. (i) Let W_{NT} be an $r \times r$ diagonal matrix consisting of the r largest eigenvalues of $\frac{1}{NT} X\Omega^{-1}X'$ in descending order. It satisfies the relation $\hat{F} = \frac{1}{NT} X\Omega^{-1}X'\hat{F}W_{NT}^{-1}$ by the definitions of eigenvalue and eigenvector. In addition, let $J = \frac{1}{NT}(\Lambda'\Omega^{-1}\Lambda)(F'\hat{F})W_{NT}^{-1}$. Then,

$$\begin{aligned}\hat{F} - FJ &= \frac{1}{NT}\left(X\Omega^{-1}X'\right)\hat{F}W_{NT}^{-1} - \frac{1}{NT}F\left(\Lambda'\Omega^{-1}\Lambda\right)\left(F'\hat{F}\right)W_{NT}^{-1} \\ &= \frac{1}{NT}\left[X\Omega^{-1}X' - F(\Lambda'\Omega^{-1}\Lambda)F'\right]\hat{F}W_{NT}^{-1} \\ &= \frac{1}{NT}\left(e\Omega^{-1}e' + e\Omega^{-1}\Lambda F' + F\Lambda'\Omega^{-1}e'\right)\hat{F}W_{NT}^{-1}.\end{aligned}$$

Note that this relation originates from Bai and Ng (2002) and Bai (2003). In vector notation, this becomes

$$\begin{aligned}\hat{F}_t - J'F_t &= \frac{1}{NT}W_{NT}^{-1}\hat{F}'\left(e\Omega^{-1}e_t + F\Lambda'\Omega^{-1}e_t + e\Omega^{-1}\Lambda F_t\right) \\ &= W_{NT}^{-1}\left(\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t + \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t\right. \\ &\qquad \left. + \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}\Lambda F_t\right) \\ &= W_{NT}^{-1}\left(a_{NTt} + b_{NTt} + c_{NTt}\right), \quad \text{say,} \end{aligned} \tag{A.1}$$

where A_s' is the sth row of the matrix A. For each fixed t,

$$a_{NTt} = O_p\left(\frac{1}{\kappa_{NT}^2}\right); \quad b_{NTt} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad c_{NTt} = O_p\left(\frac{1}{\sqrt{N}\kappa_{NT}}\right), \tag{A.2}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

where $\kappa_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. These can be proven in the same way as for Lemma A.2. of Bai (2003) once we use $e_t^* = \Omega^{-1/2} e_t$ and $\Lambda^* = \Omega^{-1/2}\Lambda^o$ in Bai's proof instead of e_t and Λ^o there, respectively. Details are not worth reporting here. Thus, when $\frac{\sqrt{N}}{T} \to 0$, $\sqrt{N}(\hat{F}_t - J'F_t) = W_{NT}^{-1}\sqrt{N} b_{NTt} + o_p(1)$. Because $\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}} \xrightarrow{d} N(0, \Sigma_{\Lambda^*})$ by Assumption 3, we have $\sqrt{N} b_{NTt} \xrightarrow{d} N(0, W)$ due to Lemma B.3. The result follows from this and Lemma B.2.

(ii) Since $\limsup \frac{T}{\sqrt{N}} \le \frac{1}{\tau} < \infty$, $T a_{NTt} = \begin{cases} O_p(1) \text{ if } T < N \\ o_p(1) \text{ if } N < T \end{cases}$, $T b_{NTt} = O_p(1)$, and $T c_{NTt} = o_p(1)$. The stated result follows from these. ■

Proof of Theorem 2. Since $\hat{\Lambda} = \frac{X'\hat{F}}{T}$, $\hat{\lambda}_i = \frac{\hat{F}F\lambda_i}{T} + \frac{\hat{F}\underline{e}_i}{T}$ where $\underline{e}_i = [e_{i1}, \ldots, e_{iT}]'$. Writing $F = F - \hat{F}J^{-1} + \hat{F}J^{-1}$ and using $\frac{\hat{F}\hat{F}}{T} = I_r$, we obtain

$$\begin{aligned} \hat{\lambda}_i &= \frac{1}{T}\hat{F}_t'\left(F - \hat{F}J^{-1} + \hat{F}J^{-1}\right)\lambda_i + \frac{1}{T}\hat{F}'\underline{e}_i \qquad \textbf{(A.3)} \\ &= J^{-1}\lambda_i + \frac{J'F'\underline{e}_i}{T} - \frac{1}{T}\hat{F}'\left(\hat{F} - FJ\right)J^{-1}\lambda_i + \frac{1}{T}\left(\hat{F} - FJ\right)'\underline{e}_i. \end{aligned}$$

Lemma B.4, below, shows that both the third and fourth terms are $O_p\left(\frac{1}{\kappa_{NT}^2}\right)$. Thus, if $\frac{\sqrt{T}}{N} \to 0$,

$$\begin{aligned} \sqrt{T}\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) &= J'\frac{1}{\sqrt{T}}\sum_{s=1}^{T} F_s e_{is} + o_p(1) \\ &\xrightarrow{d} N\left(0, W^{-1}R\Sigma_{\Lambda^*}\Phi_i\Sigma_{\Lambda^*}R'W^{-1}\right) \\ &= N\left(0, R'^{-1}\Phi_i R^{-1}\right), \end{aligned}$$

since $J \xrightarrow{p} \Sigma_{\Lambda^*}R'W^{-1}$. If $\liminf \frac{\sqrt{T}}{N} \ge \tau > 0$, $\limsup \frac{N}{\sqrt{T}} \le 1/\tau < \infty$. So, we obtain, again by using Lemma B.4,

$$\begin{aligned} N\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) &= \frac{N}{\sqrt{T}}J'\frac{1}{\sqrt{T}}\sum_{s=1}^{T} F_s e_{is} + O_p(1) \\ &= O_p(1). \end{aligned}$$

■

Proof of Theorem 3. We have

$$\begin{aligned} \hat{C}_{it} - C_{it} &= \hat{F}_t'\hat{\lambda}_i - F_t'\lambda_i \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + \hat{F}_t'\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + F_t'J\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) + \left(\hat{F}_t - J'F_t\right)'\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + F_t'J\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) + O_p\left(\frac{1}{\kappa_{NT}^2}\right), \qquad \textbf{(A.4)} \end{aligned}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

where Theorems 1 and 2 are used for the last equality. Using relations (A.1) and (A.2), we obtain

$$\kappa_{NT}\left(\hat{F}_t - J'F_t\right) \tag{A.5}$$
$$= a_{NT}W_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s F_s'\right)\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right) + O_p\left(\frac{1}{\kappa_{NT}}\right).$$

In addition, relation (A.3) and Lemma B.4 yield

$$\kappa_{NT}\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) = b_{NT}J'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is} + O_p\left(\frac{1}{\kappa_{NT}}\right). \tag{A.6}$$

Using relations (A.5) and (A.6) for (A.4), we obtain the first equality of the following relation:

$$\begin{aligned}\kappa_{NT}\left(\hat{C}_{it} - C_{it}\right) &= a_{NT}\lambda_i' J'^{-1}W_{NT}^{-1}\left(\frac{\hat{F}'F}{T}\right)\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right)\\ &\quad + b_{NT}F_t'JJ'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is} + O_p\left(\frac{1}{\kappa_{NT}}\right)\\ &= a_{NT}\lambda_i'\left(\frac{\Lambda'\Omega^{-1}\Lambda}{N}\right)^{-1}\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right)\\ &\quad + b_{NT}F_t'\Sigma_F^{-1}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is} + O_p\left(\frac{1}{\kappa_{NT}}\right)\\ &= a_{NT}c_{iN}\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right) + b_{NT}d_{tT}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is} + O_p\left(\frac{1}{\kappa_{NT}}\right).\end{aligned} \tag{A.7}$$

For the second equality, relations $J'^{-1}W_{NT}^{-1}\left(\frac{\hat{F}'F}{T}\right) = \left(\frac{\Lambda'\Omega^{-1}\Lambda}{N}\right)^{-1}$ and $JJ' \xrightarrow{p} \Sigma_F^{-1}$ are used. Since $a_{NT} \to a$, $b_{NT} \to b$, and $\{F_t\}$ and $\{e_t\}$ are independent, Assumptions 2 and 3(iii) yield, conditional on F,

$$a_{NT}c_{iN}\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right) + \left(b_{NT}d_{tT}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is}\right) \xrightarrow{d} N(0, a^2Y_i + b^2Z_{it}),$$

or, equivalently,

$$\frac{a_{NT}c_{iN}\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\right) + b_{NT}d_{tT}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_s e_{is}}{\sqrt{a^2Y_i + b^2Z_{it}}} \xrightarrow{d} N(0,1).$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Note that the limiting distribution in this relation does not depend on F any more. Since $\sqrt{a^2 Y_i + b^2 Z_{it}} - \sqrt{a_{NT}^2 Y_i + b_{NT}^2 Z_{it}} \xrightarrow{p} 0$, we may also write

$$\frac{a_{NT} c_{iN} \left(\frac{\Lambda' \Omega^{-1} e_t}{\sqrt{N}} \right) + b_{NT} d_{tT} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} F_s e_{is}}{\sqrt{a_{NT}^2 Y_i + b_{NT}^2 Z_{it}}} \xrightarrow{d} N(0,1),$$

which with relation (A.7) implies the stated result. ■

Proof of Theorem 4. Let $\hat{L}_t = \left(\hat{F}_t' \; Z_t' \right)'$ and rewrite model (11) as

$$y_{t+h} = \hat{L}_t' \delta + \varepsilon_{t+h} + \alpha' J'^{-1} \left(J' F_t - \hat{F}_t \right).$$

Then, the OLS estimator of δ is written as

$$\begin{aligned}
\sqrt{T} \left(\hat{\delta} - \delta \right) &= \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{L}_t \varepsilon_{t+h} \\
&\quad + \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{L}_t \left(J' F_t - \hat{F}_t \right)' J^{-1} \alpha \\
&= \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{L}_t \varepsilon_{t+h} + o_p(1).
\end{aligned}$$

The second equality is obtained by adapting Lemma A.1 of Bai and Ng (2006) along with relation (A.2) and the assumption $\frac{\sqrt{T}}{N} \to 0$. Moreover, as in Bai and Ng (2006, pp. 1145–1146),

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{L}_t \varepsilon_{t+h} &= \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{F}_t \varepsilon_{t+h} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} Z_t \varepsilon_{t+h} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sqrt{T}} J' \sum_{t=1}^{T-h} F_t \varepsilon_{t+h} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} Z_t \varepsilon_{t+h} \end{bmatrix} + o_p(1) \\
&\xrightarrow{d} N \left(0, \Psi \Sigma_{\varepsilon L} \Psi' \right),
\end{aligned}$$

where $\Psi = diag \left[R'^{-1} \; I \right]$, and $\frac{1}{T} \sum_{t=1}^{T-h} \hat{L}_t \hat{L}_t' \xrightarrow{p} \Psi \Sigma_L \Psi'$. Note that $J \xrightarrow{p} R^{-1}$. The result follows from these. ■

Proof of Theorem 5. Using (11), we obtain

$$\begin{aligned}
\hat{y}_{T+h|T} - y_{T+h} &= \left(\hat{\delta} - \delta \right)' \hat{L}_T + \alpha' J'^{-1} \left(\hat{F}_T - J' F_T \right) - \varepsilon_{T+h} \\
&= \sqrt{T} \left(\Psi' \left(\hat{\delta} - \delta \right) \right)' \left(\frac{1}{\sqrt{T}} \Psi^{-1} \hat{L}_T \right) \qquad \textbf{(A.8)} \\
&\quad + \frac{1}{\sqrt{N}} \alpha' J'^{-1} \left(\sqrt{N} \left(\hat{F}_T - J' F_T \right) \right) - \varepsilon_{T+h}.
\end{aligned}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Theorem 4 gives $\sqrt{T}\Psi'\left(\hat{\delta}-\delta\right) \xrightarrow{d} N\left(0, \Sigma_L^{-1}\Sigma_{\varepsilon L}\Sigma_L^{-1}\right)$. Since $\hat{L}_T = \begin{bmatrix} J'F_T + O_p\left(\frac{1}{\sqrt{N}}\right) \\ Z_T \end{bmatrix}$ and $J \xrightarrow{p} R^{-1}$, we also have

$$\Psi^{-1}\hat{L}_T = \begin{bmatrix} F_T + o_p(1) \\ Z_T \end{bmatrix}. \tag{A.9}$$

Thus, the first term in (A.8) has an approximate variance $\frac{1}{T}L_T'\Sigma_L^{-1}\Sigma_{\varepsilon L}\Sigma_L^{-1}L_T$. Moreover, since $J'^{-1}\left(\sqrt{N}\left(\hat{F}_T - J'F_T\right)\right) \xrightarrow{d} N\left(0, \Sigma_{\Lambda^*}^{-1}\right)$, the second term has an approximate variance $\frac{1}{N}\alpha'\Sigma_{\Lambda^*}^{-1}\alpha$. The first two terms are asymptotically uncorrelated, as deduced from relations (A.1), (A.2), and Assumption 4. The last term is asymptotically uncorrelated with the first term because it is so with $\sqrt{T}\Psi'\left(\hat{\delta}-\delta\right)$ due to the given assumption $\mathrm{E}\left(\varepsilon_{t+h}\varepsilon_{T+h} \mid L_1, \ldots L_{T-h}\right) = 0$ and because of the relation (A.9). It is also asymptotically uncorrelated with the second term again due to relations (A.1), (A.2), and Assumption 4. Thus, the result follows. ■

Proof of Theorem 6. We may write, as in the proof of Theorem 1,

$$\begin{aligned}
\hat{F}_t^f - J^{f\prime}F_t &= \frac{1}{NT}\hat{W}_{NT}^{-1}\hat{F}^f\left(e\hat{\Omega}^{-1}e_t + F\Lambda'\hat{\Omega}^{-1}e_t + e\Omega^{-1}\Lambda F_t\right) \\
&= \hat{W}_{NT}^{-1}\left(\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t + \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t \right. \\
&\qquad \left. + \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t\right) \\
&= \hat{W}_{NT}^{-1}\left(\hat{a}_{NTt} + \hat{b}_{NTt} + \hat{c}_{NTt}\right), \qquad \text{say.}
\end{aligned}$$

Lemma B.5 in Appendix B implies that $\hat{W}_{NT}^{-1} - W_{NT}^{-1} \xrightarrow{p} 0$. In addition, Lemma B.6 shows that $\hat{a}_{NTt}$, $\hat{b}_{NT}$, and $\hat{c}_{NTt}$ have the same probabilistic order of magnitude as a_{NTt}, b_{NT}, and c_{NTt}, respectively and that $\sqrt{N}b_{NTt}$ and $\sqrt{N}\hat{b}_{NTt}$ have the same limiting distribution. Thus, the stated result follows. ■

Proof of Theorem 7. Because $J^f - J \xrightarrow{p} 0$ due to Lemma B.6 and because the FGPCE and GPCE have the same limiting distribution, the result can be proven by using the same methods as in Theorems 4 and 5. ■

APPENDIX B: Auxiliary Lemmas

LEMMA B.1. *Assume Assumptions 1, 2, and 3. Then,*

$$\kappa_{NT}^2\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - J'F_t\right\|^2\right) = O_p(1),$$

where $\kappa_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Proof. Let $\Lambda^{o*} = \Omega^{-1/2}\Lambda^o$ and $e_t^* = \Omega^{-1/2}e_t$, and replace Λ^o and e_t in the proof of Bai and Ng's (2002) Theorem 1 with Λ^{o*} and e_t^*, respectively. It suffices to show that parts (i) and (ii) of Lemma 1 of Bai and Ng (2002) hold since Theorem 1 of Bai and Ng (2002) is essentially based on them. Obviously, part (i) of their lemma holds due to Assumption 1(ii). Let λ_i^{o*} be the ith column of Λ' and e_{it}^* the ith element of e_t. Part (ii) also holds due to the relation

$$\begin{aligned}
\mathrm{E}\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it}^* \lambda_i^{o*} \right\|^2 &= \frac{1}{N} \sum_{i=1}^{N} \lambda_i^{o*\prime} \lambda_i^{o*} \\
&= \mathrm{tr}\left(\frac{1}{N} \Lambda' \Omega^{-1} \Lambda \right) \\
&\leq ev_{\min}(\Omega) \max_i \|\lambda_i\|^2 \leq ev_{\min}(\Omega)\bar{\lambda} \qquad \textbf{(B.1)}
\end{aligned}$$

and Assumptions 1(i) and 2(iii). Note that the first inequality in (B.1) uses the Rayleigh-Ritz inequality. ■

LEMMA B.2. *Assume Assumptions 1, 2, and 3. Then,*

$$W_{NT} = \frac{1}{T} \hat{F}' \left(\frac{X\Omega^{-1}X'}{NT} \right) \hat{F} \xrightarrow{p} W,$$

where W is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda^*}\Sigma_F$.

Proof. This can be proven using the same methods as in Stock and Watson (2002a; cf. relation R10). ■

LEMMA B.3. *Assume Assumptions 1, 2, and 3. Then,*

$$\frac{1}{T} \hat{F}' F \xrightarrow{p} W^{1/2} \Theta' \Sigma_{\Lambda^*}^{-1/2} = R \quad \text{as } T, N \to \infty,$$

where $W = diag[w_1, \ldots, w_r]$, $w_1 > \cdots > w_r > 0$ are the eigenvalues of $\Sigma_{\Lambda^*}\Sigma_F$, and Θ is the corresponding eigenvector matrix such that $\Theta'\Theta = I_r$.

Proof. Use $\Omega^{-1/2}\Lambda^o$ and $X\Omega^{-1}X'$ in the proof of Bai's (2003) Proposition 1 instead of Λ^o and XX' there, respectively. ■

LEMMA B.4. *Assume Assumptions 1, 2, and 3. Then,*

(i) $\frac{1}{T}\left(\hat{F} - FJ\right)'\hat{F} = O_p\left(\frac{1}{\kappa_{NT}}\right)$.

(ii) $\frac{1}{T}\left(\hat{F} - FJ\right)'\underline{e}_i = O_p\left(\frac{1}{\kappa_{NT}^2}\right)$.

Proof. (i) Write

$$\frac{1}{T}\left(\hat{F} - FJ\right)'\hat{F} = \frac{1}{T}\left(\hat{F} - FJ\right)'FJ + \frac{1}{T}\left(\hat{F} - FJ\right)'\left(\hat{F} - FJ\right). \qquad \textbf{(B.2)}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Using (A.1), the first term is expressed as

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t - J'F_t\right)F_t'J &= W_{NT}^{-1}\left(\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t F_t'\right.\\ &\quad + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t F_t'\\ &\quad \left. + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}\Lambda F_t F_t'\right)J\\ &= W_{NT}^{-1}\left(A_{NT}+B_{NT}+C_{NT}\right)J, \qquad \text{say.}\end{aligned}$$

Write A_{NT} as

$$\begin{aligned}A_{NT} &= \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)\mathrm{E}\left(e_s'\Omega^{-1}e_t\right)F_t'\\ &\quad + J'\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s\mathrm{E}\left(e_s'\Omega^{-1}e_t\right)F_t'\\ &\quad + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)\left[e_s'\Omega^{-1}e_t - \mathrm{E}\left(e_s'\Omega^{-1}e_t\right)\right]F_t'\\ &\quad + J'\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_s\left[e_s'\Omega^{-1}e_t - \mathrm{E}\left(e_s'\Omega^{-1}e_t\right)\right]F_t'\\ &= A_{1NT} + J'A_{2NT} + A_{3NT} + J'A_{4NT}, \quad \text{say.}\end{aligned}$$

Then

$$\begin{aligned}\|A_{1NT}\| &\le \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|\left|\mathrm{E}\left(e_s'\Omega^{-1}e_t\right)\right|\|F_t\|\\ &\le \sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathrm{E}\left(\frac{e_s'\Omega^{-1}e_t}{N}\right)\right]^2\|F_t\|^2}.\end{aligned}$$

But $\sum_{s=1}^{T}\left[\mathrm{E}\left(\frac{e_s'\Omega^{-1}e_t}{N}\right)\right]^2$ is bounded for any t due to Assumption 1 and Lemma 1(i) of Bai and Ng (2002). Thus, using Assumption 2 and Lemma B.1, we obtain $\|A_{1NT}\| = O_p\left(\frac{1}{T\min\left\{\sqrt{N},\sqrt{T}\right\}}\right)$. Similarly, $\|A_{2NT}\| = O_p\left(\frac{1}{T}\right)$. Next,

$$\begin{aligned}\|A_{3NT}\| &\le \sqrt{\frac{1}{NT^2}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\left[e_s'\Omega^{-1}e_t - \mathrm{E}\left(e_s'\Omega^{-1}e_t\right)\right]F_t'\right\|^2}\\ &= O_p\left(\frac{1}{\sqrt{NT}\kappa_{NT}}\right),\end{aligned}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

where the equality holds due to Assumption 3(i). In the same manner, $A_{4NT} = O_p\left(\frac{1}{\sqrt{NT}}\right)$. Thus, $A_{NT} = O_p\left(\frac{1}{\sqrt{T}\kappa_{NT}}\right)$. Next, write

$$B_{NT} = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)F_s'\Lambda'\Omega^{-1}e_tF_t' + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}J'F_sF_s'\Lambda'\Omega^{-1}e_tF_t'$$

$$= B_{1NT} + B_{2NT}, \quad \text{say.}$$

Assumptions 2 and 3 yield

$$\|B_{1NT}\| \le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{NT}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_tF_t'\right\|^2\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2}$$

$$= O_p\left(\frac{1}{\sqrt{NT}\kappa_{NT}}\right). \tag{B.3}$$

Similarly,

$$B_{2NT} = J'\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_tF_t'\right) = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.4}$$

Last, write

$$C_{NT} = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)e_s'\Omega^{-1}\Lambda F_tF_t' + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}J'F_se_s'\Omega^{-1}\Lambda F_tF_t'$$

$$= C_{1NT} + C_{2NT}, \quad \text{say.}$$

Using the same methods as for B_{1NT} and B_{2NT}, we obtain

$$\|C_{1NT}\| \le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{N}\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\left(\frac{e_s'\Omega^{-1}\Lambda}{\sqrt{N}}\right)\right\|^2\right]\left\|\frac{1}{T}\sum_{t=1}^{T}F_tF_t'\right\|^2}$$

$$= O_p\left(\frac{1}{\sqrt{N}\kappa_{NT}}\right) \tag{B.5}$$

and

$$C_{2NT} = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.6}$$

Since $A_{NT} = O_p\left(\frac{1}{\sqrt{T}\kappa_{NT}}\right)$, as shown above, relations (B.3), (B.4), (B.5), and (B.6) show that the first term in relation (B.2) is $O_p\left(\frac{1}{\sqrt{T}\kappa_{NT}}\right)$. The second term in (B.2) converges to zero in probability at a faster rate than the first term (cf. Theorem 1), and hence the result follows.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(ii) Using relation (A.1), write $\frac{1}{T}\left(\hat{F}-FJ\right)'\underline{e}_i$ as

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t-J'F_t\right)e_{it}=W_{NT}^{-1}\left(\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t e_{it}\right.$$
$$+\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t e_{it}$$
$$\left.+\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}\Lambda F_t e_{it}\right)$$
$$=W_{NT}^{-1}\left(D_{NT}+\mathrm{E}_{NT}+G_{NT}\right),\quad \text{say.}$$

D_{NT} can be written as

$$D_{NT}=\frac{1}{NT}\sum_{s=1}^{T}\left(\hat{F}_s-J'F_s\right)\left(e_s'\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t e_{it}\right)$$
$$+J'\frac{1}{NT}\sum_{s=1}^{T}F_s e_s'\left(\frac{1}{T}\sum_{t=1}^{T}\Omega^{-1}e_t e_{it}\right)$$
$$=D_{1NT}+D_{2NT},\quad \text{say.}$$

Let ω_i be the ith column of Ω. Then, due to Assumption 1(i),

$$\left\|e_s'\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t e_{it}-e_s'\Omega^{-1}\omega_i\right\|\le\left\|\Omega^{-1}\right\|_1\|e_s\|\left\|\frac{1}{T}\sum_{t=1}^{T}e_t e_{it}-\omega_i\right\|=O_p\left(\frac{N}{\sqrt{T}}\right).\tag{B.7}$$

Since $\Omega^{-1}\omega_i=\mathbf{i}$, where $\mathbf{i}$ is an $N\times 1$ vector with the ith element being equal to 1 and the rest zeros, relation (B.7) implies $\left\|e_s'\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t e_{it}\right\|=O_p\left(\max\left\{1,\frac{N}{\sqrt{T}}\right\}\right)$. Thus,

$$\|D_{1NT}\|\le\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\left\|\hat{F}_s-J'F_s\right\|^2}\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\left|e_s'\left(\frac{1}{T}\sum_{t=1}^{T}\Omega^{-1}e_t e_{it}\right)\right|^2}$$
$$=O_p\left(\frac{1}{\sqrt{N}\kappa_{NT}}\right)\times O_p\left(\max\left\{\frac{1}{\sqrt{N}},\sqrt{\frac{N}{T}}\right\}\right)\tag{B.8}$$
$$=O_p\left(\frac{1}{\kappa_{NT}\ \min\left\{N,\sqrt{T}\right\}}\right).$$

Assumption 1(i) yields

$$\left\|\Omega^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}e_t e_t'-\Omega\right)\right\|\le\left\|\Omega^{-1}\right\|_1\left\|\frac{1}{T}\sum_{t=1}^{T}e_t e_t'-\Omega\right\|=O_p\left(\frac{N}{\sqrt{T}}\right),$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

which implies

$$\left|\left\|\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_te_{it}\right\|-1\right|\le\left\|\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_te_{it}-\mathbf{i}\right\|=O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right).$$

Thus, $\left\|\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_te_{it}\right\|=1+O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right)$. Using this, we obtain

$$\|D_{2NT}\|\le\|J\|\frac{1}{N\sqrt{T}}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}F_se_s'\right\|\left\|\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}e_te_{it}\right\| \tag{B.9}$$

$$=\frac{1}{N\sqrt{T}}\times O_p\left(\sqrt{N}\right)\times\left(1+O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right)\right)=O_p\left(\frac{1}{\sqrt{T}\kappa_{NT}}\right).$$

Thus, (B.8) and (B.9) yield

$$\|D_{NT}\|=O_p\left(\frac{1}{\kappa_{NT}\ \min\left\{N,\sqrt{T}\right\}}\right). \tag{B.10}$$

Next, write

$$\mathrm{E}_{NT}=\frac{1}{NT}\sum_{s=1}^{T}\left(\hat{F}_s-J'F_s\right)F_s'\left(\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_te_{it}\right)$$

$$+\frac{1}{NT}\sum_{s=1}^{T}J'F_sF_s'\left(\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_te_{it}\right)$$

$$=\mathrm{E}_{1NT}+\mathrm{E}_{2NT},\quad\text{say.}$$

Then, since $\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_te_{it}\xrightarrow{p}\lambda_i$,

$$\|\mathrm{E}_{1NT}\|\le\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\left\|\hat{F}_s-J'F_s\right\|^2}\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\|F_s\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_te_{it}\right\|^2}$$

$$=O_p\left(\frac{1}{\sqrt{N}\kappa_{NT}}\right)O_p\left(\frac{1}{\sqrt{N}}\right). \tag{B.11}$$

Similarly,

$$\mathrm{E}_{2NT}=J'\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)\frac{1}{NT}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_te_{it}$$

$$=O_p\left(\frac{1}{N}\right). \tag{B.12}$$

Thus, we obtain

$$\|\mathrm{E}_{NT}\|=O_p\left(\frac{1}{N}\right) \tag{B.13}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

from (B.11) and (B.12). Writing

$$G_{NT} = \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)e_s'\Omega^{-1}\Lambda F_t e_{it} + \frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}J'F_s e_s'\Omega^{-1}\Lambda F_t e_{it}$$
$$= G_{1NT} + G_{2NT}, \quad \text{say,}$$

and using Lemma B.1 and Assumption 3, we obtain

$$\|G_{1NT}\| \le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{NT^2}\left(\sum_{s=1}^{T}\left\|\frac{e_s'\Omega^{-1}\Lambda}{\sqrt{N}}\right\|^2\right)\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it}\right\|^2\right)}$$
$$= O_p\left(\frac{1}{\sqrt{NT}\kappa_{NT}}\right)$$

and

$$G_{2NT} = \frac{1}{\sqrt{N}T}J'\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}F_s e_s'\Omega^{-1}\Lambda\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it}\right)$$
$$= O_p\left(\frac{1}{\sqrt{N}T}\right),$$

which give

$$\|G_{NT}\| = O_p\left(\frac{1}{\sqrt{NT}\kappa_{NT}}\right). \tag{B.14}$$

Since $O_p\left(\frac{1}{\kappa_{NT}\min\left\{N,\sqrt{T}\right\}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}\kappa_{NT}}\right) = O_p\left(\frac{1}{\kappa_{NT}^2}\right)$, the result follows from (B.10), (B.13), and (B.14). ■

LEMMA B.5. *Assume Assumptions 1, 2, 3, and 5. Let* $\hat{W}_{NT} = \frac{1}{T}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT}\right)\hat{F}^f$. *Then,* $\hat{W}_{NT} - W_{NT} \xrightarrow{p} 0$.

Proof. Write

$$\left\|\hat{W}_{NT} - W_{NT}\right\| = \left\|\frac{1}{T}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT}\right)\hat{F}^f - \frac{1}{T}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT}\right)\hat{F}\right\|$$
$$\le \left\|\frac{1}{T}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT} - \frac{X\Omega^{-1}X'}{NT}\right)\hat{F}^f\right\|$$
$$+ \left\|\frac{1}{T}\hat{F}^{f\prime}\left(\frac{X\Omega^{-1}X'}{NT}\right)\hat{F}^f - \frac{1}{T}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT}\right)\hat{F}\right\|$$
$$= I + II. \tag{B.15}$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

For the first term we have an inequality,

$$I \leq \frac{1}{T}\left\|\hat{F}^f\right\|^2 \left\|\left(\frac{X\hat{\Omega}^{-1}X'}{NT} - \frac{X\Omega^{-1}X'}{NT}\right)\right\|_1 .$$

Due to the standardization, $\frac{1}{T}\hat{F}^{f\prime}\hat{F}^f = I_r$, $\frac{1}{T}\left\|\hat{F}^f\right\|^2 = r$. Moreover,

$$\begin{aligned}
\left\|\frac{X\hat{\Omega}^{-1}X'}{NT} - \frac{X\Omega^{-1}X'}{NT}\right\|_1 &= \left\|\frac{1}{NT}X\left(\hat{\Omega}^{-1} - \Omega^{-1}\right)X'\right\|_1 \\
&= \left\|\frac{1}{NT}\left(\hat{\Omega}^{-1} - \Omega^{-1}\right)X'X\right\|_1 \\
&\leq \left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_1 \left\|\frac{X'X}{NT}\right\|_1 \\
&\leq \left\|\hat{\Omega}^{-1}\right\|_1 \left\|\Omega^{-1}\right\|_1 \left\|\hat{\Omega} - \Omega\right\|_1 \left\|\frac{X'X}{NT}\right\|_1 , \qquad \textbf{(B.16)}
\end{aligned}$$

for which the relations $\hat{\Omega}^{-1} - \Omega^{-1} = \hat{\Omega}^{-1}(\Omega - \hat{\Omega})\Omega^{-1}$ and $\left\|\hat{\Omega} - \Omega\right\|_1 = \left\|\Omega - \hat{\Omega}\right\|_1$ are used. Since the last term in relation (B.16) is $O_p(1)$ due to Lemma A.3 of Bai (1993), Assumptions 1(i) and 5 imply that $\left\|\frac{X\hat{\Omega}^{-1}X'}{NT} - \frac{X\Omega^{-1}X'}{NT}\right\|_1 = o_p(1)$. Thus, $I = o_p(1)$. For the second term in relation (B.15), we have

$$\begin{aligned}
II &= \left\|\frac{1}{T}\left(\hat{F}^f - \hat{F}\right)'\left(\frac{X\Omega^{-1}X'}{NT}\right)\hat{F}^f + \frac{1}{T}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT}\right)\left(\hat{F}^f - \hat{F}\right)\right\| \\
&\leq \left\|\hat{F}^f - \hat{F}\right\|\left\|\frac{1}{T}\hat{F}^f\right\|\left\|\frac{X\Omega^{-1}X'}{NT}\right\|_1 + \left\|\hat{F}^f - \hat{F}\right\|\left\|\frac{1}{T}\hat{F}\right\|\left\|\frac{X\Omega^{-1}X'}{NT}\right\|_1 \\
&= 2\sqrt{\frac{r}{T}}\left\|\hat{F}^f - \hat{F}\right\|\left\|\frac{X\Omega^{-1}X'}{NT}\right\|_1 .
\end{aligned}$$

Since $\left\|\hat{F}^f - \hat{F}\right\| = o_p\left(\sqrt{T}\right)$ as can be deduced from relation (B.24) below and $\left\|\frac{X\Omega^{-1}X'}{NT}\right\|_1 = O_p(1)$ as shown above, $II = o_p(1)$. Thus, the stated result follows. ■

LEMMA B.6. *Let* $\hat{a}_{NTt} = \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t$, $\hat{b}_{NTt} = \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t$, *and* $\hat{c}_{NTt} = \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t$, *where* $\hat{F}_s^f$ *is the* s*th column of* $\hat{F}^{f\prime}$. *Under Assumptions 1, 2, 3, and 5,*

(i) $\hat{a}_{NTt} - a_{NTt} \xrightarrow{p} 0$;
(ii) $\hat{b}_{NTt} - b_{NTt} \xrightarrow{p} 0$;
(iii) $\hat{c}_{NTt} - c_{NTt} \xrightarrow{p} 0$;

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(iv) $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^f F_s' \xrightarrow{p} W^{1/2}\Theta'\Sigma_{\Lambda^*}^{-1/2}$, *where* W *and* Θ *are defined in Lemma B.3;*

(v) $\sqrt{N}\left(\hat{b}_{NTt} - b_{NTt}\right) \xrightarrow{p} 0.$

Proof. (i) First, consider the relation

$$\begin{aligned}
&\left\| \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t \right\| \\
&\quad = \left\| \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t \right\| \\
&\quad \le \frac{1}{NT}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|\left\|e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\right\| \\
&\quad \le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|^2}\sqrt{\frac{1}{N^2T}\sum_{s=1}^{T}\left\|e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\right\|^2}, \qquad \textbf{(B.17)}
\end{aligned}$$

where the Cauchy-Schwarz inequality is used for the last inequality. The first term in (B.17) is $\sqrt{r}$. The second term is less than $\sqrt{\frac{1}{N^2T}\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1^2\|e_t\|^2\Sigma_{s=1}^{T}\|e_s\|^2}$. Since $\|e_t\|^2 = O_p(N)$ and $\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1 \le \left\|\hat{\Omega}^{-1}\right\|_1\left\|\hat{\Omega}-\Omega\right\|_1\left\|\Omega^{-1}\right\|_1 = o_p(1)$ as in the proof of Lemma B.5, the second term is $o_p(1)$. Thus we have

$$\left\| \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t \right\| \xrightarrow{p} 0. \qquad \textbf{(B.18)}$$

Next, the Cauchy-Schwarz inequality gives

$$\begin{aligned}
&\left\| \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t \right\| \\
&\quad = \left\| \frac{1}{NT}\sum_{s=1}^{T}\left(\hat{F}_s^f-\hat{F}_s\right)e_s'\Omega^{-1}e_t \right\| \\
&\quad \le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f-\hat{F}_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|e_s'\Omega^{-1}e_t/N\right\|^2}.
\end{aligned}$$

The first term can be shown to be bounded above by a positive constant using the triangle inequality and the fact that $\sum_{s=1}^{T}\hat{F}_s^f\hat{F}_s^{f\prime} = \sum_{s=1}^{T}\hat{F}_s\hat{F}_s' = T\times I_r$. The second one can be shown to be $o_p(1)$ under Assumption 1(ii) using the same method as the proof of Theorem 1 of Bai and Ng (2002). Thus,

$$\left\| \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t \right\| \xrightarrow{p} 0. \qquad \textbf{(B.19)}$$

The stated result follows from (B.18) and (B.19), once the triangular inequality is used.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(ii) Write

$$\begin{aligned}\hat{b}_{NTt} - b_{NT} &= \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t \\ &= \left(\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\Omega^{-1}e_t\right) \\ &\quad + \left(\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\Omega^{-1}e_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t\right) \\ &= I + II. \end{aligned} \tag{B.20}$$

Since $\frac{1}{\sqrt{N}}\Lambda'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^f F_s' = O_p(1)$ as shown in part (iv) of this lemma, $I = o_p(1)$. Moreover,

$$\begin{aligned}\|II\| &\le \frac{1}{\sqrt{N}}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\hat{F}_s^f - \hat{F}_s\right)F_s'\right\|\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\| \\ &\le \frac{1}{\sqrt{N}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f - \hat{F}_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2}\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\| \xrightarrow{p} 0. \end{aligned} \tag{B.21}$$

Thus, the stated result follows.

(iii) Similarly to previous parts, we obtain

$$\begin{aligned}&\left\|\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}\Lambda F_t\right\| \\ &\quad\le \left\|\frac{1}{\sqrt{N}T}\sum_{s=1}^{T}\hat{F}_s^f e_s'\right\|\left\|\frac{1}{\sqrt{N}}\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)\Lambda F_t\right\| \\ &\quad\le \sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|^2}\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\|e_s\|^2}\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1\left\|\frac{1}{\sqrt{N}}\Lambda\right\|\sup_t\|F_t\| \\ &\quad= o_p(1), \end{aligned}$$

and

$$\begin{aligned}&\left\|\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}\Lambda F_t - \frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}\Lambda F_t\right\| \\ &\quad= \left\|\frac{1}{NT}\sum_{s=1}^{T}\left(\hat{F}_s^f - \hat{F}_s\right)e_s'\Omega^{-1}\Lambda F_t\right\| \\ &\quad\le \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f - \hat{F}_s\right\|\left\|\frac{1}{\sqrt{N}}e_s'\Omega^{-1}\Lambda\right\|\right)\sup_t\|F_t\| \\ &\quad\le \frac{1}{\sqrt{N}}\left(\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s^f - \hat{F}_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}e_s'\Omega^{-1}\Lambda\right\|^2}\right)\sup_t\|F_t\| \\ &\quad= o_p(1). \end{aligned}$$

These imply the stated result.

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

(iv) This will be proved by following the proof of Proposition 1 of Bai (2003). By definition, $\frac{1}{NT} X\hat{\Omega}^{-1} X' \hat{F}^f = \hat{F}^f \hat{W}_{NT}$. Multiplying this relation by $\frac{1}{T}\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2} F'$ on both sides, we have

$$\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)\left(\frac{F'\hat{F}^f}{T}\right)+d_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right) \tag{B.22}$$

$$=\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}\left(\frac{F'\hat{F}^f}{T}\right)\hat{W}_{NT},$$

where

$$d_{NT}(\hat{\Omega}^{-1},\hat{F}^f)$$

$$=\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}\frac{1}{T}F'\left[\frac{1}{NT}\left(F\Lambda'\hat{\Omega}^{-1}e'\hat{F}^f+e\hat{\Omega}^{-1}\Lambda F'\hat{F}^f+e\hat{\Omega}^{-1}e'\hat{F}^f\right)\right].$$

Let $B_{NT}(\hat{\Omega}^{-1})=\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}$, $R_{NT}(\hat{\Omega}^{-1},\hat{F}^f)=\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{1/2}\left(\frac{F'\hat{F}^f}{T}\right)$, and W^*_{NT} be a diagonal matrix consisting of the diagonal elements of $R'_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)R_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)$. Then, denoting $\Theta_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)=R_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)W^{*-1/2}_{NT}$, equation (B.22) can be written as

$$\left(B_{NT}\left(\hat{\Omega}^{-1}\right)+d_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)R^{-1}_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)\right)\Theta_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)$$
$$=\Theta_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)\hat{W}_{NT}. \tag{B.23}$$

Since $\frac{F'\hat{F}^f}{T}=\left(\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}\right)^{-1/2}\Psi_{NT}(\hat{\Omega}^{-1})W^{*1/2}_{NT}$, Bai's method amounts to obtaining the probability limit of $\frac{F'\hat{F}^f}{T}$ by treating (B.23) as an approximate eigenvalue-eigenvector relation and solving it with respect to $\Psi_{NT}\left(\hat{\Omega}^{-1}\right)$. Thus, once we show $B_{NT}\left(\hat{\Omega}^{-1}\right)-B_{NT}\left(\Omega^{-1}\right)\xrightarrow{p}0$, $d_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)-d_{NT}\left(\Omega^{-1},\hat{F}\right)\xrightarrow{p}0$, $R_{NT}\left(\hat{\Omega}^{-1},\hat{F}^f\right)=O_p(1)$, and $\hat{W}_{NT}-W_{NT}\xrightarrow{p}0$, the stated result follows from Lemma B.3. The first two relations hold as in previous parts. The third relation also holds because $\frac{\Lambda'\hat{\Omega}^{-1}\Lambda}{N}-\frac{\Lambda'\Omega^{-1}\Lambda}{N}\xrightarrow{p}0$ and $\frac{F'\hat{F}^f}{T}=O_p(1)$. The fourth relation is proven in Lemma B.5.

(v) It suffices to show that the two terms, I and II, in relation (B.20) are $o_p(1)$ when they are multiplied by $\sqrt{N}$. Parts (ii) and (iv) of this lemma imply $\sqrt{N}I=o_p(1)$. In addition, we have a triangle inequality

$$\left\|\hat{F}^f_s-\hat{F}_s\right\|=\left\|\left(\hat{F}^f_s-J^{f\prime}F_s\right)-\left(\hat{F}_s-J'F_s\right)+\left(J^f-J\right)'F_s\right\|$$
$$\leq\left\|\hat{F}^f_s-J^{f\prime}F_t\right\|+\left\|\hat{F}_s-J'F_s\right\|+\left\|\left(J^f-J'\right)F_s\right\|.$$

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press

Parts (i), (ii), and (iii) imply that the first term is $o_p(1)$ in the inequality above. The second term is also $o_p(1)$, as in the proof of Theorem 1. Using Assumption 5(ii), part (iv) of this lemma, and Lemma B.5, we also find the third term to be $o_p(1)$. Thus,

$$\left\| \hat{F}_s^f - \hat{F}_s \right\| = o_p(1), \tag{B.24}$$

which with other assumptions implies $\sqrt{N}II = o_p(1)$ (cf. relation (B.21)). ■

https://doi.org/10.1017/S0266466611000338 Published online by Cambridge University Press