Efficient Estimation of Nonstationary Factor Models*

In Choi

First Draft: June, 2011, Revised: December, 2012; June 2014

Abstract

This paper studies the generalized principal component estimator (GPCE) of Choi (2012) for the factor model $X_t = \Lambda F_t + e_t$ where F_t is a unit-root process. This paper makes the following theoretical contributions to the literature on factor analysis. First, this paper derives asymptotic distributions of the GPCEs of the factor and factor-loading spaces which show that the GPCE enjoys an efficiency gain over the conventional principal component estimator. Second, this paper extends the conventional static factor model to those with time polynomials, and studies the GPCE for the models. The GPCE continues to have an efficiency gain over the conventional principal component estimator for the extended model. Third, this paper considers the forecasting regression that uses the GPCE-based estimates of nonstationary factors and shows that the GPCE yields more accurate forecasts than the conventional principal component estimator. Last, asymptotic equivalence of the GPCE and feasible GPCE of the factor space is established. Simulation results corroborate the efficiency gain of the GPCE over the conventional principal component estimator in finite samples. Furthermore, it is shown via simulation that the mean squared forecasting errors based on level data are much lower than those based

*This paper was presented at econometrics seminars of Maastricht University and Goethe University in 2011. The author thanks the organizers and participants of the seminars, especially Jörg Breitung, Jean-Pierre Urbain and Uwe Hassler, for constructive comments. This paper has also benefitted from insightful comments of Guido Kuersteiner and two referees, whom the author thanks.
Address correspondence to In Choi, Department of Economics, Sogang University, Baekbeom-ro 35, Mapo-gu, Seoul, 121-742 Korea; e-mail: inchoi@google.com, inchoi@sogang.ac.kr.

on differenced ones when the factor-augmented predictive regression model is used.

Keywords: factor model, unit root, generalized principal component estimation, feasible generalized principal component estimation

1 Introduction

Factor models are becoming increasingly popular in economics and finance because they can utilize large data sets in an effective manner. Factor models have been used for various purposes. First, they have been used to construct economic indicators. Monthly coincident business cycle indicators such as the Chicago Fed National Activity index (CFNAI) for the US and EuroCOIN for the Euro area (cf. Altissimo *et al.* 2001) are prime examples that belong to this category. In addition, Cristadoro *et al.* (2005) and Kapetanios (2004) construct measures of core inflation by using factor models. Second, factor models have widely been used in order to forecast real and nominal economic variables. Examples are Artis *et al.* (2005), Banerjee and Marcellino (2006), Banerjee et al. (2005), Camba-Mendez and Kapetanios (2005), den Reijer (2005), Forni *et al.* (2000, 2003), Giacomini and White (2006), Huang *et al.*, (2006), Ludvigson and Ng (2007), Marcellino *et al.* (2003), Schumacher and Dreger (2004), Schumacher (2007) and Stock and Watson (1999, 2002b). These works suggest some evidence that factor models provide more accurate forecasts than more conventional models such as autoregressive and vector autoregressive models. Third, factor models have been used for monetary policy analysis. Bernanke *et al.* (2005), Favero *et al.* (2005), Forni *et al.* (2009), Giannone *et al.* (2002, 2005) and Sala (2003) show how factor models can profitably be used to study monetary policy. Fourth, factor models are used for instrumental variables estimation as in Bai and Ng (2010). Bai and Ng assume that endogenous regressors are driven by a small number of unobserved, exogenous factors and suggest using the estimated factors as instruments. Last, factor models have been used in panel regressions as a way of modelling cross-sectional correlation. Examples of this approach are Bai (2009), Bai and Ng (2004), Greenaway-McGrevy, Han and Sul (2012), Moon and Perron (2004) and Phillips and

Sul (2003). For further references on factor models, see the review articles Breitung and Eickmeier (2006) and Breitung and Choi (2013).

Chamberlain and Rothschild (1983) and Connor and Korajzcyk (1986) consider the static factor model and use the principal component estimation method to estimate it. This method is further studied by Stock and Watson (2002a, 2002b), Bai and Ng (2006) and Bai (2003, 2004). Forni *et al.* (2000, 2005) consider the dynamic factor model and suggest using the dynamic principal component method of Brillinger (1981). In addition, Kapetanios and Marcellino (2009) use methods for the state-space model to estimate the dynamic factor model.

Regarding the efficiency of the principal component estimation method for the static factor model, Choi (2012) shows that the principal component estimators of the factor space and the common component are less efficient than his generalized principal component estimators (GPCEs). It is also shown that the GPCE corresponds to the generalized least squares estimator in linear regression while the principal component estimator (see Connor and Korajzcyk, 1986; Bai, 2004), which we will call the ordinary principal component estimator (OPCE), to the ordinary least squares estimator. A paper closely related to Choi (2012) and this one is Breitung and Tenhofen (2011). Breitung and Tenhofen study efficient estimation of the factor and factor-loading spaces for the static factor model with serially correlated, heteroskedastic idiosyncratic errors. But they do not allow cross-sectional correlation as in Choi (2012) and this paper. Note that both Choi (2012) and Breitung and Tenhofen assume stationary factors.

This paper further studies the GPCE for the static factor model whose factors now have a unit root. The same model is studied in Bai (2004), but he considers the OPCE. The contributions of this paper are as follows. First, this paper derives asymptotic distributions of the GPCEs of the factor and factor-loading spaces by using the techniques of Stock and Watson (2002a) and Bai (2003, 2004). The results show that the GPCE enjoys an efficiency gain over the OPCE as in the case of stationary factors. Second, this paper extends the conventional static factor model to those with time polynomials, and studies the GPCE for the models. The GPCE continues to have an efficiency gain over the OPCE for the extended model. The model is particularly suitable for data with nonstationary factors since stochastic and nonstochastic trends are simultaneously present in many macroeconomic time series. Third, this paper considers the forecasting regression that uses the GPCE-based estimates of nonstationary factors and shows that the GPCE yields more accurate forecasts than the OPCE. Last, asymptotic equivalence of the GPCE and feasible GPCE (FGPCE) of the factor space is established. The FGPCE is the GPCE that uses the estimated variance-covariance matrix of the idiosyncratic errors in place of the true variance-covariance matrix. It is estimated by using the estimates of idiosyncratic errors from the OPCE.

This paper reports simulation results that corroborate the theoretical efficiency advantage of the GPCE in finite samples. Furthermore, it provides simulation results that the mean squared forecasting errors based on level data are much lower than those based on differenced ones when the factor-augmented predictive regression model is used. These results challenge the conventional practice of factor-augmented forecasting since most factor-augmented predictive regressions have used estimates of factors from the differenced data (e.g., Stock and Watson, 1999). In accordance with the simulation results of this paper, Choi and Hwang (2012) show by using Korean data that level data provide more accurate forecasts of inflation rates. It remains to be seen whether this result extends to other countries' data sets. In addition, Engel, Mark and West (2012) and Greenaway-McGrevy, Mark, Sul and Wu (2012) use the factor model with nonstationary factors to forecast exchange rates. They report that using estimated, nonstationary factors produce reasonably good forecasts.

In practice, researchers usually center and rescale the data by the estimated standard deviation before the principal component analysis. The centering is used because the principal component analysis assumes a zero-mean process, and the rescaling to handle the situation of heteroskedastic data. The results in this paper indicate that the rescaling should be implemented not by the estimated standard deviation of the data themselves but by the estimated standard deviation of the idiosyncratic errors when heteroskedasticity of the data is of concern. Using the estimated standard deviation of the idiosyncratic errors leads to optimal efficiency according to the results of this paper.

This paper proceeds as follows. Section 2 introduces the model and the GPCEs of the factor and factor-loading spaces. Section 3 derives asymptotic distributions of the GPCEs and compares their efficiency with that of the OPCEs. Section 4 extends the results of Section 3 to models with time polynomials. Section 5 studies the forecasting regression that uses the estimated factors based on the GPCE. Section 6 establishes asymptotic equivalence of the GPCE and FGPCE. Section 7 provides simulation results. Section 8 summarizes and concludes. All the proofs are contained in Appendices I and II.

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 the minimum of the eigenvalues of A, respectively, and $\|A\|_1 = \max_x\{\|Ax\| : \|x\| = 1\}$. $\|A\|_1$ is the positive square root of the largest eigenvalue of $A'A$ and is called the spectral norm of the matrix A. For matrices X and Y, $X \oplus Y = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$ and $P_X = X(X'X)^{-1}X'$. All the matrices in this paper are real ones. The integer part of a real number a is denoted as $[a]$. A diagonal matrix with diagonal elements $\{a_i\}_{i=1,...,N}$ is denoted as $diag(a_i)_{i=1,...,N}$. All the limits are taken as $N,\ T \to \infty$ unless otherwise stated. Convergence in probability, convergence in distribution and weak convergence are denoted as $\xrightarrow{p}$, $\xrightarrow{d}$ and $\Rightarrow$, respectively. The usual difference operator is denoted as Δ (i.e., $\Delta x_t = x_t - x_{t-1}$).

2 The model and estimator

We are concerned with the factor model

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

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 error of the model. We assume that $\{F_t\}$ is a nonstationary process represented by

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

where $\{u_t\}$ is a stationary process. It is assumed that $\{F_t\}$ is not cointegrated. More specific assumptions on $\{F_t\}$ will be given in Section 3.

Model (1) can be written in scalar notation as

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

where λ_i' is the i-th row of matrix Λ. It can also be written in matrix notation 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 $e = \begin{bmatrix} e_1' \\ \vdots \\ e_T' \end{bmatrix}$.

With the standardization $F'F = T^2 \times I_r$, the GPCE of the factor space of F, denoted by $\hat{F}$, is T times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix $X\Omega^{-1}X'$. The standardization conforms to the rate give in Assumption 2 below. The GPCE is motivated by the conditional maximum likelihood estimation with the assumption $e_t \mid F \sim iid\ N(0, \Omega)$ where $\Omega > 0$ (see Choi, 2012, for details). The GPCE of Λ is given by $\hat{\Lambda} = \frac{1}{T^2}X'\hat{F}$. Note that the OPCE (see Connor and Korajzcyk, 1986; Bai, 2004) is T times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix XX'.

The GPCEs introduced so far assume that Ω is known. It can be estimated by using the OPCE-based estimates of the idiosyncratic errors, and feasible versions of the GPCEs using this estimator can be constructed as will be discussed in Section 6.

One may consider differencing model (1) and estimate the factor-loading and factor spaces as in Bai and Ng (2004). However, differencing makes the idiosyncratic errors more serially correlated which makes the OPCEs and GPCEs less efficient. Thus, if it is known that the factors are $I(1)$ processes, it seems more proper to use the model in levels. This issue is further studied via simulation in Section 7 in relation to forecasting.

3 Asymptotic distributions of the generalized principal component estimator

This section reports asymptotic distributions of the GPCEs introduced in the previous section. A few assumptions are required for the asymptotic distributions. Regarding the idiosyncratic errors, we make the following assumption.

Assumption 1 *There exists positive constants M_1, M_2 and M such that:*
(i) $\{e_t\}$ is weakly stationary with $E(e_t) = 0$; $E(e_t e_t') = \Omega$ for all t; $\max_i \sigma_{e_i}^2 < M$ where $\sigma_{e_i}^2 = \sum_{h=-\infty}^{\infty} E(e_{i1}e_{ih})$ and $\left\| \frac{1}{T}\sum_{t=1}^{T} e_t e_t' - \Omega \right\| = O_p\left(\frac{N}{\sqrt{T}}\right)$.
(ii) For $M_1, M_2 > 0$, $M_1 < ev_{\min}(\Omega) < M_2$ and $M_1 < ev_{\max}(\Omega) < M_2$.
(iii) Let $\gamma(s,t) = E(\frac{1}{N}e_s'\Omega^{-1}e_t)$. Then, $\sum_{s=1}^{T} |\gamma(s,t)| < M$ for all t and T.
(iv) $E\left[\sqrt{N}\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right)\right]^2 < M$ for every t and s.

According to part (i) of this assumption, $\{e_t\}$ is a zero-mean process with the fixed variance–covariance matrix Ω for every t. Note that serial correlation is allowed for $\{e_t\}$, though its structure is not assumed here. Part (ii) implies that the variance-covariance matrix Ω is positive definite for every finite N and even in the limit since all of its eigenvalues are assumed to be positive and finite. Thus, Ω^{-1} exists for $N = 1, 2, ..., \infty$. Parts (iii) and (iv) are related not only to the stochastic properties of $\{e_t\}$ but also to the degree of sparsity of the covariance matrix Ω. If most of the elements of Ω are nonzero, parts (iii) and (iv) would not hold. Examples of $\{e_t\}$ and Ω that satisfy Assumption 1 are given below.

Example 1 *(Heteroskedasticity and serial correlation) Assume a moving average process of the first order* $e_{it} = v_{it} + \theta_i v_{i,t-1}$, *where* $|\theta_i| < 1,\ v_{it} \sim iid(0, \sigma_i^2)$, $\max_i E(v_{i1}^4) < C_1\ (> 0)$, $\max_i \sigma_i^2 < C_2\ (> 0)$ *and* $E(v_{it}v_{js}) = 0$ *for* $i \neq j$ *and for every* t *and* s. *Then,*

$$\Omega = diag((1+\theta_i^2)\sigma_i^2)_{i=1,...,N}.$$

Parts (i) and (ii) are easily satisfied. In addition,

$$\begin{aligned}\gamma(s,t) &= \frac{1}{N}\sum_{i=1}^{N} E\left(\frac{e_{is}e_{it}}{(1+\theta_i^2)\sigma_i^2}\right) \\ &= \left\{ \begin{array}{ll} 1, & s=t \\ \frac{1}{N}\sum_{i=1}^{N}\theta_i/(1+\theta_i^2), & |s-t| = 1 \\ 0, & |s-t| \geq 2 \end{array}\right. ,\end{aligned}$$

from which it is found that part (iii) is satisfied. Letting $\frac{e_{is}e_{it}}{(1+\theta_i^2)\sigma_i^2} = q_{i,st}$, *we have*

$$\begin{aligned} & E\left[\sqrt{N}\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right)\right]^2 \\ &= E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[q_{i,st} - E\left(q_{i,st}\right)\right]\right)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(q_{i,st} - E\left(q_{i,st}\right)\right)\left(q_{j,st} - E\left(q_{j,st}\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} E\left(q_{i,st} - E\left(q_{i,st}\right)\right)^2 < M, \end{aligned}$$

where the last equality is obtained using the fact that $\{q_{i,st}\}$ *are independent over* i *and* M *is a positive constant that depends on* C_1 *and* C_2. *Thus, part (iv) is satisfied.*

Example 2 *(Cross-sectional correlation and no serial correlation) Assume a spatial autoregressive process* $e_t = \lambda W e_t + v_t,\ |\lambda| < 1,\ v_t \sim iid\ N(0, \sigma_v^2 I_N)$, $W = W_1 \oplus ... \oplus$

W_n, where W_k $(k = 1, ..., n)$ is a square matrix of dimension a. Assume further that $N = na$. Since $e_t = (I_N - \lambda W)^{-1} v_t$, the variance-covariance matrix of $\{e_t\}$ is

$$\begin{aligned}\Omega &= \sigma_v^2 (I_N - \lambda W)^{-1} (I_N - \lambda W')^{-1} \\ &= \sigma_v^2 (I_a - \lambda W_1)^{-1} (I_a - \lambda W_1')^{-1} \oplus ... \oplus \sigma_v^2 (I_a - \lambda W_n)^{-1} (I_a - \lambda W_n')^{-1}.\end{aligned}$$

Suppose that the minimum and maximum eigenvalues of $(I_a - \lambda W_k)^{-1}(I_a - \lambda W_k')^{-1}$ satisfy part (ii) of Assumption 1 for every k. Then, parts (i) and (ii) are satisfied. Using the relation $\Omega^{-1} = \frac{1}{\sigma_v^2}(I_N - \lambda W')(I_N - \lambda W)$, we obtain

$$\gamma(s, t) = \frac{1}{N} E(v_s' v_t / \sigma_v^2) = \left\{ \begin{array}{ll} 1, & s = t \\ 0, & s \neq t \end{array} \right.,$$

which satisfies part (iii). Using the normality assumption on $\{v_{it}\}$, we have

$$\begin{aligned} & E\left[\sqrt{N}\left(\frac{1}{N} e_s' \Omega^{-1} e_t - \gamma_N(s, t)\right)\right]^2 \\ = \ & E\left[\sqrt{N}\left(\frac{1}{N} v_s' v_t / \sigma_v^2 - \gamma_N(s, t)\right)\right]^2 \\ = \ & \left\{ \begin{array}{ll} 2, & s = t \\ 1, & s \neq t \end{array} \right. .\end{aligned}$$

This shows that part (iv) is satisfied.

For the factor-loading matrix and latent factors, the following assumption is required. The factor-loading matrix Λ is assumed to be a constant matrix for simplicity.

Assumption 2 *(i) $rank(\Lambda) = r$ 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) $\frac{\sum_{t=1}^{[Ts]} \Delta F_t}{\sqrt{T}} \Rightarrow B_F(s)$, where $B_F(s)$ is an r-dimensional Brownian motion with a positive-definite covariance matrix $\Phi_F = \sum_{s=-\infty}^{\infty} E(u_1 u_s')$.
(v) The eigenvalues of the matrix $\Sigma_{\Lambda^*}^{1/2} \int_0^1 B_F(r) B_F'(r) dr \Sigma_{\Lambda^*}^{1/2}$ are distinct almost surely.

Parts (i)-(iii) of these assumptions are of standard nature in the literature of factor models. Part (iv) is the functional central limit theorem. It holds under some regularity conditions on $\{\Delta F_t\}$. For example, if $\{\Delta F_t\}$ is a mixing process or a linear process that satisfy some other conditions, part (iv) holds (cf. Phillips and Durlauf, 1986, Phillips and Solo, 1992, and Chapter 29 of Davidson, 1994). Note that it allows for serial correlation in $\{\Delta F_t\}$. But cointegration among the elements of F_t is not allowed since Φ_F is positive-definite. Part (v) is required for Lemma A.5 in Appendix II. Under part (iv), it follows that $\frac{1}{T^2}\sum_{t=1}^{T} F_t F_t' \xrightarrow{d} \int_0^1 B_F(r)B_F'(r)dr$ as discussed in Phillips and Durlauf (1986).

In addition, we make the following assumption.

Assumption 3 *(i) $\{e_t\}$ and $\{F_t\}$ are independent.*
(ii) For each i and t, as $T, N \to \infty$,

$$(a)\ \frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t \xrightarrow{d} N\left(0, \Sigma_{\Lambda^*}\right);$$

$$(b)\ \frac{1}{T}\sum_{s=1}^{T} F_s e_{is} \xrightarrow{d} \int_0^1 B_F(r)dB_{ei}(r)$$

and

$$(c)\ \left\| E\left(\frac{1}{\sqrt{N}T}\Lambda'\Omega^{-1}e_t \sum_{s=1}^{T} F_s' e_{is}\right)\right\| \to 0.$$

(iii) $\left\|\frac{1}{T\sqrt{N}}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_t F_t'\right\| = O_p(1)$ *uniformly in N and T.*

Define the Brownian motion $B_{ei}(r)$ by the weak convergence relation $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} e_{it} \Rightarrow B_{ei}(r)$ as $T \to \infty$. Under part (i), $B_F(\cdot)$ and $B_{ei}(\cdot)$ become independent (see Park and Phillips, 1988), which will be used to apply the continuous mapping theorem in the proofs of Theorems 1 and 3. Part (ii) is required to derive limiting distributions of the GPCEs. Part (ii)-(a) is the usual central limit theorem, and part (ii)-(b) a weak convergence result often used in the literature on cointegration (e.g., Phillips and Durlauf, 1986). Note that (ii)-(a) and (ii)-(b) are marginal convergence results. Part (ii)-(c) makes the random vectors in (ii)-(a) and (ii)-(b) independent in the limit since their weak limits are Gaussian processes. Thus, joint convergence result for $(\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t,\ \frac{1}{T}\sum_{s=1}^{T} F_s e_{is})$ follows from the marginal convergence results (see Theorem 4.5 of Billingsley, 1968, or Theorem 3.29 of Kallenberg, 1997). The joint convergence result is required in order to apply the continuous mapping theorem as in Theorem 3 (cf. van der Vaart, 1998, p.11). The limiting distribution of $\frac{1}{T}\sum_{s=1}^{T} F_s e_{is}$ can also be written as $\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{1/2} N(0, \sigma_{e_i}^2)$, which show that it is a mixture-normal distribution. More primitive conditions for the weak convergence result (ii)-(b) can be found in Davidson (1994) and Phillips and Durlauf

(1986). Heuristically speaking, part (iii) states that $\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}e_tF_t'\right)$ follows the central limit theorem. This can be verified if lower-level assumptions on $\{e_t\}$ and $\{F_t\}$ are given. It will be used to prove part (iii) of Lemma A.2 and part (i) of Lemma A.6.

Part (ii)-(c) of Assumption 3 is satisfied when some specific conditions are given on $\{e_t\}$ and Ω. Examples are as follows.

Example 3 *Assume the same data generating process as in Example 1. If* $\|E(F_0)\| < M\ (>0)$, *where* F_0 *is the initial variable for* $\{F_t\}$, *it follows that*

$$\begin{aligned}
&\left\|E\left(\frac{1}{\sqrt{NT}}\Lambda'\Omega^{-1}e_t\sum_{s=1}^{T}F_s'e_{is}\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\lambda_j E\left(e_{jt}e_{is}\right)E\left(F_s'\right)/\sigma_j^2\right\| \\
= &\left\|\frac{1}{\sigma_i^2\sqrt{NT}}\lambda_i\sum_{s=1}^{T}E\left(e_{it}e_{is}\right)E\left(F_0'\right)\right\| \to 0
\end{aligned}$$

Example 4 *Assume the same data generating process as in Example 2 and let* $\Omega^{-1} = \Omega_1^{-1}\oplus\ldots\oplus\Omega_n^{-1}$, $\Lambda' = [\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_n]$ *and* $e_t' = [\mathbf{e}_{1t}', ..., \mathbf{e}_{nt}']$, *where* $\boldsymbol{\lambda}_k$ *and* $\mathbf{e}_{kt}$ $(k = 1, ..., n)$ *are* $r\times a$ *and* $a\times 1$ *matrices. Assume without loss of generality that* e_{it} *is the first element of* $\mathbf{e}_{k^o t}$. *Then, if* $\|E(F_0)\| < M\ (>0)$,

$$\begin{aligned}
&\left\|E\left(\frac{1}{\sqrt{NT}}\Lambda'\Omega^{-1}e_t\sum_{s=1}^{T}F_s'e_{is}\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{n}\sum_{s=1}^{T}\boldsymbol{\lambda}_k\Omega_k^{-1}E\left(\mathbf{e}_{kt}e_{is}\right)E\left(F_s'\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\boldsymbol{\lambda}_{k^o}\Omega_{k^o}^{-1}E\left(\mathbf{e}_{k^o t}e_{is}\right)E\left(F_0'\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\boldsymbol{\lambda}_{k^o}\Omega_{k^o}^{-1}E\left(\mathbf{e}_{k^o t}e_{it}\right)E\left(F_0'\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\boldsymbol{\lambda}_{k^o}\Omega_{k^o}^{-1}\boldsymbol{\omega}_{ik^o}E\left(F_0'\right)\right\| \\
= &\left\|\frac{1}{\sqrt{NT}}\boldsymbol{\lambda}_{k^o}\mathbf{i}_a E\left(F_0'\right)\right\| \to 0,
\end{aligned}$$

where $\boldsymbol{\omega}_{ik^o}$ *is the first column of* Ω_{k^o} *and* $\mathbf{i}_a$ *is an* $a\times 1$ *vector with the first element being equal to 1 and the rest zeros.*

Asymptotic distributions of the GPCE of the factor space are reported in Theorem 1 below.

Theorem 1 *Let $\hat{F}_t$ be the t-th column of $\hat{F}'$. Suppose that Assumptions 1, 2 and 3 hold.*
(i) If $\frac{N}{T^3} \to 0$,

$$\sqrt{N}(\hat{F}_t - J'F_t) \xrightarrow{d} W^{-1/2} N(0, I_r),$$

where $J = \frac{1}{NT^2}(\Lambda'\Omega^{-1}\Lambda)(F'\hat{F})W_{NT}^{-1}$, W_{NT} is an $r \times r$ diagonal matrix consisting of the first r largest eigenvalues of $\frac{1}{NT^2}X\Omega^{-1}X'$ in descending order, $W = diag[w_1, ..., w_r]$, $w_1 > ... > w_r > 0$ are the eigenvalues of $\Sigma_{\Lambda^*}^{1/2}\int_0^1 B_F(r)B_F'(r)dr\Sigma_{\Lambda^*}^{1/2}$ and W is independent of $N(0, I_r)$.
(ii) If $\liminf \frac{N}{T^3} \geq \tau > 0$, $T^{3/2}(\hat{F}_t - J'F_t) = O_p(1)$.

There are some aspects of this theorem that are worth considering. First, $\hat{F}_t$ estimates a rotation of F_t consistently, not F_t itself.

Second, the order of convergence of the GPCE is the same as that of the GPCE for stationary factors (cf. Theorem 1 of Choi, 2012) when $\frac{N}{T^3} \to 0$. When the latter condition is not met, it is different from that of the GPCE for stationary factors.

Third, although it is hard to compare efficiency of the GPCE and OPCE directly because these are estimating different rotations of the factors, it is possible to gain some insights on the efficiency comparison of these estimators by using Theorem 1 and Corollary 1 of Bai (2004). Because $J \xrightarrow{p} R^{-1}$ ($R = W^{1/2}\Theta'\Sigma_{\Lambda^*}^{-1/2}$ and Θ is the eigenvector matrix of $\Sigma_{\Lambda^*}^{1/2}\int_0^1 B_F(r)B_F'(r)dr\Sigma_{\Lambda^*}^{1/2}$ such that $\Theta'\Theta = I_r$) and $R'W^{-1}R = \Sigma_{\Lambda^*}^{-1}$, Theorem 1 can be rewritten as

$$\sqrt{N}\left(J'^{-1}\hat{F}_t - F_t\right) \xrightarrow{d} N\left(0, \Sigma_{\Lambda^*}^{-1}\right). \tag{4}$$

In contrast, Corollary 1 of Bai (2004) implies under the assumption that $E\left(e_t e_t'\right) = \Omega$ for every t

$$\sqrt{N}\left(H_2'^{-1}\hat{F}_{ot} - F_t\right) \xrightarrow{d} N\left(0, \Sigma_\Lambda^{-1}\Sigma_{\Lambda^*}\Sigma_\Lambda^{-1}\right), \tag{5}$$

where $\hat{F}_{ot}$ is the OPCE, H_2 is as defined in Bai (2004) and $\Sigma_\Lambda = \lim_{N\to\infty}\frac{\Lambda'\Lambda}{N}$. The asymptotic variance-covariance matrices in relations (4) and (5) are similar to those of the GLS and OLS estimators in standard linear regression, respectively, once Λ is regarded as an observed regressor matrix. This means that $J'^{-1}\hat{F}_t$ is more efficient than $H'^{-1}\hat{F}_{ot}$, though these are infeasible estimators. The same is true of the GPCE and OPCE for the case of stationary factors as shown in Choi (2012).

Fourth, when $r = 1$, the GPCE and OPCE are estimating the same object up to sign, which makes it possible to compare efficiency directly. The limiting distribution of the GPCE is $\frac{1}{\sqrt{\int_0^1 B_F(r)^2 dr}} N\left(0, \Sigma_{\Lambda^*}^{-1}\right)$, while that of the OPCE is $\frac{1}{\sqrt{\int_0^1 B_F(r)^2 dr}} N\left(0, \Sigma_{\Lambda}^{-1}\Sigma_{\Lambda^*}\Sigma_{\Lambda}^{-1}\right)$. This shows that the GPCE is more efficient that the OPCE in the limit.

Asymptotic distribution of the GPCE of the factor-loading space is reported in the following theorem.

Theorem 2 *Let $\hat{\lambda}_i$ be the i-th column of $\hat{\Lambda}'$. Suppose that Assumptions 1, 2 and 3 hold. Then,*

$$T\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \xrightarrow{d} N(0, \sigma_{e_i}^2 I_r),$$

where $\sigma_{e_i}^2$ is the long-run variance of e_{it}, i.e., $\sigma_{e_i}^2 = \sum_{h=-\infty}^{\infty} E(e_{i1}e_{ih})$.

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. The independence of $\{e_t\}$ and $\{F_t\}$ assumed in Assumption 3 is essential for the asymptotic normality. Otherwise, a nonnormal distribution will ensue. The order of convergence of the GPCE and OPCE is commonly $\frac{1}{T}$ as Theorem 2 and Corollary 2 of Bai (2004) show. However, it is $\frac{1}{\sqrt{T}}$ for the factor model with stationary factors (cf. Bai, 2003 and Choi, 2012). This difference arises due to the presence of a unit root in the factors of this paper. Since the asymptotic distribution given in Corollary 2 of Bai (2004) can also be written as $N(0, \sigma_{e_i}^2 I_r)$, the GPCE and OPCE have the same level of efficiency in estimating the factor-loading space. The same is reported for the factor model with stationary factors (cf. Theorem 2 of Choi, 2012). The intuition behind these results is the equivalence of OLS and GLS in multivariate linear regression models (see Chapter 3 of Lütkephol, 1993). Because the factor-loadings are similar to regression coefficients of multivariate regression models and because the OPCE and GPCE roughly correspond to OLS and GLS, respectively, the asymptotic equivalence of the OPCE and GPCE is expected from the intuition.

The following theorem reports asymptotic distributions of the GPCE of the common component, $C_{it} = \lambda_i' F_t$. The GPCE of C_{it}, $\hat{C}_{it}$, is defined by $\hat{C}_{it} = \hat{\lambda}_i' \hat{F}_t$.

Theorem 3 *Suppose that Assumptions 1, 2 and 3 hold.*
(i) If $\frac{N}{T} \to c$ (a constant equal to or greater than zero), for each (i, t) such that $t = [T\tau]$,

$$\sqrt{N}(\hat{C}_{it} - C_{it}) \xrightarrow{d} N(0, \lambda_i' \Sigma_{\Lambda^*}^{-1}\lambda_i) + cB_F'(\tau)\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{-1}\left(\int_0^1 B_F(r)dB_e(r)dr\right).$$

(ii) If $\frac{N}{T} \to \infty$ *and* $\frac{N}{T^3} \to 0$*, for each* (i,t) *such that* $t = [T\tau]$*,*

$$\sqrt{T}(\hat{C}_{it} - C_{it}) \xrightarrow{d} B_F'(\tau) \left(\int_0^1 B_F(r) B_F'(r) dr \right)^{-1} \left(\int_0^1 B_F(r) dB_e(r) dr \right).$$

Even in the presence of nonstationary factors, the rate of convergence of the GPCE of the common component is either $\frac{1}{\sqrt{N}}$ or $\frac{1}{\sqrt{T}}$. Furthermore, depending on the rates of divergence of N and T, its limiting distribution is normal, mixture normal or a combination of normal and mixture normal distributions. Theorem 3 and Theorem 4 of Bai (2004) reveal that the GPCE is more efficient than the OPCE if $\frac{N}{T} \to c$ (a constant equal to or greater than zero). The efficiency gain of the GPCE comes from its efficient estimation of the factor space.

4 Extension to models with time polynomials

This section extends the distribution theory of the last section to models with time polynomials. We consider the models

$$X_t = \mu + \Lambda F_t + e_t, \ (t = 1, ..., T) \tag{6}$$

and

$$X_t = \mu + \beta t + \Lambda F_t + e_t, \ (t = 1, ..., T), \tag{7}$$

where μ and β are $N \times 1$ constant vectors. In matrix notation, these models can be written respectively as

$$\begin{aligned} X &= \mathbf{1}\mu' + F\Lambda' + e; \\ X &= \mathbf{1}\mu' + \Gamma\beta' + F\Lambda' + e, \end{aligned}$$

where $\mathbf{1} = [1, ..., 1]'$, $\Gamma = [1, 2, ..., T]'$ and X, F, Λ and e are as for model (1). Many macroeconomic time series are better represented by models (6) and (7) than model (1).

We first consider how to estimate the factor and factor-loading spaces for models (6) and (7) by the conditional maximum likelihood estimation method. To this end, assume $e_t \mid F \sim iid\ N(0, \Omega)$. Note that this assumption is used only to derive the GPCEs. Asymptotic results below in Theorems 4, 5 and 6 do not requires a normality assumption and are based on more general assumptions than this. The conditional log-likelihood function (multiplied by -2) for model (6) under the given assumption is:

$$-2l(\Lambda, F, \Omega) = TN\ln(2\pi) + T\ln|\Omega| + tr\left\{\Omega^{-1}(X - \mathbf{1}\mu' - F\Lambda')'(X - \mathbf{1}\mu' - F\Lambda')\right\}. \tag{8}$$

Assume for the moment that the matrix Ω is known. Then, we need to minimize

$$tr\left\{\Omega^{-1}(X - \mathbf{1}\mu' - F\Lambda')'(X - \mathbf{1}\mu' - F\Lambda')\right\} \tag{9}$$

with respect to μ, F and Λ. Using the standard theory of multivariate linear regression, we obtain

$$\hat{\Lambda} = X'(I_T - P_{\mathbf{1}})F\left(F'(I_T - P_{\mathbf{1}})F\right)^{-1} = X^{\mathbf{1}\prime}F^{\mathbf{1}}\left(F^{\mathbf{1}\prime}F^{\mathbf{1}}\right)^{-1},$$

with $X^{\mathbf{1}} = (I_T - P_{\mathbf{1}})X$ and $F^{\mathbf{1}} = (I_T - P_{\mathbf{1}})F$, and

$$\hat{\mu} = \bar{x} - \hat{\Lambda}\bar{f},$$

where $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} X_t$ and $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} F_t$. Plugging $\hat{\mu}$ and $\hat{\Lambda}$ successively into the objective function (9), we obtain

$$\begin{aligned}
& tr\left\{\Omega^{-1}\left(X - \mathbf{1}\bar{x}' - \left(F - \mathbf{1}\bar{f}'\right)\hat{\Lambda}'\right)'\left(X - \mathbf{1}\bar{x}' - \left(F - \mathbf{1}\bar{f}'\right)\hat{\Lambda}'\right)\right\} \\
= \;& tr\left\{\Omega^{-1}\left(X^{\mathbf{1}} - F^{\mathbf{1}}\hat{\Lambda}'\right)'\left(X^{\mathbf{1}} - F^{\mathbf{1}}\hat{\Lambda}'\right)\right\} \\
= \;& tr\left\{\Omega^{-1}\left(X^{\mathbf{1}} - F^{\mathbf{1}}\left(F^{\mathbf{1}}F^{\mathbf{1}}\right)^{-1}F^{\mathbf{1}\prime}X^{\mathbf{1}}\right)'\left(X^{\mathbf{1}} - F^{\mathbf{1}}\left(F^{\mathbf{1}}F^{\mathbf{1}}\right)^{-1}F^{\mathbf{1}\prime}X^{\mathbf{1}}\right)\right\} \\
= \;& tr\left\{\Omega^{-1}X^{\mathbf{1}\prime}(I_T - P_{F^{\mathbf{1}}})X^{\mathbf{1}}\right\}.
\end{aligned}$$

With the standardization $F^{\mathbf{1}\prime}F^{\mathbf{1}} = T^2 \times I_r$, the conditional maximum likelihood estimator of the demeaned factor space is obtained by maximizing $tr\left\{F^{\mathbf{1}\prime}\left(X^{\mathbf{1}}\Omega^{-1}X^{\mathbf{1}\prime}\right)F^{\mathbf{1}}\right\}$ with respect to $F^{\mathbf{1}}$. Therefore, the conditional maximum likelihood estimator of $F^{\mathbf{1}}$, denoted by $\hat{F}^{\mathbf{1}}$, is T times the matrix consisting of the eigenvectors corresponding to the r largest eigenvalues of the matrix $X^{\mathbf{1}}\Omega^{-1}X^{\mathbf{1}\prime}$. Note that $\hat{F}^{\mathbf{1}}$ estimates a rotation of $F^{\mathbf{1}}$, not that of F. In the case of zero-mean, stationary factors, $F^{\mathbf{1}}$ and F are essentially the same since $\bar{f} \xrightarrow{p} 0$. However, $\bar{f} \xrightarrow{d} \int_0^1 B_F(s)ds$ when $\{F_t\}$ is a unit root process satisfying Assumption 2 (iv), which makes $F^{\mathbf{1}}$ and F different even in the limit. In applications (e.g., forecasting regression and factor-augmented VAR), this does not seem to cause any complications. The conditional maximum likelihood estimator of Λ is given by $\hat{\Lambda} = \frac{1}{T^2}X^{\mathbf{1}\prime}\hat{F}^{\mathbf{1}}$.

For model (7), using the same argument as before, the conditional maximum likelihood estimator of the demeaned and detrended factor space is obtained by maximizing $tr\left\{F^{z\prime}\left(X^{z}\Omega^{-1}X^{z\prime}\right)F^{z}\right\}$ with respect to F^{z}, where $X^{z}=(I_T-P_z)X$ and $F^{z}=(I_T-P_z)F$ with $z=[\mathbf{1}\ \Gamma]$. The resulting estimator, $\hat{F}^{z}$, estimates the demeaned and detrended factor space F^{z}. In the case of zero-mean, stationary factors, F^{z} and F are essentially the same. When $\{F_t\}$ is a unit root process satisfying Assumption 2 (iv), F^{z} and F are different even in the limit. The conditional maximum likelihood estimator of Λ is given by $\hat{\Lambda}=\frac{1}{T^2}X^{z\prime}\hat{F}^{z}$.

The asymptotic results reported in the last section continue to hold for models (6) and (7) once some minor changes are introduced. The following theorem contains asymptotic results for models (6) and (7).

Theorem 4 *Suppose that Assumptions 1, 2 and 3 hold with F_t and $B_F(r)$ being replaced by $F_t^{\mathbf{1}}$ and $B_F^{\mathbf{1}}(r)$ and by F_t^{z} and $B_F^{z}(r)$ for models (6) and (7), respectively, where $B_F^{\mathbf{1}}(r)=B_F(r)-\int_0^1 B_F(r)dr$ and $B_F^{z}(r)=B_F(r)-4\left(\int_0^1 B_F(s)ds-\frac{3}{2}\int_0^1 sB_F(s)ds\right)+6r\left(\int_0^1 B_F(s)ds-2\int_0^1 sB_F(s)ds\right)$. Then, Theorems 1, 2 and 3 hold for models (6) and (7) if the Brownian motion $B_F(r)$ in those theorems is replaced by $B_F^{\mathbf{1}}(r)$ and $B_F^{z}(r)$, respectively, and if A is replaced by $A^{\mathbf{1}}=(I_T-P_{\mathbf{1}})A$ and $A^{z}=(I_T-P_z)A$ $(A=F,X)$, respectively.*

All the comments for Theorems 1, 2 and 3 also hold once changes in notation are made. One notable difference is that $\hat{F}^{\mathbf{1}}$ and $\hat{F}^{z}$ are estimating the factor spaces of $F^{\mathbf{1}}$ and F^{z}, not that of F. Likewise, the common component estimators estimate $\lambda_i^{\prime}F_t^{\mathbf{1}}$ and $\lambda_i^{\prime}F_t^{z}$, not $\lambda_i^{\prime}F_t$ and $\lambda_i^{\prime}F_t$.

5 Forecasting regressions

This section considers the forecasting regression equation

$$y_{t+h}=\alpha^{\prime}F_t+\beta^{\prime}Z_t+\varepsilon_{t+h},\ \ (t=1,...,T-h)\,, \tag{10}$$

where h is a forecasting horizon, $\{F_t\}$ are the factors from model (1), (6) or (7) and $\{Z_t\}$ are observable variables that may include a linear time trend and $I(1)$ variables. We estimate the parameters α and β using this model, and try to forecast y_{T+h} using the estimates. Since $\{F_t\}$ are not observed, we use their estimates to estimate model (10).

Suppose for the moment that the factors are estimated by using model (1). 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 $(\alpha' J'^{-1} \ \beta')'$ is obtained by regressing y_{t+h} on $(\hat{F}_t' \ Z_t')'$. The estimator from this regression is denoted as $\hat{\delta} = (\hat{\alpha}', \hat{\beta}')'$. Notice that the parameter vector α is not identified in this regression equation.

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

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 for all i.
(ii) $D_T^{-1} \sum_{t=1}^T L_t L_t' D_T^{-1} \xrightarrow{d} \Sigma_L \ (>0)$, where $D_T = D_{1T} \oplus D_{2T}$, $D_{1T} = diag[T, ..., T]$ and D_{2T} is a diagonal matrix whose elements are functions of T.
(iii) $D_T^{-1} \sum_{t=1}^T L_t \varepsilon_{t+h} \xrightarrow{d} \Sigma_{\varepsilon L}^{1/2} \times N(0, I)$, where $\Sigma_{\varepsilon L} > 0$ with probability one.

Part (ii) ensures that $D_T^{-1} \sum_{t=1}^{T-h} L_t L_t' D_T^{-1}$ is well behaved in the limit. Suppose that $\{Z_t\}$ are $I(0)$. Then, with $D_{2T} = diag[\sqrt{T}, ..., \sqrt{T}]$, we would have $\Sigma_L = \int_0^1 B_F(r) B_F'(r) dr \oplus \Sigma_z$ where Σ_z is the probability limit of $\frac{1}{T} \sum_{t=1}^T Z_t Z_t'$. If $\{Z_t\}$ are $I(1)$, $\Sigma_L = \int_0^1 B_{FZ}(r) B_{FZ}'(r) dr$ where $B_{FZ}(r) = (B_F'(r), B_Z'(r))'$ and $B_Z(r)$ is the weak limit of $\frac{1}{\sqrt{T}} Z_{[Tr]}$. In this case, $D_{2T} = diag[T, ..., T]$. If $Z_t = (1, t)'$, obviously $\Sigma_L = \begin{bmatrix} \int_0^1 B_F(r) B_F'(r) dr & \int_0^1 B_F(r) dr & \int_0^1 r B_F(r) dr \\ \int_0^1 B_F'(r) dr & 1 & \frac{1}{2} \\ \int_0^1 r B_F'(r) dr & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$ with $D_{2T} = diag[\sqrt{T} \ T^{3/2}]$. Part (iii) assumes the limiting distribution of $D_T^{-1} \sum_{t=1}^T L_t \varepsilon_{t+h}$. If $\{F_t\}$ are independent of $\{\varepsilon_{t+h}\}$ and if $\{Z_t, \ \varepsilon_{t+h}\}$ is a stationary and ergodic process, we have, with $D_{2T} = diag[\sqrt{T}, ..., \sqrt{T}]$,

$$\Sigma_{\varepsilon L} = \left(\int_0^1 B_F(r) B_F'(r) dr \right) \sigma_\varepsilon^2 \oplus \lim_{T \to \infty} \frac{1}{T} E\left(\sum_{t=1}^{T-h} Z_t \varepsilon_{t+h}\right)\left(\sum_{t=1}^{T-h} Z_t \varepsilon_{t+h}\right)',$$

where σ_ε^2 is the long-run variance of $\{\varepsilon_t\}$. If $\{F_t\}$ are independent of $\{\varepsilon_{t+h}\}$ and if $\{Z_t\}$ are $I(1)$ and independent of $\{\varepsilon_{t+h}\}$, we have

$$\Sigma_{\varepsilon L} = \left(\int_0^1 B_{FZ}(r) B_{FZ}'(r) dr \right) \sigma_\varepsilon^2,$$

with $D_{2T} = diag[T, ..., T]$. Other cases of $\{Z_t\}$ containing an intercept and/or a linear time trend can also be considered in a similar manner.

The following theorem reports asymptotic distribution of the OLS estimator of δ.

Theorem 5 *Let* $\delta = (\alpha' J^{-1}, \beta')'$ *and suppose that Assumptions 1, 2, 3 and 4 hold. Then, if* $\frac{T}{N} \to 0$,

$$D_T\left(\hat{\delta} - \delta\right) \xrightarrow{d} \Psi'^{-1}\Sigma_L^{-1}\Sigma_{\varepsilon L}^{1/2} \times N(0, I),$$

where $\Psi = R'^{-1} \oplus I$ *and* R *is defined right before equation 4.*

This theorem's proof utilizes the fact that $\hat{F}_t$ estimates the factor space consistently. Thus, any estimator of the factor space that retains the consistency property will yield the same result. Thus, the GPCE does not bring any efficiency gain over the OPCE in estimating the coefficient vector δ.

If the estimated factors are from model (6), the forecasting equation (10) can be rewritten as

$$y_{t+h} = \alpha' J^{\mathbf{1}\prime -1}\hat{F}_t^{\mathbf{1}} + \beta' Z_t + \varepsilon_{t+h} + \alpha' J^{\mathbf{1}\prime -1}\left(J^{\mathbf{1}\prime}F_t^{\mathbf{1}} - \hat{F}_t^{\mathbf{1}} + J^{\mathbf{1}\prime}\frac{1}{T}\sum_{t=1}^{T} F_t\right), \qquad (12)$$

where $\hat{F}_t^{\mathbf{1}}$ and $F_t^{\mathbf{1}}$ are the t-th row of $\hat{F}^{\mathbf{1}}$ and $F^{\mathbf{1}}$, respectively, and $J^{\mathbf{1}}$ is defined in the same manner as for J except that $\hat{F}^{\mathbf{1}}$, $F^{\mathbf{1}}$ and X^1 are used. Note that $F_t^{\mathbf{1}} + \frac{1}{T}\sum_{t=1}^{T} F_t = F_t$. If Z_t contains an intercept term, the effect of the term involving $\frac{1}{T}\sum_{t=1}^{T} F_t$ disappears and the asymptotic distribution of the OLS estimator of $\delta^{\mathbf{1}} = (\alpha' J^{\mathbf{1}\prime -1}, \beta')'$ from equation (12) is the same as that in Theorem 5 once the Brownian motion there is replaced by the demeaned Brownian motion $B_F^{\mathbf{1}}(r)$. Similar result involving $B_F^{z}(r)$ will hold when the factors are estimated by using model (7) as long as Z_t contains $(1\ t)'$. These discussions indicate that the GPCE does not bring any efficiency gain over the OPCE in estimating the coefficient vector of the forecasting regression equation when models (6) and (7) are used.

Now, we show that there is an efficiency gain for the forecasting error when the GPCE is used. Suppose that the factors are estimated by using model (1). Letting $\hat{L}_t = (\hat{F}_t'\ Z_t')'$, the forecast of y_{T+h} is denoted as $\hat{y}_{T+h|T} = \hat{\delta}'\hat{L}_T$. Then, the forecasting error is defined by

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

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

Theorem 6 *Suppose that Assumptions 1, 2, 3 and 4 hold. In addition, assume* $Var(\varepsilon_t) = \sigma_\varepsilon^2$ *for all* t; $E\left(\varepsilon_{t+h}\varepsilon_{T+h} \mid L_1, ..., L_{T-h}\right) = 0$ *for* $t = 1, ..., T-h$; *and that* h

is a fixed, positive integer. Then, if $\frac{T}{N} \to 0$ *and* $\frac{\sqrt{N}}{T} \to 0$*, the approximate variance of the forecasting error conditional on* $\{L_t\}_{t=1}^{T-h}$ *is*

$$L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\varepsilon L} \Sigma_L^{-1} D_T^{-1} L_T + \frac{1}{N} \alpha' \Sigma_{\Lambda^*}^{-1} \alpha + \sigma_\varepsilon^2. \tag{13}$$

The approximate variance of the OPCE-based forecasting error is as follows:

$$L_T' D_T^{-1} \Sigma_L^{-1} \Sigma_{\varepsilon L} \Sigma_L^{-1} D_T^{-1} L_T + \frac{1}{N} \alpha' \left[\Sigma_\Lambda^{-1} \Sigma_{\Lambda^*} \Sigma_\Lambda^{-1} \right] \alpha + \sigma_\varepsilon^2. \tag{14}$$

Relations (13) and (14) reveal that the GPCE of the factor space yields a smaller forecasting error variance. Obviously, this results from the efficiency gain of the GPCE. When the factors are estimated by using model (6) or (7), the same result as in Theorem 6 will hold except that $B(r)$ is replaced by $B_F^{\mathbf{1}}(r)$ or $B_F^{z}(r)$, respectively. Thus, there continues to be an efficiency gain for the forecasting error.

6 Feasible generalized principal component estimation

This section studies the FGPCE of the factor space. We focus only on the FGPCE of the factor space since the GPCE of the factor-loading space does not yield any efficiency gain over the OPCE. We assume for the moment that the factors are generated by model (1). Suppose that an estimator of Ω, $\hat{\Omega}$, is available. For example, letting $\hat{e}_{ot} = X_t - \hat{\Lambda}_o \hat{F}_{ot}$, where $\hat{\Lambda}_o$ and $\hat{F}_{ot}$ are the OPCEs of the factor loading Λ and the factor F_t, respectively, we may construct $\hat{\Omega}$ as $\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_{ot} \hat{e}_{ot}'$. When this estimator is used, assumptions for the consistency of the OPCEs should hold. Specific restrictions given to the structure of Ω should also apply to $\hat{\Omega}$. The FGPCE of the factor space, denoted as $\hat{F}^f$, is obtained by maximizing $tr\left\{F' \left(X \hat{\Omega}^{-1} X'\right) F\right\}$ with respect to F under the standardization $\hat{F}^{f\prime} \hat{F}^f = T^2 \times I_r$.

An immediate question about $\hat{F}^f$ is whether or not it has the same limiting distribution as the GPCE that uses the true variance-covariance matrix. Under the following assumption, we can establish asymptotic equivalence of the GPCE and FGPCE.

Assumption 5 *(i)* $\left\|\hat{\Omega}^{-1}\right\|_1 = O_p(1)$*. (ii)* $\left\|\hat{\Omega} - \Omega\right\|_1 \xrightarrow{p} 0$.
(iii) $\frac{1}{\sqrt{N}} \Lambda' \left(\hat{\Omega}^{-1} - \Omega^{-1}\right) e_t \xrightarrow{p} 0$.

Part (i) of this assumption is equivalent to 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. Part (ii) assumes that the largest eigenvalue of $\hat{\Omega}-\Omega$ converges to zero in probability. This implies that every element 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 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 $\hat{\Omega}$.

Assumption 5 can be verified when a special structure is given to the variance-covariance matrices Ω and $\hat{\Omega}$. When there is only heteroskedasticity, we would have $\Omega = diag[\sigma_1^2, ..., \sigma_N^2]$. We may also allow a block-diagonal structure for Ω such that $\Omega = \Omega_1 \oplus \Omega_2 ... \oplus \Omega_n$, where each Ω_k is an $a \times a$, positive definite matrix. Conditions for these structures to satisfy Assumption 5 are given in Choi (2012). In practice, structures of Ω can be found by pretesting and/or using economic theory. When there is a misspecification in the chosen structure of Ω, however, it is possible that the FGPCEs perform worse than the corresponding OPCEs.

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

Theorem 7 *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* $J^f = \frac{1}{NT^2}(\Lambda'\hat{\Omega}^{-1}\Lambda)(F'\hat{F}^f)\hat{W}_{NT}^{-1}$ *with* $\hat{W}_{NT}$ *being an* $r \times r$ *diagonal matrix consisting of the first* r *largest eigenvalues of* $\frac{1}{NT^2}X\hat{\Omega}^{-1}X'$ *in descending order.*

Because $J - J^f \xrightarrow{p} 0$ (cf. Lemmas A.8 and A.9), this theorem shows that $\hat{F}_t^f$ estimates essentially the same object as $\hat{F}_t$ and that they share the same limiting distribution. Thus, discussions after Theorem 1 about the GPCE and OPCE also apply to the FGPCE and OPCE. When the factors are generated by model (6) or (7), Theorem 7 still holds with some adaptations of its proof. The details do not seem to be worth reporting here.

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. The following theorem also holds when the factors are generated by model (6) or (7), once we make some trivial changes in its proof. The details are not reported here for brevity.

Theorem 8 *Suppose that assumptions for Theorem 6 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 6.*

7 Simulation

7.1 Efficiency comparison

This subsection reports simulation results that compare the efficiency of the OPCE, GPCE and FGPCE in estimating the factors $\{F_t\}$. As discussed in Section 3, GPCE provides higher efficiency in estimating the factors than OPCE.

Data were generated by the following data-generating process (DGP)

$$\begin{aligned} X_{it} &= \lambda_i' F_t \times \sqrt{\frac{\sigma_i^2}{1-\rho_i^2}} + e_{it}; \\ F_t &= F_{t-1} + u_t,\ u_t \sim iid\ N(0, I_r),\ u_0 = 0; \\ e_{it} &= \rho_i e_{it-1} + v_{it} \times \sqrt{\lambda_i' \lambda_i},\ v_{it} \sim iid\ N(0, \sigma_i^2), \end{aligned}$$

where $\{\lambda_i\}$ was generated according to the law of uniform distribution $U[0,1]$, $\{\rho_i\}$ was taken from $U[\phi_1, \phi_2]$ and $\{\sigma_i^2\}$ from $U[0,1]$. For $\{e_{it}\}$, we generated $T + 30$ observations and used the last T ones. Since $\{\lambda_i\}$, $\{\rho_i\}$ and $\{\sigma_i^2\}$ are assumed to be constants, they were fixed throughout iterations. Under this DGP,

$$\begin{aligned} Var\left(\lambda_i' F_t \times \sqrt{\frac{\sigma_i^2}{1-\rho_i}}\right) &= t\lambda_i' \lambda_i \times \left(\frac{\sigma_i^2}{1-\rho_i^2}\right); \\ Var\left(e_{it}\right) &= \lambda_i' \lambda_i \times \left(\frac{\sigma_i^2}{1-\rho_i^2}\right), \end{aligned}$$

so that the signal-to-noise ratio is always t for all the parameter values. In addition, the idiosyncratic errors are independent across i, serially correlated and unconditionally heteroskedastic. For $\{\rho_i\}$, we considered the three cases: $(\phi_1, \phi_2) = (0.1, 0.4), (0.3, 0.6), (0.5, 0.8)$. For sample sizes, we considered the cases $(T, N) = (50, 25),\ (100, 100),\ (50, 100),\ (100, 400)$. These are the combinations of the numbers of time series and cross-sectional observations we often encounter in reality.

The FGPCE is computed using the OPCE-based estimator of the variance–covariance matrix Ω as follows. (i) Calculate the OPCEs, $\hat{\lambda}_{oi}$ and $\hat{F}_{ot}$, and the estimates of the idiosyncratic errors $\hat{e}_{it} = X_{it} - \hat{\lambda}_{oi}'\hat{F}_{ot}$. (ii) Calculate $\hat{\omega}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}^2$. (iii) Use $\hat{\Omega} = diag(\hat{\omega}_i)_{i=1,...,N}$ for the calculation of the FGPCE.

In Table 1, we report R^2s from the regression using the estimated factor space as a dependent variable and a constant and $\{F_t\}$ as independent variables. We report the normalized R^2s that are obtained by dividing R^2s by those of the GPCE. For the GPCE, we also report unnormalized R^2s after "/" for the purpose of comparison. Results in Table 1 can be summarized as follows.

(i) The GPCE has higher R^2 than the OPCE in every case, confirming the efficiency gain reported in Theorem 1.

(ii) The FGPCE has higher R^2 than the OPCE in every case except $(T, N, r) = (50, 25, 8)$. The R^2s of the GPCE and FGPCE are comparable to each other in most cases.

(iii) For all the estimators, the R^2s decrease as the number of factors increases. This means that it becomes harder to estimate the factor space accurately as the number of factors increases.

(iv) At the same value of T, increasing N improves R^2s. Likewise, at the same value of N, increasing T improves R^2s.

(v) The values of ϕ_1 and ϕ_2 do not affect R^2s in a significant manner. But R^2s increase slightly as do the values of ϕ_1 and ϕ_2.

7.2 Mean squared forecasting errors

This subsection reports simulation results that compare mean squared forecasting errors. Data were generated by

$$y_{t+h} = \alpha' F_t + \beta y_t + \varepsilon_{t+h},\ (t = 1, ..., T)\,, \tag{15}$$

where $\alpha = [1, ..., 1]'$, $\beta = 0.1$,[1] $\varepsilon_t \sim iid\ N(0, 1)$ and $\{F_t\}$ were generated as in the last subsection. The initial values of the dependent variable are set to be zero. In this data generation, $\{F_t\}$ and $\{y_t\}$ are I(1). Furthermore, $\{e_{it}\}$, $\{w_t\}$ and $\{\varepsilon_t\}$ were generated to be independent. For $\{e_{it}\}$, we considered the same DGPs as in the last subsection. For sample sizes, we considered $(T, N) = (50, 25)$, $(100, 100)$, $(50, 100)$, $(100, 400)$ as in the last subsection. For the forecasting horizon, we tried $h = 4, 6$. The FGPCEs were computed as in the last subsection.

The target variable for forecasting is y_{T+h}, and data for $t = 1, ..., T-h$ are used to forecast it. The forecast of y_{T+h} is $\hat{\alpha}'\hat{F}_T + \hat{\beta} y_T$, where $\hat{\alpha}$ and $\hat{\beta}$ are the OLS estimates

[1] Different values of β were also tried, but gave no noticeable differences.

and $\hat{F}_T$ is an estimate of a space generated by F_T. The OPCE, GPCE and FGPCE are used to estimate the factor space.

The target variable can also be forecasted by using differenced data. That is, we may use the model

$$\Delta y_{t+m} = \alpha' \Delta F_t + \beta \Delta y_t + \Delta \varepsilon_{t+m}$$

to forecast Δy_{T+m} for $m = 1, ..., h$ and, then, the forecast of y_{T+h} becomes $\sum_{m=1}^{h} \Delta \hat{y}_{t+m}$ $+y_T$, where $\Delta \hat{y}_{t+m}$ denotes a forecast of Δy_{T+m}. To estimate $\{\Delta F_t\}$, $\{\Delta X_{it}\}$ are used. In fact, this method of forecasting has commonly been used in the literature on inflation rates (e.g., Stock and Watson, 1999). Inflation rates are I(1) in the U.S. and most European countries and their differences are often used to forecast inflation rates. We use the same data generated by equation (15), but use their differences both for the factor extraction and for forecasting. For the factor extraction we use the OPCE only.

Table 2 reports empirical mean squared forecasting errors for y_{T+h} (i.e., the empirical counterpart of $E(\hat{y}_{T+1|T} - y_{T+1})^2$) using the OPCE, GPCE, FGPCE and the OPCE using differenced data (denoted as OPCED in Table 2). We performed 2,000 iterations for the calculations of empirical mean squared forecasting errors. We report the normalized mean squared forecasting errors that are calculated by dividing the mean squared forecasting errors by those of the GPCE. For the GPCE, we also report unnormalized mean squared forecasting errors after "/" for the purpose of comparison. Results in Table 2 can be summarized as follows.

The results in Table 2 are summarized as follows.

(i) The GPCE improves empirical mean squared forecasting errors relative to the OPCE. The improvements are sometimes substantial (e.g., 1.000 versus 1.212). In addition, the OPCE tends to perform poorer than the GPCE as the number of factors increases.

(ii) The GPCE and FGPCE show similar performance in most cases and the FGPCE has lower mean squared forecasting errors than the OPCE in every case except $(h, T, N, r) = (4, 50, 25, 8)$.

(iii) The forecasts using the differenced data perform poorly relative to those using the level data. Empirical mean squared forecasting errors increase substantially in all the cases when the differenced data are used.

(iv) For all the estimators, mean squared forecasting errors increase as the number of factors increases.

(v) At the same value of T, increasing N improves mean squared forecasting errors.

Likewise, at the same value of N, increasing T improves mean squared forecasting errors.

(vi) The values of ϕ_1 and ϕ_2 do not affect mean squared forecasting errors in a significant manner. But they increase slightly as do the values of ϕ_1 and ϕ_2.

(vii) As h increases, mean squared forecasting errors increase at the same sample sizes and at the same number of factors.

8 Summary and further remarks

We have studied the GPCE for the nonstationary, static factor model. We derived asymptotic distributions of the GPCEs of the factor and factor-loading spaces, which showed that the GPCE enjoys an efficiency gain over the OPCE as in the case of stationary factors. These results were extended to the static factor models with time polynomials. The forecasting regression using the GPCE-based estimates of nonstationary factors was analyzed, the results of which showed that the GPCE yields more accurate forecasts than the OPCE. Last, asymptotic equivalence of the GPCE and feasible GPCE (FGPCE) of the factor space is established.

Simulation results confirm that the GPCE has an efficiency gain over the OPCE both in the estimation of common components and in forecasting exercises using the factor-augmented predictive regression model. Furthermore, it is shown via simulation that mean squared forecasting errors based on level data are much lower than those based on differenced ones when the factor-augmented predictive regression model is used. This experimental evidence needs to be examined further by using real data sets.

References

Altissimo, F., A. Bassanetti, R. Christadora, M. Forni, M. Hallin, M. Lippi & L. Reichlin (2001) EuroCOIN: A real time coincident indicator of the Euro area business cycle. CEPR Working Paper 3108.

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. (2004) Estimating cross–section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122, 137–183.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

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., and S. Ng (2010) Instrumental Variable Estimation in a Data Rich Environment. *Econometric Theory* 26, 1577–1606.

Banerjee, A. & M. Marcellino (2006) Are there any reliable leading indicators for US inflation and GDP growth?*International Journal of Forecasting* 22, 137–151.

Banerjee, A., M. Marcellino & I. Masten (2005) Leading indicators for euro–area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics* 67, 785–813.

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.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley & Sons.

Breitung, J. & I. Choi (2013) Factor models in *Handbook of Research Methods and Applications in Empirical Macroeconomics*, 249–265, Edward Elgar.

Breitung, J. & S. Eickmeier (2006) Dynamic factor models. Allgemeines Statistisches Archiv 90, 27–42.

Breitung, J. & J. Tenhofen (2011) GLS estimation of dynamic factor models. *Journal of the American Statistical Association* 495, 1150–1166.

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

Camba–Mendez, G. & G. Kapetanios (2005) Forecasting euro area inflation using dynamic factor measures of underlying inflation. *Journal of Forecasting* 24, 491–503.

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

Choi, I. (2012) Efficient estimation of factor models. *Econometric Theory* 28, 274–308.

Choi, I. & S.J. Hwang (2012) Forecasting Korean inflation. Mimeo, Sogang Research Institute of Market Economy (http://ideas.repec.org/s/sgo/wpaper.html).

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

Cristadoro, R., M. Forni, L. Reichlin & G. Veronese (2005) A core inflation indicator for the euro area. *Journal of Money, Credit and Banking* 37, 539–560.

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

den Reijer, A.H.J. (2005) Forecasting Dutch GDP using large scale factor models. DNB Working Papers 028, Netherlands Central Bank, Research Department.

Engel, C., Mark, N. C. & West, K. D. (2012). Factor model forecasts of exchange rates, Mimeo (No. w18382). National Bureau of Economic Research.

Favero, C.A., 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., D. Giannone, M. Lippi & L. Reichlin (2009) Opening the black box: structural factor models with large cross sections. *Econometric Theory* 25, 1319–1347.

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.

Giacomini, R. & H. White (2006) Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.

Giannone, D., L. Sala & L. Reichlin (2002) Tracking Greenspan: systematic and unsystematic monetary policy revisited. Mimeo, ECARES–ULB.

Giannone, D., L. Reichlin & L. Sala (2005) Monetary policy in real time. *NBER Macroeconomics Annual* 19, 161–200.

Greenaway–McGrevy, R., Han, C. & Sul, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. Journal of Econometrics 169, 48–53.

Greenaway–McGrevy, R., Mark, N. C., Sul, D. & Wu, J. L. (2012). Exchange rates as exchange rate common factors. Mimeo (No. 212012), Hong Kong Institure for Monetary Research.

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.

Kallenberg, O. (1997). *Foundations of Modern Probability.* New York: Springer.

Kapetanios, G. (2004) A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset. *Economics Letters* 85, 63–69.

Kapetanios, G. & M. Marcellino (2009) A Parametric estimation method for dynamic factor models of large fimensions. *Journal of Time Series Analysis* 30, 208–238.

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

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis.* Berlin: Springer–Verlag.

Lütkepohl, H. (1996) *Handbook of Matrices.* New York: John Wiley & Sons.

Marcellino, M., J.H. Stock & M.W. Watson (2003) Macroeconomic forecasting in the euro area: country specific versus area–wide information. *European Economic Review* 47, 1–18.

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

Park, J. Y., & Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. Econometric Theory, 4, 468–497.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

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

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

Sala, L. (2003) Monetary policy transmission in the euro area: a factor model approach. Mimeo, IGER Bocconi.

Schumacher, C. & C. Dreger. (2004). Estimating Large–Scale Factor Models for Economic Activity in Germany: Do they Outperform Simpler Models? *Jahrbücher für Nationalökonomie und Statitik*, Vol. 224, pp. 731–750.

Schumacher, C. (2007) Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting* 26, 271–302.

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.

Van der Vaart, A. W. (2000). *Asymptotic Statistics.* New York: Cambridge University Press.

Appendix I: 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 $\frac{1}{NT^2}X\Omega^{-1}X'$ in descending order. It satisfies the relation $\hat{F}=\frac{1}{NT^2}X\Omega^{-1}X'\hat{F}W_{NT}^{-1}$ by the definitions of eigenvalue and eigenvector. In addition, let $J=\frac{1}{NT^2}(\Lambda'\Omega^{-1}\Lambda)(F'\hat{F})W_{NT}^{-1}$. Then,

$$\begin{aligned}\hat{F}-FJ &= \frac{1}{NT^2}\left(X\Omega^{-1}X'\right)\hat{F}W_{NT}^{-1}-\frac{1}{NT^2}F(\Lambda'\Omega^{-1}\Lambda)(F'\hat{F})W_{NT}^{-1}\\ &= \frac{1}{NT^2}\left[X\Omega^{-1}X'-F(\Lambda'\Omega^{-1}\Lambda)F'\right]\hat{F}W_{NT}^{-1}\\ &= \frac{1}{NT^2}\left(e\Omega^{-1}e'+e\Omega^{-1}\Lambda F'+F\Lambda'\Omega^{-1}e'\right)\hat{F}W_{NT}^{-1}.\end{aligned}$$

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

$$\begin{aligned}\hat{F}_t-J'F_t &= \frac{1}{NT^2}W_{NT}^{-1}\hat{F}'\left(e\Omega^{-1}e_t+F\Lambda'\Omega^{-1}e_t+e\Omega^{-1}\Lambda F_t\right)\\ &= W_{NT}^{-1}\Big(\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t+\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda'\Omega^{-1}e_t\\ &\quad +\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}\Lambda F_t\Big)\\ &= W_{NT}^{-1}\left(a_{NTt}+b_{NTt}+c_{NTt}\right),\ \text{say,}\end{aligned}\tag{A.1}$$

where A_s' is the s-th row of the matrix A. When $\frac{N}{T^3}\to 0$, Lemma A.3 implies $\sqrt{N}(\hat{F}_t-J'F_t)=W_{NT}^{-1}\sqrt{N}b_{NTt}+o_p(1)$. Because $\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\xrightarrow{d}N(0,\Sigma_{\Lambda^*})$ by Assumption 3 (ii), Lemma A.4, Lemma A.5 and the continuous mapping theorem yield $W_{NT}^{-1}\sqrt{N}b_{NTt}\xrightarrow{d}W^{-1/2}N\left(0,I_r\right)$. Note that the marginal convergence results $\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}}\xrightarrow{d}N(0,\Sigma_{\Lambda^*})$ and $W_{NT}^{-1}\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_sF_s'\xrightarrow{d}W^{-1/2}\Sigma_{\Lambda^*}^{-1/2}$ imply the joint convergence result $\left(\frac{\Lambda'\Omega^{-1}e_t}{\sqrt{N}},W_{NT}^{-1}\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_sF_s'\right)\xrightarrow{d}\left(N(0,\Sigma_{\Lambda^*}),W^{-1/2}\Sigma_{\Lambda^*}^{-1/2}\right)$ as $N,T\to\infty$ under Assumption 3 (i) (see Theorem 4.5 of Billingsley, 1968, or Theorem 3.29 of Kallenberg, 1997), which is required to apply the continuous mapping theorem. Note also that W is independent of $N\left(0,I_r\right)$ again due to Assumption 3 (i).

(ii) Since $\limsup\frac{T^{3/2}}{\sqrt{N}}\le\frac{1}{\tau}<\infty$, $T^{3/2}a_{NTt}=O_p(1),T^{3/2}b_{NTt}=O_p(1)$ and $T^{3/2}c_{NTt}=o_p(1)$. The stated result follows from these.

Proof of Theorem 2: Since $\hat{\Lambda}=\frac{X'\hat{F}}{T^2}$, $\hat{\lambda}_i=\frac{\hat{F}F\lambda_i}{T^2}+\frac{\hat{F}\underline{e}_i}{T^2}$ where $\underline{e}_i=[e_{i1},...,e_{iT}]'$.

Writing $F = F - \hat{F}J^{-1} + \hat{F}J^{-1}$ and using $\frac{\hat{F}\hat{F}}{T^2} = I_r$, we obtain

$$\begin{aligned}\hat{\lambda}_i &= \frac{1}{T^2}\hat{F}'\left(F - \hat{F}J^{-1} + \hat{F}J^{-1}\right)\lambda_i + \frac{1}{T^2}\hat{F}'\underline{e}_i \\ &= J^{-1}\lambda_i + \frac{J'F'\underline{e}_i}{T^2} - \frac{1}{T^2}\hat{F}'\left(\hat{F} - FJ\right)J^{-1}\lambda_i + \frac{1}{T^2}\left(\hat{F} - FJ\right)'\underline{e}_i.\end{aligned} \tag{A.2}$$

Lemma A.6, below, shows that both the third and fourth terms are $o_p\left(\frac{1}{T}\right)$. Thus, the continuous mapping theorem gives

$$\begin{aligned}T\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) &= J'\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{1/2}\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{-1/2}\frac{1}{T}\sum_{s=1}^{T}F_s e_{is} + o_p(1) \\ &\xrightarrow{d} W^{-1}R\Sigma_{\Lambda^*}\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{1/2} N(0, \sigma_{e_i}^2 I_r) \qquad \text{(A.3)} \\ &= W^{-1/2}\Theta'\Sigma_{\Lambda^*}^{1/2}\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{1/2} N(0, \sigma_{e_i}^2 I_r),\end{aligned}$$

since $J \xrightarrow{p} \Sigma_{\Lambda^*}R'W^{-1}$ (R is defined in Lemma A.5 below). The joint convergence result $\left(\frac{1}{T^2}\sum_{t=1}^{T}F_tF_t', \frac{1}{T}\sum_{s=1}^{T}F_s e_{is}\right) \xrightarrow{d} \left(\int_0^1 B_F(r)B_F'(r)dr, \int_0^1 B_F(r)dB_{e_i}(r)\right)$ as $T \to \infty$, which is used for relation (A.3), is established in Park and Phillips (1988). Conditional on $B_F(\cdot)$, the limiting distribution can be written as

$$\begin{aligned}&N(0, W^{-1/2}\Theta'\Sigma_{\Lambda^*}^{1/2}\left(\int_0^1 B_F(r)B_F'(r)dr\right)\Sigma_{\Lambda^*}^{1/2}\Theta W^{-1/2}\sigma_{e_i}^2) \\ =\ &N(0, \sigma_{e_i}^2 I_r),\end{aligned}$$

where Θ is the eigenvector matrix of $\Sigma_{\Lambda^*}^{1/2}\left(\int_0^1 B_F(r)B_F'(r)dr\right)\Sigma_{\Lambda^*}^{1/2}$ such that $\Theta'\Theta = I_r$. But this representation does not depend on the conditioning and, hence, the result follows.

Proof of Theorem 3: We have

$$\begin{aligned}\hat{C}_{it} - C_{it} &= \hat{F}_t'\hat{\lambda}_i - F_t'\lambda_i \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + \hat{F}_t'\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + F_t'J\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) + \left(\hat{F}_t - J'F_t\right)'\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \\ &= \left(\hat{F}_t - J'F_t\right)'J^{-1}\lambda_i + F_t'J\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) + O_p\left(\frac{1}{T\min\{\sqrt{N}, T^{3/2}\}}\right),\end{aligned} \tag{A.4}$$

where Theorems 1 and 2 are used for the last equality. When $\frac{N}{T} \to c$, we also have $\frac{N}{T^3} \to 0$ so that part (i) of Theorem 1 holds. Thus, we obtain by using the continuous mapping theorem

$$\begin{aligned}
\sqrt{N}(\hat{C}_{it} - C_{it}) &= \sqrt{N}\left(\hat{F}_t - J'F_t\right)' J^{-1}\lambda_i + \left(\frac{1}{\sqrt{T}}F_t'\right) JT\left(\hat{\lambda}_i - J^{-1}\lambda_i\right)\sqrt{\frac{N}{T}} \\
&\quad + O_p\left(\frac{\sqrt{N}}{T\min\{\sqrt{N}, T^{3/2}\}}\right) \\
&\xrightarrow{d} N(0, \lambda_i'\Sigma_{\Lambda^*}^{-1}\lambda_i) + cB_F'(\tau)\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{-1} \\
&\quad \times \left(\int_0^1 B_F(r)dB_e(r)dr\right).
\end{aligned}$$

It is allowed to use the continuous mapping theorem due to parts (i) and (ii) of Assumption 3.[2] When $\frac{N}{T} \to \infty$ and $\frac{N}{T^3} \to 0$,

$$\begin{aligned}
\sqrt{T}(\hat{C}_{it} - C_{it}) &= \sqrt{\frac{T}{N}}\sqrt{N}\left(\hat{F}_t - J'F_t\right)' J^{-1}\lambda_i + \left(\frac{1}{\sqrt{T}}F_t'\right) JT\left(\hat{\lambda}_i - J^{-1}\lambda_i\right) \\
&\quad + O_p\left(\frac{\sqrt{T}}{T\min\{\sqrt{N}, T^{3/2}\}}\right) \\
&= \left(\frac{1}{\sqrt{T}}F_t'\right) JJ'\left(\frac{1}{T}\sum_{s=1}^{T} F_s e_{is}\right) + o_p(1) \\
&\xrightarrow{d} B_F'(\tau)\left(\int_0^1 B_F(r)B_F'(r)dr\right)^{-1}\left(\int_0^1 B_F(r)dB_e(r)dr\right),
\end{aligned}$$

since $JJ' \xrightarrow{d} \left(\int_0^1 B_F(r)B_F'(r)dr\right)^{-1}$.

Proof of Theorem 4: All the lemmas in Appendix II hold when F is replaced by $F^{\mathbf{1}} = (I_T - P_{\mathbf{1}})F$ and $F^z = (I_T - P_z)F$ and e by $(I_T - P_{\mathbf{1}})e$ and $(I_T - P_z)e$ for models (6) and (7), respectively. Thus, the result follows once it is noticed that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} F_t^{\mathbf{1}} \xrightarrow{d} B_F^{\mathbf{1}}(r)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} F_t^z \xrightarrow{d} B_F^z(r)$.

Proof of Theorem 5: Let $\hat{L}_t = (\hat{F}_t'\ Z_t')'$ and rewrite model (11) as

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

[2] Parts (i) and (ii) of Assumption 3 implies that $\left(\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t, \frac{1}{T}\sum_{s=1}^T F_s e_{is}, \frac{1}{T^2}\sum_{s=1}^T F_s F_s', \frac{1}{\sqrt{T}}F_s\right)$ has a joint weak limit.

Then, the OLS estimator of δ is written as

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

Note that the second equality follows from Lemma A.7. Using the relation

$$\begin{aligned} D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\varepsilon_{t+h} &= D_T^{-1}\begin{bmatrix}\sum_{t=1}^{T-h}\hat{F}_t\varepsilon_{t+h}\\ \sum_{t=1}^{T-h}Z_t\varepsilon_{t+h}\end{bmatrix} \\ &= \left(J'\oplus I\right)D_T^{-1}\begin{bmatrix}\sum_{t=1}^{T-h}F_t\varepsilon_{t+h}\\ \sum_{t=1}^{T-h}Z_t\varepsilon_{t+h}\end{bmatrix}+\begin{bmatrix}\frac{1}{T}\sum_{t=1}^{T-h}\left(\hat{F}_t-J'F_t\right)\varepsilon_{t+h}\\ 0\end{bmatrix}. \\ &\xrightarrow{d}\Psi\Sigma_{\varepsilon L}^{1/2}N\left(0,I\right), \end{aligned}$$

where $\Psi=R'^{-1}\oplus I$ (note that $J\xrightarrow{p}R^{-1}$), and Lemma A.7 (ii), we obtain the stated result.

Proof of Theorem 6: Using relation (11), we obtain

$$\begin{aligned} \hat{y}_{T+h|T}-y_{T+h} &= \left(\hat{\delta}-\delta\right)'\hat{L}_T+\alpha'J'^{-1}\left(\hat{F}_T-J'F_T\right)-\varepsilon_{T+h} \\ &= \left(\Psi'D_T\left(\hat{\delta}-\delta\right)\right)'\left(D_T^{-1}\Psi^{-1}\hat{L}_T\right) \\ &\quad +\frac{1}{\sqrt{N}}\alpha'J'^{-1}\left(\sqrt{N}\left(\hat{F}_T-J'F_T\right)\right)-\varepsilon_{T+h}. \end{aligned} \tag{A.5}$$

Theorem 5 gives $\Psi'D_T\left(\hat{\delta}-\delta\right)\xrightarrow{d}\Sigma_L^{-1}\Sigma_{\varepsilon L}^{1/2}N\left(0,I\right)$. Since $\hat{L}_T=\begin{bmatrix}J'F_T+O_p(\frac{1}{\sqrt{N}})\\ Z_T\end{bmatrix}$ and $J\xrightarrow{p}R^{-1}$, we also have

$$\Psi^{-1}\hat{L}_T=\begin{bmatrix}F_T+o_p(1)\\ Z_T\end{bmatrix}. \tag{A.6}$$

Thus, conditional on $\{L_t\}_{t=1}^{T-h}$, the first term in (A.5) has an approximate variance $L_T'D_T^{-1}\Sigma_L^{-1}\Sigma_{\varepsilon L}\Sigma_L^{-1}D_T^{-1}L_T$. 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 asymptotic distribution of the first term is driven by $\{L_t, \varepsilon_t\}_{t=1}^{T-h}$ and that of the second term by $\{e_t\}_{t=1}^{T}$. Thus, the first two terms are asymptotically uncorrelated. 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 $E\left(\varepsilon_{t+h}\varepsilon_{T+h} \mid L_1, ..., L_{T-h}\right) = 0$ for $t = 1, ..., T-h$ and because of the relation (A.6). It is also asymptotically uncorrelated with the second term due to Assumption 4. Thus, the result follows.

Proof of Theorem 7: We may write as in the proof of Theorem 1

$$\begin{aligned}
\hat{F}_t^f - J^{f\prime}F_t &= \frac{1}{NT^2}\hat{W}_{NT}^{-1}\hat{F}^f\left(e\hat{\Omega}^{-1}e_t + F\Lambda'\hat{\Omega}^{-1}e_t + e\Omega^{-1}\Lambda F_t\right) \\
&= \hat{W}_{NT}^{-1}(\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t + \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t \\
&\quad + \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t) \\
&= \hat{W}_{NT}^{-1}\left(\hat{a}_{NTt} + \hat{b}_{NTt} + \hat{c}_{NTt}\right), \text{ say.}
\end{aligned}$$

Lemma A.8 in Appendix II implies that $\hat{W}_{NT}^{-1} - W_{NT}^{-1} \xrightarrow{p} 0$. In addition, Lemma A.9 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 8: Because $J^f - J \xrightarrow{p} 0$ due to Lemma A.9 and because the FGPCE and GPCE have the same limiting distribution, the result can be proven by using the same methods as for Theorems 5 and 6.

Appendix II: Auxiliary Lemmas

Lemma A.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 $\kappa_{NT} = \min\{\sqrt{N}, T\}$.

Proof. Since

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - J'F_t\right\|^2 &\leq \left\|W_{NT}^{-1}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|a_{NTt} + b_{NTt} + c_{NTt}\right\|^2 \\
&\leq 3\left\|W_{NT}^{-1}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left(\left\|a_{NTt}\right\|^2 + \left\|b_{NTt}\right\|^2 + \left\|c_{NTt}\right\|^2\right)
\end{aligned}$$

and W_{NT} converges to a positive definite matrix as shown in Lemma A.4 below, we obtain the stated result by using Lemma A.2 given below.

Lemma A.2 *Assume Assumptions 1, 2 and 3.*
(i) $\frac{1}{T}\sum_{t=1}^{T}\left\|a_{NTt}\right\|^2 = O_p\left(\frac{1}{T^2}\right).$
(ii) $\frac{1}{T}\sum_{t=1}^{T}\left\|b_{NTt}\right\|^2 = O_p\left(\frac{1}{N}\right).$
(iii) $\frac{1}{T}\sum_{t=1}^{T}\left\|c_{NTt}\right\|^2 = O_p\left(\frac{1}{N}\right).$

Proof. (i) The Cauchy-Schwarz inequality applied to $\left\|a_{NTt}\right\|^2$ gives

$$\frac{1}{T}\sum_{t=1}^{T}\left\|a_{NTt}\right\|^2 \leq \frac{1}{T^2}\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{N}e_s'\Omega^{-1}e_t\right)^2\right).$$

The second term on the right-hand-side of this equation is r by definition. The third term is $O_p(1)$ due to Assumption 1 (iii).

(ii) Applying the Cauchy-Schwarz inequality to $\left\|b_{NTt}\right\|^2$, we obtain

$$\begin{aligned}\left\|b_{NT}\right\|^2 &= \frac{1}{N^2T^4}\left\|\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda'\Omega^{-1}e_t\right\|^2 \\ &\leq \frac{1}{N}\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\|^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right)\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|F_s\right\|^2\right).\end{aligned}$$

Thus, we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|b_{NTt}\right\|^2 \leq \frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\|^2\right)\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right)\left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|F_s\right\|^2\right), \tag{A.7}$$

from which we obtain the stated result by using Assumptions 2 and 3.

(iii) The multiplicative rule of the Euclidean norm yields

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left\|c_{NTt}\right\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_se_s'\Omega^{-1}\Lambda F_t\right\|^2 \\ &\leq \frac{1}{N}\left\|\frac{1}{T^{3/2}}\sum_{s=1}^{T}\hat{F}_s\left(\frac{1}{\sqrt{N}}e_s'\Omega^{-1}\Lambda\right)\right\|^2\left(\frac{1}{T^2}\sum_{t=1}^{T}\left\|F_t\right\|^2\right).\end{aligned} \tag{A.8}$$

The third term on the right-hand-side of equation (A.8) is $O_p(1)$ due to Assumption 2 (iv). Since

$$\left\|\frac{1}{T^{3/2}}\sum_{s=1}^{T}\hat{F}_s\left(\frac{1}{\sqrt{N}}e_s'\Omega^{-1}\Lambda\right)\right\|^2 \leq \left(\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right)\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{\sqrt{N}}e_s'\Omega^{-1}\Lambda\right\|^2\right] = O_p(1) \tag{A.9}$$

due to Assumption 3 (ii)-(a), the result follows.

Lemma A.3 *Assume Assumptions 1, 2 and 3.*
(i) $a_{NTt} = O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{\sqrt{NT}})$.
(ii) $b_{NTt} = O_p(\frac{1}{\sqrt{N}})$.
(iii) $c_{NTt} = O_p(\frac{1}{\sqrt{NT}})$.

Proof. (i) Write

$$\begin{aligned} a_{NTt} &= \frac{1}{T^2}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)\gamma_N(s,t) + J'\frac{1}{T^2}\sum_{s=1}^{T}F_s\gamma_N(s,t) \\ &\quad + \frac{1}{T^2}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right) \\ &\quad + J'\frac{1}{T^2}\sum_{s=1}^{T}F_s\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right) \\ &= a_{1NTt} + a_{2NTt} + a_{3NTt} + a_{4NTt}, \text{ say.} \end{aligned}$$

As in part (a) of Lemma B.2. of Bai (2004), we have $a_{1NTt} = O_p(\frac{1}{T^{3/2}\kappa_{NT}})$ and $a_{2NTt} = O_p(\frac{1}{T^{3/2}})$. Note that Assumption 2, Lemma A.4 and Lemma A.5 imply $J = O_p(1)$. For a_{3NTt}, we have

$$\begin{aligned} a_{3NTt} &\leq \frac{1}{T\sqrt{N}\kappa_{NT}}\sqrt{\frac{\kappa_{NT}^2}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left[\sqrt{N}\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right)\right]^2} \\ &= O_p\left(\frac{1}{T\sqrt{N}\kappa_{NT}}\right). \end{aligned}$$

Last, $a_{4NTt} = O_p(\frac{1}{\sqrt{NT}})$, due to the Cauchy-Schwarz inequality

$$\begin{aligned} &\frac{1}{T^2}\sum_{s=1}^{T}F_s\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right) \\ &\leq \frac{1}{\sqrt{NT}}\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\|F_s\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left[\sqrt{N}\left(\frac{1}{N}e_s'\Omega^{-1}e_t - \gamma_N(s,t)\right)\right]^2} \end{aligned}$$

and Assumptions 1 and 2.

(ii), (iii) These can be proven in the same manner as for Lemma B.2. (c), (d) of Bai (2004) under Assumptions 2 and 3.

Lemma A.4 *Assume Assumptions 1, 2 and 3. Then,*

$$W_{NT} = \frac{1}{T^2}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT^2}\right)\hat{F} \xrightarrow{p} W,$$

where W *is the diagonal matrix consisting of the eigenvalues of* $\Sigma_{\Lambda^*}\int_0^1 B_F(r)B_F'(r)dr$.

Proof. This can be proven using the same methods as in Stock and Watson (2002b; cf. relation R10) and Bai (2004).

Lemma A.5 *Assume Assumptions 1, 2 and 3. Then,*

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

where $W = diag[w_1, ..., w_r]$, $w_1 > ... > w_r > 0$ *are the eigenvalues of* $\Sigma_{\Lambda^*}^{1/2}\int_0^1 B_F(r)B_F'(r)dr\Sigma_{\Lambda^*}^{1/2}$ *and* Θ *is the corresponding eigenvector matrix such that* $\Theta'\Theta = I_r$.

Proof. This can be proven in the same manner as for Proposition 3 of Bai (2004).

Lemma A.6 *Assume Assumptions 1, 2 and 3. Then,*
(i) $\frac{1}{T^2}\left(\hat{F} - FJ\right)'\hat{F} = o_p\left(\frac{1}{T}\right)$;
(ii) $\frac{1}{T^2}\left(\hat{F} - FJ\right)'\underline{e}_i = o_p\left(\frac{1}{T}\right)$.

Proof. (i) Write

$$\frac{1}{T^2}\left(\hat{F} - FJ\right)'\hat{F} = \frac{1}{T^2}\left(\hat{F} - FJ\right)'FJ + \frac{1}{T^2}\left(\hat{F} - FJ\right)'(\hat{F} - FJ). \qquad \text{(A.10)}$$

Using (A.1), the first term on the right-hand-side of equation (A.10) is expressed as

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

Write A_{NT} as

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

Then,

$$\begin{aligned} \|A_{1NT}\| &\leq \frac{1}{T^4}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|\left(\sum_{t=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right|\right)\|F_t\| \\ &\leq \frac{1}{T^2}\max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{T}}F_t\right\|\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\left(\sum_{t=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right|\right)^2}. \end{aligned}$$

Using the the Cauchy-Schwarz inequality and Assumption 1 (iii), we obtain

$$\left(\sum_{t=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right|\right)^2 \leq T\sum_{t=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right|^2 = O_p(T).$$

Thus, Lemma A.1 and Assumption 2 (iv) imply $\|A_{1NT}\| = O_p(\frac{1}{\kappa_{NT}T^2})$. For A_{2NT}, we have

$$\begin{aligned} \|A_{2NT}\| &\leq \frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right|\left\|\frac{1}{\sqrt{T}}F_s\right\|\left\|\frac{1}{\sqrt{T}}F_t\right\| \\ &\leq \frac{1}{T^2}\left(\max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{T}}F_t\right\|\right)^2\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{N}e_s'\Omega^{-1}e_t\right| \\ &= O_p(\frac{1}{T^2}). \end{aligned}$$

Next, write

$$\begin{aligned} B_{NT} &= \frac{1}{NT^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)F_s'\Lambda'\Omega^{-1}e_tF_t' + J'\left(\frac{1}{NT^4}\sum_{t=1}^{T}\sum_{s=1}^{T}F_sF_s'\Lambda'\Omega^{-1}e_tF_t'\right) \\ &= B_{1NT} + J'B_{2NT}, \text{ say.} \end{aligned}$$

Lemma A.1, Assumptions 2 and 3 yield

$$\begin{aligned} \|B_{1NT}\| &\leq \frac{1}{T^{3/2}\sqrt{N}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\left\|\frac{1}{T\sqrt{N}}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_tF_t'\right\|^2\left(\frac{1}{T^2}\sum_{s=1}^{T}\|F_s\|^2\right)} \\ &= O_p(\frac{1}{\kappa_{NT}T^{3/2}\sqrt{N}}). \end{aligned} \tag{A.11}$$

Similarly,

$$B_{2NT} = \frac{1}{T\sqrt{N}} \left(\frac{1}{T^2} \sum_{s=1}^{T} F_s F_s' \right) \left(\frac{1}{T\sqrt{N}} \sum_{t=1}^{T} \Lambda' \Omega^{-1} e_t F_t' \right) = O_p \left(\frac{1}{T\sqrt{N}} \right). \quad \text{(A.12)}$$

Last, write

$$\begin{aligned} C_{NT} &= \frac{1}{NT^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{F}_s - J' F_s \right) e_s' \Omega^{-1} \Lambda F_t F_t' + J' \left(\frac{1}{NT^4} \sum_{t=1}^{T} \sum_{s=1}^{T} F_s e_s' \Omega^{-1} \Lambda F_t F_t' \right) \\ &= C_{1NT} + J' C_{2NT}, \text{ say.} \end{aligned}$$

Using the same methods as for B_{1NT} and B_{2NT}, we obtain $\|C_{1NT}\| = O_p \left(\frac{1}{\kappa_{NT} T\sqrt{N}} \right)$ and $C_{2NT} = O_p \left(\frac{1}{T\sqrt{N}} \right)$. Since $J = O_p(1)$ due to Assumption 2, Lemma A.4 and Lemma A.5 , the first term on the right-hand-side of relation (A.10) is $O_p\left(\frac{1}{T^2}\right) + O_p \left(\frac{1}{T\sqrt{N}} \right) = o_p(\frac{1}{T})$. The second term converges to zero in probability at a faster rate than the first term (cf. Theorem 1), and hence the result follows.

(ii) Using relation (A.1), write $\frac{1}{T^2} \left(\hat{F} - FJ \right)' \underline{e}_i$ as

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^{T} (\hat{F}_t - J' F_t) e_{it} &= W_{NT}^{-1} \left(\frac{1}{NT^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{F}_s e_s' \Omega^{-1} e_t e_{it} \right. \\ &\quad + \frac{1}{NT^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{F}_s F_s' \Lambda' \Omega^{-1} e_t e_{it} \\ &\quad \left. + \frac{1}{NT^4} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{F}_s e_s' \Omega^{-1} \Lambda F_t e_{it} \right) \\ &= W_{NT}^{-1} \left(D_{NT} + E_{NT} + G_{NT} \right), \text{ say.} \end{aligned}$$

D_{NT} can be written as

$$\begin{aligned} D_{NT} &= \frac{1}{NT^3} \sum_{s=1}^{T} \left(\hat{F}_s - J' F_s \right) \left(e_s' \Omega^{-1} \frac{1}{T} \sum_{t=1}^{T} e_t e_{it} \right) + J' \frac{1}{NT^3} \sum_{s=1}^{T} F_s e_s' \left(\frac{1}{T} \sum_{t=1}^{T} \Omega^{-1} e_t e_{it} \right) \\ &= D_{1NT} + D_{2NT}, \text{ say.} \end{aligned}$$

Let $\boldsymbol{\omega}_i$ be the i-th column of Ω. Then, due to Assumption 1 (i),

$$\left\| e_s' \Omega^{-1} \frac{1}{T} \sum_{t=1}^{T} e_t e_{it} - e_s' \Omega^{-1} \boldsymbol{\omega}_i \right\| \leq \left\| \Omega^{-1} \right\|_1 \|e_s\| \left\| \frac{1}{T} \sum_{t=1}^{T} e_t e_{it} - \boldsymbol{\omega}_i \right\| = O_p(\frac{N}{\sqrt{T}}). \quad \text{(A.13)}$$

Since $\Omega^{-1}\boldsymbol{\omega}_i = \mathbf{i}$, where $\mathbf{i}$ is an $N \times 1$ vector with the i-th element being equal to 1 and the rest zeros, $|e_s'\Omega^{-1}\boldsymbol{\omega}_i| = O_p(1)$. This and relation (A.13) imply $\left|e_s'\Omega^{-1}\frac{1}{T}\sum_{t=1}^T e_t e_{it}\right| = O_p(\max\{1, \frac{N}{\sqrt{T}}\})$. Thus,

$$\begin{aligned} \|D_{1NT}\| &\leq \frac{1}{NT^2}\sqrt{\frac{1}{T}\sum_{s=1}^T \left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^T \left|e_s'\Omega^{-1}\left(\frac{1}{T}\sum_{t=1}^T e_t e_{it}\right)\right|^2} \\ &= \frac{1}{NT^2} \times O_p\left(\frac{1}{\kappa_{NT}}\right) \times O_p(\max\{1, \frac{N}{\sqrt{T}}\}) \qquad (A.14) \\ &= O_p\left(\frac{1}{\kappa_{NT}\min\{NT^2, T^{5/2}\}}\right). \end{aligned}$$

Assumption 1 (i) yields

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

which implies

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

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

$$\begin{aligned} \|D_{2NT}\| &\leq \|J\| \frac{1}{NT^2}\left\|\frac{1}{T}\sum_{s=1}^T F_s e_s'\right\| \left\|\Omega^{-1}\frac{1}{T}\sum_{t=1}^T e_t e_{it}\right\| \qquad (A.15) \\ &= \frac{1}{NT^2} \times O_p\left(\sqrt{N}\right) \times O_p(\max\{1, \sqrt{\frac{N}{T}}\}). \end{aligned}$$

Thus, (A.14) and (A.15) yield

$$\|D_{NT}\| = O_p\left(\frac{1}{\kappa_{NT}\min\{NT^2, T^{5/2}\}}\right) + O_p\left(\frac{1}{T^2\min\{\sqrt{N}, \sqrt{T}\}}\right). \qquad (A.16)$$

Next, write

$$\begin{aligned} E_{NT} &= \frac{1}{NT^3}\sum_{s=1}^T \left(\hat{F}_s - J'F_s\right) F_s' \left(\frac{1}{T}\sum_{t=1}^T \Lambda'\Omega^{-1} e_t e_{it}\right) + J'\frac{1}{NT^3}\sum_{s=1}^T F_s F_s' \left(\frac{1}{T}\sum_{t=1}^T \Lambda'\Omega^{-1} e_t e_{it}\right) \\ &= E_{1NT} + E_{2NT}, \text{ say.} \end{aligned}$$

Then, since $\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_t e_{it} \xrightarrow{p} \lambda_i$ and $\|\lambda_i\| \leq \bar{\lambda} < \infty$ due to Assumption 2 (iii),

$$\begin{aligned}\|E_{1NT}\| &\leq \frac{1}{NT^{3/2}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - J'F_s\right\|^2}\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\|F_s\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_t e_{it}\right\|^2} \\ &= O_p\left(\frac{1}{NT^{3/2}\kappa_{NT}}\right). \end{aligned} \tag{A.17}$$

Similarly,

$$\begin{aligned} E_{2NT} &= \frac{1}{NT}J'\left(\frac{1}{T^2}\sum_{s=1}^{T}F_sF_s'\right)\frac{1}{T}\sum_{t=1}^{T}\Lambda'\Omega^{-1}e_t e_{it} \\ &= O_p\left(\frac{1}{NT}\right). \end{aligned} \tag{A.18}$$

Thus, we obtain

$$\|E_{NT}\| = O_p\left(\frac{1}{NT}\right) \tag{A.19}$$

from (A.17) and (A.18). Writing

$$\begin{aligned} G_{NT} &= \frac{1}{NT^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{F}_s - J'F_s\right)e_s'\Omega^{-1}\Lambda F_t e_{it} + \frac{1}{NT^4}\sum_{t=1}^{T}\sum_{s=1}^{T}J'F_s e_s'\Omega^{-1}\Lambda F_t e_{it} \\ &= G_{1NT} + G_{2NT}, \text{ say,} \end{aligned}$$

and using Lemma A.1 and Assumption 3 (ii), we obtain

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

and

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

which give

$$\|G_{NT}\| = O_p\left(\frac{1}{\sqrt{N}T^2}\right). \tag{A.20}$$

The result follows from (A.16), (A.19) and (A.20), since $O_p\left(\frac{1}{\kappa_{NT}\min\{NT^2,T^2\sqrt{T}\}}\right)+O_p\left(\frac{1}{T^2\min\{\sqrt{N},\sqrt{T}\}}\right)+O_p\left(\frac{1}{NT}\right)+O_p\left(\frac{1}{\sqrt{N}T^2}\right)=o_p\left(\frac{1}{T}\right)$.

Lemma A.7 *Assume Assumptions 1, 2, 3 and 4. Then, if* $\frac{T}{N}\to 0$,
(i) $D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(J'F_t-\hat{F}_t\right)'J^{-1}\alpha\xrightarrow{p}0$;
(ii) $D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1}\xrightarrow{d}\Psi\Sigma_L\Psi'$, *where* $\Psi=R'^{-1}\oplus I$.

Proof. (i) Let $M_t=\begin{pmatrix}J'F_t\\Z_t\end{pmatrix}$ and write

$$\begin{aligned}&D_T^{-1}\sum_{t=1}^{T-h}\hat{L}_t\left(J'F_t-\hat{F}_t\right)'J^{-1}\alpha\\=\;&D_T^{-1}\sum_{t=1}^{T-h}\left(\hat{L}_t-M_t\right)\left(J'F_t-\hat{F}_t\right)'J^{-1}\alpha\\&+D_T^{-1}\sum_{t=1}^{T-h}M_t\left(J'F_t-\hat{F}_t\right)'J^{-1}\alpha\\=\;&K_1+K_2,\text{ say.}\end{aligned}$$

Since $\hat{F}_t-J'F_t=O_p(\frac{1}{\sqrt{N}})$ as shown in Theorem 1,

$$K_1=-\begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T-h}\left(\hat{F}_t-J'F_t\right)\left(\hat{F}_t-J'F_t\right)'J^{-1}\alpha\\0\end{pmatrix}\xrightarrow{p}0.\tag{A.21}$$

Partition K_2 such Let $K_2=\begin{pmatrix}K_{21}\\K_{22}\end{pmatrix}$, where K_{21} corresponds to $J'F_t$ and K_{22} to Z_t. Then, since $\frac{1}{\sqrt{T}}F_t=O_p(1)$ for any t and $\hat{F}_t-J'F_t=O_p(\frac{1}{\sqrt{N}})$, we have under $\frac{T}{N}\to 0$,

$$K_{21}=J'\left(\frac{1}{T}\sum_{t=1}^{T-h}F_t\left(J'F_t-\hat{F}_t\right)'\right)J^{-1}\alpha\xrightarrow{p}0.\tag{A.22}$$

In addition,

$$\begin{aligned}\|K_{22}\|&=\left\|D_{2T}^{-1}\sum_{t=1}^{T-h}Z_t\left(J'F_t-\hat{F}_t\right)'\right\|\\&\le\left\|D_{2T}^{-1}\sum_{t=1}^{T-h}Z_tZ_t'D_{2T}^{-1}\right\|\left\|\sum_{t=1}^{T-h}\left(\hat{F}_t-J'F_t\right)\left(\hat{F}_t-J'F_t\right)'\right\|\\&=O_p(\frac{T}{N}).\end{aligned}\tag{A.23}$$

The stated result follows from relations (A.21), (A.22) and (A.23).

(ii) Let $\Gamma = J' \oplus I$ and write $\hat{L}_t = \Gamma L_t + \hat{L}_t - \Gamma L_t$. Then,

$$
\begin{aligned}
D_T^{-1}\left(\sum_{t=1}^{T-h}\hat{L}_t\hat{L}_t'\right)D_T^{-1} &= \Gamma D_T^{-1}\left(\sum_{t=1}^{T-h}L_tL_t'\right)D_T^{-1}\Gamma' \\
&\quad + D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t-\Gamma L_t\right)\left(\hat{L}_t-\Gamma L_t\right)'\right)D_T^{-1} \\
&\quad + \Gamma D_T^{-1}\left(\sum_{t=1}^{T-h}L_t\left(\hat{L}_t-\Gamma L_t\right)'\right)D_T^{-1} \\
&\quad + D_T^{-1}\left(\sum_{t=1}^{T-h}\left(\hat{L}_t-\Gamma L_t\right)L_t'\right)D_T^{-1}\Gamma' \\
&= P_1+P_2+P_3+P_3', \text{ say.}
\end{aligned}
$$

Note that $\Gamma D_T^{-1} = D_T^{-1}\Gamma$ thanks to the special structures of the matrices Γ and D_T. Theorem 1 and Assumption 2 yield $P_2 = o_p(1)$. In addition, $P_3 = o_p(1)$ as for (A.22) and (A.23). Thus, the result follows from Assumption 4 and the relation $\Gamma \xrightarrow{p} \Psi$.

Lemma A.8 *Assume Assumptions 1, 2, 3 and 5. Let* $\hat{W}_{NT} = \frac{1}{T^2}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT^2}\right)\hat{F}^f$. *Then,* $\hat{W}_{NT} - W_{NT} \xrightarrow{p} 0$.

Proof. Write

$$
\begin{aligned}
\left\|\hat{W}_{NT}-W_{NT}\right\| &= \left\|\frac{1}{T^2}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT^2}\right)\hat{F}^f - \frac{1}{T^2}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT^2}\right)\hat{F}\right\| \\
&\le \left\|\frac{1}{T^2}\hat{F}^{f\prime}\left(\frac{X\hat{\Omega}^{-1}X'}{NT^2}-\frac{X\Omega^{-1}X'}{NT^2}\right)\hat{F}^f\right\| \\
&\quad + \left\|\frac{1}{T^2}\hat{F}^{f\prime}\left(\frac{X\Omega^{-1}X'}{NT^2}\right)\hat{F}^f - \frac{1}{T^2}\hat{F}'\left(\frac{X\Omega^{-1}X'}{NT^2}\right)\hat{F}\right\| \\
&= I + II. \qquad \text{(A.24)}
\end{aligned}
$$

For the first term, we have an inequality

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

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

$$\begin{aligned}
\left\|\frac{X\hat{\Omega}^{-1}X'}{NT^2} - \frac{X\Omega^{-1}X'}{NT^2}\right\|_1 &= \left\|\frac{1}{NT^2}X\left(\hat{\Omega}^{-1} - \Omega^{-1}\right)X'\right\|_1 \\
&= \left\|\frac{1}{NT^2}\left(\hat{\Omega}^{-1} - \Omega^{-1}\right)X'X\right\|_1 \\
&\leq \left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_1 \left\|\frac{X'X}{NT^2}\right\|_1 \\
&\leq \left\|\hat{\Omega}^{-1}\right\|_1 \left\|\Omega^{-1}\right\|_1 \left\|\hat{\Omega} - \Omega\right\|_1 \left\|\frac{X'X}{NT^2}\right\|_1, \quad \text{(A.25)}
\end{aligned}$$

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 (A.25) is $O_p(1)$ due to Lemma B.3. of Bai (2004), Assumptions 1 and 5 imply that $\left\|\frac{X\hat{\Omega}^{-1}X'}{NT^2} - \frac{X\Omega^{-1}X'}{NT^2}\right\|_1 = o_p(1)$. Thus, $I = o_p(1)$. For the second term in relation (A.24), we have

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

Since $\left\|\hat{F}^f - \hat{F}\right\| = o_p(\sqrt{T})$ as can be deduced from relation (A.33) below and $\left\|\frac{X\Omega^{-1}X'}{NT^2}\right\|_1 = O_p(1)$ as shown above, $II = o_p(1)$. Thus, the stated result follows.

Lemma A.9 *Let* $\hat{a}_{NTt} = \frac{1}{NT^2}\sum_{s=1}^T \hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t$, $\hat{b}_{NTt} = \frac{1}{NT^2}\sum_{s=1}^T \hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t$ *and* $\hat{c}_{NTt} = \frac{1}{NT^2}\sum_{s=1}^T \hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t$, *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) $\hat{c}_{NTt} - c_{NTt} \xrightarrow{p} 0$;
(iv) $\frac{1}{T^2}\sum_{s=1}^T \hat{F}_s^f F_s' \xrightarrow{p} W^{1/2}\Theta'\Sigma_{\Lambda^*}^{-1/2}$, *where* W *and* Θ *are defined in Lemma A.5;*
(v) $\sqrt{N}\left(\hat{b}_{NTt} - b_{NTt}\right) \xrightarrow{p} 0$.

Proof. (i) First, consider the relation

$$\begin{aligned}
&\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t - \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t\right\| \\
= &\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\right\| \\
\leq &\frac{1}{NT^2}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|\left\|e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\right\| \\
\leq &\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|^2}\sqrt{\frac{1}{N^2T^2}\sum_{s=1}^{T}\left\|e_s'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\right\|^2},
\end{aligned} \tag{A.26}$$

where the Cauchy-Schwarz inequality is used for the last inequality. The first term in (A.26) is $\sqrt{r}$. The second term is less than $\sqrt{\frac{1}{N^2T^2}\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1^2\|e_t\|^2\sum_{s=1}^{T}\|e_s\|^2}$. Since $\|e_t\|^2 = O_p(N)$ and $\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1 \leq \left\|\hat{\Omega}^{-1}\right\|_1\left\|\hat{\Omega}-\Omega\right\|_1\|\Omega^{-1}\|_1 = o_p(1)$ as in the proof of Lemma A.8, the second term is $o_p(1)$. Thus, we have

$$\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}e_t - \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t\right\| \xrightarrow{p} 0. \tag{A.27}$$

Next, the Cauchy-Schwarz inequality gives

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

The first term can be shown to be bounded above by a positive constant using the triangle inequality and the fact that $\sum_{s=1}^{T}\hat{F}_s^f\hat{F}_s^{f\prime} = \sum_{s=1}^{T}\hat{F}_s\hat{F}_s' = T^2\times I_r$. The second one can be shown to be $o_p(1)$ under Assumption 1 by using the same method as for the proof of Lemma A.3 (i). Thus,

$$\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}e_t - \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s e_s'\Omega^{-1}e_t\right\| \xrightarrow{p} 0. \tag{A.28}$$

The stated result follows from (A.27) and (A.28), once the triangular inequality is used.

(ii) Write

$$\begin{aligned}
\hat{b}_{NTt}-b_{NT} &= \frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t-\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t\\
&= \left(\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\hat{\Omega}^{-1}e_t-\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\Omega^{-1}e_t\right)\\
&\quad+\left(\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'\Lambda'\Omega^{-1}e_t-\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s F_s'\Lambda'\Omega^{-1}e_t\right)\\
&= I+II. \qquad\qquad (A.29)
\end{aligned}$$

Since $\frac{1}{\sqrt{N}}\Lambda'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)e_t\xrightarrow{p}0$ and $\frac{1}{T^2}\sum_{s=1}^{T}\hat{F}_s^f F_s'=O_p(1)$ as shown in part (iv) of this lemma, $I=o_p(1)$. Moreover,

$$\begin{aligned}
\|II\| &\le \frac{1}{\sqrt{N}}\left\|\frac{1}{T^2}\sum_{s=1}^{T}\left(\hat{F}_s^f-\hat{F}_s\right)F_s'\right\|\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\|\\
&\le \frac{1}{\sqrt{N}}\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s^f-\hat{F}_s\right\|^2}\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\|F_s\|^2}\left\|\frac{1}{\sqrt{N}}\Lambda'\Omega^{-1}e_t\right\|\xrightarrow{p}0.
\end{aligned} \qquad (A.30)$$

Thus, the stated result follows.

(iii) Similarly to previous parts, we obtain

$$\begin{aligned}
&\left\|\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\hat{\Omega}^{-1}\Lambda F_t-\frac{1}{NT^2}\sum_{s=1}^{T}\hat{F}_s^f e_s'\Omega^{-1}\Lambda F_t\right\|\\
\le\ &\left\|\frac{1}{\sqrt{N}T^{3/2}}\sum_{s=1}^{T}\hat{F}_s^f e_s'\right\|\left\|\frac{1}{\sqrt{N}}\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)\Lambda\left(\frac{F_t}{\sqrt{T}}\right)\right\|\\
\le\ &\sqrt{\frac{1}{T^2}\sum_{s=1}^{T}\left\|\hat{F}_s^f\right\|^2}\sqrt{\frac{1}{NT}\sum_{s=1}^{T}\|e_s\|^2}\left\|\hat{\Omega}^{-1}-\Omega^{-1}\right\|_1\left\|\frac{1}{\sqrt{N}}\Lambda\right\|\sup_t\left\|\frac{F_t}{\sqrt{T}}\right\|\\
=\ &o_p(1)
\end{aligned}$$

and

$$
\begin{aligned}
& \left\| \frac{1}{NT^2} \sum_{s=1}^{T} \hat{F}_s^f e_s' \Omega^{-1} \Lambda F_t - \frac{1}{NT^2} \sum_{s=1}^{T} \hat{F}_s e_s' \Omega^{-1} \Lambda F_t \right\| \\
= & \left\| \frac{1}{NT^2} \sum_{s=1}^{T} \left(\hat{F}_s^f - \hat{F}_s \right) e_s' \Omega^{-1} \Lambda F_t \right\| \\
\leq & \frac{1}{\sqrt{N}} \left(\frac{1}{T^{3/2}} \sum_{s=1}^{T} \left\| \hat{F}_s^f - \hat{F}_s \right\| \left\| \frac{1}{\sqrt{N}} e_s' \Omega^{-1} \Lambda \right\| \right) \sup_t \left\| \frac{F_t}{\sqrt{T}} \right\| \\
\leq & \frac{1}{\sqrt{N}} \left(\sqrt{\frac{1}{T^2} \sum_{s=1}^{T} \left\| \hat{F}_s^f - \hat{F}_s \right\|^2} \sqrt{\frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{N}} e_s' \Omega^{-1} \Lambda \right\|^2} \right) \sup_t \left\| \frac{F_t}{\sqrt{T}} \right\| \\
= & o_p(1).
\end{aligned}
$$

These imply the stated result.

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

$$
\begin{aligned}
& \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{1/2} \left(\frac{F'F}{T^2} \right) \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right) \left(\frac{F' \hat{F}^f}{T^2} \right) + d_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) \qquad \text{(A.31)} \\
= & \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{1/2} \left(\frac{F' \hat{F}^f}{T^2} \right) \hat{W}_{NT},
\end{aligned}
$$

where

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

Let $B_{NT}(\hat{\Omega}^{-1}) = \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{1/2} \left(\frac{F'F}{T^2} \right) \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{1/2}$, $R_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) = \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{1/2} \left(\frac{F' \hat{F}^f}{T^2} \right)$ and W_{NT}^* be a diagonal matrix consisting of the diagonal elements of $R_{NT}'(\hat{\Omega}^{-1}, \hat{F}^f) R_{NT}(\hat{\Omega}^{-1}, \hat{F}^f)$. Then, denoting $\Theta_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) = R_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) W_{NT}^{*-1/2}$, equation (A.31) can be written as

$$
\left(B_{NT}(\hat{\Omega}^{-1}) + d_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) R_{NT}^{-1}(\hat{\Omega}^{-1}, \hat{F}^f) \right) \Theta_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) = \Theta_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) \hat{W}_{NT}. \qquad \text{(A.32)}
$$

Since $\frac{F' \hat{F}^f}{T} = \left(\frac{\Lambda' \hat{\Omega}^{-1} \Lambda}{N} \right)^{-1/2} \Psi_{NT}(\hat{\Omega}^{-1}) W_{NT}^{*1/2}$, Bai's method amounts to obtaining the probability limit of $\frac{F' \hat{F}^f}{T^2}$ by treating (A.32) as an approximate eigenvalue-eigenvector relation and solving it with respect to $\Psi_{NT}(\hat{\Omega}^{-1})$. Thus, once we show $B_{NT}(\hat{\Omega}^{-1}) - B_{NT}(\Omega^{-1}) \xrightarrow{p} 0$, $d_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) - d_{NT}(\Omega^{-1}, \hat{F}) \xrightarrow{p} 0$, $R_{NT}(\hat{\Omega}^{-1}, \hat{F}^f) = O_p(1)$ and $\hat{W}_{NT} - W_{NT} \xrightarrow{p} 0$, the stated result follows from Lemma A.5. 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 $\frac{F'\hat{F}^f}{T^2} = O_p(1)$. The fourth relation is proven in Lemma A.8.

(v) It suffices to show that the two terms, I and II, in relation (A.29) 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

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

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 A.8, we also find the third term to be $o_p(1)$. Thus,

$$\left\|\hat{F}_s^f - \hat{F}_s\right\| = o_p(1), \tag{A.33}$$

which and other assumptions imply $\sqrt{N}II = o_p(1)$ (cf. relation (A.30)).

Table 1. Efficiency Comparison for the Estimation of Factors

Note: 1. Data were generated by $X_{it} = \lambda_i' F_t \times \sqrt{\frac{\sigma_i^2}{1-\rho_i^2}} + e_{it}$ $(i = 1, ..., N;\ t = 1, ..., T)$; $F_t = F_{t-1} + u_t,\ u_t \sim iid\ N(0, I_r)$; $e_{it} = \rho_i e_{it-1} + v_{it} \times \sqrt{\lambda_i' \lambda_i},\ v_{it} \sim iid\ N(0, \sigma_i^2)$, where $\{\lambda_i\}$ was generated by the law of standard normal distribution, $\{\rho_i\}$ was taken from the uniform distribution $U[\phi_1, \phi_2]$ and $\{\sigma_i^2\}$ from $U[0, 1]$. 2. Parameter r denotes the number of factors and is assumed to be known. 3. Entries in this table denote the normalized R^2s from the regression using the estimated factor space as a dependent variable and a constant and $\{F_t\}$ as independent variables. The normalization is made such the R^2s of the GPCE are equal to one. For the GPCE, unnormalized R^2s are also reported after "/". 4. The number of iterations is set to be 2,000.

			$(\phi_1, \phi_2) = (0.1, 0.4)$			$(\phi_1, \phi_2) = (0.3, 0.6)$			$(\phi_1, \phi_2) = (0.5, 0.8)$		
T	N	r	OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE	OPCE	GPCE	FGPCE
50	25	1	0.9932	1.000/0.9864	1.005	0.9959	1.000/0.9881	1.004	0.9944	1.000/0.9862	1.006
		3	0.8763	1.000/0.9025	0.9532	0.9334	1.000/0.8506	0.9971	0.9148	1.000/0.8879	0.9651
		5	0.8703	1.000/0.7605	0.8958	0.8938	1.000/0.7440	0.9054	0.9082	1.000/0.7880	0.9316
		8	0.9067	1.000/0.6212	0.8917	0.8945	1.000/0.6329	0.8845	0.9487	1.000/0.6861	0.9488
100	100	1	0.9984	1.000/0.9979	1.001	0.9985	1.000/0.9982	1.001	0.9987	1.000/0.9981	1.001
		3	0.9897	1.000/0.9841	1.002	0.9858	1.000/0.9847	1.002	0.9856	1.000/0.9864	1.002
		5	0.9612	1.000/0.9585	1.002	0.9538	1.000/0.9582	1.004	0.9652	1.000/0.9608	1.002
		8	0.9276	1.000/0.8905	0.9901	0.9295	1.000/0.8973	0.9949	0.9384	1.000/0.8750	0.9935
50	100	1	0.9974	1.000/0.9964	1.001	0.9973	1.000/0.9969	1.001	0.9972	1.000/0.9968	1.001
		3	0.9753	1.000/0.9678	1.006	0.9671	1.000/0.9680	1.003	0.9595	1.000/0.9568	1.005
		5	0.9753	1.000/0.9028	1.006	0.9278	1.000/0.8933	0.9988	0.9701	1.000/0.8747	1.004
		8	0.9379	1.000/0.7777	1.008	0.9279	1.000/0.7869	0.9917	0.9549	1.000/0.8010	0.9956
100	400	1	0.9996	1.000/0.9995	1.000	0.9996	1.000/0.9995	1.000	0.9996	1.000/0.9995	1.000
		3	0.9968	1.000/0.9960	1.001	0.9974	1.000/0.9960	1.001	0.9964	1.000/0.9952	1.001
		5	0.9905	1.000/0.9888	1.001	0.9915	1.000/0.9891	1.001	0.9920	1.000/0.9853	1.002
		8	0.9764	1.000/0.9713	1.002	0.9768	1.000/0.9641	1.003	0.9727	1.000/0.9469	1.001

Table 2. Mean Squared Errors of Forecasting

Note: 1. Data were generated by $y_{t+h} = \alpha' F_t + \beta y_t + \varepsilon_{t+h}$, $(t = 1, ..., T-h)$,where $\alpha = [1, ..., 1]'$, $\beta = 0.1$, $y_1 = 0$, $\varepsilon_t \sim iid\ N(0,1)$ and $\{F_t\}$ were generated as in Table 1. 2. Entries in this table denote the normalized empirical mean squared errors from the forecasting regression $y_{t+h} = \hat{\alpha}' \hat{F}_t + \hat{\beta} y_t + \hat{\varepsilon}_{t+h}$. The normalization is made such that the normalized empirical mean squared errors of the GPCE are equal to one. For the GPCE, unnormalized empirical mean squared errors are also reported after "/". 3. The number of iterations is 2,000. 4. The factors were estimated as in Table 1.

Part 1: $h = 4$

			$(\phi_1, \phi_2) = (0.1, 0.4)$				$(\phi_1, \phi_2) = (0.3, 0.6)$				$(\phi_1, \phi_2) = (0.5, 0.8)$			
T	N	r	OPCE	GPCE	FGPCE	OPCED	OPCE	GPCE	FGPCE	OPCED	OPCE	GPCE	FGPCE	OPCED
50	25	1	1.018	1.000/1.063	0.996	1.489	1.028	1.000/1.078	0.993	1.475	1.021	1.000/1.102	0.989	1.440
		3	1.026	1.000/1.217	0.999	1.472	1.056	1.000/1.212	0.996	1.507	1.089	1.000/1.298	1.024	1.426
		5	1.053	1.000/1.410	1.018	1.411	1.073	1.000/1.371	1.037	1.456	1.089	1.000/1.476	1.047	1.366
		8	1.050	1.000/1.633	1.051	1.385	1.099	1.000/1.692	1.074	1.361	1.119	1.000/1.766	1.080	1.292
100	100	1	1.010	1.000/1.005	0.996	1.479	1.004	1.000/0.997	0.996	1.468	1.005	1.000/1.035	1.000	1.427
		3	1.027	1.000/1.052	0.998	1.540	1.021	1.000/1.054	0.995	1.493	1.017	1.000/1.077	0.999	1.456
		5	1.018	1.000/1.137	0.992	1.413	1.026	1.000/1.138	0.998	1.441	1.023	1.000/1.130	0.996	1.429
		8	1.032	1.000/1.228	0.993	1.409	1.019	1.000/1.205	1.001	1.421	1.042	1.000/1.248	1.004	1.414
50	100	1	1.011	1.000/1.042	0.997	1.475	1.004	1.000/1.076	0.995	1.416	1.007	1.000/1.047	1.000	1.436
		3	1.029	1.000/1.126	0.996	1.484	1.016	1.000/1.158	0.998	1.435	1.015	1.000/1.129	0.994	1.459
		5	1.025	1.000/1.241	1.002	1.377	1.009	1.000/1.266	0.992	1.383	1.040	1.000/1.298	1.012	1.390
		8	1.019	1.000/1.413	0.992	1.353	1.020	1.000/1.530	1.013	1.281	1.022	1.000/1.554	1.003	1.268
100	400	1	1.001	1.000/1.030	0.998	1.470	1.003	1.000/1.036	0.999	1.440	1.003	1.000/1.029	1.000	1.406
		3	1.001	1.000/1.052	0.999	1.478	1.004	1.000/1.061	0.997	1.422	1.004	1.000/1.056	0.998	1.476
		5	1.013	1.000/1.031	1.000	1.459	1.010	1.000/1.092	0.998	1.461	1.012	1.000/1.098	1.001	1.448
		8	1.018	1.000/1.173	1.001	1.394	1.010	1.000/1.149	1.003	1.431	1.022	1.000/1.156	1.000	1.482

Part 2: $h = 6$

				$(\phi_1, \phi_2) = (0.1, 0.4)$				$(\phi_1, \phi_2) = (0.3, 0.6)$				$(\gamma_1, \gamma_2) = (0.5, 0.8)$			
$\sqrt{N}/T$	T	N	r	OPCE	GPCE	FGPCE	OPCED	OPCE	GPCE	FGPCE	OPCED	OPCE	GPCE	FGPCE	OPCED
0.1	50	25	1	1.013	1.000/1.111	0.994	1.507	1.033	1.000/1.086	0.994	1.542	1.165	1.000/1.127	1.018	1.619
			3	1.068	1.000/1.303	1.009	1.574	1.064	1.000/1.280	1.003	1.508	1.066	1.000/1.337	1.010	1.455
			5	1.212	1.000/1.594	1.172	1.592	1.083	1.000/1.467	1.031	1.478	1.071	1.000/1.582	1.030	1.390
			8	1.060	1.000/1.754	1.040	1.431	1.023	1.000/1.843	1.012	1.319	1.122	1.000/1.917	1.116	1.371
	100	100	1	1.009	1.000/1.039	0.994	1.471	1.011	1.000/1.039	0.994	1.470	1.011	1.000/1.004	0.998	1.472
			3	1.017	1.000/1.080	1.001	1.465	1.014	1.000/1.075	0.994	1.523	1.017	1.000/1.062	0.996	1.528
			5	1.037	1.000/1.112	0.998	1.502	1.021	1.000/1.140	0.995	1.484	1.046	1.000/1.145	0.997	1.471
			8	1.039	1.000/1.247	1.002	1.497	1.048	1.000/1.260	0.996	1.453	1.061	1.000/1.292	1.010	1.462
0.2	50	100	1	1.009	1.000/1.079	0.995	1.491	1.014	1.000/1.075	1.000	1.508	1.018	1.000/1.044	0.995	1.500
			3	1.008	1.000/1.169	0.992	1.460	1.022	1.000/1.203	0.997	1.472	1.017	1.000/1.207	0.992	1.461
			5	1.034	1.000/1.334	0.996	1.386	1.028	1.000/1.353	1.000	1.401	1.032	1.000/1.380	1.008	1.378
			8	1.017	1.000/1.665	0.992	1.265	1.030	1.000/1.696	1.012	1.216	1.009	1.000/1.788	1.000	1.187
	100	400	1	1.000	1.000/1.002	1.000	1.486	1.004	1.000/1.011	0.999	1.447	1.003	1.000/1.009	0.997	1.499
			3	1.007	1.000/1.044	0.999	1.532	1.006	1.000/1.073	1.000	1.473	1.006	1.000/1.046	1.000	1.517
			5	1.008	1.000/1.103	0.998	1.520	1.013	1.000/1.114	1.000	1.504	1.011	1.000/1.126	1.001	1.548
			8	1.003	1.000/1.205	0.998	1.381	1.006	1.000/1.219	0.997	1.427	1.013	1.000/1.239	0.997	1.462