KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.887

"Asymptotic Efficiency in Factor Models and
Dynamic Panel Data Models"

Haruo Iwakura and Ryo Okui

February 2014

KYOTO UNIVERSITY

KYOTO, JAPAN

Asymptotic Efficiency in Factor Models and Dynamic Panel Data Models*

Haruo Iwakura† and Ryo Okui‡

February 6, 2014

Abstract

This paper studies the asymptotic efficiency in factor models with serially correlated errors and dynamic panel data models with interactive effects. We derive the efficiency bound for the estimation of factors, factor loadings and common parameters that describe the dynamic structure. We use double asymptotics under which both the cross-sectional sample size and the length of the time series tend to infinity. The results show that the efficiency bound for factors is not affected by the presence of unknown factor loadings and common parameters, and analogous results hold for the bounds for factor loadings and common parameters. The efficiency bound is derived by using an infinite-dimensional convolution theorem. Perturbation to the infinite-dimensional parameters, which consists in an important step of the derivation of the efficiency bound, is nontrivial and is discussed in detail.

Keywords: asymptotic efficiency; convolution theorem; double asymptotics; dynamic panel data model; factor model; interactive effects.

JEL classification: C13; C23.

*The authors would like to thank Hidehiko Ichimura and Masanobu Taniguchi for initially getting us interested in the problem examined in this paper. The authors acknowledge the valuable comments from Yoshihiko Nishiyama, Katsumi Shimotsu, Jushan Bai, Donggyu Sul, Marc Hallin, Tatsushi Oka, In Choi, Seung Chan Ahn, Michael Jansson and participants at the Osaka meeting of the Kansai Econometric Society, the spring meeting of the Japan Statistical Society, the Hitotsubashi Statistics seminar, the spring meeting of the Japanese Economic Association, the Asian meeting of the Econometric Society, the Kyoto Symposium on Recent Development in Statistics, Empirical Finance and Econometrics, National University of Singapore, the Tsinghua Econometrics Conference and the Hitotsubashi-Sogang Conference. Okui acknowledges financial support from the Japan Society for the Promotion of Science under KAKENHI 22730176 and 22330067. The authors are solely responsible for any errors.

†Graduate School of Economics, Kyoto University. Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: iwakura.haruo@ht8.ecs.kyoto-u.ac.jp

‡Corresponding author. Institute of Economic Research, Kyoto University. Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui@kier.kyoto-u.ac.jp

1 Introduction

This paper studies asymptotic efficiency in factor models and dynamic panel data models. We derive the efficiency bound for the estimation of parameters including infinite-dimensional parameters such as factors. We consider the setting in which we observe a variable y_{nt} for many individuals over long time periods, and y_{nt} can be written as the sum of the interaction of factors, f_t, and factor loadings, λ_n, and an idiosyncratic component, w_{nt}:

$$y_{nt} = f_t' \lambda_n + w_{nt}.$$

The factors, f_t, vary over time but are constant across individuals; and the factor loadings, λ_n are fixed over time but vary across individuals. The idiosyncratic component, w_{nt}, is assumed to be Gaussian stationary and independently and identically distributed (i.i.d.) across individuals. We further assume that its dynamics can be characterized by a finite number of parameters, θ. The vector θ consists of common parameters that do not depend on either n or t. This setting includes cases in which w_{nt} follows a stationary autoregressive-moving-average (ARMA) model, but is not restricted to them. The objective of this paper is to derive the efficiency bound for the estimation of functions of f_t, λ_n and θ.

This paper contributes to two different, although highly related, areas of the literature: factor analysis and dynamic panel data models. In factor analysis, factors are used to summarize the information contained in y_{nt} when there are many cross-sectional units.[1] See the introduction of Bai (2003), Bai (2009a), Breitung and Tenhofen (2011) or Choi (2011) for an overview of economic applications of factor models. The estimation of factors is discussed in Stock and Watson (2002), Bai (2003), Forni et al. (2005), Breitung and Tenhofen (2011) and Choi (2011) among others.

In dynamic panel data analysis, the dynamic structure of w_{nt} is of interest, and the factor specification is used as interactive effects for characterizing unobserved heterogeneity across individuals and unobserved common macro shocks. For example, when w_{nt} follows an autoregressive (AR) process, the model is called a panel AR model, which has been used in many economic applications. Note that our specification includes models with individual effects in which $f_t = 1$ and λ_n control time-invariant unobserved heterogeneity and that are commonly used in empirical applications. There have been many estimators proposed for dynamic panel

[1] This paper considers "exact factor models" in which w_{nt} is cross-sectionally independent, and any cross-sectional dependence in y_{nt} is captured by the interactive effects $f_t' \lambda_n$. In the recent literature, "approximate factor models" originated by Chamberlain and Rothschild (1983), in which cross-sectional dependence in w_{nt} is allowed, are popular. However, it would be very difficult, if not impossible, to discuss efficiency in the presence of general cross-sectional dependence in w_{nt}.

data models. For example, the estimation of panel AR models is discussed by Anderson and Hsiao (1981), Arellano and Bond (1991), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Hayakawa (2009), Lee (2012) and Han, Phillips and Sul (2014) for cases with individual effects, and by Hahn and Moon (2006) for cases with both individual and time effects. Holtz-Eakin, Newey and Rosen (1988), Bai (2009b), Moon and Weidner (2010) and Sarafidis and Yamagata (2010) discuss estimation of dynamic panel data models in the presence of interactive effects.

Our main purpose is to derive the efficiency bound; i.e., the lower bound of the asymptotic variances of any regular estimators of a function of parameters. We examine the asymptotic efficiency in factor models with serially correlated errors (or dynamic panel data models with factor structure) when both the cross-sectional sample size (N) and the time-series length (T) tend to infinity. We allow some elements of f_t and/or λ_n to be known. Thus, our analysis can be applied to simpler specifications that are popularly used in dynamic panel data analysis, such as models with only individual effects.

We derive the efficiency bound for the estimation of infinite-dimensional parameters, f_t and λ_n. The presence of the common parameters (whose dimension is finite) does not affect the efficiency bound for infinite-dimensional parameters. It is remarkable that the efficiency bound for factors is not affected by the estimation of the factor loadings and vice versa. The results indicate that the principal component analysis (PCA) estimator of factors by Stock and Watson (2002) and Bai (2003), which is arguably the most popular estimator in factor models, is efficient.[2] The PCA estimator of factor loading is efficient only if w_{nt} is serial uncorrelated. The estimator developed by Breitung and Tenhofen (2011) is shown to be efficient when it is applied to the current setting.[3]

We examine the efficiency bound for the estimation of the common parameter, θ, which is the parameter of interest in dynamic panel data analysis. Our results reveal that the efficiency bound in the presence of factor structure is the same as that in the case when factors and factor loadings are known.

We apply the efficiency results to various dynamic panel data models. First, we consider the panel AR(1) model. Note that the efficiency bound has been derived by Hahn and Kuer-

[2]More precisely, we consider the efficient bound for the estimation of some linear transformation of the factors. The factors themselves are not identified, but some linear transformation of them is identified and can be estimated by the principal component analysis.

[3]Breitung and Tenhofen (2011) show that their estimators are more efficient than the PCA estimator. However, they do not examine the efficiency bound. We also note that they consider more general settings than ours; for example, they allow heteroscedasticity. The efficiency bounds in those general settings are not known.

steiner (2002) for cases with individual effects. We derive the same efficiency bound based on our general efficiency result. However, as discussed in Section 5.1, the argument of Hahn and Kuersteiner (2002) is not complete, and we conclude their analysis. We also consider the estimation of the parameter in panel MA(1) models and the estimation of autocovariances. For the estimation of autocovariances, we provide conditions under which Okui's (2010) autocovariance estimator achieves the efficiency bound. In particular, we show that if the true data-generating process follows a Gaussian stationary ARMA(p, q) model, Okui's estimator for the k-th-order autocovariance is asymptotically efficient if and only if $p \geq q$ and $0 \leq k \leq p - q$. These results are analogous to the results of Kakizawa and Taniguchi (1994), who derive the lower bound of the variances of autocovariance estimators in a time-series setting.[4]

The notion of efficiency used in this paper is that of the convolution theorem by Hajék (1970), which is extended to the cases with infinite-dimensional parameters by van der Vaart and Wellner (1996). The derivation of the efficiency bound is nontrivial because we consider double asymptotics, and there are infinitely many nuisance parameters (i.e., there are as many factors as T and as many factor loadings as N).

The main difficulty is how to specify the parameter space for the "localized" factors and factor loadings. The convolution theorem requires that the parameter space be a linear subspace of a Hilbert space with an inner product that is consistent with the formula of the limiting process of the local log-likelihood. However, it is not trivial to construct appropriate local parameter spaces for factors and factor loadings that satisfy the conditions for a linear subspace of a Hilbert space. We discuss the difficulty of constructing an appropriate local parameter space in detail in Section 5. In particular, we argue that extending the approach taken by Hahn and Kuersteiner (2002) does not provide the efficiency bound for factors and factor loadings.

To construct a local parameter space, we adopt an approach that is based on the theory of l_2 space. We assume that the localized factors and factor loadings are square summable and prove that the local parameter space is a Hilbert space with an appropriate inner product. We then derive the efficiency bound for the estimation of factors and factor loadings, as well as the common parameters.

Our approach for constructing local parameter spaces deviates from that considered in the related literature in statistics. There have been a number of studies on asymptotic efficiency in "functional models" in statistics.[5] A functional model is one that contains nonrandom infinite-dimensional parameters. In that literature, there have been several attempts to construct appropriate local parameter spaces to derive the efficiency bound. However, this paper does

[4] See Porat (1987) and Walker (1995) for alternative derivations of the efficiency bound.

[5] See, e.g., Sprent (1966), Kumon and Amari (1984), Pfanzagl (1993) and Strasser (1996, 1998).

not take this approach because it turns out that factors are not regular parameters under this approach and we cannot derive the efficiency bound for the estimation of factors under this alternative approach.

While efficiency is an important issue, there have not been many studies on efficiency bounds in factors models or dynamic panel data models. Hahn and Kuersteiner (2002) provide the seminal contribution and derive the efficiency bounds for panel AR(1) models with individual effects. They also apply the convolution theorem. However, they do not discuss how to specify the local parameter space. While their efficiency result is valid by simply adding an assumption that the local parameter space is a linear subspace of a Hilbert space, constructing an appropriate local parameter space is far from obvious as argued above. This paper complements their analysis by providing concrete examples of local parameter spaces for models with individual effects.

Gagliardini and Gourieroux (2010) examine the asymptotic efficiency in dynamic panel data models with factor structure. They assume that the dynamics of factors can be characterized by finite-dimensional parameters. Therefore, their analysis does not need to face the difficulty associated with the presence of infinite-dimensional parameters. We treat factors as infinite-dimensional parameters and do not specify their dynamics.

The reminder of this paper is organized as follows. The next section presents the setting. Section 3 gives the summary of the results. Section 4 provides the theoretical derivation of the efficiency bound and presents the results given in Section 3 in a more mathematically formal way. Section 5 contains detailed discussion on the specification of the parameter space for localized infinite-dimensional parameters. Section 6 concludes the paper. All mathematical proofs are given in the Appendix.

2 Setup

Suppose that we have available a panel data set $\{y_{nt}\}$ for $n = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$. We assume that y_{nt} follows a factor model:

$$y_{nt} = f_t' \lambda_n + w_{nt},$$

where w_{nt} is independently and identically distributed (i.i.d.) across individual n and follows a Gaussian stationary process over time t with mean zero. We call f_t factors and λ_n factor loadings, respectively, and we regard them as parameters. The interaction, $f_t' \lambda_n$, is called interactive effects. Let p denote the number of factors so that f_t and λ_n are $p \times 1$ vectors. We assume that p is known.[6] We allow some elements of f_t and λ_n to be known.

It is assumed that the autocovariance structure of $\{w_{nt}\}_{t\in\mathbb{Z}}$ ($\mathbb{Z}$ is the set of all integers) is completely characterized by a finite-dimensional parameter $\theta \in \Theta$ where Θ is some open subset of $\mathbb{R}^L$. We denote the k-th-order autocovariance by $\gamma_k(\theta)$; i.e., $\gamma_k(\theta) := \mathbb{E}_\theta[w_{nt}w_{n,t-k}]$, where $\mathbb{E}_\theta$ denotes the expectation under θ. Note that because $\{w_{nt}\}_{t\in\mathbb{Z}}$ is a Gaussian stationary process, the parameter θ completely determines the law of the process $\{w_{nt}\}_{t\in\mathbb{Z}}$. We also impose an absolute summability condition on the autocovariance function $k \mapsto \gamma_k(\theta)$. Thus far, we have assumed the following restrictions on w_{nt}.

Assumption 1.

(i) w_{nt} is i.i.d. across individual n.

(ii) w_{nt} follows a Gaussian stationary process over t with $\mathbb{E}_\theta[w_{nt}] = 0$.

(iii) $\sum_{k=-\infty}^{\infty} |\gamma_k(\theta)| < \infty$ for every $\theta \in \Theta$.

This class of models includes many models that are popularly used in the literature. This class includes factor models in which factors, $\{f_t\}_{t=1}^T$, are the parameters of interest. It also contains many dynamic panel data models that are popularly used in applied studies. For example, if $p = 1$, $f_t = 1$ and w_{nt} follows an AR(1) process so that $w_{nt} = \alpha w_{n,t-1} + u_{nt}$ where $u_{nt} \sim i.i.d.N(0, \sigma^2)$, then a panel AR(1) model with individual effects is obtained:

$$y_{nt} = \alpha y_{n,t-1} + (1-\alpha)\lambda_n + u_{nt}.$$

If we set $p = 2$, $f_t = (1, f_{1t})'$ and $\lambda_n = (\lambda_{1n}, 1)$ and assume that w_{nt} follows an AR(1) process, we obtain the panel AR(1) model with both individual and time effects:

$$y_{nt} = \alpha y_{n,t-1} + (1-\alpha)\lambda_{1n} + (1-\alpha)f_{1t} + u_{nt}.$$

3 Summary of the results

In this section, we present the summary of the results on the efficiency bound in dynamic panel data models with factor structure. The assumptions used for the derivation and the formal theoretical results are presented in the following section.

[6] Choice of p is an important topic in the recent econometric literature. See, e.g., Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013).

3.1 Factors and factor loadings

We first present the efficiency bound for the estimation of $J_{NT}'f_t$, where J_{NT} is a nonrandom $p\times p$ invertible matrix, possibly depending on N and T, and $\lim_{N,T\to\infty} J_{NT} = J$. The reason that we consider the estimation of $J_{NT}'f_t$, not f_t itself, is that the factors f_t may not be identified. For any invertible $p \times p$ matrix A, the pair of f_t and λ_n and that of $A'f_t$ and $A^{-1}\lambda_n$ yield the same interaction effect $f_t'\lambda_n$ and the same observable distribution of y_{nt}. Therefore, without further restrictions, we cannot identify f_t and λ_n.[7] On the other hand, some linear transformation of f_t is identified and can be estimated. We thus consider the efficiency bound for $J_{NT}'f_t$. The efficiency bound is presented in Theorem 4.3. Let $\Sigma_{\lambda\lambda} = \lim_{N\to\infty}\sum_{n=1}^N \lambda_n\lambda_n'/N$. The efficiency bound is:

$$\gamma(0)J'\Sigma_{\lambda\lambda}^{-1}J,$$

and the rate of convergence is $\sqrt{N}$. This efficiency bound is the same as the asymptotic variance of the OLS estimator from the regression of y_{nt} on λ_n using observations at time t. This result indicates that the efficiency bound for factors is unaltered even when the common parameters and factor loadings are known.

Next, we give the efficiency bound for factor loadings, $J_{NT}^{-1}\lambda_n$. We introduce the following notation. Let $\Omega(\theta)$ be the variance–covariance matrix of the vector $w_n = (w_{n1}, \dots w_{nT})'$ so that $\Omega(\theta) := \mathbb{E}_\theta[w_n w_n']$. Let $F_T = (f_1, \dots, f_T)'$. Let D_T be the diagonal matrix whose diagonal elements are the square roots of the diagonal elements of $F_T'F_T$. When factors are stationary, then the order of D_T is $\sqrt{T}$. However, when some factor exhibits a deterministic trend or is slowly varying, the order of the corresponding element of D_T is different from $\sqrt{T}$. The efficiency bound for the estimation of $J_{NT}^{-1}\lambda_n$ is presented in Theorem 4.4 and is:

$$J^{-1}\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega_T(\theta)^{-1}F_T D_T^{-1}\right)^{-1} J'^{-1},$$

and the rate of convergence is D_T. Similar to the result for factors, the efficiency bounds for factor loadings are the same even when common parameters and factors are known. If θ and F_T were known, this efficiency bound would be attained by the GLS estimator of y_{nt} on f_t using observations from individual n.

Lastly, we provide the efficiency bound for an interactive effect, $f_t'\lambda_n$. Let d_k denote the k-th diagonal element of D_T, and $r_{NT} := \min\{\sqrt{N}, d_1, d_2, \dots, d_p\}$. The efficiency bound is given in

[7] Bai and Ng (2013) provide conditions under which they are identified.

Theorem 4.5. It is:

$$f_t'\left(\lim_{N,T\to\infty} D_T^{-1}r_{NT}\right)\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega_T(\theta)^{-1}F_TD_T^{-1}\right)^{-1}\left(\lim_{N,T\to\infty} D_T^{-1}r_{NT}\right)f_t$$
$$+\lambda_n'\left(\lim_{N,T\to\infty}\frac{r_{NT}}{\sqrt{N}}\right)^2\gamma(0)\Sigma_{\lambda\lambda}^{-1}\lambda_n,$$

and the rate of convergence is r_{NT}.

We now examine the efficiency of the PCA estimator. The PCA estimator is obtained as follows. Let Y be the $T\times N$ matrix whose (t,n) element is y_{nt}. The PCA estimator of factors $\hat{F}_T$ is the $\sqrt{T}$ times the eigenvectors that correspond to the p largest eigenvalues of YY'. The PCA estimator of factor loadings is $\hat{\Lambda}_N' = \hat{F}_T'Y/T$ so that the PCA estimator of λ_n is the n-th row of $\hat{\Lambda}_N$. We assume that $D_T = O(\sqrt{T})$. Bai (2003) derives the asymptotic distribution of the PCA estimator. To introduce the asymptotic distribution, we need the following notation. Υ is the eigenvector matrix of $\Sigma_{\lambda\lambda}^{1/2}(\lim_{T\to\infty}F_T'F_T/T)\Sigma_{\lambda\lambda}^{1/2}$, and V is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. Define $Q = V^{1/2}\Upsilon'\Sigma_{\lambda\lambda}^{-1/2}$. Bai (2003) considers the distribution of $\sqrt{N}(\hat{f}_t - \hat{J}_{PCA,NT}'f_t)$ for some random matrix $\hat{J}_{PCA,NT}$. In Appendix B, we show that $\hat{J}_{PCA,NT}$ can be replaced with a nonrandom matrix $J_{PCA,NT}$ so that we can apply our efficiency result. Moreover, we show that $\lim_{N,T\to\infty} J_{PCA,NT} = Q^{-1}$.

We show that the PCA estimator of factors is efficient. By Bai (2003) and Appendix B, if $N,T\to\infty$ and $\sqrt{N}/T\to 0$:

$$\sqrt{N}(\hat{f}_t - J_{PCA,NT}'f_t)\to_d N\left(0,\gamma(0)V^{-1}Q\Sigma_{\lambda\lambda}Q'V^{-1}\right).$$

Simple algebra shows that the asymptotic variance is simplified to:

$$\gamma(0)V^{-1}Q\Sigma_{\lambda\lambda}Q'V^{-1} = \gamma(0)V^{-1}V^{1/2}\Upsilon'\Sigma_{\lambda\lambda}^{-1/2}\Sigma_{\lambda\lambda}\Sigma_{\lambda\lambda}^{-1/2}\Upsilon V^{1/2}V^{-1} = \gamma(0)V^{-1}.$$

The efficiency bound is:

$$\gamma(0)(Q^{-1})'\Sigma_{\lambda\lambda}^{-1}Q^{-1} = \gamma(0)V^{-1/2}\Upsilon'\Sigma_{\lambda\lambda}^{1/2}\Sigma_{\lambda\lambda}^{-1}\Sigma_{\lambda\lambda}^{1/2}\Upsilon V^{-1/2} = \gamma(0)V^{-1}.$$

The asymptotic variance is equal to the efficiency bound, and the PCA estimator is efficient.

On the other hand, the PCA estimator of factor loadings is not efficient, in general, and is efficient only if w_{nt} is serially uncorrelated. By Bai (2003) and Appendix B, the asymptotic distribution is:

$$\sqrt{T}(\hat{\lambda}_n - J_{PCA,NT}^{-1}\lambda_n)\to_d N\left(0,(Q')^{-1}\lim_{T\to\infty}\frac{1}{T}F_T'\Omega_T(\theta)F_TQ^{-1}\right).$$

The efficiency bound is:

$$Q\left(\lim_{T\to\infty}\frac{1}{T}F_T'\Omega_T(\theta)^{-1}F_T\right)^{-1}Q'.$$

The asymptotic variance is not equal to the efficiency bound in general. Thus, the PCA estimator of factor loadings is not efficient. However, when w_{nt} is serially uncorrelated, it is efficient. In that case, $\Omega_T(\theta) = \gamma(0)I_T$, and the asymptotic variance is:

$$(Q')^{-1} \lim_{T\to\infty} \frac{1}{T} F_T' \Omega_T(\theta) F_T Q^{-1} = \gamma(0) V^{-1/2} \Upsilon' \Sigma_{\lambda\lambda}^{1/2} \lim_{T\to\infty} \frac{1}{T} F_T' F_T \Sigma_{\lambda\lambda}^{1/2} \Upsilon V^{-1/2} = \gamma(0) I_p,$$

by the properties of eigenvalues and eigenvectors. Similarly, the efficiency bound is:

$$\begin{aligned} Q \left(\lim_{T\to\infty} \frac{1}{T} F_T' \Omega_T(\theta)^{-1} F_T \right)^{-1} Q' =& \gamma(0) V^{1/2} \Upsilon' \Sigma_{\lambda\lambda}^{-1/2} \left(\lim_{T\to\infty} \frac{1}{T} F_T' F_T \right)^{-1} \Sigma_{\lambda\lambda}^{-1/2} \Upsilon V^{1/2} \\ =& \gamma(0) I_p. \end{aligned}$$

Thus, the asymptotic variance and the efficiency bound match, and the PCA estimator of factor loadings is efficient when w_{nt} is i.i.d.

We next show that the PC-GLS estimator of Breitung and Tenhofen (2011) indeed attains the efficiency bound even in the presence of serial correlation.[8] Suppose that $w_{nt} = \sum_{k=1}^{K} \alpha_k w_{n,t-k} + u_{nt}$, where $u_{nt} \sim i.i.d.N(0, \sigma^2)$. Assume also that $D_T = O(\sqrt{T})$. Consider the following approximate Gaussian log-likelihood function:

$$-\frac{N(T-p)}{2} \log \sigma^2 - \sum_{n=1}^{N} \sum_{t=p+1}^{T} \frac{((y_{nt} - f_t'\lambda_n) - \sum_{k=1}^{p} \alpha_k (y_{n,t-k} - f_{t-k}'\lambda_n))^2}{2\sigma^2}. \tag{3.1}$$

The PC-GLS estimator is the local maximum of the function (3.1) in the neighborhood of the principal component estimator by Stock and Watson (2002).[9] Let $\hat{f}_t^{GLS}$ and $\hat{\lambda}_n^{GLS}$ be the PC-GLS estimators of f_t and λ_n, respectively. Applying Theorem 1 of Breitung and Tenhofen (2011) in the current setting (see also the proof of Theorem 2 in Breitung and Tenhofen (2011)) gives that if $N, T \to \infty$ with $\sqrt{N}/T \to 0$:

$$\sqrt{N}(\hat{f}_t^{GLS} - J_{PCA,NT}' f_t) \to_d N\left(0, \gamma(0)(Q^{-1})' \Sigma_{\lambda\lambda}^{-1} Q^{-1}\right),$$

and, if $N, T \to \infty$ with $\sqrt{T}/N \to 0$:

$$\sqrt{T}(\hat{\lambda}_n^{GLS} - J_{PCA,NT}^{-1} \lambda_n) \to_d N\left(0, Q \left(\lim_{T\to\infty} \frac{1}{T} F_T' \Omega_T(\theta)^{-1} F_T \right)^{-1} Q' \right).$$

The asymptotic variances are the same as the efficiency bounds, and the PC-GLS estimator is asymptotically efficient.

[8]As written in footnote 3, Breitung and Tenhofen (2011) consider more general settings in which there is no Gaussianity assumption, and heteroscedasticity and weak cross-sectional dependence are allowed. Moreover, the AR model used for the estimation can be misspecified.

[9]The global maximum is not well defined because the likelihood function is unbounded, in general. See Breitung and Tenhofen (2011, Section 3) for the detail of this problem and how to compute the estimator.

Note that Choi (2011) and Bai and Li (2012) also develop estimators that are more efficient than the principal component estimator under the set of conditions different from ours. Choi's focus is on heteroscedasticity and cross-sectional dependence, and serial correlation is basically excluded. Bai and Li (2012) account for heteroscedasticity, but serial correlation is not allowed. On the other hand, we consider serial correlation, but we do not account for heteroscedasticity across individuals nor cross-sectional correlation. Note that they do not derive the efficiency bound, and the efficiency bounds under heteroscedasticity or cross-sectional dependence are not known.

3.2 Common parameters

This section presents the efficiency bound for the estimation of functions of the common parameters. The parameter of interest is $\beta(\theta)$, where $\beta(\cdot)$ is a differentiable function from Θ to $\mathbb{R}^M$. For example, when we are interested in the model parameter itself, we set $\beta(\theta) = \theta$. When an autocovariance is estimated, we have $\beta(\theta) = \gamma_k(\theta)$.

We first derive the result for general $\beta(\theta)$. We also discuss several applications in which specific forms of $\beta(\theta)$ are presented. Let $\dot{\beta}(\theta) = \partial\beta(\theta)/\partial\theta$. Let $g_\theta(\cdot)$ be the spectral density of w_{nt}:

$$g_\theta(s) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \gamma_m(\theta) \exp(-ims),$$

where $i := \sqrt{-1}$. Note that Assumption 1 guarantees the existence of g_θ. We also define:

$$\Gamma(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta'} g_\theta(s) \frac{\partial}{\partial\theta} g_\theta(s) \frac{ds}{g_\theta^2(s)}.$$

The efficiency bound for the estimation of $\beta(\theta)$ is given in Theorem 4.6. It is:

$$\dot{\beta}(\theta)\Gamma(\theta)^{-1}\dot{\beta}(\theta)',$$

and the rate of convergence is $\sqrt{NT}$.

The efficiency bound has the same form as the limit of the Cramér–Rao lower bounds for estimation of $\beta(\theta)$ in a time-series setting without an intercept under Gaussianity. Indeed, in the time-series literature, the matrix $\Gamma(\theta)$ is called the Gaussian Fisher information matrix associated with spectral density g_θ (see, e.g., Taniguchi and Kakizawa (2000, p. 58)). We note that the closed form of the matrix $\Gamma(\theta)$ for ARMA models is available in, e.g., Box and Jenkins (1970).

An important implication of this finding is that the presence of factors does not affect the form of the efficiency bound. This result is interesting in the sense that the sequence of factors and factor loadings is an infinite-dimensional parameter, but the efficiency bound is the same as that for the case in which they are known. An interpretation of this result is that it is related to the well-known fact that the sample average and the sample variance are independent when observations are Gaussian. Note that in factor models, the interactive effects, $f_t'\lambda_n$, determine the mean of y_{nt}, and the parameter θ determines the variance–covariance structure of y_{nt}.

We apply this efficiency result to the panel AR(1) model, the panel MA(1) model and the estimation of autocovariances.

3.2.1 Panel AR(1) model

The first application deals with the case in which w_{nt} follows an AR(1) model:

$$w_{nt} = \alpha w_{n,t-1} + u_{nt},$$

where $u_{nt} \sim i.i.d.N(0, \sigma^2)$. The parameter θ, in this case, is $\theta = (\alpha, \sigma^2)'$. In terms of y_{nt}, this specification can be written as:

$$y_{nt} = \alpha y_{n,t-1} + (f_t - \alpha f_{t-1})'\lambda_n + u_{nt}.$$

This specification, in particular the one with individual effects (i.e., $f_t = 1$), has been examined in many studies as mentioned in the introduction. Because $\psi(\theta) = \theta$ in this case, the efficiency bound is:

$$\Gamma(\theta)^{-1} = \begin{pmatrix} 1-\alpha^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

The efficiency bound for α is the same as that derived by Hahn and Kuersteiner (2002). The efficiency bound for σ^2 is a new result. Note that the estimation of σ^2 is important; for example, when we provide a forecasting interval for a future value of y_{it}.

There are many estimators that achieve the efficiency bound for α for models with individual effects. For example, all of the bias-corrected, fixed-effects estimators of Hahn and Kuersteiner (2002), the GMM estimator developed by Hayakawa (2009), and the random- effects, pseudo-maximum-likelihood estimator discussed by Alvarez and Arellano (2003) are asymptotically efficient. Therefore, to distinguish many existing estimators in terms of efficiency, the first-order efficiency result considered here is not sufficient, and we may need alternative efficiency criteria, such as higher-order efficiency.

3.2.2 Panel MA(1) model

This subsection considers the following panel MA(1) model with individual effects:

$$y_{nt} = \lambda_n + \alpha u_{n,t-1} + u_{nt},$$

where $|\alpha| < 1$ and $u_{nt} \sim$ i.i.d.$N(0, \sigma^2)$ across both n and t. The parameter is $\theta = (\alpha, \sigma^2)$. MA models are also popularly used to analyze the dynamic nature of economic variables. For example, Abowd and Card (1989) employ MA models to study income dynamics.

Noting that $\beta(\theta) = \alpha$ in this case, the efficiency bound is:

$$\left\{\frac{\partial}{\partial\theta'}\beta(\theta)\right\}\Gamma(\theta)^{-1}\left\{\frac{\partial}{\partial\theta}\beta(\theta)\right\} = \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}1-\alpha^2 & 0\\ 0 & 2\sigma^4\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = 1-\alpha^2.$$

We note that the literature on the estimation of panel MA models is scarce, and we are not aware of any estimator that achieves this efficiency bound.

3.2.3 Autocovariances

This section examines the asymptotic efficiency of the estimation of the k-th-order autocovariance, $\gamma_k(\theta)$. By using the spectral density g_θ, $\gamma_k(\theta)$ can be expressed as:

$$\gamma_k(\theta) = \int_{-\pi}^{\pi}\exp(-iks)g_\theta(s)ds = \int_{-\pi}^{\pi}\cos(ks)g_\theta(s)ds,$$

by Fourier inversion. Therefore, the efficiency bound for the estimation of γ_k is given by:

$$\left\{\int_{-\pi}^{\pi}\cos(ks)\frac{\partial}{\partial\theta'}g_\theta(s)ds\right\}\Gamma(\theta)^{-1}\left\{\int_{-\pi}^{\pi}\cos(ks)\frac{\partial}{\partial\theta}g_\theta(s)ds\right\}. \tag{3.2}$$

This expression demonstrates that the efficiency bound has the same form as the limit of the Cramér–Rao lower bounds for estimation of $\gamma_k(\theta)$ in a time-series setting obtained by Kakizawa and Taniguchi (1994).[10] We use their results to investigate the conditions under which existing estimators of autocovariances achieve the efficiency bound. In particular, we consider Okui's (2010) bias-corrected autocovariance estimator.[11]

We define Okui's (2010) estimator and derive its asymptotic distribution. We consider the case with individual effects only: $y_{nt} = \lambda_n + w_{nt}$. Okui's (2010) estimator is a bias-corrected, within-group sample autocovariance estimator and is defined as:

$$\tilde{\gamma}_k := \frac{1}{N(T-k)}\sum_{n=1}^{N}\sum_{t=k+1}^{T}(y_{nt}-\bar{y}_n)(y_{n,t-k}-\bar{y}_n) + \frac{1}{T}\hat{V}_T,$$

[10]See Porat (1987) and Walker (1995) for alternative derivations of the lower bound. Note that these three papers examine the limit of the Crámer–Rao lower bound such that it gives the lower bound of the variances of (exactly) unbiased estimators, while the convolution theorem gives the lower bound of the variances of regular estimators. Considering regular estimators allows estimators to be only asymptotically unbiased but does not require that they be unbiased in finite samples. Moreover, it is difficult to develop estimators that are unbiased in finite samples in our setting because of the presence of interactive effects.

[11]Okui (2010) analyzes models with individual effects. Cases with incidental trends and cases with both individual and time effects are considered by Okui (2011) and Okui (2013), respectively.

where $\hat{V}_T$ is an estimator of the long-run variance of w_{nt}. The first term on the right-hand-side is the within-group sample autocovariance estimator, and the second term is added to correct the bias that arises because λ_n cannot be estimated at $\sqrt{NT}$. Okui (2010, Section 4) suggests using the kernel estimators (Parzen (1957) and Andrews (1991)) for $\hat{V}_T$. Under Gaussianity, the asymptotic distribution of $\tilde{\gamma}_k$ is obtained as a corollary of Okui (2010, Theorem 4):

$$\sqrt{NT}(\tilde{\gamma}_k - \gamma_k(\theta)) \xrightarrow{d} N\left(0, \sum_{s=-\infty}^{\infty} \{\gamma_s(\theta)^2 + \gamma_{k+s}(\theta)\gamma_{k-s}(\theta)\}\right).$$

Because this asymptotic variance of $\tilde{\gamma}_k$ is exactly the same form as that of its time-series counterpart (see, e.g., Anderson (1971, Chapter 8)), we can use the results in time-series analysis to investigate its efficiency. Kakizawa and Taniguchi (1994) present the necessary and sufficient condition for this variance to be equal to (3.2). Using their result, it follows that if $\{w_{nt}\}_{t\in\mathbb{Z}}$ is a Gaussian stationary ARMA(p, q) process, then $\tilde{\gamma}_k(\theta)$ is asymptotically efficient if and only if:

$$p \geq q \quad \text{and} \quad 0 \leq k \leq p - q. \tag{3.3}$$

The condition (3.3) implies that if $\{w_{nt}\}_{t\in\mathbb{Z}}$ is a stationary AR(p) process and $k \leq p$, then we can efficiently estimate $\gamma_k(\theta)$ using $\tilde{\gamma}_k$. Porat (1987) states that this is not surprising in time-series contexts because AR coefficients can be efficiently estimated by Yule–Walker estimators, which are functions of sample autocovariances. On the other hand, if $\{w_{nt}\}_{t\in\mathbb{Z}}$ is a Gaussian MA(q) process, then none of $\tilde{\gamma}_k$ are asymptotically efficient. As an intermediate case, if $\{w_{nt}\}_{t\in\mathbb{Z}}$ is, for example, a stationary ARMA$(3, 1)$ process, then $\tilde{\gamma}_k$ is asymptotically efficient if and only if $0 \leq k \leq 2$. The results for MA models and ARMA models are nontrivial and interesting, as argued by Porat (1987).

4 Derivation of the efficiency bound

In this section, we present the derivation of the efficiency bound formally. We start with a general discussion on the convolution theorem. We then present the regularity conditions on the spectral density of w_{nt} and infinite-dimensional parameters, and we explain how to specify the parameter space for the localized factors and factor loadings. It turns out that how to specify the local parameter space is not a trivial question, and it is discussed in detail in the next section.

4.1 A convolution theorem

The efficiency bound is derived by employing the infinite-dimensional convolution theorem by van der Vaart and Wellner (1996). In this subsection, we briefly review their result.[12]

We follow the notation and terminology used by van der Vaart and Wellner (1996). Let H be a linear subspace of a Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We call H the local parameter space. For each N (sample size) and $h \in H$, let $P_{N,h}$ be a probability measure on a measurable space $(\mathcal{X}_N, \mathcal{A}_N)$. Consider a problem of estimating a parameter $\kappa_N(h) \in B$, where B is a Banach space, given a sample with law $P_{N,h}$. Let $\{\Delta_h : h \in H\}$ be an *iso-Gaussian process* indexed by H such that it is a Gaussian process with mean zero and covariance function $\mathbb{E}\Delta_{h_1}\Delta_{h_2} = \langle h_1, h_2 \rangle$. The sequence of experiments $\{\mathcal{X}_N, \mathcal{A}_N, P_{N,h} : h \in H\}$ or, simply, $\{P_{N,h} : h \in H\}$ is said to be *locally asymptotically normal* (LAN) if we can write:

$$\log \frac{dP_{N,h}}{dP_{N,0}} = \Delta_{N,h} - \frac{1}{2}\|h\|^2,$$

for a sequence of random variables $\Delta_{N,h}$ such that as $N \to \infty$:

$$\Delta_{N,h} \overset{0}{\rightsquigarrow} \Delta_h. \tag{4.1}$$

Here, $\overset{h}{\rightsquigarrow}$ denotes weak convergence under $P_{N,h}$. By the iso-Gaussianity assumption on $\{\Delta_h : h \in H\}$, the condition (4.1) is equivalent to saying that for any finite subset $\{h_1, h_2, \cdots, h_d\} \subseteq H$:

$$\begin{pmatrix} \Delta_{N,h_1} \\ \Delta_{N,h_2} \\ \vdots \\ \Delta_{N,h_d} \end{pmatrix} \overset{0}{\rightsquigarrow} N(0, (\langle h_a, h_b \rangle)), \tag{4.2}$$

as $N \to \infty$ where $(\langle h_a, h_a \rangle)$ is a $d \times d$ matrix whose (a, b)-th component is $\langle h_a, h_b \rangle$. The sequence of parameters $\kappa_N(h)$ is assumed to be *regular*, in the sense that as $N \to \infty$:

$$\mathbf{r}_N(\kappa_N(h) - \kappa_N(0)) \to \dot{\kappa}, \quad \forall h \in H,$$

for some bounded linear map $\dot{\kappa} : H \to B$ and the sequence of certain linear maps $\mathbf{r}_N : B \mapsto B$. A sequence of estimators τ_N is said to be *regular* with respect to $\mathbf{r}_N$ if, as $N \to \infty$:

$$\mathbf{r}_N(\tau_N - \kappa_N(h)) \overset{h}{\rightsquigarrow} L, \quad \forall h \in H.$$

[12] While we consider double asymptotics under which both $N \to \infty$ and $T \to \infty$, the convolution theorem stated here considers only N as the index that tends to infinity. However, it is sufficient to show the convolution theorem with $N \to \infty$ by the following argument. The theorem can be directly applied to the case of diagonal asymptotics under which T depends on N, say $T = T(N)$, and $T(N) \to \infty$ as $N \to \infty$. Phillips and Moon (1999) state in their Remark (a) after Definition 2 that if a weak convergence result holds under any diagonal asymptotics in which $T(N)$ is monotonic in N, then that result holds under double asymptotics. This condition is satisfied in our convolution theorem.

It should be emphasized that this definition requires that the limit distribution L be the same across h.[13] Let B^* denote the dual space of B. A bounded linear map $\dot{\kappa} : H \mapsto B$ has an adjoint map $\dot{\kappa}^* : B^* \mapsto \bar{H}$, where $\bar{H}$ is the completion of H. The adjoint map is determined by the relation:

$$\langle \dot{\kappa}^* b^*, h \rangle = b^* \dot{\kappa}(h),$$

for $h \in H$ and $b^* \in B^*$.

Under the setting above, van der Vaart and Wellner (1996) establish the following infinite-dimensional convolution theorem.

Theorem 4.1 (van der Vaart and Wellner, 1996, Theorem 3.11.2)**.** *: Assume that* $\{P_{N,h}, h \in H\}$ *is LAN. Furthermore, assume that the sequence of parameters* $\kappa_N(h)$ *and that of estimators* τ_N *are regular. Then, the limit distribution* L *of* $\mathbf{r}_N(\tau_N - \kappa_N(0))$ *equals the sum* $G + W$ *of independent, tight, Borel-measurable random elements in* B *such that:*

$$b^* G \sim N(0, \|\dot{\kappa}^* b^*\|^2), \quad \forall b^* \in B^*.$$

The law of G *concentrates on the closure of* $\dot{\kappa}(H)$*.*

This theorem implies that the law of G is optimal in the sense that the variance of the estimation of a linear combination of parameters cannot be smaller than the variance of the corresponding linear transformation of G. In particular, when the parameter of interest is finite dimensional, the variance of G is the efficiency bound. We apply this convolution theorem to derive the efficiency bounds for parameters in factor models. To do so, we need (1) to find some appropriate local parameter space, (2) to derive the LAN result, and (3) the parameter of interest to be regular.

4.2 Regularity conditions on the spectral density

In this subsection, we state the regularity conditions on the spectral density of w_{nt} that are needed to apply Theorem 4.1. Recall that g_θ denotes the spectral density of w_{nt} and that Assumption 1 (iii) guarantees its existence.

Assumption 2.

[13]Intuitively speaking, a sequence of estimators is regular if its limiting distribution is unaffected by a disappearing small change of the parameters (see, e.g., van der Vaart (1998, p. 115)). The regularity requirement is a desirable property for reasonable estimators to have, and it is not very restrictive. It excludes, for example, Hodge's super-efficient estimator (see van der Vaart (1998, p. 109)). For a detailed study of the regularity condition, see, e.g., Bickel et al. (1993, Chapter 2).

(i) $\theta \mapsto g_\theta(s)$ is differentiable at any point $\theta \in \Theta$.

(ii)

$$\lim_{\epsilon \to 0} \sup_s |g_{\theta+\epsilon}(s) - g_\theta(s)| = 0, \quad \forall \theta \in \Theta.$$

(iii)

$$\int_{-\pi}^{\pi} \left| \frac{\partial}{\partial \theta_m} g_\theta(s) \right|^2 ds < \infty, \quad \forall m = 1, 2, \cdots, L, \quad \forall \theta \in \Theta,$$

where θ_m is the m-th component of θ, and:

$$\lim_{\epsilon \to 0} \int_{-\pi}^{\pi} \left| \frac{\partial}{\partial \theta_m} g_{\theta+\epsilon}(s) - \frac{\partial}{\partial \theta_m} g_\theta(s) \right|^2 ds = 0, \quad \forall m = 1, 2, \cdots, L, \quad \forall \theta \in \Theta.$$

(iv) There exists a positive number $c > 0$ such that:

$$g_\theta(s) > c, \quad \forall \theta \in \Theta, \quad \forall s \in [-\pi, \pi].$$

(v) The matrix:

$$\Gamma(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta'} g_\theta(s) \frac{\partial}{\partial \theta} g_\theta(s) \frac{ds}{g_\theta^2(s)},$$

is nonsingular.

Assumptions 2(i)-(iv) are similar to the assumptions imposed in Davies (1973, A 1.1 to A 1.4). Assumption 2(v) is also assumed in Davies (1973, see, e.g., p. 482).[14]

As an example, consider the case where $\{w_{nt}\}_{t \in \mathbb{Z}}$ follows a stationary ARMA(p, q) process such that:

$$w_{nt} = a_1 w_{n,t-1} + a_2 w_{n,t-2} + \cdots + a_p w_{n,t-p} + u_{nt} + b_1 u_{n,t-1} + \cdots + b_q u_{n,t-q},$$

where $u_{nt} \sim i.i.d. N(0, \sigma^2)$ across n and t. We also assume that the polynomials $a(z) := 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p$ and $b(z) := 1 + b_1 z + b_2 z^2 + \cdots + b_q z^q$ have no common zeros and that $a(z)$ and $b(z)$ have no zeros on the unit circle. Then, the spectral density of $\{w_{nt}\}_{t \in \mathbb{Z}}$ is given by:

$$g_\theta(s) = \frac{\sigma^2}{2\pi} \frac{|b(e^{-is})|^2}{|a(e^{-is})|^2} = \frac{\sigma^2}{2\pi} \frac{b(e^{-is}) b(e^{is})}{a(e^{-is}) a(e^{is})}, \tag{4.3}$$

where $\theta = (a_1, \cdots, a_p, b_1, \cdots, b_q, \sigma^2)$. After some algebra, we can easily show that the spectral density of the ARMA model satisfies all the conditions in Assumption 2.

[14] The difference is that Davies (1973) states these conditions in terms of the autocovariance function $k \mapsto \gamma_k(\theta)$, while we state them in terms of the spectral density g_θ. The reason that we state these conditions in terms of the spectral density g_θ is that it enables us to check these conditions relatively easily.

4.3 Conditions on infinite-dimensional parameters

We first state the two sets of assumptions on the true values of factors and factor loadings. These assumptions are used to prove that the log-likelihood ratio has a well-defined limit.[15]

The first set of assumptions concerns individual effects or factor loadings.

Assumption 3. There exists some positive number $M > 0$ such that $\|\lambda_n\|_E \leq M$ for all $n \in \mathbb{Z}$, where $\|\cdot\|_E$ denotes the Euclidean norm. The matrix $\Sigma_{\lambda\lambda} = \lim_{N\to\infty} \sum_{n=1}^{N} \lambda_n \lambda_n'/N$ exists and is positive definite.

The second set of assumptions concerns time effects or factors, and these assumptions guarantee that they satisfy Grenander's conditions (see, e.g., Grenander and Rosenblatt (1957) and Anderson (1971)). Let f_{tl} be the l-th element of f_t. Let:

$$a_{lm}^T(j) := \sum_{t=1}^{T-|j|} f_{t+|j|,l} f_{tm}$$

for $l, m = 1, 2, \cdots, p$.

Assumption 4. (i) $a_{ll}^T(0) \to \infty$ as $T \to \infty$ for $l = 1, 2, \cdots, p$.

(ii) $f_{lT}/a_{ll}^T \to 0$ as $T \to \infty$ for $l = 1, 2, \cdots, p$.

(iii) The limit of:

$$r_{lm}^T(j) := \frac{a_{lm}^T(j)}{\sqrt{a_{ll}^T(0) a_{mm}^T(0)}}$$

exists and is finite for $l, m = 1, 2, \ldots, p$ and $j \in \mathbb{Z}$.

(iv) Define $\rho_{lm}(j) = \lim_{t\to\infty} r_{lm}^T(j)$, and let $R(j)$ be a $p \times p$ matrix whose (l, m)-th component is $\rho_{lm}(j)$. Assume that $R(0)$ is nonsingular.

4.4 Local parameter space

We now discuss the space of local parameters and the sequence of statistical experiments. Let ℓ_2^p denote a space of one-sided sequences of p-dimensional vectors α_n's ($n \in \mathbb{Z}$) with $\sum_{n=1}^{\infty} \|\alpha_n\|_E^2 < \infty$. Set $H_\dagger := \mathbb{R}^L \times \ell_2^p \times \ell_2^p$ and let $(\tilde{\theta}, \{\tilde{\lambda}_n\}_{n=1}^{\infty}, \{\tilde{f}_t\}_{t=1}^{\infty}) \in H_\dagger$. This $H_\dagger$ is the local parameter space used to apply the convolution theorem.

Note that the choice of a local parameter space is an important and delicate issue. The convolution theorem itself does not specify what kind of local parameter space is used as long

[15] When the additive-effects model (which is a special case of our model and is described as $y_{nt} = f_t + \lambda_n + w_{nt}$) is considered, these assumptions are not necessary. This is because f_t and λ_n do not appear in the log-likelihood ratio of the additive effects model.

as it is a linear subspace of a Hilbert space with an inner product that corresponds to the LAN result. Typically, using a "small" parameter space makes the derivation easy. However, the efficiency bound associated with a "small" parameter space may not be attainable. For example, the set $\{0\}$ is a Hilbert space, and we can apply the convolution theorem with $H = \{0\}$. However, the resulting efficiency bound is 0, which is a valid lower bound of the asymptotic variance, but it is not attainable. Using a "large" parameter space ensures that the bound is attainable, but handling a "large" space is often difficult. We thus need to find a local parameter space that is sufficiently small so that we can handle it easily and sufficiently large so that the resulting efficiency bound is attainable. It turns out that our choice of local parameter space, "$H_\dagger$", makes the analysis tractable and the resulting efficiency bound attainable.

The sequence of statistical experiments that we consider is as follows. Recall that D_T is the diagonal matrix whose l-th diagonal element is $\sqrt{a_{ll}^T(0)}$. We localize the parameters around the 'truth' as follows:

$$\theta + \frac{\tilde{\theta}}{\sqrt{NT}}, \quad \left\{\lambda_n + D_T^{-1}\tilde{\lambda}_n\right\}_{n=1}^{\infty} \quad \text{and} \quad \left\{f_t + \frac{\tilde{f}_t}{\sqrt{N}}\right\}_{t=1}^{\infty}.$$

If some elements of λ_n and/or f_t are known, then we set the corresponding elements of $\tilde{\lambda}_n$ or $\tilde{f}_t$ equal to zero. Furthermore, let $w_n = (w_{n1}, w_{n2}, \cdots, w_{nT})'$. Recall that $\Omega_T(\theta) = \mathbb{E}_\theta[w_n w_n']$. For simplicity of notation, we write $\Omega_{\tilde{\theta}} := \Omega_T\left(\theta + \tilde{\theta}/\sqrt{NT}\right)$ and $\Omega_0 := \Omega_T(\theta)$. Furthermore, we denote by $P_{NT,h}$ and $P_{NT,0}$ the laws of observations $\{\{y_{nt}\}_{t=1}^T\}_{n=1}^N$ under $(\theta + \tilde{\theta}/\sqrt{NT}, \{\lambda_n + D_T^{-1}\tilde{\lambda}_n\}_{n=1}^{\infty}, \{f_t + \tilde{f}_t/\sqrt{N}\}_{t=1}^{\infty})$ and $(\theta, \{\lambda_n\}_{n=1}^{\infty}, \{f_t\}_{t=1}^{\infty})$, respectively. Recall that $F_T = (f_1, \ldots, f_T)'$. Define $\tilde{F}_T$ similarly. Lastly, define $(\tilde{\theta}\nabla\Omega_0) = \sum_{m=1}^{L}(\partial\Omega_T(\theta))/(\partial\theta_m)$.

4.5 Main results

In this subsection, we show the convolution theorem for factor models. The log-likelihood ratio is:

$$\begin{aligned}
&\log\frac{dP_{NT,h}}{dP_{NT,0}} \\
= \; &\frac{N}{2}\log\det\Omega_0 - \frac{N}{2}\log\det\Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N}(y_n - F_T\lambda_n)'\Omega_0^{-1}(y_n - F_T\lambda_n) \\
&-\frac{1}{2}\sum_{n=1}^{N}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{N}}\right)\left(\lambda_n + D_T^{-1}\tilde{\lambda}_n\right)\right)'\Omega_{\tilde{\theta}}^{-1}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{N}}\right)\left(\lambda_n + D_T^{-1}\tilde{\lambda}_n\right)\right).
\end{aligned}$$

The limit of the log-likelihood ratio process is given by the following lemma.

Lemma 4.1. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. Then, for a statistical experiment $P_{NT,h}$, $h \in H_\dagger$, it holds that:*

$$\log\frac{dP_{NT,h}}{dP_{NT,0}} = \Delta_{NT,h}^{\dagger} - \frac{1}{2}\|h\|_\dagger^2 + o_{P_{NT,0}}(1),$$

where:

$$\Delta^{\dagger}_{NT,h} := \frac{1}{2\sqrt{NT}}\sum_{n=1}^{N}\left\{w_n'\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}w_n - \text{tr}\left[\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right]\right\}$$
$$+\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_0^{-1}w_n + \sum_{n=1}^{N}\tilde{\lambda}_n' D_T^{-1}F_T'\Omega_0^{-1}w_n, \tag{4.4}$$

and:

$$\|h\|_{\dagger}^2 := \tilde{\theta}'\Gamma(\theta)\tilde{\theta} + \sum_{n=1}^{\infty}\tilde{\lambda}_n'\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega_T(\theta)^{-1}F_T D_T^{-1}\right)\tilde{\lambda}_n$$
$$+\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n. \tag{4.5}$$

Under $P_{NT,0}$, as $N\to\infty$ and $T\to\infty$, $\Delta^{\dagger}_{NT,h}$ converges weakly to $\Delta^{\dagger}_h \sim N(0,\|h\|_{\dagger}^2)$.

Note that Assumption 4 ensures the existence of $\lim_{T\to\infty} D_T^{-1}F_T'\Omega_T(\theta)^{-1}F_T D_T^{-1}$ (see, e.g., Anderson (1971, Theorem 10.2.8)). The existence of $\lim_{N,T\to\infty}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n/N$ is guaranteed by Lemma A.9 and Assumption 3.

The following lemma shows that $H_{\dagger}$ is a Hilbert space equipped with the inner product that corresponds to the norm $\|\cdot\|_{\dagger}$ that appears in the expression of the log-likelihood ratio process. Because ℓ_2^p is a linear space, it is easy to see that $H_{\dagger}$ is a linear space. It is technically involved to show that the inner product in the lemma satisfies the conditions for an inner product and that the space is complete.

Lemma 4.2. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. Then, $H_{\dagger}$ is a Hilbert space with an inner product for $h_a, h_b \in H_{\dagger}$ given by:*

$$\langle h_a, h_b\rangle_{\dagger} := \tilde{\theta}_a'\Gamma(\theta)\tilde{\theta}_b + \sum_{n=1}^{\infty}\tilde{\lambda}_{aj}'\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_T D_T^{-1}\right)\tilde{\lambda}_{bj}$$
$$+\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\lambda_n. \tag{4.6}$$

The convolution theorem can now be applied because Lemmas 4.1 and 4.2 show that the sequence of statistical experiments $\{P_{NT,h}, h\in H_{\dagger}\}$ is LAN.

Theorem 4.2. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of statistical experiments $\{P_{NT,h}, h\in H_{\dagger}\}$. Suppose that the sequence of parameters $\kappa_{NT}(h)$ and that of estimators τ_{NT} are regular with respect to $\mathbf{r}_{NT}$. Then, the limit distribution L of $\mathbf{r}_{NT}(\tau_{NT} - \kappa_{NT}(0))$ equals the sum $G+W$ of independent, tight, Borel-measurable random elements in B such that:*

$$b^*G \sim N(0,\|\dot{\kappa}^*b^*\|_{\dagger}^2), \quad \forall b^*\in B^*,$$

where the adjoint map $\dot{\kappa}^*$ and the norm $\|\cdot\|_{\dagger}$ are defined under the inner product given in (4.6).

4.6 Applications

This section presents the findings summarized in Section 3 in a more mathematically rigorous manner. All of the results presented in this section are applications of Theorem 4.2. We do not repeat the discussions on the implications of these findings as they are already given in Section 3, so the discussion here is brief.

4.6.1 Efficiency bounds for factors and factor loadings

We first present the efficiency bound for a factor.

Theorem 4.3. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of statistical experiments* $\{P_{NT,h}, h \in H_\dagger\}$. *Fix* $t \in \mathbb{Z}$ *and take a factor* f_t. *Let* J_{NT} *be a* $p \times p$ *(nonrandom) matrix such that* $J_{NT} \to J$ *as* $N, T \to \infty$, *where* J *is invertible. The variance of the limit distribution of* $\sqrt{N}(\tau_{NT} - J'_{NT} f_t)$ *for any regular estimator* τ_{NT} *is not smaller than:*

$$\gamma(0) J' \Sigma_{\lambda\lambda}^{-1} J.$$

Next, we give the efficiency bound for a factor loading.

Theorem 4.4. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of statistical experiments* $\{P_{NT,h}, h \in H_\dagger\}$. *Fix* $n \in \mathbb{Z}$, *and take a factor loading* λ_n. *Let* J^*_{NT} *be a* $p \times p$ *(nonrandom) matrix such that* $J^*_{NT} \to J^*$ *as* $N, T \to \infty$, *where* J^* *is invertible. The variance of the limit distribution of* $D_T(\tau_{NT} - J^*_{NT}\lambda_n)$ *for any regular estimator* τ_{NT} *is not smaller than:*

$$J^* \left(\lim_{T\to\infty} D_T^{-1} F'_T \Omega_T(\theta)^{-1} F_T D_T^{-1} \right)^{-1} (J^*)'.$$

Lastly, we provide the efficiency bound for an interactive effect.

Theorem 4.5. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of statistical experiments* $\{P_{NT,h}, h \in H_\dagger\}$. *Fix* $t \in \mathbb{Z}$ *and* $n \in \mathbb{Z}$, *and take an interactive effect* $f'_t\lambda_n$. *Let* $r_{NT} = \min\{\sqrt{N}, \sqrt{a_{11}(0)}, \ldots, \sqrt{a_{pp}(0)}\}$. *The variance of the limit distribution of* $r_{NT}(\tau_{NT} - f'_t\lambda_n)$ *for any regular estimator* τ_{NT} *is not smaller than:*

$$f'_t \left(\lim_{N,T\to\infty} D_T^{-1} r_{NT} \right) \left(\lim_{T\to\infty} D_T^{-1} F'_T \Omega_T(\theta)^{-1} F_T D_T^{-1} \right)^{-1} \left(\lim_{N,T\to\infty} D_T^{-1} r_{NT} \right) f_t + \left(\lim_{N,T\to\infty} \frac{r_{NT}}{\sqrt{N}} \right)^2 \gamma(0) \lambda'_n \Sigma_{\lambda\lambda}^{-1} \lambda_n.$$

4.6.2 Efficiency bounds for common parameters

The next theorem provides the efficiency bound for any regular estimators of $\beta(\theta)$, where $\beta(\cdot)$ is a differentiable function from Θ to $\mathbb{R}^M$.

Theorem 4.6. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of experiments* $\{P_{NT,h}, h \in H_\dagger\}$*. Let* β *be a differentiable function from* Θ *into* $\mathbb{R}^M$*, and let* $\tau_{N,T}$ *be any regular estimator of* $\beta(\theta)$ *as* $N, T \to \infty$*. Let* $\dot{\beta}(\theta) = \partial\beta(\theta)/\partial\theta$*. The variance of the limit distribution of* $\sqrt{NT}(\tau_{NT} - \beta(\theta))$ *for any regular estimator* τ_{NT} *is not smaller than:*

$$\dot{\beta}(\theta)\Gamma(\theta)^{-1}\dot{\beta}(\theta)'.$$

We have discussed panel AR(1) models, panel MA(1) models and the estimation of autocovariances in Section 3. The efficiency bounds for panel AR(1) and panel MA(1) models can be derived by directly calculating the bound in Theorem 4.6. On the other hand, calculating the efficiency bound for the estimation of autocovariances requires some additional nontrivial theoretical derivation.

As argued in Section 3.2.3, we can use the results of Kakizawa and Taniguchi (1994) directly because the efficiency bound has the same form as that in the time-series context.

Theorem 4.7. *Suppose that Assumptions 1, 2, 3 and 4 hold. Consider the sequence of experiments* $\{P_{NT,h}, h \in H_\dagger\}$*.) Then, Okui's (2010) autocovariance estimator* $\tilde{\gamma}_k(\theta)$ *is asymptotically efficient if and only if there exists* $c \in \mathbb{R}^L$ *such that:*

$$g_\theta^2(s)\cos(ks) + c'\frac{\partial}{\partial\theta}g_\theta(s) = 0, \quad \forall s. \tag{4.7}$$

In particular, if $\{w_{nt}\}_{t\in\mathbb{Z}}$ *is a Gaussian stationary ARMA*(p, q) *process, then* $\tilde{\gamma}_k(\theta)$ *is asymptotically efficient if and only if:*

$$p \geq q \quad \text{and} \quad 0 \leq k \leq p - q. \tag{4.8}$$

The proof is omitted because it is essentially same as the matrix algebra given in Kakizawa and Taniguchi (1994). They comment that condition (4.7) is easy to check. They also show in Example 1 that for the case of a Gaussian stationary ARMA(p, q) process, condition (4.7) reduces to (4.8).

5 Discussion on local parameter spaces

In this section, we discuss subtle, yet important, issues concerning local parameter spaces for the infinite-dimensional parameters. The convolution theorem requires that the local parameter space be a linear subspace of a Hilbert space. However, it turns out that it is not a trivial task to construct an appropriate local parameter space. We first discuss the result of Hahn and Kuersteiner (2002), and then we discuss the difficulty arising when we apply their argument to the current situation. We also discuss an alternative approach based on the literature on "functional models" in statistics. While this alternative approach is considered in statistics, it is not possible to derive the efficiency bound for factors and factor loadings using this approach.

5.1 Models with individual effects

This subsection demonstrates the difficulty in constructing an appropriate local parameter space in models with only individual effects. Hahn and Kuersteiner (2002) provide the efficiency bound for panel AR(1) models, and their result is the seminal contribution to the literature on efficiency in large panel data. However, they do not discuss how to specify the local parameter space. We argue that specifying an appropriate local parameter space is far from trivial even in this simple model. Our analysis indicates that it is very difficult, if not impossible, to derive the efficiency bound for the estimation of individual effects (or factors and factor loadings in more general models) under their approach.

Hahn and Kuersteiner (2002) consider panel AR(1) models with individual effects. For expositional simplicity, we consider the following univariate model:

$$y_{nt} = \lambda_n + w_{nt},$$

where λ_n is a scalar individual effect.[16] While w_{nt} is assumed to follow an AR(1) process in Hahn and Kuersteiner (2002), the specification of the dynamics is not important in the following discussion, and we merely assume that the law of w_{nt} satisfies Assumptions 1 and 2. The parameter set of the model is $(\theta, \{\lambda_n\}_{n=1}^{\infty})$.

Following Hahn and Kuersteiner (2002), we perturb the parameter $(\theta, \{\lambda_n\}_{n=1}^{\infty}))$ as follows:

$$\left(\theta + \frac{1}{\sqrt{NT}}\tilde{\theta}, \left\{\lambda_n + \frac{1}{\sqrt{NT}}\tilde{\lambda}_n\right\}_{n=1}^{\infty}\right).$$

The local parameter is $h = (\tilde{\theta}, \{\tilde{\lambda}_n\}_{n=1}^{\infty}))$. Under this perturbation, the local log-likelihood ratio can be shown to be expanded in the following way; under $P_{NT,0}$, as $N, T \to \infty$:

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \Delta_{NT,h}^{+} - \frac{1}{2}\|h\|_{+}^{2} + o_{P_{NT,0}}(1),$$

where:

$$\begin{aligned} \Delta_{NT,h}^{+} &= \frac{1}{2\sqrt{NT}} \sum_{n=1}^{N} \left\{ w_n' \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} w_n - \operatorname{tr}\left(\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0)\right) \right\} \\ &\quad + \frac{1}{\sqrt{NT}} \sum_{n=1}^{N} \tilde{\lambda}_n \mathbf{1}_T' \Omega_0^{-1} w_n, \end{aligned} \tag{5.1}$$

and:

$$\|h\|_{+} := \tilde{\theta}' \Gamma(\theta) \tilde{\theta} + \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{\lambda}_n^2 \cdot \lim_{T\to\infty} \frac{1}{T} \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T.$$

[16] Hahn and Kuersteiner (2002) also consider multivariate models.

Under $P_{NT,0}$, the sequence $\Delta^+_{NT,h}$ converges weakly to $\Delta^+_h \sim N(0, \|h\|^2_+)$ as $N, T \to \infty$, provided that $\lim_{N\to\infty} \sum_{n=1}^N \tilde{\lambda}_n^2/N$ exists and is finite (the limit of $\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T/T$ always exists and equals $\sum_{k=-\infty}^{\infty} \gamma_k(\theta)$ under our assumptions on g_θ because the sequence $(1, 1, \cdots)$ satisfies Grenander's condition).[17]

While one may think that we can straightforwardly apply the convolution theorem at first glance, the situation is not that trivial. A key requirement of the convolution theorem is that the local parameter space be a linear subspace of a Hilbert space. Thus, it must be a linear space and must equip some appropriate inner product. We argue that it is not a trivial task to specify an appropriate local parameter space. Note that Hahn and Kuersteiner's (2002) argument is incomplete in the sense that they do not specify the local parameter space concretely.

Here, we present a (failed) attempt to construct a local parameter space. Let Ξ denote the local parameter space for individual effects. Because Hahn and Kuersteiner (2002) assume that the true sequence of individual effects satisfies $\sum_{n=1}^N \lambda_n^2/N = O(1)$ (Condition 4), we may consider the following restriction on Ξ:[18]

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \tilde{\lambda}_n^2 \text{ exists and finite.} \tag{5.2}$$

However, it turns out that Ξ with this restriction does not yield an appropriate local parameter space. The local parameter space given by the restriction (5.2) is not a linear subspace of a Hilbert space.[19] A sufficient condition for Ξ to be a linear space is that $\{\tilde{\lambda}_n\}_{n=1}^\infty$ is strongly Ces ⊠ ro convergent in the following sense (see Theorem 17 of Maddox (1970, p. 190)):

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} (\tilde{\lambda}_n - c)^2 = 0 \ \text{ for some } c \in \mathbb{R}.$$

[17]Our expansion and Hahn and Kuersteiner's do not coincide exactly, because of the difference of the parameterization for individual effects. More precisely, the formula in (5.1) does not reduce to that of Lemma 8 in Hahn and Kuersteiner (2002). However, a one-to-one reparametrization turns our model into Hahn and Kuersteiner's and vice versa, so the difference of the parameterization does not change the information contained in the models. In our case, we separate the parameter for the mean (λ_n) and the parameters for the variance (θ). On the other hand, Hahn and Kuersteiner's (2002) specification is $y_{nt} = \lambda_n/(1-\alpha) + w_{nt}$ and $w_{nt} = \alpha w_{n,t-1} + u_{nt}$, so the parameter α appears not only in the variance of y_{nt} but also in the mean. Because of this, their formula is more complicated than ours. Because the difference is not essential and our expansion is much simpler and easier to deal with, we make our arguments here using our expansion (5.1).

[18]We note that assuming $\sum_{n=1}^N \tilde{\lambda}_n^2/N = O(1)$ does not guarantee the existence of the limit of $\sum_{n=1}^N \tilde{\lambda}_n^2/N$. In fact, we can construct an example where $\sum_{n=1}^N \tilde{\lambda}_n^2/N$ is bounded but oscillates as $N \to \infty$ (see, e.g., Davidson (1994, p. 194)). Without assuming the existence of this limit, the first and second terms in the expansion of (5.1) may not converge.

[19]For example, take $a_n = 1$ for all i and $b_n = (-1)^k$ where k is determined by $2^k < n \le 2^{k+1}$. Then $\lim \sum_{n=1}^N a_n^2/N = \lim \sum_{n=1}^N b_n^2/N = 1$, but $\sum_{n=1}^N a_n b_n/N$ oscillates and so does $\sum_{n=1}^N (a_n + b_n)^2/N$.

However, this condition is not sufficient either because Hilbert space requires an appropriate inner product. The functional $(\{\tilde{\lambda}_{1n}\}_{n=1}^{\infty}, \{\tilde{\lambda}_{2n}\}_{n=1}^{\infty}) \mapsto \lim_{N\to\infty}\sum_{n=1}^{N}\tilde{\lambda}_{1n}\tilde{\lambda}_{2n}/N$ does not define an inner product but only defines a semi-inner product; i.e., the functional is not positive definite. A successful way to turn a semi-inner product into an inner product is to take a quotient space. In our case, however, taking a quotient of the space of strongly Ces ⊠ ro convergent sequences by $\{\{\tilde{\lambda}_n\}_{n=1}^{\infty} : \lim_{N\to\infty}\sum_{n=1}^{N}\tilde{\lambda}_n^2/N = 0\}$ implies that it is equivalent to imposing a nonperturbation to individual effects. This aspect is extremely unfavorable.

Thus, we need to specify the local parameter space in a totally different way for the convolution theorem to be applied. When there exist only individual effects, the approach considered in this paper corresponds to using the space $\{\{\tilde{\lambda}_n\}_{n=1}^{\infty} : \sum_{n=1}^{\infty}\tilde{\lambda}_n^2 < \infty\}$ and the perturbation to the parameter, $(\theta + \tilde{\theta}/\sqrt{NT}, \{\lambda_n + \tilde{\lambda}_n/\sqrt{T}\}_{n=1}^{\infty})$. Note that this approach is not adding more restrictions on Ξ but considers a different perturbation because the rate of localization is different.

On the other hand, there is an alternative approach based on the statistical literature on "functional models." We investigate this approach in more detail in the next subsection and examine its advantages and its limitations. Under this approach, the rate of localization for infinite-dimensional parameters is $\sqrt{NT}$. More precisely, in the context of Hahn and Kuersteiner (2002), we set $\Xi = \{\{\tilde{\lambda}_n\}_{n=1}^{\infty} | \tilde{\lambda}_n = \tilde{\phi}(\lambda_n)$, where $\tilde{\phi}(\cdot)$ is continuous and bounded$\}$ and the perturbation is $(\theta + \tilde{\theta}/\sqrt{NT}, \{\lambda_n + \tilde{\phi}(\lambda_n)/\sqrt{NT}\}_{n=1}^{\infty})$. Ξ is a Hilbert space, and θ is a regular parameter under this alternative approach. Therefore, Hahn and Kuersteiner's (2002) argument becomes complete by specifying the local parameter space for individual effects using the alternative approach. On the other hand, it turns out that an element of individual effects is not a regular parameter, and the convolution theorem does not provide the efficiency bound. Thus, if our objective is to derive the efficiency bound for the estimation of individual effects, the approach taken by Hahn and Kuersteiner (2002) is not appropriate.[20]

5.2 Alternative approach

In this section, we come back to the factor model and examine the approach originated in the "functional model" in the statistics literature for specifying the local parameter space. We examine how this traditional approach works and discuss why we do not use it in this paper. The main point is that under this approach, factors and factor loadings are not regular parameters, and thus, we are not able to derive the efficiency bound for their estimation using

[20]Hahn and Kuersteiner (2002) provide the efficiency bound for the estimation of the AR(1) parameter and do not discuss the bound for an individual effect.

the convolution theorem. On the other hand, they are regular parameters in our specification of the local parameter space, which enables us to apply the convolution theorem.

We first impose some conditions on the "true" value of the parameters. Let $C_b(\mathbb{R}^p)$ be the space of continuous and bounded functions from $\mathbb{R}^p$ to $\mathbb{R}^p$.

Assumption 5. 1. There exists a probability measure Φ such that:

(i) $\lim_{N\to\infty}(1/N)\sum_{n=1}^N \phi(\lambda_n)$ converges weakly to $\int \phi(\lambda)d\Phi(\lambda)$, $\forall\phi \in C_b(\mathbb{R}^p)$;

(ii) $\lim_{N\to\infty}(1/N)\sum_{n=1}^N \lambda_n\phi(\lambda_n)'$ converges weakly to $\int \lambda\phi(\lambda)d\Phi(\lambda)$, $\forall\phi \in C_b(\mathbb{R}^p)$;

(iii) $\lim_{N\to\infty}(1/N)\sum_{n=1}^N \lambda_n\lambda_n' = \int \lambda\lambda' d\Phi(\lambda) > 0$.

2. There exists some continuous and bounded function g from $[0,1]$ to $\mathbb{R}^p$ such that $f_t = \psi(t/T)$ and that $\lim_{T\to\infty}(1/T)\sum_{t=1}^T f_t = \int_0^1 \psi(a)da$; $\int_0^1 \psi(a)\psi(a)'ds > 0$.

Assumption 5.1 requires that the sequence of λ_n behave as if it were a realization of some random vector whose law is Φ. This assumption is similar to that considered by Strasser (1996, 1998). Assumption 5.2 implies that the dynamics of f_t is smooth. This specification of f_t is similar to that considered in Robinson (1989) for the space of time-varying parameters. These conditions are used to guarantee that the log-likelihood ratio process has a well-defined limit.

Next, we define the space of local parameters. Let $C_\lambda = (C_b(\mathbb{R}))^p$ such that $\tilde{\phi} \in C_\lambda$ can be written as $\tilde{\phi}(\lambda) = (\tilde{\phi}^{(1)}(\lambda_1), \ldots, \tilde{\phi}^{(p)}(\lambda_p))'$. Each $\tilde{\phi}^{(m)}$ is a function of λ_m only and does not depend on λ_l for $l \neq m$. Let $\psi^{(m)}(s)$ be the m-th element of vector $\psi(s)$. Let $C_b^{(m)}([0,1])$ be the space of continuous and bounded functions from $[0,1]$ to $\mathbb{R}$ that are orthogonal to $\psi^{(m)}$. Let $C_f = \prod_{m=1}^p C_b^{(m)}([0,1])$. The space of local parameters is $H_\ddagger = \mathbb{R}^L \times C_\lambda \times C_f$. The restrictions in C_f and C_λ are needed to make the functional appearing in the limit of the local log-likelihood ratio an inner product.

We give several remarks on the parameter space. The space of factors is different from that of factor loadings. These choices are made for technical reasons. The space C_λ is not suitable for time-varying parameters in models with serially correlated errors. The factors need to satisfy "Grenander's conditions" (see, e.g., Grenander and Rosenblatt (1957) and Anderson (1971)) to show the convergence of the log-likelihood ratio. However, the restriction imposed by C_λ does not necessarily guarantee Grenander's conditions. On the other hand, the space C_f is useful to show the convergence but implicitly requires that the parameters be naturally ordered. The factors depend on t, so it is naturally ordered. However, the factor loadings depend on n, but there is no natural ordering among cross-sectional units. Therefore, it would not be appropriate to use a space similar to C_f for the parameter space of local factor loadings.

We consider localizing the parameters around the 'truth' $(\theta, \{\lambda_n\}_{n=1}^{\infty}, \{f_t\}_{t=1}^{\infty})$ with a disappearing order of size $O(1/\sqrt{NT})$:

$$\theta + \frac{\tilde{\theta}}{\sqrt{NT}}, \quad \lambda_n + \frac{\tilde{\lambda}_n}{\sqrt{NT}} \quad \text{and} \quad f_t + \frac{\tilde{f}_t}{\sqrt{NT}}, \tag{5.3}$$

for $n = 1, \ldots, N$ and $t = 1, \ldots, T$, where $\tilde{\theta} \in \mathbb{R}^L$, $\tilde{\lambda}_n = \tilde{\phi}(\lambda_n)$ with $\tilde{\phi} \in C_\lambda$ and $\tilde{f}_t = \tilde{\psi}(t/T)$ with $\tilde{\psi} \in C_f$. Note that under this approach, the orders of localization for factors and factor loadings are different from those in the approach stated in Section 4.

We now examine the limit of the local log-likelihood ratio process. The local log-likelihood ratio process is given by:

$$\begin{aligned}
&\log \frac{dP_{NT,h}}{dP_{NT,0}} \\
= \; & \frac{N}{2} \log\det\Omega_0 - \frac{N}{2}\log\det\Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N}(y_n - F_T\lambda_n)'\Omega_0^{-1}(y_n - F_T\lambda_n) \\
& -\frac{1}{2}\sum_{n=1}^{N}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{NT}}\right)\left(\lambda_n + \frac{\tilde{\lambda}_n}{\sqrt{NT}}\right)\right)'\Omega_{\tilde{\theta}}^{-1}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{NT}}\right)\left(\lambda_n + \frac{\tilde{\lambda}_n}{\sqrt{NT}}\right)\right).
\end{aligned}$$

Under some regularity conditions, we can show that:

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \Delta_{NT,h}^{\ddagger} - \frac{1}{2}\|h\|_{\ddagger}^2 + o_{P_{NT,0}}(1),$$

where:

$$\begin{aligned}
\Delta_{NT,h}^{\ddagger} \;&:=\; \frac{1}{2\sqrt{NT}}\sum_{n=1}^{N}\left\{w_n'\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}w_n - \mathrm{tr}\left[\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right]\right\} \\
&\quad + \frac{1}{\sqrt{NT}}\sum_{n=1}^{N}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}w_n,
\end{aligned} \tag{5.4}$$

and:

$$\|h\|_{\ddagger}^2 = \tilde{\theta}'\Gamma(\theta)\tilde{\theta} + \lim_{N,T\to\infty}\frac{1}{NT}\sum_{n=1}^{N}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n).$$

The following lemma shows that $H_{\ddagger}$ is a linear subspace of a Hilbert space under the inner product that is appropriate for our purpose.

Lemma 5.1. *Suppose that Assumptions 1, 2 and 5 hold. Then, $H_{\ddagger}$ is a linear subspace of a Hilbert space with an inner product for $h_l, h_m \in H_{\ddagger}$ given by:*

$$\begin{aligned}
\langle h_l, h_m\rangle_{\ddagger} \;&:=\; \tilde{\theta}_l'\Gamma(\theta)\tilde{\theta}_m + \lim_{N,T\to\infty}\frac{1}{NT}\sum_{n=1}^{N}(\tilde{F}_{lT}\lambda_n + F_T\tilde{\lambda}_{ln})'\Omega_0^{-1}(\tilde{F}_{mT}\lambda_n + F_T\tilde{\lambda}_{mn}) \qquad (5.5)\\
&= \tilde{\theta}_l'\Gamma(\theta)\tilde{\theta}_m + \int_{\mathbb{R}^L}\int_0^1(\psi(a)'\tilde{\phi}_l(\lambda) + \lambda'\tilde{\psi}_l(a))(\psi(a)'\tilde{\phi}_m(\lambda) + \lambda'\tilde{\psi}_m(a))\,da\,d\Phi(\lambda).
\end{aligned}$$

We also show that $\{P_{NT,h}, h \in H_\ddagger\}$ is LAN.

Lemma 5.2. *Suppose that Assumptions 1, 2 and 5 hold. Then, $\{P_{NT,h}, h \in H_\ddagger\}$ is LAN in the sense that:*

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \Delta^\ddagger_{NT,h} - \frac{1}{2}\|h\|^2_\ddagger + o_{P_{NT,0}}(1),$$

where $\Delta^\ddagger_{NT,h}$ is defined in (5.4) and converges weakly under $P_{NT,0}$ to $\Delta^\ddagger_h \sim N(0, \|h\|^2_\ddagger)$ and $\Delta^\ddagger_h$ is an iso-Gaussian process with a covariance function $\mathbb{E}_\theta[\Delta^\ddagger_{h_1}\Delta^\ddagger_{h_2}] = \langle h_1, h_2\rangle_\ddagger$ and $\langle \cdot, \cdot \rangle_\ddagger$ is defined in (5.5)

Because we have shown that $H_\ddagger$ is a linear subspace of a Hilbert space and $\{P_{NT,h}, h \in H_\ddagger\}$ is LAN, the convolution theorem can be applied, and the efficiency bound is derived.

Theorem 5.1. *Suppose that Assumptions 1, 2 and 5 hold. Consider the sequence of statistical experiments $\{P_{NT,h}, h \in H_\ddagger\}$. Suppose that the sequence of parameters $\kappa_{NT}(h)$ and estimators τ_{NT} are regular with respect to $\mathbf{r}_{NT}$. Then, the limit distribution L of $\mathbf{r}_{NT}(\tau_{NT} - \kappa_{NT}(0))$ equals the sum $G + W$ of independent, tight, Borel-measurable random elements in B such that:*

$$b^* G \sim N(0, \|\dot{\kappa}^* b^*\|^2_\ddagger), \quad \forall b^* \in B^*,$$

where the adjoint map $\dot{\kappa}^*$ and the norm $\|\cdot\|_\ddagger$ are defined under the inner product given in (5.5).

The efficiency bound for the estimation of θ can thus be derived using Theorem 5.1. It is not difficult to show that the common parameter θ is a regular parameter. The efficiency bound under this approach is the same as that presented in 4.6.

The problem of this approach is that factors and factor loadings are not regular parameters, as shown in the next theorem.

Theorem 5.2. *Suppose that Assumption 5 is satisfied.*

Fix $n \in \mathbb{N}$. Let $\kappa_{NT}(h) = \lambda_n + \tilde{\phi}(\lambda_n)/\sqrt{T}$ for $\tilde{\phi} \in C_\lambda$. Suppose that measure Φ in Assumption 5 does not have a point mass on λ_n. Then, $h \mapsto \dot{\kappa}(h) = \lim_{N,T\to\infty} \sqrt{T}(\kappa_{NT}(h) - \kappa_{NT}(0)) = \tilde{\phi}(\lambda_n)$ is not continuous on $H_\ddagger$ and, thus, $\kappa_{NT}(h)$ is not regular.

Similarly, if we define $\kappa_{NT}(h) = f_t + \tilde{\psi}(f_t)/\sqrt{N}$ fixing $t \in \mathbb{N}$ for $\tilde{\psi} \in C_f$, then $h \mapsto \dot{\kappa}(h) = \lim_{N,T\to\infty} \sqrt{N}(\kappa_{NT}(h) - \kappa_{NT}(0)) = \tilde{\psi}(f_t)$ is not continuous on $H_\ddagger$, and thus $\kappa_{NT}(h)$ is not regular.

This theorem implies that we cannot derive the efficiency bound for the estimation of factors and factor loadings using this alternative approach. The convolution theorem requires that the parameter be regular. Suppose that λ_n is the parameter of interest. We then set $\kappa_{NT}(h) =$

$\lambda_n + \tilde{\phi}(\lambda_n)/\sqrt{T}$ and $r_{NT} = \sqrt{T}$. That $\kappa_{NT}(h)$ is regular means that $\dot{\kappa}$ should exist and be a *continuous* linear functional on $H_\ddagger$. It is obvious that $\dot{\kappa}(h) = \lim_{N,T\to\infty} r_{NT}(\kappa_{NT}(h) - \kappa_{NT}(0))$ exists and is given by $\phi(\lambda_n)$. Although the functional $h \mapsto \dot{\kappa}(h)$ is linear, it turns out that it is not continuous on $H_\ddagger$, in general.[21] A similar argument holds for f_t. Therefore, factors and factor loadings are not regular parameters, and the convolution theorem cannot be applied. We note that regularity depends on the choice of the local parameter space. These parameters are regular under $H_\dagger$, as shown in Section 4.

The regularity of $\kappa_{NT}(h)$ corresponds to a so-called 'differentiable parameter' assumption (see, e.g., van der Vaart (1991)), which is known to play a fundamental role in a semiparametric version of the convolution theorem. In fact, Theorem 2.1 in van der Vaart (1991) implies that when the derivative of the functional of parameters exists, the continuity of the derivative on a tangent set is necessary for the existence of regular estimators for the functional.[22] Hirano and Porter (2012) prove the same nonexistence result under slightly stronger assumptions, but in a simpler way.[23] Furthermore, Theorem 4.1 in van der Vaart (1991) implies that a differentiable parameter assumption is necessary and sufficient for the efficiency bound in a convolution theorem to be well defined. These results are derived under i.i.d. assumptions. Because we consider panel data, we cannot apply these results directly. Nonetheless, we conjecture that there is no regular estimator for λ_n and f_t and that the efficiency bounds for λ_n and f_t are not well defined under the localization (5.3) with local parameter space $H_\ddagger$.

6 Conclusion

In this paper, we investigate the asymptotic efficiency in general dynamic panel data models with factor structure when both the cross-sectional sample size and the length of the time series tend to infinity. By using the infinite-dimensional convolution theorem of van der Vaart and Wellner (1996), the efficiency bounds for the estimation of factors, factor loadings and common parameters are derived. It should be emphasized that the derivation is nontrivial and complex because of the presence of factor structure. In particular, how to define the local parameter space for the infinite- dimensional parameters is nontrivial.

We show that the efficiency bound of factors or factor loadings is not affected by the presence

[21]Our proof of Theorem 5.2 closely follows the argument in Example 3.1.1 in Bickel, et al. (1993).

[22]A 'tangent set' for semiparametric models with i.i.d. observations corresponds to a local parameter space H in the infinite-dimensional convolution theorem. For more details, see Example 3.11.1 of van der Vaart and Wellner (1996).

[23]Hirano and Porter (2012) also show no existence results of locally asymptotically unbiased estimators for nondifferentiable functionals, which is the main result of their paper.

of other parameters. These efficiency bounds are attainable in the sense that there exists an estimator whose asymptotic variance is equal to the bound. The efficiency results for the common parameters obtained here are analogous to those observed in time-series contexts, and this implies that the presence of interactive effects, which is an infinite-dimensional nuisance parameter, does not affect the form of the efficiency bound.

The theoretical results of this paper can be extended in various ways. For example, we may consider the efficiency bound for regression models with factor error structure. The estimation is considered by, for example, Bai (2009a), Moon and Weidner (2010) and Sarafidis and Yamagata (2010). It is also interesting to consider the case with heteroscedasticity and/or cross-sectional dependence. Because there exist estimators for those cases, such as Breitung and Tenhofen (2011), it would be important to investigate whether those estimators achieve the efficiency bound in more general settings.

A Appendix

A.1 Preliminaries

In this subsection, we list some properties concerning covariance matrices for stationary processes associated with spectral density g_θ. The following notation is used to state those properties. We denote the trace operator by $\mathrm{tr}[\cdot]$. For any matrix A, we define $\|A\|_E := (\mathrm{tr}(A'A))^{1/2}$ (the Euclidean norm) and $\|A\|_B := \sup_{\|x\|_E=1} \|Ax\|_E$ (the Banach norm) where x is a vector conformable with A. Note that for any Euclidean vector a, these two norms coincide, and the norm is denoted by $\|a\|_E$. We repeatedly use the fact that for any conformable matrices A and B, we have the relation $\|AB\|_E \leq \|A\|_B \|B\|_E$.

Lemma A.1. *Suppose that Assumptions 1 and 2 are satisfied. Then, the following results hold.*

(i)

$$\|\Omega_T(\theta)\|_B \leq 2\pi \sup_s g_\theta(s) \leq \sum_{k=-\infty}^{\infty} |\gamma_k(\theta)| < \infty,$$

and:

$$\|\Omega_T(\theta+\epsilon) - \Omega_T(\theta)\|_B \leq 2\pi \sup_s |g_{\theta+\epsilon}(s) - g_\theta(s)| \to 0, \quad \text{as } \epsilon \to 0.$$

(ii)

$$\|\Omega_T^{-1}(\theta)\|_B \leq \frac{1}{2\pi} \sup_s g_\theta^{-1}(s) < \infty,$$

and:

$$\sup_{\|\epsilon\|_E<\delta} \|\Omega_T^{-1}(\theta+\epsilon)\| \leq \frac{1}{2\pi} \sup_{\|\epsilon\|_E<\delta} \sup_s |g_{\theta+\epsilon}^{-1}(s)| < \infty \quad \text{for some } \delta > 0.$$

(iii)

$$\frac{1}{T}\left\|\frac{\partial}{\partial\theta_m}\Omega_T(\theta)\right\|_E^2 \leq \sum_{k=-\infty}^{\infty}\left|\frac{\partial}{\partial\theta_m}\gamma_k(\theta)\right|^2 = \int_{-\pi}^{\pi}\left|\frac{\partial}{\partial\theta_m}g_\theta(s)\right|^2 ds < \infty,$$

and:

$$\begin{aligned}\frac{1}{T}\left\|\frac{\partial}{\partial\theta_m}\Omega_T(\theta+\epsilon) - \frac{\partial}{\partial\theta_m}\Omega_T(\theta)\right\|_E^2 &\leq \sum_{k=-\infty}^{\infty}\left|\frac{\partial}{\partial\theta_m}\gamma_k(\theta+\epsilon) - \frac{\partial}{\partial\theta_m}\gamma_k(\theta)\right|^2 \\ &= \int_{-\pi}^{\pi}\left|\frac{\partial}{\partial\theta_m}g_{\theta+\epsilon}(s) - \frac{\partial}{\partial\theta_m}g_\theta(s)\right|^2 ds \to 0, \quad \text{as } \epsilon \to 0.\end{aligned}$$

(iv)

$$\frac{1}{\sqrt{T}}\left\|\frac{\partial}{\partial\theta_m}\Omega_T(\theta)\right\|_B \to 0 \quad \text{as } T \to \infty.$$

Proof. This lemma is a special case of Corollary 3.3 in Davies (1973), which shows this lemma in the setting of a multivariate Gaussian stationary process.[24] □

A.2 Proofs of the theoretical results in Section 4.3

A.2.1 Proof of Lemma 4.1

Under the perturbation induced by a local parameter in $H_\dagger$, the local log-likelihood ratio process is given by:

$$\begin{aligned}&\log\frac{dP_{NT,h}}{dP_{NT,0}} \\ =\ &\frac{N}{2}\log\det\Omega_0 - \frac{N}{2}\log\det\Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N}(y_n - F_T\lambda_n)'\Omega_0^{-1}(y_n - F_T\lambda_n) \\ &-\frac{1}{2}\sum_{n=1}^{N}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{N}}\right)\left(\lambda_n + D_T^{-1}\tilde{\lambda}_n\right)\right)'\Omega_{\tilde{\theta}}^{-1}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{N}}\right)\left(\lambda_n + D_T^{-1}\tilde{\lambda}_n\right)\right) \\ =\ &\frac{N}{2}\log\det\Omega_0 - \frac{N}{2}\log\det\Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N}w_n'\Omega_0^{-1}w_n - \frac{1}{2}\sum_{n=1}^{N}w_n'\Omega_{\tilde{\theta}}^{-1}w_n \\ &+\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}w_n - \frac{1}{2}\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{A}_n,\end{aligned}$$

[24] Note that there is a minor difference between Davies' result and ours in the expression of the spectral density g_θ: Davies (1973) uses $\{\exp(-2\pi ims)\}_{m\in\mathbb{Z}}$ as a complete orthonormal set of $L^2(0,1]$ (the set of all square-integrable functions defined on $(0,1]$), whereas we use $\left\{\frac{1}{2\pi}\exp(-ims)\right\}_{m\in\mathbb{Z}}$ as a complete orthonormal set of $L^2(-\pi,\pi]$. However, this difference only affects the expressions, but there is no essential difference.

where:

$$\tilde{A}_n = \frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n + \frac{1}{\sqrt{N}}\tilde{F}_T D_T^{-1}\tilde{\lambda}_n.$$

Lemma A.2. *Suppose that Assumptions 1 and 2 hold. Under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$:*

$$\begin{aligned}\frac{N}{2}\log\det\Omega_0 - \frac{N}{2}\log\det\Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N} w_n'\Omega_0^{-1}w_n - \frac{1}{2}\sum_{n=1}^{N} w_n'\Omega_{\tilde{\theta}}^{-1}w_n \\ = \frac{1}{2\sqrt{NT}}\sum_{n=1}^{N}\left\{w_n'\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}w_n - \mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right)\right\} \\ -\frac{1}{4T}\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right) + \mathrm{o_p}(1).\end{aligned}$$

Proof. This proof is very similar to that of Theorem 4.4 in Davies (1973), and the only difference is that we also need to consider the cross-sectional dimension. The proof is, thus, omitted. □

Lemma A.3. *Suppose that Assumption 2 holds. As $T \to \infty$:*

$$\frac{1}{2T}\mathrm{tr}\left[\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right] = \tilde{\theta}'\Gamma(\theta)\tilde{\theta} + o(1).$$

Proof. This lemma is a special case of Theorem 4.4 in Davies (1973). □

Lemma A.4. *Suppose that Assumptions 1, 2, 3 and 4 hold. Under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$:*

$$\begin{aligned}\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}w_n - \frac{1}{2}\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{A}_n = \sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}w_n \\ -\frac{1}{2}\sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right) + o_p(1).\end{aligned}$$

Proof. First observe that:

$$\begin{aligned}&\mathbb{E}_\theta\left|\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}w_n - \sum_{n=1}^{N}\tilde{A}_n'\Omega_0^{-1}w_n\right|^2 \\ =\;& \sum_{n=1}^{N}\tilde{A}_n'(\Omega_{\tilde{\theta}}^{-1} - \Omega_0^{-1})\Omega_\theta(\Omega_{\tilde{\theta}}^{-1} - \Omega_0^{-1})\tilde{A}_n \\ \leq\;& \left(\sum_{n=1}^{N}\tilde{A}_n'\tilde{A}_n\right)\|\Omega_0^{-1}\|_B^2\|\Omega_{\tilde{\theta}}^{-1}\|_B^2\|\Omega_\theta\|_B\|\Omega_0 - \Omega_{\tilde{\theta}}\|_B.\end{aligned} \tag{A.1}$$

Now, we have:

$$
\begin{aligned}
& \sum_{n=1}^{N} \tilde{A}_n' \tilde{A}_n \\
= & \sum_{n=1}^{N} \left(\frac{1}{\sqrt{N}} \tilde{F}_T \lambda_n + F_T D_T^{-1} \tilde{\lambda}_n + \frac{1}{\sqrt{N}} \tilde{F}_T D_T^{-1} \tilde{\lambda}_n \right)' \left(\frac{1}{\sqrt{N}} \tilde{F}_T \lambda_n + F_T D_T^{-1} \tilde{\lambda}_n + \frac{1}{\sqrt{N}} \tilde{F}_T D_T^{-1} \tilde{\lambda}_n \right) \\
= & \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} (\tilde{f}_t' \lambda_n)^2 + \sum_{n=1}^{N} \sum_{t=1}^{T} (f_t' D_T^{-1} \tilde{\lambda}_n)^2 + \frac{2}{\sqrt{N}} \sum_{n=1}^{N} \sum_{t=1}^{T} \tilde{f}_t' \lambda_n f_t' D_T^{-1} \tilde{\lambda}_n \\
& + \frac{2}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \tilde{f}_t' D_T^{-1} \tilde{\lambda}_n \tilde{f}_t' \lambda_n + \frac{2}{\sqrt{N}} \sum_{n=1}^{N} \sum_{t=1}^{T} \tilde{f}_t' D_T^{-1} \tilde{\lambda}_n f_t' D_T^{-1} \tilde{\lambda}_n \\
& + \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} (\tilde{f}_t' D_T^{-1} \tilde{\lambda}_n)^2.
\end{aligned}
$$

The first term is:

$$
\frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} (\tilde{f}_t' \lambda_n)^2 \leq \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \tilde{f}_t' \tilde{f}_t \lambda_n' \lambda_n = \frac{1}{N} \sum_{n=1}^{N} \lambda_n' \lambda_n \sum_{t=1}^{T} \tilde{f}_t' \tilde{f}_t = O(1),
$$

by the Cauchy–Schwarz inequality, Assumption 3 and the definition of $H_\dagger$. The second term of the right-hand side is:

$$
\sum_{n=1}^{N} \sum_{t=1}^{T} (f_t' D_T^{-1} \tilde{\lambda}_n)^2 \leq \sum_{n=1}^{N} \sum_{t=1}^{T} f_t' D_T^{-1} D_T^{-1} f_t \tilde{\lambda}_n' \tilde{\lambda}_n = \sum_{n=1}^{N} \tilde{\lambda}_n' \tilde{\lambda}_n \sum_{t=1}^{T} f_t' D_T^{-1} D_T^{-1} f_t = O(1),
$$

by the Cauchy–Schwarz inequality and the definitions of $H_\dagger$ and D_T. For the third term, we have:

$$
\begin{aligned}
\left| \frac{2}{\sqrt{N}} \sum_{n=1}^{N} \sum_{t=1}^{T} \tilde{f}_t' \lambda_n f_t' D_T^{-1} \tilde{\lambda}_n \right| & \leq 2 \left(\frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} (\tilde{f}_t' \lambda_n)^2 \right)^{1/2} \left(\frac{1}{\ } \sum_{n=1}^{N} \sum_{t=1}^{T} (f_t' D_T^{-1} \tilde{\lambda}_n)^2 \right)^{1/2} \\
& = O(1),
\end{aligned}
$$

by the Cauchy–Schwarz inequality and the arguments that show the orders of the first and the second terms. Similarly, we can show that the fourth to sixth terms are $o(1)$. Therefore, we show that $\sum_{n=1}^{N} \tilde{A}_n' \tilde{A}_n/(NT) = O(1)$ so that the extreme right-hand side of (A.1) is $o(1)$ as $N, T \to \infty$, by Lemma A.1 (i) and (ii). Thus, we have:

$$
\sum_{n=1}^{N} \tilde{A}_n' \Omega_{\tilde{\theta}}^{-1} w_n = \sum_{n=1}^{N} \tilde{A}_n' \Omega_0^{-1} w_n + o_p(1).
$$

We observe that

$$
\sum_{n=1}^{N} \tilde{A}_n' \Omega_0^{-1} w_n = \sum_{n=1}^{N} \left(\frac{1}{\sqrt{N}} \tilde{F}_T \lambda_n + F_T D_T^{-1} \tilde{\lambda}_n \right)' \Omega_0^{-1} w_n + \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tilde{F}_T D_T^{-1} \tilde{\lambda}_n)' \Omega_0^{-1} w_n.
$$

The mean of the second term of the right-hand side is zero and the variance is:

$$
\begin{aligned}
\mathbb{E}_\theta\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}w_n\right)^2 &= \frac{1}{N}\sum_{n=1}^{N}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n) \\
&\leq \frac{1}{N}||\Omega_0^{-1}||_B\sum_{n=1}^{N}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n)'(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n) \\
&= ||\Omega_0^{-1}||_B\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}(\tilde{f}_t D_T^{-1}\tilde{\lambda}_n)^2 = o(1),
\end{aligned}
$$

by the definition of $H_\dagger$ and Assumption 4. Thus, we have:

$$
\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}w_n = \sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}w_n + o_p(1). \tag{A.2}
$$

Next, observe that:

$$
\left|\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{A}_n - \sum_{n=1}^{N}\tilde{A}_n'\Omega_0^{-1}\tilde{A}_n\right| \leq \left(\sum_{n=1}^{N}\tilde{A}_n'\tilde{A}_n\right)\|\Omega_{\tilde{\theta}}^{-1}\|_B\|\Omega_0^{-1}\|_B\|\Omega_{\tilde{\theta}} - \Omega_0\|_B = o(1),
$$

because we have already shown that $\sum_{n=1}^{N}\tilde{A}_n'\tilde{A}_n = O(1)$. Hence, we have:

$$
\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{A}_n = \sum_{n=1}^{N}\tilde{A}_n'\Omega_0^{-1}\tilde{A}_n + o(1),
$$

as $N, T \to \infty$. We now have:

$$
\begin{aligned}
&\sum_{n=1}^{N}\tilde{A}_n'\Omega_0^{-1}\tilde{A}_n \\
= \;&\sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right) \\
&+2\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}\tilde{F}_T D_T^{-1}\tilde{\lambda}_n + \frac{1}{N}\sum_{n=1}^{N}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}\tilde{F}_T D_T^{-1}\tilde{\lambda}_n.
\end{aligned}
$$

By the Cauchy–Schwarz inequality, we have:

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(\frac{1}{\sqrt{N}}\tilde{F}_T\lambda_n + F_T D_T^{-1}\tilde{\lambda}_n\right)'\Omega_0^{-1}\tilde{F}_T D_T^{-1}\tilde{\lambda}_n\right| \\
\leq \;&\frac{1}{\sqrt{N}}||\Omega_0^{-1}||_B\left(\left(\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}(\tilde{f}_t'\lambda_n)^2\right)^{1/2} + \left(\sum_{n=1}^{N}\sum_{t=1}^{T}(f_t'D_T^{-1}\tilde{\lambda}_n)^2\right)^{1/2}\right) \\
&\times\left(\sum_{n=1}^{N}\sum_{t=1}^{T}(\tilde{f}_t'D_T^{-1}\tilde{\lambda}_n)^2\right)^{1/2} = o(1),
\end{aligned}
$$

by Assumption 3 and the definitions of $H_\dagger$ and D_T^{-1}. We also have:

$$
\frac{1}{N}\sum_{n=1}^{N}(\tilde{F}_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}\tilde{F}_T D_T^{-1}\tilde{\lambda}_n \leq \frac{1}{N}||\Omega_0^{-1}||_B\sum_{n=1}^{N}\sum_{t=1}^{T}(\tilde{f}_t D_T^{-1}\tilde{\lambda}_n)^2 = o(1),
$$

by Assumptions 2 and the definition of H_1. Thus, we have:

$$\frac{1}{NT}\sum_{n=1}^{N}\tilde{A}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{A}_n = \frac{1}{NT}\sum_{n=1}^{N}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n) + o(1), \tag{A.3}$$

as $N, T \to \infty$.

Combining (A.2) and (A.3) yields the desired result. □

The next two lemmas provide the expression for the "norm part" in the expansion of the local log-likelihood ratio.

Lemma A.5. *Suppose that Assumptions 1, 2, 3 and 4 hold. For $h \in H_\dagger$, as $N, T \to \infty$:*

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\tilde{\lambda}_n' D_T^{-1}F_T'\Omega(\theta)^{-1}\tilde{F}_T\lambda_n = o(1).$$

Proof. First observe that:

$$\begin{aligned}
\left|\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\tilde{\lambda}_n' D_T^{-1}F_T'\Omega(\theta)^{-1}\tilde{F}_T\lambda_n\right| &\leq \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left\|\lambda_n\tilde{\lambda}_n'\right\|_E \left\|D_T^{-1}F_T\right\|_E \left\|\Omega(\theta)^{-1}\right\|_B \left\|\tilde{F}_T\right\|_E \\
&\leq \frac{B}{\sqrt{N}}\sum_{n=1}^{N}\left\|\tilde{\lambda}_n\right\|_E \left\|D_T^{-1}F_T\right\|_E \left\|\Omega(\theta)^{-1}\right\|_B \left\|\tilde{F}_T\right\|_E .
\end{aligned}$$

Obviously, $\left\|D_T^{-1}F_T\right\|_E$, $\left\|\Omega(\theta)^{-1}\right\|_B$, and $\left\|\tilde{F}_T\right\|_E$ are $O(1)$ as $T \to \infty$. Below we show that $(1/\sqrt{N})\sum_{n=1}^{N}\|\tilde{\lambda}_n'\|_E = o(1)$ as $N \to \infty$. To show this, fix $\epsilon > 0$. Because $\{\tilde{\lambda}_n\}_{n\in\mathbb{Z}}$ is square summable, there exists some $N_1 \in \mathbb{Z}$ such that for any $s, t \geq N_1$, we have:

$$\sum_{n=s+1}^{t}\|\tilde{\lambda}_n\|^2 < \frac{\epsilon}{2}.$$

Now, take $N_2 \in \mathbb{Z}$ large enough to ensure that:

$$\frac{1}{\sqrt{N_2}}\sum_{n=1}^{N_1}\|\tilde{\lambda}_n\|_E < \epsilon/2. \tag{A.4}$$

Such a number N_2 exists because N_1 is finite. Then, for any $N > \max\{N_1, N_2\}$, we have:

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left\|\tilde{\lambda}_n\right\|_E &= \frac{1}{\sqrt{N}}\sum_{n=1}^{M}\left\|\tilde{\lambda}_n\right\|_E + \frac{1}{\sqrt{N}}\sum_{n=N_1+1}^{N}\left\|\tilde{\lambda}_n\right\|_E \\
&< \frac{\epsilon}{2} + \frac{\sqrt{N-N_1}}{\sqrt{N}}\left(\sum_{n=N_1+1}^{N}\left\|\tilde{\lambda}_n\right\|_E^2\right)^{1/2} \\
&< \epsilon,
\end{aligned}$$

where the second inequality follows from (A.4) and the Cauchy–Schwarz inequality. This completes the proof. □

Lemma A.6. *Suppose that Assumptions 1 and 2 hold. For* $\{\tilde{f}_t\}_{t=1}^{\infty} \in \ell_2^p$*, the matrix sequence* $\{\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\}_{T=1}^{\infty}$ *converges as* $T \to \infty$.

Proof. If $\tilde{f}_t = 0$ for all $t \in \mathbb{N}$, then the result trivially holds. Thus, we assume that $\{\tilde{f}_t\}_{t=1}^{\infty}$ is not the zero vector in ℓ_2^p.

Because the matrix $\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T$ is symmetric, the convergence of $\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T$ is equivalent to the convergence of $x'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T x$ for any $x \in \mathbb{R}^p$. Thus, we may, without loss of generality, assume that $\tilde{f}_t$ is one dimensional (i.e., $p = 1$), so that $\tilde{F}_T$ is a vector, rather than a matrix.

Our plan is to show that the sequence in question is a Cauchy sequence. That is, we show that for any $\epsilon > 0$, we can choose a sufficiently large $T_0 \in \mathbb{N}$ such that:

$$\left|\tilde{F}_{T_2}'\Omega_{T_2}(\theta)^{-1}\tilde{F}_{T_2} - \tilde{F}_{T_1}'\Omega_{T_1}(\theta)^{-1}\tilde{F}_{T_1}\right| < \epsilon, \quad \forall T_1, T_2 \geq T_0.$$

We now fix $\epsilon > 0$ and take a positive number δ such that:

$$\delta < \min\left\{\frac{\epsilon\pi}{3\left(\sum_{t=1}^{\infty}\tilde{f}_t^2\right)\left(\sup_s g_\theta^{-1}(s)\right)^2}, \frac{\inf_s g_\theta(s)}{2}\right\}.$$

The reason for this choice of δ will become clear below.

The main difficulty of the proof is that the sequence involves the inverse matrix $\Omega_T(\theta)^{-1}$, whose form is unclear in general and whose order grows to infinity as $T \to \infty$. We overcome technical difficulties caused by the matrix $\Omega_T(\theta)^{-1}$ by approximating it by the covariance matrix of some autoregressive process. Let $a(z) := 1 - a_1 z - a_2 z^2 - \cdots - a_m z^m$ ($z \in \mathbb{C}$ and $m \in \mathbb{N}$), and assume that $a(z)$ has no roots inside the unit circle and $A(\lambda) := (K/2\pi)|a(e^{-i\lambda})|^{-2}$ where $K > 0$. The function $A(\lambda)$ is the spectral density of an autoregressive process $\{X_t\}_{t=-\infty}^{\infty}$ of order m whose autoregressive coefficients are $a_1, a_2, \cdots, a_m$ and whose innovation variance is K. From Corollary 4.4.2. of Brockwell and Davis (1990), we can choose the polynomial $a(z)$ and the positive number K such that:

$$|g_\theta(\lambda) - A(\lambda)| < \delta \ \ \forall\lambda \in [\pi, \pi]. \tag{A.5}$$

For each $T \in \mathbb{N}$, let H_T be the covariance matrix of $X_1, X_2, \cdots, X_T$.

To prove (A.5), we observe that for any $T_1, T_2 \in \mathbb{N}$:

$$\begin{aligned}
&\left|\tilde{F}_{T_2}'\Omega_{T_2}(\theta)^{-1}\tilde{F}_{T_2} - \tilde{F}_{T_1}'\Omega_{T_1}(\theta)^{-1}\tilde{F}_{T_1}\right| \\
\leq\ & |\tilde{F}_{T_2}'\Omega_{T_2}(\theta)^{-1}\tilde{F}_{T_2} - \tilde{F}_{T_2}'H_{T_2}^{-1}\tilde{F}_{T_2}| + |\tilde{F}_{T_2}'H_{T_2}^{-1}\tilde{F}_{T_2} - \tilde{F}_{T_1}'H_{T_1}^{-1}\tilde{F}_{T_1}| \\
&+ |\tilde{F}_{T_1}'H_{T_1}^{-1}\tilde{F}_{T_1} - \tilde{F}_{T_1}'\Omega_{T_1}(\theta)^{-1}\tilde{F}_{T_1}|.
\end{aligned}$$

We show that each absolute value on the right-hand side of the inequality is bounded by $\epsilon/3$.

We first show the following result:

$$|\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T - \tilde{F}_T'H_T^{-1}\tilde{F}_T| < \frac{\epsilon}{3}, \quad \forall T \in \mathbb{N}. \tag{A.6}$$

To show this, notice that the (j,k)-th component of H_T can be written as $\int_{-\pi}^{\pi} e^{i(j-k)}A(\lambda)d\lambda$. Then, for any $T \in \mathbb{N}$ and any $x = (x_1, x_2, \cdots, x_T)' \in \mathbb{R}^T$ with $\|x\|_E = 1$, we have:

$$\begin{aligned} |x'\Omega_T(\theta)x - x'H_Tx| &= \left|\sum_{t=1}^{T}\sum_{s=1}^{T} x_t x_s \int_{-\pi}^{\pi} e^{i(t-s)}(g_\theta(\lambda) - A(\lambda))d\lambda\right| \\ &= \left|\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T} x_t e^{it\lambda}\right|^2 (g_\theta(\lambda) - A(\lambda))\, d\lambda\right| \\ &< \delta\int_{-\pi}^{\pi}\left|\sum_{t=1}^{T} x_t e^{it\lambda}\right|^2 d\lambda = 2\pi\delta. \end{aligned}$$

Consequently, we have $\|\Omega(\theta) - H_T\|_B \leq 2\pi\delta$ for all T. We also note that $\|H_T^{-1}\|_B \leq (1/2\pi)\sup_s A^{-1}(s) \leq (1/\pi)\sup_s g_\theta^{-1}(s)$.[25] Therefore, it holds that for any T:

$$\begin{aligned} |\tilde{F}_T'\Omega(\theta)^{-1}\tilde{F}_T - \tilde{F}_T'H_T^{-1}\tilde{F}_T| &= |\tilde{F}_T'H_T^{-1}(H_T - \Omega(\theta))\Omega(\theta)^{-1}\tilde{F}_T| \\ &\leq \|\tilde{F}_T\|_E^2\|H_T^{-1}\|_B\|\Omega(\theta)^{-1}\|_B\|H_T - \Omega(\theta)\|_B \\ &\leq \frac{\delta\left(\sum_{t=1}^{\infty}\tilde{f}_t^2\right)\left(\sup_s g_\theta^{-1}(s)\right)^2}{\pi} \\ &< \frac{\epsilon}{3}, \end{aligned}$$

where the last inequality follows from our choice of δ. This completes the proof of (A.6).

It remains to prove that for any sufficiently large T_1 and T_2, we have:

$$|\tilde{F}_{T_2}'H_{T_2}^{-1}\tilde{F}_{T_2} - \tilde{F}_{T_1}'H_{T_1}^{-1}\tilde{F}_{T_1}| < \frac{\epsilon}{3}.$$

To show this result, we write the matrix H_T^{-1} in a tractable form. As in Brockwell and Davis (1991, p. 381), we can use the Gram–Schmidt orthogonalization procedure to produce a white-

[25]Here is the proof of this inequality. We show that $\inf_s g_\theta(s)/2 \leq \inf_s A(s)$, which implies the desired result. Because $|g_\theta(s) - A(s)| < \delta$, we have $g_\theta(s) < A(s) + \delta$. Thus, $\inf_s g_\theta(s) - \delta \leq \inf_s A(s)$. Because $\delta < \inf_s g_\theta(s)/2$, it follows that $\inf_s g_\theta(s)/2 \leq \inf_s A(s)$.

noise process $\{W_t\}_{t=1}^{\infty}$ with variance K, as follows:

$$
\begin{aligned}
&W_1 := b_{11}X_1 \\
&W_2 := b_{21}X_1 + b_{22}X_2 \\
&\vdots \\
&W_m := b_{m1}X_1 + \cdots + b_{mm}X_m \\
&W_{m+1} := -a_m X_1 - \cdots - a_1 X_m + X_{m+1} \\
&\vdots
\end{aligned}
$$

Now, we define a $T \times T$ lower triangular matrix R_T by:

$$
R_T := \begin{pmatrix}
b_{11} & & & & & & & \\
b_{21} & b_{22} & & & & & & \\
\vdots & \vdots & & & & & & \\
b_{m1} & b_{m2} & \dots & b_{mm} & & & & \\
-a_m & -a_{m-1} & \dots & -a_1 & 1 & & & \\
 & -a_m & \dots & -a_2 & -a_1 & 1 & & \\
 & & & & & & & \\
 & & & -a_m & \dots & -a_2 & -a_1 & 1
\end{pmatrix}.
$$

Then:

$$
\begin{pmatrix} W_1 \\ \vdots \\ W_T \end{pmatrix} = R_T \begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix}.
$$

From this, it can be easily seen that $\Omega_T(\theta)^{-1} = (1/K)R_T' R_T$. Now, we define $a_0 := 1$. Because $\sum_{t=1}^{\infty} \tilde{f}_t^2 < \infty$, we can pick a natural number T_0 such that:

$$
\sum_{t=T_1+1}^{T_2} \tilde{f}_t^2 < \frac{K\epsilon}{3m^2 \max_{0 \le j \le m} a_j^2}, \quad \forall T_1, T_2 \ge T_0 - m.
$$

Thus, for any $T_1, T_2 \in \mathbb{N}$ with $T_2 > T_1 \geq T_0$:

$$\begin{aligned}
|\tilde{F}'_{T_2} H_{T_2}^{-1} \tilde{F}_{T_2} - \tilde{F}'_{T_1} H_{T_1}^{-1} \tilde{F}_{T_1}| &= \frac{1}{K} |\tilde{F}'_{T_2} R'_{T_2} R_{T_2} \tilde{F}_{T_2} - \tilde{F}'_{T_1} R'_{T_1} R_{T_1} \tilde{F}_{T_1}| \\
&= \frac{1}{K} \sum_{t=T_1+1}^{T_2} \left(\sum_{j=0}^{m} a_j \tilde{f}_{t-j} \right)^2 \\
&\leq \frac{\max_{0 \leq j \leq m} a_j^2}{K} \sum_{t=T_1+1}^{T_2} \left(\sum_{j=0}^{m} |\tilde{f}_{t-j}| \right)^2 \\
&\leq \frac{\max_{0 \leq j \leq m} a_j^2}{K} \left| \sum_{j=0}^{m} \sqrt{\sum_{t=T_1+1}^{T_2} |\tilde{f}_{t-j}|^2} \right|^2 \\
&< \frac{\epsilon}{3}.
\end{aligned}$$

where the last two inequalities follow from the Minkowski inequality and from our choice of T_0. □

The next lemma gives the asymptotic distribution of $\Delta^{\dagger}_{NT,h}$.

Lemma A.7. *Suppose that Assumptions 1, 2, 3 and 4 hold. Let $\Delta^{\dagger}_{NT,h}$ be defined as in (4.4). Under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$:*

$$\Delta^{\dagger}_{NT,h} \xrightarrow{d} N\left(0, \|h\|_{\dagger}^2\right),$$

where $\|h\|_{\dagger}^2$ is defined in (4.5).

Proof. Note that Lemmas A.5 and A.6 are used to give the expression of $\|h\|_{\dagger}^2$.

We use Theorem 2 in Phillips and Moon (1999), which is a Lindberg–Levy-type central limit theorem for double-indexed stochastic processes. We define:

$$Q_{nT} = \frac{1}{2\sqrt{T}} \left\{ w'_n \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} w_n - \operatorname{tr}\left[\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \right] \right\} + (\tilde{F}_T \lambda_n + \sqrt{N} F_T D_T^{-1} \tilde{\lambda}_n)' \Omega_0^{-1} w_n,$$

and:

$$\sigma_{NT} := \sum_{n=1}^{N} \mathbb{E}_\theta [Q_{nT}^2] \quad \text{and} \quad \xi_{n,NT} := \sigma_{NT}^{-\frac{1}{2}} Q_{nT}.$$

First, observe that:

$$
\begin{aligned}
& \frac{1}{N}\sigma_N \\
= \; & \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_\theta\left|\frac{1}{2\sqrt{T}}\left\{w_n'\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\Omega_0^{-1}w_n - \mathrm{tr}\left[\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\right]\right\} + (\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}w_n\right|^2 \\
= \; & \frac{1}{2T}\mathrm{tr}\left[\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\right] + \frac{1}{N}\sum_{n=1}^{N}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n) \\
& + \frac{1}{N\sqrt{T}}\sum_{n=1}^{N}\mathbb{E}_\theta[w_n'\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\Omega_0^{-1}w_n(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}w_n] \\
= \; & \frac{1}{2T}\mathrm{tr}\left[\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\right] + \frac{1}{N}\sum_{n=1}^{N}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n),
\end{aligned}
$$

where the last equality follows from the fact that w_n is Gaussian so that its third moment is zero. By Lemmas A.3, A.5 and A.6, it follows that as $N, T \to \infty$:

$$
\frac{1}{N}\sigma_{NT} \to \tilde\theta'\Gamma(\theta)\tilde\theta + \lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}(\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n) = \|h\|_\dagger.
$$

Hence, if:

$$
\sum_{n=1}^{N}\xi_{n,NT} \xrightarrow{d} N(0,1), \tag{A.7}
$$

as $N, T \to \infty$, then the desired convergence (A.10) is proved.

To establish the convergence (A.7), it is sufficient to show the Lindberg condition in Theorem 2 in Phillips and Moon; i.e., we show that, for each $\epsilon > 0$, the term:

$$
\sum_{n=1}^{N}\mathbb{E}_\theta\left[\xi_{n,NT}^2 1\{\xi_{n,NT}^2 > \epsilon\}\right] = \sigma_{NT}^{-1}\sum_{n=1}^{N}\mathbb{E}_\theta\left[Q_{nT}^2 1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}\right], \tag{A.8}
$$

converges to 0 as $N, T \to \infty$, where $1\{\cdot\}$ denotes an indicator function. Let:

$$
Q_{1nT} = \frac{1}{2\sqrt{T}}\left\{w_n'\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\Omega_0^{-1}w_n - \mathrm{tr}\left[\Omega_0^{-1}(\tilde\theta\nabla\Omega_0)\right]\right\},
$$

and:

$$
Q_{2nT} = (\tilde F_T\lambda_n + \sqrt{N}F_TD_T^{-1}\tilde\lambda_n)'\Omega_0^{-1}w_n.
$$

Then, we have $Q_{nT} = Q_{1nT} + Q_{2nT}$ and $Q_{nT}^2 \le 2Q_{1nT}^2 + 2Q_{2nT}^2$. It can be seen by the Cauchy–Schwarz inequality that:

$$
\begin{aligned}
\mathbb{E}_\theta\left[Q_{nT}^2 1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}\right] & \le \mathbb{E}_\theta\left[(2Q_{1nT}^2 + 2Q_{2nT}^2)1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}\right] \\
& \le 2\left(\mathbb{E}_\theta[Q_{1nT}^4] + \mathbb{E}_\theta[Q_{2nT}^4]\right)^{1/2}\left(\mathbb{E}_\theta[1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}]\right)^{1/2}.
\end{aligned}
$$

We have:

$$\begin{aligned}\mathbb{E}_\theta[1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}] &= \Pr(\sigma_{NT}^{-1}Q_{nT}^2 > \epsilon) \\ &\leq \Pr(2Q_{1nT}^2 + 2Q_{2nT}^2 > \epsilon\sigma_{NT}) \\ &\leq \Pr(2Q_{1nT}^2 > \epsilon\sigma_{NT}/2) + \Pr(2Q_{2nT}^2 > \epsilon\sigma_{NT}/2).\end{aligned}$$

By the Chebyshev inequality, it holds that:

$$\Pr(2Q_{1nT}^2 > \epsilon\sigma_{NT}/2) \leq \frac{16\mathbb{E}_\theta[Q_{1nT}^4]}{\epsilon^2\sigma_{NT}^2}.$$

Let $\alpha(a) := \Pr(|Z| > a)$, where Z is the standard Gaussian random variable. Because Q_{2nT} is Gaussian with mean zero, we have:

$$\Pr(2Q_{2nT}^2 > \epsilon\sigma_{NT}/2) = \alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4E(Q_{2nT}^2)}}\right).$$

Thus, it holds that:

$$\begin{aligned}&\sigma_{NT}^{-1}\sum_{n=1}^{N}\mathbb{E}_\theta\left[Q_{nT}^2 1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}\right] \\ \leq\ & \sigma_{NT}^{-1}\sum_{n=1}^{N} 2\left(\mathbb{E}_\theta[Q_{1nT}^4] + \mathbb{E}_\theta[Q_{2nT}^4]\right)^{1/2}\left(\frac{16\mathbb{E}_\theta[Q_{1nT}^4]}{\epsilon^2\sigma_{NT}^2} + \alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4E(Q_{2nT}^2)}}\right)\right)^{1/2} \\ \leq\ & \left(\sup_{1\leq n\leq N}\frac{16\mathbb{E}_\theta[Q_{1nT}^4]}{\epsilon^2\sigma_{NT}^2} + \sup_{1\leq n\leq N}\alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4E(Q_{2nT}^2)}}\right)\right)^{1/2}\frac{1}{\sigma_{NT}}\sum_{n=1}^{N}2\left(\mathbb{E}_\theta[Q_{1nT}^4] + \mathbb{E}_\theta[Q_{2nT}^4]\right)^{1/2} \\ \leq\ & \left(\frac{16N^2}{\epsilon^2\sigma_{NT}^2}\frac{1}{N^2}\sup_{1\leq n\leq N}\mathbb{E}_\theta[Q_{1nT}^4] + \sup_{1\leq n\leq N}\alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4E(Q_{2nT}^2)}}\right)\right)^{1/2} \\ & \times\frac{2N}{\sigma_{NT}}\left(\frac{1}{N}\sum_{n=1}^{N}(\mathbb{E}_\theta[Q_{1nT}^4] + \mathbb{E}_\theta[Q_{2nT}^4])\right)^{1/2}.\end{aligned}$$

We note that Q_{1nT} is i.i.d. across n so that:

$$\sup_{1\leq n\leq N}\mathbb{E}_\theta[Q_{1nT}^4] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_\theta[Q_{1nT}^4] = \mathbb{E}_\theta[Q_{1nT}^4].$$

Now, we have:

$$\mathbb{E}_\theta[Q_{1nT}^4] = \frac{1}{T^2}\mathbb{E}_\theta\left|w_n'\Omega_0^{-1}\left(\tilde{\theta}\nabla\Omega_0\right)\Omega_0^{-1}w_n - \mathrm{tr}\left[\Omega_0^{-1}\left(\tilde{\theta}\nabla\Omega_0\right)\right]\right|^4.$$

Let $B_T := \Omega_0^{-1}\left(\tilde{\theta}\nabla\Omega_0\right)\Omega_0^{-1}$ and observe that:

$$\begin{aligned}&\frac{1}{T^2}\mathbb{E}_\theta\left|w_n'B_Tw_n - \mathrm{tr}\left[B_T\Omega_0\right]\right|^4 \\ =\ & \frac{1}{T^2}\Big\{\mathbb{E}_\theta\left|w_n'B_Tw_n\right|^4 - 4\mathrm{tr}\left[B_T\Omega_0\right]\mathbb{E}_\theta\left|w_n'B_Tw_n\right|^3 + 6\left(\mathrm{tr}\left[B_T\Omega_0\right]\right)^2\mathbb{E}_\theta\left|w_n'B_Tw_n\right|^2 \\ & -4\left(\mathrm{tr}\left(B_T\Omega_0\right)\right)^3\mathbb{E}_\theta\left|w_n'B_Tw_n\right| + \left(\mathrm{tr}\left(B_T\Omega_0\right)\right)^4\Big\}.\end{aligned}$$

Applying the formulas for the third and fourth moments of quadratic forms in Gaussian random vectors (see, e.g., Theorem 10.21 in Schott (2006)), we see that:

$$\begin{aligned}\frac{1}{T^2}\mathbb{E}_\theta\left|w_n' B_T\Omega_0\Omega_0^{-1}w_n - \operatorname{tr}(B_T\Omega_0)\right|^4 &= \frac{12}{T^2}\left(\operatorname{tr}\left\{(B_T\Omega_0)^2\right\}\right)^2 + \frac{48}{T^2}\operatorname{tr}\left\{(B_T\Omega_0)^4\right\} \\ &\leq \frac{60}{T^2}\left\|\tilde{\theta}\nabla\Omega_0\right\|_E^4 \|\Omega_0^{-1}\|_B^4.\end{aligned}$$

The extreme right-hand side is $O(1)$ as $N, T \to \infty$ by Lemma 1 (ii) and (iii). Therefore, $\mathbb{E}_\theta[Q_{1nT}^4] = O(1)$. Next, we consider:

$$\frac{1}{N}\sum_{n=1}^N \mathbb{E}_\theta[Q_{2nT}^4] = \frac{1}{N}\sum_{n=1}^N \mathbb{E}_\theta\left|(\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}w_n\right|^4.$$

Because w_n is Gaussian, it holds that:

$$\mathbb{E}_\theta\left|(\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}w_n\right|^4 = 3\left((\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n)\right)^2.$$

We also observe that:

$$\begin{aligned}&(\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + \sqrt{N}F_T D_T^{-1}\tilde{\lambda}_n) \\ \leq\ & 2||\Omega_0^{-1}||_B(\lambda_n'\tilde{F}_T'\tilde{F}_T\lambda_n + N\tilde{\lambda}_n' D_T^{-1}F_T'F_T D_T^{-1}\tilde{\lambda}_n) \\ =\ & 2||\Omega_0^{-1}||_B\left(\sum_{t=1}^T(\tilde{f}_t'\lambda_n)^2 + N\sum_{t=1}^T(f_t'D_T^{-1}\tilde{\lambda}_n)^2\right).\end{aligned}$$

Thus, we have:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^N \mathbb{E}_\theta[Q_{2nT}^4] &\leq \frac{12}{N}\sum_{n=1}^N(||\Omega_0^{-1}||_B)^2\left(\sum_{t=1}^T(\tilde{f}_t'\lambda_n)^2 + N\sum_{t=1}^T(f_t'D_T^{-1}\tilde{\lambda}_n)^2\right)^2 \\ &\leq \frac{24}{N}\sum_{n=1}^N(||\Omega_0^{-1}||_B)^2\left(\left(\sum_{t=1}^T(\tilde{f}_t'\lambda_n)^2\right)^2 + \left(N\sum_{t=1}^T(f_t'D_T^{-1}\tilde{\lambda}_n)^2\right)^2\right) \\ &\leq 24(||\Omega_0^{-1}||_B)^2\left(\frac{1}{N}\sum_{n=1}^N(\lambda_n'\lambda_n)^2\left(\sum_{t=1}^T\tilde{f}_t'\tilde{f}_t\right)^2 + N\sum_{n=1}^N(\tilde{\lambda}_n'\tilde{\lambda}_n)^2\left(\sum_{t=1}^T f_t'D_T^{-2}f_t\right)^2\right) \\ &= O(N),\end{aligned}$$

by Assumption 5 and the definitions of D_T^{-1} and $H_\dagger$. Furthermore, we note that:

$$\sup_{1\leq n\leq N}\alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4E(Q_{2nT}^2)}}\right) = \alpha\left(\sqrt{\frac{\epsilon\sigma_{NT}}{4\sup_{1\leq n\leq N}E(Q_{2nT}^2)}}\right) = o(N^{-2}),$$

by Assumption 5 and the fact that the tail of the Gaussian distribution decreases at an exponential rate. Because σ_{NT}/N converges, it holds that:

$$\sigma_{NT}^{-1}\sum_{n=1}^N \mathbb{E}_\theta\left[Q_{nT}^2 1\{|\sigma_{NT}^{-1}Q_{nT}^2| > \epsilon\}\right] = o(1).$$

Therefore, the Lindberg condition is satisfied, and the proof is complete.

□

Lemma 4.1 follows from Lemmas A.2, A.3, A.4 and A.7. □

A.2.2 Proof of Lemma 4.2

To prove Lemma 4.2, we first show several technical lemmas.

Lemma A.8. *Suppose that Assumptions 1, 2 and 3 are satisfied. For any two elements,* $\tilde{F}_a = \{\tilde{f}_{at}\}_{t=1}^{\infty}$ *and* $\tilde{F}_b = \{\tilde{f}_{bt}\}_{t=1}^{\infty}$ *in* ℓ_2^p*, the following results hold.*

(i) The sequence $\tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}$ *converges as* $T \to \infty$*.*

(ii) The limit $\lim_{N,T\to\infty} \sum_{n=1}^{N} \lambda_n' \tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\lambda_n/N$ *exists. Furthermore, it is given by:*

$$\lim_{N,T\to\infty} \frac{1}{N}\sum_{n=1}^{N} \lambda_n' \tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\lambda_n = \mathrm{tr}\left(\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \lambda_n\lambda_n' \lim_{T\to\infty} \tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\right).$$

Proof. To show the first result, it suffices to prove that for any $x, y \in \mathbb{R}^p$, the sequence $x'\tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}y$ converges as $T \to \infty$. Note that we can write:

$$\begin{aligned} x'\tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}y =& \frac{1}{4}\left(x'\tilde{F}_{aT}' + y'\tilde{F}_{bT}'\right)\Omega_T(\theta)^{-1}\left(\tilde{F}_{aT}x + \tilde{F}_{bT}y\right) \\ &- \frac{1}{4}\left(x'\tilde{F}_{aT}' - y'\tilde{F}_{bT}'\right)\Omega_T(\theta)^{-1}\left(\tilde{F}_{aT}x - \tilde{F}_{bT}y\right). \end{aligned} \tag{A.9}$$

Because $\{x'\tilde{f}_{at}\}_{t=1}^{\infty}$ and $\{y'\tilde{f}_{bt}\}_{t=1}^{\infty}$ are both in ℓ_2^1, so are the sum $\{x'\tilde{f}_{at} + y'\tilde{f}_{bt}\}_{t=1}^{\infty}$ and the difference $\{x'\tilde{f}_{at} - y'\tilde{f}_{bt}\}_{t=1}^{\infty}$. Thus, we can apply Lemma A.6 to show that each of the two terms in the right-hand side of (A.9) converges. Hence, the left-hand side also converges. This proves the first result.

To show the second result, we write:

$$\frac{1}{N}\sum_{n=1}^{N} \lambda_n' \tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\lambda_n = \mathrm{tr}\left(\frac{1}{N}\sum_{n=1}^{N} \lambda_n\lambda_n' \tilde{F}_{Ta}'\Omega_T(\theta)^{-1}\tilde{F}_{bT}\right).$$

Because Assumption 3 ensures the existence of $\lim_{N\to\infty}\sum_{n=1}^{N}\lambda_n\lambda_n'/N$, the desired result immediately follows from the first result. □

Lemma A.9. *Suppose that Assumptions 1, 2 and 3 are satisfied. There exist two positive numbers* c *and* C *such that:*

$$c\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2 \le \lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n \le C\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2,$$

for all $\tilde{F} = \{\tilde{f}_t\}_{t=1}^{\infty} \in \ell_2^p$ *(note that the two numbers* c *and* C *do not depend on* $\tilde{F}$*).*

Proof. For any square matrix A, we denote the maximum and minimum eigenvalues of A by

$\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. Observe that for each fixed T and N:

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n &\leq \frac{\lambda_{\max}(\Omega_T(\theta)^{-1})}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\tilde{F}_T\lambda_n \\
&= \frac{1}{\lambda_{\min}(\Omega_T(\theta))N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\tilde{F}_T\lambda_n \\
&\leq \frac{1}{(\inf_s g_\theta(s))\,N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\tilde{F}_T\lambda_n,
\end{aligned}$$

where the last inequality follows from the fact that the $\inf_s g_\theta(s) \leq \lambda_{\min}(\Omega_T(\theta))$ for any T (see, for example, Proposition 4.5.3 in Brockwell and Davis (1991)). Letting $N, T \to \infty$ yields:

$$\begin{aligned}
&\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n \\
\leq\ & \frac{1}{\inf_s g_\theta(s)}\mathrm{tr}\left(\lim_{T\to\infty}\tilde{F}_T'\tilde{F}_T\left(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n\lambda_n'\right)\right) \\
\leq\ & \frac{\lambda_{\max}(\Sigma_{\lambda\lambda})}{\inf_s g_\theta(s)}\mathrm{tr}\left(\lim_{T\to\infty}\tilde{F}_T'\tilde{F}_T\right) \\
=\ & \frac{\lambda_{\max}(\Sigma_{\lambda\lambda})}{\inf_s g_\theta(s)}\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2.
\end{aligned}$$

Similar arguments show that:

$$\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n \geq \frac{\lambda_{\min}(\Sigma_{\lambda\lambda})}{\sup_s g_\theta(s)}\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2.$$

By positive definiteness of $\Sigma_{\lambda\lambda}$, the eigenvalues $\lambda_{\min}(\Sigma_{\lambda\lambda})$ and $\lambda_{\max}(\Sigma_{\lambda\lambda})$ are strictly positive. Hence, setting:

$$c := \frac{\lambda_{\min}(\Sigma_{\lambda\lambda})}{\sup_s g_\theta(s)} > 0$$

and:

$$C := \frac{\lambda_{\max}(\Sigma_{\lambda\lambda})}{\inf_s g_\theta(s)} > 0,$$

we obtain the desired result. □

Corollary A.1. *Suppose that Assumptions 1, 2 and 3 are satisfied. Let $\tilde{F} = \{\tilde{f}_t\}_{t=1}^{\infty}$ be an element in ℓ_2^p. Then, $\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n = 0$ if and only if $\tilde{F} = 0$.*

Proof. The sufficiency part is obvious. To prove the necessity part, assume that:

$$\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n = 0.$$

By the preceding lemma, there exists a positive number $c > 0$ such that:

$$c\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2 \leq \lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_T'\Omega_T(\theta)^{-1}\tilde{F}_T\lambda_n = 0.$$

This implies that $\sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2 = 0$. Hence, $\tilde{F}$ must be zero. □

Using the above results, we can prove that $\langle h_a, h_b\rangle_\dagger$ is well defined and forms an inner product on $H_\dagger$. We can also show that $H_\dagger$ itself is a Hilbert space.

It is easy to see that $H_\dagger$ is a linear space. Furthermore, for any $h_a, h_b \in H_\dagger$, the quantity $\langle h_a, h_b\rangle_\dagger$ is well defined because the second and third terms in (4.6) converge by Assumption 4 and Lemma A.8, respectively. We show that $\langle\cdot,\cdot\rangle_\dagger$ satisfies the requirements of an inner product. Note that there are five requirements: for any $h_a, h_b, h_c \in H$ (i) $\langle h_a, h_b\rangle_\dagger = \langle h_a, h_b\rangle_\dagger$; (ii) $\langle h_a + h_b, h_c\rangle_\dagger = \langle h_a, h_c\rangle_\dagger + \langle h_b, h_c\rangle_\dagger$; (iii) $\langle \alpha h_a, h_b\rangle_\dagger = \alpha\langle h_a, h_b\rangle_\dagger$, where α is a scalar; (iv) $\langle h_a, h_a\rangle_\dagger \geq 0$; (v) $\langle h_a, h_a\rangle_\dagger = 0$ if and only if $h_a = 0$. It is again easy to see that the requirements (i) to (iv) are satisfied. We verify that the requirement (v) is also satisfied. If $h_a = 0$, then it is trivial that $\langle h_a, h_a\rangle_\dagger = 0$. To show the converse implication, assume that $\langle h_a, h_a\rangle_\dagger = 0$. Then, because all the three terms in $\langle h_a, h_a\rangle_\dagger$ are nonnegative, it follows that:

$$\begin{aligned}
\tilde{\theta}_a'\Gamma(\theta)\tilde{\theta}_a &= 0 \\
\sum_{n=1}^{\infty}\tilde{\lambda}_{aj}'\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_T D_T^{-1}\right)\tilde{\lambda}_{aj} &= 0 \\
\lim_{N,T\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{aT}'\Omega_T(\theta)^{-1}\tilde{F}_{aT}\lambda_j &= 0.
\end{aligned}$$

Because the matrices $\Gamma(\theta)$ and $\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_T D_T^{-1}$ are positive definite, $\tilde{\theta}_a$ and $\tilde{F}_a$ must be zero vectors in $\mathbb{R}^L$ and ℓ_2^p, respectively. Furthermore, Corollary A.1 implies that $\tilde{F}_a = 0$. Thus, the requirement (v) is also satisfied.

Next, we turn to the proof of completeness of $(H_\dagger, \|\cdot\|_\dagger)$. The proof of this part consists of two parts. We first find another norm on $H_\dagger$ under which $H_\dagger$ is complete. Second, we show that that norm is equivalent to $\|\cdot\|_\dagger$, which, by the usual arguments in functional analysis, implies the completeness of $H_\dagger$ under $\|\cdot\|_\dagger$.

Define a norm $\|\cdot\|_{\tilde{\dagger}}$ on $H_\dagger$ by:

$$\|h\|_{\tilde{\dagger}}^2 := \|\tilde{\theta}\|_E^2 + \sum_{n=1}^{\infty}\|\tilde{\lambda}_j\|_E^2 + \sum_{t=1}^{\infty}\|\tilde{f}_t\|_E^2 \quad \forall h = (\tilde{\theta}, \tilde{\Lambda}, \tilde{F}) \in H_\dagger.$$

It is a well-known fact from functional analysis that the product space of complete normed spaces is again complete. Because $\mathbb{R}^L$ and ℓ_2^p are complete, it can be easily seen that $H_\dagger$ is also complete under the norm $\|\cdot\|_{\tilde{\dagger}}$.

Our next goal is to prove that $\|\cdot\|_\dagger$ and $\|\cdot\|_{\tilde{\dagger}}$ are equivalent norms. First notice that:

$$\lambda_{\min}(\Gamma(\theta))\|\tilde{\theta}\|_E^2 \leq \tilde{\theta}'\Gamma(\theta)\tilde{\theta} \leq \lambda_{\max}(\Gamma(\theta))\|\tilde{\theta}\|_E^2,$$

and that:

$$\begin{aligned}
&\lambda_{\min}\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_TD_T^{-1}\right)\sum_{n=1}^{\infty}\|\tilde{\lambda}_j\|_E^2 \\
&\leq \sum_{n=1}^{\infty}\tilde{\lambda}_j'\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_TD_T^{-1}\right)\tilde{\lambda}_j \\
&\leq \lambda_{\max}(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_TD_T^{-1})\sum_{n=1}^{\infty}\|\tilde{\lambda}_j\|_E^2.
\end{aligned}$$

Let c and C be as in Lemma A.9. Set:

$$m := \min\left\{c, \lambda_{\min}\left(\Gamma(\theta)\right), \lambda_{\min}\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_TD_T^{-1}\right)\right\}$$

and:

$$M := \max\left\{C, \lambda_{\max}\left(\Gamma(\theta)\right), \lambda_{\max}\left(\lim_{T\to\infty} D_T^{-1}F_T'\Omega(\theta)^{-1}F_TD_T^{-1}\right)\right\}.$$

Notice that m is strictly positive and that M is finite. It is easy to see that for every $h \in H_\dagger$, we have:

$$m\|h\|_{\tilde{\dagger}}^2 \leq \|h\|_\dagger^2 \leq M\|h\|_{\tilde{\dagger}}^2.$$

This shows the equivalence of $\|\cdot\|_\dagger$ and $\|\cdot\|_{\tilde{\dagger}}$ as norms, and hence, the proof is complete.

□

Proof of Theorem 4.2

We first show the iso-Gaussianity of $\Delta_{NT,h}^\dagger$.

Lemma A.10. *Suppose that Assumptions 1, 2, 3 and 4 hold. For $h_1, h_2, \ldots, h_d \in H_\dagger$ where d is finite, $(\Delta_{NT,h_1}^\dagger, \Delta_{NT,h_2}^\dagger, \ldots, \Delta_{NT,h_d}^\dagger)$ satisfies:*

$$(\Delta_{NT,h_1}^\dagger, \Delta_{NT,h_2}^\dagger, \ldots, \Delta_{NT,h_d}^\dagger)' \xrightarrow{d} N(0, (\langle h_l, h_m\rangle_\dagger)),$$

under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$.

Proof. To prove this lemma, we apply the argument in McNeney and Wellner (2000, p. 3). It can be easily seen that $\Delta_{NT,h}^\dagger$ is linear in h; i.e., for any positive integer d and any $a = (a_1, a_2, \cdots, a_d)' \in \mathbb{R}^d$, we have:

$$\Delta_{NT,\sum_{m=1}^d a_m h_m}^\dagger = \sum_{m=1}^{d} a_l \Delta_{NT,h_m}^\dagger.$$

Lemma A.12 implies that under $P_{NT,0}$, as $N, T \to \infty$:

$$\sum_{m=1}^{d} a_m \Delta^{\dagger}_{NT,h_m} \xrightarrow{d} N\left(0, \left\|\sum_{m=1}^{d} a_m h_m\right\|_{\dagger}^{2}\right).$$

Noting that $\langle \cdot, \cdot \rangle_{\dagger}$ defines an inner product, we see that:

$$\left\|\sum_{i=1}^{d} a_m h_m\right\|_{\dagger}^{2} = \left\langle \sum_{m=1}^{d} a_m h_m, \sum_{l=1}^{d} a_l h_l \right\rangle_{\dagger} = a'(\langle h_l, h_m \rangle)_{\dagger} a.$$

Because $a \in \mathbb{R}^d$ is arbitrary, the desired result follows from an application of the Cramér–Wold device. □

The theorem is an application of Theorem 4.1 and follows from Lemmas 4.1, 4.2 and A.10.

□

A.3 Proofs of the theoretical results in Section 4.4

A.3.1 Proof of Theorem 4.3

For simplicity of notation, we consider the case where $s = 1$. That is, we derive the efficiency bound for the factor f_1. The proof for f_s $(s \neq 1)$ is similar.

This theorem is an application of Theorem 4.2. In the situation of this theorem, $\kappa_{NT}(h) = J'_{NT} f_1 + J'_{NT} \tilde{f}_1/\sqrt{N}$, $\mathbf{r}_{NT} = \sqrt{N}$ and $\dot{\kappa}(h) = J' \tilde{f}_1$. The space of the parameter of interest is $B = \mathbb{R}^p$, so the adjoint space is $B^* = \mathbb{R}^p$. We now derive the adjoint map $\dot{\kappa}^*$. Note that $\dot{\kappa}^*$ is a map from $B^* = \mathbb{R}^p$ to $H_{\dagger}$ and satisfies:

$$\langle \dot{\kappa}^* b^*, h \rangle_{\dagger} = b^{*\prime} J' \tilde{f}_1,$$

for any $b^* \in \mathbb{R}^p$ and $h \in H_{\dagger}$. Let $\dot{\kappa}^* b^* = (\dot{\kappa}^*_1 b^*, \dot{\kappa}^*_2 b^*, \dot{\kappa}^*_3 b^*) \in H_{\dagger}$, where $\dot{\kappa}^*_1 b^* \in \mathbb{R}^L$ and $\dot{\kappa}^*_2 b^* \in l^2_p$ and $\dot{\kappa}^*_3 b^* \in l^2_p$. Set $\dot{\kappa}^*_1 b^* = 0$, $\dot{\kappa}^*_2 b^* = 0$ and:

$$\dot{\kappa}^*_3 b^* = \left(\gamma(0)\Sigma_{\lambda\lambda}^{-1} J b^*, \gamma(1)\Sigma_{\lambda\lambda}^{-1} J b^*, \gamma(2)\Sigma_{\lambda\lambda}^{-1} J b^*, \cdots\right).$$

Notice that the first T components of $\dot{\kappa}^*_3 b^*$ can be written as:

$$\left(\gamma(0)\Sigma_{\lambda\lambda}^{-1} J b^*, \cdots, \gamma(T-1)\Sigma_{\lambda\lambda}^{-1} J b^*\right) = \Sigma_{\lambda\lambda}^{-1} J \left(b^*, 0, \cdots, 0\right) \Omega_T(\theta).$$

By Lemma A.8, we have:

$$\begin{aligned}
\langle \dot{\kappa}^* b^*, h \rangle &= \operatorname{tr}\left[\Sigma_{\lambda\lambda} \lim_{T\to\infty} \Sigma_{\lambda\lambda}^{-1} J \left(\gamma(0) b^*, \cdots, \gamma(T-1) b^*\right) \Omega_T(\theta)^{-1} \tilde{F}_T\right] \\
&= \operatorname{tr}\left[J \lim_{T\to\infty} \left(b^*, 0, \cdots, 0\right) \Omega_T(\theta) \Omega_T(\theta)^{-1} \tilde{F}_T\right] \\
&= \operatorname{tr}[J b^* \tilde{f}'_1] = b^{*\prime} J' \tilde{f}_1.
\end{aligned}$$

This shows that $\dot{\kappa}^*$ is the appropriate adjoint map. Computing $\|\dot{\kappa}^* b^*\|_{\dagger}^2$ gives the efficiency bound, and the proof is complete. □

A.3.2 Proof of Theorem 4.4

This theorem is an application of Theorem 4.2. In the situation of this theorem, $\kappa_{NT}(h) = J^*_{NT}\lambda_n + D_T^{-1}J^*_{NT}\tilde{\lambda}_n$, $\mathbf{r}_{NT} = D_T$ and $\dot{\kappa}(h) = J^*\tilde{\lambda}_n$. The space of the parameter of interest is $B = \mathbb{R}^p$, so the adjoint space is $B^* = \mathbb{R}^p$. We now derive the adjoint map $\dot{\kappa}^*$. Note that $\dot{\kappa}^*$ is a map from $B^* = \mathbb{R}^p$ to $\bar{H}$, where $\bar{H}$ is a completion of $H_\dagger$ and satisfies:

$$\langle \dot{\kappa}^* b^*, h\rangle_\dagger = b^{*\prime} J^* \tilde{\lambda}_n,$$

for any $b^* \in \mathbb{R}^p$ and $h \in H_\dagger$. Let $\dot{\kappa}^* b^* = (\dot{\kappa}_1^* b^*, \dot{\kappa}_2^* b^*, \dot{\kappa}_3^* b^*)$, where $\dot{\kappa}_1^* b^* \in \mathbb{R}^L$ and $\dot{\kappa}_2^* b^* \in l_p^2$ and $\dot{\kappa}_3^* b^* \in l_p^2$. It is easy to see that setting $\dot{\kappa}_1^* b^* = 0$, $\dot{\kappa}_2^* b^* = (0, \ldots, 0, \dot{\kappa}_{2n}^* b^*, 0 \ldots)$, where:

$$\dot{\kappa}_{2n}^* b^* = \left(\lim_{T\to\infty} D_T^{-1} F_T' \Omega_T(\theta)^{-1} F_T D_T^{-1} \right)^{-1} (J^*)' b^*,$$

and $\dot{\kappa}_3^* b^* = 0$ gives the appropriate adjoint map. Computing $\|\dot{\kappa}^* b^*\|_\dagger^2$ gives the bound, and the proof is complete. □

A.3.3 Proof of Theorem 4.5

The proof merely combines the arguments in the proofs of Theorems 4.3 and 4.4, and thus is omitted. □

A.3.4 Proof of Theorem 4.6

This theorem is an application of Theorem 4.2. In the situation of this theorem, $\kappa_{NT}(h) = \beta(\theta + \tilde{\theta}/\sqrt{NT})$, $\mathbf{r}_{NT} = \sqrt{NT}$ and $\dot{\kappa}(h) = \dot{\beta}(\theta)\tilde{\theta}$. The space of the parameter of interest is $B = \mathbb{R}^M$, so the adjoint space is $B^* = \mathbb{R}^M$. We now derive the adjoint map $\dot{\kappa}^*$. Note that $\dot{\kappa}^*$ is a map from $B^* = \mathbb{R}^M$ to $\bar{H}$, where $\bar{H}$ is a completion of $H_\dagger$ and satisfies:

$$\langle \dot{\kappa}^* b^*, h\rangle_\dagger = b^{*\prime} \dot{\beta}(\theta)\tilde{\theta},$$

for any $b^* \in \mathbb{R}^M$ and $h \in H_\dagger$. Let $\dot{\kappa}^* b^* = (\dot{\kappa}_1^* b^*, \dot{\kappa}_2^* b^*)$, where $\dot{\kappa}_1^* b^* \in \mathbb{R}^L$ and $\dot{\kappa}_2^* b^*$ is in the space for $(\tilde{\lambda}, \tilde{f})$. It is easy to see that setting $\dot{\kappa}_1^* b^* = \Gamma(\theta)^{-1}\dot{\beta}(\theta)' b^*$ and $\dot{\kappa}_2^* b^* = 0$ gives the appropriate adjoint map because:

$$\langle \dot{\kappa}^* b^*, h\rangle_\dagger = (\dot{\kappa}^* b^*)' \Gamma(\theta)\tilde{\theta} = b^{*\prime}\dot{\beta}(\theta)\Gamma(\theta)^{-1}\Gamma(\theta)\tilde{\theta} = b^{*\prime}\dot{\beta}(\theta)\tilde{\theta}.$$

Therefore, it holds that G in the current setting satisfies:

$$b^{*\prime} G \sim N(0, \|\dot{\kappa}^* b^*\|_\dagger^2) = N(0, b^{*\prime}\dot{\beta}(\theta)\Gamma(\theta)^{-1}\dot{\beta}(\theta)' b^*).$$

Thus, the proof is complete. □

A.4 Proofs of the theoretical results in Section 5.2

A.4.1 Proof of Lemma 5.1

We first show the second equality in (5.5). We observe that:

$$\begin{aligned} & \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}(\tilde{F}_{lT}\lambda_n + F_T\tilde{\lambda}_{ln})'\Omega_0^{-1}(\tilde{F}_{mT}\lambda_n + F_T\tilde{\lambda}_{mn}) \\ = & \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n + \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} \\ & \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} + \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}F_T\tilde{\lambda}_{mn}. \end{aligned}$$

We have:

$$\begin{aligned} \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n &= \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N} tr\left(\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n\right) \\ &= \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N} tr\left(\lambda_n\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\right) \\ &= tr\left(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n\lambda_n' \lim_{T\to\infty}\frac{1}{T}\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\right). \end{aligned}$$

Similarly, we have:

$$\begin{aligned} \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} &= tr\left(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\tilde{\lambda}_{mn}\lambda_n' \lim_{T\to\infty}\frac{1}{T}\tilde{F}_{lT}'\Omega_0^{-1}F_T\right), \\ \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n &= tr\left(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\lambda_n\tilde{\lambda}_{ln}' \lim_{T\to\infty}\frac{1}{T}F_T'\Omega_0^{-1}\tilde{F}_{mT}\right), \end{aligned}$$

and:

$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N}\tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} = tr\left(\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\tilde{\lambda}_{mn}\tilde{\lambda}_{ln}' \lim_{T\to\infty}\frac{1}{T}F_T'\Omega_0^{-1}F_T\right).$$

Assumption 5 gives $\lim_{N\to\infty}\sum_{n=1}^{N}\lambda_n\lambda_n'/N = \int \lambda\lambda' d\Phi(\lambda)$, $\lim_{N\to\infty}\sum_{n=1}^{N}\tilde{\lambda}_{mn}\lambda_n'/N = \int \tilde{\phi}_m(\lambda)\lambda' d\Phi(\lambda)$, $\lim_{N\to\infty}\sum_{n=1}^{N}\lambda_n\tilde{\lambda}_{ln}'/N = \int \lambda\tilde{\phi}_l(\lambda)' d\Phi(\lambda)$ and $\lim_{N\to\infty}\sum_{n=1}^{N}\tilde{\lambda}_{mn}\tilde{\lambda}_{ln}'/N = \int \tilde{\phi}_m(\lambda)\tilde{\phi}_l(\lambda)' d\Phi(\lambda)$. Let $\check{F}_T = (F_T, \tilde{F}_{lT}, \tilde{F}_{mT})$ and $\check{\psi}^{(m)}$ be such that $\check{\psi}^{(m)}(t/T)$ is the (t,m)-th element of $\check{F}_T$. We now show that $\check{F}_T'\Omega_0^{-1}\check{F}_T/T$ converges using Theorem 10.2.7 of Anderson (1971). For the moment, we assume that $\check{\psi}^{(m)} \neq 0$ for all m. Note that $\check{\psi}^{i}(t/T)$ corresponds to z_{it} in Anderson (1971). Assumption 10.2.1 of Anderson (1971) is satisfied because $\sum_{t=1}^{T}\check{\psi}^{(i)}(t/T)^2/T \to \int_0^1 \check{\psi}^{(i)}(a)^2 da > 0$, which implies that $\sum_{t=1}^{T}\check{\psi}^{(i)}(t/T)^2 \to \infty$. Assumption 10.2.2 of Anderson (1971) is satisfied because $\breve{\psi}^{(i)}(1)^2$ is bounded. Assumption 10.2.3 of Anderson (1971) is also satisfied because:

$$\frac{\sum_{t=1}^T \breve{\psi}^{(i)}(t/T)\breve{\psi}^{(n)}((t+h)/T)}{\sqrt{\sum_{t=1}^T \breve{\psi}^{(i)}(t/T)^2 \sum_{t=1}^T \breve{\psi}^{(n)}(t/T)^2}} = \frac{\sum_{t=1}^T \breve{\psi}^{(i)}(t/T)\breve{\psi}^{(n)}((t+h)/T)/T}{\sqrt{(\sum_{t=1}^T \breve{\psi}^{(i)}(t/T)^2/T)(\sum_{t=1}^T \breve{\psi}^{(n)}(t/T)^2/T)}}$$
$$\to \frac{\int_0^1 \breve{\psi}^{(i)}(a)\breve{\psi}^{(n)}(a)da}{\sqrt{(\int_0^1 \breve{\psi}^{(i)}(a)^2 ds)(\int_0^1 \breve{\psi}^{(n)}(a)^2 ds)}},$$

by Assumption 5 and $\tilde{g} \in C_f$. Lastly, Assumption 2 guarantees that Assumptions 10.2.5 and 10.2.6 of Anderson (1971) are satisfied. Thus, Theorem 10.2.7 of Anderson (1971) implies that:

$$\lim_{T\to\infty} \frac{1}{T}\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT} = \frac{1}{2\pi g_\theta(0)}\int_0^1 \tilde{\psi}_l(a)\tilde{\psi}_m(a)'da,$$
$$\lim_{T\to\infty} \frac{1}{T}\tilde{F}_{lT}'\Omega_0^{-1}F_T = \frac{1}{2\pi g_\theta(0)}\int_0^1 \tilde{\psi}_l(a)\psi(a)'da,$$
$$\lim_{T\to\infty} \frac{1}{T}F_T'\Omega_0^{-1}\tilde{F}_{mT} = \frac{1}{2\pi g_\theta(0)}\int_0^1 \psi(a)\tilde{\psi}_m(a)'da,$$

and:

$$\lim_{T\to\infty} \frac{1}{T}F_T'\Omega_0^{-1}F_T/T = \frac{1}{2\pi g_\theta(0)}\int_0^1 \psi(a)\psi(a)'da.$$

We note that it is easy to see that even if $\breve{\psi}^{(m)} = 0$ for some m, the above convergences hold. Thus, we obtain:

$$\begin{aligned}
\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^N \lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n &= tr\left(\int \lambda\lambda' d\Phi(\lambda)\frac{1}{2\pi g_\theta(0)}\int_0^1 \tilde{\psi}_l(a)\tilde{\psi}_m(a)'da\right) \\
&= \frac{1}{2\pi g_\theta(0)} tr\left(\int\int_0^1 \lambda\lambda'\tilde{\psi}_l(a)\tilde{\psi}_m(a)'da d\Phi(\lambda)\right) \\
&= \frac{1}{2\pi g_\theta(0)}\int\int_0^1 \lambda'\tilde{\psi}_l(a)\tilde{\psi}_m(a)'\lambda da d\Phi(\lambda).
\end{aligned}$$

Similarly, it follows that:

$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^N \lambda_n'\tilde{F}_{lT}'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} = \frac{1}{2\pi g_\theta(0)}\int\int_0^1 \lambda'\tilde{\psi}_l(a)\psi(a)'\tilde{\phi}_m(\lambda)da d\Phi(\lambda),$$
$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^N \tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}\tilde{F}_{mT}\lambda_n = \frac{1}{2\pi g_\theta(0)}\int\int_0^1 \tilde{\phi}_l(\lambda)'\psi(a)\tilde{\psi}_m(a)'\lambda da d\Phi(\lambda),$$

and:

$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^N \tilde{\lambda}_{ln}'F_T'\Omega_0^{-1}F_T\tilde{\lambda}_{mn} = \frac{1}{2\pi g_\theta(0)}\int\int_0^1 \tilde{\phi}_l(\lambda)'\psi(a)\psi(a)'\tilde{\phi}_m(\lambda)da d\Phi(\lambda).$$

Summing up, we have:

$$\begin{aligned}
&\lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^N (\tilde{F}_{lT}\lambda_n + F_T\tilde{\lambda}_{ln})'\Omega_0^{-1}(\tilde{F}_{mT}\lambda_n + F_T\tilde{\lambda}_{mn}) \\
= &\frac{1}{2\pi g_\theta(0)}\int\int_0^1 \left(\lambda'\tilde{\psi}_l(a) + \tilde{\phi}_m(\lambda)'\psi(a)\right)\left(\lambda'\tilde{\psi}_m(a) + \tilde{\phi}_m(\lambda)'\psi(a)\right) da d\Phi(\lambda).
\end{aligned}$$

Thus, the second equality in (5.5) holds.

Now, we show that $H_\ddagger$ is a linear subspace of a Hilbert space. It is easy to see that $H_\ddagger$ is a linear space. Therefore, we only show that $\langle \cdot, \cdot \rangle_\ddagger$ satisfies the requirements of an inner product. Note that there are five requirements: for any $h_l, h_m, h_n \in H$ (i) $\langle h_l, h_m \rangle = \langle h_m, h_l \rangle$; (ii) $\langle h_l + h_m, h_n \rangle = \langle h_l, h_n \rangle + \langle h_m, h_n \rangle$; (iii) $\langle \alpha h_l, h_m \rangle = \alpha \langle h_l, h_m \rangle$, where α is a scalar; (iv) $\langle h_l, h_l \rangle \geq 0$; (v) $\langle h_l, h_l \rangle = 0$ if and only if $h_l = 0$. It is, again, easy to see that requirements (i) to (iv) are satisfied. We verify that requirement (v) is also satisfied.

Because $\Gamma(\theta)$ is positive definite, $\tilde{\theta}'\Gamma(\theta)\tilde{\theta} = 0$ if and only if $\tilde{\theta} = 0$. It is easy to see that if $\tilde{\psi} = 0$ and $\tilde{\phi} = 0$, then $\int \int_0^1 \left(\lambda'\tilde{\psi}(a) + \tilde{\phi}(\lambda)'\psi(a)\right)^2 da d\Phi(\lambda) = 0$. On the other hand, if $\int \int_0^1 \left(\lambda'\tilde{\psi}(a) + \tilde{\phi}(\lambda)'\psi(a)\right)^2 da d\Gamma(\lambda) = 0$, then we must have:

$$\lambda'\tilde{\psi}(a) + \tilde{\phi}(\lambda)'\psi(a) = 0,$$

almost surely. Because $\int \lambda\lambda' d\Phi(\lambda) > 0$ and $\int_0^1 \psi(a)\psi(a)' da > 0$, we have:

$$\tilde{\psi}(a) = -\left(\int \lambda\lambda' d\Phi(\lambda)\right)^{-1} \left(\int \lambda\tilde{\phi}(\lambda)' d\Phi(\lambda)\right) \psi(a),$$

and:

$$\tilde{\phi}(\lambda) = -\left(\int_0^1 \psi(a)\psi(a)' da\right)^{-1} \left(\int_0^1 \psi(a)\tilde{\psi}(a)' da\right) \lambda.$$

Let $A = \left(\int_0^1 \psi(s)\psi(s)' ds\right)^{-1} \left(\int_0^1 \psi(a)\tilde{\psi}(a)' da\right)$ so that $\tilde{\phi}(\lambda) = -A\lambda$. Then, we also have $\tilde{\psi}(a) = A'\psi(a)$. Because $\tilde{\phi} \in C_\lambda$, A must be diagonal. On the other hand, the diagonal elements of A must be zero because $\tilde{\psi} \in C_f$. Therefore, $A = 0$, which implies that $\tilde{\phi} = 0$ and $\tilde{\psi} = 0$.

Thus, $\langle \cdot, \cdot \rangle_\ddagger$ satisfies the requirements of an inner product, and the proof is complete. □

A.4.2 Proof of Lemma 5.2

The local log-likelihood ratio process of our panel data model with interactive effects is given by:

$$
\begin{aligned}
&\log \frac{dP_{NT,h}}{dP_{NT,0}} \\
= \; & \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N} (y_n - F_T\lambda_n)'\Omega_0^{-1}(y_n - F_T\lambda_n) \\
& -\frac{1}{2}\sum_{n=1}^{N}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{NT}}\right)\left(\lambda_n + \frac{\tilde{\lambda}_n}{\sqrt{NT}}\right)\right)' \Omega_{\tilde{\theta}}^{-1}\left(y_n - \left(F_T + \frac{\tilde{F}_T}{\sqrt{NT}}\right)\left(\lambda_n + \frac{\tilde{\lambda}_n}{\sqrt{NT}}\right)\right) \\
= \; & \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} \\
& +\frac{1}{2}\sum_{n=1}^{N} (y_n - F_T\lambda_n)'\Omega_0^{-1}(y_n - F_T\lambda_n) - \frac{1}{2}\sum_{n=1}^{N} (y_n - F_T\lambda_n)'\Omega_{\tilde{\theta}}^{-1}(y_n - F_T\lambda_n) \\
& +\frac{1}{\sqrt{NT}}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}(y_n - F_T\lambda_n) - \frac{1}{2NT}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{B}_n,
\end{aligned}
$$

where:

$$\tilde{B}_n = \tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n + \frac{\tilde{F}_T\tilde{\lambda}_n}{\sqrt{NT}}.$$

Noting that $y_n - \lambda_n' f = w_n$ under $P_{NT,0}$, we write the log-likelihood ratio under $P_{NT,0}$ as follows:

$$
\begin{aligned}
\log \frac{dP_{NT,h}}{dP_{NT,0}} \; = \; & \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} + \frac{1}{2}\sum_{n=1}^{N} w_n'\Omega_0^{-1}w_n - \frac{1}{2}\sum_{n=1}^{N} w_n'\Omega_{\tilde{\theta}}^{-1}w_n \\
& +\frac{1}{\sqrt{NT}}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}w_n - \frac{1}{2NT}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{B}_n.
\end{aligned}
$$

Note that Lemmas A.2 and A.3 provide the expansion for the first four terms in the log-likelihood ratio.

Lemma A.11. *Suppose that Assumptions 1, 2 and 5 hold. Under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$:*

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}w_n - \frac{1}{2NT}\sum_{n=1}^{N} \tilde{B}_n'\Omega_{\tilde{\theta}}^{-1}\tilde{B}_n \\
= \; & \frac{1}{\sqrt{NT}}\sum_{n=1}^{N} (\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}w_n - \frac{1}{2NT}\sum_{n=1}^{N} (\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n) + o_p(1).
\end{aligned}
$$

Proof. The proof is very similar to that of Lemma A.4 and thus is omitted. □

Lemma A.12. *Suppose that Assumptions 1, 2 and 5 hold. Let $\Delta^{\ddagger}_{NT,h}$ be defined as in (5.4). Under $P_{NT,0}$, as $N \to \infty$ and $T \to \infty$:*

$$\Delta^{\ddagger}_{NT,h} \xrightarrow{d} N\left(0, \tilde{\theta}'\Gamma(\theta)\tilde{\theta} + \lim_{N,T\to\infty} \frac{1}{NT}\sum_{n=1}^{N} (\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)'\Omega_0^{-1}(\tilde{F}_T\lambda_n + F_T\tilde{\lambda}_n)\right). \quad \text{(A.10)}$$

Proof. The proof is very similar to that of Lemma A.7 and thus is omitted. □

Lemma 5.2 follows from Lemmas A.2, A.3, A.11 and A.12. □

A.4.3 Proof of Theorem 5.1

This theorem is an application of Theorem 4.1 and follows from Lemmas 5.1 and 5.2. Note that iso-Gaussianity of $\Delta^{\ddagger}_{NT,h}$ can be shown by exactly the same argument of the proof of Lemma A.10.

□

A.4.4 Proof of Theorem 5.2

In this proof, we only consider a simpler model in which $y_{nt} = \lambda_n + w_{nt}$. The proof for the general case is similar but requires much more complicated notation.

It is a well-known fact from basic functional analysis that a linear functional on a normed linear space is continuous if and only if it is bounded on the closed unit ball of the space. Thus, to prove the discontinuity of $\dot{\kappa}$, it is sufficient to show that $\dot{\kappa}$ is *unbounded* on the closed unit ball of $H_{\ddagger}$.

It is an easy task to construct a sequence of functions $\phi_M \in C_b(\mathbb{R})$ such that (i) $\phi_M(\lambda_n) = 1$ for every $M \in \mathbb{N}$; and (ii) $\int_{\mathbb{R}} \phi_M(x)^2 d\Phi(x) \to 0$ as $M \to \infty$ with every integral being positive.[26] Using ϕ_M, we now define a sequence $\{h_M\}_{M\in\mathbb{N}}$ in $H_{\ddagger} = \mathbb{R}^L \times C_b(\mathbb{R})$ by setting $h_M := (0, \phi_M)$. Noting that the LAN norm for the model $y_{nt} = \lambda_n + w_{nt}$ is:

$$\|h\|^2 = \|(\theta,\phi)\|^2 = \tilde{\theta}'\Gamma(\theta)\tilde{\theta} + \left(\sum_{k=-\infty}^{\infty} \gamma_k(\theta)\right)^{-1} \int_{\mathbb{R}} \phi^2(x) d\Phi(x), \tag{A.11}$$

it follows that:

$$\frac{|\dot{\kappa}(h_M)|}{\|h_M\|} = \frac{|\phi_M(\lambda_n)|}{\|h_M\|} = \frac{\sum_{k=-\infty}^{\infty}\gamma_k(\theta)}{\int_{\mathbb{R}} \phi_M^2(x) d\Phi(x)} \longrightarrow \infty, \tag{A.12}$$

as $N \to \infty$. This implies that the functional $\dot{\kappa}$ cannot be bounded on the closed unit ball of $H_{\ddagger}$, so we conclude that $\dot{\kappa}$ is not continuous.

□

[26] Assume, for simplicity, that $\sum_{k=-\infty}^{\infty}\gamma_k(\theta) = 1$. For example, set:

$$\begin{aligned}\phi_M(x) \quad := \quad & \{M(x-\lambda_n)+1\}1_{[-(1/M)+(1/M^2),\, 0)}(x-\lambda_n) + \{-M(x-\lambda_n)+1\}1_{[0,\, (1/M)-(1/M^2))}(x-\lambda_n) \\ & + \frac{1}{M}1_{(-\infty,\, -(1/M)+(1/M^2))\cup[(1/M)-(1/M^2),\, \infty)}(x-\lambda_n).\end{aligned}$$

Then it can be easily checked that the sequence $\{\phi_N\}_{N\in\mathbb{N}}$, thus constructed, satisfies the conditions (i) and (ii).

B The asymptotic distribution of the PCA estimator

Bai (2003) derived the asymptotic distribution of $\sqrt{N}(\hat{f}_t - \hat{J}'_{PCA,NT} f_t)$, where $\hat{f}_t$ is the PCA estimator of f_t and $\hat{J}_{PCA,NT}$ is some invertible $p \times p$ matrix. However, $\hat{J}_{PCA,NT}$ in the formula of Bai (2003) is data dependent, and we cannot apply the convolution theorem to examine whether $\hat{f}_t$ is efficient.

In this Appendix, we show that there exists a nonrandom matrix $J_{PCA,NT}$ (that depends on N and T) such that $\sqrt{N}(\hat{f}_t - \hat{J}'_{PCA,NT} f_t) = \sqrt{N}(\hat{f}_t - J'_{PCA,NT} f_t) + o_p(1)$. Because $J_{PCA,NT} f_t$ is deterministic, we can apply the convolution theorem to see whether $\hat{f}_t$ is an efficient estimator of $J'_{PCA,NT} f_t$.

We use the following notation. Let Λ_N be the $N \times p$ matrix whose n-th row is λ'_n. Let V_{NT} be the $p \times p$ diagonal matrix whose r-th diagonal element is the r-th largest eigenvalue of $(\Lambda'_N \Lambda_N/N)^{1/2}(F'_T F_T/T)(\Lambda'_N \Lambda_N/N)^{1/2}$. We assume that all the eigenvalues in V_{NT} are different. Let Υ_{NT} be the eigenvalue matrix that corresponds to V_{NT}. By the definitions of V_{NT} and Υ_{NT}, we have:

$$\Upsilon_{NT} V_{NT} \Upsilon'_{NT} = (\Lambda'_N \Lambda_N/N)^{1/2}(F'_T F_T/T)(\Lambda'_N \Lambda_N/N)^{1/2}.$$

We show that by setting:

$$J_{PCA,NT} = \left(\frac{\Lambda'_N \Lambda_N}{N}\right)^{1/2} \Upsilon_{NT} V_{NT}^{-1/2},$$

we have:

$$\sqrt{N}(\hat{J}_{PCA,NT} - J_{PCA,NT}) = o_p(1).$$

Our argument is similar to that of Bai and Ng (2013, proof of (2)). Let $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$. We first observe that:

$$\frac{\hat{F}'_T F_T}{T} = \frac{(\hat{F}_T - F_T \hat{J}_{PCA,NT} + F_T \hat{J}_{PCA,NT})' F_T}{T} = \frac{\hat{J}'_{PCA,NT} F'_T F_T}{T} + O_p(\delta_{NT}^{-2}),$$

because Lemma B.3 of Bai (2003) shows that $(\hat{F}_T - F_T \hat{J}_{PCA,NT})' F_T/T = O_p(\delta_{NT}^{-2})$, where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. We also have:

$$\frac{\hat{F}'_T F_T \hat{J}_{PCA,NT}}{T} = \frac{\hat{F}'_T (F \hat{J}_{PCA,NT} - \hat{F}_T + \hat{F}_T)}{T} = I + O_p(\delta_{NT}^{-2}),$$

by using Lemma B.3 of Bai (2003) to show that $\hat{F}'_T (F \hat{J}_{PCA,NT} - \hat{F}_T)/T = O_p(\delta_{NT}^{-2})$. Thus, we have:

$$\frac{\hat{J}'_{PCA,NT} F'_T F_T \hat{J}_{PCA,NT}}{T} = I + O_p(\delta_{NT}^{-2}).$$

By the definitions of Υ_{NT} and V_{NT}, we have:

$$\frac{F_T' F_T}{T} = \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2} \Upsilon_{NT} V_{NT} \Upsilon_{NT}' \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2}. \tag{B.1}$$

Letting:

$$B = V_{NT}^{1/2} \Upsilon_{NT}' \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2},$$

we write:

$$\frac{F_T' F_T}{T} = B' B.$$

Let $H^* = B \hat{J}_{PCA,NT}$. Then, we have:

$$(H^*)' H^* = I + O_p(\delta_{NT}^{-2}).$$

We now show that $H^* = I + O_p(\delta_{NT}^{-2})$. Because $(H^*)' H^* = I + O_p(\delta_{NT}^{-2})$, the eigenvalues of H^* are either 1 or -1. Next, we show that the nondiagonal elements of H^* are $O_p(\delta_{NT}^{-2})$. By Bai (2003), we have:

$$(H^*)' = \hat{J}_{PCA,NT}' B' = \hat{V}_{NT}^{-1} \frac{\hat{F}_T' F_T}{T} \frac{\Lambda_N' \Lambda_N}{N} B',$$

where $\hat{V}_{NT}$ is the $p \times p$ diagonal matrix whose r-th diagonal element is the r-th largest eigenvalue of $YY'/(NT)$. By Bai (2003, Lemma B.3), $\hat{F}_T' F_T / T = \hat{J}_{PCA,NT}' F_T' F_T / T + O_p(\delta_{NT}^{-2})$. Using (B.1), we have:

$$\begin{aligned}
&(H^*)' \\
=&\hat{V}_{NT}^{-1} \hat{J}_{PCA,NT}' \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2} \Upsilon_{NT} V_{NT} \Upsilon_{NT}' \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2} \left(\frac{\Lambda_N' \Lambda_N}{N} \right) \\
&\times \left(\frac{\Lambda_N' \Lambda_N}{N} \right)^{-1/2} \Upsilon_{NT} V_{NT}^{1/2} + O_p(\delta_{NT}^{-2}) \\
=&\hat{V}_{NT}^{-1} (H^*)' V_{NT} + O_p(\delta_{NT}^{-2}).
\end{aligned}$$

Thus, we have:

$$V_{NT} H^* = H^* \hat{V}_{NT} + O_p(\delta_{NT}^{-2}).$$

It follows that H^* is the eigenvalue matrix of V_{NT} up to a negligible term. Because V_{NT} is diagonal and all the eigenvalues are assumed to be different, H^* is also a diagonal matrix (up to the order of δ_{NT}^{-2}). We assume, without loss of generality, that the eigenvalues of H^* are all 1. Therefore, it holds that $H^* = I + O_p(\delta_{NT}^{-2})$.

Therefore, we have:

$$\hat{J}_{PCA,NT} = B^{-1}H^* = B^{-1} + O_p(\delta_{NT}^{-2}) = \left(\frac{\Lambda_N'\Lambda_N}{N}\right)^{1/2} \Upsilon_{NT} V_{NT}^{-1/2} + O_p(\delta_{NT}^{-2}).$$

Setting $J_{PCA,NT} = B^{-1} = (\Lambda_N'\Lambda_N/N)^{1/2}\Upsilon_{NT}V_{NT}^{-1/2}$ and assuming that $\sqrt{N}/T \to 0$, we have:

$$\sqrt{N}(\hat{J}_{PCA,NT} - J_{PCA,NT}) = o_p(1).$$

References

[1] Abowd, J. M. & D. Card (1989). On the covariance structure of earnings and hours changes, *Econometrica* **57**(2), 411–445.

[2] Ahn, S. C. & A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**(3), 1203–1227.

[3] Alvarez, J. & M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**(4), 1121–1159.

[4] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc.

[5] Anderson, T. W. & C. Hsiao (1981). Estimation of dynamics models with error components, *Journal of the American Statistical Association* **76**(375), 598–606.

[6] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**(3), 817–858.

[7] Arellano, M. & S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**(2), 277–297.

[8] Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**(1), 135–171.

[9] Bai, J. (2009a). Panel data models with interactive fixed effects, *Econometrica* **77**(4), 1229–1279.

[10] Bai, J. (2009b). Likelihood approach to small T dynamic panel data models with interactive effects, mimeo.

[11] Bai, J. & K. Li (2012). Statistical analysis of factor models of high dimension, *Annals of Statistics* **40**(1), 436–465.

[12] Bai, J. & S. Ng (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1), 191–221.

[13] Bai, J. & S. Ng (2013). Principal components estimation and identification of the static factors, *Journal of Econometrics* **176**, 18–29.

[14] Bickel, P. J., C. A. J. Klassen, Y. Ritov & J. A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press.

[15] Box, G. & G. Jenkins (1970). *Time Series Analysis: Forecasting and Control*, Holden Day, Inc.

[16] Breitung, J. & J. Tenhofen (2011). GLS estimation of dynamic factor models, *Journal of the American Statistical Association* **106**(495), 1150–1166.

[17] Brockwell, P. J. & R. A. Davis (1991). *Time Series: Theory and Methods*, 2nd edition, Springer.

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

[19] Choi, I. (2011). Efficient estimation of factor models, *Econometric Theory* **28**, 274–308.

[20] Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series, *Advances in Applied Probability* **4**, 469–467.

[21] 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**(471), 830–840.

[22] Gagliardini, P. & C. Gourieroux (2010). Efficiency in large dynamic panel models with common factor, mimeo.

[23] Gray, R. M. (2006). *Toeplitz and Circulant Matrices: A Review*, Now Publishers.

[24] Grenander, U. & M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, Inc.

[25] Hahn, J. & G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* **70**(4), 1639–1657.

[26] Hahn, J. & H. Moon (2006). Reducing bias of MLE in a dynamic panel model, *Econometric Theory* **22**, 499–512.

[27] Hájek, J. (1970). A characterization of limiting distributions of regular estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **14**, 323–330.

[28] Han, C., P. C. B. Phillips & D. Sul (2014). X-differencing and dynamic panel model estimation, *Econometric Theory* **30**, 201–251.

[29] Hayakawa, K. (2009). A simple efficient instrumental variable estimator for panel AR(p). models when both N and T are large, *Econometric Theory* **25**, 873–890.

[30] Hirano, K. & J. R. Porter (2012). Impossibility results for nondifferentiable functionals, *Econometrica* **80**(4), 1769–1790.

[31] Holtz-Eakin, D., W. Newey & H. S. Rosen (1988). Estimating vector autoregressions with panel data, *Econometrica* **6**, 1371–1395.

[32] Kakizawa, Y. & M. Taniguchi (1994). Asymptotic efficiency of sample covariances in a Gaussian stationary process, *Journal of Time Series Analysis* **15**, 303–311.

[33] Kreyszig, E. (1989). *Introductory Functional Analysis with Applications*, John Wiley & Sons, Inc.

[34] Kumon, M. & S. Amari (1984). Estimation of a structural parameter in the presence of a large number of nuisance parameters, *Biometrika* **71**(3), 445–459

[35] Lee, Y. (2012). Bias in dynamic panel models under time series misspecification, *Journal of Econometrics* **169**(1), 54–60.

[36] Lee, Y. & P. C. B. Phillips (2013). Model selection in the presence of incidental parameters, mimeo.

[37] Maddox, I. J. (1970). *Elements of Functional Analysis*, Cambridge University Press.

[38] McNeney, B. & J. A. Wellner (2000). Application of convolution theorems in semiparametric models with non-i.i.d. data, *Journal of Statistical Planning and Inference* **91** 441–480.

[39] Moon, H. R. & M. Weidner (2010). Dynamic linear panel regression models with interactive fixed effects, mimeo.

[40] Okui, R. (2010). Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data, *Econometric Theory* **26** 1263–1304.

[41] Okui, R. (2011). Asymptotically unbiased estimation of autocovariances and autocorrelations for panel data with incidental trends, *Economics Letters* **112** 49–52.

[42] Okui, R. (2013). Asymptotically unbiased estimation of autocovariances and autocorrelations with panel data in the presence of individual and time effects, forthcoming in the *Journal of Time Series Econometrics*.

[43] Onatski, A. (2009). Testing hypothesis about the number of factors in large factor models, *Econometrica* **77**(5), 1447–1479.

[44] Parzen, E. (1957). Consistent estimates of the spectrum of a stationary time series, *Annals of Mathematical Statistics* **28**(2), 329–348.

[45] Pfanzagl, J. (1993). Incidental versus random nuisance parameters, *Annals of Statistics* **21**, 1663–1691.

[46] Phillips, P. C. B. & H. R. Moon (1999). Linear regression limit theory for nonstationary panel data, *Econometrica* **67**(5), 1057–1111.

[47] Porat, B. (1987). Some asymptotic properties of the sample covariances of Gaussian autoregressive moving average process, *Journal of Time Series Analysis* **8**, 205–220.

[48] Robinson, P. (1989). Nonparametric estimation of time-varying parameters, in *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.), Springer, 253–264.

[49] Sarafidis, V. & T. Yamagata (2010). Instrumental variable estimation of dynamic linear panel data models with defactored regressors under cross-sectional dependence, mimeo.

[50] Schott, R. (2006). *Matrix Analysis for Statistics*, John Wiley & Sons, Inc.

[51] Sprent, P (1966). A generalized least-squares approach to linear functional relationships, *Journal of the Royal Statistical Society, Series B* **28**(2), 278–297.

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

[53] Strasser, H. (1996). Asymptotic efficiency of estimates for model with incidental nuisance parameters, *Annals of Statistics* **24**(2), 879–901.

[54] Strasser, H. (1998). Perturbation invariant estimates and incidental nuisance parameters, *Mathematical Methods of Statistics* **7**, 1–26.

[55] Taniguchi, M. & Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer.

[56] van der Vaart, A. W. (1991). On differentiable functionals, *Annals of Statistics* **19**(1), 178–204.

[57] van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[58] van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer.

[59] Walker, A. M. (1995). On results of Porat concerning asymptotic efficiency of sample covariances of Gaussian ARMA processes, *Journal of Time Series Analysis* **16**(2), 237–248.