Efficient Estimation of Autocovariances for Panel Data with Individual Effects under Cross Section and Time Series Asymptotics*

Haruo Iwakura[†] and Ryo Okui[‡]

January 21, 2011

Abstract

This paper studies asymptotic efficiency of autocovariance estimation in panel data settings with individual effects when both the cross-sectional sample size and the length of time series tend to infinity. The efficiency bound for regular estimators of autocovariances is derived by using a Hajék (1970)-type convolution theorem. In view of the efficiency bound, we provide a necessary and sufficient condition under which bias-corrected autocovariance estimators developed by Okui (2010) are asymptotically efficient. In particular, we show that, when the individual dynamics follow an ARMA(p, q) process, the bias-corrected autocovariance estimator at lag k is asymptotically efficient if and only if $p \geq q$ and $0 \leq k \leq p - q$. These efficiency results are analogous to those for time series analysis obtained by Porat (1987) and Kakizawa and Taniguchi (1994).

Keywords: asymptotic efficiency; autocovariance; convolution theorem; double asymptotics; dynamic panel data model.

JEL classification: C13; C23.

1 Introduction

We are often interested in the dynamic nature of an economic variable. Panel data are useful to investigate the dynamics separately from spurious serial correlation caused by heterogeneity

*The authors would like to thank Hidehiko Ichimura and Masanobu Taniguchi for initially getting us interested in the problem. Okui acknowledges financial support from Japan Society of the Promotion of Science under KAKENHI 22730176. The authors are solely responsible for any errors.

†Graduate School of Economics, Kyoto University. Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan.

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

across individuals. For example, income is typically serially correlated since an individual who received a high income in a previous year tends to receive a high income this year too. It is interesting and important to understand the temporal dependence of the income process separately from heterogeneity in productivity. In time series analysis, a typical first step for analyzing the dynamics is to examine the autocovariances and autocorrelations. However, it is not trivial to do so in panel data settings because we need to take care of heterogeneity across individuals although some textbooks recommend such an analysis (see, e.g., Cameron and Trivedi (2005, chapter 21.3)). Okui (2010) studies the bias in the conventional autocovariance estimators and proposes asymptotically unbiased estimators of autocovariances for panel data with individual effects. This paper is concerned with asymptotic efficiency of autocovariance estimation using panel data in the presence of individual effects.

We consider the setting in which we observe an economic variable for many individuals and long time periods and the observed variable can be written as a sum of an individual effect and an idiosyncratic component. The idiosyncratic component is Gaussian stationary and its spectral density can be characterized by a finite number of parameters. We derive the lower bound of the asymptotic variances of any regular estimators of an autocovariance using double asymptotics under which both the cross-sectional sample size and the length of time series tend to infinity.

Another contribution is to 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 $0 \leq k \leq p - q$. In general, to obtain an asymptotically efficient estimator, we need to identify the correct data generating process and estimate the model parameters efficiently. The efficient autocovariance estimator can be constructed based on the efficiently estimated model parameters. However, the true data generating process is typically unknown a priori. It is not a trivial task to develop a model selection procedure to find the correct specification[1] and it is also difficult to examine the effect of model selection on the estimation. Furthermore, for panel data with individual effects, estimation procedure may become model specific.[2] Thus, if we know that Okui's autocovariance

[1]Lee (2010a) develops an information criterion to choose the lag of panel AR models and Lee (2010b) proposes asymptotically unbiased estimators for AR coefficients.

[2]There have been many different estimators proposed for panel ARMA models. For panel AR models, see Anderson and Hsiao (1981), Anderson and Hsiao (1981), Arellano and Bond (1991), Holtz-Eakin, Newey and Rosen (1988), Hahn and Kuersteiner (2002), Hahn and Moon (2006) and Lee (2008). For MA models, see Baltagi and Li (1994).

estimator is asymptotically efficient in some data generating process and if we think that the true data generating process can be described by this class of data generating process, then we can efficiently estimate autocovariances without going through a complicated and difficult process of model specification and model specific estimation.

The notion of efficiency used in this paper is that of the convolution theorem by Hajek (1970), which is extended to cases with infinite dimensional parameters by van der Vaart and Wellner (1996). We note that the number of parameters tends to infinity in our case because there are as many individual effects as the cross-sectional sample size. This paper contributes to the literature of efficient estimation in the sense that we investigate the efficiency for general dynamic panel data models with individual effects under double asymptotics.

There are three papers that are closely related to the current discussion and on which the efficiency result presented in this paper is based. Davies (1973) derives the local asymptotic normality of Gaussian stationary time series models where the spectral density is determined by a finite number of parameters.[3] Showing the local asymptotic normality is an important step in obtaining the efficiency bound and we also prove the local asymptotic normality based on Davies' (1973). The difference from Davies' (1973) is that we consider panel data settings and use double asymptotics. Another difference is that we consider individual effects whose cardinality tends to infinity while Davies (1973) only considers a finite number of parameters.

Hahn and Kuersteiner (2002) derives the efficiency bound for panel AR(1) models with Gaussian errors and individual effects. Our mathematical derivation of the efficiency bound closely follows that of Hahn and Kuersteiner (2002). On the other hand, we obtain a more general result and the efficiency bound proved by Hahn and Kuersteiner (2002) can be obtained as a special case of our result.

Kakizawa and Taniguchi (1994) derives the lower bound of the variances of autocovariance estimator in time series setting.[4] They give a necessary and sufficient condition under which the sample autocovariance estimator is asymptotically efficient in Gaussian time series models without intercept. In particular, they show that the k-th order sample autocovariance is asymptotically efficient if the process follows an ARMA(p, q) model and $p \geq q$ and $0 \leq k \leq p - q$. We basically obtain the same result in panel data settings. Again the difference is that we consider panel data in the presence of individual effects and use double asymptotics while Kakizawa and Taniguchi (1994) consider only time series models.[5]

[3]See Section A.2 for the precise definition of the local asymptotic normality. van der Vaart (1998) is an excellent textbook for the discussion of the local asymptotic normality.

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

[5]Another technical difference is that Kakizawa and Taniguchi (1994) examines the limit of the Cramer-Rao

The reminder of this paper is organized as follows. The next section presents the setting and introduces Okui's (2010) estimator and its asymptotic distribution. Section 3 gives the efficiency bound as well as conditions under which Okui's estimator achieves the efficiency bound. Section 4 discusses several possible extensions and concludes the paper. All of mathematical proofs are given in the appendix.

2 Set-up

2.1 The Model

Suppose that we have available a panel data set $\{y_{jt}\}$ for $j = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$. We assume that y_{jt} is generated as a sum of an time-invariant unobserved individual effect η_j and a time-varying component w_{jt}. Put more precisely, we consider a dynamic panel data model of the form:

$$y_{jt} = \eta_j + w_{jt},$$

where w_{jt} is independently and identically distributed (i.i.d.) across individual j and follows a Gaussian stationary process over time t with mean zero. In this paper we regard individual effects η_j's as parameters. It is assumed that the autocovariance structure of $\{w_{jt}\}_{t\in\mathbb{Z}}$, where $\mathbb{Z}$ is the set of all integers, is completely characterized by a finite-dimensional parameter $\theta \in \Theta$ where Θ is some open subset of $\mathbb{R}^L$, and we denote the k-th order autocovariance by $\gamma_k(\theta)$ i.e. $\gamma_k(\theta) := \mathbb{E}_\theta[w_{jt}w_{j,t-k}]$, where $\mathbb{E}_\theta$ denotes the expectation under θ. Note that, since $\{w_{jt}\}_{t\in\mathbb{Z}}$ is a Gaussian stationary process, the parameter θ completely determines the law of the process $\{w_{jt}\}_{t\in\mathbb{Z}}$. 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_{jt}.

Assumption 1.

(i) w_{jt} is i.i.d. across individual j.

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

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

lower bound so 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 does allow estimators to be not exactly unbiased but asymptotically unbiased. Moreover, it is difficult to develop estimators that are unbiased in finite samples in our setting because of the presence of individual effects.

Assumption 1 (iii) guarantees the existence of the spectral density f_θ and it can be written as

$$f_\theta(\lambda) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \gamma_m(\theta) \exp(-im\lambda),$$

where $i := \sqrt{-1}$. Also note that, by using the spectral density f_θ, the k-th order autocovariance $\gamma_k(\theta)$ can be expressed as

$$\begin{aligned} \gamma_k(\theta) &= \int_{-\pi}^{\pi} \exp(-ik\lambda) f_\theta(\lambda) d\lambda \\ &= \int_{-\pi}^{\pi} \cos(k\lambda) f_\theta(\lambda) d\lambda, \end{aligned}$$

by the Fourier inversion. This expression for the k-th order autocovariance will be used later in Section 3.

Under the setting above, this paper examines the asymptotic efficiency for estimation of $\gamma_k(\theta)$.

2.2 The k-th order Within-Group Sample Autocovariance Estimator

A natural estimator for $\gamma_k(\theta)$ may be a within-group (WG) sample autocovariance estimator $\hat{\gamma}_k$ which is defined by

$$\hat{\gamma}_k := \frac{1}{N(T-k)} \sum_{j=1}^{N} \sum_{t=k+1}^{T} (y_{jt} - \bar{y}_j)(y_{jt-k} - \bar{y}_j),$$

where $\bar{y}_j := \frac{1}{T}\sum_{t=1}^{T} y_{jt}$. Note that $\hat{\gamma}_k$ is the sample average of individual sample autocovariances.

Okui (2010) shows that $\hat{\gamma}_k$ is consistent for $\gamma_k(\theta)$ as $N, T \to \infty$ but has bias of order $O(1/T)$, which can be severely large when T is small relative to N. Indeed, Okui's (2010) result implies that under the Gaussianity, as $N, T \to \infty$ with $N/T^3 \to 0$,

$$\sqrt{NT}\left(\hat{\gamma}_k - \gamma_k(\theta) + \frac{1}{T}V_T\right) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty} \{\gamma_j(\theta)^2 + \gamma_{k+j}(\theta)\gamma_{k-j}(\theta)\}\right), \tag{2.1}$$

where

$$V_T := \gamma_0(\theta) + 2\sum_{k=1}^{T-1} \frac{T-k}{T} \gamma_k(\theta).$$

Here $(1/T)V_T$ is the leading term of the bias and it is easily seen that V_T converges to the long-run variance of $\{w_{jt}\}_{t\in\mathbb{Z}}$, $V = \sum_{k=-\infty}^{\infty} \gamma_k(\theta)$, as $T \to \infty$. Okui (2010) proposes to estimate V_T to alleviate the bias of $\hat{\gamma}_k$. Let $\hat{V}_T$ be an estimator of V_T and r_{NT} be the inverse of the rate of convergence of V_T such that

$$\hat{V}_T - V_T = O_p(r_{NT}) \quad \text{with } r_{NT}\sqrt{N/T} \to 0, \tag{2.2}$$

as $N, T \to \infty$. A bias-corrected WG sample autocovariance estimator, denoted $\tilde{\gamma}_k$, is obtained by simply adding $\hat{V}/T$ to $\hat{\gamma}_k$:

$$\tilde{\gamma}_k := \hat{\gamma}_k + \frac{1}{T}\hat{V}_T.$$

From (2.1) and (2.2), we obtain the following result.

Theorem 1 (Okui (2010, Theorem 3)). *Suppose that Assumption 1 is satisfied. Then, we have*

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

as $N, T \to \infty$ *with* $N/T^3 \to 0$.

Remark 2.1. Okui (2010, Section 4) constructs an estimator $\hat{V}_T$ satisfying the condition (2.2) by estimating the long-run variance of $\{w_{jt}\}_{t\in\mathbb{Z}}$ by using a kernel method developed by Parzen (1957) and Andrews (1991). The simulation study conducted in Okui (2010) shows that Okui's method effectively reduces the bias in small samples. However, since the interest of this paper is only in the asymptotic variance of $\tilde{\gamma}_k$, we will not discuss the estimation of $\hat{V}_T$ further.

Remark 2.2. The asymptotic variance of $\tilde{\gamma}_k$ presented in Theorem 1 has exactly the same form as that of its time series counterpart (see, e.g., Anderson (1971, Chapter 8)). Also note that, by Parseval's identity, it can be written as

$$\sum_{j=-\infty}^{\infty}\{\gamma_j(\theta)^2 + \gamma_{k+j}(\theta)\gamma_{k-j}(\theta)\} = 4\pi\int_{-\pi}^{\pi} f_\theta^2(\lambda)\cos^2(k\lambda)d\lambda.$$

This expression for the asymptotic variance will be used later in Section 3.2 when deriving a condition under which $\tilde{\gamma}_k$ is asymptotically efficient.

Remark 2.3. Let $\mathrm{cum}(t_1, \cdots, t_p)$ denote the p-th joint cumulant of $(w_{jt_1}, \cdots, w_{jt_p})'$. If we assume that (i)$\sum_{t_2,\cdots,t_p=-\infty}^{\infty}|\mathrm{cum}(0, t_3, \cdots, t_p)| < \infty$ for any $p \leq 8$ and (ii) there exists $M > 0$ such that $\mathbb{E}_\theta|w_{jt}w_{jk}w_{jm}w_{jl}| < M$ for any t, k, m and l , then the Gaussianity assumption on w_{jt} is not needed for the asymptotic normality of $\tilde{\gamma}_k(\theta)$. Indeed, Okui (2010) shows without imposing the Gaussianity assumption on $\{w_{jt}\}_{t\in\mathbb{Z}}$ that

$$\sqrt{NT}(\tilde{\gamma}_k - \gamma_k(\theta)) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty}\{\gamma_j(\theta)^2 + \gamma_{k+j}(\theta)\gamma_{k-j}(\theta) + \mathrm{cum}(0, -k, j, j-k)\}\right), \quad (2.3)$$

as $N, T \to \infty$ with $N/T^3 \to 0$. Because the joint cumulants of order greater than 2 are always zero for multivariate Gaussian random vectors, the asymptotic variance in (2.3) reduces to the one in Theorem 1 when $\{w_{jt}\}_{t\in\mathbb{Z}}$ is a Gaussian stationary process.

Remark 2.4. Okui (2010, Remark 2) notes that the order condition $N/T^3 \to 0$ is required only for ignoring the bias term of order $O(1/T^2)$. He also states that the condition $N/T^3 \to 0$ can be relaxed if we take into account the bias term of order $O(1/T^2)$. However he does not take such a route because it would make the form of the asymptotic bias more complicated. For the same reason, we also keep the condition $N/T^3 \to 0$.

One of the main purposes of the paper is to examine the optimality of the asymptotic distribution of the bias-corrected WG sample autocovariance estimator $\tilde{\gamma}_k$.

3 The Main Results

In this section, we first present the efficiency bound for estimation of the k-th order autocovariance $\gamma_k(\theta)$ and then provide a condition under which a bias-corrected WG autocovariance estimator achieves the efficiency bound. Lastly, we provide a brief description of the proofs of the main results.

3.1 The Efficiency Bound

This subsection gives the lower bound for asymptotic variances of any regular estimators of $\gamma_k(\theta)$.[6] To this end, the following assumptions are needed.

The first assumption is concerned with the individual effects η_j's:

Assumption 2.

(i) $\lim_{N\to\infty}(1/N)\sum_{j=1}^{N}\eta_j^2$ exists and is finite.

(ii) $(\max_{j\leq N}\eta_j^2)/N = o(1)$ as $N \to \infty$.

Next we impose some restrictions on spectral density f_θ:

Assumption 3.

(i) $\theta \mapsto f_\theta(\lambda)$ is differentiable at any point $\theta \in \Theta$.

(ii)

$$\lim_{\epsilon\to 0}\sup_{\lambda}|f(\theta+\epsilon) - f(\theta)| = 0, \quad \forall\theta \in \Theta.$$

[6]Intuitively speaking, a sequence of estimators is regular if a disappearing small change of parameters should not change its limit distribution at all (van der Vaart (1998, p115)). For a precise definition, see Appendix A.2. The regularity requirement is a desirable property for reasonable estimators to have and not so restrictive. For a detailed study of the regularity condition, see e.g. Bickel et al. (1993, Chapter 2).

(iii)

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

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

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

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

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

Finally, we impose the following restriction on the covariance matrix of $w_j := (w_{j1}, w_{j2}, \cdots, w_{jT})'$:

Assumption 4. $\lim_{T\to\infty} (1/T)\mathbf{1}_T'\Omega(\theta)^{-1}\mathbf{1}_T$ exists and is finite for every $\theta \in \Theta$, where $\Omega_T(\theta) := \mathbb{E}_\theta[w_j w_j']$ and $\mathbf{1}_T$ denotes a T-dimensional vector whose components are all 1's.[7]

Remark 3.1. Assumption 2 is slightly stronger than the assumption imposed on individual effects in Hahn and Kuersteiner (2002). They only assume that $(1/N)\sum_{j=1}^{N} \eta_j^2 = O(1)$. The condition (ii) is used to show the asymptotic normality of the score function of our dynamic panel data model with individual effects. For details, see the proof of Lemma 6 in Appendix A.3.

Remark 3.2. Assumption 3 is similar to the assumptions imposed in Davies (1973, A 1.1 to A 1.4). 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 f_θ. The reason that we state conditions in terms of spectral density f_θ is that these conditions are easier to check by doing so. As an example, let us consider the case where $\{w_{jt}\}_{t\in\mathbb{Z}}$ follows a stationary ARMA(p, q) process:

$$w_{jt} = a_1 w_{j,t-1} + a_2 w_{j,t-2} + \cdots + a_p w_{j,t-p} + u_{jt} + b_1 u_{j,t-1} + \cdots + b_{j,t-q} u_{j,t-q},$$

where u_{jt} is i.i.d.$N(0, \sigma^2)$ across j 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 $\{w_{jt}\}_{t\in\mathbb{Z}}$ is given by

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} \frac{|b(e^{-i\lambda})|^2}{|a(e^{-i\lambda})|^2}, \tag{3.1}$$

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 Assumptions 3.

[7] The matrix $\Omega_T(\theta)$ is nonsingular by Assumption 3 (iv) (See e.g. Gray (2006, Theorem 5.2)). However, Assumption 3 does not guarantee the existence of $\lim_{T\to\infty}(1/T)\mathbf{1}_T'\Omega(\theta)^{-1}\mathbf{1}_T$.

Remark 3.3. When $\{w_{jt}\}_{t\in\mathbb{Z}}$ follows a Gaussian stationary AR(p) process as specified in Remark 3.2 (i.e. set $q = 0$), Assumption 4 is automatically satisfied. In fact, the arguments in section 5.1.2 in Amemiya (1985) give

$$\lim_{T\to\infty} \frac{1}{T}\mathbf{1}_T'\Omega(\theta)^{-1}\mathbf{1}_T = \frac{(1-a_1-\cdots-a_p)^2}{\sigma^2}.$$

Note that the limit coincides with the inverse of the long-run variance of the AR(p) process.

We are now ready to provide the efficiency bound for any regular estimators of the k-th order autocovariance $\gamma_k(\theta)$.

Theorem 2. *Suppose that Assumptions 1 to 4 are satisfied. Define*

$$\Gamma(\theta) := \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta'} f_\theta(\lambda)\frac{\partial}{\partial\theta} f_\theta(\lambda)\frac{d\lambda}{f_\theta^2(\lambda)}.$$

and assume that the matrix $\Gamma(\theta)$ is nonsingular. Also suppose that $\tau_{N,T}$ is any regular estimator of $\gamma_k(\theta)$ as $N, T \to \infty$. If the limit law of τ_{NT} has variance Σ_θ, then

$$\Sigma_\theta \geq \left\{\int_{-\pi}^{\pi} \cos(k\lambda)\frac{\partial}{\partial\theta'} f_\theta(\lambda)d\lambda\right\}\Gamma(\theta)^{-1}\left\{\int_{-\pi}^{\pi} \cos(k\lambda)\frac{\partial}{\partial\theta} f_\theta(\lambda)d\lambda\right\} \quad \forall\theta\in\Theta. \tag{3.2}$$

Proof. See Appendix A.3. □

Remark 3.4. A useful sufficient condition for $\Gamma(\theta)$ to be nonsingular is that for each $a \in \mathbb{R}^L$ and each $\theta \in \Theta$, there exists some non-null set of λ such that

$$a'\frac{\partial}{\partial\theta_m} f_\theta(\lambda) \neq 0$$

(this is a special case of Theorem 4.7 in Davies (1973)). If $\{w_{jt}\}_{j\in\mathbb{Z}}$ is a stationary ARMA(p, q) process as specified in Remark 3.2 and so has spectral density (3.1), then this sufficient condition is obviously satisfied.

This theorem shows that the right hand side of (3.2) is the efficiency bound for any regular estimators of $\gamma_k(\theta)$. In the time series literature, the matrix $\Gamma(\theta)$ is called a Gaussian Fisher information matrix associated with spectral density f_θ (see e.g. Taniguchi and Kakizawa (2000, p58)). This name is after the fact that the matrix $\Gamma(\theta)$ is the limit of small-sample Fisher information matrices for a Gaussian stationary process with spectral density f_θ.[8] We also note that the efficiency bound in Theorem 2 can be rewritten as

$$\left\{\frac{\partial}{\partial\theta'}\gamma_k(\theta)\right\}\Gamma(\theta)^{-1}\left\{\frac{\partial}{\partial\theta}\gamma_k(\theta)\right\},$$

[8] The closed form of the matrix $\Gamma(\theta)$ for ARMA models is available in, e.g., Box and Jenkins (1970).

because Assumptions 1(iii) and 3(i) imply that

$$\frac{\partial}{\partial \theta}\gamma_k(\theta) = \int_{-\pi}^{\pi} \cos(k\lambda)\frac{\partial}{\partial \theta} f_\theta(\lambda) d\lambda.$$

This expression demonstrates that the efficiency bound has the same form as the limit of Crámer-Rao lower bounds for estimation of $\gamma_k(\theta)$ in the time series setting.

Kakizawa and Taniguchi (1994) study the asymptotic efficiency of sample autocovariances for a Gaussian stationary process $\{X_t\}_{t\in\mathbb{Z}}$ with mean zero and spectral density f_θ. They calculate the lower bound for asymptotic variances of any unbiased estimators of the k-th order autocovariance $\mathbb{E}_\theta[X_t X_{t-k}]$, which is the limit of small-sample Crámer-Rao lower bounds (here note that the k-th order sample autocovariance $(1/(T-k))\sum_{t=k+1}^{T} X_t X_{t-k}$ is unbiased for $\mathbb{E}_\theta[X_t X_{t-k}]$). The lower bound they derive has exactly the same form as the lower bound given in Theorem 2.

An important implication of this finding is that the presence of individual effects does not affect the form of the efficiency bound. This result is interesting in the sense that the presence of an infinite dimensional parameter (note that we treat individual effects as parameters whose cardinality tends to infinity as the sample size goes to infinity) does not affect the efficiency bound.

3.2 A Condition for $\tilde{\gamma}_k$ to be Asymptotically Efficient

We now provide a conditions under which bias-corrected WG sample autocovariance estimators $\tilde{\gamma}_k$ are asymptotically efficient. We say that a sequence of estimators is asymptotically efficient if it is regular and its asymptotic variance achieves the efficiency bound given by a convolution theorem.

Theorem 3. *Suppose that Assumptions 1 to 4 are satisfied. Then a bias-corrected WG sample autocovariance estimator $\tilde{\gamma}_k(\theta)$ is asymptotically efficient if and only if there exists $c \in \mathbb{R}^L$ such that*

$$f_\theta^2(\lambda)\cos(k\lambda) + c' \frac{\partial}{\partial \theta} f_\theta(\lambda) = 0, \quad \forall \lambda. \tag{3.3}$$

In particular, if $\{w_{jt}\}_{t\in\mathbb{Z}}$ is a Gaussian stationary ARMA(p,q) process as specified in Remark 3, then a bias-corrected WG sample autocovariance estimator $\tilde{\gamma}_k(\theta)$ is asymptotically efficient if and only if

$$p \geq q \quad \text{and} \quad 0 \leq k \leq p - q. \tag{3.4}$$

The condition (3.3) for asymptotic efficiency of $\tilde{\gamma}_k$ is the same as that for time series analysis obtained by Kakizawa and Taniguchi (1994). They comment that the condition (3.3) is easy to check. They also show that, for the case of a Gaussian stationary ARMA(p, q) process, the condition (3.3) reduces to (3.4), which is a condition first derived by Porat (1987). The condition (3.4) implies that, if $\{w_{jt}\}_{t \in \mathbb{Z}}$ is a stationary AR(p) process and $k \leq p$, we can efficiently estimate $\gamma_k(\theta)$ by using $\tilde{\gamma}_k$. Porat (1987) states that this is not so 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 $\{w_{jt}\}_{t \in \mathbb{Z}}$ is a Gaussian MA(q) process, none of the bias-corrected WG sample autocovariance estimators are asymptotically efficient. As an intermediate case, if $\{w_{jt}\}_{t \in \mathbb{Z}}$ 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 non-trivial and interesting as argued by Porat (1987).

3.3 A Brief Sketch of the Proofs of the Theorems

The efficiency bound given in Theorem 2 is derived by using a Hajék-type convolution theorem.[9] We give a brief explanation of a convolution theorem along with a short description of the steps for deriving the efficiency bound for $\gamma_k(\theta)$.

A convolution theorem is a very powerful tool for providing the efficiency bound for any regular estimators of parameters of interest. More precisely, a convolution theorem states that the asymptotic distribution of any regular estimator can be expressed as a convolution of two probability distributions: one is a normal distribution, which has mean zero and is common for all regular estimators, and the other is some 'noise' distribution, which varies with estimators. Since a convolution is the distribution of the sum of two independent random variables, a convolution theorem implies that the asymptotic variance of any regular estimator is no smaller than the variance of the normal distribution in the convolution representation. In our case, Theorem 7 in the Appendix yields

$$\sqrt{NT}(\tau_{NT} - \gamma_k(\theta)) \xrightarrow{d} N\left(0, \left\{\int_{-\pi}^{\pi} \cos(k\lambda) \frac{\partial}{\partial \theta'} f_\theta(\lambda) d\lambda\right\} \Gamma(\theta)^{-1} \left\{\int_{-\pi}^{\pi} \cos(k\lambda) \frac{\partial}{\partial \theta} f_\theta(\lambda) d\lambda\right\}\right) * W,$$

as $N, T \to \infty$, where W is some probability distribution that may be different for each τ_{NT} and "*" is the convolution operator. This expression for the limit distribution implies Theorem 2.

In our setting, there are as many individual effects as the cross-sectional sample size, so we have to deal with an infinite-dimensional parameter of the form $(\theta, \eta_1, \eta_2, \cdots)$. A conventional

[9]For convolution theorems in cross section contexts, see e.g. Bickel et al. (1993) and van der Vaart (2000). These references review not only parametric cases but also semiparametric situations. For time series contexts, see, e.g., Taniguchi and Kakizawa (2000).

finite-dimensional convolution theorem cannot be applied to obtain the efficiency bound. We follow Hahn (2002) and Hahn and Kuersteiner (2002) and apply the infinite-dimensional convolution theorem by van der Vaart and Wellner (1996), which allows a parameter space to be a Banach space.[10] We note that the infinite-dimensional convolution theorem of van der Vaart and Wellner is general enough to cover our estimation problem, but the theorem is too abstract to give us a direct guidance to calculate the explicit form of the efficiency bound for $\gamma_k(\theta)$. Hahn (2002) gives a useful characterization of the infinite-dimensional convolution theorem for a special case where the Banach-valued parameter can be decomposed into a finite-dimensional part of interest and a (possibly) infinite-dimensional part which is considered to be a nuisance parameter. However, we cannot directly apply his result because Hahn's specialization only provides the efficiency bound for estimation of the finite-dimensional parameter itself, while we are interested in a real-valued functional of the finite-dimensional parameter, i.e., $\gamma_k(\theta)$. For this reason, we slightly modify Hahn's characterization of the infinite-dimensional convolution theorem to make it appropriate for our situation (see Appendix A.2). The characterization shows that the lower bound involves the second moments of the residuals in the projection of the score function of the finite-dimensional parameter into the space of the score functions of the infinite-dimensional parameter. In Appendix A.3, we use this characterization to calculate the explicit form of the efficiency bound for $\gamma_k(\theta)$.

A key condition for applying a Hajék-type convolution theorem is the local asymptotic normality (LAN) of the model considered. A sequence of statistical experiments or models is said to be LAN if the local log likelihood ratio process admits a certain quadratic stochastic expansion with the first term being asymptotically normal and the second term being $-1/2$ times the variance of the limit distribution of the first term. For a precise definition of the local asymptotic normality, see Appendix A.2. Heuristically speaking, when the LAN property is satisfied, the sequence of statistical experiments can be approximated by a certain normal limit experiment. This means that any sequence of statistics which has a limit law can be approximated by some randomized statistic in a normal experiment. The proof of a convolution theorem proceeds based on this approximation. For details on the 'limits of experiments' arguments, see van der Vaart (1998, Chapter 9). The LAN condition required in the infinite-dimensional convolution theorem is more general than that in a conventional finite-dimensional one. In Appendix A.3, we show that our dynamic panel data model with individual effects is locally asymptotically normal in the sense of van der Vaart and Wellner (1996) (see Theorem 6 in the Appendix). As might be expected, the derivation of the LAN property of our general dynamic panel model is non-

[10] Appendix A.2 briefly reviews their results.

trivial because of the presence of the infinite-dimensional parameter $(\eta_1, \eta_2, \cdots)$ and because we consider double asymptotics. This derivation is one of the major technical contributions of the paper.

Remark 3.5. Let ψ be a differentiable function from Θ into $\mathbb{R}^M$. In Theorem 7 in the Appendix, we provide the efficiency bound for estimation of $\psi(\theta)$. Thus, the efficiency bound for estimation of θ itself can be derived from Theorem 7. Indeed, it can be easily shown that the efficiency bound for estimation of θ is given by $\Gamma(\theta)^{-1}$, the inverse of a Gaussian Fisher information matrix. Noting that our dynamic panel data model is general enough to include panel AR(p), panel MA(q) and panel ARMA(p, q) models with Gaussian innovations and individual effects, it is seen that one can use our result to assess the asymptotic efficiency of autoregressive and moving-average coefficient estimators for those models. However, such efficiency analysis goes beyond the scope of this paper.

Theorems 3 is proved by following the argument of Kakizawa and Taniguchi (1994). The representation of the asymptotic variance given in Remark 2.2 indicates that $\tilde{\gamma}_k$ achieves the efficiency bound shown in Theorem 3 when

$$\begin{aligned}\int_{-\pi}^{\pi} f_\theta^2(\lambda)\cos^2(k\lambda)d\lambda &= \left\{\int_{-\pi}^{\pi}\cos(k\lambda)\frac{\partial}{\partial\theta'}f_\theta(\lambda)d\lambda\right\}\left\{\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta'}f_\theta(\lambda)\frac{\partial}{\partial\theta}f_\theta(\lambda)\frac{d\lambda}{f_\theta^2(\lambda)}\right\}^{-1} \\ &\quad \left\{\int_{-\pi}^{\pi}\cos(k\lambda)\frac{\partial}{\partial\theta}f_\theta(\lambda)d\lambda\right\}. \qquad (3.5)\end{aligned}$$

The generalized Cauchy-Schwarz inequality, which is provided in Lemma 3 in Kakizawa and Taniguchi (1994), implies that the equality (3.5) holds if and only if the condition (3.3) is satisfied. Example 1 in Kakizawa and Taniguchi (1994) demonstrates that, if the process follows an ARMA(p, q) model so that the spectral density has the form given in Remark 3.2, the condition (3.3) reduces to $p \geq q$ and $0 \leq k \leq p - q$. We note that Kakizawa and Taniguchi's (1994) derivation of the condition (3.3) only depends on the forms of the lower bound and the asymptotic variance of sample autocovariances. Thus, we can use their argument without any modification even though we consider panel data settings while their result is for time series.

4 Conclusion

In this paper, we investigate the asymptotic efficiency of autocovariance estimation in a general dynamic panel data model with individual effects when both the cross-sectional sample size and the length of time series tend to infinity. By using the infinite dimensional convolution theorem of van der Vaart and Wellner (1996), the efficiency bound for regular estimators of the k-th order autocovariance is derived. It should be emphasized that the derivation is non-trivial and technically involved because of the presence of individual effects. In view of the lower bound, we provide a necessary and sufficient condition for Okui's (2010) bias-corrected WG sample autocovariance estimators to achieve the lower bound. In particular, we show that when the individual dynamics follows a Gaussian stationary ARMA(p,q) model, the bias-corrected WG sample autocovariance estimator at lag k is asymptotically efficient if and only if $p-q \geq 0$ and $0 \leq k \leq p-q$.

In the process of the derivation of the efficiency bound for autocovariances, we also provide the lower bound for estimation of the model parameter itself. Note that our setting is general enough to include dynamic panel data models such as panel AR(p), panel MA(q) and panel ARMA(p,q) models with Gaussian innovations and individual effects. Thus, our results can be applicable, for example, to assess the asymptotic efficiency of autoregressive and/or moving-average coefficient estimators proposed for those models. We are currently working on such a line of efficiency analysis.

A Appendix

A.1 Preliminaries

In this subsection, we list some properties concerning covariance matrices for stationary processes associated with spectral density f_θ. These are frequently used in the sequel. Before stating these properties, we begin with some notational conventions. First we denote a trace operator by 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. For any Euclidean vector a, these two norms coincide and is denoted by $\|a\|_E$. Note that, for any conformable matrices A and B, we have the relation $\|AB\|_E \leq \|A\|_B \|B\|_E$, which is also useful in the subsequent subsections.

Lemma 1. *Suppose that Assumption 1 and 3 are satisfied. Then the following hold.*

(i)

$$\|\Omega_T(\theta)\|_B \leq 2\pi \sup_\lambda f_\theta(\lambda) \leq \sum_{k=-\infty}^{\infty} |\gamma_k(\theta)| < \infty,$$

$$\|\Omega_T(\theta+\epsilon) - \Omega_T(\theta)\|_B \leq 2\pi \sup_\lambda |f_{\theta+\epsilon}(\lambda) - f_\theta(\lambda)| \to 0 \quad \text{as } \epsilon \to 0.$$

(ii)

$$\|\Omega_T^{-1}(\theta)\|_B \leq \frac{1}{2\pi} \sup_\lambda f_\theta^{-1}(\lambda) < \infty,$$

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

(iii)

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

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

(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 is a special case of Corollary 3.3 in Davies (1973), although there is a minor difference: Davies (1973) uses $\{\exp(-2\pi im\lambda)\}_{m\in\mathbb{Z}}$ as a complete orthonormal set of $L^2[0,1]$, whereas we use $\left\{\frac{1}{2\pi}\exp(-im\lambda)\right\}_{m\in\mathbb{Z}}$ as a complete orthonormal set of $L^2[-\pi,\pi]$. However, this is not an essential difference. □

A.2 A Convolution Theorem

In order to derive the efficiency bound for any regular estimators of $\gamma_k(\theta)$, we will employ the infinite dimensional convolution theorem by van der Vaart and Wellner (1996). In this subsection, we briefly review their result and then specialize it to the case appropriate to our situation (Theorem 5 below).

To begin with, we introduce some of their notation and definitions. Let H be a linear subspace of a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. For each N and 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 $\kappa_{N,h}$ given an observation with law $P_{N,h}$. Let $\{\Delta_h : h \in H\}$ be an *iso-Gaussian process* indexed by H: 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* 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 $N \to \infty$,

$$\Delta_{N,h} \overset{0}{\rightsquigarrow} \Delta_h. \tag{A.1}$$

Here $\overset{h}{\rightsquigarrow}$ denotes weak convergence under $P_{N,h}$. By the iso-Gaussianity assumption on $\{\Delta_h : h \in H\}$, the condition (A.1) is equivalent to saying that, for any finite subset $\{h_1, h_2, \cdots, h_d\} \subseteq H$,

$$\begin{pmatrix} \Delta_{N,h_1} \\ \Delta_{N,h_2} \\ \vdots \\ \Delta_{N,h_d} \end{pmatrix} \overset{0}{\rightsquigarrow} N(0, (\langle h_i, h_j \rangle)) \tag{A.2}$$

as $N \to \infty$ where $(\langle h_i, h_j \rangle)$ is a $d \times d$ matrix whose (i,j)-th component is $\langle h_i, h_j \rangle$. If $\Delta_{N,h}$ is linear in h, i.e., for any positive integer d and any $a = (a_1, a_2, \cdots, a_d)' \in \mathbb{R}^d$,

$$\Delta_{N, \sum_{i=1}^d a_i h_i} = \sum_{i=1}^d a_i \Delta_{N,h_i}, \tag{A.3}$$

then the conditon

$$\Delta_{N,h} \overset{0}{\rightsquigarrow} N(0, \|h\|^2) \quad \forall h$$

is equivalent to (A.2). This follows from an application of the Cramér-Wold device.

The sequence of parameters $\kappa_N(h)$ is assumed to take values in a Banach space B. It is also assumed to be *regular* in the sense that, as $N \to \infty$,

$$r_N(\kappa_N(h) - \kappa_N(0)) \to \dot{\kappa} \quad \forall h \in H$$

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

$$r_N(\tau_N - \kappa_N(h)) \overset{h}{\rightsquigarrow} L \quad \forall h \in H.$$

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

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

for $b^* \in B^*$.

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

Theorem 4 ((van der Vaart and Wellner, 1996, Theorem 3.11.2))**.** *Assume that $(P_{N,h} : h \in H)$ is locally asymptotically normal. Also assume that the sequence of parameters $\kappa_N(h)$ and estimators τ_N are regular. Then, the limit distribution L of $r_N(\tau_N - \kappa_N(0))$ equals a sum $G+W$ of independent, tight, Borel measurable random elements in B such that*

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

Hahn (2002) specializes the infinite-dimensional convolution theorem above for a case where a component of h is a real number, so that we can write, say, $h = (\delta, \xi)$ for some $\delta \in \mathbb{R}$.[11] In his specialization, he fix $\delta_0 \in \mathbb{R}$ and sets $\delta_0 + \delta/r_N$ as a parameter to be estimated i.e. $\kappa_N(h) = \kappa_N(\delta, \xi) = \delta_0 + \delta/r_N$. In our case, the parameter to be estimated is the k-th order autocovariance which is parameterized by $\theta_0 + \tilde{\theta}/r_N$ i.e. $\kappa_N(h) = \kappa_N(\tilde{\theta}, \tilde{\eta}) = \gamma_k(\theta_0 + \tilde{\theta}/r_N)$ where $r_N = \sqrt{NT}$ (for the details of this parameterization, see the next subsection). Obviously our situation is not covered by Hahn's theorem, so that we need to modify Hahn's result in order to cover our setup. The following theorem is sufficient for our purpose.

Theorem 5. *Suppose that $\{P_{N,h} : h \in H\}$ is locally asymptotically normal. Also suppose that (i) $h = (\delta, \xi)$ for $\delta \in \mathbb{R}^L$ and $\xi \in \Xi$ (an inner product space), i.e., $H = \mathbb{R}^L \times \Xi$, (ii) the inner product on H is a sum of inner products on $\mathbb{R}^L$ and Ξ, (iii) r_N is a real sequence with $r_N \to \infty$ and $\kappa_N(h) := \psi(\delta_0 + \delta/r_N)$ for some fixed $\delta_0 \in \mathbb{R}^L$ and a totally differentiable function $\psi : \mathbb{R}^L \to \mathbb{R}^M$ with its derivative $\dot{\psi}$, (iv) $\Delta_h = \delta'\Delta_1 + \Delta_2(\xi)$ for an L-dimensional random vector $\Delta_1 = (\Delta_1^{(1)}, \Delta_1^{(2)}, \cdots, \Delta_1^{(L)})'$ and a random variable $\Delta_2(\xi)$ and (v) $\{\Delta_2(\xi) : \xi \in \Xi\}$ is a linear subspace of a set of square-integrable random variables on the real line. Let $\tilde{\Delta}_1 = (\tilde{\Delta}_1^{(1)}, \tilde{\Delta}_1^{(2)}, \cdots, \tilde{\Delta}_1^{(L)})'$ be a vector of residuals in the projection of $\Delta_1^{(j)}$ on the closure of $\{\Delta_2(\Xi) : \xi \in \Xi\}$. We also assume that (vi) the matrix $\mathbb{E}\tilde{\Delta}_1\tilde{\Delta}_1'$ is nonsingular.Then for any regular estimator τ_N, we have, as $N \to \infty$,*

$$r_N(\tau_N - \kappa_N(0)) \xrightarrow{d} N\left(0, \dot{\psi}(\delta_0)\left(\mathbb{E}\tilde{\Delta}_1\tilde{\Delta}_1'\right)^{-1}\dot{\psi}(\delta_0)'\right) * W$$

where W is some distribution on the real line and $*$ denotes a convolution operator.

Proof. We extend the proof of Theorem 4 in Hahn (2002). First consider the case when ψ is real-valued. Since ψ is assumed to be differentiable, it is easy to see that $\kappa_N(h) = \psi(\delta_0 + \delta/r_N)$ is regular. In fact, as $N \to \infty$,

$$r_N(\kappa_N(h) - \kappa_N(0)) = r_N\left(\psi\left(\delta_0 + \frac{\delta}{r_N}\right) - \psi(\delta_0)\right) \to \dot{\psi}(\delta_0)\delta,$$

and the row vector $\dot{\psi}(\delta_0)$ obviously defines a bounded linear operator from $\mathbb{R}^L$ into $\mathbb{R}$.

For $d \in \mathbb{R}$, we write $\dot{\kappa}^* : d \mapsto \dot{\kappa}^*(d) = (\dot{k}_1^*(d), \dot{\kappa}_2^*(d))$ [12] where $\dot{\kappa}_1^*(d)$ is the L-dimensional Euclidean part of $\dot{\kappa}^*(d)$. Note that the adjoint operator $\dot{k}^*$ is determined by the relation

$$\langle(\dot{\kappa}_1^*(d), \dot{\kappa}_2^*(d)), (\delta, \xi)\rangle = d\dot{\psi}(\delta_0)\delta \quad \forall(\delta, \xi).$$

Now setting $d = 1$ and writing $(\dot{\kappa}_1^*, \dot{\kappa}_2^*) := (\dot{k}_1^*(1), \dot{\kappa}_2^*(1))$, we have

$$\langle(\dot{\kappa}_1^*, \dot{\kappa}_2^*), (\delta, \xi)\rangle = \dot{\psi}(\delta_0)\delta \quad \forall(\delta, \xi). \tag{A.4}$$

[11] Note that the notation here is slightly different from Hahn's (2002).

[12] Note that $B = \mathbb{R}$ and we may and do identify the dual space $\mathbb{R}^*$ with $\mathbb{R}$.

In particular, substituting $(\delta,\xi)=(\dot\kappa_1^*,\dot\kappa_2^*)$ yields

$$\|(\dot\kappa_1^*,\dot\kappa_2^*)\|^2=\dot\psi(\delta_0)\dot\kappa_1^*. \tag{A.5}$$

Below we will show that

$$\mathbb{E}\tilde\Delta_1\tilde\Delta_1'\dot\kappa_1^*=\dot\psi(\delta_0)'. \tag{A.6}$$

If this holds, then the equality (A.5) yields the desired result for the case of a real-valued ψ.

To show (A.6), we write $\dot\kappa_1^*=(\dot\kappa_{11}^*,\dot\kappa_{12}^*,\cdots,\dot\kappa_{1L}^*)'$ and assume for simplicity only that $\dot\kappa_{1j}^*\neq 0$ for all $j=1,2\cdots,L$ (a proof would be similar but more complicated if $\dot\kappa_{1j}^*=0$ for some j). Let e_i be the i-th column vector of an $L\times L$ identity matrix. Substituting $\delta=e_i$ into the equality (A.4), we have

$$\langle(\dot\kappa_1^*,\dot\kappa_2^*),(e_i,\xi)\rangle=\dot\psi(\delta_0)e_i\quad\forall\xi.$$

Furthermore, observe that

$$\begin{aligned}\langle(\dot\kappa_1^*,\dot\kappa_2^*),(e_i,\xi)\rangle &= \mathbb{E}\left[\left(\Delta_1'\dot\kappa_1^*+\Delta_2\left(\dot\kappa_2^*\right)\right)\left(\Delta_1^{(i)}+\Delta_2(\xi)\right)\right]\\ &= \sum_{j=1}^L\dot\kappa_{1j}^*\mathbb{E}\left[\left(\Delta_1^{(j)}+\Delta_2\left(\xi_j^*\right)\right)\left(\Delta_1^{(i)}+\Delta_2(\xi)\right)\right]\end{aligned}$$

for some ξ_j^* such that $\Delta_2(\xi_j^*)=(1/Lk_{1j}^*)\Delta_2(k_2^*)$. Thus we have

$$\sum_{j=1}^L\dot\kappa_{1j}^*\mathbb{E}\left[\left(\Delta_1^{(j)}+\Delta_2\left(\xi_j^*\right)\right)\left(\Delta_1^{(i)}+\Delta_2(\xi)\right)\right]=\dot\psi(\delta_0)e_i\quad\forall\xi. \tag{A.7}$$

From this, it follows that

$$\sum_{j=1}^L\dot\kappa_{1j}^*\mathbb{E}\left[\left(\Delta_1^{(j)}+\Delta_2\left(\xi_j^*\right)\right)\Delta_2(\xi)\right]=\text{constant}\quad\forall\xi.$$

This reveals that

$$\sum_{j=1}^L\dot\kappa_{1j}^*\mathbb{E}\left[\left(\Delta_1^{(j)}+\Delta_2\left(\xi_j^*\right)\right)\Delta_2(\xi)\right]=0\quad\forall\xi. \tag{A.8}$$

Now recalling that $\hat\Delta_1^{(j)}$ is a projection of $\Delta_1^{(j)}$ on the closure of $\{\Delta_2(\xi):\xi\in\Xi\}$, we have

$$\mathbb{E}\left[\Delta_1^{(j)}\Delta_2(\xi)\right]=\mathbb{E}\left[\hat\Delta_1^{(j)}\Delta_2(\xi)\right]\quad\forall\xi.$$

In addition, there exits a sequence $\hat\Delta_{1n}^{(i)}$ in $\{\Delta_2(\xi):\xi\in\Xi\}$ such that $\hat\Delta_{1n}^{(i)}\to\hat\Delta_1^{(i)}$ in L^2 as $n\to\infty$, and $\hat\Delta_{1n}^{(i)}+\Delta_2(\xi_i^*)$ are all in $\{\Delta_2(\xi):\xi\in\Xi\}$. Thus from (A.8) it follows that

$$\sum_{j=1}^L\dot\kappa_{1j}^*\mathbb{E}\left[\left(\hat\Delta_1^{(j)}+\Delta_2\left(\xi_j^*\right)\right)\left(\hat\Delta_{1n}^{(i)}+\Delta_2(\xi_i^*)\right)\right]=0\quad\forall n.$$

By continuity of inner product, we have

$$\sum_{j=1}^{L} \dot\kappa_{1j}^* \mathbb{E}\left[\left(\hat\Delta_1^{(j)} + \Delta_2\left(\xi_j^*\right)\right)\left(\hat\Delta_1^{(i)} + \Delta_2(\xi_i^*)\right)\right] = 0.$$

Multiplying the both hand sides of the last display by $\dot\kappa_{1i}^*$ and summing up over $i = 1, 2, \cdots, L$, we obtain

$$\mathbb{E}\left[\left(\sum_{j=1}^{L} \dot\kappa_{1i}^* \left(\hat\Delta_1^{(j)} + \Delta_2(\xi_j^*)\right)\right)^2\right] = 0.$$

This implies that, almost surely,

$$-\sum_{j=1}^{L} \dot\kappa_{1j}^* \hat\Delta_1^{(j)} = \sum_{j=1}^{L} \dot\kappa_{1j}^* \Delta_2(\xi_j^*).$$

Substite this into (A.7), replace $\Delta_2(\xi)$ by $\hat\Delta_{1n}^{(i)}$ and let $n \to \infty$ to obtain

$$\sum_{j=1}^{L} \dot\kappa_{1j}^* \mathbb{E}\left[\tilde\Delta_1^{(j)} \tilde\Delta_1^{(i)}\right] = \dot\psi(\delta_0) e_i \quad \forall i = 1, 2, \cdots, L.$$

Stacking these equalities over i yields the desired identity (A.6). This completes the proof for the case of a real-valued ψ.

Based on the above result, we next consider a more general case where ψ is M-dimensional. Since it is assumed that τ_N is regular, there exists some M-dimensional random vector L such that

$$r_N\left(\tau_N - \psi\left(\delta_0 + \frac{\delta}{r_N}\right)\right) \overset{h}{\rightsquigarrow} L \quad \forall h \in H.$$

Thus, for any fixed $a \in \mathbb{R}^M$, we have

$$r_N\left(a'\tau_N - a'\psi\left(\delta_0 + \frac{\delta}{r_N}\right)\right) \overset{h}{\rightsquigarrow} a'L \quad \forall h \in H,$$

which means that $a'\tau_N$ is regular for $a'\psi(\delta_0 + \frac{\delta}{r_N})$. It follows from the convolution result for a one-dimensional function ψ that

$$r_N\left(a'\tau_N - a'\psi\left(\delta_0 + \frac{\delta}{r_N}\right)\right) \overset{0}{\rightsquigarrow} G_a + W_a$$

where G_a and W_a are independent random variables with

$$G_a \sim N\left(0, a'\dot\psi(\delta_0)\left(\mathbb{E}\tilde\Delta_1\tilde\Delta_1'\right)^{-1}\dot\psi(\delta_0)'a\right).$$

Now we write $G_a = a'G$ where

$$G \sim N\left(0, \dot\psi(\delta_0)\left(\mathbb{E}\tilde\Delta_1\tilde\Delta_1'\right)^{-1}\dot\psi(\delta_0)'\right).$$

Because $a'L = a'G + W_a$, we can write $W_a = a'(L - G)$. Defining $W := L - G$, we have $a'L = a'(G + W)$. Since a is arbitrary, the Cramer-Wold device implies that $L = G + W$, which completes the proof for an M-dimensional case.

□

A.3 Derivation of the Lower Bound

In this subsection, we first show the local asymptotic normality of our panel data model with individual effects in the sense of van der Vaart and Wellner (1996) and then prove Theorem 2, which presents the efficiency bound for estimation of $\gamma_k(\theta)$.

To this end, we let $y_j = (y_{j1}, y_{j2}, \cdots, y_{jT})'$ and define η to be a sequence of individual effects η_j's, i.e., $\eta := (\eta_1, \eta_2, \cdots)$. Now fix (θ, η) and let $(\tilde{\theta}, \tilde{\eta})$ be local parameters. We localize the parameters around (θ, η) as follows:

$$\theta + \frac{\tilde{\theta}}{\sqrt{NT}} \quad \text{and} \quad \eta + \frac{\tilde{\eta}}{\sqrt{NT}}.$$

We also define $h := (\tilde{\eta}, \tilde{\theta})$. For simplicity of notation, we write $\Omega_{\tilde{\theta}} := \Omega_T\left(\theta + \frac{\tilde{\theta}}{\sqrt{NT}}\right)$. Note that, in this notation, $\Omega_0 = \Omega_T(\theta)$.

The local log likelihood ratio of our panel data model with individual effects is given by

$$\begin{aligned}
\log \frac{dP_{N,h}}{dP_{N,0}} &= \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} + \frac{1}{2} \sum_{j=1}^{N} (y_j - \eta_j \mathbf{1}_T)' \Omega_0^{-1} (y_j - \eta_j \mathbf{1}_T) \\
&\quad - \frac{1}{2} \sum_{j=1}^{N} \left(y_j - \left(\eta_j + \frac{\tilde{\eta}_j}{\sqrt{NT}} \right) \mathbf{1}_T \right)' \Omega_{\tilde{\theta}}^{-1} \left(y_j - \left(\eta_j + \frac{\tilde{\eta}_j}{\sqrt{NT}} \right) \mathbf{1}_T \right) \\
&= \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} \\
&\quad + \frac{1}{2} \sum_{j=1}^{N} (y_j - \eta_j \mathbf{1}_T)' \Omega_0^{-1} (y_j - \eta_j \mathbf{1}_T) - \frac{1}{2} \sum_{j=1}^{N} (y_j - \eta_j \mathbf{1}_T)' \Omega_{\tilde{\theta}}^{-1} (y_j - \eta_j \mathbf{1}_T) \\
&\quad + \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} (y_j - \eta_j \mathbf{1}_T) - \frac{1}{2NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} \mathbf{1}_T.
\end{aligned}$$

Since $y_j - \eta_j \mathbf{1}_T = w_i$ under $P_{N,0}$, we can write the log likelihood ratio as follows:

$$\begin{aligned}
\log \frac{dP_{N,h}}{dP_{N,0}} &= \frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} \\
&\quad + \frac{1}{2} \sum_{j=1}^{N} w_j' \Omega_0^{-1} w_j - \frac{1}{2} \sum_{j=1}^{N} w_j' \Omega_{\tilde{\theta}}^{-1} w_j \\
&\quad + \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} w_j - \frac{1}{2NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} \mathbf{1}_T.
\end{aligned}$$

Lemma 2. *Under $P_{N,0}$, as $N \to \infty$ and $T \to \infty$,*

$$\begin{aligned}
&\frac{N}{2} \log\det \Omega_0 - \frac{N}{2} \log\det \Omega_{\tilde{\theta}} + \frac{1}{2} \sum_{j=1}^{N} w_j' \Omega_0^{-1} w_j - \frac{1}{2} \sum_{j=1}^{N} w_j' \Omega_{\tilde{\theta}}^{-1} w_j \\
&= \frac{1}{2\sqrt{NT}} \sum_{j=1}^{N} \left\{ w_j' \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} w_j - \operatorname{tr}\left(\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \right) \right\} \\
&\qquad - \frac{1}{4T} \operatorname{tr}\left(\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \right) + o_{P_{N,0}}(1)
\end{aligned}$$

where

$$(\tilde{\theta}\nabla\Omega_0) := \sum_{m=1}^{L} \tilde{\theta}_m \frac{\partial}{\partial \theta_m}\Omega(\theta).$$

Proof. This proof is similar to that of Theorem 4.4 in Davies (1973). The difference is that we take into account cross sections and use double asymptotics. Thus several modifications are needed.

Let us define

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

It suffices to show that $\mathbb{E}_\theta[a_{NT}]$ and $\mathrm{Var}_\theta[a_{NT}]$ tend to 0 as $N, T \to \infty$.

Note that, for any symmetric matrix A, we have $\mathbb{E}_\theta[w_j'Aw_j] = \mathrm{trace}(A\Omega_0)$. From this, it follows that

$$\mathbb{E}_\theta[a_{NT}] = \frac{N}{2}\left\{\log\det(\Omega_0\Omega_{\tilde{\theta}}^{-1}) + \mathrm{tr}(I_T - \Omega_0\Omega_{\tilde{\theta}}^{-1})\right\} + \frac{1}{4T}\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right),$$

where I_T denotes a $T\times T$ identity matrix. Setting $A_{NT} := \Omega_0\Omega_{\tilde{\theta}}^{-1} - I = (\Omega_0 - \Omega_{\tilde{\theta}})\Omega_{\tilde{\theta}}^{-1}$, we have

$$\begin{aligned} \mathbb{E}_\theta a_{NT} &= \frac{N}{2}\left\{\log\det(I_T + A_{NT}) - \mathrm{tr}(A_{NT}) + \frac{1}{2}\mathrm{tr}(A_{NT}^2)\right\} \\ &+ \frac{1}{4T}\left\{\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right) - NT\mathrm{tr}\left(\Omega_{\tilde{\theta}}^{-1}(\Omega_0 - \Omega_{\tilde{\theta}})\Omega_{\tilde{\theta}}^{-1}(\Omega_0 - \Omega_{\tilde{\theta}})\right)\right\} \end{aligned} \quad (A.9)$$

It is known that, for any symmetric matrix A with $\|A\|_B < 1$, we have

$$|\log\det(I + A) - \mathrm{tr}(A) + \frac{1}{2}\mathrm{tr}(A^2)| \le \frac{1}{3}\|A\|_B\|A\|_E^2(1 - \|A\|_B)^{-3}$$

where I is the identity matrix.[13] Further, since $\|\Omega_0 - \Omega_{\tilde{\theta}}\|_B$ tends to zero and $\|\Omega_{\tilde{\theta}}^{-1}\|_B$ is bounded as $N, T \to \infty$ by Lemma 1 (i) and (ii), we see that $\|A_{NT}\|_B \le \|\Omega_0 - \Omega_{\tilde{\theta}}\|_B\|\Omega_{\tilde{\theta}}^{-1}\|_B = o(1)$ as $N, T \to \infty$. Thus, for sufficiently large N and T, the first term of the right hand side of (A.9) is bounded by

$$\frac{N}{6}\|A_{NT}\|_B\|A_{NT}\|_E^2(1 - \|A_{NT}\|_B)^{-3} \le \frac{1}{6T}\|\tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E^2\|\Omega_0 - \Omega_{\tilde{\theta}}\|_B\|\Omega_{\tilde{\theta}}^{-1}\|_B^3(1 - \|A_T\|_B)^{-3},$$

where $\bar{\theta} = u\tilde{\theta}$ and u is some real number in $(0, 1)$ that determines a mean value for Tayler's expansion of $\tilde{\theta} \mapsto \Omega_{\tilde{\theta}} = \Omega(\theta + \tilde{\theta}/\sqrt{NT})$ around 0. Because $(1/T)\|\tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E^2$ is bounded by Lemma 1 (iii), the right hand side of (A.9) turns out to be $o(1)$ as $N, T \to \infty$.

[13]For the proof of this result, see e.g. Appendix II of Davies (1973).

As for the second term of (A.9), observe that

$$
\begin{aligned}
&\frac{1}{4T}\left|\operatorname{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right) - NT\operatorname{tr}\left(\Omega_{\tilde{\theta}}^{-1}(\Omega_0-\Omega_{\tilde{\theta}})\Omega_{\tilde{\theta}}^{-1}(\Omega_0-\Omega_{\tilde{\theta}})\right)\right| \\
&\quad = \frac{1}{4T}\left|\operatorname{tr}\left(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\right) - \operatorname{tr}\left(\Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\right)\right| \\
&\quad = \frac{1}{4T}\Big|\operatorname{tr}\Big(\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0) - \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0) \\
&\qquad + \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0) - \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0) \\
&\qquad + \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_0^{-1}(\tilde{\theta}\nabla\Omega_0) - \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_0) \\
&\qquad + \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_0) - \Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Omega_{\tilde{\theta}}^{-1}(\tilde{\theta}\nabla\Omega_{\bar{\theta}})\Big)\Big| \\
&\leq \frac{1}{4T}\|\tilde{\theta}\nabla\Omega_0\|_E^2\|\Omega_0^{-1}\|_B^2\|\Omega_{\tilde{\theta}}^{-1}\|_B\|\Omega_{\tilde{\theta}}-\Omega_0\|_B + \frac{1}{4T}\|\tilde{\theta}\nabla\Omega_0 - \tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E\|\tilde{\theta}\nabla\Omega_0\|_E\|\Omega_0^{-1}\|_B\|\Omega_{\tilde{\theta}}^{-1}\|_B \\
&+ \frac{1}{4T}\|\tilde{\theta}\nabla\Omega_0\|_E\|\tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E\|\Omega_0^{-1}\|_B\|\Omega_{\tilde{\theta}}^{-1}\|_B^2\|\Omega_{\tilde{\theta}}-\Omega_0\|_B + \frac{1}{4T}\|\tilde{\theta}\nabla\Omega_0 - \tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E\|\tilde{\theta}\nabla\Omega_{\bar{\theta}}\|_E\|\Omega_{\tilde{\theta}}^{-1}\|_B^2.
\end{aligned}
$$

It follows from the same arguments as above that the extreme right hand side of the last display is $o(1)$ as $N, T \to \infty$. Thus we have $\mathbb{E}_\theta[a_{NT}] = o(1)$.

The proof that $\mathrm{Var}_\theta[a_{NT}] \to 0$ as $N, T \to \infty$ is similar to the above arguments and thus the details are omitted. □

Lemma 3. *As $T \to \infty$,*

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

where

$$\Gamma(\theta) := \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}f_\theta(\lambda)\frac{\partial}{\partial\theta'}f_\theta(\lambda)\frac{d\lambda}{f_\theta^2(\lambda)}.$$

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

Lemma 4. *Under $P_{N,0}$, as $N \to \infty$ and $T \to \infty$,*

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\tilde{\eta}_j\mathbf{1}_T'\Omega_{\tilde{\theta}}^{-1}w_j - \frac{1}{2NT}\sum_{j=1}^{N}\tilde{\eta}_j^2\mathbf{1}_T'\Omega_{\tilde{\theta}}^{-1}\mathbf{1}_T &= \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\tilde{\eta}_j\mathbf{1}_T'\Omega_0^{-1}w_j \\
&\quad - \frac{1}{2NT}\sum_{j=1}^{N}\tilde{\eta}_j^2\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T + o_{p_{N,0}}(1).
\end{aligned}
$$

Proof. First observe that

$$
\begin{aligned}
\mathbb{E}_\theta&\left|\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\tilde{\eta}_j\mathbf{1}_T'\Omega_{\tilde{\theta}}^{-1}w_j - \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\tilde{\eta}_j\mathbf{1}_T'\Omega_0^{-1}w_j\right|^2 \\
&= \frac{1}{NT}\sum_{j=1}^{N}\tilde{\eta}_j^2\mathbf{1}_T'(\Omega_{\tilde{\theta}}^{-1}-\Omega_0^{-1})\Omega_\theta(\Omega_{\tilde{\theta}}^{-1}-\Omega_0^{-1})\mathbf{1}_T \\
&\leq \left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\eta}_j^2\right)\|\Omega_0^{-1}\|_B^2\|\Omega_{\tilde{\theta}}^{-1}\|_B^2\|\Omega_\theta\|_B\|\Omega_0-\Omega_{\tilde{\theta}}\|_B.
\end{aligned}
$$

The extreme right hand side is $o(1)$ as $N, T \to \infty$, by Assumption 2 and Lemma 1 (i) and (ii). Thus we have

$$\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} w_j = \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_0^{-1} w_j + o_{P_{0,N,T}}. \tag{A.10}$$

Next,

$$\left| \frac{1}{NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} \mathbf{1}_T - \frac{1}{NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T \right| \leq \left(\frac{1}{N} \sum_{j=1}^{N} \tilde{\eta}_j^2 \right) \|\Omega_{\tilde{\theta}}^{-1}\|_B \|\Omega_0^{-1}\|_B \|\Omega_{\tilde{\theta}} - \Omega_0\|_B.$$

The extreme right hand side is also $o(1)$ as $N, T \to \infty$. Hence we have

$$\frac{1}{NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_{\tilde{\theta}}^{-1} \mathbf{1}_T = \frac{1}{NT} \sum_{j=1}^{N} \tilde{\eta}_j^2 \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T + o(1), \tag{A.11}$$

as $N, T \to \infty$. Combining (A.10) and (A.11) yields the desired result. □

Lemma 5. *Under $P_{N,0}$, as $N \to \infty$ and $T \to \infty$,*

$$\frac{1}{2\sqrt{NT}} \sum_{j=1}^{N} \left\{ w_j' \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} w_j - \operatorname{tr}\left(\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \right) \right\}$$
$$+ \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_0^{-1} w_j \xrightarrow{d} N\left(0, \tilde{\theta}' \Gamma(\theta) \tilde{\theta} + \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \tilde{\eta}_j^2 \cdot \lim_{T \to \infty} \frac{1}{T} \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T \right)$$

Proof. To prove the asymptotic normality, we use Theorem 2 in Phillips and Moon (1999).[14]

Let us denote by b_{jT} an L-dimensional random vector whose m-th component is

$$\frac{1}{2} \left\{ w_j' \Omega_0^{-1} \left(\frac{\partial}{\partial \theta_m} \Omega(\theta) \right) \Omega_0^{-1} w_j - \operatorname{tr}\left(\Omega_0^{-1} \left(\frac{\partial}{\partial \theta_{\mathrm{m}}} \Omega(\theta) \right) \right) \right\}.$$

Define

$$C_j := \begin{pmatrix} \tilde{\theta}_1' & \tilde{\eta}_{j1} \\ \tilde{\theta}_2' & \tilde{\eta}_{j2} \end{pmatrix} \quad \text{and} \quad Q_{j,T} := \frac{1}{\sqrt{T}} \begin{pmatrix} b_{jT} \\ \mathbf{1}_T' \Omega_0^{-1} w_j \end{pmatrix}. \tag{A.12}$$

Also set $\Sigma_T := \mathbb{E}_\theta[Q_{jT} Q_{jT}']$. Below we will verify the four conditions in Phillips and Moon in the order (iv), (i), (ii) and (iii).

Consider the condition (iv). Note that, for $s = 1, 2$, the (s, s)-th element of $(1/N) \sum_{j=1}^{N} C_j \Sigma_T C_j'$ is

[14] Rigorously speaking, Theorem 3 in Phillips and Moon does not cover our situation because the matrix C_j in the theorem is required to be a *square* matrix conformable with Q_{jT}, while, in our case, the matrix C_j (defined in (A.12)) is not necessarily a square matrix. However, closely examining the proof of Theorem 3 in Phillips and Moon, we can verify that the requirement is not essential and that the result is still valid even when C_j is not a square matrix.

$$\frac{1}{NT}\sum_{j=1}^{N}\mathbb{E}_\theta\left|\frac{1}{2}\left\{w_j'\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}w_j - \mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\right)\right\} + \tilde{\eta}_{js}\mathbf{1}_T'\Omega_0^{-1}w_j\right|^2$$

$$= \frac{1}{2T}\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\right) + \left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\eta}_{js}^2\right)\frac{1}{T}\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T$$

$$+\frac{1}{NT}\sum_{j=1}^{N}\tilde{\eta}_{js}\mathbb{E}_\theta[w_j'\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}w_j\mathbf{1}_T'\Omega_0^{-1}w_j]$$

$$= \frac{1}{2T}\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\right) + \left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\eta}_{js}^2\right)\frac{1}{T}\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T,$$

where the last equality follows from the observation that

$$\begin{aligned}\mathbb{E}_\theta[w_j'\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}w_j\mathbf{1}_T'\Omega_0^{-1}w_j] &= \mathrm{tr}\left[\left(\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}\right)\mathbb{E}_\theta\left[(\Omega_0^{-1}\mathbf{1}_T\otimes w_j')(w_j\otimes w_j')\right]\right] \\ &= 0.\end{aligned}$$

A similar argument shows that, for $s \neq t$, the (s,t)-th element of $(1/N)\sum_{j=1}^{N}C_j\Sigma_T C_j'$ is

$$\frac{1}{2T}\mathrm{tr}\left(\Omega_0^{-1}(\tilde{\theta}_s\nabla\Omega_0)\Omega_0^{-1}(\tilde{\theta}_t\nabla\Omega_0)\right) + \left(\frac{1}{N}\sum_{j=1}^{N}\tilde{\eta}_{js}\tilde{\eta}_{jt}\right)\frac{1}{T}\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T$$

Thus we see that, as $N, T \to \infty$, $(1/N)\sum_{j=1}^{N}C_j\Sigma_T C_j'$ converges to the covariance matrix (**??**) by Lemma 3 and Assumptions 2 and 4. To show the positive definiteness of the covariance matrix (**??**), observe that it can be written as

$$\begin{pmatrix}\tilde{\theta}_1' \\ \tilde{\theta}_2'\end{pmatrix}\Gamma(\theta)\begin{pmatrix}\tilde{\theta}_1 & \tilde{\theta}_2\end{pmatrix} + \rho\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\begin{pmatrix}\tilde{\eta}_{j1}^2 & \tilde{\eta}_{j1}\tilde{\eta}_{j2} \\ \tilde{\eta}_{j1}\tilde{\eta}_{j2} & \tilde{\eta}_{j2}^2\end{pmatrix}.$$

Since $\Gamma(\theta)$ is positive definite and θ_1 and θ_2 are linearly independent, we deduce that the covariance matrix is positive definite. Thus the condition (iv) is satisfied.

Next turn to the condition (i). From the same arguments as above, we see that

$$\Sigma_T \to \begin{pmatrix}\Gamma(\theta) & 0 \\ 0' & \rho\end{pmatrix},$$

as $T \to \infty$. Obviously the limit matrix is positive definite. Thus the continuity of eigen values (see e.g. Schott (2005, Theorem 3.14 and Example 3.11)) implies that $\liminf_{T\to\infty}\lambda_{\min}(\Sigma_T) > 0$ where $\lambda_{\min}(\cdot)$ denotes the minimum of the eigenvalues. Hence the condition (i) is satisfied.

To verify the condition (ii), first observe that

$$\frac{1}{N}\sum_{j=1}^{N}C_jC_j' = \begin{pmatrix}\tilde{\theta}_1' \\ \tilde{\theta}_2'\end{pmatrix}\begin{pmatrix}\tilde{\theta}_1 & \tilde{\theta}_2\end{pmatrix} + \frac{1}{N}\sum_{j=1}^{N}\begin{pmatrix}\tilde{\eta}_{j1}^2 & \tilde{\eta}_{j1}\tilde{\eta}_{j2} \\ \tilde{\eta}_{j1}\tilde{\eta}_{j2} & \tilde{\eta}_{j2}^2\end{pmatrix}$$

and note that the limit of $\frac{1}{N}\sum_{j=1}^{N} C_j C_j'$ is positive definite. From this it follows that

$$\liminf_{N\to\infty} \lambda_{\min}\left(\frac{1}{N}\sum_{j=1}^{N} C_j C_j'\right) > 0.$$

Thus, for sufficiently large N,

$$\frac{\max_{j\le N}\|C_j\|_E^2}{\lambda_{\min}\left(\sum_{j=1}^{N} C_j C_j'\right)} \le \frac{\|\tilde\theta_1\|_E^2 + \|\tilde\theta_2\|_E^2 + \max_{j\le N}\tilde\eta_{j1}^2 + \max_{j\le N}\tilde\eta_{j2}^2}{N \liminf_{N\to\infty}\lambda_{\min}\left(\frac{1}{N}\sum_{j=1}^{N} C_j C_j'\right)} = O\left(\frac{1}{N}\right),$$

by Assumption 2 (ii). This verifies the condition (ii) in Phillips and Moon.

The condition (iii) in Phillips and Moon (1999) requires that the process $\{\|Q_{jT}\|_E^2\}_{T\in\mathbb{Z}}$ be uniformly integrable. A sufficient condition for this is that each component of Q_{jT} is bounded in its fourth moment. In other words, it is enough to prove that, for each $m = 1, 2, \cdots L$,

$$\frac{1}{T^2}\mathbb{E}_\theta\left|w_j'\Omega_0^{-1}\left(\frac{\partial}{\partial\theta_m}\Omega(\theta)\right)\Omega_0^{-1}w_j - \mathrm{tr}\left(\Omega_0^{-1}\left(\frac{\partial}{\partial\theta_\mathrm{m}}\Omega(\theta)\right)\right)\right|^4 = O(1) \tag{A.13}$$

and

$$\frac{1}{T^2}\mathbb{E}_\theta\left|\mathbf{1}_T'\Omega_0^{-1}w_j\right|^4 = O(1).$$

We first show the latter. Observe that

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

This is O(1) since $\frac{1}{T}\mathbf{1}_T'\Omega_0^{-1}\mathbf{1}_T$ is convergent as $T\to\infty$ by Assumption 4.

Next we turn to the fourth moment in (A.13). Let $B_T^{(m)} := \Omega_0^{-1}\left(\frac{\partial}{\partial\theta_m}\Omega(\theta)\right)$ and observe that

$$\begin{aligned}&\frac{1}{T^2}\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j - \mathrm{tr}\left(B_T^{(m)}\right)\right|^4 \\ &= \frac{1}{T^2}\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j\right|^4 - 4\mathrm{tr}\left(B_T^{(m)}\right)\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j\right|^3 + 6\left(\mathrm{tr}\left(B_T^{(m)}\right)\right)^2\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j\right|^2 \\ &\qquad - 4\left(\mathrm{tr}\left(B_T^{(m)}\right)\right)^3\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j\right| + \left(\mathrm{tr}\left(B_T^{(m)}\Omega_0\right)\right)^4\end{aligned}$$

Applying the formulae for the third and fourth moments of quadratic forms in Gaussian random vectors (see e.g. Theorem.10.21 in Schott (2005)), we see that

$$\begin{aligned}\frac{1}{T^2}\mathbb{E}_\theta\left|w_j'B_T^{(m)}\Omega_0^{-1}w_j - \mathrm{tr}\left(B_T^{(m)}\right)\right|^4 &= \frac{12}{T^2}\left(\mathrm{tr}\left\{\left(B_T^{(m)}\right)^2\right\}\right)^2 + \frac{48}{T^2}\mathrm{tr}\left\{\left(B_T^{(m)}\right)^4\right\} \\ &\le \frac{60}{T^2}\left\|\frac{\partial}{\partial\theta_m}\Omega(\theta)\right\|_E^4\|\Omega_0^{-1}\|_B^4.\end{aligned}$$

The extreme right hand side is $O(1)$ as $N, T \to \infty$, again by Lemma 1 (ii) and (iii). Therefore $\|Q_{jT}\|_E^2$ is uniformly integrable in T. Now that the conditions of Theorem 3 in Phillips and Moon are all verified, the asymptotic normality is shown. □

Theorem 6. *The local log likelihood ratio of our panel model is asymptotically normal in the sense of van der Vaart and Wellner (1996, p412). That is, under $P_{N,0}$, as $N \to \infty$ and $T \to \infty$, we have*

$$\log \frac{dP_{N,h}}{dP_{N,0}} = \Delta_{N,h} - \frac{1}{2}\|h\|^2 + o_{p_{n,0}}(1),$$

where

$$\Delta_{N,h} = \Delta_N(\tilde{\theta}, \tilde{\eta}) = \frac{1}{2\sqrt{NT}} \sum_{j=1}^{N} \left\{ w_j' \Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \Omega_0^{-1} w_j - \mathrm{tr}\left(\Omega_0^{-1} (\tilde{\theta} \nabla \Omega_0) \right) \right\} + \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_0^{-1} w_j$$

converges weakly under $P_{N,0}$ to $\Delta_h \sim N(0, \|h\|^2)$. Here, $\|h\|^2 = \langle h, h \rangle$ and the inner product $\langle \cdot\,, \cdot \rangle$ is defined by

$$\langle (\tilde{\theta}_1, \tilde{\eta}_1), (\tilde{\theta}_2, \tilde{\eta}_2) \rangle_H := \tilde{\theta}_1' \Gamma(\theta) \tilde{\theta}_2 + \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \tilde{\eta}_{j1} \tilde{\eta}_{j2} \cdot \lim_{T \to \infty} \frac{1}{T} \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T.$$

Proof. This theorem follows from Lemma 2 to 6. □

Theorem 7. *Suppose that Assumption 1 to 4 are satisfied. Let ψ be a differentiable function from Θ into $\mathbb{R}^M$ and let $\tau_{N,T}$ be any regular estimator of $\psi(\theta)$ as $N, T \to \infty$. Then,*

$$\sqrt{NT}(\tau_{NT} - \psi(\theta)) \xrightarrow{d} N\left(0, \dot{\psi}(\theta) \Gamma(\theta)^{-1} \dot{\psi}(\theta)'\right) * W,$$

as $N, T \to \infty$, where $*$ denotes a convolution of probability measures, W is some distribution on the real line. In particular, if the limit law of τ_{NT} has variance Σ_θ, then

$$\Sigma_\theta \geq \dot{\psi}(\theta) \Gamma(\theta)^{-1} \dot{\psi}(\theta)'.$$

Proof. In Theorem 6, We have established the LAN property of our panel data model, so that, in order to apply Theorem 5 into our estimation problem for $\psi(\theta)$, we have to decompose $\Delta(\tilde{\theta}, \tilde{\eta})$ into a sum of random variables of the form as in the condition (iii) in Theorem 5. To this end, let us denote by $\Delta_{1,N}$ an L-dimensional random vector whose m-th component is

$$\frac{1}{2\sqrt{NT}} \sum_{j=1}^{N} \left\{ w_j' \Omega_0^{-1} \left(\frac{\partial}{\partial \theta_m} \Omega(\theta) \right) \Omega_0^{-1} w_j - \mathrm{tr}\left(\Omega_0^{-1} \left(\frac{\partial}{\partial \theta_m} \Omega(\theta) \right) \right) \right\},$$

and also set

$$\Delta_{2,N}(\tilde{\eta}) := \frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \tilde{\eta}_j \mathbf{1}_T' \Omega_0^{-1} w_j.$$

Then we can write $\Delta_N(\tilde{\theta}, \tilde{\eta}) = \tilde{\theta}' \Delta_{1,N} + \Delta_{2,N}(\tilde{\eta})$. Now letting $(\tilde{\theta}_1, \tilde{\eta}_1) = (\tilde{\theta}, 0)$ and $(\tilde{\theta}_2, \tilde{\eta}_2) = (0, \tilde{\eta})$ in Lemma 6, we have, as $N \to \infty$ and $T \to \infty$,

$$\begin{pmatrix} \tilde{\theta}' \Delta_{1,N} \\ \Delta_{2,N}(\tilde{\eta}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \tilde{\theta}' \Delta_1 \\ \Delta_2(\tilde{\eta}) \end{pmatrix},$$

where

$$\Delta_1 = (\Delta_1^{(1)}, \Delta_1^{(2)}, \cdots, \Delta_1^{(L)})' \sim N(0, \Gamma(\theta))$$

and

$$\Delta_2(\tilde{\eta}) \sim N(0, \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \tilde{\eta}_j^2 \cdot \lim_{T\to\infty} \frac{1}{T} \mathbf{1}_T' \Omega_0^{-1} \mathbf{1}_T)$$

with $\mathbb{E}\Delta_1 \Delta_2(\tilde{\eta}) = 0$. Using this joint convergence in distribution, together with the continuous mapping theorem, we easily see that $\Delta(\tilde{\theta}, \tilde{\eta})$ admits a desired decomposition $\Delta(\tilde{\theta}, \tilde{\eta}) = \tilde{\theta}' \Delta_1 + \Delta_2(\tilde{\eta})$.

Finally we calculate $\mathbb{E}\tilde{\Delta}_1 \tilde{\Delta}_1'$. Noting that $\Delta_1^{(m)}$ is orthogonal to $\Delta_2(\tilde{\eta})$ for each $m = 1, 2, \cdots, L$, we can deduce that the projection of $\Delta_1^{(m)}$ on the linear space $\{\Delta_2(\tilde{\eta})\}$ is just 0 (see e.g. Kreyszig (1989, p148)) and the residual $\tilde{\Delta}_1^{(m)}$ in the projection is $\Delta_1^{(m)}$ itself. Hence, it follows that $\mathbb{E}\tilde{\Delta}_1 \tilde{\Delta}_1' = \mathbb{E}\Delta_1 \Delta_1' = \Gamma(\theta)$.

This result and Theorem 5 yield the desired result.

□

Proof of Theorem 2

Setting $\psi(\theta) = \gamma_k(\theta)$ in Theorem 7, we have the desired result.

References

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

[2] Amemiya, T. (1985) *Advanced Econometrics* Harvard University Press, Cambridge, Massachusetts.

[3] Anderson, T. W. *The Statistical Analysis of Time Series,* New York: Wiley.

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

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

[6] Arellano, M. (2003) *Panel Data Econometrics*, Oxford University Press.

[7] Baltagi, B. H. & Q. Li (1994) Estimating error component models with general MA(q) disturbances, *Econometric Theory* **10**, 396–408.

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

[9] Box, G. & G. Jenkins (1970) *Time series analysis: Forecasting adn control*, Holden Day. Inc.

[10] Brillinger, D. R. (1981) *Time Series: Data Analysis and Theory*, Holden Day. Inc.

[11] Cameron, A. C. & P. K. Trivedi (2005) *Microeconometrics, Methods and Applications*, Cambridge University Press.

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

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

[14] 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.

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

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

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

[18] Kakizawa, Y. (1999) Notes on the asymptotic efficiency of sample covariances in Gaussian vector stationary process, *Journal of Time Series Analysis* **20**(5), 551–558.

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

[20] Lee, Y. (2010a) Bias in dynamic panels under time series misspecification, mimeo.

[21] Lee, Y. (2010b) Model selection in the presence of incidental parameters, mimeo.

[22] Magnus, J., & H. Neudecker (1988) *Matrix Differential Calculus.* New York: Wiley.

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

[24] Okui, R. (2010) Asymptotically Unbiased Estimation of Autocovariances and Autocorrelations with Long Panel Data *Econometric Theory* **26** 1263-1304.

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

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

[27] Schott, R.(2006) *Matrix Analysis for Statistics.* New York: Wiley.

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

[29] Van der Vaart, A. W. (1998) *Asymptotic Statistics* Cambridge University Press, New York.

[30] Van der Vaart, A. W. & Wellner, J. A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics* Springer Verlag New York.