Estimation of the Long-run Average Relationship in Nonstationary Panel Time Series*

Yixiao Sun

Department of Economics
University of California, San Diego

First Version: May 2002
This Version: September 2003

*I am grateful to Bruce Hansen, Peter Phillips, Zhijie Xiao, and three anonymous referees for constructive comments and suggestions. Correspondence to: Yixiao Sun, Department of Economics, 0508, University of California, San Diego, La Jolla, CA 92093-0508, USA. E-mail: yisun@ucsd.edu. Tel: (858) 534-4692. Fax: (858)534-7040.

Abstract

This paper proposes a new class of estimators of the long-run average relationship in nonstationary panel time series. The estimators are based on the long-run average variance estimate using bandwidth equal to T. The new estimators include the pooled least squares estimator and the fixed effects estimator as special cases. It is shown that the new estimators are consistent and asymptotically normal under both the sequential limit, wherein $T \to \infty$ followed by $n \to \infty$, and the joint limit where $T, n \to \infty$ simultaneously. The rate condition for the joint limit to hold is relaxed to $\sqrt{n}/T \to 0$, which is less restrictive than the rate condition $n/T \to 0$, as imposed by Phillips and Moon (1999). By exponentiating existing kernels, this paper introduces a new approach to generating kernels and shows that these exponentiated kernels can deliver more efficient estimates of the long-run average coefficient.

JEL Classification: C32; C33

Keywords: Long-run average relationship, long-run variance matrix, multidimensional limits, panel spurious regression, panel cointegration, exponentiated kernels.

1 Introduction

Nonstationary panel data with large cross section (n) and time series dimension (T) have attracted much attention in recent years (e.g. Pedroni 1995; Kao 1999; Phillips and Moon; 1999). Financial and macroeconomic panel data sets that cover many firms, regions or countries over a relatively long time period are familiar examples. Such panels have been used to study growth and convergence, the Feldstein-Horioka puzzle, purchasing power parity, among others. Phillips and Moon (2000) and Baltagi and Kao (2000) provided recent surveys of this rapidly growing research area. When both n and T are large, we can allow the parameters in the data generating process to be different across different individuals, which is not possible in traditional panels. Such a panel data structure also enables us to define an interesting long run average relationship for both panel spurious models and panel cointegration models. Phillips and Moon (1999) showed that both the pooled least squares (PLS) regression and the fixed effects (FE) regression provide consistent estimates of this long-run average relationship.

In this paper, we propose a new class of estimators of the long-run average relationship. Our estimators are motivated from the definition of the long-run average relationship. As shown by Phillips and Moon (1999), the long-run average relationship can be parametrized in terms of the matrix regression coefficient derived from the cross sectional average of the long-run variance (LRV) matrices. A natural way to estimate this coefficient is to first estimate the LRV matrices directly and then use these matrices to construct an estimate of the coefficient. This leads to our LRV-based estimators of the long-run average relationship. In this paper, we use kernel estimators of the LRV matrices (e.g. White 1980; Newey and West 1987; Andrews 1991; Hansen 1992; de Jong and Davidson 2000). The new estimator thus depends on the kernel used to construct the LRV matrices.

We show that the new estimator converges to the long-run average relationship under the sequential limit, in which $T \to \infty$ followed by $n \to \infty$. To develop a joint limit theory, in which T and n go to infinity simultaneously, we need to exercise some control over the relative rate that T and n diverge to infinity. The rate condition is required to eliminate the effect of the bias. For example, Phillips and Moon (1999) imposed the rate condition $n/T \to 0$ in order to establish the joint limit of the PLS and FE estimators. This rate condition is likely to hold when n is moderate and T is large. However, in many financial panels, the number of firms (n) is either of the same magnitude as the time series dimension (T) or far greater. To relax the rate condition, we need an LRV estimator that achieves the greatest bias reduction. It turns out that the kernel LRV estimator with the bandwidth equal to the time series dimension fits our purpose. We show that the bias of this particular estimator is of order $O(1/T)$, which is the best obtainable rate in the nonparametric estimation of the LRV matrix. On the other hand, the variance of this estimator does not vanish. Therefore, such an estimator is necessarily inconsistent, reflecting the usual bias-variance trade-off.

Using a kernel LRV estimator with full bandwidth (the bandwidth is set equal to the time series dimension), we show that the new estimator is consistent and asymptotically normal as n and T go to infinity simultaneously such that $\sqrt{n}/T \to 0$. This rate condition is obviously less restrictive than the rate condition $n/T \to 0$. The so-derived joint limit theory therefore allows for a possibly wide cross section relative to the time series dimension.

We show that the PLS and FE estimators are special cases of the LRV-based estimator. These two estimators implicitly use kernel LRV estimates with full bandwidth. The underlying kernels are $K(s,t) = 1 - \max(s,t)$ and $K(s,t) = \min(s,t) - st$, respectively. As a consequence, our joint limit theory is also applicable to these two estimators. Hence, our work reveals that the rate condition $n/T \to 0$ is only sufficient, but not necessary for the joint limit theory and that it can be weakened to $\sqrt{n}/T \to 0$.

The new estimator is consistent under both the sequential limit and the joint limit, even though the LRV estimator is inconsistent. The reason is that the LRV estimator is proportional to the true LRV matrix up to an additive noise term. If the noise is assumed to be independent, then by averaging across independent individuals, we can recover a matrix that is proportional to the long-run average variance matrix. The consistency of the new estimator follows from the fact that it is not affected by the proportional factor.

We find that the new estimators with exponentiated kernels are more efficient than the PLS and FE estimators. The exponentiated kernels are obtained by taking powers of the popular Bartlett and Parzen kernels. In fact, the asymptotic variance of the new estimator can be made as small as possible by choosing a large exponent. This is not surprising as a larger exponent leads to LRV estimates with less variability. Variance reduction usually comes at the cost of bias inflation. We show that the bias inflation is small when T is large. In addition, for exponentiated Parzen kernels, the bias inflation occurs only to the second dominating bias term but not to the first dominating bias term. Therefore, the bias inflation is likely to factor in only when T is too small.

The kernel LRV estimator with full bandwidth has been used in hypothesis testing by Kiefer and Vogelsang (2002a, 2002b). Our paper provides another instance that the kernel LRV estimator with full bandwidth is useful. Other papers that investigated the new LRV estimator include Jansson (2002), Sun (2002), and Phillips, Sun and Jin (2003a, 2003b). In particular, the latter two papers considered consistent long-run variance estimation using exponentiated kernels.

The use of the LRV matrix to estimate the long-run average relationship has been explored by Makela (2002). He followed the traditional approach to construct the long-run variance matrix. His estimator therefore depends on the truncation lag and is not fully operational. In contrast, our estimator, like the PLS and FE estimators, does not involve the choice of any additional parameter and seems to be appealing to empirical analysts.

The rest of the paper is organized as follows: Section 2 describes the basic model, lays out the assumptions and introduces the new estimator. Section 3 establishes the asymptotic properties of the kernel LRV estimator when the bandwidth is equal to the sample size. Section 4 considers the spurious panel model and investigates the asymptotic properties of the LRV-based estimator. Section 5 extends the results to the cointegration case. Section 6 concludes. Proofs are collected in the Appendix.

Throughout the paper, $\mathrm{vec}(\cdot)$ is the column-by-column vectorization function, $\mathrm{tr}(\cdot)$ is the trace function, and $\otimes$ is the tensor (or Kronecker) product. K_{mm} denotes the $m^2 \times m^2$ commutation matrix that transforms $\mathrm{vec}(A)$ into $\mathrm{vec}(A')$, i.e. $K_{mm} = \sum_{i=1}^{m} \sum_{j=1}^{m} e_i e_j' \otimes e_j e_i'$, where e_i is the unit vector (e.g. Magnus and Neudecker, 1979). For a matrix $A = (a_{ij})$, $\|A\|$ is the Euclidean norm $(tr(A'A))^{1/2}$ and $|A|$ is the matrix $(|a_{ij}|)$. "$A < \infty$" means all the elements of matrix A are finite. The symbol "$\Rightarrow$" signifies weak convergence, ":=" is definitional equivalence, "$\equiv$" signifies equivalence in distribution. For a matrix Z_n, "$Z_n \Rightarrow N(0, \Sigma)$" means "$\mathrm{vec}(Z_n) \Rightarrow N(0, \Sigma)$". M is a generic constant.

2 Model and Estimator

This section introduces notation, specifies the data generating process, defines the estimator and relates it to the existing ones.

2.1 The Model

The model we consider is the same as that in Phillips and Moon (1999). For completeness, we briefly describe the data generating process. The panel data model is based on the vector integrated process:

$$Z_{i,t} = Z_{i,t-1} + U_{i,t}, \ t = 1, ..., T; i = 1, ..., n \tag{2.1}$$

with common initialization $Z_{i,0} = 0$ for all i. The zero initialization is maintained for simplicity. All the results in the paper hold if we assume

$$Z_{i,0} \text{ is iid across } i \text{ with } E||Z_{i,0}||^4 < \infty. \tag{2.2}$$

We partitioned the m-vectors $Z_{i,t}$ and $U_{i,t}$ into m_y and m_x components ($m = m_x + m_y$) as $Z_{i,t}' = (Y_{i,t}', X_{i,t}')$ and $U_{i,t}' = (U_{y_i,t}', U_{x_i,t}')$. The error term $U_{i,t}$ is assumed to be generated by the random coefficient linear process

$$U_{i,t} = \sum_{s=0}^{\infty} C_{i,s} V_{i,t-s}, \tag{2.3}$$

where: (i) $\{C_{i,t}\}$ is a double sequence of $m \times m$ random matrices across i and t; (ii) the m-vectors $V_{i,t}$ are iid across i and t with $EV_{i,t} = 0$, $EV_{i,t}V_{i,t}' = I_m$ and $EV_{a,i,t}^4 = v^4$ for all i and t, where $V_{a,i,t}$ is the a-th element of $V_{i,t}$. (iii) $C_{i,s}$ and $V_{j,t}$ are independent for all i, j, s, t.

Let $C_{a,i,s}$ be the a-th element of $\mathrm{vec}(C_{i,s})$ and $\sigma_{kas} = EC_{a,i,s}^k$. We make two further assumptions on the random coefficients.

Assumption 1 (Random Coefficient Condition): $C_{i,s}$ is iid across i for all s.

Assumption 2 (Summability Condition): $\sum_{s=0}^{\infty} s^4 \left(\sigma_{4as}\right)^{1/4} < \infty$.

Assumptions 1 and 2 are the same as Assumptions 1(i) and 2(ii) of Phillips and Moon (1999). Note that their Assumptions 1(ii) and 2(i) are both implied by their Assumption 2(ii), so there is no need to state their Assumptions 1(ii) and 2(i) here. Assumption 1 and the assumption that $V_{i,t}$ is iid imply cross sectional independence, an assumption that may be restrictive for some economic applications. However, because of the lack of natural ordering, there is no completely satisfactory and general way of modelling cross-sectional dependence, although some important progresses have been made, see Conley (1999), Phillips and Sul (2002) and Andrews (2003). In this paper, we follow the large panel data literature and maintain the assumption of cross sectional independence.

Let $C_i(1) = \sum_{s=0}^{\infty} C_{i,s}$, $\tilde{C}_{i,s} = \sum_{t=s+1}^{\infty} C_{i,t}$ and $\tilde{U}_{i,t} = \sum_{s=0}^{\infty} \tilde{C}_{i,s} V_{i,t-s}$. Under Assumptions 1 and 2, we can prove the following Lemma, which ensures the integrability of the terms that appear frequently in our development.

Lemma 1 *Let Assumptions 1 and 2 hold, then*

(a) $\sum_{s=0}^{\infty} s^2 E \left\| C_{i,s} \right\| < \infty$,

(b) $E \left\| U_{i,t} \right\|^2 < M$ *for some* $M < \infty$ *and all* t,

(c) $E \left\| C_i(1) \right\|^4 < \infty$,

(d) $E||\tilde{U}_{i,t}||^4 < M$ *for some* $M < \infty$ *and all* t,

(e) $\sum_{s=0}^{\infty} [E(||\tilde{C}_{i,s}||^4)]^{1/4} < \infty$.

Under Assumptions 1 and 2, the processes $U_{i,t}$ admit the following BN decomposition almost surely:

$$U_{i,t} = C_i(1) V_{i,t} + \tilde{U}_{i,t-1} - \tilde{U}_{i,t}. \tag{2.4}$$

Using this decomposition and following Phillips and Solo (1992), we can prove that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} U_{i,t} \overset{c}{\Rightarrow} C_i(1) W_i(r), \text{ as } T \to \infty \text{ for all } i, \tag{2.5}$$

where $W_i(r)$ is a standard Brownian Motion with $\text{var}(W_i(r)) = r I_m$ and '$\overset{c}{\Rightarrow}$' signifies the weak convergence conditional on $\mathcal{F}_{c_i} = \sigma(C_{i,0}, ..., C_{i,t}, ...)$, the sigma field generated by the sequence $\{C_{i,t}\}_{t=0}^{\infty}$. Let $S_T(r)$ be the partial sum process, then a formal definition of the above conditional weak convergence is that

$$\lim_{T \to \infty} E\left(h(S_T) | \mathcal{F}_{c_i}\right) = E\left(h\left(C_i(1) W_i\right) | \mathcal{F}_{c_i}\right) \tag{2.6}$$

for all continuous and bounded functionals on $D[0,1]$.

2.2 Definition and Estimation of Long-run Average Relationship

Let Ω_i be the long-run variance matrix of $Z_{i,t}$ conditional on $\mathcal{F}_{c_i}$. It is well known that Ω_i is proportional to the conditional spectral density matrix $f_{U_iU_i}(\lambda)$ of $U_{i,t}$ evaluated at the origin, i.e. $\Omega_i = 2\pi f_{U_iU_i}(0)$. Partitioning Ω_i conformably, we have

$$\Omega_i = \begin{pmatrix} \Omega_{yyi} & \Omega_{yxi} \\ \Omega_{xyi} & \Omega_{xxi} \end{pmatrix}. \tag{2.7}$$

By Lemma 1(c), Ω_i is integrable and

$$\Omega = E\Omega_i = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix}, \tag{2.8}$$

which is called the long-run average variance matrix of $Z_{i,t}$. Following a classical regression approach, we can analogously define a long-run regression coefficient between Y and X by $\beta = \Omega_{yx}\Omega_{xx}^{-1}$. For more discussion on this analogy, see Phillips and Moon (2000).

To construct an estimate of β, we first estimate Ω_i as follows:

$$\widehat{\Omega}_i = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} U_{i,s} K(\frac{s}{T}, \frac{t}{T}) U'_{i,t}, \tag{2.9}$$

where $U_{i,t} = Z_{i,t} - Z_{i,t-1}$, $K(\cdot,\cdot)$ is a kernel function. When $K(x,y)$ depends only on $x-y$, i.e. $K(x,y)$ is translation invariant, we write $K(x,y) = k(x-y)$. In this case, $\widehat{\Omega}_i$ reduces to

$$\widehat{\Omega}_i = \sum_{j=-T+1}^{T-1} k(\frac{j}{T})\widehat{\Gamma}_i(j), \tag{2.10}$$

$$\widehat{\Gamma}_i(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} U_{i,t+j}U'_{i,t} & \text{for } j \geq 0, \\ \frac{1}{T}\sum_{t=-j+1}^{T} U_{i,t+j}U'_{i,t} & \text{for } j < 0. \end{cases} \tag{2.11}$$

From the above formulation, it is clear that $\widehat{\Omega}_i$ is the usual kernel LRV estimator using the full bandwidth. It should be noted that translation invariant kernels are commonly used in the estimation of the LRV matrix. We consider the kernels other than the translation invariant ones in order to include some existing estimators of the long-run average relationship as special cases. This will be made clear in Section 2.3.

Based on the above estimate, we can estimate Ω by

$$\widehat{\Omega} = \begin{pmatrix} \widehat{\Omega}_{yy} & \widehat{\Omega}_{yx} \\ \widehat{\Omega}_{xy} & \widehat{\Omega}_{xx} \end{pmatrix} = n^{-1}\sum_{i=1}^{n}\widehat{\Omega}_i. \tag{2.12}$$

The long-run average relationship parameter β can then be estimated by

$$\widehat{\beta}_{LRV} = \widehat{\Omega}_{yx}\widehat{\Omega}_{xx}^{-1}, \tag{2.13}$$

which is called the LRV-based estimator.

Note that the LRV-based estimator $\widehat{\beta}_{LRV}$ depends on the observations $Z_{i,t}$ only through their first order difference. Therefore, when the model contains individual effects such that

$$
\begin{aligned}
Z_{i,t} &= A_{i,0} + Z_{i,t}^{0} &(2.14)\\
Z_{i,t}^{0} &= Z_{i,t-1}^{0} + U_{i,t}, &(2.15)
\end{aligned}
$$

where $Z_{i,0}^{0} = 0$, and $U_{i,t}$ follows the linear process defined in (2.3), the LRV-based estimator $\widehat{\beta}_{LRV}$ can be computed exactly the same as before. In other words, the LRV-based estimator is robust to the presence of the individual effects.

2.3 Relationship between New and Existing Estimators

Phillips and Moon (1999) showed that both PLS and FE estimators are consistent and asymptotically normal. In this subsection, we examine the relationships between the LRV-based estimator and the PLS and FE estimators.

The PLS estimator is

$$\widetilde{\beta}_{PLS} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{i,t}X_{i,t}'\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}X_{i,t}'\right)^{-1}. \tag{2.16}$$

Some simple algebraic manipulations show that

$$
\begin{aligned}
\widetilde{\beta}_{PLS} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{PLS,T}(\frac{s}{T},\frac{t}{T})U_{y_i,s}U_{x_i,t}'\right)\\
&\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{PLS,T}(\frac{s}{T},\frac{t}{T})U_{x_i,s}U_{x_i,t}'\right)^{-1}, &(2.17)
\end{aligned}
$$

where

$$
\begin{aligned}
K_{PLS,T}(\frac{s}{T},\frac{t}{T}) &= 1-\frac{(s+1)\vee(t+1)}{T} \text{ and}\\
(s+1)\vee(t+1) &= \max(s+1,t+1). &(2.18)
\end{aligned}
$$

Hence, the PLS estimator is a special case of the LRV-based estimator. Note that the kernel for the PLS estimator depends on T. If we replace $K_{PLS,T}(s,t)$ by $K_{PLS}(s,t) = 1-(s\vee t)$, then we get an asymptotically equivalent estimator $\widehat{\beta}_{PLS}$. In view of (2.9), we see that $\widehat{\beta}_{PLS}$ is an LRV-based estimator with kernel $K(s,t) = 1-(s\vee t)$.

We now consider the FE estimator, viz.

$$\widetilde{\beta}_{FE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{i,t}-\bar{Y}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\right)^{-1}, \tag{2.19}$$

where $\bar{Y}_{i,\cdot} = 1/T \sum_{t=1}^{T} Y_{i,t}$ and $\bar{X}_{i,\cdot} = 1/T \sum_{t=1}^{T} X_{i,t}$. Again, some algebraic manipulations yield

$$\begin{aligned}\widetilde{\beta}_{FE} &= \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} K_{FE,T} \left(\frac{s}{T}, \frac{t}{T} \right) U_{y_i,s} U'_{x_i,t} \right) \\ &\quad \times \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} K_{FE,T} \left(\frac{s}{T}, \frac{t}{T} \right) U_{x_i,s} U'_{x_i,t} \right)^{-1}, \end{aligned} \tag{2.20}$$

where

$$K_{FE,T} \left(\frac{s}{T}, \frac{t}{T} \right) = \frac{T - (s \vee t) + 1}{T} - \left(\frac{T - s + 1}{T} \right) \left(\frac{T - t + 1}{T} \right). \tag{2.21}$$

The kernel function $K_{FE,T}(s,t)$ depends on T. As before, we can replace $K_{FE,T}(s,t)$ by $K_{FE}(s,t) = \min(s,t) - st$ to obtain an estimator $\widehat{\beta}_{FE}$ that is asymptotically equivalent to $\widetilde{\beta}_{FE}$. The resulting estimator $\widehat{\beta}_{FE}$ is an LRV-based estimator with kernel $K(s,t) = \min(s,t) - st$.

In summary, the existing estimators or their asymptotically equivalent forms are special cases of the LRV-based estimator. The underlying LRV estimators use kernels that are not translation invariant. This sharply contrasts with the usual LRV estimators where translation invariant kernels are commonly used.

3 Asymptotic Properties of the New LRV Estimator

The properties of $\widehat{\beta}_{LRV}$ evidently depend on those of the long-run variance matrix estimator $\widehat{\Omega}_i$. In this section, we consider the asymptotic properties of $\widehat{\Omega}_i$. We first examine the bias and variance of $\widehat{\Omega}_i$ for fixed T and then establish its asymptotic distribution.

The bias of $\widehat{\Omega}_i$ depends on the smoothness of $f_{U_iU_i}(\lambda)$ at zero and the properties of the kernel function. Following Parzen (1957), Hannan (1970), and Andrews (1991), we define

$$f_{U_iU_i}^{(q)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_i(j). \tag{3.22}$$

The smoothness of the spectral density at zero is indexed by q for which $f_{U_iU_i}^{(q)}$ is finite almost surely. The larger is q such that $f_{U_iU_i}^{(q)} < \infty$ a.s., the smoother is the spectral density at zero.

The following lemma establishes the smoothness of the spectral density at $\lambda = 0$.

Lemma 2 *Let Assumptions 1 and 2 hold, then*

(a) $E \sum_{j=-\infty}^{\infty} j^2 \|\Gamma_i(j)\| = \sum_{j=-\infty}^{\infty} j^2 E \|\Gamma_i(j)\| < \infty.$

(b) $E(2\pi f_{U_iU_i}^{(2)}) = E \sum_{j=-\infty}^{\infty} j^2 \Gamma_i(j) < \infty.$

When $K(s,t) = k(s-t)$, the bias of Ω_i depends on the smoothness of $k(x)$ at zero. To define the degree of smoothness, we let

$$k_q = \lim_{x \to 0} \frac{1-k(x)}{|x|^q} < \infty \text{ for } q \geq 0. \tag{3.23}$$

The largest q for which k_q is finite is defined to be the Parzen characteristic exponent q^*. The smoother is $k(x)$ at zero, the larger is q^*. The values of q^* for various kernels can be found in Andrews (1991).

To investigate the asymptotic properties of $\widehat{\Omega}_i$, we assume the kernel function $K(s,t)$ satisfies the following conditions.

Assumption 3 (Kernel Conditions) $K(s,t) \in \mathcal{K}_1 \cup \mathcal{K}_2$ where

$$\mathcal{K}_1 = \{K(s,t) : K(s,t) = 1 - (s \vee t), \text{ or } \min(s,t) - st\}$$

and $\mathcal{K}_2 = \{K(s,t) : K(s,t) = k(s-t)$ and

(i) $k(x) : [-1,1] \to [0,1]$ is symmetric and piecewise smooth with $k(0) = 1$.

(ii) The Parzen characteristic exponent satisfies $q^* \geq 1$.

(iii) $k(x)$ is positive semi-definite, i.e., for any square integrable function $f(x)$, $\int_0^1 \int_0^1 k(s-t) f(s) f(t) ds dt \geq 0\}$

Note that the two kernels in $\mathcal{K}_1$ are positive semi-definite. When $K(s,t) = 1 - (s \vee t)$,

$$\int_0^1 \int_0^1 K(s,t) f(s) f(t) ds dt = \int_0^1 \left(\int_0^t f(s) ds \right)^2 ds \geq 0. \tag{3.24}$$

When $K(s,t) = \min(s,t) - st$,

$$\int_0^1 \int_0^1 K(s,t) f(s) f(t) ds dt = \int_0^1 F^2(s) ds - \left(\int_0^1 F(s) ds \right)^2 \geq 0, \tag{3.25}$$

where $F(s) = \int_0^s f(r) dr$. Therefore, the kernels satisfying Assumption 3 are positive semi-definite. As shown by Newey and West (1987) and Andrews (1991), the positive semi-definiteness guarantees the positive semi-definiteness of $\widehat{\Omega}_i$.

We proceed to investigate the bias and variance of $\widehat{\Omega}_i$. The following two lemmas establish the limiting behaviors of the bias and variance of $\widehat{\Omega}_i$ as $T \to \infty$.

Lemma 3 *Let Assumptions* $1-3$ *hold. Define* $\mu = \int_0^1 K(s,s) ds$.

(a) If $K(s,t)$ *is translation invariant with* $q^* = 1$, *then*

$$\lim_{T \to \infty} TE \left[E \left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu \Omega_i \right] = -2\pi (k_1 + 1) E f^{(1)}_{U_i U_i}. \tag{3.26}$$

(b) If $K(s,t)$ *is translation invariant with* $q^* \geq 2$, *then*

$$\lim_{T \to \infty} TE \left[E \left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu \Omega_i \right] = -2\pi E f^{(1)}_{U_i U_i}. \tag{3.27}$$

(c) If $K(s,t) \in \mathcal{K}_1$, *then* $E \left(E \left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu \Omega_i \right) = O(1/T)$.

Remarks: (i) When $K(s,t)$ is translation invariant, $K(s,s)=1$, so $\mu=1$. In this case, Lemmas 3(a) and (b) show that $\widehat{\Omega}_i$ is centered around a matrix that is equal to the true long-run variance matrix up to a small additive error. The error has a finite expectation and is independent across i. As a consequence, the average long-run variance matrix can be estimated by averaging $\widehat{\Omega}_i$ over $i=1,2,...,n$. When $K(s,t)\in\mathcal{K}_1$, $\widehat{\Omega}_i$, scaled by $\int_0^1 K(s,s)ds$, is equal to the true variance matrix plus a noise term. The average long-run variance matrix can be estimated by averaging $(\int_0^1 K(s,s)ds)^{-1}\widehat{\Omega}_i$ over $i=1,2,...,n$.

(ii) For the conventional LRV estimator with a truncation parameter S_T, the bias is of order $O\left(1/S_T^{q*}\right)$ under the assumption that $S_T/T+S_T^{q*}/T+1/S_T\rightarrow 0$ (e.g. Hannan 1970; Andrews 1991). The bias of the conventional estimator is thus of a larger order than the estimator without truncation. This is not surprising as truncation is used in the conventional estimator to reduce the variance at the cost of the bias inflation.

(iii) When $K(s,t)$ is translation invariant, the dominating bias term depends on the kernel through k_1 if $q^*=1$. In contrast, when $q^*\geq 2$, the dominating bias term does not depend on the kernel. From the proof of the Lemma, we see that when $q^*=2$, the next dominating bias term is $-2\pi T^{-2}k_2 Ef_{U_iU_i}^{(2)}$. Therefore, when $q^*\geq 2$, the kernels exert their bias effects only through high order terms. This has profound implications for the asymptotic bias of $\widehat{\beta}_{LRV}$ considered in section 4.2.

Lemma 4 *Let Assumptions* $1-3$ *hold. Then we have:*

(a) $lim_{T\rightarrow\infty} var(vec(\widehat{\Omega}_i-\widetilde{\Omega}_i))=0$, *where*

$$\widetilde{\Omega}_i=T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\left(C_i(1)V_{i,t}\right)K\left(\frac{t}{T},\frac{\tau}{T}\right)\left(C_i(1)V_{i,\tau}\right)'; \tag{3.28}$$

(b) $\lim_{T\rightarrow\infty} var(vec(\widehat{\Omega}_i))=\mu^2 var(vec(\Omega_i))+\delta^2\left(I_{m^2}+K_{mm}\right)E\left(\Omega_i\otimes\Omega_i\right)$, *where*

$$\delta^2=\int_0^1\int_0^1 K^2(r,s)drds. \tag{3.29}$$

Remarks: (i) Lemma 4(b) gives the expression for the unconditional variance. It is easy to see from the proof that the conditional variance has a limit given by $\lim_{T\rightarrow\infty}\text{var}(\text{vec}(\widehat{\Omega}_i)|\mathcal{F}_{c_i})=\delta^2\left(I_{m^2}+K_{mm}\right)\left(\Omega_i\otimes\Omega_i\right)$. Therefore, the magnitude of the asymptotic variance depends on δ^2. This suggests using the kernel that has the smallest δ^2 value when the variance of $\widehat{\Omega}_i$ is the main concern.

(ii) Lemma 4(b) calculates the limit of the finite sample variance of $\widehat{f}_{U_iU_i}(\lambda)$ when $\lambda=0$. Following the same procedure and using a frequency domain BN decomposition, we can calculate the limit of the finite sample variance of $\widehat{f}_{U_iU_i}(\lambda)$ for other values of λ when the full bandwidth is used in smoothing. This extension may be needed to investigate seasonally integrated processes. This extension is straightforward but tedious and is beyond the scope of this paper.

Lemma 5 *Let Assumptions* 1 –3 *hold. Then*

(a) Conditional on $\mathcal{F}_{c_i}$, $\widehat{\Omega}_i \Rightarrow C_i(1)\Xi_i C_i'(1)$*;*

(b) $E(C_i(1)\Xi_i C_i(1)'|\mathcal{F}_{c_i}) = \mu\Omega_i$, *where*

$$\Xi_i = \int_0^1 \int_0^1 K(r,s) dW_i(r) dW_i'(s). \tag{3.30}$$

Remarks: (i) When $K(s,t)$ is translation invariant, $\mu = 1$. In this case, Lemma 5 shows that $\widehat{\Omega}_i$ is asymptotically unbiased, even though it is inconsistent. For other kernels, $\widehat{\Omega}_i$ is asymptotically proportional to the true LRV matrix. We will show that the consistency of $\widehat{\beta}_{LRV}$ inherits from this asymptotic proportionality.

(ii) Kiefer and Vogelsang (2002a, 2002b) established asymptotic results similar to Lemma 5(a) under different assumptions. Specifically, they assumed the kernels were continuously differentiable to the second order. As a consequence, they had to treat the Bartlett kernel separately. They obtained different representations of the asymptotic distributions for these two cases. The unified representation in Lemma 5 is very valuable. It helps us shorten the proof and enables us to prove the asymptotic properties of $\widehat{\beta}_{LRV}$ in a coherent way.

(iii) When $K(r,s) \in \mathcal{K}_1$, the limiting distribution in Lemma 5(a) is the same as that obtained by using (2.5) and the continuous mapping theorem.

4 Panel Spurious Regression

This section considers the case where the two component random vectors $Y_{i,t}$ and $X_{i,t}$ of $Z_{i,t}$ have no cointegrating relation for any i. This case is characterized by the following assumption:

Assumption 4 (Rank Condition) $\text{rank}(\Omega_i) = m$ almost surely for all $i = 1, ..., n$.

Define $\beta_i = \Omega_{yxi} (\Omega_{xxi})^{-1}$. Assumption 4 implies that

$$Y_{i,t} \overset{a.s.}{=} \beta_i X_{i,t} + W_{i,t} \tag{4.1}$$

where $W_{i,t}$ is a unit root process and the long run covariance between $X_{i,t}$ and $W_{i,t}$ is zero, i.e. $\sum_{j=-\infty}^{\infty} E\Delta W_{i,t-j}\Delta X_{i,t}' = 0$. Our interest lies in the long run average coefficient $\beta = E\Omega_{yxi} (E\Omega_{xxi})^{-1}$, which is in general different from the 'average long run coefficient' defined by $E\beta_i$. For more discussions on this, see Phillips and Moon (1999).

Before investigating the asymptotic properties of the LRV-based estimate, we first define some notation. The sequential approach adopted in the paper is to fix n and allow T to pass to infinity, giving an intermediate limit, then by letting n pass to infinity subsequently to obtain the sequential limit. As in Phillips and Moon (1999), we write the sequential limit of this type as $(T, n \to \infty)_{seq}$. The joint approach adopted in the paper allows both indexes, n and T, to pass to infinity simultaneously. We write the joint limit of this type as $(T, n \to \infty)$.

4.1 Sequential Limit Theory and Joint Limit Theory

The following theorem establishes the consistency of $\widehat{\beta}_{LRV}$ as either $(T, n \to \infty)_{seq}$ or $(T, n \to \infty)$.

Theorem 6 *Let Assumptions 1–4 hold, then*

(a) $\widehat{\Omega}_{xx} \to_p \mu\Omega_{xx}$,

(b) $\widehat{\Omega}_{yx} \to_p \mu\Omega_{yx}$,

(c) $\widehat{\beta}_{LRV} \to_p \beta$,

as either $(T, n \to \infty)_{seq}$ *or* $(T, n \to \infty)$.

REMARK: $\widehat{\beta}_{LRV}$ is consistent even though $\widehat{\Omega}_i$ is inconsistent. This is not surprising as $\widehat{\Omega}_i$ equals $\mu\Omega_i$ plus a noise term. Although the noise in the time series estimation is strong, we can weaken the strong effect of noise by averaging across independent individuals. This is reflected in Lemma 6(a) and (b), which shows that $\widehat{\Omega}_{xx}$ and $\widehat{\Omega}_{yx}$ are respective consistent estimates of Ω_{xx} and Ω_{yx} up to a multiplicative scalar.

Now we proceed to investigate the asymptotic distribution of $\widehat{\beta}_{LRV}$. We consider the sequential asymptotics first and then extend the result to the joint asymptotics. In order to get a definite joint limit, we need to control the relative rate of expansion of the two indexes. Write $\sqrt{n}(\widehat{\beta}_{LRV} - \beta) = \sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx})\widehat{\Omega}_{xx}^{-1}$. Theorem 6 describes the asymptotic behavior of $\widehat{\Omega}_{xx}$ under the sequential and joint limits. Under Assumption 4, Ω_{xx} has full rank, which implies that $\widehat{\Omega}_{xx}^{-1}$ converge to $\mu^{-1}\Omega_{xx}^{-1}$. Therefore, it suffices to consider the limiting distribution of $\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx})$.

Under the sequential limit, we first let $T \to \infty$ for fixed n. The intermediate limit is

$$\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx}) \Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_i, \tag{4.2}$$

where

$$Q_i = C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1), \tag{4.3}$$

$C_{yi}(1)$ is the $m_y \times m$ matrix consisting of the first m_y rows of $C_i(1)$, and $C_{xi}(1)$ is the $m_x \times m$ matrix consisting of the last m_x rows of $C_i(1)$. In view of Lemma 5, the mean of the summand is

$$E(Q_i) = \mu(E\Omega_{yxi} - \beta E\Omega_{xxi}) = \mu\left(\Omega_{yx} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xx}\right) = 0,$$

and the covariance matrix Θ is $E\text{vec}(Q_i)\text{vec}(Q_i)'$. An explicit expression for Θ is established in the following lemma.

Lemma 7 *Let Assumptions 1–4 hold. Then* Θ *is equal to*

$$\begin{aligned}
&\mu^2 E vec(\Omega_{yxi} - \beta\Omega_{xxi})\, vec(\Omega_{yxi} - \beta\Omega_{xxi})' \\
&+\delta^2 E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\
&+\delta^2\left(E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})\right) K_{m_y m_x},
\end{aligned}$$

where $K_{m_y m_x}$ is the $m_y m_x \times m_y m_x$ commutation matrix.

The sequence of random matrices $C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1)$ is iid $(0, \Theta)$ across i. From the multivariate Linderberg-Levy theorem, we then get, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1) \right) \Rightarrow N(0, \Theta). \tag{4.4}$$

Combining (4.4) with the limit $\lim \widehat{\Omega}_{xx}^{-1} = \mu^{-1}\Omega_{xx}^{-1}$, we establish the sequential limit in the following theorem.

Theorem 8 *Let Assumptions 1 – 4 hold. Then, as $(T, n \to \infty)_{seq}$,*

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1} \otimes I_{m_y})), \tag{4.5}$$

where Θ_{LRV} is

$$\begin{aligned} & E\, vec\,(\Omega_{yxi} - \beta\Omega_{xxi})\, vec\,(\Omega_{yxi} - \beta\Omega_{xxi})' \\ & + \mu^{-2}\delta^2 E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\ & + \mu^{-2}\delta^2 \left(E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})\right) K_{m_y m_x}. \end{aligned} \tag{4.6}$$

We now show that the limiting distribution continues to hold in the joint asymptotics as $(T, n \to \infty)$. Write $\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx})$ as

$$\begin{aligned} \sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{i,T} + b_{nT}, \end{aligned} \tag{4.7}$$

where

$$Q_{i,T} = \widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi} - E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}) \tag{4.8}$$

and

$$b_{nT} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}). \tag{4.9}$$

Because of Lemma 3, the term b_{nT} vanishes under the sequential limit. However, under the joint limit, we need to exercise some control over the relative expansion rate of (T, n) so that b_{nT} vanishes as $(T, n \to \infty)$. When this occurs, the term $1/\sqrt{n}\sum_{i=1}^{n} Q_{i,T}$ will deliver the asymptotic distribution as $(T, n \to \infty)$.

Using Lemma 3, we have

$$\begin{aligned} b_{nT} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left(E\left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi})|\mathcal{F}_{c_i}\right)\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[E\left(\Omega_{yxi} - \beta\Omega_{xxi}\right) + O(1/T)\right] = O\left(\sqrt{n}/T\right), \end{aligned} \tag{4.10}$$

because the $O(\cdot)$ terms in the summand are independent across i. Therefore, in order to eliminate the asymptotic bias, we need to assume the two indexes pass to infinity in such a way that $\sqrt{n}/T \to 0$. Under this condition, we can prove the following theorem, which provides the asymptotic distribution under the joint limit.

Theorem 9 *Let Assumptions 1 – 4 hold. Then, as* $(T, n \to \infty)$ *such that* $\sqrt{n}/T \to 0$,

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1} \otimes I_{m_y})). \tag{4.11}$$

REMARKS: (i) For the PLS estimator, $K(r,s) = 1 - (r \vee s)$. Therefore, $\mu^2 = (\int_0^1 (1-s)ds)^2 = 1/4$, $\delta^2 = \int_0^1 \int_0^1 K^2(r,s)drds = 1/6$, and $\mu^{-2}\delta^2 = 2/3$. Hence, the PLS estimator satisfies, under both the sequential and joint limits,

$$\sqrt{n}(\widehat{\beta}_{PLS} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{PLS}(\Omega_{xx}^{-1} \otimes I_{m_y})) \tag{4.12}$$

with

$$\begin{aligned} \Theta_{PLS} &= E\text{vec}\,(\Omega_{yxi} - \beta\Omega_{xxi})\,\text{vec}\,(\Omega_{yxi} - \beta\Omega_{xxi})' \\ &+2/3E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\ &+2/3E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})K_{m_y m_x}. \end{aligned} \tag{4.13}$$

The above limiting distribution is identical to that obtained by Phillips and Moon (1999).

(ii) For the FE estimator, $K(s,t) = \min(s,t) - st$. In this case, it is easy to see that $\mu^2 = 1/36$ and $\delta^2 = 1/90$. So $\mu^{-2}\delta^2 = 2/5$. Hence $\widehat{\beta}_{FE}$ has the limiting distribution given in (4.12) and (4.13) but with 2/3 replaced by 2/5. Once again, the asymptotic result is consistent with Phillips and Moon (1999).

(iii) The efficiency of $\widehat{\beta}_{LRV}$ depends only on $\mu^{-2}\delta^2$. The smaller $\mu^{-2}\delta^2$ is, the more efficient the estimator is. This is because the sum of the last two terms in (4.6) is

$$E\left(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))\right) \times (I_{m^2} + K_{mm})\left(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))\right)',$$

which is positive semi-definite. Therefore, $\widehat{\beta}_{FE}$ is more efficient than $\widehat{\beta}_{PLS}$. But $\widehat{\beta}_{FE}$ is less efficient than $\widehat{\beta}_{LRV}$ if $\kappa = (\int_0^1 K(s,s)ds)^{-2} \int_0^1 \int_0^1 K^2(r,s)drds < 2/5$. In the next subsection, we consider a class of new kernels that have smaller κ values.

If we assume that $C_{i,t}$ are the same across individuals, then $\Omega_i = \Omega$ and $\beta_i = \beta$ for some β and all i. In this case, $\Omega_{yxi} - \beta\Omega_{xxi} = 0$. As a consequence, Θ_{LRV} reduces to

$$\mu^{-2}\delta^2\left(\Omega_{xx} \otimes \left(\Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}\right)\right),$$

and we obtain the following corollary.

Corollary 10 *Let Assumptions 1 – 4 hold. If* $C_{i,t} =_{a.s} C_t$ *where* C_t *is an* $m \times m$ *nonrandom matrix for all* t, *then, as* $(T, n \to \infty)_{seq}$, *or as* $(T, n \to \infty)$ *with* $\sqrt{n}/T \to 0$,

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, \mu^{-2}\delta^2\left(\Omega_{xx}^{-1} \otimes \left(\Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}\right)\right)). \tag{4.15}$$

REMARKS: (i) The corollary generalizes a result of Kao (1999). He considered the homogeneous spurious regression and showed that under the sequential limit, the fixed effects estimator satisfies (4.15) with $\mu^{-2}\delta^2 = 2/5$.

(ii) Note that the matrix $\Omega_{xx}^{-1} \otimes \left(\Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}\right)$ is positive semi-definite. Therefore, the efficiency of $\widehat{\beta}_{LRV}$ depends only on $\mu^{-2}\delta^2$ regardless of whether $C_{i,t}$ is heterogeneous or not.

4.2 LRV-Based Estimator with Exponentiated Kernels

In this subsection, we exponentiate some commonly-used kernels and investigate the asymptotic properties of the LRV-based estimators that these exponentiated kernels deliver.

We first consider the sharp kernels defined by $k(x) = k_{Bart}^{\rho}(x)$, where $k_{Bart}(\cdot)$ is the Bartlett kernel and $\rho \in \mathbb{Z}^+$. These kernels, as so defined, exhibit a sharp peak at the origin. Sharp kernels are positive semi-definite, as they are equal to the products of the positive semi-definite kernels. To see this, we may use equation (7.11) in the appendix and represent the Bartlett kernel by

$$k_{Bart}(r-s) = \sum_{m=1}^{\infty} \frac{1}{\lambda_m} f_m(r) f_m(s), \text{ for } (r,s) \in [0,1]^2. \tag{4.16}$$

Then

$$k_{Bart}^2(r-s) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{1}{\lambda_n \lambda_m} f_n(r) f_m(r) f_n(s) f_m(s). \tag{4.17}$$

So, for any function $g(x) \in L^2[0,1]$, we have

$$\int_0^1 \int_0^1 g(r) k_{Bart}^2(r-s) g(s) dr ds = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{1}{\lambda_n \lambda_m} \left(\int_0^1 g(r) f_n(r) f_m(r) dr \right)^2 \geq 0,$$

which implies that $k_{Bart}^2(r-s)$ is indeed positive semi-definite. Iterating the above procedure leads to the positive semi-definiteness of $k_{Bart}^{\rho}(r-s)$ for any $\rho \in \mathbb{Z}^+$.

For sharp kernels, the Parzen characteristic exponent is $q^* = 1$ and $k_1 = \rho$. The value of κ is $\kappa = 1/(\rho+1)$. Therefore, κ is a decreasing function of the exponent ρ. In principle, we can choose ρ to make κ as small as possible. However, the finite sample performance can be hurt when ρ is too large for a moderate time series dimension. This is because the bias of $\widehat{\Omega}_i$ increases as ρ increases, as shown by Lemma 3. In fact, when $\sqrt{n}/T \to \alpha$, the asymptotic distribution of $\sqrt{n}\left(\widehat{\beta}_{LRV} - \beta\right)$ under the joint limit is

$$N\left(b,\ (\Omega_{xx}^{-1} \otimes I_{m_y}) \Theta_{PLS} (\Omega_{xx}^{-1} \otimes I_{m_y})\right), \tag{4.19}$$

where $b = -2\pi\alpha\left(\rho+1\right)(\Omega_{xx}^{-1} \otimes I_{m_y})\text{vec}(Ef_{U_{y_i}U_{x_i}}^{(1)} - \beta Ef_{U_{x_i}U_{x_i}}^{(1)})$. Therefore, the squared asymptotic bias $b'b$ is increasing in ρ while the asymptotic variance is decreasing in ρ. This observation implies that there exists an optimal ρ that minimizes the mean squared errors. The optimal ρ depends on the ratio α and the average spectral density of U_i. We can estimate the optimal ρ along the lines of Andrews (1991), but we do not pursue this analysis in the present paper.

Next, we consider the steep kernels defined by $k(x) = (k_{PR}(x))^{\rho}$ where $k_{PR}(x)$ is the Parzen kernel. These kernels decay to zero as x approaches 1. The speed of decay depends on ρ. The larger ρ is, the faster the decay and the steeper the kernel. Steep kernels are positive semi-definite because the Parzen kernel is positive semi-definite. The difference between the sharp kernels and the steep kernels is that the former are not differentiable at the origin while the latter are. For steep kernels, the Parzen characteristic exponent is $q^* = 2$ and $k_2 = 6\rho$. The value of κ can be calculated using numerical integration. They are given in Table 1 for $\rho = 1, ..., 6$. Obviously, κ decreases as ρ increases. This is expected because $(k_{PR}(x))^{\rho_1} \leq (k_{PR}(x))^{\rho_2}$ if $\rho_1 \geq \rho_2$. Therefore, the steep kernel can deliver an LRV-based estimator $\widehat{\beta}_{LRV}$ that is more efficient than $\widehat{\beta}_{FE}$, as long as the exponent is greater than 1 (see Table 1).

When the steep kernel is employed, the dominating bias of $\widehat{\Omega}_i$ is independent of the exponent. If $(n, T \to \infty)$ such that $\sqrt{n}/T \to \alpha$, then the asymptotic distribution of $\sqrt{n}(\widehat{\beta}_{LRV} - \beta)$ is

$$N\left(b,\ (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1} \otimes I_{m_y}\right), \tag{4.20}$$

where $b = -2\pi\alpha(\Omega_{xx}^{-1} \otimes I_{m_y})\text{vec}(Ef^{(1)}_{U_{y_i}U_{x_i}} - \beta Ef^{(1)}_{U_{x_i}U_{x_i}})$. This limiting distribution seems to imply that we can choose ρ to make κ as small as possible without inflating the asymptotic bias. This is true in large samples. But in finite samples, a large κ may lead to a poor performance. The reason is that the second dominating bias term in $\widehat{\Omega}_i$ is $T^{-2}2\pi k_2 Ef^{(2)}_{U_iU_i}$, which depends on k_2. As a consequence, the asymptotic bias of $\widehat{\beta}_{LRV}$ under the joint limit is

$$-2\pi\sqrt{n}/T(\Omega_{xx}^{-1} \otimes I_{m_y})vec(Ef^{(1)}_{U_{y_i}U_{x_i}} - \beta Ef^{(1)}_{U_{x_i}U_{x_i}}) + O\left(k_2\sqrt{n}/T^2\right).$$

The $O(\cdot)$ term vanishes when $(n, T \to \infty)$ such that $\sqrt{n}/T \to \alpha$. But in finite samples, the $O(\cdot)$ term may have an adverse effect on the performance of $\widehat{\beta}_{LRV}$. Nevertheless, the effect is expected to be small, especially when T is large.

Finally, we may take powers of the kernels in $\mathcal{K}_1$ and obtain more efficient estimates. Although Assumption 3 does not cover exponentiated kernels of this sort. Theorems 8 and 9 go through without modification.

Table 1 summarizes the values of κ for different exponentiated kernels. The table clearly shows that for a given 'mother' kernel, the value of κ decreases as the exponent increases. Recall that the smaller κ is, the more efficient the LRV-based estimator is. We can thus conclude that a larger exponent (ρ) gives rise to a more efficient estimator.

Table 1: The Values of κ for Some Kernels*

	$\rho=1$	$\rho=2$	$\rho=3$	$\rho=4$	$\rho=5$	$\rho=6$
$k^{\rho}_{Bartlett}$	0.5000	0.3333	0.2500	0.2000	0.1666	0.1429
k^{ρ}_{Parzen}	0.4473	0.3359	0.2806	0.2459	0.2216	0.2033
$(1-r\vee s)^{\rho}$	0.6666	0.6000	0.5714	0.5556	0.5455	0.5385
$(\min(r,s)-rs)^{\rho}$	0.4000	0.2857	0.2331	0.2016	0.1800	0.1642

* $\kappa = (\int_0^1 K(s,s)ds)^{-2}(\int_0^1\int_0^1 K^2(r,s)drds)$

5 Heterogeneous Panel Cointegration

This section assumes that the variables in $Z_{i,t}$ are cointegrated. As discussed in Engle and Granger (1987), the long run covariance matrix is singular in this case. We consider the case that the cointegration relationships are different for different individuals.

Following Phillips and Moon (1999), we strengthen the summability condition and impose additional conditions.

Assumption 5 (Summability Conditions')
(i) $\sum_{s=0}^{\infty} s^4 (\sigma_{4as})^{1/4} < \infty$.
(ii) $\sum_{s=0}^{\infty} s^2 (\sigma_{8as})^{1/8} < \infty$. (iii) $\sum_{s=0}^{\infty} (\sigma_{16as})^{1/16} < \infty$.

Assumption 6 (Rank Conditions): $\text{rank}(\Omega_i) = \text{rank}(\Omega_{xxi}) = m_x$ and $\text{rank}(\Omega_{yyi}) = m_y$ almost surely for all $i = 1, ..., n$.

Assumption 7 (Tail Conditions): The random matrix Ω_{xxi} has continuous density function f with
(i) $f(\Omega) = O(\exp\{tr(-c\Omega\})$ for some $c > 0$ when $\text{tr}(\Omega) \to \infty$.
(ii) $f(\Omega) = O((\det(\Omega)^{\gamma}))$ from some $\gamma > 7$ when $\det(\Omega) \to \infty$.

Note that Assumption 5 is stronger than Assumption 2. Therefore, under Assumptions 1, 3 and 5, all results in Section 3 continue to hold. Let $\alpha_i = (I_{m_y}, -\beta_i)$, where $\beta_i = \Omega_{yxi}\Omega_{xxi}^{-1}$. Assumption 6 implies that $\alpha_i C_i(1) C'_{y_i}(1) = 0$. As a consequence, $\alpha_i C_i(1) = 0$, i.e. $C_{yi}(1) = \beta_i C_{xi}(1)$. Define $E_{i,t} = \alpha_i Z_{i,t} = Y_{i,t} - \beta_i X_{i,t}$. Then, using $\alpha_i C_i(1) = 0$, we have:

$$E_{i,t} = \alpha_i \sum_{s=1}^{t} \left(C_i(1) V_{i,s} + \tilde{U}_{i,s-1} - \tilde{U}_{i,s} \right) = \alpha_i \tilde{U}_{i,0} - \alpha_i \tilde{U}_{i,t}.$$

Therefore, Assumption 6 implies the existence of the following panel cointegration relationship with probability one:

$$\begin{aligned} Y_{i,t} &\overset{a.s.}{=} \beta_i X_{i,t} + E_{i,t} \\ X_{i,t} &= X_{i,t-1} + U_{x_i t} \end{aligned} \quad (5.1)$$

where

$$F_{i,t}=\left(\begin{array}{c}E_{i,t}\\ U_{x_it}\end{array}\right)=\sum_{s=0}^{\infty}G_{i,s}V_{i,t-s},\ G_{i,s}=\left(\begin{array}{c}-\alpha_i\tilde{D}_{i,s}\\ C_{xi,s}\end{array}\right), \tag{5.2}$$

and

$$\tilde{D}_{i,s}=\begin{cases}-\alpha_i\tilde{C}_{i,s} & \text{if } s<t,\\ -\alpha_i(\tilde{C}_{i,s}-\tilde{C}_{i,s-t}) & \text{if } s\geq t.\end{cases} \tag{5.3}$$

Let $G_i(1)=\sum_{s=0}^{\infty}G_{i,s}$, $\tilde{G}_{i,s}=\sum_{t=s+1}^{\infty}G_{i,t}$ and $\tilde{F}_{i,t}=\sum_{s=0}^{\infty}\tilde{G}_{i,s}V_{i,t-s}$. As shown by Phillips and Moon (1999), Assumptions 5 and 7 ensure that quantities analogous to those in Lemma 1 are bounded. Specifically, $E\sum_{s=0}^{\infty}s^2\left\|G_{i,s}\right\|^2$, $E\left\|F_{i,t}\right\|^2$, $E\left\|G_i(1)\right\|^4$, $E||\tilde{F}_{i,t}||^4$ and $\sum_{s=0}^{\infty}(E||\tilde{G}_{i,s}||^4)^{1/4}$ are all bounded.

Using the long run covariance matrix, we can estimate the individual cointegration relationship by $\widehat{\beta}_i=\widehat{\Omega}_{yxi}\widehat{\Omega}_{xxi}^{-1}$. It follows from Lemma 5 that

$$\widehat{\Omega}_{yxi}\Rightarrow\beta_iC_{x_i}(1)\Xi_iC_{x_i}'(1),\ \widehat{\Omega}_{xxi}\Rightarrow C_{x_i}(1)\Xi_iC_{x_i}'(1). \tag{5.4}$$

As a consequence, $\widehat{\beta}_i\Rightarrow\beta_i$, which implies that $\widehat{\beta}_i\rightarrow_p\beta_i$. This is because β_i is a constant conditional on $\mathcal{F}_{c_i}$.

The following theorem establishes the rate of convergence of $\widehat{\beta}_i$. Before stating the theorem, we define the Lipschitz continuity. A function $f(\cdot)\colon\Gamma\rightarrow\mathbb{R}$ is Lipschitz continuous if there exists a constant $M>0$ such that $\|f(x)-f(y)\|\leq M\|x-y\|$ for all x and y in Γ. It is easy to see that the kernels satisfying Assumption 3 are Lipschitz continuous.

Lemma 11 *Let Assumptions 5-7 hold. Assume that the kernel function $K(\cdot,\cdot)$ is symmetric and Lipschitz continuous. Then*

(a) $T(\widehat{\beta}_i-\beta_i)=O_p(1)$ if $K(1,r)=0$ for almost all r.

(b) $\sqrt{T}(\widehat{\beta}_i-\beta_i)=O_p(1)$ if $K(1,r)\neq 0$ for some r in a set with positive Lesbegue measure.

Remarks: (i) The Lemma shows that $\widehat{\beta}_i$ is not only consistent but also converges to the true value at the rate of $\sqrt{T}$ or T. This result is particularly interesting. Although both $\widehat{\Omega}_{yxi}$ and $\widehat{\Omega}_{xxi}$ are inconsistent, the linear combination $\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{yxi}$ is consistent, reflecting the singularity of the long run covariance matrix Ω_i. In fact, the proof of the lemma shows that $\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{yxi}=O_p(1/\sqrt{T})$ or $O_p(1/T)$, depending on the kernel used.

(ii) The kernel $K(\cdot,\cdot)$ may be called a 'tied down' kernel if $K(1,s)=K(r,1)=0$ for any r and s. Since both kernels in $\mathcal{K}_1$ are 'tied down' kernels, $\widehat{\beta}_i$ converges to β at the rate of T if $K\in\mathcal{K}_1$. This is of course a well-known result. Lemma 11(a) has more implications. Given any kernel function $K(r,s)$, we can construct a new kernel $K^*(r,s)=K(r,s)-K(1,s)-K(r,1)+K(1,1)$ such that $K^*(1,s)=K^*(r,1)=0$ for any r and s. The new kernel is then able to deliver an estimator that is super-consistent.

(iii) For translation invariance kernels, $K(1,r) = k(1-r) \neq 0$ in general. So the estimator that they deliver is only $\sqrt{T}$-consistent. The difference in the rate of convergence arises because the dominated terms are different for different types of kernels.

We now investigate the asymptotic distribution of $\widehat{\beta}_{LRV}$ in the heterogeneous panel cointegration model. We first consider the sequential limit of $\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx})$. The intermediate limit for large T is the same as that given by (4.2). More explicitly,

$$\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx}) \Rightarrow \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1)\right).$$

Following exactly the same arguments, we can show that the summands are iid $(0,\Theta)$. Invoking the multivariate Linderberg-Levy Theorem and using the consistency of $\widehat{\Omega}_{xx}$, we have, as $(T,n\to\infty)_{seq}$

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1}\otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1}\otimes I_{m_y})). \tag{5.5}$$

The next theorem shows that the asymptotic distribution is applicable to the case of joint limit. The proof of the theorem follows steps similar to that of Theorem 9 and is omitted.

Theorem 12 *Suppose Assumptions 1-3 and 6 hold. Then, as* $(T,n\to\infty)_{seq}$, *or as* $(T,n\to\infty)$ *with* $\sqrt{n}/T \to 0$,

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1}\otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1}\otimes I_{m_y})). \tag{5.6}$$

Remarks: (i) Note that Assumption 7 is not needed for the theorem to hold. The strong summability conditions in Assumption 5 are also not necessary. The asymptotic distribution not only has precisely the same form as in the spurious regression case, but also holds under the same conditions. However, Assumptions 5 and 7 are required for Lemma 11, as it replies on the panel BN decomposition of the error term $E_{i,t}$.

(ii) Since the limiting distribution is the same as that in Theorem 9, the remarks given there and the efficiency analyses presented in Section 4.2 remain valid. Therefore, in the presence of heterogeneity, the LRV-based estimator is more efficient than the PLS and FE estimators if exponentiated kernels are used.

(iii) The asymptotic theory developed above allows us to test hypotheses about the long-run average coefficient β. To test the null hypothesis $H_0 : \psi(\beta) = 0$, where $\psi(\cdot)$ is a p-vector of smooth function on a subset $\mathbb{R}^{m_y\times m_x}$ such that $\partial\psi/\partial\beta'$ has full rank p $(\leq m_y m_x)$, we construct the Wald statistic: $W_\psi = n\psi(\widehat{\beta}_{LRV})\widehat{V}_\psi^{-1}\psi(\widehat{\beta}_{LRV})$, where

$$\widehat{V}_\psi = \partial\psi(\widehat{\beta}_{LRV})/\partial\beta'\widehat{V}_\beta^{-1}\partial\psi(\widehat{\beta}_{LRV})/\partial\beta \tag{5.7}$$

$$\widehat{V}_\beta = \left(\widehat{\Omega}_{xx}^{-1}\otimes I_{m_y}\right)\widehat{\Theta}_{LRV}\left(\widehat{\Omega}_{xx}^{-1}\otimes I_{m_y}\right) \tag{5.8}$$

and $\widehat{\Theta}_{LRV}$ is the sample analogue of (4.6). Some simple manipulations show that this test statistic converges to a χ_p^2 random variable under both the sequential and joint limits.

6 Conclusion

In this paper, we have proposed an LRV-based estimator of the long-run average relationship. Our estimator includes the pooled least squares and fixed-effects estimators as special cases. We show that the LRV-based estimator is consistent and asymptotically normal under both the sequential limit and the joint limit. The joint limit is derived under the rate condition $\sqrt{n}/T \to 0$, which is less restrictive than the rate condition $n/T \to 0$, as required by Phillips and Moon (1999). A central result is that, using exponentiated kernels introduced in this paper, the LRV-based estimator is asymptotically more efficient than the existing ones.

It should be pointed out that we have not considered the homogeneous panel cointegration model. When the long run relations are the same across individuals, the LRV-based estimator may have a slower rate of convergence than the PLS and FE estimators. We have shown that, when translation invariant kernels are used, $\widehat{\beta}_i$ is only $\sqrt{T}$ consistent. Because of the slower rate of convergence, we expect that the LRV-based estimator converges at the rate of $\sqrt{nT}$ in homogeneous panel cointegration models. The $\sqrt{nT}$-rate is slower than the $\sqrt{n}T$-rate that is attained by the PLS and FE estimators. However, the $\sqrt{n}T$-rate can be restored if 'tied down' kernels are used. The efficiency of the LRV-based estimator with other 'tied down' kernels is an open question.

This paper can be extended in several directions. First, the power parameter ρ for the sharp and steep kernels is fixed in the paper. We may extend the results to the case that ρ grows to infinity at a suitable rate with N and T along the lines of Phillips, Sun and Jin (2003a, 2003b). Second, the LRV-based estimator can be employed in implementing residual-based tests for cointegration in panel data. Following the lines of Kao (1999), we can use the LRV-based estimator to construct the residuals and test for unit roots in the residuals. Since the LRV-based estimator is more efficient than the fixed effects estimator employed by Kao (1999), the test using the LRV-based residuals may have better power properties. Finally, we generate the new kernels by exponentiating existing ones. An alternative approach to generating kernels is to start from a mother kernel k and consider the class $\{k_b(s,t)\} = \left\{k(b^{-1}r, b^{-1}s) : b \in (0,1]\right\}$ (Kiefer and Vogelsang 2003). For this approach, Theorems 8, 9, and 12 go through but with μ and δ^2 defined by

$$\mu = \int_0^1 k(b^{-1}r, b^{-1}r)dr \text{ and } \delta^2 = \int_0^1 \int_0^1 k^2(b^{-1}r, b^{-1}s)drds. \tag{6.9}$$

With the above extension, we may analyze the efficiency of the LRV-based estimators for different values of b. In general, the efficiency will not be monotonic in $b \in (0,1]$.

7 Appendix of Proofs

Proof of Lemma 1. Parts (a)–(d) are the same as Lemma 1 of Phillips and Moon (1999). It remains to prove part (e). From Lemma 9(a) of Phillips and Moon (1999), for any $\rho \geq 1$ and any $p \times q$ matrix $A = (a_{ij})$, we have

$$||A||^{\rho} \leq M \sum_{i=1}^{p} \sum_{j=1}^{q} |a_{ij}|^{\rho} \tag{7.1}$$

for some constant M. Therefore, to evaluate the order of $\sum_{s=0}^{\infty}[E(\left\|\tilde{C}_{i,s}\right\|^4)]^{1/4}$, it suffices to consider $\sum_{p=0}^{\infty}[E(\tilde{C}_{a,i,p}^4)]^{1/4}$. By the generalized Minkowski inequality and the Cauchy inequality, we have, for some constant M,

$$\begin{aligned}
& \sum_{p=0}^{\infty} \left[E\left(\tilde{C}_{a,i,p}^4\right)\right]^{1/4} \\
= & \sum_{p=0}^{\infty} \left[E\left(\sum_{t=p+1}^{\infty} C_{a,i,t}\right)^4\right]^{1/4} \leq \sum_{p=0}^{\infty}\sum_{t=p+1}^{\infty} \left[E\left(C_{a,i,t}^4\right)\right]^{1/4} \\
= & \sum_{p=0}^{\infty}\sum_{t=p+1}^{\infty} \left(\sigma_{4it}^{1/8} t^2\right)\left(\sigma_{4it}^{1/8} t^{-2}\right) \leq \sum_{p=0}^{\infty}\left(\sum_{t=p+1}^{\infty}\sigma_{4it}^{1/4}t^4\right)^{1/2}\left(\sum_{t=p+1}^{\infty}\sigma_{4it}^{1/4}t^{-4}\right)^{1/2} \\
\leq & M\sum_{p=0}^{\infty}\left(\sum_{t=p+1}^{\infty}\sigma_{4it}^{1/4}t^4\right)\left(\frac{1}{(p+1)^{3/2}}\right) \leq M\left(\sum_{t=0}^{\infty}\sigma_{4it}^{1/4}t^4\right)\left(\sum_{p=1}^{\infty}\frac{1}{p^{3/2}}\right) \\
< & \infty
\end{aligned} \tag{7.2}$$

where the last line follows from Assumption 2. This completes the proof of the Lemma. ■

Proof of Lemma 2. Since part (b) follows from part (a), it suffices to prove part (a). Write $E\sum_{j=0}^{\infty} j^2 \left\|\Gamma_i(j)\right\|$ as

$$\begin{aligned}
& E\sum_{j=0}^{\infty} j^2 \left\|E\left(U_{i,t+j}U_{i,t}'|\mathcal{F}_{c_i}\right)\right\| = E\sum_{j=0}^{\infty} j^2 \left\|E\left(\sum_{p,q=0}^{\infty} C_{i,q}V_{i,t+j-q}V_{i,t-p}'C_{i,p}'|\mathcal{F}_{c_i}\right)\right\| \\
= & E\sum_{j=0}^{\infty} j^2 \left\|E\left(\sum_{p=0}^{\infty}\sum_{k=-j}^{\infty} C_{i,j+k}V_{i,t-k}V_{i,t-p}'C_{i,p}'|\mathcal{F}_{c_i}\right)\right\| = E\sum_{j=0}^{\infty} j^2 \left\|\sum_{p=0}^{\infty} C_{i,j+p}C_{i,p}'\right\| \\
\leq & E\sum_{j=0}^{\infty} j^2 \sum_{p=0}^{\infty}\left\|C_{i,j+p}\right\|\left\|C_{i,p}'\right\| = E\sum_{p=0}^{\infty}\sum_{j=0}^{\infty} j^2 \left\|C_{i,j+p}\right\|\left\|C_{i,p}'\right\| \\
\leq & E\sum_{p=0}^{\infty}\left(\sum_{j=0}^{\infty}(j+p)^2\left\|C_{i,j+p}\right\|\right)\left\|C_{i,p}'\right\| \leq E\sum_{p=0}^{\infty}\left(\sum_{j=0}^{\infty} j^2\left\|C_{i,j}\right\|\right)\left\|C_{i,p}'\right\|
\end{aligned}$$

Therefore, $E\sum_{j=0}^{\infty} j^2 \|\Gamma_i(j)\|$ is bounded by

$$\begin{aligned}
& \sum_{p=0}^{\infty}\sum_{j=0}^{\infty} j^2 E\left\|C_{i,j}\right\| \left\|C_{i,p}'\right\| \\
\leq \; & \sum_{p=0}^{\infty}\sum_{j=0}^{\infty} j^2 \left(E\left\|C_{i,j}\right\|^2\right)^{1/2} E\left(\left\|C_{i,p}'\right\|^2\right)^{1/2} \\
= \; & \sum_{j=0}^{\infty} j^2 \left(E\left\|C_{i,j}\right\|^2\right)^{1/2} \sum_{p=0}^{\infty} E\left(\left\|C_{i,p}\right\|^2\right)^{1/2} < \infty
\end{aligned}$$

where the last line follows from (7.1) and Assumption 2. This completes the proof of part (a). ■

Proof of Lemma 3. We first consider the case that $K(s,t)$ is translation invariant, i.e. $K(s,t) = k(s-t)$. The proof follows closely those of Parzen (1957) and Hannan (1970). We decompose $E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) - \Omega_i$ into three terms as follows:

$$\begin{aligned}
E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) - \Omega_i &= \sum_{j=-T+1}^{T-1} k(\frac{j}{T}) E(\widehat{\Gamma}_i(j)|\mathcal{F}_{c_i}) - \Omega_i \\
&= \sum_{j=-T+1}^{T-1} k(\frac{j}{T})(1-\frac{|j|}{T})\Gamma_i(j) - \sum_{j=-\infty}^{\infty}\Gamma_i(j) \\
&= \sum_{j=-T+1}^{T-1}\left(k(\frac{j}{T})-1\right)\Gamma_i(j) - \sum_{j=-T+1}^{T-1} k(\frac{j}{T})\frac{|j|}{T}\Gamma_i(j) - \sum_{|j|\geq T}\Gamma_i(j) \\
&= \Omega_{i1}^e + \Omega_{i2}^e + \Omega_{i3}^e, \text{ say.}
\end{aligned}$$

We consider the expectations of the three terms in turn. First, for $q = \min(q^*, 2)$, $E\Omega_{i1}^e$ is,

$$\begin{aligned}
& T^{-q} E \sum_{j=-T+1}^{T-1}\left(\frac{k(j/T)-1}{|j/T|^q}\right)|j|^q\,\Gamma_i(j) = T^{-q}\sum_{j=-T+1}^{T-1}\left(\frac{k(j/T)-1}{|j/T|^q}\right)|j|^q\,E\Gamma_i(j) \\
= \; & T^{-q}\sum_{j=-\infty}^{\infty} 1\left\{-T+1 \leq j \leq T-1\right\}\left(\left|\frac{k(j/T)-1}{|j/T|^q}\right|\right)|j|^q\,E\Gamma_i(j) \\
= \; & -T^{-q}k_q\left(\sum_{j=-\infty}^{\infty}|j|^q\,E\Gamma_i(j)\right)(1+o(1)).
\end{aligned}$$

The last inequality follows because $(k(j/T)-1)\,|j/T|^{-q}$ converges boundedly to k_q for each fixed j.

Second, $E\Omega_{i2}^e$ is

$$-\sum_{j=-T+1}^{T-1} k(\frac{j}{T})\frac{|j|}{T}E\Gamma_i(j) = -T^{-1}\sum_{j=-\infty}^{\infty}|j|E\Gamma_i(j)(1+o(1))$$

using Lemma 2.

Finally, $\|E\Omega_{i3}^e\|$ is bounded by

$$\left\| \sum_{|j|\geq T} E\Gamma_i(j) \right\| \leq T^{-2} \sum_{|j|\geq T} |j|^2 \, E \left\| \Gamma_i(j) \right\| = o(T^{-2}). \tag{7.3}$$

Let $\Omega_i^e = \left(\Omega_{i1}^e + \Omega_{i2}^e + \Omega_{i3}^e\right)$, then we have shown that, when $q^* = 1$, $\lim_{T\to\infty} TE\Omega_i^e = -2\pi\left(k_1+1\right) Ef_{U_iU_i}^{(1)}$ and when $q^* \geq 2$, $\lim_{T\to\infty} TE\Omega_i^e = -2\pi Ef_{U_iU_i}^{(1)}$.

Next, we consider the case that $K \in \mathcal{K}_1$. Some algebraic manipulations show that

$$E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} K(\frac{s}{T},\frac{t}{T})\Gamma_i(s-t)$$

$$= \frac{1}{T}\sum_{s=1}^{T} K(\frac{s}{T},\frac{s}{T})\Gamma_i(0) + \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{s-1} K(\frac{s}{T},\frac{t}{T})\left[\Gamma_i(s-t)+\Gamma_i(t-s)\right]$$

$$= \frac{1}{T}\sum_{s=1}^{T} K(\frac{s}{T},\frac{s}{T})\Gamma_i(0) + \frac{1}{T}\sum_{j=1}^{T}\sum_{s=j+1}^{T} K(\frac{s}{T},\frac{s-j}{T})\left[\Gamma_i(j)+\Gamma_i(-j)\right].$$

When $K(s,t) = 1-(s\vee t)$,

$$\begin{aligned} \frac{1}{T}\sum_{s=1}^{T} K(\frac{s}{T},\frac{s}{T}) &= \frac{1}{2}\frac{T-1}{T}, \\ \frac{1}{T}\sum_{s=j+1}^{T} K(\frac{s}{T},\frac{s-j}{T}) &= \frac{1}{2}\frac{T^2-2jT+j^2+j-T}{T^2}. \end{aligned}$$

Combining the above calculation with the steps for the translation invariant case, we can get $E(E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) - 1/2\Omega_i) = O(1/T)$. Similarly, we can show that when $K(s,t) = \min(s,t) - st$,

$$\begin{aligned} \frac{1}{T}\sum_{s=1}^{T} K(\frac{s}{T},\frac{s}{T}) &= \frac{1}{T}\sum_{s=1}^{T}(\frac{s}{T}-\frac{s^2}{T^2}) = \frac{1}{6}\frac{\left(T^2-1\right)}{T^2} \\ \frac{1}{T}\sum_{s=j+1}^{T} K(\frac{s}{T},\frac{s-j}{T}) &= \frac{1}{6}\frac{j-T-3jT^2+T^3-j^3+3j^2T}{T^3}. \end{aligned}$$

and $E(E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) - 1/6\Omega_i) = O(1/T)$.

The proof of the theorem is completed by noting that $\int_0^1 k(0)ds = 1$, $\int_0^1 \left(1-(s\vee s)\right)ds = 1/2$ and $\int_0^1(\min(s,s)-s^2)ds = 1/6$. ■

Proof of Lemma 4. Plugging the BN decomposition

$$U_{i,t} = C_i(1)V_{i,t} + \tilde{U}_{i,t-1} - \tilde{U}_{i,t} \tag{7.4}$$

into

$$\widehat{\Omega}_i = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} U_{i,t} K(\frac{t}{T},\frac{\tau}{T}) U_{i,\tau}^{\prime}, \tag{7.5}$$

we get

$$\widehat{\Omega}_i = \widetilde{\Omega}_i + R_i, \tag{7.6}$$

where $R_i = R_{i1} + R_{i2} + R_{i3}$ with

$$\begin{aligned}
R_{i1} &= \frac{1}{T} C_i(1) \sum_{t=1}^{T}\sum_{\tau=1}^{T} V_{i,t} K(\frac{t}{T},\frac{\tau}{T}) \left(\tilde{U}_{i,\tau-1} - \tilde{U}_{i,\tau}\right)^{\prime}, \\
R_{i2} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} \left(\tilde{U}_{i,t-1} - \tilde{U}_{i,t}\right) K(\frac{t}{T},\frac{\tau}{T}) V_{i,\tau}^{\prime} C_i^{\prime}(1) = R_{i1}^{\prime}, \\
R_{i3} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} \left(\tilde{U}_{i,t-1} - \tilde{U}_{i,t}\right) K(\frac{t}{T},\frac{\tau}{T}) \left(\tilde{U}_{i,\tau-1} - \tilde{U}_{i,\tau}\right)^{\prime}.
\end{aligned}$$

We proceed to show that $Etr\left(\text{vec}(R_{i1})\text{vec}(R_{i1})^{\prime}\right) = o(1)$. It is easy to see that R_{i1} is

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} C_i(1) V_{i,t} K(\frac{t}{T},\frac{1}{T}) \tilde{U}_{i,0}^{\prime} - \frac{1}{T}\sum_{t=1}^{T} C_i(1) V_{i,t} K(\frac{t}{T},\frac{T}{T}) \tilde{U}_{iT}^{\prime} \\
& + \frac{1}{T}\sum_{t=1}^{T} C_i(1) V_{i,t} \sum_{\tau=1}^{T-1} \left(K(\frac{t}{T},\frac{\tau+1}{T}) - K(\frac{t}{T},\frac{\tau}{T}) \right) \tilde{U}_{i,\tau}^{\prime} \qquad (7.7) \\
: & \equiv R_{i1}^{(1)} + R_{i1}^{(2)} + R_{i1}^{(3)}, \text{ say.}
\end{aligned}$$

But $Etr(\text{vec}(R_{i1}^{(1)})\text{vec}(R_{i1}^{(1)})^{\prime})$ is

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K(\frac{t}{T},\frac{1}{T}) K(\frac{s}{T},\frac{1}{T}) Etr\left(\text{vec}\left(C_i(1) V_{i,t} \tilde{U}_{i,0}^{\prime}\right) \text{vec}\left(C_i(1) V_{i,s} \tilde{U}_{i,0}^{\prime}\right)^{\prime}\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K(\frac{t}{T},\frac{1}{T}) K(\frac{s}{T},\frac{1}{T}) Etr\left(\left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,t}\right)\left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,s}\right)^{\prime}\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K(\frac{t}{T},\frac{1}{T}) K(\frac{s}{T},\frac{1}{T}) Etr\left(\left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,t}\right)\left(\tilde{U}_{i,0}^{\prime} \otimes V_{i,s}^{\prime} C_i^{\prime}(1)\right)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K(\frac{t}{T},\frac{1}{T}) K(\frac{s}{T},\frac{1}{T}) Etr\left(\tilde{U}_{i,0} \tilde{U}_{i,0}^{\prime} \otimes C_i(1) V_{i,t} V_{i,s}^{\prime} C_i^{\prime}(1)\right) \qquad (7.8)
\end{aligned}$$

where the first equality follows from the fact that for $m \times 1$ vectors A and B, $\text{vec}(AB^{\prime}) = B \otimes A$, the third equality follow from the rule that $(A \otimes B)(C \otimes D) = AC \otimes BD$. In view of the fact that $\text{tr}(C \otimes D) = tr(C) tr(D)$, we write $Etr(\text{vec}(R_{i1}^{(1)})\text{vec}(R_{i1}^{(1)})')$ as

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{1}{T})Etr\left(\tilde{U}_{i,0}\tilde{U}_{i,0}' \otimes C_i(1)C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{1}{T})Etr\left(\tilde{U}_{i,0}\tilde{U}_{i,0}'\right) tr\left(C_i(1)C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{1}{T})E\left\|\tilde{U}_{i,0}\right\|^2 \|C_i(1)\|^2 \\
\leq & \frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{1}{T})\left(E\left\|\tilde{U}_{i,0}\right\|^4\right)^{1/2} E\left(\|C_i(1)\|^4\right)^{1/2} \\
= & \frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{1}{T})O(1) = O(\frac{1}{T}), \qquad (7.9)
\end{aligned}$$

where the last two equalities follow from Lemma 1(c) and (d) and the boundedness of $K(\cdot,\cdot)$.

The proofs of $Etr(\text{vec}(R_{i1}^{(2)})\text{vec}(R_{i1}^{(2)})') = o_p(1)$ and $Etr(\text{vec}(R_{i1}^{(3)})\text{vec}(R_{i1}^{(3)})') = o_p(1)$ are rather lengthy. They are given in Sun (2003). The details are omitted here.

Given that $Etr(\text{vec}(R_{i1}^{(k)})\text{vec}(R_{i1}^{(k)}))'$, $k = 1, 2, 3,$ we have $Etr\left(\text{vec}\,(R_{i1})\,\text{vec}\,(R_{i1})'\right) = o(1)$. As a consequence, we also have $Etr\left(\text{vec}\,(R_{i2})\,\text{vec}\,(R_{i2})'\right) = o(1)$. Similarly, we can prove $Etr\left(\text{vec}\,(R_{i3})\,\text{vec}\,(R_{i3})'\right) = o(1)$. Again, details are omitted.

Part (c) From part (b), we deduce immediately that

$$\begin{aligned}
\text{var}(\text{vec}(\widehat{\Omega}_i) &= E\text{vec}(\widehat{\Omega}_i - E\widehat{\Omega}_i)\text{vec}(\widehat{\Omega}_i - E\widehat{\Omega}_i)' \\
&= E\text{vec}(\widetilde{\Omega}_i - E\widetilde{\Omega}_i)\text{vec}(\widetilde{\Omega}_i - E\widetilde{\Omega}_i)' + o(1).
\end{aligned}$$

Note that $E\text{vec}(\widetilde{\Omega}_i)\text{vec}(\widetilde{\Omega}_i)'$ equals

$$\begin{aligned}
& E\frac{1}{T^2}\sum_{t,\tau,p,q=1}^{T} K(\frac{t}{T},\frac{\tau}{T})K(\frac{p}{T},\frac{q}{T})\left(C_i(1)\otimes C_i(1)\right)\left(V_{i,\tau}V_{i,q}'\otimes V_{i,t}V_{i,p}'\right)\left(C_i'(1)\otimes C_i'(1)\right) \\
= & \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K^2(\frac{t}{T},\frac{\tau}{T})\right) E\left(C_i(1)\otimes C_i(1)\right)\left(C_i'(1)\otimes C_i'(1)\right) \\
& + \left(\frac{1}{T}\sum_{t=1}^{T} K(\frac{t}{T},\frac{t}{T})\right)^2 E\text{vec}\left(C_i(1)C_i'(1)\right)\text{vec}\left(C_i(1)C_i'(1)\right) \\
& + \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K^2(\frac{t}{T},\frac{\tau}{T})\right) E\left(C_i(1)\otimes C_i(1)\right)K_{mm}\left(C_i'(1)\otimes C_i'(1)\right) \\
& + \left(\frac{1}{T^2}\sum_{t=1}^{T} K^2(\frac{t}{T},\frac{t}{T})\right)\zeta E\left(C_i(1)\otimes C_i(1)\right)\left(\sum_{l=1}^{m} e_{ll}\otimes e_{ll}\right)\left(C_i'(1)\otimes C_i'(1)\right) \qquad (7.10)
\end{aligned}$$

and

$$
\begin{aligned}
& \left(E\mathrm{vec}(\widetilde{\Omega}_i)\right)\left(E\mathrm{vec}(\widetilde{\Omega}_i)\right)' \\
= & \left(\frac{1}{T}\sum_{t=1}^{T} K(\frac{t}{T},\frac{t}{T})\right)^2 E\mathrm{vec}\left(C_i(1)C_i'(1)\right) E\mathrm{vec}\left(C_i(1)C_i'(1)\right)',
\end{aligned}
$$

so $E\mathrm{vec}(\widehat{\Omega}_i - E\widehat{\Omega}_i)\mathrm{vec}(\widehat{\Omega}_i - E\widehat{\Omega}_i)'$ is

$$
\begin{aligned}
& \left(\frac{1}{T}\sum_{t=1}^{T} K(\frac{t}{T},\frac{t}{T})\right)^2 \mathrm{var}\left(\mathrm{vec}\left(C_i(1)C_i'(1)\right)\right) + \\
& \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K^2(\frac{t}{T},\frac{\tau}{T})\right) E\left(C_i(1)\otimes C_i(1)\right)\left(I_{m^2}+K_{mm}\right)\left(C_i'(1)\otimes C_i'(1)\right) + o(1).
\end{aligned}
$$

Letting $T \to \infty$ completes the proof. ■

Proof of Lemma 5. Part (a) Lemma 3 has shown that $\widehat{\Omega}_i = \widetilde{\Omega}_i + o_p(1)$. To establish the asymptotic distribution of $\widehat{\Omega}_i$, we only need to consider $\widetilde{\Omega}_i$. Since the kernels are assumed to be continuous and positive semi-definite, it follows from Mercer's theorem that $K(r,s)$ can be represented as

$$
K(r,s) = \sum_{m=1}^{\infty} \frac{1}{\lambda_m} f_m(r) f_m(s), \tag{7.11}
$$

where $\lambda_m > 0$ are the eigenvalues of the kernel and $f_m(x)$ are the corresponding eigenfunctions, i.e. $f_m(s) = \lambda_m \int_0^1 K(r,s) f_m(r)dr$, and the right hand side converges uniformly over $(r,s) \in [0,1]\times[0,1]$. In fact, for the two kernels in $\mathcal{K}_1$, we have

$$
\begin{aligned}
\min(r,s) - rs &= \sum_{m=1}^{\infty} \frac{2}{\pi^2 m^2} \sin \pi m r \sin \pi m s, & (7.12) \\
1 - \max(r,s) &= \sum_{m=1}^{\infty} \frac{8}{\pi^2 (2m-1)^2} \cos\frac{\pi(2m-1)r}{2} \cos\frac{\pi(2m-1)s}{2}. & (7.13)
\end{aligned}
$$

For kernels in $\mathcal{K}_2$, we have the Fourier series representation:

$$
k(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos m\pi x, \tag{7.14}
$$

where $a_m = \int_{-1}^{1} k(x)\exp\left(-im\pi x\right)dx$, $\sum_{m=0}^{\infty}|a_m| < \infty$, and the right side of (7.14) converges uniformly over $x \in [-1,1]$. It follows from the above representation that for any $r, s \in [0,1]$,

$$
k(r-s) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos m\pi r \cos m\pi s + \sum_{m=1}^{\infty} a_m \sin m\pi r \sin m\pi s. \tag{7.15}
$$

Hence, under Assumption 3, the kernels can be represented by (7.11) with smooth eigenfunctions.

Using (7.11), we have, for any T,

$$\begin{aligned} K(\frac{t}{T},\frac{\tau}{T}) &= \sum_{m=1}^{\infty}\frac{1}{\lambda_m}f_m(\frac{t}{T})f_m(\frac{\tau}{T}) \\ &= \sum_{m=1}^{M_0}\frac{1}{\lambda_m}f_m(\frac{t}{T})f_m(\frac{\tau}{T})+\sum_{m=M_0+1}^{\infty}\frac{1}{\lambda_m}f_m(\frac{t}{T})f_m(\frac{\tau}{T}). \end{aligned} \tag{7.16}$$

Therefore, $\widetilde{\Omega}_i = C_i(1)\left(\widetilde{\Omega}_{i,1}+\widetilde{\Omega}_{i,2}\right)C_i'(1)$ where

$$\widetilde{\Omega}_{i,1} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}V_{i,t}\sum_{m=1}^{M_0}\frac{1}{\lambda_m}f_m(\frac{t}{T})f_m(\frac{\tau}{T})V_{i,\tau}', \tag{7.17}$$

$$\widetilde{\Omega}_{i,2} = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}V_{i,t}\sum_{m=M_0+1}^{\infty}\frac{1}{\lambda_m}f_m(\frac{t}{T})f_m(\frac{\tau}{T})V_{i,\tau}'. \tag{7.18}$$

It is easy to see that, for a fixed M_0,

$$\begin{aligned} \widetilde{\Omega}_{i,1} &= \sum_{m=1}^{M_0}\frac{1}{\lambda_m}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{i,t}f_m(\frac{t}{T})\right)\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}f_m(\frac{\tau}{T})V_{i,\tau}'\right) \\ &\Rightarrow \sum_{m=1}^{M_0}\frac{1}{\lambda_m}\int_0^1 f_m(r)dW_i(r)\int_0^1 f_m(s)dW_i'(s) \\ &= \int_0^1\int_0^1\left(\sum_{m=1}^{M_0}\frac{1}{\lambda_m}f_m(r)f_m(s)\right)dW_i(r)dW_i'(s). \end{aligned} \tag{7.19}$$

The above weak convergence result follows from integration and summation by parts and the continuous mapping theorem. Note that the integral $\int_0^1 f_m(r)dW_i(r)$ is well defined because $f_m(\cdot)$ is of bounded variation.

Following the same argument as in (7.10), we have, as $M_0\to\infty$,

$$E(\text{vec}(\widetilde{\Omega}_{i2})\text{vec}(\widetilde{\Omega}_{i2})') = o(\frac{1}{T^2})\sum_{t=1}^{T}\sum_{\tau=1}^{T}E\text{vec}\left(V_{i,t}V_{i,\tau}'\right)\text{vec}\left(V_{i,t}V_{i,\tau}'\right)' = o(1), \tag{7.20}$$

which implies that $\widetilde{\Omega}_{i2} = o_p(1)$ for any T as $M_0\to\infty$. Combining the above results (e.g. Nabeya and Tanaka, 1988), we obtain

$$\begin{aligned} \widehat{\Omega}_i &\Rightarrow C_i(1)\int_0^1\int_0^1 K(r,s)dW_i(r)dW_i'(s)C_i'(1) \\ &= C_i(1)\Xi_i C_i'(1). \end{aligned} \tag{7.21}$$

Part (b) The mean of any off-diagonal element of Ξ_i is obviously zero. It suffices to consider the means of the diagonal elements. They are $\int_0^1 K(s,s)ds$. So

$E\Xi_i = \int_0^1 K(s,s)dsI_m$. As a consequence $EC_i(1)\Xi_i C_i'(1) = C_i(1)C_i'(1)\int_0^1 K(s,s)ds = \Omega_i \int_0^1 K(s,s)ds$. ■

Proof of Theorem 6. By Assumption 3, Ω_{xxi} is positive definite almost surely, and $c'\Omega_{xxi}c > 0$ for any $c \neq 0$ in $\mathbb{R}^{m_x}$. Thus $Ec'\Omega_{xxi}c = c'\Omega_{xx}c > 0$, which implies Ω_{xx} is positive definite. Hence Ω_{xx}^{-1} exists, and part (c) follows from parts (a) and (b). It remains to prove parts (a) and (b). We first consider the joint probability limits. To prove $\widehat{\Omega}_{xx} \to_p \mu\Omega_{xx}$ and $\widehat{\Omega}_{yx} \to_p \mu\Omega_{yx}$ as $(T, n \to \infty)$, it is sufficient to show that $\text{plim}_{(T,n\to\infty)} n^{-1}\sum_{i=1}^n \lim \widehat{\Omega}_i = \mu\Omega$. Note that $E(\widehat{\Omega}_i|\mathcal{F}_{c_i}) = \mu\Omega_i + \Omega_i^e$ where $\Omega_i^e = \Omega_{i1}^e + \Omega_{i2}^e + \Omega_{i3}^e$ and Ω_{ik}^e, $k = 1, 2, 3$ are defined in the proof of Lemma 3. We can write $\widehat{\Omega}_i$ as $\widehat{\Omega}_i = \mu\Omega_i + \Omega_i^e + \Omega_i^\varepsilon$, where Ω_i^e is iid across i with $E\Omega_i^e = O(1/T)$ and Ω_i^ε is iid across i with $E\Omega_i^\varepsilon = 0$. Therefore,

$$
\begin{aligned}
\text{plim}_{(T,n\to\infty)} \frac{1}{n}\sum_{i=1}^n \widehat{\Omega}_i &= \text{plim}_{(T,n\to\infty)} \frac{1}{n}\sum_{i=1}^n \left(\mu\Omega_i + \Omega_i^\varepsilon + \Omega_i^e\right) \\
&= \text{plim}_{(T,n\to\infty)} \left(\frac{\mu}{n}\sum_{i=1}^n \Omega_i\right) + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^\varepsilon\right) \\
&\quad + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^e\right) \\
&= \mu\Omega + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^e\right), \qquad (7.22)
\end{aligned}
$$

by the law of large numbers. The last line holds because Ω_i and Ω_i^ε do not depends on T. In this case, the joint limits as $(T, n \to \infty)$ reduces to the limits as $n \to \infty$. It remains to show that $\text{plim}_{(T,n\to\infty)} n^{-1}\sum_{i=1}^n \Omega_i^e = 0$. To save space, we only present the proof for $\text{plim}_{(T,n\to\infty)} n^{-1}\sum_{i=1}^n \Omega_{i1}^e = 0$. A sufficient condition is that $\lim_{(T,n\to\infty)} E\left\|n^{-1}\sum_{i=1}^n \Omega_{i1}^e\right\| = 0$. Using Lemma 2, we have

$$
\begin{aligned}
& E\left\|\frac{1}{n}\sum_{i=1}^n \Omega_{i1}^e\right\| = E\left\|\frac{1}{n}\sum_{i=1}^n \sum_{j=-T+1}^{T-1} \left(k(\frac{j}{T}) - 1\right)\Gamma_i(j)\right\| \\
\leq\ & \frac{1}{n}\sum_{i=1}^n \sum_{j=-T+1}^{T-1} j^{-1}\left|k(\frac{j}{T}) - 1\right| jE\left\|\Gamma_i(j)\right\| \\
\leq\ & \frac{1}{n}\sum_{i=1}^n \left(\sum_{j=-T+1}^{T-1} j^{-2}\left|k(\frac{j}{T}) - 1\right|^2\right)^{1/2} \left(\sum_{j=-\infty}^{\infty} j^2 E\left\|\Gamma_i(j)\right\|^2\right)^{1/2} \\
\leq\ & \frac{M}{n\sqrt{T}}\left(\frac{1}{T}\sum_{j=-T+1}^{T-1} (\frac{j}{T})^{-2}\left|k(\frac{j}{T}) - 1\right|^2\right)^{1/2} = O_p\left(\frac{1}{n\sqrt{T}}\right) \qquad (7.23)
\end{aligned}
$$

as $(T, n \to \infty)$. By the Markov inequality, we get $\text{plim}_{(T,n\to\infty)} n^{-1}\sum_{i=1}^n \Omega_{i1}^e = 0$, which completes the proof of the joint limits.

Next, we consider the sequential probability limits. By Lemma 5(a) of Phillips and Moon (1999), it suffices to show that, for fixed n, the probability limit $\text{plim}_{T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\widehat{\Omega}_i$ exists. But the latter is true by Lemma 4(b). ■

Proof of Lemma 7. Note that

$$\begin{aligned}
& E\text{vec}\left(C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1)\right)\text{vec}\left(C_{yi}(1)\Xi_i C'_{xi}(1) - \beta C_{xi}(1)\Xi_i C'_{xi}(1)\right)' \\
= \; & E\left(\text{vec}(C_{yi}(1) - \beta C_{xi}(1))\Xi_i C'_{xi}(1)\right)\text{vec}\left((C_{yi}(1) - \beta C_{xi}(1))\Xi_i C'_{xi}(1)\right)' \\
= \; & E\left(C_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))\text{vec}(\Xi_i)\right)\left(\text{vec}(\Xi_i)' C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))'\right) \\
= \; & EC_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1)) E\left(\text{vec}(\Xi_i)\text{vec}(\Xi_i)'\right) C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))',
\end{aligned}$$

and $E\left(\text{vec}(\Xi_i)\text{vec}(\Xi_i)'\right)$ can be written as

$$E\left(\int_0^1\int_0^1\int_0^1\int_0^1 k(r,s)k(p,q)\text{vec}\left(dW_m(r)dW'_m(s)\right)\text{vec}\left(dW_m(p)dW'_m(q)\right)'\right).$$

Some calculations show that $E\left(\text{vec}\left(dW_m(r)dW'_m(s)\right)\text{vec}\left(dW_m(p)dW'_m(q)\right)\right)$ is

$$\begin{cases} \text{vec}(I_m)\text{vec}(I_m)' drdp, & \text{if } r = s \neq p = q, \\ I_{m^2} drds, & \text{if } r = p \neq s = q, \\ K_{mm} drds, & \text{if } r = q \neq s = p, \\ 0, & \text{otherwise.} \end{cases} \tag{7.24}$$

Using the above result, we have

$$E\left(\text{vec}(\Xi_i)\,\text{vec}(\Xi_i)'\right) = \mu^2\text{vec}(I_m)\text{vec}(I_m)' + \delta^2\left(I_{m^2} + K_{mm}\right).$$

Consequently,

$$\begin{aligned}
& EC_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1)) E\left(\text{vec}(\Xi_i)\text{vec}(\Xi_i)'\right) C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))' \\
= \; & \mu^2 C_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))\text{vec}(I_m)\text{vec}(I_m)' C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))' \\
& + \delta^2 E\left(C_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))\right)\left(C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))'\right) \\
& + \delta^2 E\left(C_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))\right)\left((C_{yi}(1) - \beta C_{xi}(1))' \otimes C'_{xi}(1)\right) K_{m_y m_x} \\
= \; & \mu^2 E\text{vec}\left((C_{yi}(1) - \beta C_{xi}(1)) I_m C'_{xi}(1)\right)\text{vec}\left((C_{yi}(1) - \beta C_{xi}(1)) I_m C'_{xi}(1)\right)' \\
& + \delta^2 E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\
& + \delta^2\left(E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})\right) K_{m_y m_x} \\
= \; & \mu^2 E\text{vec}\left(\Omega_{yxi} - \beta\Omega_{xxi}\right)\text{vec}\left(\Omega_{yxi} - \beta\Omega_{xxi}\right)' \\
& + \delta^2 E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\
& + \delta^2\left(E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})\right) K_{m_y m_x}.
\end{aligned}$$

Here we have used the identity that

$$K_{mm}\left(C'_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))'\right) = \left((C_{yi}(1) - \beta C_{xi}(1))' \otimes C'_{xi}(1)\right) K_{m_y m_x},$$

(see Part (viii) of Theorem 3.1 in Magnus and Neudecker (1979)). ■

Proof of Theorem 9. Under the joint limit, we have shown $\widehat{\Omega}_{xx} \to_p \mu\Omega_{xx}$ and $b_{nT} \to_p 0$ as $(n, T \to \infty)$ and $\sqrt{n}/T \to 0$. To prove the theorem, it suffices to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{i,T} \Rightarrow N(0, \Theta)$$

under the joint limit. Note that $Q_{i,T}$ are iid random matrices across i with zero mean and covariance matrix $\Theta_T = E\text{vec}(Q_{i,T})\text{vec}(Q_{i,T})'$. To calculate Θ_T, let

$$\begin{aligned} G_m &= \begin{pmatrix} 0 & 0 \\ 0 & I_{m_x} \end{pmatrix} \text{ and} \\ \mu_T &= \frac{1}{T}\sum_{t=1}^{T} K(\frac{t}{T}, \frac{t}{T}),\ \delta_T^2 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T} K^2(\frac{t}{T}, \frac{\tau}{T}). \end{aligned}$$

Then, by Lemma 4 (b)**,** Θ_T is

$$\begin{aligned} & E\text{vec}(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi} - E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi})\text{vec}(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{yxi} - E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{yxi})' \\ = \ & E\text{vec}\left[\left(I_{m_y}, -\beta\right)\left(\widehat{\Omega}_i - E\widehat{\Omega}_i\right)G_m\right]\text{vec}\left[\left(I_{m_y}, -\beta\right)\left(\widehat{\Omega}_i - E\widehat{\Omega}_i\right)G_m\right]' \\ = \ & \left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right] E\text{vec}\left(\widehat{\Omega}_i - E\widehat{\Omega}_i\right)\text{vec}\left(\widehat{\Omega}_i - E\widehat{\Omega}_i\right)'\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]' \\ = \ & \mu_T^2\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right] E\text{vec}\left(C_i(1)C_i'(1)\right)\text{vec}\left(C_i(1)C_i'(1)\right)\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]' \\ & -\mu_T^2\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right] E\text{vec}\left(C_i(1)C_i'(1)\right) E\text{vec}\left(C_i(1)C_i'(1)\right)'\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]' \\ & +\delta_T^2\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]\left(C_i(1) \otimes C_i(1)\right)\left(C_i'(1) \otimes C_i'(1)\right)\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]' \\ & +\delta_T^2\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]\left(C_i(1) \otimes C_i(1)\right) K_{mm}\left(C_i'(1) \otimes C_i'(1)\right)\left[G_m' \otimes \left(I_{m_y}, -\beta\right)\right]' \\ & +o(1). \end{aligned}$$

A few more calculations give us

$$\begin{aligned} \Theta_T = \ & \mu_T^2 E\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\left(\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\right)' \\ & +\delta_T^2 E\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right) \\ & +\delta_T^2 E\left(\Omega_{xyi} - \Omega_{xxi}\beta'\right) \otimes \left(\Omega_{yxi} - \beta\Omega_{xxi}\right) K_{m_y m_x} + o(1). \end{aligned}$$

So $\{Q_{i,T}\}_i$ is an iid sequence with mean zero and covariance matrix Θ_T.

Next we apply Theorem 3 of Phillips and Moon (1999) with $C_i = I_{m_y m_x}$ to establish $1/\sqrt{n}\sum_{i=1}^{n} Q_{i,T} \Rightarrow N(0, \Theta)$. Conditions (i), (ii) and (iv) of the theorem are obviously satisfied in view of the facts that $C_i = I_{m_y m_x}$ and $\Theta_T \to \Theta$ as $T \to \infty$. To prove the uniform integrability of $\|Q_{i,T}\|$, we use Theorem 3.6 of Billingsley(1999). Put in our context, the theorem states that if $\|Q_{i,T}\| \Rightarrow \|Q_i\|$ and $E\|Q_{i,T}\| \to E\|Q_i\|$, then $\|Q_{i,T}\|$ is uniformly integrable. Note that, using the continuous mapping theorem, we have, as $T \to \infty$,

$$\begin{aligned} \|Q_{i,T}\|^2 \Rightarrow \ & \|Q_i\|^2 = \left\|C_{yi}(1)\Xi_i C_{xi}'(1) - \beta C_{xi}(1)\Xi_i C_{xi}'(1)\right\|^2 \\ = \ & \left\|\left(C_{yi}(1) - \beta C_{xi}(1)\right)\int_0^1\int_0^1 K(r,s)dW_i(r)dW_i(s)C_{xi}'(1)\right\|^2, \end{aligned}$$

and

$$
\begin{aligned}
E\left\|Q_{i,T}\right\|^2 &= Etr\left(\mathrm{vec}(Q_{i,T})\mathrm{vec}(Q_{i,T})'\right) = tr(\Theta_T) \\
&\to tr(\Theta) = E\left\|Q_i\right\|^2.
\end{aligned}
$$

Therefore, $\|Q_{i,T}\|$ is uniformly integrable. Invoking Theorem 3 of Phillips and Moon (1999) to complete the proof. ■

Proof of Lemma 11. Note that $\widehat{\beta}_i-\beta_i=(\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{xxi})\widehat{\Omega}_{xxi}^{-1}$, we first consider the stochastic order of $\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{xxi}$. By definition,

$$
\begin{aligned}
\widehat{\Omega}_{yxi} &= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T}\beta_i U_{x_i,t}K(\frac{t}{T},\frac{\tau}{T})U'_{x_i\tau}+\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T}\left(E_{i,t}-E_{i,t-1}\right)K(\frac{t}{T},\frac{\tau}{T})U'_{x_i\tau} \\
&= \beta_i\widehat{\Omega}_{xxi}+\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T}\left(E_{i,t}-E_{i,t-1}\right)K(\frac{t}{T},\frac{\tau}{T})U'_{x_i\tau}. \\
&= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T-1}E_{i,t}(K(\frac{t}{T},\frac{\tau}{T})-K(\frac{t+1}{T},\frac{\tau}{T}))U'_{x_i\tau} \\
&\quad +\beta_i\widehat{\Omega}_{xxi}+\frac{1}{T}E_{iT}\sum_{\tau=1}^{T}K(1,\frac{\tau}{T})U'_{x_i\tau} \qquad (7.25)
\end{aligned}
$$

where the last equality follows from summation by parts.

Therefore, when $K(1,r)=K(s,1)=0$ for any r and s,

$$
T(\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{xxi})=\sum_{\tau=1}^{T}\sum_{t=1}^{T-1}E_{i,t}(K(\frac{t}{T},\frac{\tau}{T})-K(\frac{t+1}{T},\frac{\tau}{T}))U'_{x_i\tau} \qquad (7.26)
$$

Following the same steps as the proof of Lemma 4(a), we can prove that

$$
\sum_{\tau=1}^{T}\sum_{t=1}^{T-1}E_{i,t}(K(\frac{t}{T},\frac{\tau}{T})-K(\frac{t+1}{T},\frac{\tau}{T}))U'_{x_i\tau}=O_p(1), \qquad (7.27)
$$

provided that $K(\cdot,\cdot)$ is Lipschitz continuous. As a consequence, we get $T(\widehat{\beta}_i-\beta_i)=O_p(1)$.

When $K(1,s)\neq 0$, $\sqrt{T}(\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{xxi})$ equals

$$
\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\sum_{t=1}^{T-1}E_{i,t}\left(K(\frac{t}{T},\frac{\tau}{T})-K(\frac{t+1}{T},\frac{\tau}{T})\right)U'_{x_i\tau}+\frac{1}{\sqrt{T}}E_{iT}\sum_{\tau=1}^{T}K(1,\frac{\tau}{T})U'_{x_i\tau}. \qquad (7.28)
$$

In view of (7.27), the first term $o_p(1)$. The second term is $O_p(1)$ because

$$
1/\sqrt{T}\sum_{\tau=1}^{T}K(1,\frac{\tau}{T})U'_{x_i\tau}\Rightarrow\int_0^1 K(1,r)dW_i'(r)C_{xi}'(1). \qquad (7.29)
$$

Hence $\sqrt{T}(\widehat{\Omega}_{yxi}-\beta_i\widehat{\Omega}_{xxi})=O_p(1)$, which implies that $\sqrt{T}(\widehat{\beta}_i-\beta_i)=O_p(1)$. ■

REFERENCES

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817–854.

——— (2003): "Cross-section Regression with Common Shocks," Cowles Foundation Discussion Paper No. 1428. Yale University.

Baltagi, B. H. and Kao, C. (2000): "Nonstationary Panels, Panel Cointegration, and Dynamic Panels: A Survey," *Advances in Econometrics* 15, JAI Press.

Billingsley, P. (1999): *Convergence of Probability Measures*, Wiley: New York.

Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92 1-45.

de Jong, R. M. and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices," *Econometrica,* 68, 407–424.

Engle, R. F. and C. W. J. Granger, (1987): "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

Hannan, E. J. (1970): *Multiple Time Series*, New York, Wiley.

Hansen, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes," *Econometrica,* 60, 967–972.

Jansson, M. (2002): "Autocorrelation robust tests with good size and power," Department of Economics, University of California, Berkeley.

Kao, C. (1999): "Spurious Regression and Residual-based Tests for Cointegration in Panel Data," *Journal of Econometrics,* 90, 1–44.

Kiefer, N. M. and T. J. Vogelsang (2002a): "Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory,* 18, 1350–1366.

——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

——— (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Department of Economics, Cornell University.

Magnus, J. R. and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7, 381–394.

Makela, T. (2002): "Long run Covariance Based Inference in Nonstationary Panels with Large Cross Section," Department of Economics, Yale University.

Nabeya, S. and K. Tanaka (1988): "Asymptotic Theory of a Test for the Consistency of Regression Coefficients against the Random Walk Alternative," *Annals of Statistics,* 16, 218–235.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55, 703–708.

Parzen, E. (1957): "On the Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

Pedroni, P. (1995): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests, With an Application to the PPP Hypothesis," Indiana University Working Papers in Economics no. 95–013.

Phillips, P. C. B. and H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

——— (2000): "Nonstationary Panel Data Analysis: an Overview of Some Recent Developments," *Econometric Reviews*, 19(3), 263–286.

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

Phillips, P. C. B. and D. Sul (2002): "Dynamic Panel Estimation and Homogeneity Testing under Cross Sectional Dependence," Cowles Foundation Discussion Paper No. 1362. Yale University.

Phillips, P. C. B., Y. Sun and S. Jin (2003a): "Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," Department of Economics, UCSD.

Phillips, P. C. B., Y. Sun and S. Jin (2003b): "Long Run Variance Estimation with Steep Origin Kernels without Truncation," Department of Economics, UCSD.

Sun, Y. (2002): "A Convergent t-statistic in Spurious Regressions," Department of Economics, UCSD.

Sun, Y. (2003): "Estimation of the Long-run Average Relationship in Nonstationary Panel Time Series," Department of Economics Working Paper 2003-06, UCSD.

Tanaka, K. (1996): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley: New York.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.