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

Author
Sun, Yixiao

Publication Date
2003-04-01

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

Yixiao Sun

Department of Economics
University of California, San Diego

This Version: March 2003

*My thanks go to Roger Moon, Peter Phillips and Zhijie Xiao for helpful comments.

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 when there is no individual time series cointegration. Using panel data with large cross section (n) and time series dimensions (T), the estimators are based on the long-run average variance estimate using bandwidth equal to T. The new estimators include the panel pooled least squares estimators and the limiting cross sectional least squares 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 taking powers of the Bartlett and Parzen kernels, this paper introduces two new classes of kernels, the sharp kernels and steep kernels, and shows that these new kernels deliver new estimators of the long-run average relationship that are more efficient than the existing ones. A simulation study supports the asymptotic results.

JEL Classification: C32; C33

KEYWORDS: Long-run average relationship, long-run variance matrix, multidimensional limits, panel spurious regression, sharp kernels, steep kernels.

1 Introduction

Nonstationary panel data with large cross section and time series dimensions have attracted much attention in recent years (e.g. Pedroni 1995; Kao 1999; Phillips and Moon; 2000 and Baltagi 2000). Financial and macroeconomic panel data sets that cover many firms, regions or countries over a relatively long time period are familiar examples. Such a panel data structure allows us to identify the long-run average relationship between two I(1) random vectors that conventional panel data can not identify. When the two I(1) random vectors are not cointegrated for any given individual, the noise in the time series regression is as strong as the signal. As a consequence, we can not identify the long-run relationship using time series data alone. However, the noise can often be characterized as independent across individuals. By pooling the cross section and time series observations, we can attenuate the strong effect of the noise while retaining the strength of the signal. Phillips and Moon (1999) showed that both the panel pooled least squares (PLS) regression and the limiting cross-sectional least squares (CLS) regression provide consistent estimates of the long-run average relationship.

In this paper, we propose a new class of estimators of the long-run average relationship. The estimators considered include the PLS and CLS estimators as special cases and are generally more efficient than both the PLS and CLS estimators. 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 the kernel estimators of the LRV matrices (e.g. White 1980, 2001; Newey and West 1987, 1999; Andrews 1991; Andrews and Monahan, 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 estimator. 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 $O_p(1/T)$, 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 the bandwidth 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. Since the new estimator incorporates the PLS and CLS estimators as special cases, our joint limit theory is also applicable to these estimators. The joint limit theory is the same as that obtained by Phillips and Moon (1999). 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. Even if the effect of the noise in the time series estimation does not die out as $T \to \infty$, the noise is independent across individuals. Hence, by averaging across all 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 the Bartlett and Parzen kernels are more efficient than the PLS and CLS estimators designed for a model without individual effects. However, the new estimators with the Bartlett and Parzen kernels are less efficient than the PLS estimator for a model with possible individual effects. As shown by Phillips and Moon (1999), the PLS estimator becomes more efficient when intercepts are allowed in time series regressions to capture the possible individual effects. To develop an LRV-based estimator that is more efficient than the latter estimator, we introduce two new classes of kernels – sharp kernels and steep kernels. A sharp kernel is defined to be the Bartlett kernel raised to some integer power (sharpness index), while a steep kernel is defined to be the Parzen kernel raised to some integer power (steepness index). We show that the new estimator using the sharp kernel or steep kernel is more efficient than the PLS estimator even if intercepts are included in time series regressions. In fact, the asymptotic variance of the new estimator can be made as small as possible by choosing a large sharpness index or steepness index. 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 steep 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.

In this paper, we compare the finite sample performances of the LRV-based estimators using sharp and steep kernels with the PLS estimators. The Monte Carlo simulations show that the LRV-based estimator using the steep kernel with power parameter 2 or 4 has a smaller RMSE than the PLS estimators for almost all parameter configurations and (N,T) combinations considered. The superior performance may be explained by observing that the LRV-based estimator incorporates cross section and time series information in a delicate and sophisticated way. The cross section observations are used to reduce variance while the time series observations, together with the untruncated kernel, are used to reduce bias. So the cross section and time series observations can be used together to improve the overall estimation of the LRV matrix and LR average coefficient. The presence of the power parameter ρ makes this more flexible. In fact, there is an opportunity for optimal choice of ρ, but this is beyond the scope of the current paper.

The kernel LRV estimator using the full bandwidth (the bandwidth is equal to the sample size) has been suggested by Kiefer and Vogelsang (2002a, 2002b) and Vogelsang (2000) in other settings. Specifically, they considered this type of estimators in hypothesis testing in the presence of nonparametric autocorrelation. Their motivation is to develop asymptotically valid tests that are free from bandwidth selection and have good size and power properties. Our motivation is quite different. Our paper provides another instance that the kernel LRV estimator using the full bandwidth is useful. In addition, the PLS and CLS estimators are shown to be special cases of the new estimator. The LRV matrices implicitly used in these estimators are kernel LRV estimates using the full bandwidth. Therefore, the new long-run variance estimator has been employed implicitly in previous papers. Other papers that use or investigate 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 the sharp and steep 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 estimator, 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. It also proves that the new estimator includes the PLS and CLS estimators as special cases. Section 3 establishes the asymptotic properties of the kernel LRV estimator when the bandwidth is equal to the sample size. Section 4 develops the sequential and joint limit theory for the LRV-based estimator. This section also introduces the sharp and steep kernels and examines the asymptotic properties of the resulting LRV-based estimators. Section 5 provides the Monte Carlo simulation results. Section 6 concludes. Proofs are collected in the Appendix.

Throughout the paper, $\text{vec}(\cdot)$ is the column-by-column vectorization function, $\text{tr}(\cdot)$ 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 $\text{vec}(A)$ into $\text{vec}(A')$, 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, 1988). 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 "$\text{vec}(Z_n) \Rightarrow$

$N(0, \Sigma)$". 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 $EV^4_{a,i,t} = v^4$ 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 vec$(C_{i,s})$ and $\sigma_{kas} = EC^k_{a,i,s}$. We make two further assumptions on the moments of the random coefficients and the summability of these moments.

Assumption 1 (Random Coefficient Conditions)
(i) $\{C_{i,s}\}$ is *iid* across i for all s. (ii) $E \|C_{i,s}\|^4 < \infty$.

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

Assumptions 1 and 2(ii) are the same as those in Phillips and Moon (1999). Assumption 2(i) is slightly stronger than their Assumption 2(i). 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) $\sum_{s=0}^{\infty} s^2 E \|C_{i,s}\| < \infty,$

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

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

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

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

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 $\text{var}(W_i(r)) = rI_m$ 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 $\{C_{i,t}\}_{t=0}^{\infty}$.

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 equals the conditional spectral density matrix $f_{U_iU_i}(\lambda)$ of $U_{i,t}$ evaluated at the origin, i.e. $\Omega_i = f_{U_iU_i}(0)$. Partitioning Ω_i conformably, we have

$$\Omega_i = \begin{pmatrix} \Omega_{yyi} & \Omega_{yxi} \\ \Omega_{xyi} & \Omega_{xxi} \end{pmatrix}. \tag{2.6}$$

We assume that Ω_i satisfies the following rank condition:

Assumption 3 (Rank Condition) $\text{rank}(\Omega_i) = m$ almost surely for all $i = 1, ..., n$.

Assumption 3 implies that the two component random vectors $Y_{i,t}$ and $X_{i,t}$ of $Z_{i,t}$ are not cointegrated for any individual i (Engle and Granger 1987). However, the assumption does not exclude the existence of an interesting long-run relationship between the panel vectors $Y_{i,t}$ and $X_{i,t}$. This relationship is defined below.

By Lemma 1(c), Ω_i is integrable and

$$\Omega = E\Omega_i = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix}, \tag{2.7}$$

which we 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:

$$\widehat{\Omega}_i = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} U_{i,s} K(\frac{s}{T}, \frac{t}{T}) U'_{i,t}, \tag{2.8}$$

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

$$\widehat{\Omega}_i = \sum_{j=-T+1}^{T-1} k(\frac{j}{T}) \widehat{\Gamma}_i(j), \tag{2.9}$$

$$\widehat{\Gamma}_i(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} U_{i,t+j}U'_{i,t} & \text{for } j \geq 0, \\ \frac{1}{T}\sum_{t=-j+1}^{T} U_{i,t+j}U'_{i,t} & \text{for } j < 0. \end{cases} \tag{2.10}$$

From the above formulation, it is clear that $\widehat{\Omega}_i$ is the usual kernel LRV estimator using the full bandwidth (the bandwidth is equal to the time series dimension). 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 only to include some existing estimators of the long-run average relationship as special cases. This will be made clear in subsection 2.3.

Based on the above estimate, we can estimate Ω by

$$\widehat{\Omega} = \begin{pmatrix} \widehat{\Omega}_{yy} & \widehat{\Omega}_{yx} \\ \widehat{\Omega}_{xy} & \widehat{\Omega}_{xx} \end{pmatrix} = n^{-1}\sum_{i=1}^{n} \widehat{\Omega}_i. \tag{2.11}$$

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

$$\widehat{\beta}_{LRV} = \widehat{\Omega}_{yx}\widehat{\Omega}_{xx}^{-1}, \tag{2.12}$$

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

$$Z_{i,t} = A_{i,0} + Z^0_{i,t} \tag{2.13}$$

$$Z^0_{i,t} = Z^0_{i,t-1} + U_{i,t}, \tag{2.14}$$

where $Z^0_{i,0} = 0$, 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 the new estimator and existing estimators

Phillips and Moon (1999) considered the PLS and CLS estimators for the model with no individual effects and the PLS estimator for the model with individual effects.

They showed that these estimators are consistent and asymptotically normal. In this subsection, we examine the relationships between the LRV-based estimator and the estimators considered by Phillips and Moon (1999).

When the model contains no individual effects, the PLS estimator is defined as

$$\widetilde{\beta}_{PLS} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} Y_{i,t}X_{i,t}'\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}X_{i,t}'\right)^{-1}. \tag{2.15}$$

Some simple algebraic manipulations show that

$$\begin{aligned}\sum_{t=1}^{T} Y_{i,t}X_{i,t}' &= \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{\tau=1}^{t} U_{y_i,s}U_{x_i,\tau}' \\ &= \sum_{s=1}^{T}\sum_{t=1}^{T}\left(T-(s+1)\vee(t+1)\right)U_{y_i,s}U_{x_i,t}' \end{aligned}\tag{2.16}$$

and

$$\sum_{t=1}^{T} X_{i,t}X_{i,t}' = \sum_{s=1}^{T}\sum_{t=1}^{T}\left(T-(s+1)\vee(t+1)\right)U_{x_i,s}U_{x_i,t}'. \tag{2.17}$$

Therefore,

$$\begin{aligned}\widetilde{\beta}_{PLS} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{PLS,T}(\frac{s}{T},\frac{t}{T})U_{y_i,s}U_{x_i,t}'\right) \\ &\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{PLS,T}(\frac{s}{T},\frac{t}{T})U_{x_i,s}U_{x_i,t}'\right)^{-1}, \end{aligned}\tag{2.18}$$

where

$$\begin{aligned}K_{PLS,T}(\frac{s}{T},\frac{t}{T}) &= 1-\frac{(s+1)\vee(t+1)}{T} \text{ and} \\ (s+1)\vee(t+1) &= \max\left(s+1,t+1\right). \end{aligned}\tag{2.19}$$

Hence, the PLS estimator is a special case of the LRV-based estimator. Note that the kernel for the PLS estimator depends on T. An asymptotically equivalent LRV-based estimator is

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

where $K_{PLS}(s,t) = 1-(s\vee t)$. The underlying LRV estimator is (2.8) with $K(s,t) = 1-(s\vee t)$. Therefore, the PLS estimator is based on an LRV estimate with the bandwidth equal to the time series dimension.

For models containing no individual effects, the CLS estimator is defined by

$$\widehat{\beta}_{CLS,r_0} = \left(\sum_{i=1}^{n} Y_{i,\tau} X'_{i,\tau}\right)\left(\sum_{i=1}^{n} X_{i,\tau} X'_{i,\tau}\right)^{-1}, \tag{2.21}$$

where $\tau = [Tr_0]$. Plugging in $Y_{i,t} = Y_{i,t-1} + U_{y_i,t}$ and $X_{i,t} = X_{i,t-1} + U_{x_i,t}$, we have

$$\begin{aligned}
\widetilde{\beta}_{CLS,r_0} &= \left(\sum_{i=1}^{n}\sum_{s=1}^{[Tr_0]}\sum_{t=1}^{[Tr_0]} U_{y_i,s}U'_{x_i,t}\right)\left(\sum_{i=1}^{n}\sum_{s=1}^{[Tr_0]}\sum_{t=1}^{[Tr_0]} U_{x_i,s}U'_{x_i,t}\right)^{-1} \\
&= \left(\sum_{i=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{CLS}(\frac{s}{T},\frac{t}{T})U_{y_i,s}U'_{x_i,t}\right) \\
&\quad \times \left(\sum_{i=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T} K_{CLS}(\frac{s}{T},\frac{t}{T})U_{x_i,s}U'_{x_i,t}\right)^{-1},
\end{aligned} \tag{2.22}$$

where $K_{CLS}(s,t) = 1\left\{s \le r_0, t \le r_0\right\}$. Therefore, the CLS estimator is also a special case of the LRV-based estimator. The underlying variance matrix estimator is (2.8) with $K(s,t) = 1\left\{s \le r_0, t \le r_0\right\}$.

We now consider the PLS estimator when intercepts are allowed in the time series regressions. In this case, the PLS estimator is defined as

$$\widetilde{\beta}_{PLS,c} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{i,t} - \bar{Y}_{i,\cdot}\right)\left(X_{i,t} - \bar{X}_{i,\cdot}\right)'\right)\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(X_{i,t} - \bar{X}_{i,\cdot}\right)\left(X_{i,t} - \bar{X}_{i,\cdot}\right)'\right)^{-1}, \tag{2.23}$$

where $\bar{Y}_{i,\cdot} = 1/T\sum_{t=1}^{T} Y_{i,t}$ and $\bar{X}_{i,\cdot} = 1/T\sum_{t=1}^{T} X_{i,t}$. But

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\left(Y_{i,t} - \bar{Y}_{i,\cdot}\right)\left(X_{i,t} - \bar{X}_{i,\cdot}\right)' = \frac{1}{T^2}\sum_{t=1}^{T} Y_{i,t}X'_{i,t} - \frac{1}{T}\bar{Y}_{i,\cdot}\bar{X}'_{i,\cdot} \\
&= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(T - (s \vee t) + 1)U_{y_i,s}U'_{x_i,t} - \frac{1}{T}\sum_{s=1}^{T}\frac{T-s+1}{T}U_{y_i,s}\sum_{t=1}^{T}\frac{T-t+1}{T}U'_{x_i,t} \\
&= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\{\frac{T - (s \vee t) + 1}{T} - \left(\frac{T-s+1}{T}\right)\left(\frac{T-t+1}{T}\right)\right\}U_{y_i,s}U'_{x_i,t}.
\end{aligned} \tag{2.24}$$

Similarly,

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\left(X_{i,t} - \bar{X}_{i,\cdot}\right)\left(X_{i,t} - \bar{X}_{i,\cdot}\right)' = \frac{1}{T^2}\sum_{t=1}^{T} X_{i,t}X'_{i,t} - \frac{1}{T}\bar{X}_{i,\cdot}\bar{X}'_{i,\cdot} \\
&= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left\{\frac{T - (s \vee t) + 1}{T} - \left(\frac{T-s+1}{T}\right)\left(\frac{T-t+1}{T}\right)\right\}U_{x_i,s}U'_{x_i,t}
\end{aligned} \tag{2.25}$$

Therefore,

$$\widetilde{\beta}_{PLS,c} = \left(\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}K_{PLS,c}^{T}\left(\frac{s}{T},\frac{t}{T}\right)U_{y_i,s}U_{x_i,t}'\right) \times\left(\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}K_{PLS,c}^{T}\left(\frac{s}{T},\frac{t}{T}\right)U_{x_i,s}U_{x_i,t}'\right)^{-1}, \tag{2.26}$$

where

$$K_{PLS,c}^{T}\left(\frac{s}{T},\frac{t}{T}\right)=\frac{T-(s\vee t)+1}{T}-\left(\frac{T-s+1}{T}\right)\left(\frac{T-t+1}{T}\right). \tag{2.27}$$

The kernel function $K_{PLS,c}^{T}(s,t)$ depends on T. An asymptotically equivalent estimator with a kernel function that is independent of T is

$$\widehat{\beta}_{PLS,c} = \left(\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}K_{PLS,c}\left(\frac{s}{T},\frac{t}{T}\right)U_{y_i,s}U_{x_i,t}'\right) \times\left(\sum_{i=1}^{n}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}K_{PLS,c}\left(\frac{s}{T},\frac{t}{T}\right)U_{x_i,s}U_{x_i,t}'\right)^{-1}, \tag{2.28}$$

where

$$K_{PLS,c}(s,t)=1-(s\vee t)-(1-s)(1-t)=\min(s,t)-st. \tag{2.29}$$

From the above expression, it is clear that $\widehat{\beta}_{PLS,c}$ 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 which are not translation invariant. This sharply contrasts with the usual LRV estimator 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

$$f_{U_iU_i}^{(q)}=\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}|j|^{q}\Gamma_i(j). \tag{3.30}$$

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

COMMENT: Lemma 2(b) reveals that $f_{U_iU_i}^{(2)} < \infty$ almost surely for $i = 1, 2, ..., n$. When q is even,

$$f_{U_iU_i}^{(q)} = (-1)^{q/2} d^q f_{U_iU_i}(\lambda)/d\lambda^q \Big|_{\lambda=0}. \tag{3.31}$$

Therefore, the boundedness of $f_{U_iU_i}^{(2)}$ implies that $f_{U_iU_i}(\lambda)$ is differentiable to the second order at $\lambda = 0$.

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

$$k_q = \lim_{x \to 0} \frac{1 - k(x)}{|x|^q} < \infty \text{ for } q \geq 0. \tag{3.32}$$

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

$$\mathcal{K}_1 = \{K(s,t) : K(s,t) = 1 - (s \vee t)\,,\ 1\{s \leq r_0, t \leq r_0\}\,,\ \text{or}\ \min(s,t) - st\}$$

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

(i) $k(x) : [-1,1] \to [0,1]$ is continuous and satisfies $k(x) = k(-x)$ and $k(0) = 1$.

(ii) the Parzen characteristic exponent of $k(x)$ is greater than or equal to one.

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

Assumption 4 guarantees the boundedness of $\int_0^1 \int_0^1 K^2(r,s) dr ds$, a quantity that appears in the expressions for the asymptotic variances. Assumption 4(i)–(iii) is not as restrictive as it seems. Examples of commonly used kernels satisfying Assumption 4(i)–(iii) include the Bartlett and Parzen kernels:

$$\begin{aligned} &\text{Bartlett} \quad k_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1, \\ 0 & \text{otherwise,} \end{cases} \\ &\text{Parzen} \quad k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.33}$$

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

$$\int_0^1 \int_0^1 K(s,t)f(s)f(t)dsdt = \int_0^1 \left(\int_0^t f(s)ds \right)^2 ds \geq 0. \qquad (3.34)$$

When $K(s,t) = \{s \leq r_0, t \leq r_0\}$,

$$\int_0^1 \int_0^1 K(s,t)f(s)f(t)dsdt = \left(\int_0^{r_0} f(s)ds \right)^2 \geq 0. \qquad (3.35)$$

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

$$\int_0^1 \int_0^1 K(s,t)f(s)f(t)dsdt = \int_0^1 F^2(s)ds - \left(\int_0^1 F(s)ds \right)^2 \geq 0, \qquad (3.36)$$

where $F(s) = \int_0^s f(r)dr$. Therefore, the kernels satisfying Assumption 4 are positive semi-definite. As shown by Newey and West (1987) and Andrews (1991), the positive semi-definiteness guarantees the positive semi-definiteness of $\widehat{\Omega}_i$. In addition, it enables us to use the Mercer's theorem. We present a modified version below, which helps establish the asymptotic distribution of $\widehat{\Omega}_i$. For more discussion on the Mercer's Theorem, see Tanaka (1996).

Mercer's Theorem *If $k(x)$ is positive semi-definite, then*

$$k(r-s) = \sum_{m=1}^{\infty} \frac{1}{\lambda_m} f_m(r) f_m(s),$$

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 $(r,s) \in [0,1] \times [0,1]$.

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 $1-4$ hold.*

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

$$\lim_{T\to\infty} TE\left[E\left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu \Omega_i \right] = -2\pi(k_1+1)Ef_{U_iU_i}^{(1)}. \qquad (3.37)$$

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

$$\lim_{T\to\infty} TE\left[E\left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu \Omega_i \right] = -2\pi Ef_{U_iU_i}^{(1)}. \qquad (3.38)$$

(c) If $K(s,t) \in \mathcal{K}_1$, then $E\left(E\left(\widehat{\Omega}_i | \mathcal{F}_{c_i} \right) - \mu\Omega_i \right) = O(1/T)$, where $\mu = \int_0^1 K(s,s)ds$.

COMMENTS: (i) When $K(s,t)$ is translation invariant, $K(s,s) = k(0) = 1$, so $\int_0^1 K(s,s)ds = 1$. 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 $\left(\int_0^1 K(s,s)ds\right)^{-1} \widehat{\Omega}_i$ 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 $S_T/T + S_T^q/T + 1/S_T \to 0$ (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 $-2\pi T^{-2} k_2 E f_{U_i U_i}^{(2)}$. 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 subsection 4.2.

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

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

$$\widetilde{\Omega}_i = T^{-1} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \left(C_i(1)V_{i,t}\right) K\left(\frac{t}{T}, \frac{\tau}{T}\right) \left(C_i(1)V_{i,\tau}\right)'; \tag{3.39}$$

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

$$\delta^2 = \int_0^1 \int_0^1 K^2(r,s) dr ds. \tag{3.40}$$

COMMENTS: (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 $\lim_{T\to\infty} \text{var}\Big(\text{vec}(\widehat{\Omega}_i) | \mathcal{F}_{c_i}\Big) = \delta^2 \left(I_{m^2} + K_{mm}\right) \left(\Omega_i \otimes \Omega_i\right)$. 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) shows that the conditional covariance between the (a,b) and (c,d) elements of $\widehat{\Omega}_i$ converges to $\delta^2 \left(\Omega_{iac}\Omega_{ibd} + \Omega_{iad}\Omega_{ibc}\right)$, where Ω_{iab} denotes the (a,b) element of Ω_i.

(iii) Lemma 4(c) calculates the limit of the finite sample variance of $\widehat{f}_{U_i U_i}(\lambda)$ 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 $\widehat{f}_{U_i U_i}(\lambda)$ for other values of λ when the full bandwidth is used in the smoothing. This extension may be needed to investigate seasonally integrated processes. This extension is straightforward but tedious and beyond the scope of this paper.

Lemma 5 *Let Assumptions* 1 –4 *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*

$$\Xi_i = \int_0^1 \int_0^1 K(r,s)dW_i(r)dW_i'(s). \tag{3.41}$$

COMMENTS: (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 $K(r,s) = 1 - (r \vee s)$, some simple calculations show that $\Xi_i = \int_0^1 W_i(s)W_i'(s)ds$. So

$$T^{-2}\sum_{t=1}^{T} Z_{i,t}Z_{i,t}' \Rightarrow C_i(1)\left(\int_0^1 W_i(s)W_i'(s)ds\right)C_i'(1). \tag{3.42}$$

When $K(r,s) = \{r \le r_0, s \le r_0\}$, we have $\Xi_i = W_i(r_0)W_i'(r_0)$. So

$$T^{-2}\sum_{t=1}^{T} Z_{i,[Tr_0]}Z_{i,[Tr_0]}' \Rightarrow C_i(1)W_i(r_0)W_i'(r_0)C_i'(1). \tag{3.43}$$

When $K(r,s) = \min(s,t) - st$, we have $\Xi_i = \int_0^1 W_i(s)W_i'(s)ds - \left(\int W_i(s)ds\right)^2$. So

$$\begin{aligned} & T^{-2}\sum_{t=1}^{T} \left(Z_{i,t} - Z_{i,.}\right)\left(Z_{i,t} - Z_{i,.}\right)' \\ \Rightarrow \quad & C_i(1)\left(\int_0^1 W_i(s)W_i'(s)ds - \left(\int W_i(s)ds\right)^2\right)C_i'(1). \end{aligned} \tag{3.44}$$

The above weak convergence results are consistent with (2.5) and the continuous mapping theorem.

4 Asymptotic Properties of the LRV-based Estimator

This section considers the asymptotic properties of the LRV-based estimator. Before proceeding, 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*

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

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

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

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

COMMENT: $\widehat{\beta}_{LRV}$ 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. Even if the noise in the time series estimation is strong, the noise can be characterized as independent across individuals. Hence, by averaging across individuals, we may weaken the strong effect of noise while maintaining the strength of the signal. This is reflected in Lemma 6(i) and (ii), which show that $\widehat{\Omega}_{xx}$ and $\widehat{\Omega}_{yx}$ are respective consistent estimates of Ω_{xx} and Ω_{yx} up to a multiplicative scaler.

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 3, 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

$$\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx}) \Rightarrow \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q_i, \tag{4.1}$$

where

$$Q_i = C_{yi}(1)\Xi_i C_{xi}'(1) - \beta C_{xi}(1)\Xi_i C_{xi}'(1), \tag{4.2}$$

$C_{yi}(1)$ 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

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

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

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

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

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 $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(C_{yi}(1)\Xi_i C_{xi}'(1) - \beta C_{xi}(1)\Xi_i C_{xi}'(1)\right) \Rightarrow N(0, \Theta). \tag{4.3}$$

Combining (4.3) 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 $(T, n \to \infty)_{seq}$,*

$$\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1} \otimes I_{m_y}), \tag{4.4}$$

where Θ_{LRV} is

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

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

$$
\begin{aligned}
\sqrt{n}(\widehat{\Omega}_{yx} - \beta\widehat{\Omega}_{xx}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi} - E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi})\right) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_{i,T} + b_{nT}, \qquad (4.5)
\end{aligned}
$$

where

$$
Q_{i,T} = \widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi} - E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}) \qquad (4.6)
$$

and

$$
b_{nT} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi}). \qquad (4.7)
$$

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

$$
\begin{aligned}
b_{nT} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left(E\left(\widehat{\Omega}_{yxi} - \beta\widehat{\Omega}_{xxi})|\mathcal{F}_{c_i}\right)\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[E\left(\Omega_{yxi} - \beta\Omega_{xxi}\right) + O(1/T)\right] = O\left(\sqrt{n}/T\right), \qquad (4.8)
\end{aligned}
$$

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 $\sqrt{n}/T \to 0$,*

$$
\sqrt{n}(\widehat{\beta}_{LRV} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1} \otimes I_{m_y}). \qquad (4.9)
$$

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

$$\sqrt{n}(\widehat{\beta}_{PLS} - \beta) \Rightarrow N(0, (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{PLS}(\Omega_{xx}^{-1} \otimes I_{m_y}) \tag{4.10}$$

with

$$\begin{aligned}\Theta_{PLS} &= E\text{vec}\left(\Omega_{yxi} - \beta\Omega_{xxi}\right)\text{vec}\left(\Omega_{yxi} - \beta\Omega_{xxi}\right)' \\ &+2/3E\left(\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right)\right) \\ &+2/3\left(E(\Omega_{xyi} - \Omega_{xxi}\beta') \otimes (\Omega_{yxi} - \beta\Omega_{xxi})\right)K_{m_y m_x}.\end{aligned} \tag{4.11}$$

The above limiting distribution is identical to that obtained by Phillips and Moon (1999). Similarly, we can show that the CLS estimator has the same asymptotic distribution as in (4.10) but with 2/3 replaced by 1. Again, the limiting distribution is the same as that obtained by Phillips and Moon (1999).

(ii) When the time series regression includes an intercept, the kernel for the PLS estimator is $K(s,t) = \min(s,t) - st$. In this case,

$$\mu_c = \int_0^1 K(r,r)dr = \int_0^1 \left(r - r^2\right)dr = 1/6, \tag{4.12}$$

and

$$\begin{aligned}\delta_c^2 &= \int_0^1 K^2(r,s)drds = 2\int_0^1\int_0^s (r - rs)^2\, drds \\ &= \int_0^1 \left(\frac{2}{3}(-1+s)^2 s^3\right)ds = 1/90.\end{aligned} \tag{4.13}$$

So $\mu_c^{-2}\delta_c^2 = 36/90 = 2/5$. Hence $\widehat{\beta}_{PLS,c}$ has the limiting distribution in (4.10) with 2/3 replaced by 2/5. Once again, we obtain the asymptotic result that was established in Phillips and Moon (1999).

(iii) $\widehat{\beta}_{PLS,c}$ is more efficient than the $\widehat{\beta}_{PLS}$ and $\widehat{\beta}_{PLS,r_0}$. But $\widehat{\beta}_{PLS,c}$ is less efficient than $\widehat{\beta}_{LRV}$ if $\kappa = \left(\int_0^1 K(s,s)ds\right)^{-2}\left(\int_0^1\int_0^1 K^2(r,s)drds\right) < 2/5$. This is apparent because asymvar$(\widehat{\beta}_{LRV})-$asymvar$(\widehat{\beta}_{PLS,c})$ is

$$\begin{aligned}&(2/5-\kappa)\,E\left(\Omega_{xxi} \otimes \Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right) \\ &+(2/5-\kappa)\,E\left(\Omega_{xyi} - \Omega_{xxi}\beta'\right) \otimes \left(\Omega_{yxi} - \beta\Omega_{xxi}\right)K_{m_y m_x} \\ =\;&(2/5-\kappa)\,E\left(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))\right) \\ &\times(I_{m^2} + K_{mm})\left(C_{x_i}(1) \otimes (C_{y_i}(1) - \beta C_{x_i}(1))\right)',\end{aligned}$$

which is negative definite if $\kappa < 2/5$.

The values of κ for the Bartlett and Parzen kernels are 0.5 and 0.4473, respectively. Therefore, $\widehat{\beta}_{LRV}$ is asymptotically more efficient than $\widehat{\beta}_{PLS}$ and $\widehat{\beta}_{CLS,r_0}$ if the Bartlett or Parzen kernel is used. However, for these two kernels, $\widehat{\beta}_{LRV}$ is less efficient than $\widehat{\beta}_{PLS,c}$. The next subsection proposes two new classes of kernels which deliver more efficient estimators than $\widehat{\beta}_{PLS,c}$.

4.2 LRV Based Estimator with Sharp and Steep Kernels

In this subsection, we consider two new classes of kernels and the asymptotic properties of the LRV-based estimators that these new kernels delivered. These new kernels have also been considered by Phillips, Sun and Jin (2003a, 2003b).

The first class consists of the sharp kernels defined by $k(x) = (1-|x|)^{\rho}1\{|x| \leq 1\}$, where $\rho \in \mathbb{Z}^{+}$ is the sharpness index. These kernels, as so defined, exhibit a sharp peak at the origin and include the Bartlett kernel as a special case. The positive semi-definiteness of sharp kernels inherits from that of the Bartlett kernel as they are equal to the products of the Bartlett kernels.

For sharp kernels, the Parzen characteristic exponent is $q^{*} = 1$ and

$$k_1 = \lim_{x \to 0} \frac{1-(1-|x|)^{\rho}}{|x|} = \rho. \tag{4.14}$$

The value of κ is

$$\kappa = \int_0^1 \int_0^1 (1-|r-s|)^{2\rho}\, dr\, ds = \frac{2}{2\rho+2}. \tag{4.15}$$

Therefore, κ is a decreasing function of the sharpness index 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

$$N\left(b,\ (\Omega_{xx}^{-1} \otimes I_{m_y})\Theta_{PLS}(\Omega_{xx}^{-1} \otimes I_{m_y}\right), \tag{4.16}$$

where $b = -2\pi\alpha\,(\rho+1)\,(\Omega_{xx}^{-1} \otimes I_{m_y})\mathrm{vec}(Ef^{(1)}_{U_{y_i}U_{x_i}} - \beta E f^{(1)}_{U_{x_i}U_{x_i}})$. 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 following the lines in Andrews (1991), but we do not pursue this line of analysis in the present paper

The second class consists of the steep kernels defined by $k(x) = (k_{PR}(x))^{\rho}$ where $k_{PR}(x)$ is the Parzen kernel and $\rho \in \mathbb{Z}^{+}$ is the steepness index. 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 kernels. 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

$$k_2 = \lim_{x \to 0} \frac{1-(1-6x^2+6\,|x|^3)^{\rho}}{x^2} = 6\rho. \tag{4.17}$$

The value of κ can be calculated using numerical integration. For $\rho = 1, 2, 3, 4, 5, 6, 7, 8$, the values of κ are 0.4473, 0.3359, 0.2806, 0.2459, 0.2216, 0.2033, 0.1890, 0.1772, respectively. Obviously, κ decreases as ρ increases. This is expected because the steep kernel with a larger ρ is smaller than that with a smaller ρ. Therefore, the steep kernel can deliver an LRV-based estimator $\widehat{\beta}_{LRV}$ that is more efficient than $\widehat{\beta}_{LPS,c}$, as long as the steepness index is greater than 1.

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

$$N\left(b,\ (\Omega_{xx}^{-1}\otimes I_{m_y})\Theta_{LRV}(\Omega_{xx}^{-1}\otimes I_{m_y}\right), \tag{4.18}$$

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

$$-2\pi\alpha(\Omega_{xx}^{-1}\otimes I_{m_y})vec(Ef^{(1)}_{U_{y_i}U_{x_i}}-\beta Ef^{(1)}_{U_{x_i}U_{x_i}})+O_p\left(k_2\sqrt{n}/T^2\right). \tag{4.19}$$

The $O_p(\cdot)$ vanishes when $(n,T\to\infty)$ such that $\sqrt{n}/T\to\alpha$. But in finite samples, the $O_p(\cdot)$ 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.

4.3 Hypothesis Testing

The asymptotic theory developed above allows us to test hypotheses about the long-run average relationship. To examine the existence of a nonlinear (or linear) relationship about the components of the long-run coefficient, we can perform the usual Wald test. Specifically, the null hypothesis is $H_0:\psi(\beta)=0$, 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:

$$W_\psi=n\psi(\widehat{\beta}_{LRV})\widehat{V}_\psi^{-1}\psi(\widehat{\beta}_{LRV}), \tag{4.20}$$

where

$$\widehat{V}_\psi = \partial\psi(\widehat{\beta}_{LRV})/\partial\beta'\widehat{V}_\beta^{-1}\partial\psi(\widehat{\beta}_{LRV})/\partial\beta \tag{4.21}$$

$$\widehat{V}_\beta = \left(\widehat{\Omega}_{xx}^{-1}\otimes I_{m_y}\right)\widehat{\Theta}_{LRV}\left(\widehat{\Omega}_{xx}^{-1}\otimes I_{m_y}\right) \tag{4.22}$$

and

$$\widehat{\Theta}_{LRV}=\frac{1}{n}\sum_{i=1}^{n}\text{vec}\left(\widehat{\Omega}_{yxi}-\widehat{\beta}_{LRV}\widehat{\Omega}_{xxi}\right)\text{vec}\left(\widehat{\Omega}_{yxi}-\widehat{\beta}_{LRV}\widehat{\Omega}_{xxi}\right)'. \tag{4.23}$$

Some simple manipulations show that this test statistic converges to a χ^2 random variable under both the sequential and joint limits. The details are omitted here.

Theorem 10 *Let Assumptions 1 – 4 hold, Then under H_0, $W_\psi\Rightarrow\chi^2_p$ as $(T,n\to\infty)_{seq}$ or $(T,n\to\infty)$ with $\sqrt{n}/T\to 0$.*

5 Finite Sample Performances

In this section, we investigate the finite sample performances of the LRV-based estimators. The data generating process is

$$\begin{pmatrix} \Delta Y_{i,t} \\ \Delta X_{i,t} \end{pmatrix} = U_{i,t} := \begin{pmatrix} u_{y,it} \\ u_{x,it} \end{pmatrix} \tag{5.24}$$

with $U_{i,t}$ following a VAR(1) process:

$$U_{i,t} = A_i U_{i,t-1} + V_{i,t}, \tag{5.25}$$

where A_i is a 2×2 matrix and $V_{i,t}$ is a 2×1 random vector that is iid across i and t with distribution $N(0, I_2)$, i.e.

$$V_{i,t} = \begin{pmatrix} v_{y,it} \\ v_{x,it} \end{pmatrix} \overset{iid}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}. \tag{5.26}$$

For simplicity, we assume that there is no cross sectional heterogeneity in VAR(1) coefficients so that $A_i = A$ for all i. In addition, we assume A is symmetric so that

$$A = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \tag{5.27}$$

for some scalars a and b. It is easy to see that the eigenvalues of matrix A are $\lambda_1 = a+b$ and $\lambda_2 = a - b$. We choose a and b such that $\lambda_1, \lambda_2 \in (0,1)$, i.e.

$$(a,b) \in \mathcal{A} = \{(a,b) : 0 < a+b < 1 \text{ and } 0 < a-b < 1\}. \tag{5.28}$$

The process $U_{i,t}$ as defined in (5.25)–(5.28) is stationary and exhibits some persistence. Note that the set $\mathcal{A}$ is a square, which is divided into four triangles by the two diagonals. The parameter configurations we consider are the centers of these four triangles and the center of the square itself. More specifically, we consider the following constellation of parameter configurations:

$$(a,b) = \left(\frac{2}{3}, \frac{1}{6}\right), \left(\frac{1}{3}, \frac{1}{6}\right), \left(\frac{2}{3}, -\frac{1}{6}\right), \left(\frac{1}{3}, -\frac{1}{6}\right), \left(\frac{1}{2}, 0\right). \tag{5.29}$$

For the above (a,b) values, the true long run average slope coefficients are $\beta = 4/5, 8/17, -4/5, -8/17$ and 0, respectively.

In the Monte-Carlo simulations, the initial values were set to zero whenever they were called for. The data were generated by creating $T+100$ observations and discarding the first 100 observations to alleviate the effect of initialization. We examine different (N,T) combinations and consider three sets of estimators: $\widehat{\beta}_{PLS}, \widehat{\beta}_{PLS,c}$; $\widehat{\beta}_{b1}, \widehat{\beta}_{b2}, \widehat{\beta}_{b4}$, the LRV-based estimators with sharp kernels ($\rho = 1, 2$ and 4), and $\widehat{\beta}_{p1}, \widehat{\beta}_{p2}, \widehat{\beta}_{p4}$, the LRV-based estimators with steep kernels ($\rho = 1, 2$ and 4). For each (N,T) and (a,b) combination, we calculate the bias, standard error and root mean squared error (RMSE) of each estimator based on 5000 replications.

Tables 1–3 report the simulation results when $(a, b) = (2/3, 1/6)$. We first consider the case $N = T$. As is apparent from Table 1, the absolute biases and standard errors of all the estimators decrease as N and T increase. As a consequence, the RMSE's of all estimators decrease as N and T increase. For any given (N, T), the biases of the LRV-based estimators with sharp and steep kernels increase with the power parameter ρ, while the standard errors decrease with the power parameter ρ. This pattern is consistent with our asymptotic theory in Section 4.2. Table 1 also reveals that $\widehat{\beta}_{p2}$ dominates $\widehat{\beta}_{PLS,c}$ based on the RMSE criterion. The dominance of $\widehat{\beta}_{p2}$ over $\widehat{\beta}_{PLS}$ is also obvious for a large T ($T \geq 25$). In addition, for a given power parameter ρ, the LRV-based estimator with the steep kernel ($\widehat{\beta}_{p\rho}$) dominates that with the sharp kernel ($\widehat{\beta}_{b\rho}$).

Next, we consider the case that N is fixed at 25 and T is allowed to take various values (Table 2). Again, for all estimators considered, the absolute bias and standard error decrease as T increases. The former is consistent with the asymptotic result that the asymptotic bias is of order $O(1/T)$. The latter may be explained by the fact that the time series estimate of the LRV becomes less variable as T increases. The decrease of the standard error is accentuated as T continues to increase, reflecting that the asymptotic variance is independent of T. As in Table 1, the biases of $\widehat{\beta}_{b\rho}$ and $\widehat{\beta}_{p\rho}$ increase with ρ while their standard errors decrease with ρ. For a given power index ρ, the RMSE of $\widehat{\beta}_{p\rho}$ is uniformly smaller than that of $\widehat{\beta}_{b\rho}$ over different values of T. In general, for both $\widehat{\beta}_{p\rho}$ and $\widehat{\beta}_{b\rho}$, a larger power parameter ρ may be employed for larger T. It is also clear from Table 2 that $\widehat{\beta}_{p2}$ has a smaller RMSE than $\widehat{\beta}_{PLS,c}$ for all T and $\widehat{\beta}_{PLS}$ for $T > 25$.

Finally, we turn to the case that T is fixed at 25 and N is allowed to change. Table 3 shows that for each estimator the bias is not sensitive to the width of the cross sectional dimension. This is expected because the asymptotic theory shows that the asymptotic bias depends only on T but not on N. It is not surprising that the standard error is decreasing in N as it is of order $1/\sqrt{N}$, according to the asymptotic distributions. Comparing different estimators, we find that $\widehat{\beta}_{PLS}$ is superior to other estimators in terms of the RMSE.

The observations made above apply to the case $(a, b) = (2/3, -1/6)$. To save space, we do not reproduce the results for this case. The results for the two cases $(a, b) = (1/3, 1/6)$ and $(a, b) = (1/3, -1/6)$ are similar. For brevity, we report only the RMSE's for the case $(a, b) = (1/3, 1/6)$, since they are representative of the results found in the case $(a, b) = (1/3, -1/6)$. Table 4 contains the results. We draw attention to three aspects of this table. First, when $N = T$, the RMSE's of $\widehat{\beta}_{b\rho}$ and $\widehat{\beta}_{p\rho}$ decrease as ρ increases. This is because the biases (not reported) are on average much smaller than those in Table 1, and the standard error components determine the relative magnitudes of the RMSE's. It is clear from the table that $\widehat{\beta}_{p4}$ has a smaller RMSE than $\widehat{\beta}_{b4}$, which in turn has a smaller RMSE than the rest of the estimators. Second, when N is fixed, the performances of $\widehat{\beta}_{b\rho}$ and $\widehat{\beta}_{p\rho}$ improve as ρ increases, just as in the case $N = T$. When T is less than 50, $\widehat{\beta}_{p4}$ outperforms other estimators while when T is greater than 50, $\widehat{\beta}_{b4}$ outperforms others. Regardless of the time series dimension, $\widehat{\beta}_{b2}$, $\widehat{\beta}_{b4}$, $\widehat{\beta}_{p2}$, $\widehat{\beta}_{p4}$ have smaller RMSE's than $\widehat{\beta}_{PLS}$ and $\widehat{\beta}_{PLS,c}$. Third, when T is fixed at 25, $\widehat{\beta}_{p2}$ is the best estimator, having a smaller RMSE than any other estimators. This is in contrast with the results in Table 3 where $\widehat{\beta}_{PLS}$ is the best estimator.

We now consider the case $(a, b) = (1/2, 0)$. Table 5 reports the RMSE's. The table shows that either $\widehat{\beta}_{p4}$ or $\widehat{\beta}_{b4}$ has the smallest RMSE for all (N, T) combinations considered. The reason is that the biases are very small for this (a, b) configuration. As a consequence, variance reduction outweighs bias inflation for the LRV-based estimators and a large power parameter is justified. Note that for all (N, T) combinations, $\widehat{\beta}_{b2}$, $\widehat{\beta}_{b4}$, $\widehat{\beta}_{p2}$, $\widehat{\beta}_{p4}$ have smaller RMSE's than $\widehat{\beta}_{PLS}$ and $\widehat{\beta}_{PLS,c}$.

To sum up, the simulation results reveal that the steep kernel with power parameter 2 or 4 delivers estimates that have the best performance in an overall sense. Comparing with the existing estimators $\widehat{\beta}_{PLS}$ and $\widehat{\beta}_{PLS,c}$, the estimator $\widehat{\beta}_{p2}$ has a smaller RMSE for almost all the parameter configurations and (N, T) combinations. The only possible exception is when T is small ($T \leq 25$).

6 Conclusion

In this paper, we have proposed a unified framework for the estimation of the long-run average relationship. This framework includes the panel PLS estimators and the limiting CLS estimator, as well as a class of new estimators. We show that the new estimators are 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 sharp kernels and steep kernels introduced in this paper, the new estimators are asymptotically more efficient than the existing ones. A simulation study shows that steep kernel with the exponent parameter 2 or 4 produces estimators that dominate the PLS estimators for various (N, T) combinations.

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 approach can be used to estimate the long-run average relationship when there is a cointegrating relationship in the time series regression. This problem is currently being investigated and the results will be reported in a future paper. Finally, 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. Kao (1999) used the least squares dummy variable (LSDV) estimator to construct the residuals. The LSDV estimator is the same as the PLS estimator when intercepts are included in the time series regressions. Since the LRV-based estimator is more efficient than the LSDV estimator, the test using the LRV-based residuals may have better power properties.

Table 1: Finite Sample Performances of LRC-based Estimators
with N=T, (a,b)=(2/3,1/6), and 5000 Replications

N&T	$\widehat{\beta}_{PLS}$	$\widehat{\beta}_{PLS,c}$	$\widehat{\beta}_{b1}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{b4}$	$\widehat{\beta}_{p1}$	$\widehat{\beta}_{p2}$	$\widehat{\beta}_{p4}$
				Bias				
25	-0.0120	-0.0898	-0.0663	-0.0897	-0.1272	-0.0749	-0.0978	-0.1284
50	-0.0112	-0.0469	-0.0325	-0.0459	-0.0699	-0.0319	-0.0434	-0.0614
75	-0.0091	-0.0302	-0.0206	-0.0302	-0.0477	-0.0184	-0.0252	-0.0365
100	-0.0083	-0.0236	-0.0168	-0.0239	-0.0372	-0.0140	-0.0182	-0.0256
125	-0.0077	-0.0186	-0.0131	-0.0188	-0.0297	-0.0103	-0.0131	-0.0183
150	-0.0059	-0.0150	-0.0104	-0.0152	-0.0246	-0.0076	-0.0098	-0.0139
				Standard Error				
25	0.1259	0.1109	0.1096	0.0991	0.0892	0.1019	0.0942	0.0875
50	0.0843	0.0670	0.0687	0.0603	0.0523	0.0633	0.0570	0.0516
75	0.0674	0.0515	0.0533	0.0457	0.0386	0.0489	0.0433	0.0385
100	0.0552	0.0432	0.0457	0.0389	0.0325	0.0421	0.0371	0.0326
125	0.0494	0.0381	0.0408	0.0345	0.0285	0.0376	0.0331	0.0290
150	0.0437	0.0335	0.0364	0.0305	0.0249	0.0336	0.0293	0.0254
				Root Mean Squared Error				
25	0.1265	0.1427	0.1281	0.1337	0.1553	0.1265	0.1358	0.1554
50	0.0850	0.0818	0.0760	0.0758	0.0874	0.0709	0.0717	0.0802
75	0.0680	0.0597	0.0571	0.0548	0.0614	0.0523	0.0501	0.0531
100	0.0559	0.0492	0.0487	0.0457	0.0494	0.0444	0.0413	0.0415
125	0.0500	0.0424	0.0428	0.0393	0.0412	0.0390	0.0356	0.0343
150	0.0441	0.0367	0.0378	0.0341	0.0350	0.0344	0.0309	0.0290

Table 2: Finite Sample Performances of LRC-based Estimators with $N = 25, (a, b) = (2/3, 1/6)$, and 5000 Replications

T	$\widehat{\beta}_{PLS}$	$\widehat{\beta}_{PLS,c}$	$\widehat{\beta}_{b1}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{b4}$	$\widehat{\beta}_{p1}$	$\widehat{\beta}_{p2}$	$\widehat{\beta}_{p4}$
				Bias				
25	-0.0120	-0.0898	-0.0663	-0.0897	-0.1272	-0.0749	-0.0978	-0.1284
50	-0.0105	-0.0455	-0.0308	-0.0446	-0.0690	-0.0304	-0.0421	-0.0602
75	-0.0119	-0.0320	-0.0219	-0.0312	-0.0482	-0.0194	-0.0257	-0.0366
100	-0.0063	-0.0209	-0.0139	-0.0213	-0.0352	-0.0109	-0.0153	-0.0232
125	-0.0075	-0.0201	-0.0139	-0.0195	-0.0302	-0.0108	-0.0134	-0.0185
150	-0.0070	-0.0146	-0.0100	-0.0148	-0.0243	-0.0069	-0.0091	-0.0133
				Standard Error				
25	0.1259	0.1109	0.1096	0.0991	0.0892	0.1019	0.0942	0.0875
50	0.0843	0.0670	0.0687	0.0603	0.0523	0.0633	0.0570	0.0516
75	0.0674	0.0515	0.0533	0.0457	0.0386	0.0489	0.0433	0.0385
100	0.0552	0.0432	0.0457	0.0389	0.0325	0.0421	0.0371	0.0326
125	0.0494	0.0381	0.0408	0.0345	0.0285	0.0376	0.0331	0.0290
150	0.0437	0.0335	0.0364	0.0305	0.0249	0.0336	0.0293	0.0254
				Root Mean Squared Error				
25	0.1265	0.1427	0.1281	0.1337	0.1553	0.1265	0.1358	0.1554
50	0.1226	0.1063	0.1037	0.0973	0.1016	0.0959	0.0918	0.0950
75	0.1182	0.0948	0.0964	0.0861	0.0831	0.0880	0.0800	0.0765
100	0.1143	0.0900	0.0938	0.0820	0.0749	0.0867	0.0775	0.0708
125	0.1143	0.0873	0.0925	0.0791	0.0695	0.0851	0.0747	0.0662
150	0.1104	0.0861	0.0927	0.0787	0.0676	0.0857	0.0750	0.0659

Table 3: Finite Sample Performances of LRC-based Estimators with $T = 25, (a, b) = (2/3, 1/6)$, and 5000 replications

N	$\widehat{\beta}_{PLS}$	$\widehat{\beta}_{PLS,c}$	$\widehat{\beta}_{b1}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{p1}$	$\widehat{\beta}_{p2}$	$\widehat{\beta}_{p4}$
				Bias				
25	-0.0120	-0.0898	-0.0663	-0.0897	-0.1272	-0.0749	-0.0978	-0.1284
50	-0.0109	-0.0870	-0.0638	-0.0868	-0.1239	-0.0721	-0.0947	-0.1251
75	-0.0108	-0.0881	-0.0655	-0.0881	-0.1248	-0.0736	-0.0959	-0.1259
100	-0.0101	-0.0866	-0.0635	-0.0863	-0.1232	-0.0718	-0.0942	-0.1245
125	-0.0111	-0.0874	-0.0645	-0.0872	-0.1241	-0.0728	-0.0952	-0.1255
150	-0.0100	-0.0876	-0.0643	-0.0870	-0.1239	-0.0726	-0.0950	-0.1252
				Standard Error				
25	0.1259	0.1109	0.1096	0.0991	0.0892	0.1019	0.0942	0.0875
50	0.0868	0.0774	0.0772	0.0699	0.0627	0.0717	0.0662	0.0614
75	0.0710	0.0617	0.0611	0.0554	0.0500	0.0569	0.0527	0.0490
100	0.0617	0.0544	0.0540	0.0490	0.0442	0.0503	0.0466	0.0433
125	0.0534	0.0486	0.0483	0.0438	0.0394	0.0449	0.0415	0.0386
150	0.0487	0.0441	0.0433	0.0393	0.0355	0.0403	0.0374	0.0348
				Root Mean Squared Error				
25	0.1265	0.1427	0.1281	0.1337	0.1553	0.1265	0.1358	0.1554
50	0.0875	0.1165	0.1002	0.1114	0.1389	0.1017	0.1155	0.1394
75	0.0718	0.1076	0.0896	0.1041	0.1344	0.0931	0.1094	0.1351
100	0.0625	0.1022	0.0833	0.0992	0.1308	0.0877	0.1051	0.1318
125	0.0545	0.1000	0.0806	0.0976	0.1302	0.0856	0.1039	0.1313
150	0.0497	0.0980	0.0775	0.0955	0.1288	0.0831	0.1021	0.1299

Table 4: RMSE's of LRC-based Estimators
with (a,b)=(1/3,1/6) and 5000 Replications

	$\widehat{\beta}_{PLS}$	$\widehat{\beta}_{PLS,c}$	$\widehat{\beta}_{b1}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{b4}$	$\widehat{\beta}_{p1}$	$\widehat{\beta}_{p2}$	$\widehat{\beta}_{p4}$
N&T				Root Mean Squared Error (N=T)				
25	0.1804	0.1347	0.1375	0.1221	0.1183	0.1267	0.1145	0.1081
50	0.1203	0.0872	0.0930	0.0799	0.0724	0.0860	0.0757	0.0674
75	0.0963	0.0691	0.0741	0.0624	0.0542	0.0688	0.0598	0.0521
100	0.0793	0.0589	0.0644	0.0540	0.0460	0.0600	0.0521	0.0450
125	0.0708	0.0520	0.0569	0.0475	0.0399	0.0532	0.0464	0.0401
150	0.0635	0.0467	0.0520	0.0430	0.0355	0.0485	0.0419	0.0359
T				Root Mean Squared Error (N=25)				
25	0.1804	0.1347	0.1375	0.1221	0.1183	0.1267	0.1145	0.1081
50	0.1726	0.1201	0.1309	0.1103	0.0944	0.1217	0.1063	0.0930
75	0.1671	0.1167	0.1289	0.1064	0.0869	0.1196	0.1032	0.0889
100	0.1638	0.1167	0.1290	0.1069	0.0862	0.1216	0.1059	0.0913
125	0.1655	0.1143	0.1269	0.1036	0.0820	0.1189	0.1023	0.0872
150	0.1597	0.1164	0.1298	0.1064	0.0840	0.1221	0.1055	0.0903
N				Root Mean Squared Error (T=25)				
25	0.1804	0.1347	0.1375	0.1221	0.1183	0.1267	0.1145	0.1081
50	0.1254	0.0991	0.0994	0.0917	0.0969	0.0909	0.0838	0.0831
75	0.1015	0.0854	0.0824	0.0792	0.0894	0.0751	0.0711	0.0735
100	0.0873	0.0770	0.0729	0.0718	0.0845	0.0665	0.0640	0.0680
125	0.0777	0.0723	0.0671	0.0680	0.0829	0.0609	0.0596	0.0654
150	0.0701	0.0684	0.0620	0.0644	0.0808	0.0560	0.0559	0.0627

Table 5: RMSE's of LRC-based Estimators
with $(a, b) = (1/2, 0)$ and 5000 Replications

	$\widehat{\beta}_{PLS}$	$\widehat{\beta}_{PLS,c}$	$\widehat{\beta}_{b1}$	$\widehat{\beta}_{b2}$	$\widehat{\beta}_{b4}$	$\widehat{\beta}_{p1}$	$\widehat{\beta}_{p2}$	$\widehat{\beta}_{p4}$
N&T				Root Mean Squared Error (N=T)				
25	0.2016	0.1433	0.1504	0.1277	0.1059	0.1386	0.1219	0.1068
50	0.1352	0.0955	0.1044	0.0869	0.0702	0.0966	0.0841	0.0728
75	0.1081	0.0778	0.0844	0.0699	0.0557	0.0786	0.0682	0.0587
100	0.0894	0.0655	0.0725	0.0597	0.0471	0.0675	0.0584	0.0500
125	0.0793	0.0580	0.0636	0.0525	0.0415	0.0597	0.0519	0.0445
150	0.0722	0.0526	0.0588	0.0482	0.0376	0.0550	0.0474	0.0404
T				Root Mean Squared Error (N=25)				
25	0.2016	0.1433	0.1504	0.1277	0.1059	0.1386	0.1219	0.1068
50	0.1928	0.1341	0.1473	0.1230	0.0996	0.1373	0.1198	0.1036
75	0.1863	0.1328	0.1469	0.1208	0.0961	0.1364	0.1178	0.1012
100	0.1855	0.1324	0.1467	0.1210	0.0959	0.1378	0.1196	0.1025
125	0.1861	0.1290	0.1428	0.1165	0.0914	0.1334	0.1150	0.0983
150	0.1820	0.1319	0.1454	0.1195	0.0941	0.1369	0.1187	0.1017
N				Root Mean Squared Error (T=25)				
25	0.2016	0.1433	0.1504	0.1277	0.1059	0.1386	0.1219	0.1068
50	0.1421	0.1011	0.1071	0.0909	0.0751	0.0986	0.0865	0.0756
75	0.1138	0.0828	0.0869	0.0736	0.0607	0.0798	0.0699	0.0609
100	0.0976	0.0721	0.0757	0.0647	0.0537	0.0701	0.0617	0.0539
125	0.0890	0.0635	0.0670	0.0571	0.0472	0.0618	0.0543	0.0473
150	0.0798	0.0579	0.0604	0.0515	0.0428	0.0557	0.0491	0.0430

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 $\sum_{s=0}^{\infty} \left[E\left(\left\| \tilde{C}_{i,s} \right\|^4 \right) \right]^{1/4}$, it suffices to consider $\sum_{p=0}^{\infty} \left[E\left(\tilde{C}_{a,i,p}^4 \right) \right]^{1/4}$. 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(ii). 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 $E \sum_{j=0}^{\infty} j^2 \|\Gamma_i(j)\|$ as

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

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

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

where the last line follows from (7.1) and Assumption 2(i). 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 $E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right) - \Omega_i$ into three terms as follows:

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

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

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

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 $\Omega_i^e = (\Omega_{i1}^e + \Omega_{i2}^e + \Omega_{i3}^e)$, then we have shown that, when $q^* = 1$, $\lim_{T\to\infty} TE\Omega_i^e = -2\pi (k_1+1) Ef_{U_iU_i}^{(1)}$ 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 $K(s,t) = 1-(s\vee t)$. Using $EU_i(s)U_i'(t) = \Gamma_i(s-t)$, we have

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

$$= \frac{1}{T}\sum_{s=1}^{T}\sum_{j=s-T}^{0}(1-\frac{s\vee (s-j)}{T})\Gamma_i(j) + \frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{s-1}(1-\frac{s\vee (s-j}{T})\Gamma_i(j)$$

$$= \frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{T-s}(1-\frac{s+j}{T})\Gamma_i(-j) + \frac{1}{T}\sum_{p=1}^{T}\sum_{j=1}^{s-1}(1-\frac{s-j}{T})\Gamma_i(j)$$

$$= \frac{1}{T}\sum_{j=0}^{T-1}\sum_{s=1}^{T-j}(1-\frac{s+j}{T})\Gamma_i(-j) + \frac{1}{T}\sum_{j=1}^{T-1}\sum_{s=j+1}^{T}(1-\frac{s-j}{T})\Gamma(j)$$

$$= \sum_{j=0}^{T-1}\frac{1}{2}\frac{T^2-2jT+j^2+j-T}{T^2}\Gamma_i(-j) + \sum_{j=1}^{T-1}(\frac{1}{2}\frac{T^2+j-T-j^2}{T^2})\Gamma_i(j). \tag{7.4}$$

Therefore $E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right) - 1/2\Omega_i$ is

$$1/2\sum_{j=0}^{T-1}\left(-2j/T + (j/T)^2 + j/T^2 - 1/T\right)\Gamma_i(-j)$$
$$+1/2\sum_{j=1}^{T-1}\left(j/T^2 - (j/T)^2 - 1/T\right)\Gamma_i(j) - \sum_{|j|\geq T}\Gamma_i(j). \tag{7.5}$$

From the above equation and Lemma 2, it is easy to see that

$$E\left[E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right) - 1/2\Omega_i\right] = O(1/T). \tag{7.6}$$

We now consider the case $K(s,t) = \{s \leq r_0, t \leq r_0\}$. Note that

$$E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right) - r_0\Omega_i = \frac{1}{T}\sum_{s=1}^{[Tr_0]}\sum_{t=1}^{[Tr_0]}\Gamma_i(s-t) - r_0\Omega_i$$

$$= r_0\left(\frac{[Tr_0]}{Tr_0}\sum_{j=-[Tr_0]+1}^{[Tr_0]-1}\left(1-\frac{Tr_0}{[Tr_0]}-\frac{|j|}{[Tr_0]}\right)\Gamma_i(j)\right) - \sum_{|j|>[Tr_0]}\Gamma_i(j). \tag{7.7}$$

Again, it is easy to see that $E\left[E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right)-r_0\Omega_i\right]=O(1/T)$.

Finally, for the case $K(s,t)=\min(s,t)-st$, $E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right)-1/6\Omega_i$ is

$$
\begin{aligned}
& \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(T^{-1}\min(s,t)-T^{-2}st\right)\Gamma_i(s-t)-1/6\Omega_i \\
= & \frac{1}{T}\sum_{s=1}^{T}\sum_{j=s-T}^{0}\left(T^{-1}\min(s,s-j)-T^{-2}s(s-j)\right)\Gamma_i(j) \\
& +\frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{s-1}\left(T^{-1}\min(s,s-j)-T^{-2}s(s-j)\right)\Gamma_i(j)-1/6\Omega_i \\
= & \frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{T-s}\left[T^{-1}s-T^{-2}s(s+j)\right]\Gamma_i(-j) \\
& +\frac{1}{T}\sum_{p=1}^{T}\sum_{j=1}^{s-1}\left[T^{-1}(s-j)-T^{-2}s(s-j)\right]\Gamma_i(j)-1/6\Omega_i \\
= & \frac{1}{T}\sum_{j=0}^{T-1}\sum_{s=1}^{T-j}\left(T^{-1}s-T^{-2}s(s+j)\right)\Gamma_i(-j) \\
& +\frac{1}{T}\sum_{j=1}^{T-1}\sum_{s=j+1}^{T}\left(T^{-1}(s-j)-T^{-2}s(s-j)\right)\Gamma(j)-1/6\Omega_i \\
= & \sum_{j=0}^{T-1}-\frac{1}{6}\frac{-j-3Tj^2+j^3+3jT^2+T}{T^3}\Gamma_i(-j) \\
& +\sum_{j=1}^{T-1}-\frac{1}{6}\frac{-j-3Tj^2+j^3+3jT^2+T}{T^3}\Gamma_i(j)+\sum_{|j|\geq T}\Gamma_i(j). \qquad (7.8)
\end{aligned}
$$

It follows from the above equation and Lemma 2 that

$$E\left[E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right)-1/6\Omega_i\right]=O(1/T). \qquad (7.9)$$

The proof of the theorem is completed by noting that $\int_0^1 k(0)ds=\int_0^1 ds=1$, $\int_0^1(1-(s\vee s))\,ds=1/2$, $\int_0^1\{s\leq r_0,s\leq r_0\}\,ds=r_0$ and $\int_0^1(\min(s,s)-s^2)ds=\int_0^1\left(s-s^2\right)ds=1/6$. ■

Proof of Lemma 4. Plugging the BN decomposition

$$U_{i,t}=C_i(1)V_{i,t}+\tilde{U}_{i,t-1}-\tilde{U}_{i,t} \qquad (7.10)$$

into

$$\widehat{\Omega}_i=\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}U_{i,t}K(\frac{t}{T},\frac{\tau}{T})U_{i,\tau}', \qquad (7.11)$$

we get

$$\widehat{\Omega}_i = \widetilde{\Omega}_i + R_i, \tag{7.12}$$

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

$$\begin{aligned}
R_{i1} &= \frac{1}{T} C_i(1) \sum_{t=1}^{T} \sum_{\tau=1}^{T} V_{i,t} K(\frac{t}{T}, \frac{\tau}{T}) \left(\tilde{U}_{i,\tau-1} - \tilde{U}_{i,\tau} \right)', \\
R_{i2} &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \left(\tilde{U}_{i,t-1} - \tilde{U}_{i,t} \right) K(\frac{t}{T}, \frac{\tau}{T}) V'_{i,\tau} C'_i(1) = R'_{i1}, \\
R_{i3} &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \left(\tilde{U}_{i,t-1} - \tilde{U}_{i,t} \right) K(\frac{t}{T}, \frac{\tau}{T}) \left(\tilde{U}_{i,\tau-1} - \tilde{U}_{i,\tau} \right)'.
\end{aligned}$$

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

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} C_i(1) V_{i,t} K(\frac{t}{T}, \frac{1}{T}) \tilde{U}'_{i,0} - \frac{1}{T} \sum_{t=1}^{T} C_i(1) V_{i,t} K(\frac{t}{T}, \frac{T}{T}) \tilde{U}'_{iT} \\
& + \frac{1}{T} \sum_{t=1}^{T} C_i(1) V_{i,t} \sum_{\tau=1}^{T-1} \left(K(\frac{t}{T}, \frac{\tau+1}{T}) - K(\frac{t}{T}, \frac{\tau}{T}) \right) \tilde{U}'_{i,\tau} \qquad (7.13) \\
: & \equiv R_{i1}^{(1)} + R_{i1}^{(2)} + R_{i1}^{(3)}, \text{ say.}
\end{aligned}$$

But $Etr\left(\text{vec}\left(R_{i1}^{(1)}\right)\text{vec}\left(R_{i1}^{(1)}\right)'\right)$ is

$$\begin{aligned}
& \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} K(\frac{t}{T}, \frac{1}{T}) K(\frac{s}{T}, \frac{1}{T}) Etr \left(\text{vec}\left(C_i(1) V_{i,t} \tilde{U}'_{i,0} \right) \text{vec}\left(C_i(1) V_{i,s} \tilde{U}'_{i,0} \right)' \right) \\
= & \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} K(\frac{t}{T}, \frac{1}{T}) K(\frac{s}{T}, \frac{1}{T}) Etr \left(\left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,t} \right) \left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,s} \right)' \right) \\
= & \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} K(\frac{t}{T}, \frac{1}{T}) K(\frac{s}{T}, \frac{1}{T}) Etr \left(\left(\tilde{U}_{i,0} \otimes C_i(1) V_{i,t} \right) \left(\tilde{U}'_{i,0} \otimes V'_{i,s} C'_i(1) \right) \right) \\
= & \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} K(\frac{t}{T}, \frac{1}{T}) K(\frac{s}{T}, \frac{1}{T}) Etr \left(\tilde{U}_{i,0} \tilde{U}'_{i,0} \otimes C_i(1) V_{i,t} V'_{i,s} C'_i(1) \right) \qquad (7.14)
\end{aligned}$$

where the first equality follows from the fact that for $m \times 1$ vectors A and B, $\text{vec}(AB') = B \otimes A$, the third equality follow from the rules 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 $Etr\left(\text{vec}\left(R_{i1}^{(1)}\right)\text{vec}\left(R_{i1}^{(1)}\right)'\right)$ as

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\tilde{U}_{i,0}\tilde{U}_{i,0}'\otimes C_i(1)C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\tilde{U}_{i,0}\tilde{U}_{i,0}'\right)tr\left(C_i(1)C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{1}{T})E\left\|\tilde{U}_{i,0}\right\|^2\left\|C_i(1)\right\|^2 \\
\leq & \frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{1}{T})\left(E\left\|\tilde{U}_{i,0}\right\|^4\right)^{1/2}E\left(\left\|C_i(1)\right\|^4\right)^{1/2} \\
= & \frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{1}{T})O(1)=O(\frac{1}{T}), \qquad (7.15)
\end{aligned}$$

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

For notational simplicity, let

$$K_T(t,s)=K(\frac{t}{T},\frac{1}{T})K(\frac{s}{T},\frac{1}{T}). \qquad (7.16)$$

Then $Etr\left(\text{vec}\left(R_{i1}^{(2)}\right)\text{vec}\left(R_{i1}^{(2)}\right)'\right)$ is equal to

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t,s=1}^{T}K_T(t,s)Etr\left(\tilde{U}_{iT}\tilde{U}_{iT}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{p,q=0}^{\infty}K_T(t,s)Etr\left(\tilde{C}_{i,p}V_{i,T-p}V_{i,T-q}'\tilde{C}_{i,q}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right) \\
= & \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{p,q=T}^{\infty}K_T(t,s)Etr\left(\tilde{C}_{i,p}V_{i,T-p}V_{i,T-q}'\tilde{C}_{i,q}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right) \\
& +\frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{p,q=0}^{T-1}K_T(t,s)Etr\left(\tilde{C}_{i,p}V_{i,T-p}V_{i,T-q}'\tilde{C}_{i,q}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right) \\
= & J_1+J_2 \qquad (7.17)
\end{aligned}$$

where

$$\begin{aligned}
J_1 &= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{p,q=0}^{\infty}K_T(t,s)Etr\left(\tilde{C}_{i,p-T}V_{i,-p}V_{i,-q}'\tilde{C}_{i,q-T}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right), \\
J_2 &= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{p,q=1}^{T}K_T(t,s)Etr\left(\tilde{C}_{i,T-p}V_{i,p}V_{i,q}'\tilde{C}_{i,T-q}'\otimes C_i(1)V_{i,t}V_{i,s}'C_i'(1)\right).
\end{aligned}$$

It is easy to show that J_1 is

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=}^{\infty}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\tilde{C}_{i,p-T}\tilde{C}'_{i,p-T}\otimes C_i(1)C'_i(1)\right) \\
\leq & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=0}^{\infty}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\left(\tilde{C}_{i,p-T}\tilde{C}'_{i,p-T}\right)tr\left(C_i(1)C'_i(1)\right)\right) \\
\leq & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=0}^{\infty}K^2(\frac{t}{T},\frac{1}{T})\left[E\left(\left\|\tilde{C}_{i,p-T}\right\|^4\right)\right]^{1/2}\left[E\left(\|C_i(1)\|^4\right)\right]^{1/2} \\
\leq & \frac{M}{T}\sum_{p=0}^{\infty}\left[E\left(\left\|\tilde{C}_{i,p}\right\|^4\right)\right]^{1/2} = O_p(T^{-1}) \qquad (7.18)
\end{aligned}$$

by Lemma 1(e), where M is a generic constant.

We now consider J_2. Note that

$$\begin{aligned}
& \tilde{C}_{i,T-p}V_{i,p}V'_{i,q}\tilde{C}'_{i,T-q}\otimes C_i(1)V_{i,t}V'_{i,s}C'_i(1) \\
= & \left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\left(V_{i,p}V'_{i,q}\otimes V_{i,t}V'_{i,s}\right)\left(\tilde{C}'_{i,T-q}\otimes C'_i(1)\right).
\end{aligned}$$

Some tedious calculations show that $E\left(V_{i,p}V'_{i,q}\otimes V_{i,t}V'_{i,s}\right)$ is

$$\begin{cases}
I_{m^2}, & \text{if } p=q\neq t=s \\
\text{vec}(I_m)\text{vec}(I_m)', & \text{if } p=t\neq q=s \\
K_{mm}, & \text{if } p=s\neq q=t \\
I_{m^2}+\text{vec}(I_m)\text{vec}(I_m)'+K_{mm}+\zeta\sum_{l=1}^{m}e_{ll}\otimes e_{ll} & \text{if } p=s=q=t
\end{cases}$$

where $\zeta=(v^4-3)$, e_{ll} is the $m\times m$ matrix with the $(l,l)^{th}$ element being one and the other elements being zeros. Then, in view of the independence between $\{\tilde{C}_{i,T-p}$, $C_i(1)\}$ and $V_{i,t}$, we can write J_2 as $J_{21}+J_{22}+J_{23}+J_{24}$ where

$$J_{21}=\frac{1}{T^2}\sum_{t,p=1}^{T}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\left(\tilde{C}'_{i,T-p}\otimes C'_i(1)\right)\right)$$

$$J_{22}=\frac{1}{T^2}\sum_{p,q=1}^{T}K_T(p,q)Etr\left(\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\text{vec}(I_m)\text{vec}(I_m)'\left(\tilde{C}'_{i,T-q}\otimes C'_i(1)\right)\right)$$

$$J_{23}=\frac{1}{T^2}\sum_{p,q=1}^{T}K_T(p,q)Etr\left(\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)K_{mm}\left(\tilde{C}'_{i,T-q}\otimes C'_i(1)\right)\right)$$

$$J_{24}=\frac{\zeta}{T^2}\sum_{p=1}^{T}K^2(\frac{p}{T},\frac{1}{T})Etr\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\sum_{l=1}^{m}(e_{ll}\otimes e_{ll})\left(\tilde{C}'_{i,T-p}\otimes C'_i(1)\right)$$

We now consider the above four terms one by one. For J_{21}, we have

$$\begin{aligned}
J_{21} &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=1}^{T}K^2(\frac{t}{T},\frac{1}{T})Etr\left(\left(\tilde{C}_{i,T-p}\tilde{C}'_{i,T-p}\right)\otimes\left(C_i(1)C_i'(1)\right)\right)\\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=1}^{T}K^2(\frac{t}{T},\frac{1}{T})E\left(tr\left(\tilde{C}_{i,T-p}\tilde{C}'_{i,T-p}\right)tr\left(C_i(1)C_i'(1)\right)\right)\\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{p=1}^{T}K^2(\frac{t}{T},\frac{1}{T})\left(E\left(\left\|\tilde{C}_{i,T-p}\right\|^4\right)E\left(\left\|C_i(1)\right\|^4\right)\right)^{1/2}\\
&\leq \frac{1}{T}\sum_{p=1}^{T}\left(E\left(\left\|\tilde{C}_{i,T-p}\right\|^4\right)\right)^{1/2}\left(E\left(\left\|C_i(1)\right\|^4\right)\right)^{1/2}\\
&\leq \frac{1}{T}\sum_{p=1}^{\infty}\left(E\left(\left\|\tilde{C}_{i,p}\right\|^4\right)\right)^{1/2} = O(1/T),
\end{aligned}$$

by Lemma 1(c) and (e).

For J_{22}, we have, using $tr(AB) = \text{vec}(A')'\text{vec}(B)$ and $tr(CD) \leq \|C\|\,\|D\|$ for square matrices C and D,

$$\begin{aligned}
J_{22} &= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}K_T(p,q)Etr\left(\text{vec}(\tilde{C}_{i,T-p}C_i'(1))\text{vec}(\tilde{C}_{i,T-q}C_i'(1))'\right)\\
&= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}K_T(p,q)E\left(\text{vec}\left(\tilde{C}_{i,T-p}C_i'(1)\right)'\text{vec}\left(\tilde{C}_{i,T-q}C_i'(1)\right)\right)\\
&= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}K_T(p,q)Etr\left(\tilde{C}'_{i,T-p}\tilde{C}_{i,T-q}C_i'(1)C_i(1)\right)\\
&\leq \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}K_T(p,q)E\left(\left\|\tilde{C}'_{i,T-p}\tilde{C}_{i,T-q}\right\|\left\|C_i(1)\right\|^2\right)\\
&= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T}K_T(p,q)\left(E\left\|\tilde{C}'_{i,T-p}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}_{i,T-q}\right\|^4\right)^{1/4}\left(E\left\|C_i(1)\right\|^4\right)^{1/2}\\
&\leq \frac{M}{T^2}\left(\sum_{p=0}^{\infty}\left(E\left\|\tilde{C}'_{i,p}\right\|^4\right)^{1/4}\right)^2 = O(1/T^2),
\end{aligned}$$

by Lemma 1(c) and (e).

For J_{23}, we have, using $tr\left((C\otimes D)K_{mm}\right) \leq \|C\|\,\|D\|$ for square matrices C and

D,

$$\begin{aligned}
J_{23} &= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} K_T(p,q) Etr\left(\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\left(\tilde{C}'_{i,T-q}\otimes C'_i(1)\right)K_{mm}\right) \\
&= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} K_T(p,q) Etr\left(\left(\tilde{C}_{iT-p}\tilde{C}'_{iT-q}\right)\otimes\left(C_i(1)C'_i(1)\right)K_{mm}\right) \\
&\leq \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} K_T(p,q) E\left(\left\|\tilde{C}_{iT-p}\tilde{C}'_{iT-q}\right\|\left\|C_i(1)\right\|^2\right) \\
&= \frac{1}{T^2}\sum_{p=1}^{T}\sum_{q=1}^{T} K_T(p,q)\left(E\left\|\tilde{C}_{i,T-p}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}_{i,T-q}\right\|^4\right)^{1/4}\left(E\left\|C_i(1)\right\|^4\right)^{1/2} \\
&= O(1/T^2).
\end{aligned}$$

Finally, $J_{24} = 0$ if $\zeta = 0$. Otherwise, $\zeta^{-1}J_{24}$ is

$$\begin{aligned}
& \frac{1}{T^2}\sum_{p=1}^{T}\sum_{l=1}^{m} K^2\left(\frac{p}{T},\frac{1}{T}\right) Etr\left(\left(\tilde{C}_{i,T-p}\otimes C_i(1)\right)\left(e_{ll}\otimes e_{ll}\right)\left(\tilde{C}'_{i,T-p}\otimes C'_i(1)\right)\right) \\
=\; & \frac{1}{T^2}\sum_{p=1}^{T}\sum_{l=1}^{m} K^2\left(\frac{p}{T},\frac{1}{T}\right) Etr\left(\tilde{C}_{i,T-p}e_{ll}\tilde{C}'_{i,T-p}\right)\otimes\left(C_i(1)e_{ll}C'_i(1)\right) \\
=\; & \frac{1}{T^2}\sum_{p=1}^{T}\sum_{l=1}^{m} K^2\left(\frac{p}{T},\frac{1}{T}\right)\left(E\left\|\tilde{C}_{i,T-p}e_{ii}\right\|^4 E\left\|C_i(1)e_{ll}\right\|^4\right)^{1/2} \\
\leq\; & \frac{1}{T^2}\sum_{p=1}^{T}\sum_{l=1}^{m} K^2\left(\frac{p}{T},\frac{1}{T}\right)\left(E\left\|\tilde{C}_{i,T-p}\right\|\left\|e_{ll}\right\|^4 E\left\|C_i(1)\right\|^4\left\|e_{ll}\right\|^4\right)^{1/2} \\
\leq\; & \frac{m}{T^2}\sum_{p=1}^{T} K^2\left(\frac{p}{T},\frac{1}{T}\right)\left(E\left\|\tilde{C}_{i,T-p}\right\|^4 E\left\|C_i(1)\right\|^4\right)^{1/2} = O(1/T^2),
\end{aligned}$$

where we have used $\|e_{ll}\| = 1$.

Combining the above results yields $Etr\left(\text{vec}\left(R_{i2}^{(2)}\right)\text{vec}\left(R_{i2}^{(2)}\right)'\right) = O(1/T)$.

Let

$$\Delta_T(t,\tau,p,q) = \left[K\left(\frac{t}{T},\frac{\tau+1}{T}\right) - K\left(\frac{t}{T},\frac{\tau}{T}\right)\right]\left[K\left(\frac{p}{T},\frac{q+1}{T}\right) - K\left(\frac{p}{T},\frac{q}{T}\right)\right], \tag{7.19}$$

then $Etr\left(vec\left(R_{i1}^{(3)}\right)\text{vec}\left(R_{i1}^{(3)}\right)'\right)$ can be written as

$$
\begin{aligned}
& Etr\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T-1}\sum_{p=1}^{T}\sum_{q=1}^{T-1}\Delta_T(t,\tau,p,q)\text{vec}(C_i(1)V_{i,t}\tilde{U}'_{i,\tau})\text{vec}(C'_i(1)V_{i,p}\tilde{U}'_{i,q})' \\
= & Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\left(\tilde{U}_{i,\tau}\tilde{U}'_{i,q}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1)\right) \\
= & Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k,j=0}^{\infty}\tilde{C}_{i,k}V_{i,\tau-k}V'_{i,q-j}\tilde{C}'_{i,j}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1) \\
= & Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=0}^{\tau-1}\sum_{j=0}^{q-1}\tilde{C}_{i,k}V_{i,\tau-k}V'_{i,q-j}\tilde{C}'_{i,j}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1) \\
& +Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=\tau}^{\infty}\sum_{j=q}^{\infty}\tilde{C}_{i,k}V_{i,\tau-k}V'_{i,q-j}\tilde{C}'_{i,j}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1) \\
= & H_1+H_2
\end{aligned}
$$

where

$$
\begin{aligned}
H_1 &= Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau}\sum_{j=1}^{q}\tilde{C}_{i,\tau-k}V_{i,k}V'_{i,j}\tilde{C}'_{i,q-j}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1), \\
H_2 &= Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta(t,\tau)\Delta(p,q)\sum_{k=\tau}^{\infty}\sum_{j=q}^{\infty}\tilde{C}_{i,k}V_{i,\tau-k}V'_{i,q-j}\tilde{C}'_{i,j}\otimes C_i(1)V_{i,t}V'_{i,p}C'_i(1).
\end{aligned}
$$

It is easy to see that H_2 is bounded by

$$
\begin{aligned}
& Etr\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau,q=1}^{T-1}|\Delta_T(t,\tau,t,q)|\sum_{k=0}^{\infty}\left(\tilde{C}_{i,\tau+k}\tilde{C}'_{i,q+k}\right)\otimes C_i(1)C'_i(1) \\
\leq & \frac{M}{T^2}\sum_{t=1}^{T}\sum_{\tau,q=1}^{T-1}|\Delta_T(t,\tau,t,q)|\sum_{k=0}^{\infty}\left(E\left\|\tilde{C}_{i,\tau+k}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}'_{i,q+k}\right\|^4\right)^{1/4} \\
\leq & \frac{M}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T-1}|[K(t/T,(\tau+1)/T)-K(t/T,\tau/T)]|\,O(1)=o(1),
\end{aligned}
$$

where the last line follows from the observation that the kernel function $K(\cdot,\cdot)$ is continuous almost surely.

Write H_1 as

$$
Etr\frac{1}{T^2}\sum_{t,p=1}^{T}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau}\sum_{j=1}^{q}\tilde{C}_{i,\tau-k}\otimes C_i(1)\left(V_{i,k}V'_{i,j}\otimes V_{i,t}V'_{i,p}\right)\tilde{C}'_{i,q-j}\otimes C'_i(1),
$$

which is the sum of four terms, say, $H_{11}+H_{12}+H_{13}+H_{14}$. Here H_{11} is

$$\begin{aligned}
& Etr\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T-1}\sum_{q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=0}^{\tau\wedge q}\left(\tilde{C}_{i,\tau+k}\otimes C_i(1)\right)\left(\tilde{C}'_{i,q+j}\otimes C'_i(1)\right) \\
= & Etr\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T-1}\sum_{q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=0}^{\tau\wedge q}\left(\tilde{C}_{i,\tau+k}\tilde{C}'_{i,q+j}\right)\otimes\left(C_i(1)C'_i(1)\right) \\
= & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T-1}\sum_{q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=0}^{\tau\wedge q}\left(E\left\|\tilde{C}_{i,\tau+k}\right\|^4\right)^{1/4} \\
& \times\left(E\left\|\tilde{C}'_{i,q+k}\right\|^4\right)^{1/4}E\left(\|C_i(1)\|^4\right)^{1/2} \\
= & o(1),
\end{aligned}$$

H_{12} is

$$\begin{aligned}
& Etr\frac{1}{T^2}\sum_{\tau,q=1}^{T-1}\sum_{p=1}^{T}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau}\sum_{j=1}^{q}\tilde{C}_{i,\tau-k}\otimes C_i(1)\left(\text{vec}(I_m)\text{vec}(I_m)'\right)\tilde{C}'_{i,q-j}\otimes C'_i(1) \\
\leq & Etr\frac{1}{T^2}\sum_{\tau,q=1}^{T-1}\sum_{k=1}^{\tau}\sum_{j=1}^{q}\Delta_T(t,\tau,p,q)\left(E\left\|\tilde{C}_{i,\tau-k}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}'_{i,q-j}\right\|^4\right)^{1/4}E\left(\|C_i(1)\|^4\right)^{1/2} \\
= & o(1),
\end{aligned}$$

H_{13} is

$$\begin{aligned}
& Etr\frac{1}{T^2}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau}\sum_{j=1}^{q}\tilde{C}_{i,\tau-k}\tilde{C}'_{i,q-j}\otimes C_i(1)C'_i(1)K_{mm} \\
\leq & \frac{1}{T^2}\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau}\sum_{j=1}^{q}\left(E\left\|\tilde{C}_{i,\tau-k}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}'_{i,q-j}\right\|^4\right)^{1/4}E\left(\|C_i(1)\|^4\right)^{1/2} \\
= & o(1),
\end{aligned}$$

and H_{14} is

$$\begin{aligned}
& \zeta Etr\frac{1}{T^2}\sum_{\tau,q=1}^{T-1}\sum_{l=1}^{m}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau\wedge q}\left(\tilde{C}_{i,\tau-k}\otimes C_i(1)\right)(e_{ll}\otimes e_{ll})\left(\tilde{C}'_{i,q-j}\otimes C'_i(1)\right) \\
\leq & \frac{M}{T^2}Etr\sum_{\tau,q=1}^{T-1}\Delta_T(t,\tau,p,q)\sum_{k=1}^{\tau\wedge q}\left(E\left\|\tilde{C}_{i,\tau-k}\right\|^4\right)^{1/4}\left(E\left\|\tilde{C}'_{i,q-j}\right\|^4\right)^{1/4}E\left(\|C_i(1)\|^4\right)^{1/2} \\
= & o(1).
\end{aligned}$$

We have therefore proved $Etr\left(\text{vec}\left(R_{i1}\right)\text{vec}\left(R_{i1}\right)'\right)=o(1)$. As a consequence, we also have $Etr\left(\text{vec}\left(R_{i2}\right)\text{vec}\left(R_{i2}\right)'\right)=o(1)$. Similarly, we can prove $Etr\left(\text{vec}\left(R_{i3}\right)\text{vec}\left(R_{i3}\right)'\right)=o(1)$. Details are omitted.

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

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

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

$$\begin{aligned}& E\frac{1}{T^2}\sum_{t,\tau,p,q=1}^{T} K(\frac{t}{T},\frac{\tau}{T})K(\frac{p}{T},\frac{q}{T})\text{vec}\left(C_i(1)V_{i,t}V_{i,\tau}'C_i'(1)\right)\text{vec}\left(C_i(1)V_{i,p}V_{i,q}'C_i'(1)\right)' \\ =\; & E\frac{1}{T^2}\sum_{t,\tau,p,q=1}^{T} K(\frac{t}{T},\frac{\tau}{T})K(\frac{p}{T},\frac{q}{T})\left(C_i(1)V_{i,\tau}\otimes C_i(1)V_{i,t}\right)\left(V_{i,q}'C_i'(1)\otimes V_{i,p}'C_i'(1)\right) \\ =\; & E\frac{1}{T^2}\sum_{t,\tau,p,q=1}^{T} K(\frac{t}{T},\frac{\tau}{T})K(\frac{p}{T},\frac{q}{T})\left(C_i(1)V_{i,\tau}V_{i,q}'C_i'(1)\right)\otimes\left(C_i(1)V_{i,t}V_{i,p}'C_i'(1)\right) \\ =\; & E\frac{1}{T^2}\sum_{t,\tau,p,q=1}^{T} K(\frac{t}{T},\frac{\tau}{T})K(\frac{p}{T},\frac{q}{T})\left(C_i(1)\otimes C_i(1)\right)\left(V_{i,\tau}V_{i,q}'\otimes V_{i,t}V_{i,p}'\right)\left(C_i'(1)\otimes C_i'(1)\right) \\ =\; & \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}K^2(\frac{t}{T},\frac{\tau}{T})\right)E\left(C_i(1)\otimes C_i(1)\right)\left(C_i'(1)\otimes C_i'(1)\right) \\ & + \left(\frac{1}{T}\sum_{t=1}^{T}K(\frac{t}{T},\frac{t}{T})\right)^2 E\text{vec}\left(C_i(1)C_i'(1)\right)\text{vec}\left(C_i(1)C_i'(1)\right) \\ & + \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}K^2(\frac{t}{T},\frac{\tau}{T})\right)E\left(C_i(1)\otimes C_i(1)\right)K_{mm}\left(C_i'(1)\otimes C_i'(1)\right) \\ & + \left(\frac{1}{T^2}\sum_{t=1}^{T}K^2(\frac{t}{T},\frac{t}{T})\right)\zeta E\left(C_i(1)\otimes C_i(1)\right)\left(\sum_{l=1}^{m}e_{ll}\otimes e_{ll}\right)\left(C_i'(1)\otimes C_i'(1)\right) \qquad (7.20)\end{aligned}$$

and

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

so $E\text{vec}(\widehat{\Omega}_i - E\widehat{\Omega}_i)\text{vec}\left(\widehat{\Omega}_i - E\widehat{\Omega}_i\right)'$ is

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

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 $\widetilde{\Omega}_i$. Using Mercer's Theorem, we have, for any T,

$$\begin{aligned} K(\frac{t}{T}, \frac{\tau}{T}) &= \sum_{m=1}^{\infty} \frac{1}{\lambda_m} f_m(\frac{t}{T}) f_m(\frac{\tau}{T}) \\ &= \sum_{m=1}^{M_0} \frac{1}{\lambda_m} f_m(\frac{t}{T}) f_m(\frac{\tau}{T}) + \sum_{m=M_0+1}^{\infty} \frac{1}{\lambda_m} f_m(\frac{t}{T}) f_m(\frac{\tau}{T}). \quad (7.21) \end{aligned}$$

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

$$\widetilde{\Omega}_{i,1} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} V_{i,t} \sum_{m=1}^{M_0} \frac{1}{\lambda_m} f_m(\frac{t}{T}) f_m(\frac{\tau}{T}) V_{i,\tau}', \quad (7.22)$$

$$\widetilde{\Omega}_{i,2} = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} V_{i,t} \sum_{m=M_0+1}^{\infty} \frac{1}{\lambda_m} f_m(\frac{t}{T}) f_m(\frac{\tau}{T}) V_{i,\tau}'. \quad (7.23)$$

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

$$\begin{aligned} \widetilde{\Omega}_{i,1} &= \sum_{m=1}^{M_0} \frac{1}{\lambda_m} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{i,t} f_m(\frac{t}{T}) \right) \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} f_m(\frac{\tau}{T}) V_{i,\tau}' \right) \\ &\Rightarrow \sum_{m=1}^{M_0} \frac{1}{\lambda_m} \int_0^1 f_m(r) dW_i(r) \int_0^1 f_m(s) dW_i'(s) \\ &= \int_0^1 \int_0^1 \left(\sum_{m=1}^{M_0} \frac{1}{\lambda_m} f_m(r) f_m(s) \right) dW_i(r) dW_i'(s). \quad (7.24) \end{aligned}$$

Following the same argument as in (7.20), we have, as $M_0 \to \infty$,

$$\begin{aligned} E\left(\text{vec}(\widetilde{\Omega}_{i2}) \text{vec}(\widetilde{\Omega}_{i2})' \right) &= o(\frac{1}{T^2}) \sum_{t=1}^{T} \sum_{\tau=1}^{T} E\text{vec}\left(V_{i,t} V_{i,\tau}' \right) \text{vec}\left(V_{i,t} V_{i,\tau}' \right)' \\ &= o(1), \quad (7.25) \end{aligned}$$

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

$$\begin{aligned} \widehat{\Omega}_i &\Rightarrow C_i(1) \int_0^1 \int_0^1 K(r,s) dW_i(r) dW_i'(s) C_i'(1) \\ &= C_i(1) \Xi_i C_i'(1). \quad (7.26) \end{aligned}$$

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 $E\left(\widehat{\Omega}_i|\mathcal{F}_{c_i}\right) = \mu\Omega_i + \Omega_i^e$ 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,

$$
\begin{aligned}
\text{plim}_{(T,n\to\infty)} \frac{1}{n}\sum_{i=1}^n \widehat{\Omega}_i &= \text{plim}_{(T,n\to\infty)} \frac{\mu}{n}\sum_{i=1}^n \left(\Omega_i + \Omega_i^\varepsilon + \Omega_i^e\right) \\
&= \text{plim}_{(T,n\to\infty)} \left(\frac{\mu}{n}\sum_{i=1}^n \Omega_i\right) + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^\varepsilon\right) \\
&\quad + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^e\right) \\
&= \mu\Omega + \text{plim}_{(T,n\to\infty)} \left(\frac{1}{n}\sum_{i=1}^n \Omega_i^e\right), \qquad (7.27)
\end{aligned}
$$

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

$$
\begin{aligned}
E\left\|\frac{1}{n}\sum_{i=1}^n \Omega_{i1}^e\right\| &= E\left\|\frac{1}{n}\sum_{i=1}^n \sum_{j=-T+1}^{T-1}\left(k(\frac{j}{T}) - 1\right)\Gamma_i(j)\right\| \\
\leq\ & \frac{1}{n}\sum_{i=1}^n \sum_{j=-T+1}^{T-1} j^{-1}\left|k(\frac{j}{T}) - 1\right| jE\left\|\Gamma_i(j)\right\| \\
\leq\ & \frac{1}{n}\sum_{i=1}^n \left(\sum_{j=-T+1}^{T-1} j^{-2}\left|k(\frac{j}{T}) - 1\right|^2\right)^{1/2} \left(\sum_{j=-\infty}^{\infty} j^2 E\left\|\Gamma_i(j)\right\|^2\right)^{1/2} \\
\leq\ & \frac{M}{n\sqrt{T}}\left(\frac{1}{T}\sum_{j=-T+1}^{T-1} (\frac{j}{T})^{-2}\left|k(\frac{j}{T}) - 1\right|^2\right)^{1/2} = O_p\left(\frac{1}{n\sqrt{T}}\right) \qquad (7.28)
\end{aligned}
$$

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

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

We need to calculate $E\left(\text{vec}(\Xi_i)\text{vec}(\Xi_i)'\right)$. Write $E\left(\text{vec}(\Xi_i)\text{vec}(\Xi_i)'\right)$ as

$$
\begin{aligned}
& E\left(\int_0^1\int_0^1 K(r,s)\text{vec}\left(dW_i(r)dW'_i(s)\right)\int_0^1\int_0^1 K(p,q)\text{vec}\left(dW_i(p)dW'_i(q)\right)'\right) \\
= \; & \int_0^1\int_0^1\int_0^1\int_0^1 K(r,s)K(p,q)E\left(\text{vec}\left(dW_i(r)dW'_i(s)\right)\text{vec}\left(dW_i(p)dW'_i(q)\right)'\right) \\
= \; & \int_0^1\int_0^1 \; I(r\neq s)\;\; K(r,r)K(s,s)E\left(\text{vec}\left(dW_i(r)dW'_i(r)\right)\text{vec}\left(dW_i(s)dW'_i(s)\right)'\right) \\
& +\int_0^1\int_0^1 \; I(r\neq s)\;\; K^2(r,s)E\left(\text{vec}\left(dW_i(r)dW'_i(s)\right)\text{vec}\left(dW_i(r)dW'_i(s)\right)'\right) \\
& +\int_0^1\int_0^1 \; I(r\neq s)\;\; K^2(r,s)E\left(\text{vec}\left(dW_i(r)dW'_i(s)\right)\text{vec}\left(dW_i(s)dW'_i(r)\right)'\right) \\
& +\int_0^1 K^2(r,r)E\left(\text{vec}\left(dW_i(r)dW'_i(r)\right)\text{vec}\left(dW_i(r)dW'_i(r)\right)'\right).
\end{aligned}
$$

We consider the four terms one by one. The first term is

$$
\begin{aligned}
& \int_0^1\int_0^1 \; I(r\neq s)\;\; K(r,r)K(s,s)E\left(\text{vec}\left(dW_i(r)dW'_i(r)\right)\text{vec}\left(dW_i(s)dW'_i(s)\right)'\right) \\
= \; & \int_0^1\int_0^1 \; I(r\neq s)\;\; K(r,r)K(s,s)E\left(\left(dW_i(r)\otimes dW_i(r)\right)\left(dW'_i(s)\otimes dW'_i(s)\right)\right) \\
= \; & \int_0^1\int_0^1 \; I(r\neq s)\;\; K(r,r)K(s,s)\left(EdW_i(r)\otimes dW_i(r)\right)\left(EdW'_i(s)\otimes dW'_i(s)\right) \\
= \; & \int_0^1\int_0^1 \; I(r\neq s)\;\; K(r,r)K(s,s)\text{vec}(I_m)\text{vec}(I_m)'drds \\
= \; & \mu^2\text{vec}(I_m)\text{vec}(I_m)'.
\end{aligned}
$$

The second term is

$$
\begin{aligned}
& \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\operatorname{vec}\left(dW_i(r) dW_i'(s)\right) \operatorname{vec}\left(dW_i(r) dW_i'(s)\right)'\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\left(dW_i(s) \otimes dW_i(r)\right)\left(dW_i'(s) \otimes dW_i'(r)\right)\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\left(dW_i(s) dW_i'(s)\right) \otimes \left(dW_i(r) dW_i'(r)\right)\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s)\left(I_m \otimes I_m\right) dr ds \\
= & \; I_{m^2} \int_0^1 \int_0^1 K^2(r,s) dr ds = I_{m^2} \delta^2 .
\end{aligned}
$$

The third term is

$$
\begin{aligned}
& \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\operatorname{vec}\left(dW_i(r) dW_i'(s)\right) \operatorname{vec}\left(dW_i(s) dW_i'(r)\right)'\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\left(dW_i(s) \otimes dW_i(r)\right)\left(dW_i'(r) \otimes dW_i'(s)\right)\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(\left(dW_i(s) \otimes dW_i(r)\right)\left(dW_i'(s) \otimes dW_i'(r) K_{mm}\right)\right) \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s) E\left(dW_i(s) dW_i'(s) \otimes dW_i(r) dW_i'(r)\right) K_{mm} \\
= & \int_0^1 \int_0^1 \; I(r \neq s) \;\; K^2(r,s)\left(I_m \otimes I_m\right) K_{mm} dr ds \\
= & \; K_{mm} \int_0^1 \int_0^1 K^2(r,s) dr ds = K_{mm} \delta^2 ,
\end{aligned}
$$

where the third line follows from a property of the commutation matrix, i.e. $dW_i'(r) \otimes dW_i'(s) = dW_i'(s) \otimes dW_i'(r) K_{mm}$, See Part (ix) of Theorem 3.1 in Magnus and Neudecker (1979). Finally, the fourth term is

$$
\begin{aligned}
& \int_0^1 K^2(r,r) E\left(\operatorname{vec}\left(dW_i(r) dW_i'(r)\right) \operatorname{vec}\left(dW_i(r) dW_i'(r)\right)'\right) \\
= & \int_0^1 K^2(r,r) E\left(dW_i(r) \otimes dW_i(r)\right)\left(dW_i'(r) \otimes dW_i'(r)\right) \\
= & \int_0^1 K^2(r,r) E\left(dW_i(r) dW_i'(r)\right) \otimes \left(dW_i(r) dW_i'(r)\right) = 0 .
\end{aligned}
$$

Therefore

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

Consequently,

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

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 $\Theta_T=E\operatorname{vec}(Q_{i,T})\operatorname{vec}(Q_{i,T})'$. 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

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

Hence,

$$\begin{aligned}
\Theta_T &= \mu_T^2 E\text{vec}\left((I_{m_y}, -\beta)C_i(1)C_i'(1)G_m\right)\text{vec}\left((I_{m_y}, -\beta)C_i(1)C_i'(1)G_m\right)' + o(1) \\
&\quad + \delta_T^2 C_{xi}(1) \otimes \left(C_{yi}(1) - \beta C_{xi}(1)\right)\left(I_{m^2} + K_{mm}\right) C_{xi}'(1) \otimes \left(C_{yi}(1) - \beta C_{xi}(1)\right)' \\
&= \mu_T^2 E\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\left(\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\right)' + o\left(1\right) \\
&\quad + \delta_T^2 \left(C_{xi}(1)C_{xi}'(1)\right) \otimes \left((C_{yi}(1) - \beta C_{xi}(1))(C_{yi}(1) - \beta C_{xi}'(1))\right) \\
&\quad + \delta_T^2 \left\{C_{xi}(1) \otimes (C_{yi}(1) - \beta C_{xi}(1))\right\}\left\{(C_{yi}(1) - \beta C_{xi}(1))' \otimes C_{xi}'(1)\right\} K_{m_y m_x} \\
&= \mu_T^2 E\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\left(\text{vec}(\Omega_{yxi} - \beta\Omega_{xxi})\right)' \\
&\quad + \delta_T^2 E\Omega_{xxi} \otimes \left(\Omega_{yyi} - \beta\Omega_{xyi} - \Omega_{yxi}\beta' + \beta\Omega_{xxi}\beta'\right) \\
&\quad + \delta_T^2 E\left(\Omega_{xyi} - \Omega_{xxi}\beta'\right) \otimes \left(\Omega_{yxi} - \beta\Omega_{xxi}\right) K_{m_y m_x} + o\left(1\right).
\end{aligned}$$

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 $T \to \infty$,

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

and

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

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

REFERENCES

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

Andrews, D. W. K. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica,* 60, 953–966.

Baltagi, B. H. (Ed.) (2000): *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, JAI Press.

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

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.

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

——— (1988): *Matrix Differential Calculus with Application in Statistics and Econometrics*, New York, Wiley.

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.

——— (1994): "Automatic Lag Selection in Covariance Estimation," *Review of Economic Studies*, 61, 631–654.

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., 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, Yale University.

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

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

Vogelsang, T. J. (2000): "Testing in GMM Models Without Truncation," Center for Analytic Economics, Working Paper 01-12, Cornell University.

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

——— (2001): *Asymptotic Theory for Econometricians*, revised edition. Academic Press: New York.