

Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data*

Ryo Okui[†]

Hong Kong University of Science and Technology

May 27, 2009

*The author would like to thank two anonymous referees, Guido Kuersteiner, In Choi, Songnian Chen, Masanobu Taniguchi, Hidehiko Ichimura, Eiji Kurozumi, Peter Phillips, Takashi Yamagata, Katsumi Shimotsu, Stephane Bonhomme, Kohtaro Hitomi, Yoshihiko Nishiyama and seminar participants at Kobe, Hokkaido, Hitotsubashi, Yokohama National, Tokyo and Kyoto Universities, and attendees at the Kansai Econometric Society Meeting held in Yokohama, the Third Symposium on Econometric Theory and Applications held in Hong Kong, the Far Eastern Meeting of the Econometric Society held in Taipei and the 14th Panel Data Conference held in Xiamen. The author also acknowledges financial support from the Hong Kong University of Science and Technology under Project No. DAG05/06.BM16 and from the Research Grants Council under Project No. HKUST643907. The author is solely responsible for all errors.

[†]Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.
Email: okui@ust.hk

Abstract

An important reason for analyzing panel data is to observe the dynamic nature of an economic variable separately from its time-invariant unobserved heterogeneity. This paper examines how to estimate the autocovariances of a variable separately from its time-invariant unobserved heterogeneity. When both cross-sectional and time series sample sizes tend to infinity, we show that the within-group autocovariances are consistent, although that they are severely biased when the time series length is short. The biases have the leading term that converges to the long-run variance of the individual dynamics. This paper develops methods to estimate the long-run variance in panel data settings and to alleviate the biases of the within-group autocovariances based on the proposed long-run variance estimators. Monte Carlo simulations reveal that the procedures developed in this paper effectively reduce the biases of the estimators for small samples.

Proposed running head: Unbiased estimation of autocovariances

Corresponding author: Ryo Okui, Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: okui@ust.hk

1 Introduction

An important reason for analyzing panel data is to observe the dynamic nature of an economic variable separately from its time-invariant unobserved heterogeneity. In time series analysis, the first step in investigating the dynamics of a variable may be to examine its correlogram. However, in panel data analysis, it is difficult to analyze autocovariances and autocorrelations, although some textbooks, such as Cameron and Trivedi (2005, Chapter 21.3), suggest such an analysis. The difficulty comes from the fact that sample autocovariances and autocorrelations are contaminated by spurious correlations caused by unobserved heterogeneity. This paper develops statistical tools to estimate autocovariances and autocorrelations of economic variables using panel data separately from their time-invariant unobserved heterogeneity.

When the length of the time series of a panel is short, some restrictions on the autocovariance structure are necessary, otherwise the autocovariances of the individual dynamics are not identified (see, for example, Arellano (2003, Chapter 5)) and the conventional autocovariance estimators are even asymptotically biased, as pointed out by Solon (1984). For example, early studies on income dynamics (e.g., Lillard and Willis (1978), MaCurdy (1982), and Abowd and Card (1989)) model the time-varying components as ARMA processes. Researchers have developed methods to estimate those models. For autoregressive models, the within-group estimator is severely biased when the length of the time series is short (Nickell (1981)). Anderson and Hsiao (1981) have proposed instrumental variable estimation of first-order autoregressive (AR(1)) models. Their methods have been extended by Arellano and Bond (1992) and Holtz-Eakin, Newey and Rosen (1988) to generalized methods of moments estimation. Baltagi and Li (1994) consider the estimation of the moving average models. Alternatively, the minimum distance estimator (see Chamberlain (1984)) may be employed as considered by Abowd and Card (1989).

Recently, panel data with moderately long time lengths have become available, and researchers have developed mathematical tools to handle asymptotic sequences under which two indexes tend to infinity. These panels and mathematical tools have motivated researchers to look into the asymptotic properties of the statistics in the case of long panel data. Alvarez and Arellano (2003) and Hahn and Kuersteiner (2002) study the asymptotic properties of the within-group estimator for panel AR(1) models when both the cross-sectional sample size (N) and the length of the time series (T) are large. Kiviet (1995) and Bun and Kiviet (2006) consider more general (but still AR(1)-type) models that include covariates. Hahn and Kuersteiner (2002) also develop a bias-corrected within-group estimator for panel AR(1) models. Lee (2008a) and Hansen (2007) consider AR(p) models and develop methods to correct the biases of the within-group estimators. Lee (2008a) also considers cases in which the lag order is misspecified and proposes methods to choose the lag order. While AR(p) models can capture many kinds of dynamics, these methods still suffer from model misspecification. Moreover, the focus of these articles is on the estimation of the coefficients in autoregressive models, and the results in the existing literature are not readily applicable to this project's purpose.

This paper addresses a basic, unanswered question of how to estimate the autocovariance structure of the individual dynamic component of a variable without imposing a specific structure. The statistical methods developed in this paper have several potential impacts. They should yield a better understanding of the dynamic nature of key economic variables. They are also useful for the purpose of finding appropriate models in empirical applications, even if we desire a model-based analysis. Moreover, many important quantities in dynamic panel data analysis, such as autocorrelations and coefficients in panel VAR models, are written as a function of autocovariances and understanding how to estimate autocovariances is helpful in developing methods to estimate those quantities.

We study the asymptotic properties of the within-group autocovariances, using double asymptotics, under which both N and T tend to infinity. We show that the within-group autocovariances are consistent for the autocovariances of individual dynamics, but that these estimators are heavily biased when T is only moderately large. The key finding is that the leading terms of the biases of these estimators are proportional to the long-run variance of the individual dynamics. The presence of long-run variances in the bias caused by the incidental parameters problem is also observed by Hahn and Kuersteiner (2004) and Lee (2008a, 2008b).

We consider the estimation of the biases and propose bias-corrected estimators. The key is the estimation of the long-run variance of individual dynamics. There have been numerous procedures proposed for the estimation of long-run variances in the time series literature. (See, e.g., den Haan and Levin (1997) for a review, although a large number of articles on this issue has been published since.) We extend the kernel long-run variance estimators to panel data settings. We then develop methods to alleviate the biases of the within-group autocovariances using the proposed long-run variance estimator.

We examine the mean squared error (MSE) of the long-run variance estimator and the result reveals that the bias in the autocovariance estimators also causes bias in the long-run variance estimator. To address this problem, we consider iterative procedures in which we estimate the long-run variance based on the bias-corrected estimators of the autocovariances, and we correct the bias using the new long-run variance estimator. We may repeat this iteration many times. The iteration converges under a mild condition and the autocovariance estimator obtained as the limit of the iteration has a closed form, which makes it easy to implement. The theoretical and simulation results show that this iteration reduces the bias in the long-run estimator and improves the performance of the bias-corrected autocovariance estimators.

The remainder of the paper is organized as follows. Section 2 introduces the theoretical framework. In Section 3, we study the asymptotic properties of the within-group autocovariance estimators. Methods to alleviate the biases of the within-group autocovariance estimators are discussed in Section 4. In Section 5, we report the results of Monte Carlo experiments. We consider several extensions in Section 6.

2 Setting

Suppose that panel data $\{y_{it}\}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$ are available. We assume that y_{it} is generated by the sum of the time-invariant individual effect, η_i , and the time-varying stationary process, w_{it} :

$$y_{it} = \eta_i + w_{it},$$

where $\{\{w_{it}\}_{t=1}^T\}_{i=1}^N$ are independently and identically distributed (i.i.d.) across individuals and stationary over time with mean $E(w_{it}) = 0$. We do not impose any specific model on the autocovariance structure of w_{it} .

Let γ_k denote the k -th order autocovariance of w_{it} (i.e., $\gamma_k = E(w_{it}w_{it-k})$). Our main question is how to estimate γ_k s when relatively long panel data sets are available.

3 Asymptotic properties of the within-group autocovariances

We examine the asymptotic properties of the k -th within-group autocovariance:

$$\hat{\gamma}_k = \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (y_{it} - \bar{y}_i)(y_{i,t-k} - \bar{y}_i),$$

which may be a natural estimator of γ_k , where $\bar{y}_i = \sum_{t=1}^T y_{it}/T$. When T is fixed, $\hat{\gamma}_k$ is not consistent for γ_k (Solon (1984)). The main source of the inconsistency is that we cannot consistently estimate η_i when T is fixed. On the other hand, it is shown below that $\hat{\gamma}_k$ is consistent for γ_k when both N and T tend to infinity under the following assumption.

Assumption 1. 1. $\{\{w_{it}\}_{t=1}^T\}_{i=1}^N$ are i.i.d. across individuals.

2. w_{it} is strictly stationary within individuals and $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$.

3. There exists $M < \infty$ such that $E(|w_{it}w_{ik}w_{im}w_{il}|) < M$ for any t, k, m and l .

This set of assumptions is standard. Note that Assumption 1 does not impose any restriction on the probabilistic nature of η_i , as η_i does not appear in $\hat{\gamma}_k$. The following theorem shows the consistency of $\hat{\gamma}_k$.

Theorem 1. Suppose that Assumption 1 is satisfied. As $N \rightarrow \infty$ and $T \rightarrow \infty$, we have $\hat{\gamma}_k \rightarrow_p \gamma_k$ for any k .

However, $\hat{\gamma}_k$ may be severely biased when T is not very large relative to N . To see this, we observe that $\hat{\gamma}_k$ may be decomposed in the following form (see the proof of Theorem 1):

$$\hat{\gamma}_k = \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T w_{it}w_{i,t-k} - \frac{1}{N} \sum_{i=1}^N (\bar{w}_i)^2 + \text{small}.$$

The term $\bar{w}_i (= \bar{y}_i - \eta_i)$ can be understood as the estimation error for η_i . This estimation error is the main source of the bias, even when T tends to infinity. Now, we have:

$$E \left\{ \frac{1}{N} \sum_{i=1}^N (\bar{w}_i)^2 \right\} = E \{ (\bar{w}_i)^2 \} = \frac{1}{T} \left(\gamma_0 + 2 \sum_{j=1}^{T-1} \frac{T-j}{T} \gamma_j \right),$$

which is of order $O(1/T)$ when $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$. Thus, the estimator $\hat{\gamma}_k$ exhibits bias of order $O(1/T)$, which may be severe when T is not very large.

To make the argument more formal, we present the theorem below concerning the asymptotic distribution of $\hat{\gamma}_k$. We make the following assumption that concerns the cumulants of w_{it} . Let $cum(t_1, \dots, t_p)$ denote the p -th-order cumulant of $(w_{i,t_1}, \dots, w_{i,t_p})$.

Assumption 2. $\sum_{j_2, \dots, j_p = -\infty}^{\infty} |cum(0, j_2, \dots, j_p)| < \infty$, for $p \leq 8$.

We use Theorem 3 of Phillips and Moon (1999) to prove the next theorem. Assumption 2 is used to guarantee the uniform integrability condition of $\{\sum_{t_1=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) / \sqrt{T}\}^2$, which is one of the key conditions of Theorem 3 of Phillips and Moon (1999). To prove the asymptotic normality, Assumption 2 may be relaxed as long as the uniform integrability condition is met. Assumption 2 is also used to guarantee the existence of the asymptotic variance of $\hat{\gamma}_k$ and is used later to show the asymptotic properties of the long-run variance estimator.

Theorem 2. Suppose that Assumptions 1 and 2 are satisfied. Then, as $N \rightarrow \infty$, $T \rightarrow \infty$ and $N/T^3 \rightarrow 0$, we have

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

where

$$V_T \equiv \gamma_0 + 2 \sum_{j=1}^{T-1} \frac{T-j}{T} \gamma_j.$$

Remark 1. Let $V \equiv \sum_{j=-\infty}^{\infty} \gamma_j$ denote the long-run variance of w_{it} . We have $V_T \rightarrow V$ as $T \rightarrow \infty$. The leading term of the bias of $\hat{\gamma}_k$ converges to the long-run variance of w_{it} . The next section examines the possibility of correcting the bias by estimating the long-run variance. This observation also implies that the bias is large if w_{it} is highly persistent. Note that $V_T > 0$, which implies that the bias is downward and $\hat{\gamma}_k$ is, on average, smaller than γ_k . It is also notable that the leading term of the bias does not depend on the order of the autocovariance, k .

Remark 2. The condition $N/T^3 \rightarrow 0$ is required to ignore the bias term of order $1/T^2$. This condition can be relaxed if the bias term of order $1/T^2$ is taken into account. However, it makes the expression of the asymptotic bias complicated and we shall keep the condition $N/T^3 \rightarrow 0$.

Remark 3. Although this paper is about the bias correction, the efficiency issue might deserve some discussion. In time series analysis (i.e., when $N = 1$), the conventional autocovariance estimators up

to p -th-order are efficient when the process follows a Gaussian AR(p) model (see, e.g., Porat (1987), Kakizawa and Taniguchi (1994) and Kakizawa (1999)). Since $\hat{\gamma}_k$ is the sample average of individual autocovariances and we assume that we have an i.i.d. sample, we may expect that the variance of $\hat{\gamma}_k$ is the smallest possible when w_{it} follows a Gaussian AR(p) model and $k \leq p$. This may be proved by following the steps used for the efficiency result in Hahn and Kuersteiner (2002). However, this is beyond the scope of this paper.

Remark 4. Theorem 2 presents the asymptotic distribution of $\hat{\gamma}_k$ for each k . It is easy to find the joint asymptotic distribution of $\hat{\gamma}_k$ and $\hat{\gamma}_j$ for $k \neq j$, because $\hat{\gamma}_k$ has an asymptotic linear form:

$$\sqrt{NT} \left(\hat{\gamma}_k - \gamma_k + \frac{1}{T} V_T \right) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) + o_p(1).$$

Note that the asymptotic covariance between $\hat{\gamma}_k$ and $\hat{\gamma}_j$ is:

$$\sum_{t=-\infty}^{\infty} \{ \gamma_t \gamma_{t-k+j} + \gamma_{t+j} \gamma_{t-k} + \text{cum}(0, -k, t, t-j) \}.$$

4 Bias correction

In this section, we consider ways to alleviate the bias of γ_k . We propose to use an estimate of V_T to mitigate the bias of γ_k . Before discussing how to estimate V_T , we show that this idea of bias correction works at least theoretically. Let \hat{V}_T denote an estimator of V_T . The bias-corrected estimator of γ_k , denoted as $\tilde{\gamma}_k$, is obtained by adding \hat{V}_T/T to $\hat{\gamma}_k$:

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

Let $r_{N,T}$ be the inverse of the rate of convergence of \hat{V}_T such that $\hat{V}_T - V_T = O_p(r_{N,T})$. The next theorem shows that the asymptotic distribution of $\tilde{\gamma}_k$ is centered around zero.

Theorem 3. *Suppose that Assumptions 1 and 2 are satisfied. Suppose also that $N/T^3 \rightarrow 0$ and $r_{N,T} \sqrt{N/T} \rightarrow 0$. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$,*

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

The proof is omitted as it is trivial. This theorem implies that we may obtain estimates of the autocovariances whose biases are small if we get some estimates of V_T . Thus, the main question of this section is how to construct a good estimator of the long-run variance of w_{it} and discover the rate of convergence of the long-run variance estimator.

4.1 Estimating the long-run variance

This subsection considers the estimation of the long-run variance of w_{it} . As is known in the time series literature, while the long-run variance is the sum of the autocovariances, simply summing sample

autocovariances does not yield a consistent estimator. We must weight the effect of the higher-order autocovariances downward in order to obtain a consistent estimator for the long-run variance. Following Parzen (1957) and Andrews (1991), we consider the kernel estimators:

$$\tilde{V}_T = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{T-|j|}{T} \hat{\gamma}_j = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \hat{\gamma}_j^+,$$

where

$$\hat{\gamma}_j^+ = \frac{1}{NT} \sum_{i=1}^N \sum_{t=|j|+1}^T (y_{it} - \bar{y}_i)(y_{i,t-|j|} - \bar{y}_i) = \frac{T-|j|}{T} \hat{\gamma}_j$$

is the within-group autocovariance using T in the denominator instead of $T-|j|$, $k(\cdot)$ is a kernel function and the scalar, S , is the bandwidth chosen by the researcher. We assume that the kernel function belongs to the class \mathcal{K}_1 :

$$\begin{aligned} \mathcal{K}_1 = & \left\{ k(\cdot) : R \rightarrow [-1, 1] | k(0) = 1, k(x) = k(-x) \forall x \in R, \right. \\ & \left. \int_{-\infty}^{\infty} k^2(x) dx < \infty, k(\cdot) \text{ is continuous almost everywhere and at } 0 \right\}. \end{aligned}$$

An example of a kernel that belongs to \mathcal{K}_1 is the quadratic spectrum (QS) kernel:

$$k(x) = \frac{3}{(6\pi x/5)^2} \left\{ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right\}.$$

Andrews (1991) demonstrates several attractive properties of the QS kernel function. Note that \tilde{V}_T is always nonnegative with the QS kernel, which also means that $\tilde{\gamma}_0$ is nonnegative with the QS kernel. Later, we also consider the truncated kernel:

$$k(x) = \begin{cases} 1 & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

We also assume that the kernel function satisfies $\int k(x) dx < \infty$ and $\int |x|k(x) dx < \infty$.

The following theorem shows the consistency of \tilde{V}_T and gives the rate of convergence of \hat{V}_T . The MSE formula given in the theorem also serves as the device used to choose the bandwidth parameter.

Theorem 4. *Suppose that Assumptions 1 and 2 are satisfied. Assume that $k(\cdot) \in \mathcal{K}_1$, $\int k(x) dx < \infty$ and $\int |x|k(x) dx < \infty$. If $S \rightarrow \infty$ and $S/T \rightarrow 0$, then,*

$$\tilde{V}_T - V_T \rightarrow_p 0.$$

Let $k_q \equiv \lim_{x \rightarrow 0} \{1 - k(x)\}/|x|^q$ and $V^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \gamma_j$. Suppose also that $N^{q+1}/T^q \rightarrow 0$ and $S^{2q+1}/(NT) \rightarrow \tau$, where $0 < \tau < \infty$, for some $0 < q < \infty$, for which k_q and $|V^{(q)}|$ are finite. Then,

$$\lim_{N, T \rightarrow \infty} \frac{NT}{S} \text{MSE}(\tilde{V}_T) = k_q^2 \left(V^{(q)} \right)^2 \tau^{-1} + 2V^2 \int k^2(x) dx.$$

On the other hand, suppose that $N^{q+1}/T^q \rightarrow \infty$ and $S^{q+1}/T \rightarrow \tau$, where $0 < \tau < \infty$, for some $0 < q < \infty$, for which k_q and $|V^{(q)}|$ are finite. Then,

$$\lim_{N, T \rightarrow \infty} \frac{T^2}{S^2} \text{MSE}(\tilde{V}_T) = \left\{ -k_q V^{(q)} \tau^{-1} - V \int k(x) dx \right\}^2.$$

The value of q in the theorem represents the smoothness of the kernel function at the origin. A large value of q for which k_q is finite indicates that the kernel function is smooth at zero. For example, the QS kernel has $q = 2$.

Remark 5. There are two bias terms that are relevant to this result. The first bias term that is proportional to $k_q V^{(q)}$ comes from the fact that we use a kernel function. The other bias term that is proportional to $V \int k(x) dx$ stems from the result that each $\hat{\gamma}_k$ is biased. When T is sufficiently large relative to N (i.e., $N^{q+1}/T^q \rightarrow 0$), the MSE has a similar form to that presented by Andrews (1991). When T is not very large compared with N (i.e., $N^{q+1}/T^q \rightarrow \infty$), the second term of the bias becomes more important than the variance term. Note that the estimator, \hat{V}_T , is the sample average of the long-run variance estimators across individuals and that N affects the variance, but not the bias, of \hat{V}_T . Therefore, when N is large, the variance becomes small relative to the biases and the leading term in the MSE is the square of the leading terms of the biases. This phenomenon happens when N is proportional to T and is relevant in practice.

Remark 6. The theorem gives the rate of convergence of \hat{V}_T , which is useful in examining the conditions on the relationship between N and T for asymptotically unbiased estimation of autocovariances. Theorem 3 assumes two conditions, $N/T^3 \rightarrow 0$ and $r_{N,T} \sqrt{N/T} \rightarrow 0$. Theorem 4 gives $r_{N,T} = (NT)^{-q/(2q+1)}$ if $N^{q+1}/T^q \rightarrow 0$, and $r_{N,T} = T^{-q/(q+1)}$ if $N^{q+1}/T^q \rightarrow \infty$. When $N^{q+1}/T^q \rightarrow 0$, the condition $N/T^3 \rightarrow 0$ is automatically satisfied and $r_{N,T} \sqrt{N/T} = (N/T^{4q+1})^{1/(4q+2)}$, which is also $o(1)$ under $N^{q+1}/T^q \rightarrow 0$. On the other hand, $N/T^3 \rightarrow 0$ is stronger than $N^{q+1}/T^q \rightarrow \infty$ and the condition $r_{N,T} \sqrt{N/T} \rightarrow 0$ becomes $(N^{q+1}/T^{3q+1})^{1/(2q+2)} \rightarrow 0$, which is stronger than $N/T^3 \rightarrow 0$. Therefore, we need $N^{q+1}/T^{3q+1} \rightarrow 0$ for asymptotically unbiased estimation of autocovariances if we correct the bias by using \tilde{V}_T .

4.2 Choosing the bandwidth parameter

We choose the bandwidth parameter by minimizing the MSE of \tilde{V}_T . Let $\xi = V^{(q)}/V$. Then, the value of the bandwidth parameter that minimizes the MSE is:

$$S^* = \begin{cases} [\{qk_q^2 / \int k^2(x) dx\} \xi^2 T N]^{1/(2q+1)}, & \text{when } N^{q+1}/T^q \rightarrow 0, \\ [\{qk_q / \int k(x) dx\} \xi T]^{1/(q+1)}, & \text{when } N^{q+1}/T^q \rightarrow \infty \text{ and } V^{(q)} \geq 0, \\ [\{k_q / \int k(x) dx\} \xi |T|]^{1/(q+1)}, & \text{when } N^{q+1}/T^q \rightarrow \infty \text{ and } V^{(q)} < 0. \end{cases}$$

We need to obtain an estimate of ξ . We follow the strategy proposed by Andrews (1991): we estimate ξ based on the formula that is valid when w_{it} follows an AR(1) process. When w_{it} follows the AR(1)

process with coefficient δ , then the parameter ξ can be written as:

$$\xi = \frac{2\delta}{(1-\delta)^2}.$$

There are many ways to estimate the parameter δ . Here, we consider the estimator in Hahn and Kuersteiner (2002):

$$\hat{\delta} = \frac{T}{T-1} \frac{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{i-})(y_{i,t-1} - \bar{y}_{i+})}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{i-})^2} + \frac{1}{T-1},$$

where $\bar{y}_{i-} = \sum_{t=1}^{T-1} y_{it}/(T-1)$ and $\bar{y}_{i+} = \sum_{t=2}^T y_{it}/(T-1)$. Then, we estimate ξ by $\hat{\xi} = 2\hat{\delta}/\{(1-\hat{\delta})^2\}$.

We use the following estimated bandwidth:

$$\hat{S}^* = \begin{cases} \min \left\{ \left[\{qk_q^2 / \int k^2(x)dx\} \hat{\xi}^2 TN \right]^{1/(2q+1)}, \left[\{qk_q / \int k(x)dx\} \hat{\xi} T \right]^{1/(q+1)} \right\}, & \text{if } \hat{\xi} \geq 0, \\ \min \left\{ \left[\{qk_q^2 / \int k^2(x)dx\} \hat{\xi}^2 TN \right]^{1/(2q+1)}, \left[\{k_q / \int k(x)dx\} |\hat{\xi}| T \right]^{1/(q+1)} \right\}, & \text{if } \hat{\xi} < 0. \end{cases}$$

Note that $\hat{\delta}$ converges to the first-order autocorrelation of w_{it} and is bounded in probability. Thus, the estimation of $\hat{\delta}$ (and $\hat{\xi}$) does not affect the rate of the bandwidth asymptotically. We see that $C_1(TN)^{1/(2q+1)} < C_2(T)^{1/(q+1)}$ for T and N sufficiently large for any constants C_1 and C_2 if $N^{q+1}/T^q \rightarrow 0$ and the opposite result holds if $N^{q+1}/T^q \rightarrow \infty$. These observations imply that

$$\begin{aligned} \Pr \left\{ \hat{S}^* = \left[\left\{ qk_q^2 / \int k^2(x)dx \right\} \hat{\xi}^2 TN \right]^{1/(2q+1)} \right\} &\rightarrow 1 \quad \text{if } N^{q+1}/T^q \rightarrow 0, \\ \Pr \left\{ \hat{S}^* = \left[\left\{ qk_q^2 / \int k^2(x)dx \right\} \hat{\xi}^2 TN \right]^{1/(2q+1)} \right\} &\rightarrow 0 \quad \text{if } N^{q+1}/T^q \rightarrow \infty. \end{aligned}$$

Thus, the bandwidth has an appropriate rate in large samples.

In the simulations, we use the QS kernel function, for which we have $q = 2$, $k_q \approx 1.4212$, $\int k(x)dx \approx 1.2930$ and $\int k^2(x)dx = 1$. The bandwidth is:

$$\hat{S}^* = \begin{cases} \min\{1.3221(\hat{\xi}^2 TN)^{1/5}, 1.3002(\hat{\xi} T)^{1/3}\}, & \text{if } \hat{\xi} \geq 0, \\ \min\{1.3221(\hat{\xi}^2 TN)^{1/5}, 1.0320(|\hat{\xi}| T)^{1/3}\}, & \text{if } \hat{\xi} < 0. \end{cases} \quad (1)$$

Remark 7. One may alternatively consider choosing the bandwidth parameter by minimizing the MSE of $\tilde{\gamma}_k$. In general, the bandwidth parameter that minimizes the MSE of $\tilde{\gamma}_k$ is different from that minimizing the MSE of \tilde{V}_T . However, the benefit of using this alternative criterion is limited. To see this, consider the MSE of $\tilde{\gamma}_k$:

$$E\{(\tilde{\gamma}_k - \gamma_k)^2\} = E\left\{\left(\hat{\gamma}_k - \gamma_k + \frac{1}{T}V_T\right)^2\right\} + 2\frac{1}{T}E\left\{\left(\hat{\gamma}_k - \gamma_k + \frac{1}{T}V_T\right)(\tilde{V}_T - V_T)\right\} + \frac{1}{T^2}E\{(\tilde{V}_T - V_T)^2\}.$$

While the cross term, $E\{(\hat{\gamma}_k - \gamma_k + V_T/T)(\tilde{V}_T - V_T)\}/T$, depends on the bandwidth, its leading term does not. The cross term is approximately equal to $\sum_{j=-T+1}^{T-1} k(j/S) \text{cov}(\hat{\gamma}_k, \hat{\gamma}_j)/T$. Since $NT \sum_{j=-T+1}^{T-1} \text{cov}(\hat{\gamma}_k, \hat{\gamma}_j)$ converges, we have

$$\frac{1}{T}E\left\{\left(\hat{\gamma}_k - \gamma_k + \frac{1}{T}V_T\right)(\tilde{V}_T - V_T)\right\} = O\left(\frac{1}{NT^2}\right),$$

which implies that the bandwidth does not affect the order of the cross term. Moreover, when N is relatively large, the cross term is small compared with the MSE of \tilde{V}_T . For example, when $N^3/T^4 \rightarrow \infty$, the order of the term $E\{(\tilde{V}_T - V_T)^2\}/T^2$ with the QS kernel is $T^{-2-4/3} = T^{-10/3}$ and is of an order larger than $O(1/(NT^2))$.

4.3 Iterative procedures

Theorem 4 demonstrates that the bias in each $\hat{\gamma}_k$ is relevant even in the estimation of the long-run variance. To address this problem, we use an iterative procedure. We update the estimate of V_T using the bias-corrected estimators for γ_k for $k = 0, \dots, T-1$. Then, we reestimate γ_k based on the updated estimate of V_T . This iteration may be repeated many times. As $\tilde{\gamma}_k$ s are bias-corrected, we expect that the long-run variance estimator based on $\tilde{\gamma}_k$ is also bias-corrected. The iteration is expressed in the following way. Let

$$\tilde{V}_T(m+1) = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_m}\right) \frac{T-|j|}{T} \tilde{\gamma}_j(m),$$

and

$$\tilde{\gamma}_k(m) = \hat{\gamma}_k + \frac{1}{T} \tilde{V}_T(m), \quad k = 0, \dots, T-1,$$

where m denotes the number of iterations, S_m is the bandwidth parameter for the m -th iteration and $\tilde{\gamma}_k(0) = \hat{\gamma}_k$ for $k = 0, \dots, T-1$.

Let $\tilde{\gamma}(m) = (\tilde{\gamma}_0(m), \dots, \tilde{\gamma}_{T-1}(m))'$, $\hat{\gamma} = (\hat{\gamma}_0, \dots, \hat{\gamma}_{T-1})'$, I_T be the $T \times T$ identity matrix and ι_T be the $T \times 1$ vector of ones. We consider using the same bandwidth throughout the iterations. Let S denote the bandwidth parameter. Let

$$K_T = \left(k(0), 2\frac{T-1}{T}k\left(\frac{1}{S}\right), 2\frac{T-2}{T}k\left(\frac{2}{S}\right), \dots, 2\frac{1}{T}k\left(\frac{T-1}{S}\right) \right)'.$$

We can write the iteration formula in the following way:

$$\tilde{\gamma}(m+1) = \hat{\gamma} + \frac{1}{T} \iota_T K_T' \tilde{\gamma}(m).$$

If $\iota_T' K_T < T$, this iteration converges and the limit, $\tilde{\gamma}(\infty)$, can be written as

$$\tilde{\gamma}(\infty) = \left(I_T + \frac{1}{T - \iota_T' K_T} \iota_T K_T' \right) \hat{\gamma} = \left(I_T - \frac{1}{T} \iota_T K_T' \right)^{-1} \hat{\gamma}.$$

Note that $\iota_T' K_T < T$ is satisfied when $k(x) < 1$ for $x \neq 0$. Most commonly used kernel functions, including the QS kernel, satisfy this condition. The long-run variance estimator obtained as the limit of the iteration is

$$\tilde{V}_T(\infty) = K_T' \tilde{\gamma}(\infty) = K_T' \left(I_T + \frac{1}{T - \iota_T' K_T} \iota_T K_T' \right) \hat{\gamma} = \left(1 + \frac{\iota_T' K_T}{T - \iota_T' K_T} \right) \tilde{V}_T.$$

The following theorem presents the MSE of $\tilde{V}_T(\infty)$. Since $\tilde{V}_T(\infty)$ is based on the bias-corrected autocovariance estimators, the second term in the bias becomes small, and we have the usual bias-variance trade-off. Note that the iterations do not alter the asymptotic distribution of $\tilde{\gamma}$ s, while the rate conditions for N and T would be affected.

Theorem 5. *Suppose that Assumptions 1 and 2 are satisfied. Assume that $k(\cdot) \in \mathcal{K}_1$, $\int k(x)dx < \infty$ and $\int |x|k(x)dx < \infty$. Suppose also that $S \rightarrow \infty$ and $S/T \rightarrow 0$. Then, $\tilde{V}_T(\infty) - V_T \rightarrow_p 0$. Let q be a number that satisfies $0 < q < \infty$, for which k_q and $|V^{(q)}|$ are finite. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$ with $N^{q+2}/T^{3q} \rightarrow 0$ and $S^{2q+1}/(NT) \rightarrow \tau$, we have:*

$$\lim_{N,T \rightarrow \infty} \frac{NT}{S} \text{MSE}\{\tilde{V}_T(\infty)\} = k_q^2 \left(V^{(q)}\right)^2 \tau^{-1} + 2V^2 \int k^2(x)dx.$$

Remark 8. One of the conditions for $\tilde{\gamma}_k$ to be asymptotically unbiased is $r_{N,T}\sqrt{N/T} = o(1)$, which is automatically satisfied since $r_{N,T} = (NT)^{-q/(2q+1)}$ and $N^{q+1}/T^{3q} \rightarrow 0$ under the conditions of the theorem. On the other hand, $N^{q+2}/T^{3q} \rightarrow 0$ is typically stronger than $N/T^3 \rightarrow 0$ because commonly used kernel functions have $q = 1$ or $q = 2$. For example, the QS kernel has $q = 2$ and asymptotically unbiased estimation of autocovariances is possible under $N^2/T^3 \rightarrow 0$ when the bias correction is done using $\tilde{V}_T(\infty)$ with the QS kernel.

Remark 9. The estimator $\tilde{\gamma}(\infty)$ may be motivated as a continuously updated estimator of $\gamma = (\gamma_0, \dots, \gamma_{T-1})'$. We note that $E(\hat{\gamma}) \approx \gamma - \iota_T V_T / T$. By replacing V_T with $K'_T \gamma$, we get $E(\hat{\gamma}) \approx \gamma - \iota_T K'_T \gamma / T$. The estimator $\tilde{\gamma}(\infty)$ can be obtained by solving the following equation: $\hat{\gamma} = \tilde{\gamma}(\infty) - \iota_T K'_T \tilde{\gamma}(\infty) / T$.

As before, we use the MSE formula as the device to choose the bandwidth parameter. The bandwidth parameter that minimizes the MSE formula is

$$S^* = \left[\left\{ q k_q^2 / \int k^2(x)dx \right\} \xi^2 T N \right]^{1/(2q+1)}.$$

For the QS kernel function, the bandwidth parameter may be chosen to be:

$$\hat{S}^* = 1.3221(\hat{\xi}^2 T N)^{1/5}. \quad (2)$$

4.4 The truncated kernel

In this subsection, we discuss the bandwidth choice rule for the truncated kernel. Let \check{V}_T and $\check{V}_T(\infty)$, respectively, be the long-run variance estimator and its infinitely iterated version based on the truncated kernel such that:

$$\begin{aligned} \check{V}_T &= \sum_{j=-S}^S \frac{T-|j|}{T} \hat{\gamma}_j = \sum_{j=-S}^S \hat{\gamma}_j^+, \\ \check{V}_T(\infty) &= \left(1 + \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \right) \check{V}_T, \end{aligned}$$

where $K_T^* = (1, 2(T-1)/T, \dots, 2(T-S)/T, 0, \dots, 0)$. The truncated kernel has not been commonly used in long-run variance estimation. The main reason is that the truncated kernel does not guarantee the positive definiteness of the estimator. However, as pointed out by Hahn and Kuersteiner (2007), ensuring the positive definiteness may not be important if the purpose of estimating the long-run variance is the bias correction. Given that the truncated kernel provides a good estimate when w_{it} is an M-dependent process, it is worthwhile to consider the truncated kernel in our context.

The bandwidth choice rules in the previous subsections are not useful for the truncated kernel. We note that $k_q = 0$ for any finite q for the truncated kernel since it is flat around the origin. Therefore, the bias term of order S^{-q} disappears from the MSE formula and the bandwidth choice rule presented above recommends that the bandwidth be as small as possible, which obviously does not work in practice.

This observation implies that we need alternative MSE formulas. The following theorem presents the leading terms of the MSEs of \tilde{V}_T and $\tilde{V}_T(\infty)$. The proof is in the Appendix.

Theorem 6. *Suppose that Assumptions 1 and 2 are satisfied. Suppose also that $S \rightarrow \infty$ and $S/T \rightarrow 0$. Then, $\tilde{V}_T - V_T \rightarrow_p 0$ and*

$$MSE(\tilde{V}_T) = \left(-2 \sum_{j=S+1}^{\infty} \gamma_j - 2V \frac{S}{T} \right)^2 + 4V^2 \frac{S}{NT} + o \left(\left(\sum_{j=S+1}^{\infty} \gamma_j \right)^2 + \frac{S^2}{T^2} + \frac{S}{NT} \right).$$

We also have $\tilde{V}_T(\infty) - V_T \rightarrow_p 0$ and

$$MSE\{\tilde{V}_T(\infty)\} = 4 \left(\sum_{j=S+1}^{\infty} \gamma_j \right)^2 + 4V^2 \frac{S}{NT} + o \left(\left(\sum_{j=S+1}^{\infty} \gamma_j \right)^2 + \frac{S}{NT} \right) + O \left(\frac{S^4}{T^4} \right).$$

The theorem does not explicitly give the rate of convergence of the estimators because it is difficult to evaluate the order of the term $\sum_{j=S+1}^{\infty} \gamma_j$.

We choose the bandwidth using the MSE formulas. As before, we estimate the approximate MSEs based on the formula that is valid when w_{it} follows the panel AR(1) process. Let δ be the AR(1) coefficient and $\hat{\delta}$ be Hahn and Kuersteiner's (2002) estimator. In the AR(1) model, we have $V = (1 + \delta)/(1 - \delta)$ and $\sum_{j=S+1}^{\infty} \gamma_j = \delta^S/(1 - \delta)$. Thus, the bandwidth choice rule for \tilde{V}_T is

$$\hat{S}^* = \arg \min_{S \in \{1, \dots, T-1\}} \left(-2 \frac{\hat{\delta}^S}{1 - \hat{\delta}} - 2 \frac{1 + \hat{\delta}}{1 - \hat{\delta}} \frac{S}{T} \right)^2 + 4 \left(\frac{1 + \hat{\delta}}{1 - \hat{\delta}} \right)^2 \frac{S}{NT}, \quad (3)$$

and that for $\tilde{V}_T(\infty)$ is

$$\hat{S}^* = \arg \min_{S \in \{1, \dots, T-1\}} 4 \left(\frac{\hat{\delta}^S}{1 - \hat{\delta}} \right)^2 + 4 \left(\frac{1 + \hat{\delta}}{1 - \hat{\delta}} \right)^2 \frac{S}{NT}. \quad (4)$$

These bandwidth choice rules are similar to that considered by Hahn and Kuersteiner (2007). In particular, $\sum_{j=S+1}^{\infty} \gamma_j$ is difficult to estimate in panel data settings and the idea of using an AR(1) model to approximate this term is proposed by Hahn and Kuersteiner (2007).

5 Monte Carlo simulations

This section reports the results of Monte Carlo simulations. The simulations are conducted on Ox 5.10 (Doornik (2007)).

5.1 Design

The data-generating process used in the experiments is the following:

$$y_{it} = w_{it} + \eta_i,$$

where $\eta_i \sim i.i.d.N(0, \sigma_\eta^2)$, and w_{it} follows an AR(1) process:

$$w_{it} = \alpha w_{i,t-1} + \epsilon_{it},$$

and $\epsilon_{it} \sim i.i.d.N(0, \sigma^2)$. The initial observations are generated from the stationary distribution. Specifically, we generate (w_{i0}, ϵ_{i0}) from:

$$\begin{pmatrix} w_{i0} \\ \epsilon_{i0} \end{pmatrix} \sim N \left(0, \begin{pmatrix} \sigma^2 \frac{1}{1-\alpha^2} & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix} \right).$$

We set the value of σ^2 such that $\gamma_0 = 1$ (i.e., $\sigma^2 = 1 - \alpha^2$) and fix the value of σ_η^2 at $\sigma_\eta^2 = 1$. Note that σ_η^2 does not affect the results as η_i is eliminated in the estimation of autocovariances. The value of σ^2 only affects the scale of the estimator and does not have any essential effect on the Monte Carlo results. Each experiment is characterized by the vector of (N, T, α) . We set $N = 20$; $T = 5, 10, 25, 50$; and $\alpha = 0, 0.5, 0.9$. We consider several different procedures. The first procedure considered is the within-group autocovariances (i.e., $\hat{\gamma}_k$; we call these “WG”). The other procedures are bias-corrected estimators. The QS kernel and the truncated kernel are used in the bias correction. For each kernel, we consider three different procedures: the one-time bias-corrected autocovariance (i.e., $\tilde{\gamma}_k(1)$; we call these “BWG”); the two-time bias-corrected autocovariance (i.e., $\tilde{\gamma}_k(2)$; we call these “BWG2”); the autocovariance estimators obtained after infinite iterations (i.e., $\tilde{\gamma}_k(\infty)$; we call these “IB”). The bandwidth parameters for the QS kernel are chosen using formula (1) for “BWG” and formula (2) for “IB”. For “BWG2”, the first iteration uses formula (1) and the second iteration uses formula (2). Similarly, the bandwidth parameters for the truncated kernel are chosen using formula (3) for “BWG” and formula (4) for “IB”. Formula (3) is used for the first iteration of “BWG2” and formula (4) is used for the second iteration. The number of replications is 5000.

We have also tried other specifications whose results are not reported here. Those results are obtained from the author upon request and are briefly summarized here. We have tried cases where $(w_{i0}, \epsilon_{i0}) = 0$ for all i throughout the simulations. However, the specification of the initial observation does not appear to affect the simulation results. We have considered other cross-sectional sample sizes. The results regarding the biases of the estimators are similar across different cross-sectional sample sizes, while the

cross-sectional sample size affects the standard deviations of the estimators. We have also considered cases in which w_{it} follows an ARMA model. The results from the ARMA model are similar to the results we present here.

5.2 Results

Tables 1 and 2 summarize the results of the experiments. Table 1 presents the results for the QS kernel and Table 2 presents the results for the truncated kernel. For each procedure, we report the biases and standard deviations (std) of the estimates of the zeroth-, first- and second-order autocovariances. We also report the theoretical approximation of the bias of $\hat{\gamma}_k$ (i.e., V_T/T) in the column entitled “Tbias”.

[Tables 1-2 about here]

We first examine the results for “WG”. The biases of “WG” are large when the length of the time series is short and when the degree of persistence is large ($\alpha = 0.9$). These findings are consistent with our theoretical results. Moreover, “Tbias” and the biases of “WG” are reasonably similar.

Next, we investigate the performance of the procedures developed in this paper that have bias-reducing properties. While the “BWG” procedure alleviates the bias, the “BWG2” procedure mitigates the bias more effectively than does “BWG”. The gain from iterating the bias correction is substantial, particularly when T is small ($T = 5$ and 10). The “IB” procedure eliminates the bias even more effectively than does “BWG2”, although “IB” exhibits somewhat larger standard deviations when $T = 5$ or $\alpha = 0.9$. The effectiveness of our bias correction crucially depends on T and α . (In the current setting, α measures the persistence of individual dynamics.) When there is no persistence in individual dynamics ($\alpha = 0$), our bias correction works very well and can completely eliminate the bias even if T is small. Moreover, when $\alpha = 0$, the bias correction does not inflate the standard deviation by much. However, when there is strong persistence ($\alpha = 0.9$), a long time series is required to obtain estimates that are mostly unbiased and the standard deviations of the bias-corrected estimators are somewhat large compared with “WG”. Nevertheless, our procedures (in particular, “IB”) are able to improve the within-group autocorrelation estimators substantially. Compared with the QS kernel, the truncated kernel typically yields a better bias correction. On the other hand, the choice of the kernel function does not have a large impact on the standard deviations.

[Figures 1-3 about here]

Lastly, we evaluate the quality of the normal approximation. We compare the distribution of each estimator with the QS kernel for γ_0 with the normal distribution with same mean and variance using QQ plots. These QQ plots are presented in Figures 1-3, where we consider three cases; $(N, T, \alpha) = (20, 5, 0)$, $(20, 5, 0.9)$ and $(200, 5, 0.9)$, respectively. Figure 1 shows that the normal approximation works reasonably well when the degree of persistence is low. On the contrary, when $\alpha = 0.9$ (Figure 2), the normal approximation may not be accurate. In particular, the distribution of “IB” deviates from the

normal distribution considerably. Nonetheless, in Figure 3, we see that the normal approximation works (although the distributions are not centered around the true value due to the bias) when the sample size is reasonably large even if the degree of persistence is high.

To sum up, we observe that the procedures developed in this paper effectively reduce the biases. They provide reliable estimates of the autocovariances, particularly when the time dimension is moderately large or when the persistence is not very large. On the other hand, when the length of the time series is short and the persistence is large, our procedures may not be able to eliminate the biases completely, although they perform remarkably better than does the conventional procedure. We also see that the asymptotic normal approximation is accurate in sample sizes that we often encounter. Given the results of the experiments, we believe that applied researchers could benefit by using the procedures developed in this analysis. In particular, the “IB” procedure with the truncated kernel works remarkably well.

6 Extensions

In this section, we consider several extensions of the methods developed in this paper.

6.1 Other related quantities

We consider the estimation of other related quantities and see how the estimators developed in the previous sections are useful for this purpose.

Let ρ_k be the k -th-order autocorrelation of w_{it} (i.e., $\rho_k = \gamma_k/\gamma_0$). We consider estimating ρ_k based on estimates of γ_k and γ_0 . Let $\tilde{\rho}_k$ be the estimator for ρ_k based on bias-corrected estimators:

$$\tilde{\rho}_k = \frac{\tilde{\gamma}_k}{\tilde{\gamma}_0}.$$

It is easy to see that $\tilde{\rho}_k$ is consistent by the continuous mapping theorem. It is also easy to see that, by the Delta method, $\tilde{\rho}_k$ is asymptotically normal with zero mean.

Partial autocorrelation is another popular measure of dependence over time. Let α_k signify the k th partial autocorrelation. Note that α_k is the population value of the coefficient on $w_{i,t-k}$ in the regression of w_{it} on $w_{i,t-1}, \dots, w_{i,t-k}$ (this does not mean that w_{it} follows an AR(k) model). We recommend using bias-corrected estimators of γ s to estimate α_k . The estimator, $\tilde{\alpha}_k$, is obtained by solving the following equation:

$$\begin{pmatrix} * \\ * \\ \dots \\ \tilde{\alpha}_k \end{pmatrix} = \begin{pmatrix} \tilde{\gamma}_0 & \tilde{\gamma}_1 & \dots & \tilde{\gamma}_{k-1} \\ \tilde{\gamma}_1 & \tilde{\gamma}_0 & \dots & \tilde{\gamma}_{k-2} \\ \dots & \dots & \dots & \dots \\ \tilde{\gamma}_{k-1} & \tilde{\gamma}_{k-2} & \dots & \tilde{\gamma}_0 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \\ \dots \\ \tilde{\gamma}_k \end{pmatrix},$$

where *s are elements of the vector irrelevant to our discussion. The consistency of $\tilde{\alpha}_k$ is proved by a simple application of the continuous mapping theorem. The asymptotic distribution of $\tilde{\alpha}_k$ can be derived easily with the Delta method and it is centered around zero.

Remark 10. The results in Lee (2008a) can be used to find the probability limit and the asymptotic distribution of a partial autocorrelation coefficient estimator based on $\hat{\gamma}$ s (not $\tilde{\gamma}$). However, we cannot use the bias correction method by Lee (2008a) for the estimation of the partial autocorrelations. The strategy that Lee (2008a) adopts is to select the correct order of the autoregression and then mitigate the bias of the estimates of the coefficients in correctly specified AR (p) models.

Lastly, we consider the variance of individual effects. Let σ_η^2 be the variance of η_i . A natural estimator of σ_η^2 may be the between-group variance:

$$\hat{\sigma}_\eta^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{y}_i - \bar{y})^2,$$

where $\bar{y} = \sum_{i=1}^N \sum_{t=1}^T y_{it} / (NT)$. As for $\hat{\gamma}_k$, we show below that $\hat{\sigma}_\eta^2$ exhibits bias whose leading term converges to the long-run variance of w_{it} . However, it turns out that the direction of the bias of $\hat{\sigma}_\eta^2$ is upward. A bias-corrected estimator of σ_η^2 may be given as

$$\tilde{\sigma}_\eta^2 = \hat{\sigma}_\eta^2 - \frac{1}{T} \hat{V}_T.$$

We need assumptions on the distribution of η_i in addition to the assumptions on w_{it} to study the asymptotic properties of the $\hat{\sigma}_\eta^2$ and $\tilde{\sigma}_\eta^2$.

Assumption 3. 1. $\{\eta_i\}_{i=1}^N$ are i.i.d. across individuals.

2. $E(\eta_i^4) < \infty$.

3. w_{it} and η_i are independent for any t .

Theorem 7. 1. Suppose that Assumptions 1 and 3 are satisfied. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$, it follows that $\hat{\sigma}_\eta^2 \rightarrow_p \sigma_\eta^2$.

Suppose that Assumptions 1, 2 and 3 are satisfied. Then, as $N \rightarrow \infty$, $T \rightarrow \infty$,

$$\sqrt{N} \left(\hat{\sigma}_\eta^2 - \sigma_\eta^2 - \frac{1}{T} V_T \right) \rightarrow_d N \left(0, [E\{(\eta_i - \mu)^4\} - \sigma_\eta^4] \right).$$

2. Suppose that Assumptions 1, 2 and 3 are satisfied. Suppose also that $r_{N,T} T^{-1} \sqrt{N} \rightarrow 0$. Then, as $N \rightarrow \infty$, $T \rightarrow \infty$,

$$\sqrt{N} (\tilde{\sigma}_\eta^2 - \sigma_\eta^2) \rightarrow_d N \left(0, [E\{(\eta_i - \mu)^4\} - \sigma_\eta^4] \right).$$

Remark 11. While $T \rightarrow \infty$ is required for the consistency of $\hat{\sigma}_\eta^2$, the rate of convergence of $\hat{\sigma}_\eta^2$ is \sqrt{N} , not \sqrt{NT} . Roughly speaking, this is because we can observe only one η for each individual. Note also that, contrary to the result for $\hat{\gamma}_k$, we do not need a condition on the relationship between the rates of N and T for the results for $\hat{\sigma}_\eta^2$. The reason is that $\hat{\sigma}_\eta^2$ does not possess a bias term of order $1/T^2$.

6.2 Fixed-effects regression models

Another extension is the estimation of the autocovariance structure of error terms in panel regression models. We consider the following panel regression model:

$$z_{it} = x'_{it}\beta + \eta_i + w_{it},$$

where x_{it} is the vector of regressors and β is the vector of parameters to be estimated. Let $\hat{\beta}$ be an estimator of β . Our analysis is based on the residuals from this estimation:

$$\hat{y}_{it} = z_{it} - x'_{it}\hat{\beta}.$$

Let $\hat{\gamma}_k^*$ be the within-group estimator of the k -th-order autocovariance of w_{it} computed using the residuals

$$\hat{\gamma}_k^* = \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (\hat{y}_{it} - \bar{\hat{y}}_i)(\hat{y}_{i,t-k} - \bar{\hat{y}}_i),$$

where $\bar{\hat{y}}_i = \sum_{t=1}^T \hat{y}_{it}/T$. We also consider how the estimation error in $\hat{\beta}$ affects the long-run variance estimation. Let \tilde{V}_T^* be a long-run variance estimator based on \hat{y}_{it} s so that

$$\tilde{V}_T^* = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{T-|j|}{T} \hat{\gamma}_j^* = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \hat{\gamma}_j^{*+},$$

where

$$\hat{\gamma}_j^{*+} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=|j|+1}^T (\hat{y}_{it} - \bar{\hat{y}}_i)(\hat{y}_{i,t-|j|} - \bar{\hat{y}}_i) = \frac{T-|j|}{T} \hat{\gamma}_j^*.$$

We rely on the following assumption to study the asymptotic properties of $\hat{\gamma}_k^*$ and \tilde{V}_T^* .

Assumption 4. 1. $\hat{\beta} - \beta = O_p(1/\sqrt{NT})$.

2. $\{w_{it}, x'_{it} - E_i(x_{it})'\}$ are i.i.d. across individual and strictly stationary over t , where $E_i(x_{it})$ is the expectation of x_{it} given η_i .
3. Let v_{at} be the a -th element of the vector $(w_{it}, x'_{it} - E_i(x_{it})')'$. We have $\sum_{j=-\infty}^{\infty} |E(v_{at}v_{b,t-j})| < \infty$ for any a, b .
4. Let $\text{cum}_{a,b,c,d}(0, j_1, j_2, j_3)$ be the fourth-order cumulant of $(v_{a0}, v_{bj_1}, v_{cj_2}, v_{dj_3})$. For any (a, b, c, d) ,

$$\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \sum_{j_3=-\infty}^{\infty} |\text{cum}_{a,b,c,d}(0, j_1, j_2, j_3)| < \infty.$$

Assumption 4.1 states that $\hat{\beta}$ is \sqrt{NT} -consistent. For example, the fixed-effects estimator satisfies this assumption when the regressors are strictly exogenous. Assumption 4.2 allows the individual effect, η_i , to enter the regressor, x_{it} , in an additive fashion. Assumption 4.3 states that the serial correlation in $(w_{it}, x'_{it} - E_i(x_{it})')'$ vanishes sufficiently fast as the time difference increases. Assumption 4.4 is a technical assumption and it restricts the magnitude of fourth-order moments.

Let $\hat{\gamma}_k$ and \tilde{V}_T denote the k -th within-group autocovariance and the long-run variance estimator, respectively, where we ignore the estimation of β , as considered in previous sections.

Theorem 8. *Suppose that Assumption 4 is satisfied.*

1. *As $N \rightarrow \infty$ and $T \rightarrow \infty$, we have:*

$$\sqrt{NT}(\hat{\gamma}_k^* - \hat{\gamma}_k) = (E[w_{it}\{x_{i,t-k} - E_i(x_{it})\}'] + E[w_{i,t-k}\{x_{it} - E_i(x_{it})\}']) \sqrt{NT}(\hat{\beta} - \beta) + o_p(1).$$

2. *Assume that $k(\cdot) \in \mathcal{K}_1$. As $N \rightarrow \infty$, $T \rightarrow \infty$, $S \rightarrow \infty$ and $S^2/T \rightarrow 0$, we have*

$$\tilde{V}_T^* - \tilde{V}_T = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

The proof is included in the Appendix. When the regressors are strictly exogenous such that $E[w_{i,t_1}\{x_{it_2} - E_i(x_{it_2})\}'] = 0$ for any t_1 and t_2 , the theorem implies that all the asymptotic results for $\hat{\gamma}_k$ presented in previous sections hold for $\hat{\gamma}_k^*$. However, when the regressors are not strictly exogenous (e.g., when the regressors are merely predetermined), the asymptotic distributions of $\hat{\gamma}_k^*$ and $\hat{\gamma}_k$ are different and the estimation error of $\hat{\beta}$ affects the asymptotic behavior of $\hat{\gamma}_k^*$. This observation is well known in the time series literature (see, e.g., Hayashi (2000, pp. 144-146)). On the other hand, the estimation error in $\hat{\beta}$ does not affect the asymptotic behavior of \tilde{V}_T^* because the rate of convergence of \tilde{V}_T is slower than $1/\sqrt{NT}$ when the bandwidth is optimally chosen. This implies that we can apply the bias correction developed for $\hat{\gamma}_k$ s to $\hat{\gamma}_k^*$ s without any modification.

An application of this procedure is the GLS estimation of fixed effects regression models with strictly exogenous regressors. Here, the regressors must be strictly exogenous because the GLS estimator is not necessarily consistent when the regressors are merely predetermined (see, e.g., Hayashi (2000, p. 416)). The GLS estimation of these models is investigated by Kiefer (1980), Hansen (2007) and Hausman and Kuersteiner (2008). Let

$$\Upsilon = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & & \gamma_{T-2} \\ \vdots & & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{pmatrix}, \quad \tilde{\Upsilon} = \begin{pmatrix} \tilde{\gamma}_0 & \frac{T-1}{T}\tilde{\gamma}_1 & \cdots & \frac{1}{T}\tilde{\gamma}_{T-1} \\ \frac{T-1}{T}\tilde{\gamma}_1 & \tilde{\gamma}_0 & & \frac{2}{T}\tilde{\gamma}_{T-2} \\ \vdots & & \ddots & \vdots \\ \frac{1}{T}\tilde{\gamma}_{T-1} & \frac{2}{T}\tilde{\gamma}_{T-2} & \cdots & \tilde{\gamma}_0 \end{pmatrix}.$$

We note that using the QS kernel guarantees that $\tilde{\Upsilon}$ is positive definite. To see this, we observe the following decomposition:

$$\tilde{\Upsilon} = \begin{pmatrix} \hat{\gamma}_0 & \frac{T-1}{T}\hat{\gamma}_1 & \cdots & \frac{1}{T}\hat{\gamma}_{T-1} \\ \frac{T-1}{T}\hat{\gamma}_1 & \hat{\gamma}_0 & & \frac{2}{T}\hat{\gamma}_{T-2} \\ \vdots & & \ddots & \vdots \\ \frac{1}{T}\hat{\gamma}_{T-1} & \frac{2}{T}\hat{\gamma}_{T-2} & \cdots & \hat{\gamma}_0 \end{pmatrix} + \frac{1}{T^2} \begin{pmatrix} T & T-1 & \cdots & 1 \\ T-1 & T & & 2 \\ \vdots & & \ddots & \vdots \\ 1 & 2 & \cdots & T \end{pmatrix} \tilde{V}_T.$$

We note that it is well known in the time series literature that the first term on the right-hand side is positive semi-definite. Now, we have $\tilde{V}_T > 0$ when we use the QS kernel, and the matrix between $1/T^2$ and \tilde{V}_T can be written as:

$$\iota_T \iota_T' + \sum_{j=1}^{T-1} \iota_{1,j} \iota_{1,j}' + \sum_{j=1}^{T-1} \iota_{j+1,k+1} \iota_{j+1,k+1}',$$

where $\iota_{a,b}$ is the $(k+1) \times 1$ vector whose a -th to b -th elements are one and other elements are zero, and this formula tells us that the matrix is positive definite. It therefore follows that $\tilde{\Upsilon}$ is positive definite.

The GLS transformation of the panel regression model gives

$$\Upsilon^{-1/2} z_i = \Upsilon^{-1/2} x_i \beta + \Upsilon^{-1/2} \iota_T \eta_i + \Upsilon^{-1/2} w_i,$$

for some choice of $\Upsilon^{-1/2}$, where $z_i = (z_{i1}, \dots, z_{iT})'$ and x_i and w_i are defined similarly. This transformation yields a serially uncorrelated error term. The (infeasible) GLS estimator is obtained by eliminating the fixed effects by multiplying the annihilator matrix of $\Upsilon^{-1/2} \iota_T$ and then applying the OLS estimator. A feasible GLS estimator may be obtained by replacing the matrix Υ by $\tilde{\Upsilon}$ such that

$$\begin{aligned} \hat{\beta}_{FGLS} = & \left\{ \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} x_i - \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} \iota_T (\iota_T' \tilde{\Upsilon}^{-1} \iota_T)^{-1} \iota_T' \tilde{\Upsilon}^{-1} x_i \right\}^{-1} \\ & \times \left\{ \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} z_i - \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} \iota_T (\iota_T' \tilde{\Upsilon}^{-1} \iota_T)^{-1} \iota_T' \tilde{\Upsilon}^{-1} z_i \right\}. \end{aligned}$$

The asymptotic variance of $\hat{\beta}_{FGLS}$ may be estimated by

$$\left\{ \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} x_i - \frac{1}{NT} \sum_{i=1}^N x_i' \tilde{\Upsilon}^{-1} \iota_T (\iota_T' \tilde{\Upsilon}^{-1} \iota_T)^{-1} \iota_T' \tilde{\Upsilon}^{-1} x_i \right\}^{-1}. \quad (5)$$

We examine the properties of the feasible GLS estimator through simulations. We consider the case in which x_{it} is scalar. The data are generated in the following way: $x_{it} \sim i.i.d. U[-1, 1]$; and y_{it} is generated in the same way as in the experiments in Section 5. We fix the value of β at $\beta = 1$. We set $N = 50$, $T = 5, 10, 20$, $\alpha = 0, 0.5, 0.9$, $\sigma_\eta^2 = 1$ and σ^2 is set such that $\gamma_0 = 1$. We examine the following four estimators of β : the within-group estimator (“WG”); the (infeasible) GLS estimator (“GLS”); the feasible GLS estimator with $\tilde{\Upsilon}$ based on $\tilde{\gamma}_k(\infty)$ s (“FGLS”); the estimator considered by Kiefer (1980) (“KGLS”) which is the feasible GLS estimator applied to the equation transformed by the fixed effects transformation. For each estimator, we compute the bias and the standard deviation (std). We also give the mean of the standard error (meanse) for each estimator and the coverage probability of the 95% confidence interval based on each estimator, where the confidence interval is constructed by the standard formula: estimate ± 1.96 (standard error). The standard errors are computed using formula (5.2.8) of Hayashi (2000) for “WG”, formula (5) with Υ instead of $\tilde{\Upsilon}$ for “GLS”, formula (5) for “FGLS”, and the formula in page 199 of Kiefer (1980) for “KGLS”. We note that the standard error for “WG” allows serial dependence but it assumes homoskedasticity conditional on the regressor.

Table 3 summarizes the results. The biases of all the estimators are negligible and the estimators should be compared in terms of their standard deviations. Naturally, “GLS” exhibits the smallest standard deviation among the estimators compared. Both “FGLS” and “KGLS” exhibit lower standard deviations than does “WG”. Among these feasible GLS estimators, “FGLS” has the smallest standard deviation and its standard deviation is similar to that of “GLS”. Moreover, the standard error and the standard deviation of “FGLS” are reasonably similar. The coverage rate of the 95% confidence interval

based on “FGLS” is close to 0.95. Although the confidence intervals based on “WG” and “GLS” have better coverage rates than that of “FGLS” does when $\alpha = 0.9$, the confidence interval based on “FGLS” performs better than does that based on “KGLS”. These results indicate the usefulness of “FGLS”. It has a small standard deviation and its standard error is reliable. These results imply that the asymptotically unbiased autocovariance estimators developed in this paper are useful for the GLS estimation of fixed effects regression models.

[Table 3 about here]

References

- Abowd, J. M. & D. Card (1989) On the covariance structure of earnings and hours changes, *Econometrica* **57**(2), 411–445.
- Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**(4), 1121–1159.
- Anderson, T. W. & C. Hsiao (1981) Estimation of dynamics models with error components, *Journal of the American Statistical Association* **76**(375), 598–606.
- Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**(3), 817–858.
- Arellano, M. (2003) *Panel Data Econometrics*, Oxford University Press.
- Arellano, M. & S. Bond (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economics Studies* **58**, 277–297.
- Baltagi, B. H. & Q. Li (1994) Estimating error component models with general MA(q) disturbances, *Econometric Theory* **10**, 396–408.
- Brillinger, D. R. (1981) *Time Series: Data Analysis and Theory*, Holden Day. Inc.
- Bun, M. J. & J. F. Kiviet (2006) The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models, *Journal of Econometrics* **132**, 409–444.
- Cameron, A. C. & P. K. Trivedi (2005) *Microeconometrics, Methods and Applications*, Cambridge University Press.
- Chamberlain, G. (1984) Panel data, in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, chapter 22, pp. 1247–1318. Elsevier.
- den Haan, W. J. & A. T. Levin (1997) A practitioner’s guide to robust covariance matrix estimation, in G. S. Maddala and C. R. Rao (eds.), *Handbook of Statistics*, Vol. 15, pp. 299–342. Elsevier.
- Doornik, J. A. (2007) *Ox, - An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.
- Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* **70**(4), 1639–1657.
- Hahn, J. & G. Kuersteiner (2004) Bias reduction for dynamic nonlinear panel models with fixed effects, mimeo.
- Hahn, J. & G. Kuersteiner (2007) Bandwidth choice for bias estimators in dynamic nonlinear panel models, mimeo.
- Hansen, C. B. (2007) Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects, *Journal of Econometrics* **140**, 670–694.
- Hausman, J. & G. Kuersteiner (2008) Difference in difference meets generalized least squares: Higher order properties of hypotheses tests, *Journal of Econometrics* **144**, 371–391.
- Hayashi, F. (2000) *Econometrics*, Princeton University Press.
- Holtz-Eakin, D., W. Newey & H. S. Rosen (1988) Estimating vector autoregressions with panel data, *Econometrica* **6**, 1371–1395.
- Kakizawa, Y. (1999) Notes on the asymptotic efficiency of sample covariances in gaussian vector stationary process, *Journal of Time Series Analysis* **20**(5), 551–558.

- Kakizawa, Y. & M. Taniguchi (1994) Asymptotic efficiency of sample covariances in a gaussian stationary process, *Journal of Time Series Analysis* **15**, 303–311.
- Kiefer, N. M. (1980) Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance, *Journal of Econometrics* **14**, 195–202.
- Kiviet, J. F. (1995) On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* **68**, 53–78.
- Lee, Y. (2008a) Bias correction in dynamic panel models under time series misspecification, mimeo.
- Lee, Y. (2008b) Nonparametric estimation of dynamic panel models with fixed effects, mimeo.
- Lillard, L. A. & R. J. Willis (1978) Dynamic aspects of earning mobility, *Econometrica* **46**(5), 985–1012.
- MaCurdy, T. E. (1982) The use of time series processes to model the error structure of earnings in a longitudinal data analysis, *Journal of Econometrics* **18**, 83–114.
- Nickell, S. (1981) Biases in dynamic models with fixed effects, *Econometrica* **49**(6), 1417–1426.
- Parzen, E. (1957) Consistent estimates of the spectrum of a stationary time series, *Annals of Mathematical Statistics* **28**(2), 329–348.
- Phillips, P. C. B. & H. R. Moon (1999) Linear regression limit theory for nonstationary panel data, *Econometrica* **67**(5), 1057–1111.
- Porat, B. (1987) Some asymptotic properties of the sample covariances of gaussian autoregressive moving average process, *Journal of Time Series Analysis* **8**, 205–220.
- Solon, G. (1984) Estimating autocorrelations in fixed-effects models, NBER, Technical Working Paper No. 32.

A Technical appendix

A.1 Proof of Theorem 1

Proof. We have the following decomposition:

$$\begin{aligned}\hat{\gamma}_k &= \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T w_{it} w_{i,t-k} - \frac{1}{N} \sum_{i=1}^N (\bar{w}_i)^2 \\ &\quad - 2 \frac{k}{N(T-k)} \sum_{i=1}^N (\bar{w}_i)^2 + \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i + \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=T-k+1}^T w_{it} \bar{w}_i.\end{aligned}$$

The first term on the right-hand side of the equation converges to γ_k by Lemma 1. The second and third terms are $o_p(1)$ by Lemma 2 and the fourth and fifth terms are $o_p(1)$ by Lemma 3. It follows that $\hat{\gamma}_k \rightarrow_p \gamma_k$. ■

A.2 Proof of Theorem 2

Proof. We have the following decomposition:

$$\begin{aligned}&\sqrt{NT} \left(\hat{\gamma}_k - \gamma_k + \frac{1}{T} V_T \right) \\ &= \sqrt{NT} \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) - \sqrt{NT} \left\{ \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 - \frac{1}{T} V_T \right\} \\ &\quad + 2\sqrt{NT} \frac{k}{N(T-k)} \sum_{i=1}^N (\bar{w}_i)^2 + \sqrt{NT} \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i + \sqrt{NT} \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=T-k+1}^T w_{it} \bar{w}_i.\end{aligned}$$

The first term on the right-hand side is asymptotically normal by Lemma 1, and the second and third terms are $o_p(1)$ by Lemma 2. The last two terms on the right-hand side are $o_p(1)$ by Lemma 3. ■

A.3 Proof of Theorem 4

Proof. First, we consider the bias:

$$E(\tilde{V}_T) = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) E(\hat{\gamma}_j^+).$$

Note that:

$$E(\hat{\gamma}_j^+) = \frac{T-|j|}{T} \gamma_j - \frac{T-|j|}{T} \frac{1}{T} V_T + B_{jT},$$

where

$$B_{jT} = E \left\{ -2 \frac{|j|}{NT} \sum_{i=1}^N (\bar{w}_i)^2 + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{|j|} w_{it} \bar{w}_i + \frac{1}{NT} \sum_{i=1}^N \sum_{t=T-|j|+1}^T w_{it} \bar{w}_i \right\}.$$

We have

$$\begin{aligned} E(\tilde{V}_T - V_T) &= \sum_{j=-T+1}^{T-1} \left\{ k\left(\frac{j}{S}\right) - 1 \right\} \frac{T-|j|}{T} \gamma_j \\ &\quad - \frac{1}{T} V_T \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{T-|j|}{T} + \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) B_{jT}. \end{aligned}$$

As shown by Parzen (1957), S^q times the first term on the right-hand side converges to $-k_q V^{(q)}$. This implies that the first term is of order $O(S^{-q})$. Next, we consider the second term. Observing that $V_T \rightarrow V$ and $S^{-1} \sum_{j=-T+1}^{T-1} k(j/S) \rightarrow \int_{-1}^1 k(x) dx$, the second term is of order $O(S/T)$. The first term is therefore of an order larger than the second term when $S^{q+1}/T \rightarrow 0$, which is satisfied when $N^{q+1}/T^q \rightarrow 0$ and $S^{2q+1}/(NT) \rightarrow \tau$. The first term and the second term are of the same order when $S^{q+1}/T \rightarrow \tau$.

We consider the third term on the right-hand side. We observe that

$$\begin{aligned} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) E \left\{ 2 \frac{|j|}{NT} \sum_{i=1}^N (\bar{w}_i)^2 \right\} &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{2|j|}{T^2} V_T \\ &\leq \frac{2S}{T^2} V_T \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) = O\left(\frac{S^2}{T^2}\right), \end{aligned}$$

and that

$$\begin{aligned} &\left| \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{|j|} w_{it} \bar{w}_i \right) \right| \\ &\leq \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{|j|}{T^2} \sum_{m=1}^T |\gamma_m| \leq \frac{S}{T^2} \sum_{m=1}^T |\gamma_m| \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) = O\left(\frac{S^2}{T^2}\right), \end{aligned}$$

by Lemma 3. Similarly, we can show that

$$\left| \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=T-|j|+1}^T w_{it} \bar{w}_i \right) \right| = O\left(\frac{S^2}{T^2}\right).$$

Therefore, we have

$$\left| \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) B_{jT} \right| = O\left(\frac{S^2}{T^2}\right).$$

Therefore, we have that $E(\tilde{V}_T - V_T) \rightarrow 0$ if $S \rightarrow \infty$ and $S/T \rightarrow 0$. Moreover, when $S^{q+1}/T \rightarrow 0$,

$$S^q E(\tilde{V}_T - V_T) \rightarrow -k_q V^{(q)},$$

and, when $S^{q+1}/T \rightarrow \tau$, where $0 < \tau < \infty$, we have

$$S^q E \left(\tilde{V}_T - V_T \right) \rightarrow -k_q V^{(q)} - \tau V \int k(x) dx.$$

Next, we consider the variance. We note that \hat{V}_T is the sample average across cross-sections of the long-run variance estimator for each time series. Let

$$\tilde{V}_T = \frac{1}{N} \sum_{i=1}^N \tilde{V}_{T,i},$$

where

$$\tilde{V}_{T,i} \equiv \sum_{j=-T+1}^{T-1} k \left(\frac{j}{S} \right) \frac{1}{T} \sum_{t=|j|+1}^T (y_{it} - \bar{y}_i)(y_{i,t-|j|} - \bar{y}_i).$$

Therefore, we have

$$\text{var}(\tilde{V}_T) = \frac{1}{N} \text{var}(\tilde{V}_{T,i}).$$

We verify Assumptions B, C and D in Andrews (1991), under which we can use the variance formula for $\tilde{V}_{T,i}$ provided by Andrews (1991). Note that θ , $\hat{\theta}$ and $V_i(\theta)$ in Assumptions B, C and D of Andrews (1991) are η_i , \bar{y}_i and $y_{it} - \eta_i$, respectively, in our case. Observing that $\partial(y_{it} - \eta_i)/(\partial \eta_i) = -1$, we can easily verify that Assumptions B, C and D are satisfied. Therefore, we have

$$\frac{NT}{S} \text{var}(\tilde{V}_T) \rightarrow 2V^2 \int k(x)^2 dx.$$

This also implies that $\text{var}(\tilde{V}_T) \rightarrow 0$ if $S/(NT) \rightarrow 0$.

For the first bias term and the variance term to be of the same order, we need $S^{2q+1}/(NT) \rightarrow \tau$. For these two terms to be of larger order than the second bias term, we need $S^{q+1}/T \rightarrow 0$, which is equivalent to $N^{q+1}/T^q \rightarrow 0$ when $S = O((NT)^{\frac{1}{2q+1}})$. Therefore, when $N^{q+1}/T^q \rightarrow 0$ and $S^{2q+1}/(NT) \rightarrow 0$, the asymptotic MSE is

$$\lim_{N,T \rightarrow \infty} \frac{NT}{S} \text{MSE}(\tilde{V}_T) = k_q^2 \left(V^{(q)} \right)^2 \tau^{-1} + 2V^2 \int k^2(x) dx.$$

On the other hand, the first and second bias terms are of the same order when $S^{q+1}/T \rightarrow \tau$. These terms are of larger order than the variance term when $\{S/(NT)\}/(S/T)^2 \rightarrow 0$, which is equivalent to $N^{q+1}/T^q \rightarrow \infty$ when $S = O(T^{\frac{1}{q+1}})$. Therefore, when $N^{q+1}/T^q \rightarrow \infty$ and $S^{q+1}/T \rightarrow \tau$, the asymptotic MSE is

$$\lim_{N,T \rightarrow \infty} \frac{T^2}{S^2} \text{MSE}(\tilde{V}_T) = \left\{ -k_q V^{(q)} \tau^{-1} - V \int k(x) dx \right\}^2.$$

■

A.4 Proof of Theorem 5

Proof. In this proof, we use the notation defined in the proof of Theorem 4.

First, note that the asymptotic variance of $\tilde{V}_T(\infty)$ is the same as that of \tilde{V}_T because

$$\tilde{V}_T(\infty) = \left(1 + \frac{\iota'_T K_T}{T - \iota'_T K_T} \right) \tilde{V}_T = \frac{T}{T - \iota'_T K_T} \tilde{V}_T$$

and $T/(T - \iota'_T K_T) \rightarrow 1$. Therefore, we have

$$\frac{NT}{S} \text{var}\{\tilde{V}_T(\infty)\} \rightarrow 2V^2 \int k(x)^2 dx.$$

Next, we consider the bias of $\tilde{V}_T(\infty)$:

$$\begin{aligned} E\{\tilde{V}_T(\infty) - V_T\} &= \sum_{j=-T+1}^{T+1} \left\{ k \left(\frac{j}{S} \right) - 1 \right\} \frac{T - |j|}{T} \gamma_j \\ &\quad - \frac{1}{T} V_T \iota'_T K_T + \frac{\iota'_T K_T}{T - \iota'_T K_T} \sum_{j=-T+1}^{T+1} k \left(\frac{j}{S} \right) \frac{T - |j|}{T} \gamma_j \\ &\quad - \frac{1}{T} V_T \frac{\iota'_T K_T}{T - \iota'_T K_T} \iota'_T K_T + \frac{T}{T - \iota'_T K_T} \sum_{j=-T+1}^{T+1} k \left(\frac{j}{S} \right) B_{jT}. \end{aligned}$$

As in the proof of Theorem 4, we have

$$S^q \sum_{j=-T+1}^{T+1} \left\{ k\left(\frac{j}{S}\right) - 1 \right\} \frac{T-|j|}{T} \gamma_j \rightarrow -k_q V^{(q)}$$

and

$$\frac{T}{T - \iota'_T K_T} \sum_{j=-T+1}^{T+1} k\left(\frac{j}{S}\right) B_{jT} = O(1) O\left(\frac{S^2}{T^2}\right) = O\left(\frac{S^2}{T^2}\right),$$

as $T/(T - \iota'_T K_T) \rightarrow 1$.

We have

$$\begin{aligned} & -\frac{1}{T} V_T \iota'_T K_T + \frac{\iota'_T K_T}{T - \iota'_T K_T} \sum_{j=-T+1}^{T+1} k\left(\frac{j}{S}\right) \frac{T-|j|}{T} \gamma_j - \frac{1}{T} V_T \frac{\iota'_T K_T}{T - \iota'_T K_T} \iota'_T K_T \\ &= \frac{\iota'_T K_T}{T - \iota'_T K_T} \sum_{j=-T+1}^{T+1} \left\{ k\left(\frac{j}{S}\right) - 1 \right\} \frac{T-|j|}{T} \gamma_j = O\left(\frac{S}{T}\right) O(S^{-q}) = o(S^{-q}), \end{aligned}$$

as $\iota'_T K_T / (T - \iota'_T K_T) = O(S/T)$.

Therefore, when $(S^2/T^2)/S^{-q} = S^{q+2}/T^2 \rightarrow 0$, we have

$$S^q E \left\{ \tilde{V}_T(\infty) - V_T \right\} = S^q \sum_{j=-T+1}^{T+1} \left\{ k\left(\frac{j}{S}\right) - 1 \right\} \frac{T-|j|}{T} \gamma_j + S^q o(S^{-q}) + S^q O\left(\frac{S^2}{T^2}\right) \rightarrow -k_q V^{(q)}.$$

Note that, when $S = O((NT)^{\frac{1}{2q+1}})$, the square of the bias and the variance are of the same order and $S^{q+2}/T^2 \rightarrow 0$ is equivalent to $N^{q+2}/T^{3q} \rightarrow 0$. ■

A.5 Proof of Theorem 6

Proof. We note that we can use the results of Theorems 4 and 5 on the variance of the estimators. Therefore, observing that $\int k(x)^2 dx = \int_{-1}^1 1 dx = 2$ for the truncated kernel, we have:

$$\text{var}(\check{V}_T) = 4V^2 \frac{S}{NT} + o\left(\frac{S}{NT}\right), \quad \text{and} \quad \text{var}(\check{V}_T(\infty)) = 4V^2 \frac{S}{NT} + o\left(\frac{S}{NT}\right).$$

The bias of \check{V}_T is

$$E(\check{V}_T - V_T) = -2 \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j - \frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} + \sum_{j=-S}^S B_{jT},$$

where B_{jT} is defined in the proof of Theorem 4.

We examine the order of each term in the bias. First, we have

$$-2 \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j = -2 \sum_{j=S+1}^{\infty} \gamma_j + o\left(\sum_{j=S+1}^{\infty} \gamma_j\right).$$

Second, it is easy to see that

$$\frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} = \frac{1}{T} V_T \left(2S + 1 - \frac{S^2 + S}{T} \right) = 2V \frac{S}{T} + o\left(\frac{S}{T}\right).$$

Next, we consider B_{jT} . We have

$$\sum_{j=-S}^S E \left(2 \frac{|j|}{NT} \sum_{i=1}^N (\bar{w}_i)^2 \right) = \sum_{j=-S}^S \frac{2|j|}{T^2} V_T = \frac{2(S^2 + S)}{T^2} V_T = O\left(\frac{S^2}{T^2}\right)$$

and

$$\left| \sum_{j=-S}^S E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T w_{it} \bar{w}_i \right) \right| \leq \sum_{j=-S}^S \frac{|j|}{T^2} \sum_{m=1}^T |\gamma_m| = \frac{S^2 + S}{T^2} \sum_{m=1}^T |\gamma_m| = O\left(\frac{S^2}{T^2}\right).$$

Similarly,

$$\left| \sum_{j=-S}^S E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=T-|j|+1}^T w_{it} \bar{w}_i \right) \right| = O \left(\frac{S^2}{T^2} \right).$$

Therefore, we have

$$\sum_{j=-S}^S B_{jT} = O \left(\frac{S^2}{T^2} \right).$$

To sum up, the bias of the estimator is

$$E(\check{V}_T - V_T) = -2 \sum_{j=S+1}^{\infty} \gamma_j - 2V \frac{S}{T} + o \left(\sum_{j=S+1}^{\infty} \gamma_j + \frac{S}{T} \right).$$

The mean squared error of \check{V}_T is given as the square of the bias plus the variance.

Next, we consider the bias of $\check{V}_T(\infty)$. First, we note that

$$\frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} = \frac{1 + 2S + (S^2 + S)/T}{T - 1 - 2S + (S^2 + S)/T} = O \left(\frac{S}{T} \right).$$

Given this, we have the decomposition:

$$\begin{aligned} E\{\check{V}_T(\infty) - V_T\} &= E(\check{V}_T - V_T) + \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} E(\check{V}_T) \\ &= -2 \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j - \frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} + \sum_{j=-S}^S B_{jT} \\ &\quad + \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \left(\sum_{j=-S}^S \frac{T-|j|}{T} \gamma_j - \frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} + \sum_{j=-S}^S B_{jT} \right). \end{aligned}$$

We have already shown that $\sum_{j=-S}^S B_{jT} = O(S^2/T^2)$. It holds that

$$\begin{aligned} & -\frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} + \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \sum_{j=-S}^S \frac{T-|j|}{T} \gamma_j - \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \frac{1}{T} V_T \sum_{j=-S}^S \frac{T-|j|}{T} \\ &= -V_T \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} + \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \left(V_T - 2 \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j \right) \\ &= -2 \frac{\iota'_T K_T^*}{T - \iota'_T K_T^*} \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j = o \left(\sum_{j=S+1}^{\infty} \gamma_j \right). \end{aligned}$$

Note that here we use the fact that $\iota'_T K_T^* = \sum_{j=-S}^S (T-|j|)/T$. Therefore, the bias of $\check{V}_T(\infty)$ is

$$\begin{aligned} E\{\check{V}_T(\infty) - V_T\} &= -2 \sum_{j=S+1}^{T-1} \frac{T-j}{T} \gamma_j + o \left(\sum_{j=S+1}^{\infty} \gamma_j \right) + O \left(\frac{S^2}{T^2} \right) \\ &= -2 \sum_{j=S+1}^{\infty} \gamma_j + o \left(\sum_{j=S+1}^{\infty} \gamma_j \right) + O \left(\frac{S^2}{T^2} \right). \end{aligned}$$

The mean squared error formula in the theorem is obtained by the sum of the squares of the bias and the variance. ■

A.6 Proof of Theorem 7

Proof. We have the following decomposition:

$$\hat{\sigma}_\eta^2 = \frac{1}{N} \sum_{i=1}^N (\eta_i - \mu)^2 + \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 - (\bar{y} - \mu)^2 + \frac{2}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)(\eta_i - \mu).$$

The first term on the right-hand side of the equation converges to σ_η^2 by applying the Khintchine law of large numbers, noting that $\{(\eta_i - \mu)^2\}_{i=1}^N$ is an i.i.d. sequence. The second term is $o_p(1)$ by Lemma 2.

We now consider the third term. We have $E(\bar{y} - \mu) = 0$. The variance is

$$\begin{aligned} \text{var}(\bar{y} - \mu) &= \text{var} \left\{ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mu) \right\} = \frac{1}{N} E \left[\left\{ \frac{1}{T} \sum_{t=1}^T (y_{it} - \mu) \right\}^2 \right] \\ &= \frac{1}{N} [E\{(\bar{w}_i)^2\} + \sigma_\eta^2] \leq \frac{1}{N} \left(\frac{1}{T} \sum_{j=-\infty}^{\infty} |\gamma_j| + \sigma_\eta^2 \right) = O\left(\frac{1}{N}\right). \end{aligned}$$

Thus, by the Chebyshev inequality, we have

$$\bar{y} - \mu = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Thus, the third term is $o_p(1)$.

Next, we consider the fourth term.

$$\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)(\eta_i - \mu) = \frac{1}{N} \sum_{i=1}^N \bar{w}_i(\eta_i - \mu).$$

Its expectation is 0, and its variance is:

$$\text{var} \left\{ \frac{1}{N} \sum_{i=1}^N \bar{w}_i(\eta_i - \mu) \right\} = \frac{1}{N} E\{(\bar{w}_i)^2\} E\{(\eta_i - \mu)^2\} = \frac{1}{NT} V_T \sigma_\eta^2 = O\left(\frac{1}{NT}\right).$$

By the Chebyshev inequality, we have

$$\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)(\eta_i - \mu) = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Thus, the fourth term is $o_p(1)$.

Therefore, it follows that $\hat{\sigma}_\eta^2 \rightarrow_p \sigma_\eta^2$.

Next, we examine the asymptotic distribution of $\hat{\sigma}_\eta^2$. We have the following decomposition:

$$\begin{aligned} \sqrt{N} \left(\hat{\sigma}_\eta^2 - \sigma_\eta^2 - \frac{1}{T} V_T \right) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \{(\eta_i - \mu)^2 - \sigma_\eta^2\} + \sqrt{N} \left\{ \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 - \frac{1}{T} V_T \right\} \\ &\quad - \sqrt{N} (\bar{y} - \mu)^2 + \frac{2}{\sqrt{N}} \sum_{i=1}^N (\bar{y}_i - \eta_i)(\eta_i - \mu). \end{aligned}$$

The first term converges to $N(0, [E\{(\eta_i - \mu)^4\} - \sigma_\eta^4])$ by the Lindberg–Levy central limit theorem noting that $\{(\eta_i - \mu)^2\}_{i=1}^N$ is an i.i.d. sequence. The other terms are $o_p(1)$ by Lemma 2 and the arguments given above.

The proof of part 2 is omitted as it is trivial given part 1. ■

A.7 Proof of Theorem 8

Proof. Part 1). Observing that:

$$\hat{y}_{it} = z_{it} - x'_{it} \hat{\beta} = x'_{it} (\beta - \hat{\beta}) + \eta_i + w_{it},$$

we have

$$\begin{aligned} \hat{\gamma}_k^* - \hat{\gamma}_k &= \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' (\beta - \hat{\beta}) \\ &\quad + \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{i,t-k} - \bar{w}_i)(x_{it} - \bar{x}_i)' (\beta - \hat{\beta}) \\ &\quad + (\beta - \hat{\beta})' \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (x_{it} - \bar{x}_i)(x_{i,t-k} - \bar{x}_i)' (\beta - \hat{\beta}). \end{aligned}$$

We consider the first term on the right-hand side. We observe

$$\begin{aligned}
& E \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \right\} \\
&= E[w_{it}\{x_{i,t-k} - E_i(x_{it})'\}] - \frac{1}{T(T-k)} \sum_{t_1=k+1}^T \sum_{t_2=1}^T E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}] \\
&\quad - \frac{1}{T(T-k)} \sum_{t_1=1}^T \sum_{t_2=k+1}^T E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}] + \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_2=1}^T E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}].
\end{aligned}$$

By Assumption 4.3, the second term on the right-hand side is

$$\begin{aligned}
& \left\| \frac{1}{T(T-k)} \sum_{t_1=k+1}^T \sum_{t_2=1}^T E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}] \right\| \\
&\leq \frac{1}{T(T-k)} \sum_{t_1=k+1}^T \sum_{t_2=1}^T \|E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}]\| = o(1).
\end{aligned}$$

Similarly, the third and fourth terms are $o(1)$ such that

$$E \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \right\} = E[w_{it_1}\{x_{it_2} - E_i(x_{it})'\}] + o(1).$$

Next, we consider the variance:

$$var \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \right\} = \frac{1}{N} var \left\{ \frac{1}{T-k} \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \right\}.$$

We have the the following decomposition:

$$\begin{aligned}
& \frac{1}{T-k} \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \\
&= \frac{1}{T-k} \sum_{t=k+1}^T w_{it}\{x_{i,t-k} - E_i(x_{it})'\} + \frac{1}{T-k} \sum_{t=k+1}^T w_{it}\{\bar{x}_i - E_i(x_{it})'\} \\
&\quad + \frac{1}{T-k} \sum_{t=k+1}^T \bar{w}_i\{x_{i,t-k} - E_i(x_{it})'\} + \bar{w}_i\{\bar{x}_i - E_i(x_{it})'\}.
\end{aligned}$$

Let u_{ait} be the a -th element of the vector $\{x_{i,t-k} - E_i(x_{it})'\}$. We have the following:

$$\begin{aligned}
var \left(\frac{1}{T-k} \sum_{t=k+1}^T w_{it}u_{ait} \right) &= \frac{1}{(T-k)^2} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T E(w_{it_1}w_{it_2})E(u_{ait_1}u_{ait_2}) \\
&\quad + \frac{1}{(T-k)^2} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T E(w_{it_1}u_{ait_2})E(w_{it_2}u_{ait_1}) \\
&\quad + \frac{1}{(T-k)^2} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T cum_{1,1,a,a}(t_1, t_2, t_1-k, t_1-k) \\
&= O(1)
\end{aligned}$$

by Assumption 4.3-4. Similarly, the variances of $\sum_{t=k+1}^T w_{it}\{\bar{x}_i - E_i(x_{it})'\}/(T-k)$, $\sum_{t=k+1}^T \bar{w}_i\{x_{i,t-k} - E_i(x_{it})'\}/(T-k)$ and $\bar{w}_i\{\bar{x}_i - E_i(x_{it})'\}$ are also $O(1)$. It therefore follows that

$$var \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)' \right\} = O\left(\frac{1}{N}\right).$$

Summing up, we have by the Chebyshev inequality

$$\frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} - \bar{w}_i)(x_{i,t-k} - \bar{x}_i)'(\beta - \hat{\beta}) = E[w_{it}\{x_{i,t-k} - E_i(x_{it})\}'](\beta - \hat{\beta}) + o_p\left(\frac{1}{\sqrt{NT}}\right).$$

Similarly, we have

$$\frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{i,t-k} - \bar{w}_i)(x_{it} - \bar{x}_i)'(\beta - \hat{\beta}) = E[w_{i,t-k}\{x_{it} - E_i(x_{it})\}'](\beta - \hat{\beta}) + o_p\left(\frac{1}{\sqrt{NT}}\right),$$

and

$$(\beta - \hat{\beta})' \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (x_{it} - \bar{x}_i)(x_{i,t-k} - \bar{x}_i)'(\beta - \hat{\beta}) = O_p\left(\frac{1}{NT}\right).$$

Therefore, we have

$$\sqrt{NT}(\hat{\gamma}_k^* - \hat{\gamma}_k) = (E[w_{it}\{x'_{i,t-k} - E_i(x_{it})\}] + E[w_{i,t-k}\{x'_{it} - E_i(x_{it})\}])\sqrt{NT}(\hat{\beta} - \beta) + o_p(1).$$

Part 2). We have the following decomposition:

$$\begin{aligned} \tilde{V}_T^* - \tilde{V}_T &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right)(\hat{\gamma}_j^{*+} - \hat{\gamma}_j^+) \\ &= V^{wx}(\beta - \hat{\beta}) + V^{xw}(\beta - \hat{\beta}) + (\beta - \hat{\beta})'V^{xx}(\beta - \hat{\beta}), \end{aligned}$$

where

$$\begin{aligned} V^{wx} &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{1}{NT} \sum_{i=1}^N \sum_{t=j+1}^T (w_{it} - \bar{w}_i)(x_{i,t-j} - \bar{x}_i)', \\ V^{xw} &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{1}{NT} \sum_{i=1}^N \sum_{t=j+1}^T (w_{i,t-j} - \bar{w}_i)(x_{it} - \bar{x}_i)', \\ V^{xx} &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{1}{NT} \sum_{i=1}^N \sum_{t=j+1}^T (x_{it} - \bar{x}_i)(x_{i,t-j} - \bar{x}_i)'. \end{aligned}$$

We will show that $V^{wx} = O_p(1)$, $V^{xw} = O_p(1)$ and $V^{xx} = O_p(1)$. We use the results of Andrews (1991) to prove these equalities. We note that

$$V^{wx} = \frac{1}{N} \sum_{i=1}^N V^{wx,i},$$

where

$$V_i^{wx} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S}\right) \frac{1}{T} \sum_{t=j+1}^T (w_{it} - \bar{w}_i)(x_{i,t-j} - \bar{x}_i)'.$$

We use Theorem 1 of Andrews (1991) to show that $V_i^{wx} \rightarrow_p \sum_{j=-\infty}^{\infty} E[w_{it}\{x_{i,t-j} - E_i(x_{it})\}]$ as $T \rightarrow \infty$. For this purpose, we show that Assumptions A and B of Andrews (1991) are satisfied in our case. Note that $\theta_0, \hat{\theta}, V'_t$ and $(\partial/\partial\theta')V_t(\theta)$ in Assumptions A and B of Andrews (1991) are $(0, E_i(x'_{it}))$, (\bar{w}_i, \bar{x}'_i) , $(w_{it}, x'_{it} - E_i(x_{it}))'$ and $(-1, -1, \dots, -1)$, respectively, in our case. Therefore, the requirements for Theorem 1 of Andrews (1991) are satisfied in our case by Assumptions 1 and 4. Now, we have shown that $V_i^{wx} \rightarrow_p \sum_{j=-\infty}^{\infty} E[w_{it}\{x_{i,t-j} - E_i(x_{it})\}]$ as $T \rightarrow \infty$. Since the mean and variance of V_i^{wx} exist and are bounded uniformly over T , $\|V_i^{wx}\|$ is uniformly integrable. Note also that V_i^{wx} is i.i.d. across individuals. Therefore, Corollary 1 of Phillips and Moon (1999) implies that $V^{wx} = O_p(1)$ as $N, T \rightarrow \infty$. We can show that $V^{xw} = O_p(1)$ and $V^{xx} = O_p(1)$ in the exactly same manner.

Therefore, given that $(\beta - \hat{\beta}) = O_p(1/\sqrt{NT})$, we have $\tilde{V}_T^* - \tilde{V}_T = O_p(1/\sqrt{NT})$. ■

A.8 Lemmas

Lemma 1. Suppose that Assumption 1 is satisfied. Then, for any k , as $N \rightarrow \infty$ and $T \rightarrow \infty$,

$$\frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T w_{it} w_{i,t-k} \rightarrow_p \gamma_k.$$

Suppose that Assumptions 1 and 2 are satisfied. Then, for any k , as $N \rightarrow \infty$ and $T \rightarrow \infty$,

$$\sqrt{NT} \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) \rightarrow_d N \left(0, \sum_{j=-\infty}^{\infty} \{ \gamma_j^2 + \gamma_{k+j} \gamma_{k-j} + \text{cum}(0, -k, j, j-k) \} \right).$$

Proof. We observe that

$$E \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T w_{it} w_{i,t-k} \right\} = E(w_{it} w_{i,t-k}) = \gamma_k.$$

The variance term is

$$\begin{aligned} \text{var} \left\{ \frac{1}{N(T-k)} \sum_{i=1}^N \sum_{t=k+1}^T w_{it} w_{i,t-k} \right\} &= \frac{1}{N(T-k)^2} E \left[\left\{ \sum_{t=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) \right\}^2 \right] \\ &< \frac{M}{N} \rightarrow 0. \end{aligned}$$

By the Chebyshev inequality, the consistency is proven.

We apply Phillips and Moon's (1999) Theorem 3 to show the asymptotic normality. It is easy to see that Conditions (i), (ii) and (iv) of Theorem 3 in Phillips and Moon (1999) are satisfied. Let $a_{it} = w_{it} w_{i,t-k} - \gamma_k$. The sufficient condition for Condition (iii) is

$$E \left[\left\{ \sqrt{T} \frac{1}{T-k} \sum_{t=k+1}^T (w_{it} w_{i,t-k} - \gamma_k) \right\}^4 \right] = \frac{T^2}{(T-k)^4} E \left[\left(\sum_{t=k+1}^T a_{it} \right)^4 \right] < \infty.$$

We note that $E(a_{it}) = 0$. Now,

$$\begin{aligned} E \left\{ \left(\sum_{t=k+1}^T a_{it} \right)^4 \right\} &= \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T \sum_{t_3=k+1}^T \sum_{t_4=k+1}^T E(a_{it_1} a_{it_2} a_{it_3} a_{it_4}) \\ &= 3 \left\{ \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T E(a_{it_1} a_{it_2}) \right\}^2 + \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T \sum_{t_3=k+1}^T \sum_{t_4=k+1}^T \text{cum}_a(t_1, t_2, t_3, t_4), \end{aligned}$$

where $\text{cum}_a(t_1, t_2, t_3, t_4)$ is the fourth-order cumulant of $(a_{it_1}, a_{it_2}, a_{it_3}, a_{it_4})$. We have

$$\begin{aligned} \frac{1}{T} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T E(a_{it_1} a_{it_2}) &= \frac{1}{T} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T \{ \gamma_{t_1-t_2}^2 + \gamma_{t_1-t_2+k} \gamma_{t_1-t_2-k} + \text{cum}(t_1, t_1-k, t_2, t_2-k) \} \\ &\rightarrow \sum_{j=-\infty}^{\infty} \{ \gamma_j^2 + \gamma_{k+j} \gamma_{k-j} + \text{cum}(0, -k, j, j-k) \} \end{aligned}$$

by Assumptions 1.2 and 2. Theorem 2.3.2 of Brillinger (1981) states that $\text{cum}_a(t_1, t_2, t_3, t_4)$ can be written as a sum of products of at most eighth-order cumulants of w_{it} . See Brillinger (1981, pp20-21) for how it can be written. This implies that

$$\left| \sum_{t_2=k+1}^T \sum_{t_3=k+1}^T \sum_{t_4=k+1}^T \text{cum}_a(t_1, t_2, t_3, t_4) \right| \leq \sum_{t_2=-\infty}^{\infty} \sum_{t_3=-\infty}^{\infty} \sum_{t_4=-\infty}^{\infty} |\text{cum}_a(0, t_2, t_3, t_4)| < \infty$$

by Assumption 2. Therefore, we have $E\{(\sum_{t=k+1}^T a_{it})^4\}/T^2 = O(1)$. Thus, the asymptotic normality is shown.

Now, using the result for $E(a_{it_1} a_{it_2})$, the asymptotic variance of $\hat{\gamma}_k$ is

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t_1=k+1}^T \sum_{t_2=k+1}^T \{ E(w_{it_1} w_{i,t_1-k} w_{it_2} w_{i,t_2-k}) - \gamma_k^2 \} = \sum_{j=-\infty}^{\infty} \{ \gamma_j^2 + \gamma_{k+j} \gamma_{k-j} + \text{cum}(0, -k, j, j-k) \}.$$

■

Lemma 2. Suppose that Assumption 1 is satisfied. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 \rightarrow_p 0.$$

Suppose that Assumption 1 is satisfied. Suppose also that Assumption 2 is satisfied. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$,

$$\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 - \frac{1}{T} V_T = O_p \left(\frac{1}{\sqrt{NT}} \right).$$

Proof. First, we note that $\sum_{i=1}^N (\bar{y}_i - \eta_i)^2 / N \geq 0$. Its expectation is

$$E \left\{ \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 \right\} = E \{ (\bar{w}_i)^2 \} = \frac{1}{T} V_T \leq \frac{1}{T} \sum_{j=-\infty}^{\infty} |\gamma_j| \rightarrow 0.$$

By the Markov inequality, $\sum_{i=1}^N (\bar{y}_i - \eta_i)^2 / N = o_p(1)$. Now, the variance is

$$\begin{aligned} \text{var} \left\{ \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - \eta_i)^2 \right\} &= \frac{1}{N} E \left[\left\{ (\bar{w}_i)^2 - \frac{1}{T} V_T \right\}^2 \right] = \frac{1}{N} \left[E \{ (\bar{w}_i)^4 \} - \frac{1}{T^2} V_T^2 \right] \\ &= \frac{1}{N} \left[\frac{1}{T^4} E \left\{ \left(\sum_{t=1}^T w_{it} \right)^4 \right\} - \frac{1}{T^2} V_T^2 \right] = O \left(\frac{1}{NT^2} \right) \end{aligned}$$

by Assumption 2. By the Chebyshev inequality, we obtain the desired result. ■

Lemma 3. Suppose that Assumptions 1 and 2 are satisfied. Then, as $N \rightarrow \infty$ and $T \rightarrow \infty$,

$$E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i \right) = O \left(\frac{k}{T^2} \right) \text{ and } \text{var} \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i \right) = O \left(\frac{k}{NT^3} \right).$$

Proof. The mean is

$$\left| E \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i \right) \right| = \left| \frac{1}{T^2} \left[k \sum_{j=1}^{T-k} \gamma_j + \sum_{j=T-k}^T \{k - (T - j)\} \gamma_j \right] \right| \leq \frac{k}{T^2} \sum_{j=1}^T |\gamma_j| = O \left(\frac{k}{T^2} \right).$$

The variance is

$$\begin{aligned} &\text{var} \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^k w_{it} \bar{w}_i \right) \\ &= \frac{1}{N} \text{var} \left\{ \left(\frac{1}{T} \sum_{t=1}^k w_{it} \right) \left(\frac{1}{T} \sum_{t=1}^T w_{it} \right) \right\} \\ &= \frac{1}{NT^4} E \left[\left\{ \sum_{t_1=1}^k \sum_{t_2=1}^T \sum_{t_3=1}^k \sum_{t_4=1}^T (w_{it_1} w_{it_2} - \gamma_{|t_1-t_2|}) (w_{it_3} w_{it_4} - \gamma_{|t_3-t_4|}) \right\}^2 \right] \\ &= \frac{1}{NT^4} \sum_{t_1=1}^k \sum_{t_3=1}^k \gamma_{|t_1-t_3|} \sum_{t_2=1}^T \sum_{t_4=1}^T \gamma_{|t_2-t_4|} + \frac{1}{NT^4} \sum_{t_1=1}^k \sum_{t_4=1}^T \gamma_{|t_1-t_4|} \sum_{t_2=1}^T \sum_{t_3=1}^k \gamma_{|t_2-t_3|} \\ &\quad + \frac{1}{NT^4} \sum_{t_1=1}^k \sum_{t_2=1}^T \sum_{t_3=1}^k \sum_{t_4=1}^T \text{cum}(t_1, t_2, t_3, t_4) \\ &= O \left(\frac{k}{NT^3} \right). \end{aligned}$$

The last equality follows by Assumptions 1.2 and 2. ■

Table 1: Monte Carlo results: $N = 20$, the QS kernel.

T	α	k	TRUE	Tbias	WG		BWG		BWG2		IB	
					bias	std	bias	std	bias	std	bias	std
5	0	0	1	-0.2	-0.2	0.13	-0.064	0.15	-0.028	0.15	-0.0037	0.16
		1	0	-0.2	-0.2	0.088	-0.061	0.097	-0.026	0.1	-0.0012	0.11
		2	0	-0.2	-0.2	0.11	-0.063	0.094	-0.027	0.09	-0.003	0.09
10	0	0	1	-0.1	-0.1	0.095	-0.018	0.1	-0.0054	0.11	-0.0018	0.11
		1	0	-0.1	-0.1	0.067	-0.017	0.073	-0.0036	0.076	0.00	0.077
		2	0	-0.1	-0.1	0.07	-0.017	0.068	-0.0043	0.068	-0.00079	0.069
25	0	0	1	-0.04	-0.04	0.061	-0.0031	0.064	-0.00038	0.064	-0.00012	0.064
		1	0	-0.04	-0.04	0.044	-0.0026	0.046	0.00011	0.046	0.00036	0.046
		2	0	-0.04	-0.04	0.045	-0.0031	0.045	-0.00033	0.045	0.00	0.045
50	0	0	1	-0.02	-0.02	0.044	-0.00092	0.044	-0.00017	0.045	-0.00014	0.045
		1	0	-0.02	-0.019	0.032	0.0002	0.033	0.00095	0.033	0.00098	0.033
		2	0	-0.02	-0.02	0.032	-0.00054	0.032	0.00021	0.032	0.00024	0.032
5	0.5	0	1	-0.44	-0.45	0.097	-0.35	0.11	-0.32	0.12	-0.21	0.17
		1	0.5	-0.44	-0.47	0.057	-0.38	0.058	-0.34	0.063	-0.23	0.14
		2	0.25	-0.44	-0.46	0.066	-0.37	0.064	-0.34	0.065	-0.22	0.093
10	0.5	0	1	-0.26	-0.26	0.093	-0.17	0.11	-0.14	0.11	-0.078	0.14
		1	0.5	-0.26	-0.28	0.073	-0.18	0.087	-0.15	0.092	-0.09	0.12
		2	0.25	-0.26	-0.28	0.056	-0.18	0.067	-0.15	0.073	-0.094	0.099
25	0.5	0	1	-0.11	-0.11	0.072	-0.045	0.081	-0.025	0.085	-0.014	0.088
		1	0.5	-0.11	-0.12	0.062	-0.048	0.071	-0.028	0.076	-0.016	0.08
		2	0.25	-0.11	-0.12	0.053	-0.049	0.063	-0.029	0.067	-0.018	0.071
50	0.5	0	1	-0.058	-0.058	0.054	-0.015	0.059	-0.0059	0.06	-0.0034	0.061
		1	0.5	-0.058	-0.058	0.048	-0.015	0.053	-0.0063	0.054	-0.0038	0.055
		2	0.25	-0.058	-0.059	0.042	-0.016	0.047	-0.007	0.049	-0.0046	0.049
5	0.9	0	1	-0.8	-0.85	0.031	-0.83	0.033	-0.82	0.036	-0.74	0.079
		1	0.9	-0.8	-0.87	0.017	-0.85	0.018	-0.84	0.021	-0.75	0.067
		2	0.81	-0.8	-0.86	0.014	-0.84	0.014	-0.83	0.015	-0.75	0.049
10	0.9	0	1	-0.72	-0.73	0.049	-0.7	0.049	-0.69	0.049	-0.58	0.11
		1	0.9	-0.72	-0.75	0.042	-0.72	0.041	-0.71	0.041	-0.59	0.1
		2	0.81	-0.72	-0.75	0.028	-0.73	0.028	-0.72	0.028	-0.6	0.09
25	0.9	0	1	-0.49	-0.49	0.078	-0.44	0.082	-0.42	0.083	-0.29	0.14
		1	0.9	-0.49	-0.5	0.075	-0.45	0.079	-0.43	0.08	-0.3	0.14
		2	0.81	-0.49	-0.51	0.07	-0.45	0.073	-0.44	0.074	-0.31	0.13
50	0.9	0	1	-0.31	-0.31	0.086	-0.22	0.098	-0.2	0.1	-0.12	0.14
		1	0.9	-0.31	-0.31	0.085	-0.23	0.097	-0.2	0.1	-0.12	0.13
		2	0.81	-0.31	-0.32	0.082	-0.23	0.095	-0.2	0.098	-0.12	0.13

Note: This table presents the performances of various estimators with the quadratic spectrum kernel for autocovariances when $N = 20$. T refers to the length of time series. α is the AR(1) coefficient in the data-generating process. k is the order of autocovariance. "TRUE" is the true value of autocovariance. "Tbias" is the theoretical value of the bias. "bias" and "std" refer to bias and standard deviation. "WG", "BWG", "BWG2" and "IB" refer to the within-group estimator, the one-time bias-corrected autocovariance estimator, the two-time bias-corrected autocovariance estimator and the autocovariance estimator obtained after infinite iterations, respectively.

Table 2: Monte Carlo results: $N = 20$, the truncated kernel.

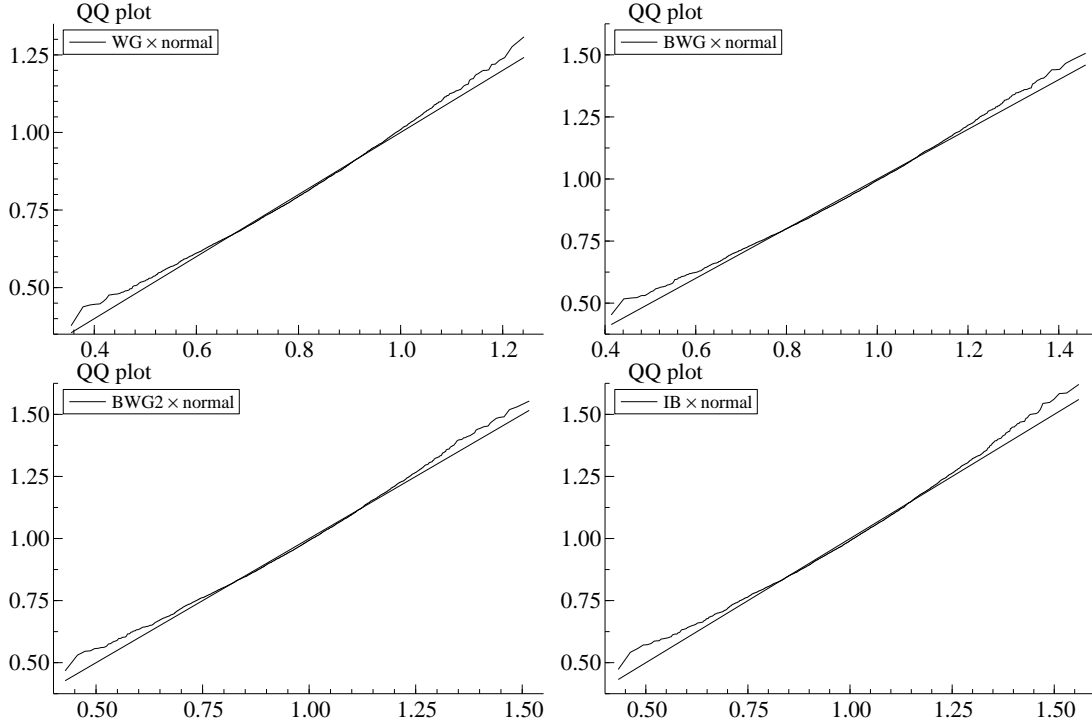
T	α	k	TRUE	Tbias	WG		BWG		BWG2		IB	
					bias	std	bias	std	bias	std	bias	std
5	0	0	1	-0.2	-0.2	0.13	-0.043	0.15	-0.022	0.16	-0.011	0.16
		1	0	-0.2	-0.2	0.087	-0.04	0.083	-0.019	0.1	-0.0088	0.12
		2	0	-0.2	-0.2	0.11	-0.04	0.1	-0.018	0.092	-0.0081	0.094
10	0	0	1	-0.1	-0.098	0.095	-0.0075	0.1	-0.00038	0.11	0.00087	0.11
		1	0	-0.1	-0.1	0.067	-0.0092	0.067	-0.0021	0.077	-0.00083	0.08
		2	0	-0.1	-0.1	0.072	-0.011	0.071	-0.0038	0.069	-0.0025	0.069
25	0	0	1	-0.04	-0.04	0.061	-0.0012	0.063	0.00	0.063	0.00	0.063
		1	0	-0.04	-0.04	0.044	-0.0017	0.045	-0.00054	0.047	-0.00048	0.047
		2	0	-0.04	-0.04	0.045	-0.0017	0.045	-0.00054	0.045	-0.00048	0.045
50	0	0	1	-0.02	-0.02	0.044	0.00	0.045	0.00026	0.045	0.00027	0.045
		1	0	-0.02	-0.02	0.031	-0.00067	0.031	-0.00043	0.032	-0.00042	0.032
		2	0	-0.02	-0.02	0.032	-0.00018	0.032	0.00	0.032	0.00	0.032
5	0.5	0	1	-0.44	-0.45	0.095	-0.34	0.11	-0.29	0.12	-0.17	0.19
		1	0.5	-0.44	-0.47	0.056	-0.36	0.067	-0.31	0.075	-0.2	0.16
		2	0.25	-0.44	-0.46	0.066	-0.35	0.057	-0.3	0.058	-0.19	0.11
10	0.5	0	1	-0.26	-0.26	0.095	-0.14	0.11	-0.099	0.13	-0.052	0.15
		1	0.5	-0.26	-0.27	0.074	-0.16	0.095	-0.11	0.11	-0.066	0.13
		2	0.25	-0.26	-0.28	0.057	-0.16	0.068	-0.12	0.082	-0.071	0.11
25	0.5	0	1	-0.11	-0.11	0.072	-0.039	0.083	-0.019	0.087	-0.011	0.089
		1	0.5	-0.11	-0.12	0.062	-0.042	0.073	-0.022	0.078	-0.015	0.08
		2	0.25	-0.11	-0.12	0.052	-0.043	0.063	-0.024	0.068	-0.016	0.071
50	0.5	0	1	-0.058	-0.059	0.054	-0.013	0.059	-0.0051	0.06	-0.0036	0.06
		1	0.5	-0.058	-0.06	0.047	-0.014	0.052	-0.0061	0.053	-0.0046	0.054
		2	0.25	-0.058	-0.06	0.041	-0.014	0.046	-0.0062	0.048	-0.0047	0.048
5	0.9	0	1	-0.8	-0.85	0.031	-0.82	0.037	-0.81	0.038	-0.73	0.078
		1	0.9	-0.8	-0.87	0.018	-0.84	0.023	-0.83	0.024	-0.74	0.066
		2	0.81	-0.8	-0.86	0.015	-0.83	0.012	-0.82	0.013	-0.74	0.052
10	0.9	0	1	-0.72	-0.73	0.05	-0.69	0.052	-0.68	0.05	-0.54	0.13
		1	0.9	-0.72	-0.74	0.043	-0.71	0.043	-0.7	0.041	-0.55	0.12
		2	0.81	-0.72	-0.75	0.029	-0.72	0.029	-0.71	0.028	-0.56	0.11
25	0.9	0	1	-0.49	-0.49	0.078	-0.42	0.076	-0.39	0.078	-0.25	0.16
		1	0.9	-0.49	-0.5	0.074	-0.43	0.072	-0.4	0.075	-0.26	0.16
		2	0.81	-0.49	-0.51	0.068	-0.44	0.066	-0.41	0.069	-0.27	0.15
50	0.9	0	1	-0.31	-0.31	0.084	-0.2	0.1	-0.16	0.11	-0.092	0.15
		1	0.9	-0.31	-0.31	0.083	-0.21	0.1	-0.16	0.11	-0.097	0.15
		2	0.81	-0.31	-0.32	0.081	-0.21	0.1	-0.17	0.11	-0.1	0.14

Note: This table presents the performances of various estimators with the truncated kernel for autocovariances when $N = 20$. T refers to the length of time series. α is the AR(1) coefficient in the data-generating process. k is the order of autocovariance. "TRUE" is the true value of autocovariance. "Tbias" is the theoretical value of the bias. "bias" and "std" refer to bias and standard deviation. "WG", "BWG", "BWG2" and "IB" refer to the within-group estimator, the one-time bias-corrected autocovariance estimator, the two-time bias-corrected autocovariance estimator and the autocovariance estimator obtained after infinite iterations, respectively.

Table 3: Monte Carlo results: The performances of the GLS estimators.

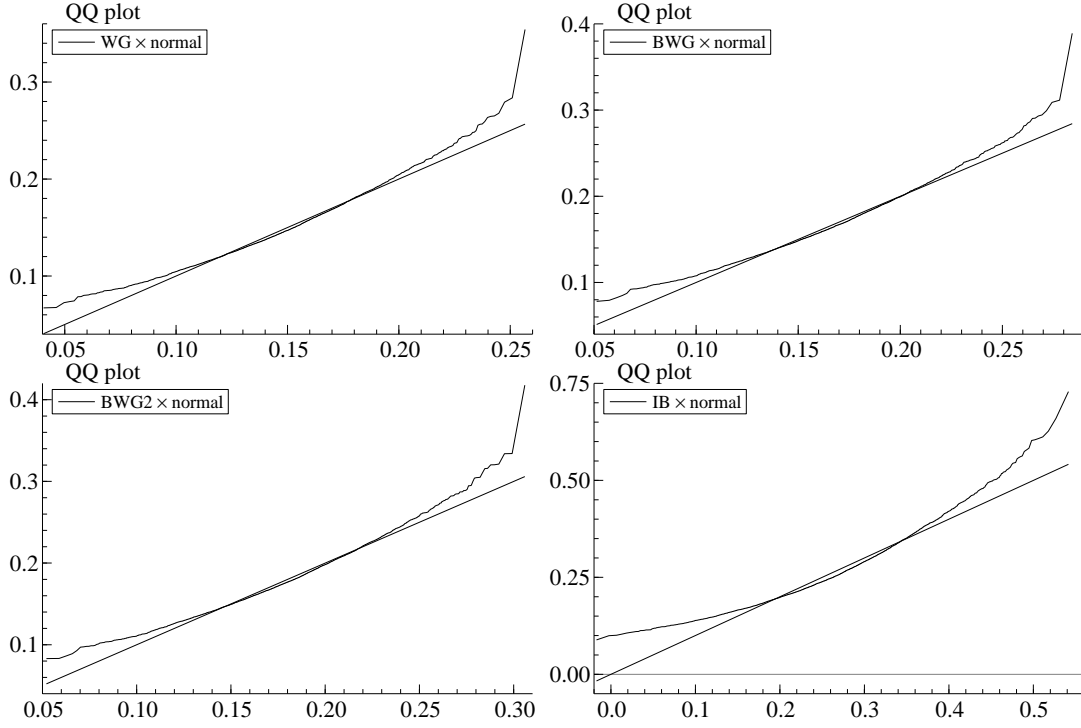
T	α		WG	GLS	FGLS	KGLS
5	0	bias	0.00063	0.00063	0.00082	0.0012
		std	0.12	0.12	0.12	0.13
		meanse	0.12	0.12	0.12	0.12
		cov. rate	0.95	0.95	0.95	0.93
10	0	bias	-0.00048	-0.00048	-0.00039	0.00
		std	0.081	0.081	0.081	0.086
		meanse	0.082	0.082	0.081	0.073
		cov. rate	0.95	0.95	0.95	0.9
20	0	bias	-0.00022	-0.00022	-0.00024	-0.00023
		std	0.056	0.056	0.056	0.063
		meanse	0.056	0.056	0.056	0.044
		cov. rate	0.95	0.95	0.95	0.83
5	0.5	bias	0.00011	0.00025	0.00032	0.001
		std	0.1	0.092	0.093	0.095
		meanse	0.1	0.092	0.098	0.088
		cov. rate	0.95	0.96	0.96	0.93
10	0.5	bias	-0.0013	-0.00038	-0.00053	-0.00068
		std	0.074	0.061	0.062	0.066
		meanse	0.074	0.062	0.065	0.056
		cov. rate	0.95	0.95	0.96	0.89
20	0.5	bias	-0.00036	-0.00011	-0.00013	-0.00013
		std	0.053	0.043	0.043	0.05
		meanse	0.053	0.043	0.044	0.034
		cov. rate	0.95	0.95	0.95	0.82
5	0.9	bias	0.00052	0.00023	0.00032	0.00063
		std	0.053	0.039	0.041	0.041
		meanse	0.053	0.04	0.049	0.038
		cov. rate	0.95	0.96	0.98	0.93
10	0.9	bias	-0.00048	0.00	-0.00014	-0.00028
		std	0.045	0.026	0.027	0.029
		meanse	0.045	0.026	0.033	0.024
		cov. rate	0.95	0.95	0.98	0.88
20	0.9	bias	-0.00016	0.00	0.00	0.00
		std	0.039	0.018	0.018	0.024
		meanse	0.038	0.018	0.022	0.014
		cov. rate	0.95	0.95	0.98	0.75

Note: This table presents the performances of various estimators for a fixed effects regression model when $N = 50$. T refers to the length of time series. α is the AR(1) coefficient in the process of the disturbance term. “bias”, “std”, “meanse” and “cov. rate” refer to bias, standard deviation, mean of the standard error and coverage rate of the 95% confidence interval, respectively. “WG”, “GLS”, “FGLS” and “KGLS” refer to within-group estimator, infeasible GLS estimator, feasible GLS estimator proposed in this paper and the GLS estimator considered by Kiefer (1980), respectively.



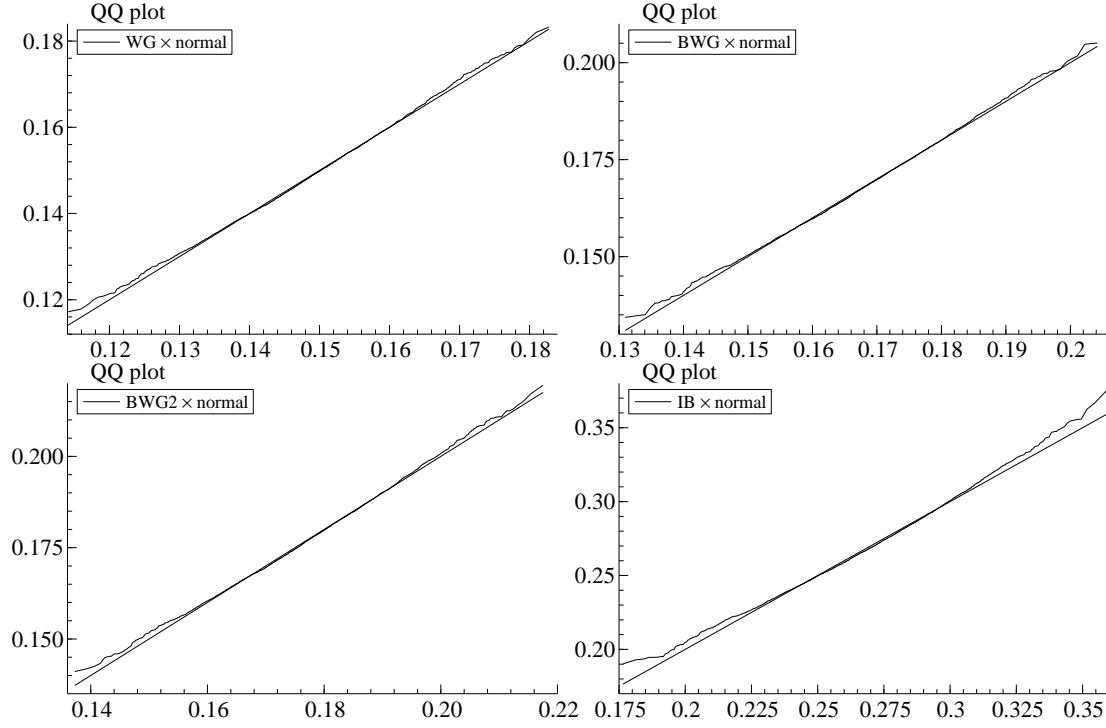
Note: This figure presents four QQ plots, each of which compares the distribution of an estimator for γ_0 (the variance) with the normal distribution with the same mean and the same variance as those of the estimator, when $N = 20$ (cross-sectional sample size), $T = 5$ (time series length) and $\alpha = 0$ (the AR(1) parameter). The true value of γ_0 is 1. “WG”, “BWG”, “BWG2” and “IB” refer to the within-group estimator, the one-time bias-corrected autocovariance estimator, the two-time bias-corrected autocovariance estimator and the autocovariance estimator obtained after infinite iterations, respectively. The QS kernel is used for all the estimators.

Figure 1: QQ plots, $N = 20$, $T = 5$, $\alpha = 0$



Note: This figure presents four QQ plots, each of which compares the distribution of an estimator for γ_0 (the variance) with the normal distribution with the same mean and the same variance as those of the estimator, when $N = 20$ (cross-sectional sample size), $T = 5$ (time series length) and $\alpha = 0.9$ (the AR(1) parameter). The true value of γ_0 is 1. “WG”, “BWG”, “BWG2” and “IB” refer to the within-group estimator, the one-time bias-corrected autocovariance estimator, the two-time bias-corrected autocovariance estimator and the autocovariance estimator obtained after infinite iterations, respectively. The QS kernel is used for all the estimators.

Figure 2: QQ plots. $N = 20$, $T = 5$, $\alpha = 0.9$



Note: This figure presents four QQ plots, each of which compares the distribution of an estimator for γ_0 (the variance) with the normal distribution with the same mean and the same variance as those of the estimator, when $N = 200$ (cross-sectional sample size), $T = 5$ (time series length) and $\alpha = 0.9$ (the AR(1) parameter). The true value of γ_0 is 1. “WG”, “BWG”, “BWG2” and “IB” refer to the within-group estimator, the one-time bias-corrected autocovariance estimator, the two-time bias-corrected autocovariance estimator and the autocovariance estimator obtained after infinite iterations, respectively. The QS kernel is used for all the estimators.

Figure 3: QQ plots, $N = 200$, $T = 5$, $\alpha = 0.9$