DEPARTMENT OF ECONOMICS

Working Paper

CONSISTENT COVARIANCE MATRIX
ESTIMATION FOR LINEAR PROCESSES

Michael Jansson

Working Paper No. 1999-29

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

INSTITUT FOR ØKONOMI

AFDELING FOR NATIONALØKONOMI - AARHUS UNIVERSITET - BYGNING 350
8000 AARHUS C - ☎ 89 42 11 33 - TELEFAX 86 13 63 34

WORKING PAPER

CONSISTENT COVARIANCE MATRIX ESTIMATION FOR LINEAR PROCESSES

Michael Jansson

Working Paper No. 1999-29

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Consistent Covariance Matrix Estimation for Linear Processes

Michael Jansson
Department of Economics and CDME, University of Aarhus
DK-8000 Aarhus C, Denmark
Email: mjansson@econ.au.dk

December 18, 1999

Abstract. This note establishes consistency of kernel estimators of the long-run covariance matrix of a linear process under weak moment and memory conditions. In addition, it is pointed out that some published consistency proofs are in error as they stand.

Keywords: Covariance Matrix Estimation, Kernel Estimator, Linear Process
JEL Classification: C13

1. Introduction

This note establishes consistency of kernel estimators of the long-run covariance matrix of a linear process under weak moment and memory conditions. The best such consistency results currently known to the author require substantially more restrictive moment and/or memory conditions than needed for the functional central limit theorem (FCLT).[1] In contrast, our conditions are only moderately stronger than those of the FCLT of Davidson (1999).[2]

2. Results

Consider a sequence of n-dimensional random vectors $\{V_t\}_{t\geq 1} = \left\{(V_{t1}, \ldots, V_{tn})'\right\}_{t\geq 1}$ generated by the linear process

$$V_t = C(L)\, e_t, \tag{1}$$

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$ is an $n \times n$ matrix polynomial in the lag operator. For all $t \in \mathbb{Z}$, let $\mathcal{F}_t = \sigma(e_s : s \leq t)$ and for any $m \times n$ matrix $A = (a_{ij})$ and any $p > 0$,

[1] Specifically, Robinson (1991, Theorem 2.1) requires at least 2.5 finite moments when the bandwidth expansion rates recommended by Andrews (1991) are employed, while de Jong and Davidson (1999, Theorem 2.1) require near epoch dependence of size -1/2. As discussed by Davidson (1999), the latter condition is excessively stringent for the FCLT. In particular, it is stronger than our condition ($\mathcal{V}1$) (i).

[2] Davidson (1999, Theorem 3) requires square summability rather than absolute summability of the MA coefficients.

let $\|A\|_p = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^p\right)^{1/p}$. We impose the following condition on $C(L)$ and $\{e_t\}_{t\in\mathbb{Z}} = \left\{(e_{t1},\ldots,e_{tn})'\right\}_{t\in\mathbb{Z}}$:

$$\begin{array}{ll} \text{(i)} & \sum_{i=0}^\infty \|C_i\|_2 < \infty, \\ \text{(ii)} & \text{For all } t\in\mathbb{Z}, E(e_t \mid \mathcal{F}_{t-1}) = 0 \text{ (a.s.) and } E(e_te_t' \mid \mathcal{F}_{t-1}) = I_n \text{ (a.s.)}, \\ \text{(iii)} & \{e_{ti}e_{tj}\}_{t\in\mathbb{Z}} \text{ is uniformly integrable for } 1\le i\le j\le n. \end{array} \qquad (\mathcal{V}1)$$

Notice that C_0 is not necessarily the identity matrix. Therefore, the assumption $E(e_te_t' \mid \mathcal{F}_{t-1}) = I_n$ simply restricts $E(V_tV_t' \mid \mathcal{F}_{t-1})$ to be constant. Under condition $(\mathcal{V}1)$ (ii), $(\mathcal{V}1)$ (iii) holds whenever $\{e_t\}_{t\in\mathbb{Z}}$ is *i.i.d.* or $\max_{1\le i\le n}\sup_{t\in\mathbb{Z}} E|e_{it}|^r < \infty$ for some $r > 2$.

When $(\mathcal{V}1)$ holds, the long-run covariance matrix of V_t,

$$\Omega = \lim_{T\to\infty} T^{-1}\sum_{s=1}^T\sum_{t=1}^T E(V_sV_t'),$$

can be written as follows:

$$\Omega = \Gamma + \Gamma' - \Sigma_0, \qquad (2)$$

where

$$\Gamma = \sum_{i=0}^\infty \Sigma_i, \qquad (3)$$

$$\Sigma_i = E\left(V_tV_{t-i}'\right), \qquad t\ge i+1, i\ge 0.$$

In some applications, such as the cointegration procedures of Phillips and Hansen (1990) and Park (1992), the one-sided long-run covariance matrix Γ in (3) is of interest in it's own right. In recognition of this fact, we focus explicitly on Γ. Of course, in view of (2) and the fact that Σ_0 is easy to estimate, a consistent estimator of Ω is readily constructed given a consistent estimator of Γ. We consider the class of kernel estimators of Γ given by

$$\hat{\Gamma}_T = \sum_{i=0}^{T-1} k\left(\frac{i}{s_T}\right)\hat{\Sigma}_{i,T},$$

where

$$\hat{\Sigma}_{i,T} = T^{-1} \sum_{t=i+1}^{T} V_t V_{t-i}', \qquad 0 \leq i \leq T-1,$$

and $k(\cdot)$ is a kernel. The corresponding estimator of Ω is

$$\hat{\Omega}_T = \hat{\Gamma}_T + \hat{\Gamma}_T' - \hat{\Sigma}_{0,T}.$$

The kernel $k(\cdot)$ and the sequence $\{s_T\}_{T\geq 1}$ of (positive) bandwidth parameters are assumed to satisfy the following conditions:

For all $x \in \mathbb{R}, |k(x)| \leq 1$ and $k(x) = k(-x); k(0) = 1; k(\cdot)$ is continuous at zero; $\int_{[0,\infty)}^{*} \bar{k}(x)\,dx < \infty$, where $\bar{k}(x) = \sup_{y\geq|x|} |k(y)|$ for all $x \geq 0$. $(\mathcal{K})$

$s_T \to \infty$ and $T^{-1/2} s_T \to 0$ as $T \to \infty$. $(\mathcal{S})$

In $(\mathcal{K})$, $\int_{[0,\infty)}^{*} \bar{k}(x)\,dx$ denotes the outer integral of $\bar{k}(\cdot)$ over $[0,\infty)$.[3] Condition $(\mathcal{K})$ resembles Assumption A2(0) of Robinson (1991) and is satisfied by most kernels considered in the literature.[4] In particular, $(\mathcal{K})$ holds for the truncated kernel and for all kernels in the class $\mathcal{K}_3$ of Andrews (1991) and Andrews and Monahan (1992). Likewise, $(\mathcal{S})$ is satisfied whenever the bandwidth expansion rate coincides with the optimal rate reported in Andrews (1991, p. 830). Our main result is the following.

Theorem 1. *Suppose* $(\mathcal{K}), (\mathcal{S})$ *and* $(\mathcal{V}1)$ *hold. Then* $\hat{\Gamma}_T - \Gamma \to_p 0$ *and* $\hat{\Omega}_T - \Omega \to_p 0$ *as* $T \to \infty$.

Although condition $(\mathcal{K})$ is satisfied by most kernels in actual use, some kernels in the class $\mathcal{K}_1$ of Andrews (1991) and Andrews and Monahan (1992) do not satisfy $(\mathcal{K})$. Similarly, condition $(\mathcal{K})$ can be violated under Hansen's (1992) Condition (K) and Assumption 1.1 of de Jong (1998). As explained in the appendix, however, some of the proofs in Andrews (1991), Hansen (1992) and de Jong (1998) are in error as

[3] We state the condition in terms of the outer integral in order to avoid measurability complications.

[4] Unlike Robinson (1991) and de Jong and Davidson (1999, Assumption 1), we do not require $\int_{\mathbb{R}} |K(\lambda)|\,d\lambda < \infty$, where $K(\lambda) = (2\pi)^{-1} \int_{\mathbb{R}} k(x) \exp(i\lambda x)\,dx$. This enables us to accommodate the truncated kernel.

they stand, precisely because a condition like $(\mathcal{K})$ is needed for a key step in these proofs to be valid.[5]

In applications, the vectors $\{V_t\}_{t\geq 1}$ are often functions of an unknown parameter vector θ (say), $V_t = V_t(\theta)$. Given an estimator $\hat{\theta}_T$ of θ_0 (the true value of θ), we can construct the estimators $\hat{\Omega}_T\left(\hat{\theta}_T\right)$ and $\hat{\Gamma}_T\left(\hat{\theta}_T\right)$, where

$$\hat{\Omega}_T\left(\hat{\theta}_T\right) = \hat{\Gamma}_T\left(\hat{\theta}_T\right) + \hat{\Gamma}_T\left(\hat{\theta}_T\right)' - \hat{\Sigma}_{0,T}\left(\hat{\theta}_T\right),$$

$$\hat{\Gamma}_T\left(\hat{\theta}_T\right) = \sum_{i=0}^{T-1} k\left(\frac{i}{s_T}\right)\hat{\Sigma}_{i,T}\left(\hat{\theta}_T\right),$$

$$\hat{\Sigma}_{i,T}\left(\hat{\theta}_T\right) = T^{-1}\sum_{t=i+1}^{T} V_t\left(\hat{\theta}_T\right)V_{t-i}\left(\hat{\theta}_T\right)', \qquad 0 \leq i \leq T-1.$$

To establish consistency of $\hat{\Omega}_T\left(\hat{\theta}_T\right)$ and $\hat{\Gamma}_T\left(\hat{\theta}_T\right)$, we impose the following condition:

Either

(i) $T^{1/2}\left(\hat{\theta}_T - \theta_0\right) = O_p(1)$ and for some neighborhood $\mathcal{N}$ of θ_0, $\sup_{t\geq 1} E\left(\sup_{\theta\in\mathcal{N}}\left(\left\|\frac{\partial}{\partial\theta'}V_t(\theta)\right\|_2\right)^2\right) < \infty$,

or $(\mathcal{V}2)$

(ii) $V_t(\theta) = V_t(\theta_0) - (\theta-\theta_0)' X_t, \sup_{1\leq t\leq T}\|\delta_T X_t\|_2 = O_p(1)$, and $T^{1/2}\left(\hat{\theta}_T - \theta_0\right)\delta_T^{-1} = O_p(1)$, where $\{\delta_T\}_{T\geq 1}$ is a sequence of nonsingular matrices.

Condition $(\mathcal{V}2)$ (ii) is Hansen's (1992) Condition (V3), while $(\mathcal{V}2)$ (i) is equivalent to Assumption B of Andrews (1991) under $(\mathcal{V}1)$. As in Hansen (1992), the following is an immediate consequence of Theorem 1.

Corollary 2. *Suppose* $(\mathcal{K}), (\mathcal{S}), (\mathcal{V}1)$ *and* $(\mathcal{V}2)$ *hold. Then* $\hat{\Gamma}_T\left(\hat{\theta}_T\right) - \Gamma \to_p 0$ *and* $\hat{\Omega}_T\left(\hat{\theta}_T\right) - \Omega \to_p 0$ *as* $T \to \infty$.

[5] Theorem 1 of Andrews and Monahan (1992) is true as stated, since the kernel k is assumed to belong to the class $\mathcal{K}_3$. On the other hand, the claim that the consistency results (Theorem 1) hold for all $k \in \mathcal{K}_1$ when the sequence of bandwidth parameters is fixed (Andrews and Monahan, 1992, p. 956) would appear to be incorrect.

Sample-dependent bandwidth parameters can also be accommodated. Let $\hat{\Omega}_T\left(\hat{\theta}_T, \hat{s}_T\right)$ and $\hat{\Gamma}_T\left(\hat{\theta}_T, \hat{s}_T\right)$ denote $\hat{\Omega}_T\left(\hat{\theta}_T\right)$ and $\hat{\Gamma}_T\left(\hat{\theta}_T\right)$ evaluated at the possibly stochastic bandwidth $\hat{s}_T$. The following assumption on $\{\hat{s}_T\}$ suffices:

$$\hat{s}_T = \hat{\alpha}_T s_T, \text{ where } \hat{\alpha}_T = O_p(1), 1/\hat{\alpha}_T = O_p(1), \text{ and } s_T \text{ satifies } (\mathcal{S}). \tag{$\mathcal{S}'$}$$

Theorem 3. *Suppose* $(\mathcal{K}), (\mathcal{S}'), (\mathcal{V}1)$ *and* $(\mathcal{V}2)$ *hold. Then* $\hat{\Gamma}_T\left(\hat{\theta}_T, \hat{s}_T\right) - \Gamma \to_p 0$ *and* $\hat{\Omega}_T\left(\hat{\theta}_T, \hat{s}_T\right) - \Omega \to_p 0$ *as* $T \to \infty$.

3. Proofs

3.1. Proof of Theorem 1.

Notice that

$$\left\|\hat{\Gamma}_T - \Gamma\right\|_2 \leq \left\|\hat{\Gamma}_T - E\left(\hat{\Gamma}_T\right)\right\|_2 + \left\|E\left(\hat{\Gamma}_T\right) - \Gamma\right\|_2,$$

$$\left\|\hat{\Omega}_T - \Omega\right\|_2 \leq \left\|\hat{\Omega}_T - E\left(\hat{\Omega}_T\right)\right\|_2 + \left\|E\left(\hat{\Omega}_T\right) - \Omega\right\|_2.$$

Continuity of $k(\cdot)$ at zero and $\sum_{i=0}^{\infty} \|\Sigma_i\|_2 < \infty$ implies $\left\|E\left(\hat{\Gamma}_T\right) - \Gamma\right\|_2 \to 0$ and $\left\|E\left(\hat{\Omega}_T\right) - \Omega\right\|_2 \to 0$. Moreover,

$$\left\|\hat{\Omega}_T - E\left(\hat{\Omega}_T\right)\right\|_2 \leq 2 \cdot \left\|\hat{\Gamma}_T - E\left(\hat{\Gamma}_T\right)\right\|_2 + \left\|\hat{\Sigma}_{0,T} - E\left(\hat{\Sigma}_{0,T}\right)\right\|_2,$$

while

$$\left\|\hat{\Gamma}_T - E\left(\hat{\Gamma}_T\right)\right\|_2 \leq \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right| \cdot \left\|\hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right)\right\|_2.$$

Suppose we can show that

$$s_T^{-1} \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right| = O(1). \tag{4}$$

Moreover, suppose we can find non-negative sequences $\{\beta_i\}_{i\geq 0}$ and $\{\psi_T, \eta_T\}_{T\geq 1}$ such that

$$E\left(\left\|\hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right)\right\|_2\right) \leq \beta_i \psi_T + \eta_T, \qquad 0 \leq i \leq T-1, \tag{5}$$

where $\sum_{i=0}^{\infty} \beta_i < \infty, \psi_T = o(1)$ and $\eta_T = O\left(T^{-1/2}\right)$. Then

$$E\left(\left\|\hat{\Sigma}_{0,T} - E\left(\hat{\Sigma}_{0,T}\right)\right\|_2\right) \leq \beta_0 \psi_T + \eta_T \to 0,$$

and

$$\begin{aligned} E\left(\left\|\hat{\Gamma}_T - E\left(\hat{\Gamma}_T\right)\right\|_2\right) &\leq \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right| \cdot E\left(\left\|\hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right)\right\|_2\right) \\ &\leq \psi_T \sum_{i=0}^{T-1} \beta_i + \eta_T \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right| \\ &\leq \psi_T \left(\sum_{i=0}^{\infty} \beta_i\right) + (\eta_T s_T)\left(s_T^{-1} \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right|\right) \\ &\to 0, \end{aligned}$$

since $|k(x)| \leq 1$ for all $x \in \mathbb{R}$ and $\psi_T s_T = o(1)$ under $(\mathcal{K}), (\mathcal{S})$. The two lemmas that follow establish (4) and (5), hereby completing the proof. ■

Lemma 4. *Suppose* $(\mathcal{K})$ *and* $(\mathcal{S})$ *hold. Then* $\limsup_{T\to\infty} s_T^{-1} \sum_{i=0}^{T-1} |k(i/s_T)| < \infty$.

Remark 1. In the proof of Theorem 1(a) in Andrews (1991), it is claimed that the conclusion of Lemma 4 holds under the weaker condition that $k \in \mathcal{K}_1$, where[6]

$$\begin{aligned} \mathcal{K}_1 = \ & \{k(\cdot) : \mathbb{R} \to [-1, 1], k(0) = 1, k(x) = k(-x) \ \forall x \in \mathbb{R}, \\ & \textstyle\int_{\mathbb{R}} |k(x)|\, dx < \infty, \ k(\cdot) \text{ is continuous at 0 and at all but} \\ & \text{a finite number of other points}\}. \end{aligned}$$

A similar claim has been made by Hansen (1992, pp. 970-972). As we now show, these claims are invalid. Take any $k \in \mathcal{K}_1$ such that $k(x) = 1 \ \forall x \in \mathbb{Z}$ and take any $\{s_T\} \subseteq \mathbb{N}$ such that $(\mathcal{S})$ holds. Then, as $T \to \infty$,

$$\begin{aligned} s_T^{-1} \sum_{i=0}^{T-1} \left|k\left(\frac{i}{s_T}\right)\right| &\geq s_T^{-1} \sum_{x=0}^{[(T-1)/s_T]} |k(x)| \\ &= s_T^{-1} \left(\left[\frac{T-1}{s_T}\right] + 1\right) \\ &\to \infty, \end{aligned}$$

[6] As pointed out by Andrews and Monahan (1992, p. 955), the class $\mathcal{K}_1$ in Andrews (1991) should be defined with $\int_{\mathbb{R}} k(x)^2\, dx < \infty$ replaced by $\int_{\mathbb{R}} |k(x)|\, dx < \infty$.

where $[(T-1)/s_T]$ is the largest integer not exceeding $(T-1)/s_T$. As a consequence, the proofs of Theorem 1 of Andrews (1991) and Theorems 1 and 3 of Hansen (1992) are in error as they stand. Likewise,

$$\begin{aligned} s_T^{-2}\sum_{i=0}^{T-1}\left|k\left(\frac{i}{s_T}\right)\right|\cdot i &\geq s_T^{-1}\sum_{x=0}^{[(T-1)/s_T]}|k(x)|\cdot x \\ &= s_T^{-1}\cdot\frac{\left[\frac{T-1}{s_T}\right]\left(\left[\frac{T-1}{s_T}\right]+1\right)}{2} \\ &\to \infty, \end{aligned}$$

contradicting a claim made by de Jong (1998, Proof of Theorem 2). Therefore, de Jong's (1998) corrected proof of Hansen's (1992) incorrect consistency proof is in error as it stands.

One kernel $k\in\mathcal{K}_1$ such that $k(x)=1\ \forall x\in\mathbb{Z}$ is

$$k(\cdot)=\sum_{i=0}^{\infty}k_i(\cdot),$$

where, for each $i\geq 0$ and $x\geq 0$,

$$k_i(x)=\begin{cases} 1-(i+1)^2(x-i), & \text{if } i\leq x\leq i+(i+1)^{-2}, \\ 1-(i+1)^2(i+1-x), & \text{if } i+1-(i+1)^{-2}\leq x\leq i+1, \\ 0, & \text{otherwise.}\end{cases}$$

Letting $k(x)=k(-x)\ \forall x<0$, it easily seen that $k\in\mathcal{K}_1$. In particular, k is continuous and

$$\begin{aligned} \int_{\mathbb{R}}|k(x)|\,dx &= 2\int_{[0,\infty)}k(x)\,dx \\ &= 2\sum_{i=0}^{\infty}\left(\int_{[0,\infty)}k_i(x)\,dx\right) \\ &= 2\sum_{i=0}^{\infty}\frac{1}{(i+1)^2} \\ &= \frac{\pi^2}{3}<\infty. \end{aligned}$$

Remark 2. Any kernel $k \in \mathcal{K}_1$ satisfies Hansen's (1992) Condition (K), which is identical to Assumption 1.1 of de Jong (1998). In Remark 1, we therefore only considered the case where $k \in \mathcal{K}_1$. We notice, though, that matters can get even worse under Hansen's (1992) Condition (K), since that condition allows k to be discontinuous at countably many points. The sequence $\left\{ s_T^{-1} \sum_{i=0}^{T-1} |k(i/s_T)| \right\}_{T \geq 1}$ only depends on $k(\cdot)$ through $\{k(x) : x \in \mathcal{D}\}$, where $\mathcal{D} = \cup_{T \geq 1} \cup_{1 \leq i \leq T-1} \{i/s_T\}$. Since $\mathcal{D}$ is countable, $s_T^{-1} \sum_{i=0}^{T-1} |k(i/s_T)|$ can take on any value in $\left[s_T^{-1}; s_T^{-1} \cdot T \right]$ (for each T) and still satisfy Hansen's (1992) Condition (K). In particular, we can have $s_T^{-1} \sum_{i=0}^{T-1} |k(i/s_T)| = s_T^{-1} \cdot T$, which diverges (as $T \to \infty$) whenever $s_T = o(T)$.

Proof of Lemma 4. For any $1 \leq i \leq T - 1$, we have

$$\left| k\left(\frac{i}{s_T}\right) \right| \leq \bar{k}\left(\frac{i}{s_T}\right) \leq \bar{k}(x), \qquad \frac{i-1}{s_T} \leq x \leq \frac{i}{s_T},$$

where $\bar{k}(x) = \sup_{y \geq |x|} |k(y)|$ for all $x \geq 0$. Therefore,

$$s_T^{-1} \left| k\left(\frac{i}{s_T}\right) \right| = \int_{[(i-1)/s_T, i/s_T)} \left| k\left(\frac{i}{s_T}\right) \right| dx \leq \int^{*}_{[(i-1)/s_T, i/s_T)} \bar{k}(x)\, dx,$$

and hence

$$\begin{aligned} s_T^{-1} \sum_{i=0}^{T-1} \left| k\left(\frac{i}{s_T}\right) \right| &= s_T^{-1} + s_T^{-1} \sum_{i=1}^{T-1} \left| k\left(\frac{i}{s_T}\right) \right| \\ &\leq s_T^{-1} + \int^{*}_{[0,(T-1)/s_T)} \bar{k}(x)\, dx \\ &\leq s_T^{-1} + \int^{*}_{[0,\infty)} \bar{k}(x)\, dx. \end{aligned}$$

The lemma follows by taking the lim sup (as $T \to \infty$) on each side since $s_T^{-1} \to 0$ and $\int^{*}_{[0,\infty)} \bar{k}(x)\, dx < \infty$. ■

Lemma 5. *Suppose* $\{V_t\}$ *is generated by* (1) *and satisfies* $(\mathcal{V}1)$. *Then* (5) *holds.*

Proof of Lemma 5. We have:

$$\begin{aligned} V_t V_{t-i}' &= \left(\sum_{j=0}^{\infty} C_j e_{t-j}\right)\left(\sum_{k=0}^{\infty} C_k e_{t-i-k}\right)' \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} C_j e_{t-j} e_{t-i-k}' C_k' \\ &= \sum_{k=0}^{\infty} C_{i+k} e_{t-i-k} e_{t-i-k}' C_k' + \sum_{k=0}^{\infty} \sum_{\substack{j=0 \\ j\neq i+k}}^{\infty} C_j e_{t-j} e_{t-i-k}' C_k'. \end{aligned}$$

Clearly, $E\left(V_t V_{t-i}'\right) = \sum_{k=0}^{\infty} C_{i+k} C_k'$, so

$$\begin{aligned} \hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right) &= T^{-1} \sum_{t=i+1}^{T} \left(\sum_{k=0}^{\infty} C_{i+k}\left(e_{t-i-k} e_{t-i-k}' - I_n\right) C_k'\right) \\ &\quad + T^{-1} \sum_{t=i+1}^{T} \left(\sum_{k=0}^{\infty} \sum_{\substack{j=0 \\ j\neq i+k}}^{\infty} C_j e_{t-j} e_{t-i-k}' C_k'\right) \\ &= \sum_{k=0}^{\infty} \left(T^{-1} \sum_{t=i+1}^{T} \left(C_{i+k}\left(e_{t-i-k} e_{t-i-k}' - I_n\right) C_k'\right)\right) \\ &\quad + \sum_{k=0}^{\infty} \sum_{\substack{j=0 \\ j\neq i+k}}^{\infty} \left(T^{-1} \sum_{t=i+1}^{T} \left(C_j e_{t-j} e_{t-i-k}' C_k'\right)\right). \end{aligned}$$

As a consequence,

$$\begin{aligned} \left\|\hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right)\right\|_2 &\leq \sum_{k=0}^{\infty} \left\|T^{-1} \sum_{t=i+1}^{T} \left(e_{t-i-k} e_{t-i-k}' - I_n\right)\right\|_2 \cdot \|C_k\|_2 \cdot \|C_{i+k}\|_2 \\ &\quad + \sum_{k=0}^{\infty} \sum_{\substack{j=0 \\ j\neq i+k}}^{\infty} \left\|T^{-1} \sum_{t=i+1}^{T} e_{t-j} e_{t-i-k}'\right\|_2 \cdot \|C_k\|_2 \cdot \|C_j\|_2, \end{aligned}$$

since $\|AB\|_2 \leq \|A\|_2 \|B\|_2$ for conformable A and B. Therefore,

$$E\left(\left\|\hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T}\right)\right\|_2\right) \leq \beta_i \psi_T + \eta_T,$$

where

$$\beta_i = \sum_{k=0}^{\infty} \|C_k\|_2 \cdot \|C_{i+k}\|_2 \,,$$

$$\psi_T = \sup_{k\geq 0} \max_{0\leq i\leq T-1} E\left(\left\| T^{-1} \sum_{t=i+1}^{T} \left(e_{t-i-k}e'_{t-i-k} - I_n\right)\right\|_2\right),$$

$$\eta_T = \left(\max_{0\leq i\leq T-1} \sup_{\substack{j,k\geq 0 \\ j\neq i+k}} E\left(\left\| T^{-1} \sum_{t=i+1}^{T} e_{t-j}e'_{t-i-k}\right\|_2\right)\right)\left(\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \|C_k\|_2 \cdot \|C_j\|_2\right).$$

By $(\mathcal{V}1)$ (i),

$$\sum_{i=0}^{\infty} \beta_i = \sum_{i=0}^{\infty}\sum_{k=0}^{\infty} \|C_k\|_2 \cdot \|C_{i+k}\|_2 \leq \left(\sum_{k=0}^{\infty} \|C_k\|_2\right)^2 < \infty.$$

Next,

$$\begin{aligned} E\left(\left\| T^{-1} \sum_{t=i+1}^{T} \left(e_{t-i-k}e'_{t-i-k} - I_n\right)\right\|_2\right) &= E\left(\left\| T^{-1} \sum_{t=1}^{T-i} \left(e_{t-k}e'_{t-k} - I_n\right)\right\|_2\right) \\ &\leq E\left(\left\| T^{-1} \sum_{t=1}^{T-i} \left(e_{t-k}e'_{t-k} - I_n\right)\right\|_1\right), \end{aligned}$$

since $\|A\|_2 \leq \|A\|_1$ for any matrix A. Each element of $\sum_{t=1}^{T-i} \left(e_{t-k}e'_{t-k} - I_n\right)$ is a martingale, so

$$E\left(\left\| T^{-1} \sum_{t=1}^{T-i} \left(e_{t-k}e'_{t-k} - I_n\right)\right\|_1\right) \leq E\left(\left\| T^{-1} \sum_{t=1}^{T} \left(e_{t-k}e'_{t-k} - I_n\right)\right\|_1\right),$$

and

$$\psi_T \leq \sup_{k\geq 0} E\left(\left\| T^{-1} \sum_{t=1}^{T} \left(e_{t-k}e'_{t-k} - I_n\right)\right\|_1\right) = o\left(1\right),$$

as in Andrews (1988, Proof of Lemma) since each element of $\{e_t e_t'\}_{t\in\mathbb{Z}}$ is uniformly integrable under ($\mathcal{V}1$) (iii). Finally,

$$\begin{aligned}\left(\left\|T^{-1}\sum_{t=i+1}^{T} e_{t-j}e_{t-i-k}'\right\|_2\right)^2 &= T^{-2}\cdot tr\left(\left(\sum_{t=i+1}^{T} e_{t-j}e_{t-i-k}'\right)'\left(\sum_{s=i+1}^{T} e_{s-j}e_{s-i-k}'\right)\right)\\ &= T^{-2}\cdot\sum_{t=i+1}^{T}\sum_{s=i+1}^{T} e_{t-j}'e_{s-j}e_{s-i-k}'e_{t-i-k},\end{aligned}$$

since $(\|A\|_2)^2 = tr\,(A'A)$ for any matrix A. If $s \neq t$ and $j \neq i+k$ then

$$E\left(e_{t-j}'e_{s-j}e_{s-i-k}'e_{t-i-k}\right) = 0,$$

since e.g.

$$E\left(e_{t-j}'e_{s-j}e_{s-i-k}'e_{t-i-k}\right) = E\left(E\left(e_{t-j}'e_{s-j}e_{s-i-k}'e_{t-i-k} \mid \mathcal{F}_{t-j-1}\right)\right) = 0,$$

when $s < t$ and $j < i+k$. If $s = t$ and $j \neq i+k$ then

$$E\left(e_{t-j}'e_{s-j}e_{s-i-k}'e_{t-i-k}\right) = n^2,$$

since e.g.

$$\begin{aligned}E\left(e_{t-j}'e_{t-j}e_{t-i-k}'e_{t-i-k}\right) &= E\left(E\left(e_{t-j}'e_{t-j}e_{t-i-k}'e_{t-i-k} \mid \mathcal{F}_{t-j-1}\right)\right)\\ &= E\left(n\cdot e_{t-i-k}'e_{t-i-k}\right) = n^2,\end{aligned}$$

when $j < i+k$. Therefore,

$$E\left(\left(\left\|T^{-1}\sum_{t=i+1}^{T} e_{t-j}e_{t-i-k}'\right\|_2\right)^2\right) = T^{-2}\cdot\sum_{t=i+1}^{T} n^2 = \frac{T-i}{T^2}n^2 \leq \frac{n^2}{T},$$

whenever $j \neq i+k$. By the Cauchy-Schwarz inequality,

$$E\left(\left\|T^{-1}\sum_{t=i+1}^{T}e_{t-j}e'_{t-i-k}\right\|_2\right) \le \left(E\left(\left(\left\|T^{-1}\sum_{t=i+1}^{T}e_{t-j}e'_{t-i-k}\right\|_2\right)^2\right)\right)^{1/2}$$
$$\le \frac{n}{T^{1/2}},$$

and as a consequence,

$$\begin{aligned}
\eta_T &= \left(\max_{0\le i\le T-1}\sup_{\substack{j,k\ge 0\\ j\ne i+k}} E\left(\left\|T^{-1}\sum_{t=i+1}^{T}e_{t-j}e'_{t-i-k}\right\|_2\right)\right)\left(\sum_{k=0}^{\infty}\sum_{\substack{j=0\\ j\ne i+k}}^{\infty}\|C_k\|_2\cdot\|C_j\|_2\right)\\
&\le \frac{n}{T^{1/2}}\left(\sum_{k=0}^{\infty}\sum_{\substack{j=0\\ j\ne i+k}}^{\infty}\|C_k\|_2\cdot\|C_j\|_2\right)\\
&\le \frac{n}{T^{1/2}}\left(\sum_{k=0}^{\infty}\|C_k\|_2\right)^2\\
&= O\left(T^{-1/2}\right),
\end{aligned}$$

since $\sum_{k=0}^{\infty}\|C_k\|_2<\infty$ under $(\mathcal{V}1)$ (i). ■

3.2. Proof of Corollary 2. Since $\hat{\Gamma}_T(\theta_0)-\Gamma\to_p 0$ and $\hat{\Omega}_T(\theta_0)-\Omega\to_p 0$ (Theorem 1), it suffices to show that $\hat{\Gamma}_T\left(\hat{\theta}_T\right)-\hat{\Gamma}_T(\theta_0)\to_p 0$ and $\hat{\Omega}_T\left(\hat{\theta}_T\right)-\hat{\Omega}_T(\theta_0)\to_p 0$. As in the proofs of Theorems 2 and 3 in Hansen (1992), Condition $(\mathcal{V}2)$ implies that for some $Q_T=O_p(1)$, where Q_T does not depend on s_T,

$$\left\|\hat{\Gamma}_T\left(\hat{\theta}_T\right)-\hat{\Gamma}_T(\theta_0)\right\|_2\le\left(T^{-1/2}s_T\right)\left(s_T^{-1}\sum_{i=0}^{T-1}\left|k\left(\frac{i}{s_T}\right)\right|\right)\cdot Q_T.$$

Now, $T^{-1/2}s_T=o(1)$ under $(\mathcal{S})$ and $s_T^{-1}\sum_{i=0}^{T-1}|k(i/s_T)|=O(1)$ under $(\mathcal{K}),(\mathcal{S})$ (Lemma 4). As a consequence, $\left\|\hat{\Gamma}_T\left(\hat{\theta}_T\right)-\hat{\Gamma}_T(\theta_0)\right\|_2=o_p(1)$. An analogous argument can be used to show that $\left\|\hat{\Omega}_T\left(\hat{\theta}_T\right)-\hat{\Omega}_T(\theta_0)\right\|_2=o_p(1)$ and the desired result follows. ■

3.3. Proof of Theorem 3. Under $(\mathcal{S}')$, $\alpha_l \leq \alpha \leq \alpha_u$ with probability arbitrarily close to unity for sufficiently large T and appropriately chosen $\alpha_l > 0$ and $\alpha_u < \infty$. Consequently, it suffices to show that for any $0 < \alpha_l < \alpha_u < \infty$,

$$\sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \Gamma \right\|_2 = o_{p^*}(1),$$

which is easily shown to imply $\sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Omega}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \Omega \right\|_2 = o_{p^*}(1)$, where $o_{p^*}(1)$ denotes convergence to zero in outer probability.[7] Notice that

$$\begin{aligned} \sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \Gamma \right\|_2 \leq & \sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \hat{\Gamma}_T (\theta_0, \alpha \cdot s_T) \right\|_2 \\ & + \sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T (\theta_0, \alpha \cdot s_T) - E \left(\hat{\Gamma}_T (\theta_0, \alpha \cdot s_T) \right) \right\|_2 \\ & + \sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| E \left(\hat{\Gamma}_T (\theta_0, \alpha \cdot s_T) \right) - \Gamma \right\|_2 . \end{aligned}$$

We shall show that each term on the right hand side is $o_{p^*}(1)$. As in the proof of Corollary 2, we can show that

$$\left\| \hat{\Gamma}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \hat{\Gamma}_T (\theta_0, \alpha \cdot s_T) \right\|_2 \leq \left(T^{-1/2} \alpha \cdot s_T \right) \left(\frac{1}{\alpha \cdot s_T} \sum_{i=0}^{T-1} \left| k \left(\frac{i}{\alpha \cdot s_T} \right) \right| \right) \cdot Q_T,$$

for any $\alpha > 0$, where $Q_T = O_p(1)$ and Q_T does not depend on α or s_T. Now, $T^{-1/2} \alpha \cdot s_T = o(1)$ and

$$\begin{aligned} \frac{1}{\alpha \cdot s_T} \sum_{i=0}^{T-1} \left| k \left(\frac{i}{\alpha \cdot s_T} \right) \right| & \leq \frac{1}{\alpha \cdot s_T} \sum_{i=0}^{T-1} \bar{k} \left(\frac{i}{\alpha \cdot s_T} \right) \\ & \leq \frac{\alpha_u}{\alpha} \cdot \frac{1}{\alpha_u \cdot s_T} \sum_{i=0}^{T-1} \bar{k} \left(\frac{i}{\alpha_u \cdot s_T} \right) \\ & \leq \frac{\alpha_u}{\alpha_l} \cdot \frac{1}{\alpha_u \cdot s_T} \sum_{i=0}^{T-1} \bar{k} \left(\frac{i}{\alpha_u \cdot s_T} \right) \\ & = O(1), \end{aligned}$$

[7] To avoid measurability complications, we consider convergence in outer probability rather than convergence in probability.

for any $0 < \alpha_l \leq \alpha \leq \alpha_u < \infty$ under $(\mathcal{K})$ and $(\mathcal{S})$ (Lemma 4), so

$$\sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T \left(\hat{\theta}_T, \alpha \cdot s_T \right) - \hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) \right\|_2 = o_{p^*}(1).$$

Next,

$$\begin{aligned} \left\| \hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) - E\left(\hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) \right) \right\|_2 &\leq \sum_{i=0}^{T-1} \left| k\left(\frac{i}{\alpha \cdot s_T} \right) \right| \cdot \left\| \hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T} \right) \right\|_2 \\ &\leq \sum_{i=0}^{T-1} \bar{k}\left(\frac{i}{\alpha_u \cdot s_T} \right) \cdot \left\| \hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T} \right) \right\|_2, \end{aligned}$$

for any $0 < \alpha_l \leq \alpha \leq \alpha_u < \infty$ under $(\mathcal{K})$. It follows from the proof of Theorem 1 that

$$\sum_{i=0}^{T-1} \bar{k}\left(\frac{i}{\alpha_u \cdot s_T} \right) \cdot \left\| \hat{\Sigma}_{i,T} - E\left(\hat{\Sigma}_{i,T} \right) \right\|_2 = o_p(1),$$

establishing

$$\sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| \hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) - E\left(\hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) \right) \right\|_2 = o_{p^*}(1).$$

Finally,

$$\sup_{\alpha_l \leq \alpha \leq \alpha_u} \left\| E\left(\hat{\Gamma}_T \left(\theta_0, \alpha \cdot s_T \right) \right) - \Gamma \right\| \to 0,$$

by continuity of $\bar{k}(\cdot)$ at zero and $\sum_{i=0}^{\infty} \|\Sigma_i\|_2 < \infty$. ■

REFERENCES

ANDREWS, D. W. K. (1988): "Laws of Large Numbers for Non-Identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

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

Davidson, J. (1999): "When is a Time Series I(0)? Evaluating the Memory Properties of Nonlinear Dynamic Models," Working Paper, Cardiff University.

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

de Jong, R. M. (1998): "A Strong Consistency Proof for Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimators," Econometric Theory, forthcoming.

de Jong, R. M., and J. Davidson (1999): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," Econometrica, forthcoming.

Park, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.

Phillips, P. C. B., and B. E. Hansen (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Variables," *Review of Economic Studies*, 57, 99–125.

Robinson, P. M. (1991): "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica*, 59, 1329–1363.

Working Paper