

Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices*

James Davidson
Cardiff University

Robert M. de Jong
Michigan State University

Forthcoming in *Econometrica*, 2000

Abstract

Conditions are derived for the consistency of kernel estimators of the covariance matrix of a sum of vectors of dependent heterogeneous random variables, which match those of the currently best-known conditions for the central limit theorem, as required for a unified theory of asymptotic inference. These include finite moments of order no more than $2 + \delta$ for $\delta > 0$, trending variances, and variables which are near-epoch dependent on a mixing process, but not necessarily mixing. The results are also proved for the case of sample-dependent bandwidths.

1 Introduction

This paper derives conditions for the consistency of kernel estimators of the covariance matrix of a weighted sum of vectors of dependent heterogeneous random variables. This is a problem which has been studied recently by, among others, Newey and West (1987), Gallant and White (1988), Andrews (1991), Pötscher and Prucha (1991b), Andrews and Monahan (1992), and Hansen (1992). A leading example of its application is where an estimator $\hat{\theta}_n$ ($r \times 1$) of a parameter θ_0 is known to satisfy

$$n^{1/2}(\hat{\theta}_n - \theta_0) - B_n^{-1} \sum_{t=1}^n X_{nt}(\theta_0) \xrightarrow{p} 0 \quad (1.1)$$

where the random function $X_{nt}(\theta)$ ($p \times 1$) has mean zero at θ_0 , and B_n ($r \times p$) is some nonrandom matrix that is usually easily estimated.¹ Applying a central limit theorem (CLT) to the second term in (1.1) leads to

$$(B_n \Omega_n B_n')^{-1/2} n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_r) \quad (1.2)$$

where, letting $X_{nt} = X_{nt}(\theta_0)$,

$$\Omega_n = \sum_{t=1}^n \sum_{s=1}^n E X_{nt} X'_{ns}. \quad (1.3)$$

A complete asymptotic distribution theory for $\hat{\theta}_n$ must incorporate whatever conditions are needed to ensure consistent estimation of Ω_n when the array X_{nt} is dependently and heterogeneously

*We thank Benedict Pötscher, Peter Robinson and three anonymous referees for their comments on the paper. The first author thanks Herman Bierens for discussion that eventually led to the proof of Lemma A-2. Responsibility for any errors is ours alone.

¹For example, in the linear regression model $y_t = z_t' \theta_0 + u_t$, we would have $X_{nt}(\theta_0) = n^{-1/2} z_t u_t$ and $B_n = n^{-1} \sum_{t=1}^n E z_t z_t'$.

distributed. An undesirable feature of the above-cited studies is that they impose conditions stronger in important respects than are known to be required for the application of a CLT to the same variables. All impose either a form of stationarity or uniform boundedness in L_p -norms for some $p \geq 2$, precluding trending moments. All except Pötscher and Prucha (1991b) assume strong or uniform mixing, and also that the true covariance matrix converges to some well-defined limit. Finally, all except Hansen (1992) assume finite fourth moments. However, while the latter paper was the first to relax the moment restrictions, the proof contains an error (see De Jong 1999), and the correction involves strengthening the necessary conditions on the bandwidth parameter, thereby limiting the scope of Hansen's results.

In this paper, we will bridge the gap between asymptotic normality and consistent covariance matrix estimation by obtaining conditions for the latter similar to those obtained for the CLTs in Davidson (1992, 1993) and De Jong (1997). These are the best such results for heterogeneous processes currently known to us. The weak dependence is characterized by near-epoch dependence on a mixing process, a more general property than mixing. We prove our results for stochastic (sample-dependent) bandwidths for the kernel estimators, and also show that in the standard cases, a sufficient condition on the bandwidth for consistency of the variance estimator is that its ratio with the sample size converges to zero. Of the above-cited references only Andrews (1991) gives results that allow for such a behavior of the bandwidth, although under stronger assumptions. Our central result shows convergence to zero of the difference between the elements of the estimated and the true covariance matrix. There is no need to assume that the true covariance matrix itself converges to a well-defined limit. We will argue that relaxing the so-called size conditions on the sequences measuring dependence for the case of covariance matrix estimation for root- n consistent minimization estimators is not possible, and in that sense, our dependence conditions are the best possible. Finally, we are able to generalise from the root- n consistency represented in (1.2), by giving results subject to the condition

$$n^{1/2}\kappa_n(\hat{\theta} - \theta_0) = O_p(1) \quad (1.4)$$

where κ_n is a diagonal matrix function of n . This permits such applications as the Phillips-Perron (1988) unit root test and the Phillips-Hansen (1990) fully modified least squares estimator, but our assumptions also permit nonlinear models involving deterministic or stochastic trends.

Section 2 of the paper will present our main results. The proofs can be found in the Appendix.

2 Main results

The consistency results are inspired by the proofs of the CLTs for possibly trending-variance processes in Davidson (1992, 1993), and De Jong (1997), in which showing the consistency of a certain variance estimate is an essential step. Let

$$\hat{\Omega}_n = \sum_{t=1}^n \sum_{s=1}^n X_{nt} X'_{ns} k((t-s)/\gamma_n), \quad (2.1)$$

where $k(\cdot)$ is called the kernel function, and γ_n , the bandwidth or lag truncation parameter, is an increasing function of n . X_{nt} ($p \times 1$) is to be thought of as a random function of parameters θ ($r \times 1$), which unless we indicate otherwise is evaluated at the true parameter point $\theta_0 \in \Theta$, where Θ is a convex, compact subset of \mathbb{R}^r . More generally we write $X_{nt}(\theta)$, and our operational estimator is

$$\hat{\Omega}_n(\hat{\theta}_n) = \sum_{t=1}^n \sum_{s=1}^n X_{nt}(\hat{\theta}_n) X'_{ns}(\hat{\theta}_n) k((t-s)/\gamma_n). \quad (2.2)$$

The array notation allows us to generalize our results, but direct comparability with those of Andrews (1991) and Hansen (1992) is obtained by considering the case $X_{nt} = n^{-1/2}X_t$. In this

case, (2.1) becomes

$$\hat{\Omega}_n = n^{-1} \sum_{t=1}^n \sum_{s=1}^n X_t X'_s k((t-s)/\gamma_n). \quad (2.3)$$

We assume that $k(\cdot)$ belongs to the function class \mathcal{K} , defined as follows

Assumption 1 $k(\cdot) \in \mathcal{K}$, where

$$\mathcal{K} = \{k(\cdot) : \mathbb{R} \rightarrow [-1, 1] \mid k(0) = 1, k(x) = k(-x) \forall x \in \mathbb{R},$$

$$|quad \int_{-\infty}^{\infty} |k(x)| dx < \infty, \int_{-\infty}^{\infty} |\psi(\xi)| d\xi < \infty,$$

$$k(\cdot) \text{ is continuous at } 0 \text{ and at all but a finite number of points}\}, \quad (2.4)$$

where

$$\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x) e^{i\xi x} dx. \quad (2.5)$$

In Andrews (1991), function classes \mathcal{K}_1 and \mathcal{K}_2 are defined.² Our class \mathcal{K} contains Andrews' class \mathcal{K}_2 , since his requirement $\psi(\xi) \geq 0$ is replaced by $\int_{-\infty}^{\infty} |\psi(\xi)| d\xi < \infty$, a weaker condition in view of the fact that $\int_{-\infty}^{\infty} \psi(\xi) d\xi = k(0) = 1$. As Andrews notes, the class \mathcal{K}_2 corresponds to that for which $\hat{\Omega}_n(\theta)$ is positive semidefinite with probability one, and hence class \mathcal{K} contains these cases. The Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernels, although not the truncated kernel, are all included in \mathcal{K} .³

We need the following assumptions linking the dependence and heterogeneity of the processes and the permitted bandwidth. First, recall that X_{nt} is said to be L_2 -near epoch dependent (L_2 -NED) on an array V_{nt} if for $m \geq 0$, $\|X_{nt} - E(X_{nt} | \mathcal{V}_{n,t-m}^{t+m})\|_2 \leq d_{nt} \nu(m)$ where $\mathcal{V}_{n,t-m}^{t+m} = \sigma(\{V_{n,t-m}, \dots, V_{n,t+m}\})$, d_{nt} is a nonstochastic triangular array, and $\nu(m) \rightarrow 0$ as $m \rightarrow \infty$.⁴ The sequence $\nu(m)$ is said to be of size $-\lambda$ if $\nu(m) = O(m^{-\lambda-\varepsilon})$ for some $\varepsilon > 0$. See, e.g. Gallant and White (1988), Davidson (1994), for additional details.

Assumption 2 X_{nt} is L_2 -near epoch dependent of size $-1/2$ on a strong mixing (resp. uniform mixing) random array V_{nt} of size $-r/(r-2)$ (resp. $-r/(2r-2)$), and there exists a triangular constant array c_{nt} such that

$$\sup_{n \geq 1} \sup_{1 \leq t \leq n} (\|X_{nt}\|_r + d_{nt}) / c_{nt} < \infty, \quad (2.6)$$

for some $r > 2$, and

$$\sup_{n \geq 1} \sum_{t=1}^n c_{nt}^2 < \infty. \quad (2.7)$$

Under uniform mixing, $r = 2$ is also permitted if $|X_{nt}|^2 / c_{nt}^2$ is uniformly integrable.

²Note that the definition of class \mathcal{K}_1 in Andrews (1991) specifies square-integrability of k , which is incorrect. This is corrected to absolute integrability (as here) in Andrews and Monahan (1992), see the footnote on page 955.

³The Tukey-Hanning and truncated kernels belong to Andrews' (1991) \mathcal{K}_1 class, but not to \mathcal{K}_2 . See Andrews (1991) for definitions of all these kernel functions.

⁴In this paper, the L_q -norm of a random matrix X is defined as $\|X\|_q = (\sum_i \sum_j E|X_{ij}|^q)^{1/q}$ for $q \geq 1$. Similarly, $|\cdot|$ denotes the Euclidean norm of the appropriate dimension.

Assumption 3 $\lim_{n \rightarrow \infty} (\gamma_n^{-1} + \gamma_n \max_{1 \leq t \leq n} c_{nt}^2) = 0$.

The reference case of the constant array specified here is simply $c_{nt} = n^{-1/2}$, corresponding to $X_{nt} = n^{-1/2} X_t$ where X_t is a stationary sequence. See Davidson (1992, 1993), de Jong (1997) for a discussion of other examples involving trends. These dependence conditions match those of the best central limit theorems for NED/mixing arrays. It also can be shown (see De Jong 1997 and Doukhan, Massart and Rio 1994) that relaxing the strong mixing size requirement on V_{nt} , or the L_2 -NED size requirement on X_{nt} , implies that the latter need not be covariance summable. In the reference case of (1.1) with $B_n = B$, this would imply that $n^{-1} E(\sum_{t=1}^n X_t)(\sum_{t=1}^n X_t') \rightarrow \infty$ as $n \rightarrow \infty$, so that (1.1) would be incompatible with root- n consistency of $\hat{\theta}_n$. For this case our result is in effect the best possible with respect to dependence conditions, when dependence is characterised by Assumption 2.

In the reference case, Assumption 3 permits $\gamma_n = o(n)$. For comparison, order of magnitude limits placed on the bandwidth in previous consistency proofs (under differing regularity conditions, although generally stronger than ours) include $o(n^{1/4})$ by Newey and West (1987), $o(n^{1/3})$ by Pötscher and Prucha (1991b), $o(n^{1/2})$ by Kool (1988) and Hansen (1992), and $o(n)$ by Andrews (1991).⁵

These assumptions suffice for consistency of $\hat{\Omega}_n$ in (2.1), in which θ_0 is implicitly known.

Theorem 2.1 *Under Assumptions 1, 2 and 3,*

$$\hat{\Omega}_n - \Omega_n \xrightarrow{p} 0.$$

The next step is to show that $\hat{\Omega}_n(\hat{\theta}_n)$ is asymptotically equivalent to $\hat{\Omega}_n$. For this we need the following extra assumption.

Assumption 4 (a) $n^{1/2} \kappa_n(\hat{\theta}_n - \theta_0) = O_p(1)$, where $\kappa_n = \text{diag}(\kappa_{1n}, \dots, \kappa_{rn})$ is a deterministic $r \times r$ matrix;

(b) $n^{-1/2} \kappa_n^{-1} \sum_{t=1}^n E \partial X_{nt}(\theta) / \partial \theta$ is continuous at θ_0 uniformly in n ;

(c) there exists $\mathcal{N} \subset \Theta$, an open neighbourhood of θ_0 , such that,

$$\limsup_{n \rightarrow \infty} \sum_{t=1}^n E \sup_{\theta \in \mathcal{N}} \left| \kappa_n^{-1} \frac{\partial X_{nt}(\theta)}{\partial \theta'} \right|^2 < \infty, \quad (2.8)$$

and either (i)

$$\sup_{\theta \in \mathcal{N}} \left\| n^{-1/2} \kappa_n^{-1} \sum_{t=1}^n e^{i\xi t / \gamma_n} \left(\frac{\partial X_{nt}(\theta)}{\partial \theta'} - E \frac{\partial X_{nt}(\theta)}{\partial \theta'} \right) \right\|_2 \rightarrow 0 \quad (2.9)$$

for all $\xi \in \mathbb{R}$, or (ii), $n^{-1/2} \gamma_n = o(1)$ and

$$\sup_{\theta \in \mathcal{N}} \left| \sum_{t=1}^n \kappa_n^{-1} \frac{\partial X_{nt}'(\theta)}{\partial \theta} \frac{\partial X_{nt}(\theta)}{\partial \theta'} \kappa_n^{-1} \right| = O_p(1). \quad (2.10)$$

In case (1.1) for example, we would set $\kappa_n = I$. For a different choice of κ_n , Assumption 4(a) allows the elements of $\hat{\theta}$ to converge at different rates.

Theorem 2.2 *Under Assumptions 1, 2, 3 and 4,*

$$\hat{\Omega}_n(\hat{\theta}_n) - \Omega_n \xrightarrow{p} 0.$$

⁵ Andrews (1991) also shows that bandwidths growing faster than $o(n^{1/2})$ can never be optimal under a mean squared error optimality criterion. However, the MSE condition is only relevant when fourth moments of the data exist, and hence for only a subset of the cases covered by our assumptions.

If Assumption 4(i) applies, Assumption 3 represents the only restrictions on the bandwidth for this result. The complex-valued weights in (2.9) lie on the unit circle, and, if it holds in their absence, a uniform law of large numbers for heterogeneous processes could usually be adapted to this case by considering the real and imaginary parts separately. Lower-level assumptions sufficient for the convergence in L_2 cannot be cited without referring specifically to the functional form of $X_{nt}(\theta)$, but we note that arbitrary mixing and NED sizes are often sufficient, and various results exist for extending the NED property to transformations of an NED process.⁶ It might suffice merely to extend Assumption 2 to hold uniformly on \mathcal{N} , for example. Also note, in this connection, that conditions similar to Assumption 4(i) often need to be invoked to prove the convergence represented in (1.1).

Assumption 4(ii) accommodates processes with deterministic or stochastic trends, and in a linear model, corresponds to Assumption V3 of Hansen (1992). Thus, letting W_t ($r \times 1$) represent an I(1) vector of cointegrated variables, suppose $X_{nt}(\hat{\theta}_n) = n^{-1/2}\hat{\theta}'_n W_t$ is the normalised OLS residual. $\hat{\theta}_n$ is n -consistent and we can set $\kappa_n = n^{1/2}I$. In this case the assumption is trivially satisfied under standard regularity conditions, although note that Assumption 4(i) fails because with I(1) variables, the array in (2.9) converges weakly to a random limit, but not in probability. However, unlike Hansen's, our assumption still allows the gradient to depend on θ and accordingly permits nonlinear models involving trends.

We lastly establish a result that allows bandwidths to be stochastic. See Andrews and Monahan (1992) and Newey and West (1994) for discussion of such procedures. While the optimal bandwidths discussed in those papers are justified on the MSE criterion, which is available only when fourth moments exist, our result can be thought of as reassuring practitioners of the consistency of these estimators even if the fourth moment condition fails. Let $\hat{\Omega}_n(\hat{\theta}_n, \hat{\delta}_n)$ denote $\hat{\Omega}_n(\hat{\theta}_n)$ as before, but evaluated at the possibly stochastic bandwidth $\hat{\delta}_n$ instead of γ_n .

Theorem 2.3 *Assume that Assumptions 1, 2, 3, and 4 hold. In addition, assume that $\hat{\delta}_n = \hat{\alpha}_n \gamma_n$, where $\hat{\alpha}_n = O_p(1)$ and $1/\hat{\alpha}_n = O_p(1)$, γ_n satisfies the bandwidth conditions of the assumptions, and for all $\varepsilon \in (0, 1)$, $k(\cdot)$ satisfies*

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |k(\alpha x)| dx < \infty. \quad (2.11)$$

and

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |\psi(\alpha \xi)| d\xi < \infty. \quad (2.12)$$

Then

$$\hat{\Omega}_n(\hat{\theta}_n, \hat{\delta}_n) - \Omega_n \xrightarrow{P} 0. \quad (2.13)$$

Note that the conditions on $k(\cdot)$ that are imposed in Theorem 2.3 are satisfied for the Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernels.

A Appendix: Proofs

In the proofs and associated lemmas we assume for simplicity that X_{nt} is real-valued (i.e. $p = 1$). As pointed out by Newey and West (1987), the results can be extended directly to vector-valued X_{nt} , provided the assumptions hold for the scalars $\lambda' X_{nt}$ for all conformable nonzero fixed vectors λ .

⁶See Davidson (1994), Chapters 17 and 19, for further information on these issues.

The proofs require the *mixingale* concept. Let \mathcal{H}_{nt} denote an array of σ -fields that is non-decreasing in t for each n . The array $\{Y_{nt}, \mathcal{H}_{nt}\}$, is called an L_2 -mixingale of size $-\lambda$ if for a triangular constant array c_{nt} (the mixingale magnitude indices)

$$\|Y_{nt} - E(Y_{nt}|\mathcal{H}_{n,t+m})\|_2 \leq c_{nt}\psi(m+1), \quad (\text{A-1})$$

$$\|E(Y_{nt}|\mathcal{H}_{n,t-m})\|_2 \leq c_{nt}\psi(m), \quad (\text{A-2})$$

for $m \geq 0$, and $\psi(m) = O(m^{-\lambda-\varepsilon})$ for some $\varepsilon > 0$. Assumption 2 implies that $\{Y_{nt}, \mathcal{H}_{nt}\}$ is a mixingale of size $-1/2$ with mixingale magnitude indices c_{nt} as defined therein, for $\mathcal{H}_{nt} = \sigma(\{V_{nt}, V_{n,t-1}, \dots\})$. This follows by Theorem 17.5 of Davidson (1994). In this Appendix, \mathcal{H}_{nt} will denote this sigma field. We will make use of the following result (Theorem 1.6 of McLeish 1975a):

Lemma A.1 *Let $\{Y_{nt}, \mathcal{H}_{nt}\}$ be a real valued L_2 -mixingale of size $-1/2$ with mixingale magnitude indices c_{nt} . Then $E(\sum_{t=1}^n Y_{nt})^2 = O(\sum_{t=1}^n c_{nt}^2)$.*

Proof of Theorem 2.1

Define $b_n = \lceil \gamma_n/\delta \rceil$ and $r_n = \lfloor n/b_n \rfloor$ where $\lfloor x \rfloor$ denotes the largest integer not exceeding x ,

$$\eta_\delta(x) = (\delta^2 2\pi)^{-1/2} \exp(-x^2 \delta^{-2} 2^{-1}), \quad (\text{A-3})$$

$$\Omega_{1n\delta} = \sum_{t=-n+1}^{2n} \left(\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,n]}(|l|) \right) \left(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) \right), \quad (\text{A-4})$$

$$\Omega_{2n\delta} = \sum_{t=-n+1}^{2n} \left(\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,b_n]}(|l|) 1_{[0,n]}(|l|) \right) \left(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) \right) \quad (\text{A-5})$$

where $1_{[a,b]}(\cdot)$ denotes the indicator function of the interval $[a, b]$, and

$$\Omega_{3n\delta} = \sum_{t=-n+1}^{2n} \left(\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,b_n]}(|l|) \right) \left(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) 1_{[0,b_n]}(|j|) \right). \quad (\text{A-6})$$

Note that

$$\begin{aligned} \|\hat{\Omega}_n - \Omega_n\|_1 &\leq \|\hat{\Omega}_n - \Omega_{1n\delta}\|_1 + \|\Omega_{1n\delta} - \Omega_{2n\delta}\|_1 + \|\Omega_{2n\delta} - \Omega_{3n\delta}\|_1 \\ &\quad + \|\Omega_{3n\delta} - E\Omega_{3n\delta}\|_1 + \|E\Omega_{3n\delta} - E\Omega_{2n\delta}\|_1 + \|E\Omega_{2n\delta} - E\Omega_{1n\delta}\|_1 \\ &\quad + \|E\Omega_{1n\delta} - E\hat{\Omega}_n\|_1 + |E\hat{\Omega}_n - \Omega_n| \\ &\leq 2\|\hat{\Omega}_n - \Omega_{1n\delta}\|_1 + 2\|\Omega_{1n\delta} - \Omega_{2n\delta}\|_1 + 2\|\Omega_{2n\delta} - \Omega_{3n\delta}\|_1 \\ &\quad + \|\Omega_{3n\delta} - E\Omega_{3n\delta}\|_1 + |E\hat{\Omega}_n - \Omega_n| \end{aligned} \quad (\text{A-7})$$

The four lemmas that follow show that each of the terms on the right-hand side of (A-7) vanishes if we first take the limsup as n approaches infinity and then, if required, the limit as $\delta \rightarrow 0$. \blacksquare

Lemma A.2 Under Assumptions 1, 2 and 3,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \|\hat{\Omega}_n - \Omega_{1n\delta}\|_1 = 0.$$

Proof. First note that $k(x) = \int_{-\infty}^{\infty} \exp(i\xi x) \psi(\xi) d\xi$, by the inversion formula for Fourier transforms. Therefore, using the fact that $k(\cdot)$ and hence $\psi(\cdot)$ are even functions,

$$\hat{\Omega}_n = \int_{-\infty}^{\infty} \left(\left(\sum_{t=1}^n X_{nt} \cos(t\xi/\gamma_n) \right)^2 + \left(\sum_{t=1}^n X_{nt} \sin(t\xi/\gamma_n) \right)^2 \right) \psi(\xi) d\xi. \quad (\text{A-8})$$

Next, note that

$$\Omega_{1n\delta} = \sum_{t=1}^n \sum_{s=1}^n X_{nt} X_{ns} k_{n\delta}((t-s)/\gamma_n), \quad (\text{A-9})$$

where

$$k_{n\delta}(x) = \gamma_n^{-1} \sum_{l=-n}^n k(l/\gamma_n) \eta_{\delta}(x + l/\gamma_n). \quad (\text{A-10})$$

The inverse Fourier transform of $k_{n\delta}(\cdot)$ can be shown to be

$$(2\pi)^{-1} \int_{-\infty}^{\infty} e^{-i\xi x} k_{n\delta}(x) dx = \psi_n(\xi) e^{-\delta^2 \xi^2 / 2} \quad (\text{A-11})$$

where

$$\psi_n(\xi) = (\gamma_n 2\pi)^{-1} \sum_{l=-n}^n k(l/\gamma_n) e^{i\xi l/\gamma_n}, \quad (\text{A-12})$$

and note that for all $x \in \mathbb{R}$, $\lim_{n \rightarrow \infty} \psi_n(x) = \psi(x)$. From the representation of equation (A-8) and the properties of $\|\cdot\|_1$, it follows that

$$\begin{aligned} \|\hat{\Omega}_n - \Omega_{1n\delta}\|_1 &\leq \int_{-\infty}^{\infty} \left(E \left(\sum_{t=1}^n X_{nt} \cos(t\xi/\gamma_n) \right)^2 + E \left(\sum_{t=1}^n X_{nt} \sin(t\xi/\gamma_n) \right)^2 \right) \\ &\quad \times \left| \psi(\xi) - \psi_n(\xi) e^{-\delta^2 \xi^2 / 2} \right| d\xi. \end{aligned} \quad (\text{A-13})$$

Therefore, by Lemma A.1,

$$\|\hat{\Omega}_n - \Omega_{1n\delta}\|_1 = O \left(\int_{-\infty}^{\infty} \left| \psi(\xi) - \psi(\xi) e^{-\delta^2 \xi^2 / 2} \right| d\xi + \int_{-\infty}^{\infty} |\psi(\xi) - \psi_n(\xi)| e^{-\delta^2 \xi^2 / 2} d\xi \right) \quad (\text{A-14})$$

and by first taking the limsup as $n \rightarrow \infty$ and then the limit as $\delta \rightarrow 0$, the result follows by dominated convergence. ■

Lemma A.3 Under Assumptions 1, 2 and 3,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \|\Omega_{1n\delta} - \Omega_{2n\delta}\|_1 = 0, \quad (\text{A-15})$$

and

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \|\Omega_{2n\delta} - \Omega_{3n\delta}\|_1 = 0. \quad (\text{A-16})$$

Proof. We prove (A-15), the case of (A-16) being similar.

$$\begin{aligned}
\|\Omega_{1n\delta} - \Omega_{2n\delta}\|_1 &= \left\| \sum_{t=-n+1}^{2n} \left(\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) \mathbf{1}_{[0,n]}(|l|) \mathbf{1}_{[b_n+1,\infty)}(|l|) \right) \right. \\
&\quad \times \left. \left(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) \right) \right\|_1 \\
&\leq \left(\sum_{t=-n+1}^{2n} \gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 k(l/\gamma_n)^2 \mathbf{1}_{[b_n+1,\infty)}(|l|) \mathbf{1}_{[0,n]}(|l|) \right)^{1/2} \\
&\quad \times \left(\sum_{t=-n+1}^{2n} \gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 \eta_\delta(j/\gamma_n)^2 \right)^{1/2} \\
&= O \left(\left(\int_{1/\delta}^{\infty} k(x)^2 dx \right)^{1/2} \right) \rightarrow 0 \tag{A-17}
\end{aligned}$$

as $\delta \rightarrow 0$, where the inequality is got by successively the Cauchy-Schwarz inequality, Lemma A.1, and a second application of Cauchy-Schwarz. ■

Lemma A.4 Under Assumptions 1, 2 and 3,

$$\lim_{n \rightarrow \infty} \|\Omega_{3n\delta} - E\Omega_{3n\delta}\|_1 = 0 \tag{A-18}$$

for all $\delta > 0$.

Proof. Write

$$\Omega_{3n\delta} = \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} \tag{A-19}$$

where

$$Y_{nt} = \gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) \mathbf{1}_{[0,b_n]}(|l|), \tag{A-20}$$

$$Z_{nt} = \gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) \mathbf{1}_{[0,b_n]}(|j|). \tag{A-21}$$

Note that these are sums of at most b_n terms. Also define

$$\phi_{nt}^2 = \gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 k(l/\gamma_n)^2 \mathbf{1}_{[0,b_n]}(|l|), \tag{A-22}$$

$$\psi_{nt}^2 = \gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 \eta_\delta(j/\gamma_n)^2 \mathbf{1}_{[0,b_n]}(|j|). \tag{A-23}$$

Note that Y_{nt}^2/ϕ_{nt}^2 and Z_{nt}^2/ψ_{nt}^2 are uniformly integrable by Lemma 3.2 of Davidson (1992) and the discussion following that lemma. Define $h(K, x) = x \mathbf{1}_{[-K, K]}(x) + K \mathbf{1}_{(K, \infty)}(x) - K \mathbf{1}_{(-\infty, -K)}(x)$,

and let K_η be some constant to be chosen. Let $\tilde{Y}_{nt} = h(K_\eta \phi_{nt}, Y_{nt})$ and $\tilde{Z}_{nt} = h(K_\eta \psi_{nt}, Z_{nt})$. Then we have

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \left\| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} - \tilde{Y}_{nt} \tilde{Z}_{nt} \right\|_1 \\
& \leq \limsup_{n \rightarrow \infty} \left(\left\| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} 1_{(K_\eta \phi_{nt}, \infty)}(|Y_{nt}|) \right\|_1 + \left\| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} 1_{(K_\eta \psi_{nt}, \infty)}(|Z_{nt}|) \right\|_1 \right) \\
& \leq C \left(\sup_{t, n \geq 1} E \left(\frac{Y_{nt}^2}{\phi_{nt}^2} \right) 1_{(K_\eta^2, \infty)} \left(\frac{Y_{nt}^2}{\phi_{nt}^2} \right) \right)^{1/2} + C \left(\sup_{t, n \geq 1} E \left(\frac{Z_{nt}^2}{\psi_{nt}^2} \right) 1_{(K_\eta^2, \infty)} \left(\frac{Z_{nt}^2}{\psi_{nt}^2} \right) \right)^{1/2} \\
& < \eta
\end{aligned} \tag{A-24}$$

for some constant $C > 0$ by a large enough choice of K_η and because

$$\sup_{n \geq 1} \sum_{t=-n+1}^{2n} (\psi_{nt}^2 + \phi_{nt}^2) < \infty \tag{A-25}$$

by assumption. Next, note that for $\tilde{Y}_{nt} \tilde{Z}_{nt}$ we have for $m \geq 0$

$$\begin{aligned}
\left\| \tilde{Y}_{nt} \tilde{Z}_{nt} - E(\tilde{Y}_{nt} \tilde{Z}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right\|_2 & \leq \left\| \left(\tilde{Y}_{nt} - E(\tilde{Y}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right) \tilde{Z}_{nt} \right\|_2 \\
& \quad + \left\| E(\tilde{Y}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \left(\tilde{Z}_{nt} - E(\tilde{Z}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right) \right\|_2 \\
& \leq \left\| \tilde{Z}_{nt} \right\|_\infty \left\| \tilde{Y}_{nt} - E(\tilde{Y}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right\|_2 \\
& \quad + \left\| E(\tilde{Y}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right\|_\infty \left\| \tilde{Z}_{nt} - E(\tilde{Z}_{nt} | \mathcal{V}_{n, t-b_n m}^{t+b_n m}) \right\|_2 \\
& = O \left(m^{-1/2-\varepsilon} \gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n, t+j}^2 1_{[0, b_n]}(|j|) \right),
\end{aligned} \tag{A-26}$$

for $\varepsilon > 0$, in view of the near-epoch dependence assumption on X_{nt} .

We now introduce a blocking scheme. Define $r'_n = \lceil 3n/2b_n \rceil$, and note that

$$\begin{aligned}
\sum_{t=-n+1}^{2n} \tilde{Y}_{nt} \tilde{Z}_{nt} & = \sum_{i=1}^{r'_n} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \tilde{Y}_{nt} \tilde{Z}_{nt} + \sum_{i=1}^{r'_n} \sum_{t=(2i-1)b_n-n+1}^{2ib_n-n} \tilde{Y}_{nt} \tilde{Z}_{nt} + \sum_{t=r'_n b_n-n+1}^{2n} \tilde{Y}_{nt} \tilde{Z}_{nt} \\
& \equiv \sum_{i=1}^{r'_n} U_{ni} + \sum_{i=1}^{r'_n} U'_{ni} + \sum_{t=r'_n b_n-n+1}^{2n} \tilde{Y}_{nt} \tilde{Z}_{nt}.
\end{aligned} \tag{A-27}$$

For the last term, we have

$$\left\| \sum_{t=r'_n b_n-n+1}^{2n} \tilde{Y}_{nt} \tilde{Z}_{nt} \right\|_1 = O \left(b_n \max_{1 \leq t \leq n} c_{nt}^2 \right) = o(1) \tag{A-28}$$

by assumption. We will analyze the sum of the U_{ni} , noting that the case of the U'_{ni} is analogous. The assertion of the lemma follows if we can show that the first term of Equation (A-27) obeys a law of large numbers. Define

$$\mathcal{W}_{n, i-m}^{i+m} = \sigma(\{V_{n, (2i-2m-2)b_n-n+1}, \dots, V_{n, (2i+2m-1)b_n-n}\})$$

and note that U_{ni} is near epoch dependent on \mathcal{W}_{ni} because for $m \geq 1$

$$\begin{aligned}
\|U_{ni} - E(U_{ni}|\mathcal{W}_{n,i-m}^{i+m})\|_2 &\leq \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \left\| \tilde{Y}_{nt} \tilde{Z}_{nt} - E(\tilde{Y}_{nt} \tilde{Z}_{nt} | \mathcal{W}_{n,i-m}^{i+m}) \right\|_2 \\
&\leq \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \left\| \tilde{Y}_{nt} \tilde{Z}_{nt} - E(\tilde{Y}_{nt} \tilde{Z}_{nt} | \mathcal{V}_{n,t-mb_n}^{t+mb_n}) \right\|_2 \\
&= O \left(m^{-1/2-\varepsilon} \gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \right) \quad (\text{A-29})
\end{aligned}$$

where the second inequality follows because $\mathcal{V}_{n,t-mb_n}^{t+mb_n} \subseteq \mathcal{W}_{n,i-m}^{i+m}$ for $t \in [(2i-2)b_n-n+1, (2i-1)b_n-n]$ and $m \geq 1$, and the last step is by (A-26). For $m = 0$, the relevant result is

$$\|U_{ni} - E(U_{ni}|\mathcal{W}_{ni}^i)\|_2 \leq 2 \|U_{ni}\|_2 = O \left(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \right) \quad (\text{A-30})$$

We conclude that U_{ni} is also an L_2 -mixingale of size $-1/2$ with respect to the $\mathcal{F}_{ni} = \sigma(\{V_{n,2ib_n-n}, V_{n,2ib_n-n-1}, \dots\})$ because for all $m \geq 1$

$$\|E(U_{ni}|\mathcal{F}_{n,i-2m}) - EU_{ni}\|_2 = O \left(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) m^{-1/2-\varepsilon} \right) \quad (\text{A-31})$$

similarly to Davidson (1994), Theorem 17.5. For $m = 0$, the result from Equation (A-30) can again be used. Finally, note that from Lemma A.1 it follows that

$$E \left| \sum_{i=1}^{r'_n} U_{ni} - EU_{ni} \right|^2 = O \left(\sum_{i=1}^{r'_n} h_{ni}^2 \right) \quad (\text{A-32})$$

where the h_{ni} denote the mixingale magnitude indices of U_{ni} . Therefore, the lemma holds because

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \sum_{i=1}^{r'_n} \left(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \right)^2 \\
&\leq \lim_{n \rightarrow \infty} \sum_{i=1}^{r'_n} \gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \max_{1 \leq i \leq r'_n} \gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \\
&= O \left(\gamma_n \max_{1 \leq t \leq n} c_{nt}^2 \right) = o(1) \quad (\text{A-33})
\end{aligned}$$

under the conditions stated. ■

Lemma A.5 *Under Assumptions 1, 2 and 3,*

$$\lim_{n \rightarrow \infty} \left| E\hat{\Omega}_n - \Omega_n \right| = 0. \quad (\text{A-34})$$

Proof. Let $Y_{tk} = E(X_{nt}|\mathcal{H}_{n,t-k}) - E(X_{nt}|\mathcal{H}_{n,t-k-1})$, and note that

$$X_{nt} = \sum_{k=-\infty}^{\infty} Y_{tk} \quad (\text{A-35})$$

and also that

$$EY_{tk}Y_{t+m,i+m} = 0 \quad (\text{A-36})$$

unless $i = k$. Further, letting

$$\xi_{tk} = \|E(X_{nt}|\mathcal{H}_{n,t-k})\|_2 \quad (\text{A-37})$$

and

$$\zeta_{tk} = \|X_{nt} - E(X_{nt}|\mathcal{H}_{n,t+k})\|_2, \quad (\text{A-38})$$

we have

$$\|Y_{tk}\|_2^2 = \xi_{tk}^2 - \xi_{t,k+1}^2 = \zeta_{t,k-1}^2 - \zeta_{t,k}^2. \quad (\text{A-39})$$

Substituting from De Jong (1997), Lemma 3, we have

$$\begin{aligned} |EX_{nt}X_{n,t+j}| &\leq \sum_{k=-\infty}^0 (\xi_{tk}^2 - \xi_{t,k+1}^2)^{1/2} (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2)^{1/2} \\ &\quad + \sum_{k=1}^{\infty} (\zeta_{t,k-1}^2 - \zeta_{tk}^2)^{1/2} (\zeta_{t+j,k-j-1}^2 - \zeta_{t+j,k-j}^2)^{1/2} \end{aligned} \quad (\text{A-40})$$

where on the assumptions, $\xi_{tk} \leq c_{nt}\psi(k)$ for $k \geq 0$, and $\zeta_{tk} \leq c_{nt}\psi(k+1)$ for $k \geq 0$, where the numbers $\psi(k)$ are from the mixingale definition. Therefore,

$$\begin{aligned} |E\hat{\Omega}_n - \Omega_n| &\leq 2 \left| \sum_{t=1}^n \sum_{j=0}^{n-t} EX_{nt}X_{n,t+j}(1 - k(j/\gamma_n)) \right| \\ &\leq 2T_{n1} + 2T_{n2} \end{aligned} \quad (\text{A-41})$$

where

$$T_{n1} = \sum_{k=-\infty}^0 \sum_{t=1}^n (\xi_{tk}^2 - \xi_{t,k+1}^2)^{1/2} \sum_{j=0}^{n-t} |1 - k(j/\gamma_n)| (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2)^{1/2} \quad (\text{A-42})$$

and

$$T_{n2} = \sum_{k=1}^{\infty} \sum_{t=1}^n (\zeta_{t,k-1}^2 - \zeta_{tk}^2)^{1/2} \sum_{j=0}^{n-t} |1 - k(j/\gamma_n)| (\zeta_{t+j,k-j-1}^2 - \zeta_{t+j,k-j}^2)^{1/2}. \quad (\text{A-43})$$

Take T_{n1} as representative. Define a summable sequence $\{\beta_m\}_{m=0}^{\infty}$ by setting $\beta_0 = 1$ and $\beta_m = m^{-1} \log(m+1)^{-2}$ for $m \geq 1$. Two applications of the Cauchy-Schwarz inequality yield

$$\begin{aligned} T_{n1} &= \sum_{k=-\infty}^0 \sum_{t=1}^n \left((\xi_{tk}^2 - \xi_{t,k+1}^2)^{1/2} \right) \left(\sum_{j=0}^{n-t} |1 - k(j/\gamma_n)| (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2)^{1/2} \right) \\ &\leq \sum_{k=-\infty}^0 \left(\sum_{t=1}^n (\xi_{tk}^2 - \xi_{t,k+1}^2) \right)^{1/2} \end{aligned}$$

$$\begin{aligned}
& \times \left(\sum_{t=1}^n \left(\sum_{j=0}^{n-t} \left(\beta_j^{1/2} |1 - k(j/\gamma_n)| \right) \left(\beta_j^{-1/2} (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2)^{1/2} \right) \right)^2 \right)^{1/2} \\
& \leq \sum_{k=-\infty}^0 \left(\sum_{t=1}^n (\xi_{tk}^2 - \xi_{t,k+1}^2) \right)^{1/2} \\
& \times \left(\sum_{t=1}^n \left(\sum_{j=0}^{n-t} \beta_j^{-1} (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2) \right) \sum_{j=0}^{\infty} \beta_j (1 - k(j/\gamma_n))^2 \right)^{1/2}. \quad (\text{A-44})
\end{aligned}$$

Note that for any array $\{h_{sj}\}$, $\sum_{t=1}^n \sum_{j=0}^{n-t} h_{t+j,j} = \sum_{s=1}^n \sum_{j=0}^{s-1} h_{sj}$, and therefore in particular

$$\begin{aligned}
\sum_{t=1}^n \sum_{j=0}^{n-t} \beta_j^{-1} (\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2) &= \sum_{s=1}^n \sum_{j=0}^{s-1} \beta_j^{-1} (\xi_{s,k+j}^2 - \xi_{s,k+j+1}^2) \\
&\leq \sum_{s=1}^n \left(\sum_{j=0}^{s-2} (\beta_{j+1}^{-1} - \beta_j^{-1}) \xi_{s,k+j+1}^2 + \beta_0^{-1} \xi_{s,k+1}^2 \right) \\
&= O(1) \quad (\text{A-45})
\end{aligned}$$

by assumption. Therefore,

$$\begin{aligned}
T_{n1} &= O \left(\sum_{k=-\infty}^0 \left(\sum_{t=1}^n (\xi_{tk}^2 - \xi_{t,k+1}^2) \right)^{1/2} \left(\sum_{j=0}^{\infty} \beta_j (1 - k(j/\gamma_n))^2 \right)^{1/2} \right) \\
&= o(1) \quad (\text{A-46})
\end{aligned}$$

where it is used that $\lim_{n \rightarrow \infty} \gamma_n^{-1} = 0$ and $k(0) = 1$. This concludes the proof. \blacksquare

Proof of Theorem 2.2

In view of Theorem 2.1, the result follows if $|\hat{\Omega}_n(\hat{\theta}_n) - \hat{\Omega}_n| \xrightarrow{p} 0$. Define the r -vector

$$W_{nt}(\theta) = \kappa_n^{-1} \frac{\partial X_{nt}}{\partial \theta}. \quad (\text{A-47})$$

Also define

$$f_n(\theta_1, \theta_2) = \sum_{t=1}^n \sum_{s=1}^n X_{nt}(\theta_1) X_{ns}(\theta_2)' k((t-s)/\gamma_n)$$

such that $\hat{\Omega}_n(\theta) = f_n(\theta, \theta)$. First-order Taylor's expansions around θ_0 , successively of f_n with respect to θ_1 , and then of f_n and $\partial f_n / \partial \theta_1$ with respect to θ_2 , yield

$$\begin{aligned}
\hat{\Omega}_n(\hat{\theta}_n) - \hat{\Omega}_n &= 2 \sum_{t=1}^n \sum_{s=1}^n (\hat{\theta}_n - \theta_0)' \kappa_n \tilde{W}_{nt} X_{ns} k \left(\frac{t-s}{\gamma_n} \right) \\
&\quad + \sum_{t=1}^n \sum_{s=1}^n (\hat{\theta}_n - \theta_0)' \kappa_n \tilde{W}_{ns} \tilde{W}'_{nt} \kappa_n (\hat{\theta}_n - \theta_0) k \left(\frac{t-s}{\gamma_n} \right) \quad (\text{A-48})
\end{aligned}$$

where \tilde{W}_{nt} represents W_{nt} with its elements evaluated at unspecified points (in general different in each term) on the line segment joining θ_0 and $\hat{\theta}_n$. The last term here converges in probability

to zero because $n^{1/2}\kappa_n(\hat{\theta}_n - \theta_0) = O_p(1)$ by assumption, and for n large enough,

$$\begin{aligned} \left\| n^{-1} \sum_{t=1}^n \sum_{s=1}^n \lambda' \tilde{W}_{ns} \tilde{W}'_{nt} \lambda k \left(\frac{t-s}{\gamma_n} \right) \right\|_1 &\leq n^{-1} \sum_{j=-n+1}^{n-1} \left| k \left(\frac{j}{\gamma_n} \right) \right| \sum_{t=1}^n \left\| \sup_{\theta \in \mathcal{N}} |\lambda' W_{nt}(\theta)| \right\|_2^2 \\ &= O \left(n^{-1} \sum_{j=-n+1}^{n-1} |k(j/\gamma_n)| \right) \\ &= O(\gamma_n/n) = o(1) \end{aligned} \quad (\text{A-49})$$

for any $\lambda \neq 0$, by assumption. Note that consistency of $\hat{\theta}_n$ implies that the inequality holds for any \mathcal{N} , with n large enough.

To show that the first term of (A-48) also converges in probability to zero, suppose first that Assumption 4(i) holds. Letting W_{nt} denote $W_{nt}(\theta_0)$, note that

$$\begin{aligned} \sum_{t=1}^n \sum_{s=1}^n (\hat{\theta} - \theta_0)' \kappa_n \tilde{W}_{nt} X_{ns} k \left(\frac{t-s}{\gamma_n} \right) &= \sum_{t=1}^n \sum_{s=1}^n (\hat{\theta} - \theta_0)' \kappa_n (\tilde{W}_{nt} - EW_{nt}) X_{ns} k \left(\frac{t-s}{\gamma_n} \right) \\ &\quad + \sum_{t=1}^n \sum_{s=1}^n (\hat{\theta} - \theta_0)' \kappa_n EW_{nt} X_{ns} k \left(\frac{t-s}{\gamma_n} \right). \end{aligned} \quad (\text{A-50})$$

The first term of (A-50) converges in probability to zero because $n^{1/2}\kappa_n(\hat{\theta} - \theta_0) = O_p(1)$ and

$$\begin{aligned} &\left\| n^{-1/2} \sum_{t=1}^n \sum_{s=1}^n \lambda' (\tilde{W}_{nt} - EW_{nt}) X_{ns} k \left(\frac{t-s}{\gamma_n} \right) \right\|_1 \\ &\leq \int_{-\infty}^{\infty} \left\| \left(\sum_{s=1}^n X_{ns} e^{-i\xi s/\gamma_n} \right) \left(n^{-1/2} \sum_{t=1}^n \lambda' (\tilde{W}_{nt} - EW_{nt}) e^{i\xi t/\gamma_n} \right) \right\|_1 |\psi(\xi)| d\xi \\ &\leq \int_{-\infty}^{\infty} \left\| \sum_{t=1}^n X_{nt} e^{i\xi t/\gamma_n} \right\|_2 \left(\left\| \sup_{\theta \in \mathcal{N}} \left| n^{-1/2} \sum_{t=1}^n \lambda' (W_{nt} - EW_{nt}) e^{i\xi t/\gamma_n} \right| \right\|_2 \right. \\ &\quad \left. + \left\| n^{-1/2} \sum_{t=1}^n \lambda' (E\tilde{W}_{nt} - EW_{nt}) e^{i\xi t/\gamma_n} \right\|_2 \right) |\psi(\xi)| d\xi \\ &= o(1) \end{aligned} \quad (\text{A-51})$$

for any $\lambda \neq 0$, by assumption. Also,

$$\left(\sum_{t=1}^n \sum_{s=1}^n (\hat{\theta}_n - \theta_0)' \kappa_n EW_{nt} X_{ns} k \left(\frac{t-s}{\gamma_n} \right) \right)^2 = n(\hat{\theta}_n - \theta_0)' \kappa_n M_n \kappa_n (\hat{\theta}_n - \theta_0) \quad (\text{A-52})$$

where

$$M_n = n^{-1} \sum_{t=1}^n \sum_{s=1}^n \sum_{i=1}^n \sum_{j=1}^n EW_{nt} EW'_{nj} X_{ns} X_{ni} k \left(\frac{t-s}{\gamma_n} \right) k \left(\frac{i-j}{\gamma_n} \right). \quad (\text{A-53})$$

Since M_n is a positive semidefinite matrix, the result now follows if we can show that $\lambda' M_n \lambda \rightarrow 0$ for any $\lambda \neq 0$. To prove this, first note that from the proof of Lemma A.5,

$$\left| \sum_{i=1}^n \sum_{s=1}^n EX_{ni} X_{ns} a_{nij} a_{nst} \right| = O \left(\left(\sum_{i=1}^n \hat{c}_{ni}^2 a_{nij}^2 \sum_{s=1}^n \hat{c}_{ns}^2 a_{nst}^2 \right)^{1/2} \right) \quad (\text{A-54})$$

for L_2 -mixingale random variables X_{nt} of size $-1/2$, and any positive deterministic array a_{nst} . Therefore, under the assumptions,

$$\begin{aligned}
\lambda' M_n \lambda &= C \sum_{t=1}^n \sum_{j=1}^n |\lambda' EW_{nt}| |\lambda' EW_{nj}| \left(n^{-2} \sum_{i=1}^n c_{ni}^2 k \left(\frac{i-j}{\gamma_n} \right)^2 \sum_{s=1}^n c_{ns}^2 k \left(\frac{t-s}{\gamma_n} \right)^2 \right)^{1/2} \\
&= C \left(\sum_{t=1}^n |\lambda' EW_{nt}| \left(n^{-1} \sum_{s=1}^n c_{ns}^2 k \left(\frac{t-s}{\gamma_n} \right)^2 \right)^{1/2} \right)^2 \\
&\leq C \left(\sum_{t=1}^n |\lambda' EW_{nt}|^2 \right) \left(n^{-1} \sum_{s=1}^n c_{ns}^2 \sum_{t=1}^n k \left(\frac{t-s}{\gamma_n} \right)^2 \right) \\
&= O(\gamma_n/n) = o(1),
\end{aligned} \tag{A-55}$$

for $C > 0$, where the second inequality follows from the Cauchy-Schwarz inequality.

Now suppose that Assumption 4(ii) holds. In this case we may argue similarly to Hansen's (1992, p. 971) inequality (A4), except that we replace Hansen's $\delta_T X_t$ by \tilde{W}_{nt} , to argue that

$$\sum_{t=1}^n \sum_{s=1}^n (\hat{\theta}_n - \theta_0)' \kappa_n \tilde{W}_{nt} X_{ns} k \left(\frac{t-s}{\gamma_n} \right) = O_p(n^{-1/2} \gamma_n) = o_p(1). \tag{A-56}$$

This concludes the proof. ■

Proof of Theorem 2.3

By assumption, $\hat{\alpha}$ can be made an element of the interval $[\varepsilon, 1/\varepsilon]$ for some small $\varepsilon \in (0, 1)$ with arbitrarily large probability. Furthermore it is well-known (see e.g. Newey (1991)) that

$$\sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \left| \hat{\Omega}_n(\hat{\theta}_n, \alpha \gamma_n) - \Omega_n \right| \xrightarrow{p} 0 \tag{A-57}$$

if $|\hat{\Omega}_n(\hat{\theta}_n, \alpha \gamma_n) - \Omega_n|$ converges to zero pointwise in α and $\hat{\Omega}_n(\hat{\theta}_n, \alpha \gamma_n) - \Omega_n$ is stochastically equicontinuous on $[\varepsilon, 1/\varepsilon]$. Pointwise convergence follows from the results that were established earlier. Stochastic equicontinuity follows if both of the terms $\hat{\Omega}_n(\theta_0, \alpha \gamma_n) - \Omega_n$ and $\hat{\Omega}_n(\hat{\theta}_n, \alpha \gamma_n) - \hat{\Omega}_n(\theta_0, \alpha \gamma_n)$ are stochastically equicontinuous on $[\varepsilon, 1/\varepsilon]$. The first of these results follows because

$$\begin{aligned}
\hat{\Omega}_n(\theta_0, \alpha \gamma_n) &= \int_{-\infty}^{\infty} \left(\left(\sum_{t=1}^n X_{nt} \sin(t\xi/\alpha \gamma_n) \right)^2 + \left(\sum_{t=1}^n X_{nt} \cos(t\xi/\alpha \gamma_n) \right)^2 \right) \psi(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \left(\left(\sum_{t=1}^n X_{nt} \sin(t\phi/\gamma_n) \right)^2 + \left(\sum_{t=1}^n X_{nt} \cos(t\phi/\gamma_n) \right)^2 \right) \psi(\alpha\phi) \alpha d\phi
\end{aligned} \tag{A-58}$$

so

$$\begin{aligned}
&\sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{\alpha' \in [\varepsilon, 1/\varepsilon]: |\alpha - \alpha'| < \eta} \left| \hat{\Omega}_n(\theta_0, \alpha \gamma_n) - \hat{\Omega}_n(\theta_0, \alpha' \gamma_n) \right| \\
&\leq \int_{-\infty}^{\infty} \left(E \left(\sum_{t=1}^n X_{nt} \sin(t\phi/\gamma_n) \right)^2 + E \left(\sum_{t=1}^n X_{nt} \cos(t\phi/\gamma_n) \right)^2 \right) \\
&\quad \times \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{\alpha' \in [\varepsilon, 1/\varepsilon]: |\alpha - \alpha'| < \eta} |\psi(\alpha\phi)\alpha - \psi(\alpha'\phi)\alpha'| d\phi.
\end{aligned} \tag{A-59}$$

Therefore, using Lemma A.1 again, it is easily seen that

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} E \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{\alpha' \in [\varepsilon, 1/\varepsilon]; |\alpha - \alpha'| < \eta} \left| \hat{\Omega}_n(\theta_0, \alpha \gamma_n) - \hat{\Omega}_n(\theta_0, \alpha' \gamma_n) \right| = 0, \quad (\text{A-60})$$

which implies stochastic equicontinuity of $\hat{\Omega}_n(\theta_0, \alpha \gamma_n) - \Omega_n$. To show the second result, consider equation (A-48). By the reasoning leading up to equation (A-49), it is easily shown that

$$\begin{aligned} \left\| \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \left| n^{-1} \sum_{t=1}^n \sum_{s=1}^n \tilde{W}_{nt} \tilde{W}'_{nt} k \left(\frac{t-s}{\alpha \gamma_n} \right) \right| \right\|_1 &= O \left(n^{-1} \sum_{j=-n+1}^{n-1} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \left| k \left(\frac{j}{\alpha \gamma_n} \right) \right| \right) \\ &= O(\gamma_n/n) = o(1) \end{aligned} \quad (\text{A-61})$$

by assumption, where W_{nt} is defined as before. Therefore, for proving stochastic equicontinuity of $\hat{\Omega}_n(\hat{\theta}_n, \alpha \gamma_n) - \hat{\Omega}_n(\theta_0, \alpha \gamma_n)$ it is sufficient since $\kappa_n(\hat{\theta}_n - \theta_0) = O_p(n^{-1/2})$ to show that

$$T_n(\alpha) = \sum_{t=1}^n \sum_{s=1}^n \lambda' \tilde{W}_{nt} X_{ns} k((t-s)/\alpha \gamma_n) \quad (\text{A-62})$$

is stochastically equicontinuous, where λ is any nonzero r -vector. This is shown by noting that

$$T_n(\alpha) = \int_{-\infty}^{\infty} \left(\sum_{s=1}^n X_{nt} e^{-is\xi/\gamma_n} \right) \left(\sum_{t=1}^n \lambda' \tilde{W}_{nt} e^{it\xi/\gamma_n} \right) \alpha \psi(\alpha \xi) d\xi, \quad (\text{A-63})$$

and therefore for some $C > 0$ and n large enough,

$$\begin{aligned} E \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{|\alpha - \alpha'| < \eta} |T_n(\alpha) - T_n(\alpha')| &\leq \int_{-\infty}^{\infty} \left\| \sum_{t=1}^n X_{nt} e^{it\xi/\gamma_n} \right\|_2 \left\| \sum_{t=1}^n \sup_{\theta \in \mathcal{N}} |\lambda' W_{nt}(\theta)| e^{it\xi/\gamma_n} \right\|_2 \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{|\alpha - \alpha'| < \eta} |\alpha \psi(\alpha \xi) - \alpha' \psi(\alpha' \xi)| d\xi \\ &\leq C \int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} \sup_{|\alpha - \alpha'| < \eta} |\alpha \psi(\alpha \xi) - \alpha' \psi(\alpha' \xi)| d\xi \rightarrow 0 \end{aligned} \quad (\text{A-64})$$

as $\eta \rightarrow 0$. ■

References

- [1] Andrews, D.W.K. (1991): "Heteroscedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.
- [2] Andrews, D.W.K. and J.C. Monahan (1992): "An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953-966 .
- [3] Davidson, J. (1992): "A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes," *Econometric Theory*, 8, 313-334.
- [4] Davidson, J. (1993): "The central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case," *Econometric Theory*, 9, 402-412.
- [5] Davidson, J. (1994): *Stochastic Limit Theory*. Oxford: Oxford University Press.

- [6] Doukhan, P., Massart, P. and E. Rio (1994): "The functional central limit theorem for strongly mixing processes," *Annales de l'Institut Henri Poincaré, Probabilités et statistiques*, 10, 63-82.
- [7] Gallant, A.R. and H. White (1988): *A Unified Theory of Estimation and Inference for Non-linear Dynamic Models*. New York: Basil Blackwell.
- [8] Hall, P. and C.C. Heyde (1980): *Martingale Limit Theory and its Applications*. New York: Academic Press.
- [9] Hansen, B.E. (1992): "Consistent covariance matrix estimation for dependent heterogeneous processes," *Econometrica*, 60, 967-972.
- [10] de Jong, R.M. (1997): "Central limit theorems for dependent heterogeneous random variables" *Econometric Theory* 13,3, 353-367.
- [11] de Jong, R.M. (1999) "A strong consistency proof for heteroscedasticity and autocorrelation consistent covariance matrix estimators", Working paper.
- [12] Kool, H. (1988): "A note on consistent estimation of heteroscedastic and autocorrelated covariance matrices," Research memorandum 1988-21, Department of Econometrics, Free University Amsterdam
- [13] McLeish, D.L. (1975a) "A maximal inequality and dependent strong laws". *Annals of Probability* 3, 829-839.
- [14] McLeish, D.L. (1975b): "Invariance principles for dependent variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 33, 165-178.
- [15] McLeish, D.L. (1977): "On the invariance principle for nonstationary mixingales," *Annals of Probability*, 5, 616-621.
- [16] Newey, W.K. (1991): "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 59, 1161-1167.
- [17] Newey, W.K. and K.D. West (1987): "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703-708.
- [18] Newey, W.K. and K.D. West (1994): "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61, 631-654.
- [19] Phillips, P. C. B., and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-125.
- [20] Phillips, P. C. B., and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika*, 75, 335-46.
- [21] Pötscher, B.M. and I.R. Prucha (1991a): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part 1: consistency and approximation concepts," *Econometric Reviews*, 10, 125-216.
- [22] Pötscher, B.M. and I.R. Prucha (1991b): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part 2: asymptotic normality," *Econometric Reviews*, 10, 253-325.
- [23] White, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.