Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices

by Robert M. de Jong* and James Davidson

May 1996

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 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). Interest in it is motivated typically by the fact that many estimators $\hat{\theta}_n$ of a parameter θ_0 are known to satisfy

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

where $X_{nt}(\theta)$ is a random vector of dimension p, defined on a probability space $(\Omega,\mathcal{F},P)$, that has mean zero at the point $\theta=\theta_0$, and B_n is some nonrandom matrix of dimension $r\times p$ that is usually easily estimated. [1] Applying a central limit theorem to the second of the terms in (1) leads to

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

*The first author thanks Herman Bierens for discussion that eventually led to the proof of Lemma 1.

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

where

$$\Omega_n = \sum_{t=1}^{n} \sum_{s=1}^{n} E X_{nt} X'_{ns} \tag{3}$$

where we defined $X_{nt} = X_{nt}(\theta_0)$. 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 distributed. However, an undesirable feature of all the above-cited studies is that they impose conditions stronger than are known to be required for the application of a central limit theorem (CLT) to the same variables. All except the last-mentioned assume that the random variables under consideration possess finite fourth moments, and all impose either a form of stationarity, or uniform boundedness in L_p-norms for some $p \geq 2$, precluding the possibility of trending moments. Further, all except Pötscher and Prucha (1991b) assume that the random variables considered are strong or uniform mixing and that the true covariance matrix converges to some well-defined limit.

In this paper, we will bridge the gap between asymptotic normality and covariance matrix estimation by obtaining conditions for the latter similar to those obtained for the CLTs in Davidson(1992, 1993) and De Jong (1995), which are the best such results currently known to us. These theorems, which develop techniques pioneered by McLeish (1975,1977), permit globally nonstationary data processes and require the existence only of $2 + \delta$-order moments for some $\delta > 0$. The weak dependence is characterized by near-epoch dependence on a mixing process, a more general concept than strong or uniform mixing; for example, under general regularity conditions ARMA processes are near epoch dependent, but need not satisfy the strong mixing condition. We will prove our results for stochastic (sample-dependent) bandwidths for the kernel estimators, and also show that 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, and there is no need to assume that the true covariance matrix itself converges to a well-defined limit. Finally, 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.

Newey-West type estimators also occur in the presence of unit roots, as in the variance estimation for Phillips or Phillips-Perron tests for example, and we will also state results for such cases.

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

2 Main results

The main consistency results of the paper are, in fact, inspired by the proofs of the CLTs for possibly trending-variance processes given in Davidson (1992), Davidson (1993), and De Jong (1995), in which showing the consistency of a certain variance estimate is an essential step. However, the role of these results is relatively obscure, and the statistics considered in those papers do not allow for an easy interpretation. Moreover, the conditions given are those for the CLTs to hold, and stronger in some respects than the conditions required for convergence of the variance estimates alone. We therefore follow the approach of Andrews (1991) and Hansen (1992). Similarly to the latter authors, define

$$\hat{\Omega}_n = \sum_{j=-n+1}^{n-1} k(j/\gamma_n)\hat{\Gamma}_n(j), \tag{4}$$

$$\hat{\Gamma}_n(j) = \sum_{t=1}^{n-j} X_{nt}X'_{n,t+j} \qquad j \geq 0, \tag{5}$$

$$\hat{\Gamma}_n(j) = \hat{\Gamma}_n(-j)' \qquad j < 0. \tag{6}$$

The function $k(.)$ is called the kernel function, and the sequence γ_n is called the bandwidth or the lag truncation parameter. It is assumed that $\gamma_n \to \infty$ as $n \to \infty$. Note that (5) adopts an array notation which allows us to generalize our results, but direct comparability with the results of Andrews and Hansen is obtained by considering the case $X_{nt} = n^{-1/2}X_t$. In this case, (5) becomes

$$\hat{\Gamma}_n(j) = n^{-1}\sum_{t=1}^{n-j} X_t X'_{t+j} \qquad j \geq 0. \tag{7}$$

The variance estimator of Newey and West (1987) can be obtained by choosing $k(x) = (1-|x|)I(-1 < x \leq 1)$ (the Bartlett kernel). For that case, Newey and West have proven that a consistent covariance matrix estimator results under regularity conditions if $\gamma_n = o(n^{1/4})$. Pötscher and Prucha (1991b) require $\gamma_n = o(n^{1/3})$. Kool (1988) and Hansen (1992) have shown that it is sufficient that $\gamma_n = o(n^{1/2})$ under regularity conditions, while the results of Andrews (1991) imply that we can choose $\gamma_n = o(n)$. From Andrews (1991), however, it can be seen that choosing γ_n such that $\gamma_n = o(n)$ but not $o(n^{1/2})$ can never be optimal under a mean squared error criterion function. Alternatively, we can write

$$\hat{\Omega}_n = \sum_{t=1}^{n}\sum_{s=1}^{n} X_{nt}X'_{ns}k((t-s)/\gamma_n). \tag{8}$$

Finally, define

$$\hat{\Omega}_n(\hat{\theta}_n) = \sum_{t=1}^{n}\sum_{s=1}^{n} X_{nt}(\hat{\theta}_n)X_{ns}(\hat{\theta}_n)'k((t-s)/\gamma_n), \tag{9}$$

which is our operational estimator for Ω_n. The representations of (8) and (9) illustrate the idea behind the estimator. While we cannot set $k(.) = 1$ because that would introduce too much variance into the estimator, we require the weights $k((t-s)/\gamma_n)$ to approach unity as $n \to \infty$ for each fixed value of $t-s$. We assume that the kernel function $k(.)$ is an element of the function class $\mathcal{K}$:

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

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

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

$$k(.) \textit{ is continuous at 0 and at all but a finite number of points}\}\,, \tag{10}$$

where

$$\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x) \exp(-i\xi x) dx. \tag{11}$$

In Andrews (1991), function classes $\mathcal{K}_1$ and $\mathcal{K}_2$ are defined. Our definition of $\mathcal{K}$ is identical to Andrews' definition of $\mathcal{K}_2$, except that Andrews' condition that $k(.) \in L_2(-\infty,\infty)$ is replaced by the condition that $k(.) \in L_1(-\infty,\infty)$, and the requirement that $\psi(\xi) \geq 0$ from Andrews (1991) is replaced by the requirement that $\int_{-\infty}^{\infty} |\psi(\xi)|d\xi < \infty$. The integrability condition on $\psi(.)$ that we impose is weaker than Andrews' requirement. This follows because for all functions $k(.)$ that satisfy the conditions for $\mathcal{K}$ except for the integrability condition on $\psi(.)$, $\int_{-\infty}^{\infty} \psi(\xi)d\xi = k(0) = 1$ (see the proof of Theorem 2 of Andrews (1991)), and therefore $\psi(\xi) \geq 0$ implies that $\int_{-\infty}^{\infty} |\psi(\xi)|d\xi < \infty$. As Andrews (1991) notes, his function class $\mathcal{K}_2$ corresponds to the function class for which $\hat{\Omega}_n$ and $\hat{\Omega}_n(\hat{\theta}_n)$ necessarily are positive semidefinite matrices with probability one. It is clear that this property is desirable. In view of this, the fact that the function class $\mathcal{K}$ does not contain the truncated kernel (see Andrews (1991) for definitions) does not seem an important restriction of the analysis that is provided here. The Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernels are included in $\mathcal{K}$. Note that the Tukey-Hanning kernel is not included in Andrews' (1991) $\mathcal{K}_2$ class, while the truncated kernel is an element of Andrews' $\mathcal{K}_1$ class. Again the reader is referred to Andrews (1991) for definitions of these kernel functions.
The concept of weak dependence that we employ is that of near epoch dependence. Let V_{nt} denote a triangular array of random variables. The L_q-norm of a random matrix X in this paper will be defined as $\| X \|_q = (\sum_i \sum_j E|X_{ij}|^q)^{1/q}$ for $q \geq 1$.

Definition 1 *A triangular array of random variables X_{nt} is called L_2-near epoch dependent on an array V_{nt} if for $m \geq 0$*

$$\| X_{nt} - E(X_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2 \leq d_{nt}\nu(m) \tag{12}$$

where $\mathcal{V}_{n,t-m}^{t+m} = \sigma(\{V_{n,t-m}, \ldots, V_{n,t+m}\})$, and $\nu(m) \to 0$ as $m \to \infty$.

The reader is referred to Gallant and White (1988), Pötscher and Prucha (1991a, 1991b) or Davidson (1994) for more details about the concept of near epoch dependence. Furthermore, we will say that the sequence $\nu(m)$ is of *size* $-\lambda$ if $\psi(m) = O(m^{-\lambda-\varepsilon})$ for some $\varepsilon > 0$. For the proof of our main result, we need the following assumption.

Assumption 2 *X_{nt} is L_2-near epoch dependent on V_{nt}, where V_{nt} are strong or uniform mixing random variables; for some triangular array c_{nt} we have*

$$\sup_{n\geq 1}\sup_{1\leq t\leq n}(\| X_{nt}\|_r + d_{nt})/c_{nt} < \infty \tag{13}$$

for some $r > 2$ and $\nu(m)$ is of size -1/2 and either $\alpha(m)$ is of size $-r/(r-2)$, or $\phi(m)$ is of size $-r/(2(r-1))$, and

$$\sup_{n\geq 1}\sum_{t=1}^{n} c_{nt}^2 < \infty. \tag{14}$$

Also note that from the proof it can be seen that in the ϕ-mixing case it is allowed that we set $r = 2$, but in that case we have to assume uniform integrability of X_{nt}^2/c_{nt}^2 in addition. These dependence conditions match those of the best-known central limit theorem in both the ϕ-mixing and α-mixing cases. Also, from the discussion in De Jong (1995) and the covariance inequalities in Doukhan, Massart and Rio (1994), it can be shown that relaxing the size requirements on the $\alpha(m)$ or $\nu(m)$ sequences of the X_{nt} implies that X_{nt} need no longer be covariance summable. Considering the standard case in which $X_{nt} = n^{-1/2}X_t$ where X_t is a stationary sequence, and $B_n = B$, this would imply that $n^{-1}E(\sum_{t=1}^{n} X_t)(\sum_{t=1}^{n} X_t') \to \infty$ as $n \to \infty$, implying that the result of Equation (1) would be incompatible with root-n consistency of $\hat{\theta}_n$. For such applications, our results are in effect the best possible with respect to the size conditions on the $\alpha(m)$ and $\nu(m)$ sequences.
Our assumption on the behavior of the bandwidth is as follows:

Assumption 3

$$\lim_{n\to\infty}(\gamma_n^{-1} + \gamma_n \max_{1\leq t\leq n} c_{nt}^2) = 0. \tag{15}$$

The following assumption is needed in order to show that $\hat{\Omega}_n(\hat{\theta}_n)$ is asymptotically equivalent to $\hat{\Omega}_n$.

Assumption 4 *For each deterministic triangular array a_{nt} such that $0 \leq a_{nt} \leq 1$ for all $t, n \geq 1$, and for all j, $j = 1, \ldots, r$, we have*

$$\sup_{\theta\in\Theta}|n^{-1/2}\sum_{t=1}^{n} a_{nt}\left((\partial/\partial\theta_j)X_{nt}(\theta) - E(\partial/\partial\theta_j)X_{nt}(\theta)\right)| \xrightarrow{p} 0, \tag{16}$$

$$\limsup_{n\to\infty}\sum_{t=1}^{n} E\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)X_{nt}(\theta)|^2 < \infty, \tag{17}$$

$$n^{1/2}(\hat{\theta}_n - \theta_0) = O_P(1), \tag{18}$$

$$n^{-1/2} \sum_{t=1}^{n} E(\partial/\partial\theta_j) X_{nt}(\theta) \quad \textit{is continuous at } \theta_0 \textit{ uniformly in } n. \tag{19}$$

Note that, although the above uniform convergence requirement of Equation (16) is nonstandard, proofs of uniform laws of large numbers are usually not affected by the presence of the a_{nt}. Therefore, usually the requirement of Equation (16) will hold if it holds for the case $a_{nt} = 1$. For the case of covariance matrix estimation for minimization estimators, the resulting condition then is usually proven as a part of the asymptotic normality proof of the minimization estimator.
The following alternative assumption is similar to one that has been introduced in Hansen (1992) and can be found in Andrews and Monahan (1992) as well. It is used for nonlinear dynamic models with deterministic or stochastic trends.

Assumption 5 $\lim_{n\to\infty} \gamma_n n^{-1/2} = 0$, *and for some sequence* δ_n *of nonsingular matrices and for random variables* W_{nt} *we have*

$$X_{nt}(\theta) = X_{nt}(\theta_0) - (\theta - \theta_0) W_{nt}; \tag{20}$$

$$\delta_n \sum_{t=1}^{n} W_{nt} W'_{nt} \delta'_n = O_P(1); \tag{21}$$

$$n^{1/2}(\hat{\theta}_n - \theta_0)\delta_n^{-1} = O_P(1). \tag{22}$$

We will state three lemmas that provide the tools for showing consistency of $\hat{\Omega}_n$. The first shows the asymptotic equivalence of $\hat{\Omega}_n$ to its expectation.

Lemma 1 *Under Assumptions 1, 2 and 3,*

$$\hat{\Omega}_n - E\hat{\Omega}_n \xrightarrow{p} 0. \tag{23}$$

The asymptotic bias of our covariance matrix estimator can be shown to disappear as well:

Lemma 2 *Under Assumptions 1, 2 and 3,*

$$\lim_{n\to\infty} (E\hat{\Omega}_n - \Omega_n) = 0. \tag{24}$$

The third lemma states that the effect of estimation of θ_0 is asymptotically negligible under regularity conditions:

Lemma 3 *Under Assumptions 1, 2 and 3, and either Assumption 4 or Assumption 5,*

$$\hat{\Omega}_n - \hat{\Omega}_n(\hat{\theta}_n) \xrightarrow{p} 0. \tag{25}$$

The main result that follows from Lemma 1, 2 and 3 is the following:

Theorem 1 *Under Assumptions 1, 2 and 3, and either Assumption 4 or Assumption 5,*

$$\hat{\Omega}_n(\hat{\theta}_n) - \Omega_n \xrightarrow{p} 0. \tag{26}$$

Finally, we establish a result that allows bandwidths to be stochastic. See Andrews and Monahan (1992) and Newey and West (1994) for such procedures. 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 *Assume that Assumptions 1, 2 and 3, and either Assumption 4 or Assumption 5 hold. In addition, assume that $\hat{\delta}_n = \hat{\alpha}_n \gamma_n$, where $\hat{\alpha}_n = O_P(1)$ and $1/\hat{\alpha}_n = O_P(1)$. Moreover, assume that for all $\varepsilon \in (0,1)$ the kernel function $k(.)$ satisfies*

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |\psi(\alpha\xi)| d\xi < \infty \tag{27}$$

and

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |k(\alpha x)| dx < \infty, \tag{28}$$

and γ_n satisfies the bandwidth conditions of the assumptions. Then

$$\hat{\Omega}_n(\hat{\theta}_n, \hat{\delta}_n) - \Omega_n \xrightarrow{p} 0. \tag{29}$$

Finally, note that the conditions on $k(.)$ and $\psi(.)$ that are imposed in Theorem 2 are satisfied for the Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernels.

Department of Econometrics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

and

Department of Economics, Cardiff Business School, Aberconway Building, Colum Drive, Cardiff CF1 3EU, U.K.

Appendix

For simplicity, in the proofs and the lemmas that follow, we assume that $X_{nt}(\theta)$ is real-valued, i.e. $p = 1$. From the reasoning as in Newey and West (1987), it follows that convergence in probability then has to hold for the case of vector-valued $X_{nt}(\theta)$ as well. In what follows, we will need the *mixingale* concept. Let $\mathcal{H}_{nt}$ denote an array of σ-fields that is nondecreasing in t for each n. Mixingales are defined as follows:

Definition 2 *$\{X_{nt}, \mathcal{H}_{nt}\}$ is called an L_2-mixingale if for $m \geq 0$*

$$\| X_{nt} - E(X_{nt}|\mathcal{H}_{n,t+m}) \|_2 \leq c_{nt}\psi(m+1), \tag{30}$$

$$\| E(X_{nt}|\mathcal{H}_{n,t-m}) \|_2 \leq c_{nt}\psi(m), \tag{31}$$

and $\psi(m) \to 0$ as $m \to \infty$.

The c_{nt} are usually referred to as the mixingale magnitude indices, and we will refer to $\{X_{nt}, \mathcal{H}_{nt}\}$ as a mixingale of szie -1/2 if the associated $\psi(m)$ sequence is of size -1/2. Also, note that Assumption 2 implies that $\{X_{nt}, \mathcal{H}_{nt}\}$ is a mixingale of size -1/2 and mixingale magnitude indices c_{nt} for $\mathcal{H}_{nt} = \sigma(\{V_{nt}, V_{n,t-1}, \ldots\})$ by Theorem 17.5 of Davidson (1994), and in this Appendix $\mathcal{H}_{nt}$ will denote this sigma field. Before proving our main results, we will state the following result (see e.g. Lemma 2.1 of Hall and Heyde(1980)):

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

Proof of Lemma 1:

Define

$$b_n = [\gamma_n/\delta] \text{ and } r_n = [n/b_n], \tag{32}$$

$$\eta_\delta(x) = (\delta^2 2\pi)^{-1/2} \exp(-x^2\delta^{-2}2^{-1}), \tag{33}$$

$$\Omega_{1n\delta} = \sum_{t=-n+1}^{2n} (\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,n]}(|l|)) \times$$

$$(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n)), \tag{34}$$

$$\Omega_{2n\delta} = \sum_{t=-n+1}^{2n} (\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,b_n]}(|l|) 1_{[0,n]}(|l|)) \times$$

$$(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n)), \tag{35}$$

and

$$\Omega_{3n\delta} = \sum_{t=-n+1}^{2n} (\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,b_n]}(|l|)) \times$$

$$(\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) 1_{[0,b_n]}(|j|)). \tag{36}$$

Lemma 1 will be proven by noting that

$$\begin{aligned}\| \hat{\Omega}_n - E\hat{\Omega}_n \|_1 &\leq \| \hat{\Omega}_n - \Omega_{1n\delta} \|_1 + \| \Omega_{1n\delta} - \Omega_{2n\delta} \|_1 + \| \Omega_{2n\delta} - \Omega_{3n\delta} \|_1 \\ &+ \| \Omega_{3n\delta} - E\Omega_{3n\delta} \|_1 + \| E\Omega_{3n\delta} - E\Omega_{2n\delta} \|_1 \\ &+ \| E\Omega_{2n\delta} - E\Omega_{1n\delta} \|_1 + \| E\Omega_{1n\delta} - E\hat{\Omega}_n \|_1 \\ &\leq 2 \| \hat{\Omega}_n - \Omega_{1n\delta} \|_1 + 2 \| \Omega_{1n\delta} - \Omega_{2n\delta} \|_1 \\ &+ 2 \| \Omega_{2n\delta} - \Omega_{3n\delta} \|_1 + \| \Omega_{3n\delta} - E\Omega_{3n\delta} \|_1 . \end{aligned} \tag{37}$$

The lemmas that follow show that each of the four terms on the right-hand side of the last equation vanish if we first take the 'limsup' as n approaches infinity and then take the limit as δ approaches zero. ▮

Lemma A. 2 *Under the conditions of Lemma 1,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \| \hat{\Omega}_n - \Omega_{1n\delta} \|_1 = 0. \tag{38}$$

Proof:

Firstly note that under the conditions stated,

$$k(x) = \int_{-\infty}^{\infty} \exp(i\xi x)\psi(\xi) d\xi \tag{39}$$

by the inversion formula for Fourier transforms. Therefore, using the fact that $\psi(.)$ is an even function (which is a consequence of the fact that $k(.)$ is an even function), we can write

$$\begin{aligned}\hat{\Omega}_n &= \int_{-\infty}^{\infty} \sum_{t=1}^{n} \sum_{s=1}^{n} X_{nt} X_{ns} \exp(i\xi(t-s)/\gamma_n)\psi(\xi) d\xi \\ &= \int_{-\infty}^{\infty} \sum_{t=1}^{n} \sum_{s=1}^{n} X_{nt} X_{ns} \cos(\xi(t-s)/\gamma_n)\psi(\xi) d\xi \\ &= \int_{-\infty}^{\infty} \left((\sum_{t=1}^{n} X_{nt} \cos(t\xi/\gamma_n))^2 + (\sum_{t=1}^{n} X_{nt} \sin(t\xi/\gamma_n))^2 \right) \psi(\xi) d\xi. \end{aligned} \tag{40}$$

Next, note that we can rewrite the expression for $\Omega_{1n\delta}$ as

$$\gamma_n^{-1} \sum_{s=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} X_{n,s+j} X_{n,s+l} k(l/\gamma_n) \eta_\delta(j/\gamma_n) 1_{[0,n]}(|l|) 1_{[1,n]}(s+l) 1_{[1,n]}(s+j)$$

$$= \gamma_n^{-1} \sum_{t=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} X_{nt} X_{n,t+l-j} k(l/\gamma_n) \eta_\delta(j/\gamma_n) 1_{[0,n]}(|l|) 1_{[1,n]}(t+l-j) 1_{[1,n]}(t)$$

$$= \gamma_n^{-1} \sum_{t=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} X_{nt} X_{ns} k(l/\gamma_n) \eta_\delta((t-s)/\gamma_n + l/\gamma_n) 1_{[0,n]}(|l|) 1_{[1,n]}(s) 1_{[1,n]}(t)$$

$$= \sum_{t=1}^{n} \sum_{s=1}^{n} X_{nt} X_{ns} k_{n\delta}((t-s)/\gamma_n), \tag{41}$$

where

$$k_{n\delta}(x) = \gamma_n^{-1} \sum_{l=-n}^{n} k(l/\gamma_n) \eta_\delta(x + l/\gamma_n). \tag{42}$$

The inverse Fourier transform of $k_{n\delta}(.)$ equals

$$(2\pi)^{-1} \int_{-\infty}^{\infty} \exp(-i\xi x) \gamma_n^{-1} \sum_{l=-n}^{n} k(l/\gamma_n) \eta_\delta(x + l/\gamma_n) dx$$

$$= (\gamma_n 2\pi)^{-1} \sum_{l=-n}^{n} k(l/\gamma_n) \exp(i\xi l/\gamma_n) \left(\int_{-\infty}^{\infty} \exp(-i\xi(x + l/\gamma_n)) \eta_\delta(x + l/\gamma_n) dx \right)$$

$$= (\gamma_n 2\pi)^{-1} \sum_{l=-n}^{n} k(l/\gamma_n) \exp(i\xi l/\gamma_n) \exp(-\delta^2 \xi^2/2)$$

$$\equiv \psi_n(\xi) \exp(-\delta^2 \xi^2/2), \tag{43}$$

and note that for all $x \in I\!R$, $\lim_{n\to\infty} \psi_n(x) = \psi(x)$. Next, note that from the representation of Equation (40) and the properties of $\| \, . \, \|_1$, it follows that

$$\| \hat{\Omega}_n - \Omega_{1n\delta} \|_1$$

$$\leq \int_{-\infty}^{\infty} \left(E(\sum_{t=1}^{n} X_{nt} \cos(t\xi/\gamma_n))^2 + E(\sum_{t=1}^{n} X_{nt} \sin(t\xi/\gamma_n))^2 \right) \times$$

$$(\psi(\xi) - \psi_n(\xi) \exp(-\xi^2 \delta^2/2)) d\xi, \tag{44}$$

and therefore by Lemma A.1,

$$\| \hat{\Omega}_n - \Omega_{1n\delta} \|_1$$

$$= O(\int_{-\infty}^{\infty} (|\psi(\xi) - \psi(\xi) \exp(-\xi^2 \delta^2/2)| + |\psi(\xi) - \psi_n(\xi)| \exp(-\xi^2 \delta^2/2)) d\xi) \tag{45}$$

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

Lemma A. 3 *Under the conditions of Lemma 1,*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \| \Omega_{1n\delta} - \Omega_{2n\delta} \|_1 = 0. \tag{46}$$

Proof:

Note that

$$
\begin{aligned}
&\| \Omega_{1n\delta} - \Omega_{2n\delta} \|_1 \\
&= \| \sum_{t=-n+1}^{2n} (\gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,n]}(|l|) 1_{[b_n+1,\infty)}(|l|)) \times \\
&\quad (\gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n)) \|_1 \\
&\le \sum_{t=-n+1}^{2n} \| \gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[b_n+1,\infty)}(|l|) 1_{[0,n]}(|l|) \|_2 \times \\
&\quad \| \gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) \|_2 \\
&\le 3(\sum_{t=-n+1}^{2n} \gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 k(l/\gamma_n)^2 1_{[b_n+1,\infty)}(|l|) 1_{[0,n]}(|l|))^{1/2} \times \\
&\quad (\sum_{t=-n+1}^{2n} \gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 \eta_\delta(j/\gamma_n)^2)^{1/2} \\
&= O((\sum_{l=b_n+1}^{\infty} \gamma_n^{-1} k(l/\gamma_n)^2 (\sup_{n\ge 1} \sum_{t=1}^{n} c_{nt}^2))^{1/2} (\sum_{j=0}^{\infty} \gamma_n^{-1} \eta_\delta(j/\gamma_n)^2 (\sup_{n\ge 1} \sum_{t=1}^{n} c_{nt}^2))^{1/2}) \\
&= O((\int_{1/\delta}^{\infty} k(x)^2 dx)^{1/2}) \to 0 \qquad (47)
\end{aligned}
$$

as $\delta \to 0$, where the first two inequalities are Cauchy-Schwartz's. ∎

Lemma A. 4 *Under the conditions of Lemma 1,*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \| \Omega_{2n\delta} - \Omega_{3n\delta} \|_1 = 0. \qquad (48)$$

Proof:

This proof is similar to the previous proof. ∎

Lemma A. 5 *Under the conditions of Lemma 1, for all $\delta > 0$,*

$$\lim_{n\to\infty} \| \Omega_{3n\delta} - E\Omega_{3n\delta} \|_1 = 0. \qquad (49)$$

Proof:

Note that

$$\Omega_{3n\delta} = \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt}, \tag{50}$$

where

$$Y_{nt} = \gamma_n^{-1/2} \sum_{l=1-t}^{n-t} X_{n,t+l} k(l/\gamma_n) 1_{[0,b_n]}(|l|), \tag{51}$$

$$Z_{nt} = \gamma_n^{-1/2} \sum_{j=1-t}^{n-t} X_{n,t+j} \eta_\delta(j/\gamma_n) 1_{[0,b_n]}(|j|), \tag{52}$$

and define

$$\phi_{nt}^2 = \gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 k(l/\gamma_n)^2 1_{[0,b_n]}(|l|), \tag{53}$$

$$\psi_{nt}^2 = \gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 \eta_\delta(j/\gamma_n)^2 1_{[0,b_n]}(|j|). \tag{54}$$

Next, 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. Next, define $h(K,x) = x1_{[-K,K]}(x) + K1_{(K,\infty)}(x) - K1_{(-\infty,-K)}(x)$, and let K_η be some constant that will be defined later on. 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

$$\limsup_{n\to\infty} \| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} - \tilde{Y}_{nt} \tilde{Z}_{nt} \|_1$$

$$\leq \limsup_{n\to\infty} (\| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} 1_{(K_\eta \phi_{nt},\infty]}(|Y_{nt}|) \|_1 + \| \sum_{t=-n+1}^{2n} Y_{nt} Z_{nt} 1_{(K_\eta \psi_{nt},\infty)}(|Z_{nt}|) \|_1)$$

$$\leq C \left(\sup_{t,n\geq 1} E(Z_{nt}^2/\psi_{nt}^2) 1_{(K_\eta^2,\infty)}(Z_{nt}^2/\psi_{nt}^2) \right)^{1/2}$$

$$+ C \left(\sup_{t,n\geq 1} E(Y_{nt}^2/\phi_{nt}^2) 1_{(K_\eta^2,\infty)}(Y_{nt}^2/\phi_{nt}^2) \right)^{1/2} < \eta \tag{55}$$

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

$$\sup_{n\geq 1} \sum_{t=-n+1}^{2n} (\psi_{nt}^2 + \phi_{nt}^2) < \infty \tag{56}$$

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

$$\| \tilde{Y}_{nt}\tilde{Z}_{nt} - E(\tilde{Y}_{nt}\tilde{Z}_{nt} | \mathcal{V}_{n,t-m}^{t+m}) \|_2$$

$$\leq \| \tilde{Y}_{nt}\tilde{Z}_{nt} - E(\tilde{Y}_{nt}|\mathcal{V}_{n,t-m}^{t+m})\tilde{Z}_{nt} \|_2$$

$$+ \| E(\tilde{Y}_{nt}|\mathcal{V}_{n,t-m}^{t+m})\tilde{Z}_{nt} - E(\tilde{Y}_{nt}|\mathcal{V}_{n,t-m}^{t+m})E(\tilde{Z}_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2$$

$$\leq \| \tilde{Z}_{nt} \|_\infty \| \tilde{Y}_{nt} - E(\tilde{Y}_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2$$

$$+ \| E(\tilde{Y}_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_\infty \| \tilde{Z}_{nt} - E(\tilde{Z}_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2$$

$$= O((\gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|))^{1/2} \| Y_{nt} - E(Y_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2)$$

$$+O((\gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 1_{[0,b_n]}(|l|))^{1/2} \| Z_{nt} - E(Z_{nt}|\mathcal{V}_{n,t-m}^{t+m}) \|_2)$$

$$= O((\gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|))\gamma_n^{1/2}\nu(m)). \tag{57}$$

Next, we introduce our blocking scheme. Define $r_n' = [3n/2b_n]$, and note that

$$\sum_{t=-n+1}^{2n} \tilde{Y}_{nt}\tilde{Z}_{nt} = \sum_{i=1}^{r_n'} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \tilde{Y}_{nt}\tilde{Z}_{nt}$$

$$+\sum_{i=1}^{r_n'} \sum_{t=(2i-1)b_n-n+1}^{2ib_n-n} \tilde{Y}_{nt}\tilde{Z}_{nt} + \sum_{t=r_n'b_n-n+1}^{2n} \tilde{Y}_{nt}\tilde{Z}_{nt}$$

$$\equiv \sum_{i=1}^{r_n'} U_{ni} + \sum_{i=1}^{r_n'} U_{ni}' + \sum_{t=r_n'b_n-n+1}^{2n} \tilde{Y}_{nt}\tilde{Z}_{nt}. \tag{58}$$

For the last term, we have

$$\| \sum_{t=r_n'b_n-n+1}^{2n} \tilde{Y}_{nt}\tilde{Z}_{nt} \|_1 = O(b_n \max_{1\leq t\leq n} c_{nt}^2) = o(1) \tag{59}$$

by assumption. We will analyze the sum of the U_{ni}, noting that the case of the U_{ni}' is analogous, and note that the assertion of the lemma follows if we can show that the first term of Equation (58) obeys a law of large numbers. Define

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

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

$$\| U_{ni} - E(U_{ni}|\mathcal{W}_{n,i-m}^{i+m}) \|_2 \leq \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \| \tilde{Y}_{nt}\tilde{Z}_{nt} - E(\tilde{Y}_{nt}\tilde{Z}_{nt}|\mathcal{W}_{n,i-m}^{i+m}) \|_2$$

$$\leq \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \| \tilde{Y}_{nt}\tilde{Z}_{nt} - E(\tilde{Y}_{nt}\tilde{Z}_{nt}|\mathcal{V}_{n,t-mb_n}^{t+mb_n}) \|_2$$

$$= O(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) m^{-1/2-\varepsilon}) \tag{60}$$

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 $m \geq 1$. For $m = 0$, the relevant result is

$$\| U_{ni} - E(U_{ni}|\mathcal{W}_{ni}^i) \|_2 \leq 2 \| U_{ni} \|_2$$

$$= O(\sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} (\gamma_n^{-1} \sum_{l=1-t}^{n-t} c_{n,t+l}^2 k(l/\gamma_n)^2)^{1/2} (\gamma_n^{-1} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 \eta_\delta(j/\gamma_n)^2)^{1/2})$$

$$= O(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|)). \tag{61}$$

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

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

$$+O(\alpha(mb_n)^{1/2-1/r} \| U_{ni} \|_r)$$

$$= O(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) m^{-1/2-\varepsilon}) \tag{62}$$

similarly to the argument in De Jong (1995). For $m = 0$, the result from Equation (61) can again be used. Finally, note that from Lemma A.1 it follows that

$$E|\sum_{i=1}^{r_n'} U_{ni} - EU_{ni}|^2 = O(\sum_{i=1}^{r_n'} h_{ni}^2) \tag{63}$$

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

$$\lim_{n\to\infty} \sum_{i=1}^{r_n'} \left(\gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \right)^2$$

$$\leq \lim_{n\to\infty} \sum_{i=1}^{r_n'} \gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|) \times$$

$$\max_{1\leq i\leq r_n'} \gamma_n^{-1} \sum_{t=(2i-2)b_n-n+1}^{(2i-1)b_n-n} \sum_{j=1-t}^{n-t} c_{n,t+j}^2 1_{[0,b_n]}(|j|)$$

$$= O(\gamma_n \max_{1\leq t\leq n} c_{nt}^2) = o(1) \tag{64}$$

under the conditions stated. ∎

Proof of Lemma 2:

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

$$X_{nt} = \sum_{k=-\infty}^{\infty} Y_{tk} \tag{65}$$

and also that

$$EY_{tk}Y_{t+m,i+m} = 0 \tag{66}$$

unless $i = k$. Further, letting

$$\xi_{tk} = \|E(X_{nt}|\mathcal{H}_{n,t-k})\|_2 \tag{67}$$

and

$$\zeta_{tk} = \|X_{nt} - E(X_{nt}|\mathcal{H}_{n,t+k})\|_2\,, \tag{68}$$

we have

$$\| Y_{tk} \|_2^2 = \xi_{tk}^2 - \xi_{t,k+1}^2 = \zeta_{t,k-1}^2 - \zeta_{t,k}^2. \tag{69}$$

Substituting from De Jong (1995), Lemma 3 we have

$$|EX_{nt}X_{n,t+j}| \leq |\sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} EY_{tk}Y_{t+j,k+j}| = |\sum_{k=-\infty}^{\infty} EY_{tk}Y_{t+j,k+j}|$$

$$\leq \sum_{k=-\infty}^{\infty} \|Y_{tk}\|_2 \|Y_{t+j,k+j}\|_2 = \sum_{k=-\infty}^{0}\left(\xi_{tk}^2 - \xi_{t,k+1}^2\right)^{1/2}\left(\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2\right)^{1/2}$$

$$+\sum_{k=1}^{\infty}\left(\zeta_{t,k-1}^2 - \zeta_{tk}^2\right)^{1/2}\left(\zeta_{t+j,k-j-1}^2 - \zeta_{t+j,k-j}^2\right)^{1/2} \tag{70}$$

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,

$$|E\hat{\Omega}_n - \Omega_n| = |\sum_{t=1}^{n}\sum_{s=1}^{n} EX_{nt}X_{ns}(1 - k((t-s)/\gamma_n))|$$

$$\leq 2|\sum_{t=1}^{n}\sum_{j=0}^{n-t} EX_{nt}X_{n,t+j}(1 - k(j/\gamma_n))| \leq 2T_{n1} + 2T_{n2} \tag{71}$$

where

$$T_{n1} = \sum_{k=-\infty}^{0}\sum_{t=1}^{n}\left(\xi_{tk}^2 - \xi_{t,k+1}^2\right)^{1/2}\sum_{j=0}^{n-t}|1 - k(j/\gamma_n)|\left(\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2\right)^{1/2} \tag{72}$$

and

$$T_{n2} = \sum_{k=1}^{\infty}\sum_{t=1}^{n}\left(\zeta_{t,k-1}^2 - \zeta_{tk}^2\right)^{1/2}\sum_{j=0}^{n-t}|1 - k(j/\gamma_n)|\left(\zeta_{t+j,k-j-1}^2 - \zeta_{t+j,k-j}^2\right)^{1/2}. \tag{73}$$

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-Schwartz inequality yield

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

$$\leq \sum_{k=-\infty}^{0}\left(\sum_{t=1}^{n}(\xi_{tk}^2 - \xi_{t,k+1}^2)\right)^{1/2} \times$$

$$\sum_{t=1}^{n}\left(\sum_{j=0}^{n-t}(\beta_j^{1/2}|1-k(j/\gamma_n)|)(\beta_j^{-1/2}\left(\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2\right)^{1/2})\right)^2$$

$$\leq \sum_{k=-\infty}^{0}\left(\sum_{t=1}^{n}(\xi_{tk}^2 - \xi_{t,k+1}^2)\right)^{1/2} \times$$

$$\left(\sum_{t=1}^{n}(\sum_{j=0}^{n-t}\beta_j^{-1}(\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2))(\sum_{j=0}^{\infty}\beta_j(1-k(j/\gamma_n))^2\right)^{1/2}. \tag{74}$$

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

$$\sum_{t=1}^{n}\sum_{j=0}^{n-t}\beta_j^{-1}(\xi_{t+j,k+j}^2 - \xi_{t+j,k+j+1}^2) = \sum_{s=1}^{n}\sum_{j=0}^{s-1}\beta_j^{-1}(\xi_{s,k+j}^2 - \xi_{s,k+j+1}^2)$$

$$\leq \sum_{s=1}^{n}\left(\sum_{j=0}^{s-2}(\beta_{j+1}^{-1} - \beta_j^{-1})\xi_{s,k+j+1}^2 + \beta_0^{-1}\xi_{s,k+1}^2\right) = O(n) \tag{75}$$

by assumption. Therefore,

$$T_{n1} = O(\sum_{k=-\infty}^{0}(\sum_{t=1}^{n}(\xi_{tk}^2 - \xi_{t,k+1}^2))^{1/2}(\sum_{j=0}^{\infty}\beta_j(1-k(j/\gamma_n))^2)^{1/2})$$

$$= O(1) \times o(1) = o(1) \tag{76}$$

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

Proof of Lemma 3:

For the case that Assumptions 1, 2, 3 and 5 hold, the proof is identical to that in Hansen (1992, p. 971-972). For the case that Assumptions 1, 2, 3 and 4 hold, first note that by Taylor's theorem,

$$\hat{\Omega}_n(\hat{\theta}_n) - \hat{\Omega}_n = 2\sum_{t=1}^{n}\sum_{s=1}^{n}(\partial/\partial\theta)X_{nt}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)X_{ns}k((t-s)/\gamma_n)$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{n}(\partial/\partial\theta)X_{ns}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)(\partial/\partial\theta)X_{nt}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)k((t-s)/\gamma_n) \tag{77}$$

for some $\tilde{\theta}_n$ that is on the line between θ_0 and $\hat{\theta}_n$. The last term converges in probability to zero because $n^{1/2}(\hat{\theta}_n - \theta_0) = O_P(1)$ by assumption and because

$$\| n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}(\partial/\partial\theta')X_{ns}(\tilde{\theta}_n)(\partial/\partial\theta)X_{nt}(\tilde{\theta}_n)k((t-s)/\gamma_n) \|_1$$

$$= O(n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}|k((t-s)/\gamma_n)| \; \| (\partial/\partial\theta)X_{ns}(\tilde{\theta}_n) \|_2 \| (\partial/\partial\theta)X_{nt}(\tilde{\theta}_n) \|_2)$$

$$= O(\sum_{j=-n+1}^{n-1}|k(j/\gamma_n)|n^{-1}\sum_{t=1}^{n}E\sup_{\theta\in\Theta}|(\partial/\partial\theta)X_{nt}(\theta)|^2)$$

$$= O(\sum_{j=-n+1}^{n-1}|k(j/\gamma_n)|/n) = O(\gamma_n/n) = o(1) \tag{78}$$

by assumption. Next, write $\lambda_t = E(\partial/\partial\theta)X_{nt}(\theta_0)$ and note that

$$\sum_{t=1}^{n}\sum_{s=1}^{n}(\partial/\partial\theta)X_{nt}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)X_{ns}k((t-s)/\gamma_n)$$

$$= \sum_{t=1}^{n}\sum_{s=1}^{n}((\partial/\partial\theta)X_{nt}(\tilde{\theta}_n) - \lambda_t)(\hat{\theta}_n - \theta_0)X_{ns}k((t-s)/\gamma_n)$$

$$+n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}\lambda_t(\hat{\theta}_n - \theta_0)X_{ns}k((t-s)/\gamma_n), \tag{79}$$

and the first term converges in probability to zero because $n^{1/2}(\hat{\theta} - \theta_0) = O_P(1)$ and because

$$\| n^{-1/2}\sum_{t=1}^{n}\sum_{s=1}^{n}((\partial/\partial\theta)X_{nt}(\tilde{\theta}_n) - \lambda_t)X_{ns}k((t-s)/\gamma_n) \|_1$$

$$\leq \int_{-\infty}^{\infty} \| \left(\sum_{s=1}^{n} X_{ns}\exp(-i\xi s/\gamma_n)\right) \times$$

$$\left(n^{-1/2}\sum_{t=1}^{n}((\partial/\partial\theta)X_{nt}(\tilde{\theta}_n)-\lambda_t)\exp(i\xi t/\gamma_n)\right)\|_1\,|\psi(\xi)|d\xi$$

$$\leq\int_{-\infty}^{\infty}\|\sum_{s=1}^{n}X_{ns}\exp(-i\xi s/\gamma_n)\|_2\times$$

$$\Bigg(\|\sup_{\theta\in\Theta}|\sum_{t=1}^{n}((\partial/\partial\theta)X_{nt}(\theta)-E(\partial/\partial\theta)X_{nt}(\theta))\exp(i\xi t/\gamma_n)|\,\|_2+$$

$$\|\sum_{t=1}^{n}(E(\partial/\partial\theta)X_{nt}(\tilde{\theta}_n))-\lambda_t)\exp(i\xi t/\gamma_n)\|_2\Bigg)|\psi(\xi)|d\xi=o(1) \tag{80}$$

by assumption. Finally, note that

$$(\sum_{t=1}^{n}\sum_{s=1}^{n}\lambda_t(\hat{\theta}_n-\theta_0)X_{ns}k((t-s)/\gamma_n))^2$$

$$=n(\hat{\theta}_n-\theta_0)'M(\hat{\theta}_n-\theta_0) \tag{81}$$

where

$$M=n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_t'X_{ns}X_{ni}\lambda_j k((t-s)/\gamma_n)k((i-j)/\gamma_n), \tag{82}$$

and because M is a positive definite matrix and because $n^{1/2}(\hat{\theta}_n-\theta_0)=O_P(1)$, the result now follows if we can show that

$$En^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_t\lambda_j'X_{ns}X_{ni}k((t-s)/\gamma_n)k((i-j)/\gamma_n)=o(1).$$

In order to prove this, first note that from the proof of Lemma 2 it can be seen that for L_2-mixingale random variables X_{nt} of size -1/2 and a positive deterministic array a_{nst} we have

$$|\sum_{i=1}^{n}\sum_{s=1}^{n}EX_{ni}X_{ns}a_{nij}a_{nst}|$$

$$=O((\sum_{i=1}^{n}c_{ni}^2a_{nij}^2\sum_{s=1}^{n}c_{ns}^2a_{nst}^2)^{1/2}). \tag{83}$$

Therefore, we have

$$En^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_t'X_{ns}X_{ni}k((t-s)/\gamma_n)k((i-j)/\gamma_n)$$

$$=O\left(n^{-2}\sum_{t=1}^{n}\sum_{j=1}^{n}|\lambda_t||\lambda_j|(\sum_{i=1}^{n}c_{ni}^2k((i-j)/\gamma_n)^2\sum_{s=1}^{n}c_{ns}^2k((s-t)/\gamma_n)^2)^{1/2}\right)$$

$$= O\left(n^{-2}\left(\sum_{t=1}^{n}|\lambda_t|(\sum_{s=1}^{n}c_{ns}^2 k((t-s)/\gamma_n)^2)^{1/2}\right)^2\right)$$

$$= O\left((\sum_{t=1}^{n}|\lambda_t|^2)(n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}c_{ns}^2 k((t-s)/\gamma_n)^2)\right)$$

$$= O(\gamma_n/n) = o(1), \tag{84}$$

where the third equality follows from the Cauchy-Schwartz inequality. This concludes the proof. ∎

Proof of Theorem 2:

For the case that Assumptions 1, 2, 3 and 5 hold, the proof is identical to that in Hansen (1992, p. 971-972). For the case that Assumptions 1, 2, 3 and 4 hold, note that by assumption, with arbitrary large probability $\hat{\alpha}$ can be made an element of the interval $[\varepsilon, 1/\varepsilon]$ for some small $\varepsilon \in (0,1)$. Furthermore note that it is well-known (see e.g. Newey (1991)) that

$$\sup_{\alpha\in[\varepsilon,1/\varepsilon]}|\hat{\Omega}_n(\hat{\theta}_n,\alpha\gamma_n)-\Omega_n| \xrightarrow{p} 0 \tag{85}$$

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

$$\hat{\Omega}_n(\theta_0,\alpha\gamma_n)$$

$$= \int_{-\infty}^{\infty}\left((\sum_{t=1}^{n}X_{nt}\sin(t\xi/\alpha\delta_n))^2+(\sum_{t=1}^{n}X_{nt}\cos(t\xi/\alpha\delta_n))^2\right)\psi(\xi)d\xi$$

$$= \int_{-\infty}^{\infty}\left((\sum_{t=1}^{n}X_{nt}\sin(t\phi/\delta_n))^2+(\sum_{t=1}^{n}X_{nt}\cos(t\phi/\delta_n))^2\right)\psi(\alpha\phi)\alpha d\phi \tag{86}$$

so

$$\sup_{\alpha\in[\varepsilon,1/\varepsilon]}\sup_{\alpha\in[\varepsilon,1/\varepsilon]:|\alpha-\alpha'|<\eta}|\hat{\Omega}_n(\theta_0,\alpha\gamma_n)-\hat{\Omega}_n(\theta_0,\alpha'\gamma_n)|$$

$$\leq \int_{-\infty}^{\infty}\left(E(\sum_{t=1}^{n}X_{nt}\sin(t\phi/\delta_n))^2+E(\sum_{t=1}^{n}X_{nt}\cos(t\phi/\delta_n))^2\right)\times$$

$$\sup_{\alpha\in[\varepsilon,1/\varepsilon]}\sup_{\alpha'\in[\varepsilon,1/\varepsilon]:|\alpha-\alpha'|<\eta}|\psi(\alpha\phi)\alpha-\psi(\alpha'\phi)\alpha'|d\phi. \tag{87}$$

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

$$\lim_{\eta\to 0}\limsup_{n\to\infty}E\sup_{\alpha\in[\varepsilon,1/\varepsilon]}\sup_{\alpha'\in[\varepsilon,1/\varepsilon]:|\alpha-\alpha'|<\eta}|\hat{\Omega}_n(\theta_0,\alpha\gamma_n)-\hat{\Omega}_n(\theta_0,\alpha'\gamma_n)|=0, \tag{88}$$

which implies stochastic equicontinuity of $\hat{\Omega}_n(\theta_0, \alpha\gamma_n) - \Omega_n$. For showing the other result, consider Equation (77). From copying the reasoning leading up to Equation (78), it is easily shown that

$$\| \sup_{\alpha\in[\varepsilon,1/\varepsilon]} |n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (\partial/\partial\theta) X_{ns}(\tilde{\theta}_n)(\partial/\partial\theta') X_{nt}(\tilde{\theta}_n) k((t-s)/\gamma_n)| \|_1$$

$$= O(n^{-1} \sum_{j=-n+1}^{n-1} \sup_{\alpha\in[\varepsilon,1/\varepsilon]} |k(\alpha x)| dx) = O(\gamma_n/n) = o(1) \qquad (89)$$

by assumption. 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 to show that

$$T_n(\alpha) = \sum_{t=1}^{n} \sum_{s=1}^{n} (\partial/\partial\theta) X_{nt}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) X_{ns} k((t-s)/\gamma_n) \qquad (90)$$

is stochastically equicontinuous. This is shown by noting that

$$T_n(\alpha) = \int_{-\infty}^{\infty} (\sum_{s=1}^{n} X_{nt} \exp(-is\xi/\gamma_n))(\sum_{t=1}^{n} (\partial/\partial\theta) X_{nt}(\tilde{\theta}_n) \exp(it\xi/\gamma_n))\alpha\psi(\alpha\xi) d\xi, (91)$$

and therefore

$$E \sup_{\alpha\in[\varepsilon,1/\varepsilon]} \sup_{|\alpha-\alpha'|<\eta} |T_n(\alpha) - T_n(\alpha')|$$

$$\leq \int_{-\infty}^{\infty} \| \sum_{t=1}^{n} X_{nt} \exp(it\xi/\gamma_n) \|_2 \times$$

$$\| \sum_{t=1}^{n} \sup_{\theta\in\Theta} (\partial/\partial\theta) X_{nt}(\theta) \exp(it\xi/\gamma_n) \|_2 \sup_{\alpha\in[\varepsilon,1/\varepsilon]} \sup_{|\alpha-\alpha'|<\eta} |\alpha\psi(\alpha\xi) - \alpha'\psi(\alpha'\xi)| d\xi$$

$$= O(\int_{-\infty}^{\infty} \sup_{\alpha\in[\varepsilon,1/\varepsilon]} \sup_{|\alpha-\alpha'|<\eta} |\alpha\psi(\alpha\xi) - \alpha'\psi(\alpha'\xi)| d\xi) \longrightarrow 0 \qquad (92)$$

as $\eta \to 0$. ∎

References

Andrews, D.W.K. (1991): "Heteroscedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858. Andrews, D.W.K. and J.C. Monahan (1992): "An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953-966 .

Davidson, J. (1992): "A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes," *Econometric theory*, 8, 313-334.

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.

Davidson, J. (1994): *Stochastic limit theory.* Oxford: Oxford University Press.

Doukhan, P., Massart, P. and E. Rio (1994): "The functional central limit theorem for strongly mixing processes," *Annales de l'institute Henri Poincaré, Probabilités et statistiques*, 10, 63-82.

Gallant, A.R. and H. White (1988): *A unified theory of estimation and inference for nonlinear dynamic models.* New York: Basil Blackwell.

Hall, P. and C.C. Heyde (1980): *Martingale limit theory and its applications.* New York: Academic Press.

Hansen, B.E. (1992): "Consistent covariance matrix estimation for dependent heterogeneous processes," *Econometrica*, 60, 967-972.

De Jong, R.M. (1995): "Central limit theorems for dependent heterogeneous processes," Working paper.

Kool, H. (1988): "A note on consistent estimation of heteroscedastic and autocorrelated covariance matrices," Research memorandum 1988-21, Department of Econometrics, Free University Amsterdam.

McLeish, D.L. (1975): "Invariance principles for dependent variables," *Zeitschrift für Wahrscheinlichskeitstheorie und Verwandte Gebiete*, 33, 165-178.

McLeish, D.L. (1977): "On the invariance principle for nonstationary mixingales," *Annals of probability*, 5, 616-621.

Newey, W.K. (1991): "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 59, 1161-1167.

Newey, W.K. and K.D. West (1987): "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703-708.

Newey, W.K. and K.D. West (1994): "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61, 631-654.

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.

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.

White, H. (1984): *Asymptotic theory for econometricians.* New York: Academic Press.