A new consistency proof for HAC variance estimators

James Davidson
University of Exeter

Revised, August 2019

Abstract

A consistency theorem for kernel HAC variance estimators was originally proposed by Hansen (1992) but corrected under stronger conditions on the order of existing moments by de Jong (2000). The present result restores and also generalizes the conditions of Hansen's result by assuming the process to be adapted to a filtration. It allows for nonstationarity, and dependence is modelled by the assumption of near-epoch dependence on a mixing process.

1 Introduction

Research on the consistency of heteroscedastic and autocorrelation consistent (HAC) nonparametric variance estimators in econometrics appears to date from White's (1984) monograph. Subsequent contributions have notably included Newey and West (1987), Gallant and White (1988), Andrews (1991), Hansen (1992), de Jong (2000), de Jong and Davidson (2000), Davidson and de Jong (2002) and Jansson (2002). With the exception of de Jong (2000) these studies all deal with weak consistency. They pose various restrictions on the form of the weighting kernel and a variety of dependence conditions on the data, including strong and uniform mixing, near-epoch dependence (NED) on a mixing process, and linearity. The challenge in this literature has been to find conditions matching those sufficient for a central limit theorem to hold, given the usual interdependence of the respective convergences of the statistic and the studentizing factor. Theorem 2.1 of de Jong and Davidson (2000) appears by general consent to approach this goal most nearly in the 'NED-on-mixing' framework. However, this theorem is not very attractive from a pedagogical point of view since the proof is lengthy and technical.

Hansen's (1992) result is pedagogically attractive since it benefits from a simple proof and quite general conditions, in particular not requiring the existence of fourth moments. Unfortunately the proof contained an error, noted by de Jong (2000) who gave a corrected version that significantly strengthened the sufficient conditions. Letting γ_n denote the kernel bandwidth, the consistency condition $\gamma_n = o(n^{1-2/p})$ became $\gamma_n = o(n^{1/2-1/p})$ where p is strictly smaller than the minimum order of existing moments. Even with fourth moments existing, this may rule out in particular $\gamma_n = O(n^{1/3})$, the usual bandwidth choice in conjunction with the popular Bartlett kernel. Thus, the corrected form of the Hansen (1992) theorem is really too restrictive to be very useful.

The present paper gives an alternative modification of Hansen's theorem that restores the original convergence rate. The new result generalizes the dependence characterization to allow NED on a mixing process, and is formulated to allow the processes to exhibit nonstationary variations in second moments, but with the extra conditions that the data series are adapted to a filtration as well as an extension of the NED property to apply to sequences of conditional means. The new trick is to show that under these assumptions, sequences of the form X_tX_{t+s} can possess the mixingale property where the rate of convergence is independent of s.

2 The Result

Let $(X_1, \ldots, X_n)$ be a real stochastic sequence with zero mean and define $\sigma_{t,t+k} = \mathrm{E}(X_t X_{t+k})$ so that the variance of the sum is

$$s_n^2 = \mathrm{E}\left(\sum_{t=1}^{n} X_t\right)^2 = \sum_{k=1-n}^{n-1} \sum_{t\in n(k)} \sigma_{t,t+k} \tag{2.1}$$

where $n(k) = \{t : \max(1, 1-k) \le t \le \min(n, n-k)\}$. The object is to show that $\hat{s}_n^2/s_n^2$ converges in probability to 1 where

$$\hat{s}_n^2 = \sum_{k=1-n}^{n-1} w_{nk} \sum_{t\in n(k)} X_t X_{t+k} \tag{2.2}$$

for kernel function $w_{nk} = w(k/\gamma_n)$ and bandwidth sequence γ_n. Considering a scalar variance incurs little loss of generality since a legitimate case is $X_t = \boldsymbol{\alpha}'\boldsymbol{Z}_t$ for a vector of random variables $\boldsymbol{Z}_t$ and conformable weight vector $\boldsymbol{\alpha}$. This setup is almost invariably needed to establish multivariate weak convergence via the Cramér-Wold device.

Some well-known definitions in forms appropriate to the present setup are as follows. If $(\Omega, \mathcal{F}, \boldsymbol{F}, P)$ is a filtered probability space with filtration $\boldsymbol{F} = \{\mathcal{F}_t\}_{t=-\infty}^{\infty}$, a sequence $\{Y_t\}$ is said to be adapted to $\boldsymbol{F}$ if Y_t is $\mathcal{F}_t$-measurable for each t so that $\mathrm{E}_s Y_t = Y_t$ for $s \ge t$.[1] An adapted sequence $\{Y_t\}$ will be called a L_p-mixingale of size $-a$ if $\|\mathrm{E}_{t-k}Y_t\|_p \le \|Y_t\|_r \zeta_k$ for $r \ge p$ where $\zeta_k = O(k^{-a-\delta})$ for $\delta > 0$. Also, letting $\{\boldsymbol{V}_t\}$ be an $\boldsymbol{F}$-adapted sequence where $\boldsymbol{V}_t$ is vector-valued in general, the adapted sequence $\{Y_t\}$ will be called L_p-NED of size $-a$ on $\{\boldsymbol{V}_t\}$ if $\|Y_t - \mathrm{E}_{t-k}^t Y_t\|_p \le \|Y_t\|_p \nu_k$ where $\mathrm{E}_{t-k}^t(\cdot) = \mathrm{E}(\cdot|\sigma(\boldsymbol{V}_{t-k}, \ldots, \boldsymbol{V}_t))$ and $\nu_k = O(k^{-a-\delta})$ for $\delta > 0$.[2]

Assumption 1 $\{X_t\}$ is a L_r-bounded zero-mean $\boldsymbol{F}$-adapted sequence. For each $s \ge 0$, $\{\mathrm{E}_t X_{t+s}\}$ is L_p-NED of size $-a$ where $\{\boldsymbol{V}_t\}$ is either α-mixing of size $-apr/(r-p)$ with $r > p$, or ϕ-mixing of size $-ar/(r-1)$ with $r \ge p$. Either

(a) $p \ge 4$ and $a = -\frac{1}{2}$, or

(b) $4 > p > 2$ and $a = -1$. □

While the adaptation assumption is unusual in the literature, it is not a strong assumption in most econometric contexts. Such processes are called 'causal', with the familiar implication that the arrow of time is unidirectional. The martingale difference is only the most prominent case. Hansen (1992) and de Jong (2000), as well as the oft-cited result of Newey and West (1987), specify dependence in the form of strong or uniform mixing conditions, Hansen's mixing size conditions improving on Newey-West and being comparable to our own. The more general NED-on-mixing characterization (but without adaptation) is used by de Jong and Davidson (2000) and Davidson and de Jong (2002) as well as Gallant and White (1988).

Another feature of Assumption 1 that is novel, albeit quite natural given the adapted process framework, is the NED attribute of the sequence of conditional means. This subsumes the usual NED property for X_t which holds as the case $s = 0$. Under adaptation, by the law of iterated expectations the L_p-NED condition bounds $\|\mathrm{E}_t X_{t+s} - \mathrm{E}_{t-k}^t X_{t+s}\|_p$, which clearly converges to zero as $k \to \infty$ for any s. The essence of Assumption 1 is therefore that the NED size does not depend adversely on s. The assumption implies by Lemma 2.1 (see below) that $\{X_t\}$ is an L_p-mixingale

[1] The notation E_t is used as a short form of $\mathrm{E}(\cdot|\mathcal{F}_t)$, and is equivalent to $\mathrm{E}_{-\infty}^t(\cdot)$. Note that the role of the subscript depends on the presence of a superscript.

[2] Conventionally these definitions specify scaling constants c_t or d_t in the majorants, but for simplicity these are equated with the L_p- or L_r-norms in this application. The notations ζ_k and ν_k are generic, simply denoting sequences with the specified orders of magnitude.

with $\|\mathrm{E}_t X_{t+s}\|_p \leq \|X_{t+s}\|_r \zeta_s$. Therefore, it further implies $\|\mathrm{E}_t X_{t+s} - \mathrm{E}_{t-k}^t X_{t+s}\|_p \leq \|X_{t+s}\|_r \zeta_s \nu_k$. For illustration, consider the linear adapted process $\{X_t, \mathcal{F}_t\}$ where $X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$, with $\mathcal{F}_t = \sigma(\varepsilon_u, u \leq t)$ and $\{\varepsilon_t\}$ an L_2-bounded independent sequence with zero mean. Here, $\mathrm{E}_t X_{t+s} = \sum_{j=0}^{\infty} \theta_{j+s} \varepsilon_{t-j}$ and if $|\theta_j| = O(j^{-1-\varphi})$ for $\varphi > \varphi_0 \geq 0$, then for $p \leq 2$,

$$\begin{aligned} \|\mathrm{E}_t X_{t+s} - \mathrm{E}_{t-k}^t X_{t+s}\|_p &= \left\| \textstyle\sum_{j=k+1}^{\infty} \theta_{j+s} \varepsilon_{t-j} \right\|_p \\ &\leq \sup_{u \leq t-k} \|\varepsilon_u\|_p \textstyle\sum_{j=s+k+1}^{\infty} |\theta_j| \\ &= O((s+k)^{-\varphi}). \end{aligned}$$

The sequences $\{\mathrm{E}_t X_{t+s}\}$ are therefore L_p-NED on $\{\varepsilon_t\}$ of size φ_0 for all $s \geq 0$.

Finite moments strictly greater than order two must exist, but there is no requirement under Assumption 1(b) for finite fourth moments. The penalty incurred is a further restriction on dependence, since the NED size needs to be -1 in place of $-\frac{1}{2}$. De Jong and Davidson (2000) avoid this sort of penalty by introducing a truncation, so that all moments of the truncated variables exist yet under uniform integrability assumptions the remainders can be neglected. However, implementation of this trick requires a different decomposition of the estimator from that adopted here and results in a complicated argument.

Assumption 1 permits stationary processes in which variances are constant and $s_n^2 = O(n)$, but also globally nonstationary processes where the variance sequence may either diverge or degenerate within limits, specifically, $\sigma_{tt} \simeq t^{\beta}$ for $-1 < \beta < \infty$.[3] However, the dependence restrictions ensure that $s_n^2 \simeq n^{1+\beta}$ in such a case. It is also evident, on considering the replacement of X_t by the normalized sequence $t^{-\beta/2} X_t$, that $\|X_t\|_r^2 = O(t^{\beta})$ in the same case.

Assumption 2 The kernel function w satisfies the following conditions.

(i) $w(0) = 1$, $w(-x) = w(x)$, all $x \geq 0$ and $\sup_{x \geq 0} |w(x)| < \infty$.
(ii) w is continuous at 0.
(iii) $\int_{-\infty}^{\infty} \bar{w}(x) \mathrm{d}x < \infty$ where $\bar{w}(x) = \sup_{y \geq x} |w(y)|$. □

Assumption 2 is taken from Jansson (2002) who points out that the conditions on the kernel function adopted by Andrews (1991) and Hansen (1992) among others are not sufficient to ensure the convergence

$$\frac{1}{\gamma_n} \sum_{k=1-n}^{n-1} |w_{nk}| \to \int_{-\infty}^{\infty} |w(x)| \mathrm{d}x < \infty \text{ as } n \to \infty \tag{2.3}$$

as is required by the proof of consistency. The condition of Assumption 2(iii) is stronger than the 'continuity except at a finite set of points' commonly specified in this literature, although the condition is satisfied by all the kernel functions in common use and is violated only in extreme cases unlikely to be adopted in practice.

Assumption 3 $\gamma_n = o(n^{1-2/\min\{p,4\}})$. □

If $p \geq 4$ as in Assumption 1(a), Assumption 3 specifies $\gamma_n = o(n^q)$ for any $q \leq \frac{1}{2}$. With $p \geq 3$ which is allowed in case of Assumption 1(b), $q \leq \frac{1}{3}$ is allowed. Andrews (1991) showed that given the existence of fourth moments, $\gamma_n = O(n^{1/3})$ is the optimal bandwidth choice for the Bartlett kernel on the minimum mean-squared error (MSE) criterion, while other cases are optimized by $\gamma_n = O(n^{1/5})$. Both these choices are compatible with consistency under Assumption 1(a).

The consistency proof is an application of mixingale theory resting primarily on the following lemmas, the first one being a standard result.

[3] $b_t \simeq a_t$ denotes that for $N < \infty$ and $0 < A < B < \infty$, $\inf_{n \geq N}(X_n/a_n) \geq A$ and $\sup_{n \geq N}(X_n/a_n) \leq B$.

Lemma 2.1 *Under Assumption 1, $\{X_t\}$ is a $\mathcal{F}_t$-measurable L_p-mixingale of size $-a$ with respect to constants $\|X_t\|_r$.*

Lemma 2.2 *Under Assumption 1, $\{X_t X_{t+s} - \sigma_{t,t+s}\}$ is a $\mathcal{F}_{t+s}$-measurable $L_{p/2}$-mixingale of size $-a$ for each $s \geq 0$.* □

Lemma 2.3 *Under Assumption 1,*

$$\frac{1}{s_n^2}\sum_{t=1}^{n}\sum_{k=0}^{n-t}|\sigma_{t,t+k}| = O(1) \text{ as } n \to \infty. \quad \square \tag{2.4}$$

Lemma 2.2 uses the adaptation condition to show the key result that the mixingale size (matching the NED size) is preserved under the formation of the product $X_t X_{t+s}$. The argument combines the mixingale attribute of NED-on-mixing with the fact that NED is preserved under the product transformation. Lemma 2.3 is based on Lemma 4 of de Jong (1997). Hansen (1992) cited Gallant and White (1988) Lemma 6.6 for this step of the argument, but that result assumes an NED size of -1, and so does not apply under Assumption 1(a). The summability holds in both the stationary and the nonstationary cases, noting that the sum $\sum_{k=0}^{n-1}\sum_{t=k+1}^{n}|\sigma_{t,t-k}|$ contains the same terms as (2.4) and hence Lemma 2.3 applies equally to the forward and backward autocovariances.

Theorem 2.1 *Under Assumptions 1, 2 and 3, $\hat{s}_n^2/s_n^2 - 1 \xrightarrow{\text{pr}} 0$.* □

3 Proofs

Proof of Lemma 2.1 By Theorem SLT17.5.[4] ∎

Proof of Lemma 2.2 $\{X_t\}$ is a L_p-mixingale of size $-a$ by Lemma 2.1. Applying successively the law of iterated expectations and linearity, the modulus, Liapunov and Cauchy-Schwarz inequalities and the mixingale property, the s^{th} autocovariance is bounded as

$$\begin{aligned}|\sigma_{t,t+s}| &= |\mathrm{E}(X_t \mathrm{E}_t X_{t+s})| \\ &\leq \|X_t \mathrm{E}_t X_{t+s}\|_{p/2} \\ &\leq \|X_t\|_p \|\mathrm{E}_t X_{t+s}\|_p \\ &\leq \|X_t\|_p \|X_{t+s}\|_r \zeta_s.\end{aligned} \tag{3.1}$$

First, suppose that $k \leq s$. Applying linearity, the Minkowski and Cauchy-Schwarz inequalities and then (3.1) and the mixingale property gives

$$\begin{aligned}\|\mathrm{E}_{t+s-k}(X_t X_{t+s}) - \sigma_{t,t+s}\|_{p/2} &\leq \|X_t \mathrm{E}_{t+s-k} X_{t+s}\|_{p/2} + |\sigma_{t,t+s}| \\ &\leq \|X_t\|_p \|\mathrm{E}_{t+s-k} X_{t+s}\|_p + |\sigma_{t,t+s}| \\ &\leq \|X_t\|_p \|X_{t+s}\|_r (\zeta_k + \zeta_s).\end{aligned} \tag{3.2}$$

Next, suppose that $k > s$. To show that the sequence $\{X_t \mathrm{E}_t X_{t+s}\}$ is $L_{p/2}$-NED of size $-a$, let $P_t = X_t - \mathrm{E}_{t+s-k}^t X_t$ and $Q_t = \mathrm{E}_t X_{t+s} - \mathrm{E}_{t+s-k}^t \mathrm{E}_t X_{t+s}$, where

$$\|P_t\|_p \leq \|X_t\|_p \nu_{k-s} \tag{3.3}$$

[4] Here and henceforth, the prefix SLT denotes a result cited from Davidson (1994).

and

$$\|Q_t\|_p \le \|\mathrm{E}_t X_{t+s}\|_p \nu_{k-s} \le \|X_{t+s}\|_r \nu_{k-s}\zeta_s \tag{3.4}$$

by Assumption 1 and Lemma 2.1. Hence,

$$\begin{aligned}\|X_t \mathrm{E}_t X_{t+s} - \mathrm{E}^t_{t+s-k}(X_t \mathrm{E}_t X_{t+s})\|_{p/2} &= \|X_t Q_t + P_t \mathrm{E}^t_{t+s-k}\mathrm{E}_t X_{t+s} - \mathrm{E}^t_{t+s-k}(P_t Q_t)\|_{p/2} \\ &\le \|X_t Q_t\|_{p/2} + \|P_t \mathrm{E}^t_{t+s-k}\mathrm{E}_t X_{t+s}\|_{p/2} + \|P_t Q_t\|_{p/2} \\ &\le \|X_t\|_p \|Q_t\|_p + \|P_t\|_p \|\mathrm{E}_t X_{t+s}\|_p + \|P_t\|_p \|Q_t\|_p \\ &\le \|X_t\|_p \|X_{t+s}\|_r (2+\nu_{k-s})\nu_{k-s}\zeta_s. \end{aligned} \tag{3.5}$$

The first inequality of (3.5) applies the Minkowski and conditional Jensen inequalities, the second applies Cauchy-Schwarz and the conditional Jensen a second time, and the last substitutes from (3.3) and (3.4). The expected value of $X_t \mathrm{E}_t X_{t+s}$ is $\sigma_{t,t+s}$ and $\{X_t \mathrm{E}_t X_{t+s} - \sigma_{t,t+s}\}$ is a $L_{p/2}$ mixingale of size $-a$ with respect to scale constants $\|X_t \mathrm{E}_t X_{t+s} - \sigma_{t,t+s}\|_{r/2} \le 2\|X_t\|_r \|X_{t+s}\|_r$ by (3.5) and Lemma 2.1, noting $\nu_{k-s}\zeta_s = O(k^{-a-\delta})$ for $\delta > 0$. The conclusion

$$\begin{aligned}\|\mathrm{E}_{t+s-k}(X_t X_{t+s}) - \sigma_{t,t+s}\|_{p/2} &= \|\mathrm{E}_{t+s-k}(X_t \mathrm{E}_t X_{t+s} - \sigma_{t,t+s})\|_{p/2} \\ &\le 2\|X_t\|_r \|X_{t+s}\|_r \zeta_k \end{aligned} \tag{3.6}$$

now follows by the law of iterated expectations. Combining (3.2) and (3.6) shows that

$$\|\mathrm{E}_{t+s-k}(X_t X_{t+s}) - \sigma_{t,t+s}\|_{p/2} = O(k^{-a-\delta}). \quad \blacksquare$$

Proof of Lemma 2.3 The telescoping sum representation of an adapted process is $X_t = \sum_{j=0}^{\infty} Y_{tj}$ where $Y_{tj} = \mathrm{E}_{t-j}X_t - \mathrm{E}_{t-j-1}X_t$ a.s. and likewise, $X_{t+k} = \sum_{j=-k}^{\infty} Y_{t+k,j+k}$. Note that

$$\sigma_{t,t+k} = \sum_{j=0}^{\infty}\sum_{i=-k}^{\infty} \mathrm{E}(Y_{tj}Y_{t+k,i+k}) = \sum_{j=0}^{\infty} \mathrm{E}(Y_{tj}Y_{t+k,j+k}). \tag{3.7}$$

The second equality of (3.7) follows since if $i > j$ then $\mathrm{E}(Y_{tj}Y_{t+k,i+k}) = \mathrm{E}(Y_{t+k,i+k}\mathrm{E}_{t-i}Y_{tj}) = 0$, if $j > i$ then $\mathrm{E}(Y_{tj}Y_{t+k,i+k}) = \mathrm{E}(Y_{tj}\mathrm{E}_{t-j}Y_{t+k,i+k}) = 0$, and if $j < 0$ then $\mathrm{E}(Y_{tj}Y_{t+k,j+k}) = 0$ since $Y_{tj} = 0$ a.s. in that case. For given t and $k \ge 0$, let

$$p_{t+k,j} = \|Y_{t+k,j+k}\|_2 = \sqrt{\mathrm{E}\mathrm{E}^2_{t-j}X_{t+k} - \mathrm{E}\mathrm{E}^2_{t-j-1}X_{t+k}}.$$

Then by (3.7),

$$|\sigma_{t,t+k}| \le \sum_{j=0}^{\infty} |\mathrm{E}(Y_{tj}Y_{t+k,j+k})| \le \sum_{j=0}^{\infty} p_{tj}p_{t+k,j}$$

where the second inequality applies the Cauchy-Schwarz inequality to each term. Substitute this expression into the formula in (2.4). Define a positive, summable sequence $\{\eta_k\}_{k=0}^{\infty}$ by setting $\eta_0 = \eta_1 = 1$ and $\eta_k = k^{-1}\log(k)^{-2}$ for $k > 1$, and then a double application of the Cauchy-Schwarz inequality for sums yields

$$\begin{aligned}\frac{1}{s_n^2}\sum_{t=1}^{n}\sum_{k=0}^{n-t}|\sigma_{t,t+k}| &\le \frac{1}{s_n^2}\sum_{j=0}^{\infty}\left(\sum_{t=1}^{n} p_{tj}\sum_{k=0}^{n-t} p_{t+k,j}\right) \\ &\le \sum_{j=0}^{\infty}\left(\frac{1}{s_n^2}\sum_{t=1}^{n} p_{tj}^2\right)^{1/2}\left(\frac{1}{s_n^2}\sum_{t=1}^{n}\left(\sum_{k=0}^{n-t} p_{t+k,j}\eta_k^{-1/2}\eta_k^{1/2}\right)^2\right)^{1/2}\end{aligned}$$

$$\leq \sum_{j=0}^{\infty}\left(\frac{1}{s_n^2}\sum_{t=1}^{n}p_{tj}^2\right)^{1/2}\left(\frac{B}{s_n^2}\sum_{s=1}^{n}\sum_{k=0}^{s-1}p_{s,j+k}^2\eta_k^{-1}\right)^{1/2} \tag{3.8}$$

where $B = \sum_{k=0}^{\infty}\eta_k < \infty$. Note the rearrangement of terms in the last member of (3.8), where letting $s = t + k$ it can be verified that $\sum_{t=1}^{n}\sum_{k=0}^{n-t}p_{t+k,j}^2\eta_k^{-1} = \sum_{s=1}^{n}\sum_{k=0}^{s-1}p_{s,j+k}^2\eta_k^{-1}$.

Next, in the majorant of (3.8) consider the j^{th} term of the sum. Applying the Abelian partial summation formula to the second factor gives

$$\begin{aligned}\frac{1}{s_n^2}\sum_{s=1}^{n}\sum_{k=0}^{s-1}p_{s,j+k}^2\eta_k^{-1} &= \frac{1}{s_n^2}\sum_{s=1}^{n}\sum_{k=0}^{s-1}\mathrm{E}\big(\mathrm{E}_{s-j-k}^2X_s - \mathrm{E}_{s-j-k-1}^2X_s\big)\eta_k^{-1} \\ &= \frac{1}{s_n^2}\sum_{s=1}^{n}\mathrm{E}\left(\mathrm{E}_{s-j}^2X_s\eta_0^{-1} - \mathrm{E}_{-j}^2X_s\eta_{s-1}^{-1} + \sum_{k=0}^{s-2}(\eta_{k+1}^{-1}-\eta_k^{-1})\mathrm{E}_{s-k-j-1}^2X_s\right) \\ &= O\big(j^{-2\delta}\big). \end{aligned} \tag{3.9}$$

Noting that $\eta_k^{-1} = k\log^2 k$ so that $\eta_{k+1}^{-1}-\eta_k^{-1} = O(\log^2 k)$, the order of magnitude of (3.9) follows since $\mathrm{E}\mathrm{E}_{s-k-j}^2X_s \leq \|X_s\|_r^2\zeta(k+j)^2$ where $\sum_{s=1}^{n}\|X_s\|_r^2 = O(s_n^2)$ and $\zeta(k+j)^2 = O\big((k+j)^{-1-2\delta}\big)$ for $\delta > 0$, since the mixingale size is (at worst) $-\frac{1}{2}$. A third application of the Cauchy-Schwarz inequality for sums, similarly to (3.8), gives

$$\sum_{j=0}^{\infty}\left(\frac{1}{s_n^2}\sum_{t=1}^{n}p_{tj}^2\right)^{1/2}\eta_j^{-1/2}\eta_j^{1/2} \leq B^{1/2}\left(\sum_{j=0}^{\infty}\frac{1}{s_n^2}\sum_{t=1}^{n}p_{tj}^2\eta_j^{-1}\right)^{1/2} \tag{3.10}$$

and Abelian summation then gives

$$\begin{aligned}\sum_{j=0}^{\infty}\frac{1}{s_n^2}\sum_{t=1}^{n}p_{tj}^2\eta_j^{-1} &= \frac{1}{s_n^2}\sum_{t=1}^{n}\mathrm{E}\left(\mathrm{E}_t^2X_t + \sum_{j=0}^{\infty}(\eta_{j+1}^{-1}-\eta_j^{-1})\mathrm{E}_{t-j-1}^2X_t\right) \\ &= O(1). \end{aligned} \tag{3.11}$$

Putting (3.9) and (3.10) into (3.8) and applying (3.11) yields the result. ∎

Proof of Theorem 2.1 $\left|\hat{s}_n^2/s_n^2 - 1\right| \leq A_1 + A_2$ where

$$A_1 = \frac{1}{s_n^2}\left|\sum_{k=1-n}^{n-1}w_{nk}\sum_{t\in n(k)}(X_tX_{t+k}-\sigma_{t,t+k})\right|$$

$$A_2 = \frac{1}{s_n^2}\left|\sum_{k=1-n}^{n-1}(w_{nk}-1)\sum_{t\in n(k)}\sigma_{t,t+k}\right|$$

We show that $A_1 \to_{\text{pr}} 0$ and that $A_2 = o(1)$.

Lemma 2.2 and the assumptions imply that for $k \geq 0$, $X_tX_{t+k} - \sigma_{t,t+k}$ is a $L_{p/2}$ mixingale of size either $-\frac{1}{2}$ when $p \geq 4$ or -1 when $2 \leq p < 4$. In case (a) of Assumption 1, the Liapunov inequality and SLT16.10 give for each $k \geq 0$ and $K < \infty$,

$$\begin{aligned}\frac{1}{s_n^2}\mathrm{E}\left|\sum_{t\in n(k)}(X_tX_{t+k}-\sigma_{t,t+k})\right| &\leq K\frac{1}{s_n^2}\left|\sum_{t\in n(k)}\|X_t\|_p^2\|X_{t+k}\|_p^2\right|^{1/2} \\ &= O\big(n^{-1/2}\big). \end{aligned} \tag{3.12}$$

The order of magnitude of (3.12) is transparent in the stationary case with $\|X_t\|_r = O(1)$ and $s_n^2 \simeq n$. In the nonstationary case where $\sigma_{tt} \simeq t^\beta$ and $s_n^2 \simeq n^{1+\beta}$, the terms of the sum in the majorant are of order $t^{2\beta}$. In case (b) of Assumption 1 the mixingale is of order $p/2 < 2$, and in this case the Liapunov inequality and SLT16.11 give, for some $K < \infty$,

$$\frac{1}{s_n^2}\mathrm{E}\left|\sum_{t\in n(k)}(X_tX_{t+k} - \sigma_{t,t+k})\right| \le K\frac{1}{s_n^2}\Bigg(\sum_{t\in n(k)}\|X_t\|_r^{p/2}\|X_{t+k}\|_r^{p/2}\Bigg)^{2/p}$$
$$= O\big(n^{2/p-1}\big). \tag{3.13}$$

The order of magnitude in (3.13) is shown by similar reasoning to that of (3.12). Applying the limit in (2.3) which holds under Assumption 2, we obtain in case (a),

$$\mathrm{E}(A_1) = O\big(\gamma_n n^{-1/2}\big)$$

and in case (b),

$$\mathrm{E}(A_1) = O\big(\gamma_n n^{2/p-1}\big).$$

The assumption on γ_n implies $\mathrm{E}(A_1) = o(1)$ in each case, which is sufficient for convergence in probability.

Next, it follows by Lemma 2.3 that there exists a summable sequence of nonnegative weights $\{\mu(k), k \ge 1\}$, not depending on n, where $s_n^{-2}|\sum_{t\in n(k)}\sigma_{t,t+k}| \le \mu(k)$ for each k and n so that

$$A_2 \le \sum_{k=1-n}^{n-1}|w_{nk} - 1|\left|\frac{1}{s_n^2}\sum_{t\in n(k)}\sigma_{t,t+k}\right|$$
$$\le \sum_{k=1-n}^{n-1}|w_{nk} - 1|\mu(k). \tag{3.14}$$

The $\mu(k)$ define a finite measure that is absolutely continuous with respect to counting measure on the positive integers, and the majorant of (3.14) is the integral of the function $|w(k/\gamma_n) - 1|$ with respect to this measure. All but a finite number of these weights must be arbitrarily close to zero. Since $w(k/\gamma_n) \to 1$ as $n \to \infty$ for all finite k, and in view of Assumption 2(i), it follows by the dominated convergence theorem that $A_2 = o(1)$ as $n \to \infty$. This completes the proof. ∎

References

Andrews, Donald W. K. (1991), 'Heteroscedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* 59, 817-58.

Davidson, James (1994) *Stochastic Limit Theory*, Oxford University Press

Davidson, James and Robert M. de Jong (2002) 'Consistency of kernel variance estimators for sums of semiparametric linear processes' *Econometrics Journal* 5, 160-175.

de Jong, Robert M. (1997) 'Central limit theorems for dependent heterogeneous random variables' *Econometric Theory* 13, 353-367.

de Jong, Robert M. (2000) 'A Strong Consistency Proof for Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimators', *Econometric Theory* 16 (2), 262-268.

de Jong, Robert M. and James Davidson (2000) 'Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices' *Econometrica* 68, 407-424.

Gallant, A. Ronald and White Halbert (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* Basil Blackwell, Oxford.

Hansen, Bruce E. (1992b), 'Consistent covariance matrix estimation for dependent heterogeneous processes', *Econometrica* 60, 967-72.

Jansson, Michael (2002) 'Consistent Covariance Matrix Estimation for Linear Processes' *Econometric Theory* 18 (6) 1449-1459.

Newey, W. K. and West, K. (1987), 'A simple positive definite heteroskedasticity and correlation consistent covariance matrix', *Econometrica* 55, 703-8.

White, Halbert (1984), *Asymptotic Theory for Econometricians,* Academic Press, New York.