Econometrica, Vol. 60, No. 4 (July, 1992), 967–972

CONSISTENT COVARIANCE MATRIX ESTIMATION FOR DEPENDENT HETEROGENEOUS PROCESSES

BY BRUCE E. HANSEN[1]

1. INTRODUCTION

MANY TIME SERIES APPLICATIONS in econometrics—such as hypothesis tests from generalized method of moments estimation (Hansen (1982)) or other dynamic models (Andrews and Fair (1988))—require covariance matrix estimation robust to serial correlation and heteroskedasticity. Robust variance estimation also plays a large role in the recent "unit root" literature, including the unit root tests of Phillips (1987) and Phillips and Perron (1987), the cointegration tests of Phillips and Ouliaris (1990), and the fully modified estimator for cointegrated systems of Phillips and Hansen (1990).

Current consistency proofs which allow for general serial correlation require the restrictive assumption of finite fourth moments. See, for example, White (1984), Newey and West (1987), Gallant and White (1988), Keener, et. al. (1991), and Andrews (1991). Robinson (1991) allows for violation of finite fourth moments in the restrictive case that the data are covariance stationary and generated by a linear process with homoskedastic martingale difference innovations. These assumptions exclude many useful applications.

This note presents a simple consistency proof for general kernel-based covariance estimators, requiring the existence of only slightly more than second moments. Covariance stationarity is not required. Instead, the data are assumed to satisfy either an α-mixing or a φ-mixing condition. The proof is quite straightforward, relying on recent developments in the theory of mixingales. These results considerably broaden the scope for application of robust covariance matrix estimation.

2. CONSISTENCY

We stay largely in the notational framework of Andrews (1991). Consider a sequence of random vectors $\{V_t(\theta)\}$, which are (possibly) a function of an unknown parameter vector θ. We wish to estimate the "long-run" covariance matrix of the underlying process $V_t = V_t(\theta_o)$, where θ_o denotes the true (or pseudo-true) value of θ. This matrix is

$$\Omega = \lim_{T\to\infty} \sum_{i=1}^{T} \sum_{j=1}^{T} E(V_i V_j') < \infty.$$

Assume that θ has some consistent estimator $\hat{\theta}$, yielding the estimates $\hat{V}_t = V_t(\hat{\theta})$. Consider the class of kernel estimates of Ω given by

$$\hat{\Omega} = \sum_{j=-T}^{T} k(j/S_T)\hat{\Gamma}(j),$$

$$\hat{\Gamma}(j) = \frac{1}{T}\sum_{t=1}^{T-j} \hat{V}_t \hat{V}_{t+j}', \quad j \geq 0,$$

$$\hat{\Gamma}(j) = \hat{\Gamma}(-j)', \quad j < 0.$$

[1] I would like to thank two referees and Hal White for helpful comments, and the National Science Foundation for financial support.

The kernel weights $k(\cdot)$ are assumed to satisfy the following condition.

(K) For all $x \in \mathbb{R}$, $|k(x)| \leqslant 1$ and $k(x) = k(-x)$; $k(0) = 1$; $k(x)$ is continuous at zero and for almost all $x \in \mathbb{R}$; $\int_{\mathbb{R}} |k(x)| dx < \infty$.

Kernels which satisfy (K) include most considered in the literature. For example, the Bartlett kernel is given by: $k(x) = 1 - |x|$ for $|x| \leqslant 1$. Some authors refer to the number S_T as the band-width parameter, and others call S_T the lag truncation parameter. It is indexed by T to denote its dependence upon sample size. S_T is assumed to go to infinity as a power of sample size:

(S) $S_T \to \infty$, and for some $q \in (1/2, \infty)$, $S_T^{1+2q}/T = O(1)$.

Andrews (1991) documents[2] that for variables with finite fourth moments, the kernel will dictate the appropriate choice of q. For the Bartlett kernel, $q = 1$, and for the Parzen and Quadratic Spectral (QS) kernels, $q = 2$.

It is most convenient if we start by analyzing the idealized estimator

$$\tilde{\Omega} = \sum_{j=-T}^{T} k(j/S_T)\tilde{\Gamma}(j)$$

where $\tilde{\Gamma}(j)$ is defined as $\hat{\Gamma}(j)$ but with the random variables V_t instead of the estimates $\hat{V}_t$.

Assume that $E(V_t) = 0$. Let $\|V_t\|_\beta = (\Sigma_i E|V_{ti}|^\beta)^{1/\beta}$, let $\{\alpha_m\}_{m=1}^\infty$ denote the α-mixing (strong mixing) coefficients for $\{V_t\}_{t=1}^\infty$, and let $\{\varphi_m\}_{m=1}^\infty$ denote the φ-mixing coefficients for $\{V_t\}$. Define $\mathscr{F}_t = \sigma(V_i\colon i \leqslant t)$, the sigma-field generated by the past history of $\{V_t\}$, and set $E_{t-m}X = E(X|\mathscr{F}_{t-m})$. We start with mixingale-type bounds[3] for the products V_tV_{t+j}':

Lemma 1: *For* $j \geqslant 0$, $\gamma > \beta \geqslant 1$,

(a) $\|E_{t-m}V_tV'_{t+j} - EV_tV'_{t+j}\|_\beta \leqslant 12\alpha_m^{1/\beta - 1/\gamma}\|V_t\|_{2\gamma}\|V_{t+j}\|_{2\gamma}$,

(b) $\|E_{t-m}V_tV'_{t+j} - EV_tV'_{t+j}\|_\beta \leqslant 4\varphi_m^{1-1/\gamma}\|V_t\|_{2\gamma}\|V_{t+j}\|_{2\gamma}$.

We assume the following conditions are satisfied by the process $\{V_t\}$:

(V1) For some $r \in (2,4]$ such that $r > 2 + 1/q$, and some $p > r$,

(i) $12\sum_{m=1}^{\infty} \alpha_m^{2(1/r - 1/p)} = A < \infty$, or $4\sum_{m=1}^{\infty} \varphi_m^{1-2/p} = A < \infty$,

(ii) $\sup_{t \geqslant 1} \|V_t\|_p = C < \infty$.

(V1) is sufficient to uniformly bound the autocovariance matrix estimators:

Lemma 2: *(V1) implies that* $\|\tilde{\Gamma}(j) - \Gamma(j)\|_{r/2} \leqslant (36A[r/(r-2)]^{3/2}C^2)T^{-1+2/r}$.

Given these inequalities, it is straightforward to establish our main result.

Theorem 1: *Under (K), (S), and (V1),* $\tilde{\Omega} \to_p \Omega$ *as* $T \to \infty$.

[2] Andrews (1991) uses an asymptotic truncated mean square error criterion.
[3] Mixingales in L^2 were introduced by McLeish (1975), and extended to L^β, $\beta \geqslant 1$, by Andrews (1988).

To analyze $\hat{\Omega}$, we will invoke Theorem 1(a) of Andrews (1991). Set $\mathscr{N}$ to be some neighborhood of θ_o. Let $\|\cdot\|$ denote the Euclidean norm.

(V2)(i) $\sqrt{T}\left(\hat{\theta}-\theta_o\right)=O_p(1)$;

(ii) $$\sup_{t\geqslant 1} E\left(\sup_{\theta\in\mathscr{N}} \|V_t(\theta)\|^2\right)<\infty;$$

(iii) $$\sup_{t\geqslant 1} E\left(\sup_{\theta\in\mathscr{N}} \left\|\frac{\partial}{\partial\theta'}V_t(\theta)\right\|^2\right)<\infty.$$

THEOREM 2: *Under* (K), (S), $(V1)$, *and* $(V2)$, $\hat{\Omega}\to_p \Omega$ *as* $T\to\infty$.

This is the first set of results which do not require finite fourth moments. In the inequality $p>2+1/q$, a trade-off appears between finite moments and the rate of divergence for the band-width parameter. If S_T grows just more slowly than $T^{1/2}$, then q is just larger than $1/2$ and we require $p>4$. If S_T grows like $T^{1/3}$ (as Andrews (1991) recommends for the Bartlett kernel) then $q=1$ and we require $p>3$. If S_T grows like $T^{1/5}$ (as Andrews (1991) recommends for the Parzen and QS kernels) then $q=2$ and we require $p>5/2$. This provides an argument for the choice of the Parzen or QS kernel rather than the more popular Bartlett kernel. It is important to remember that the recommendations of Andrews (1991) were based upon the criterion of asymptotic mean squared error, and therefore are not necessarily applicable when $p<4$. Hence the optimality of the bandwidth choices cannot be extended to the present context.

Even for the case of finite fourth moments, Theorems 1 and 2 are less restrictive than the existing literature. First note that (V1) (i) is implied if α_m are of size $-rp/[2(p-r)]$. For S_T growing close to $T^{1/2}$ (so q is close to $1/2$), then we have to set $r=4$ and thus require that α_m be of size $-2p/(p-4)$, which is the size requirement used by Newey and West (1987). They required, however, the strict condition $S_T=o(T^{1/4})$. On the other hand, if we specify that S_T grows like $T^{1/5}$, then $q=2$ and we need $r>5/2$. We could then set $r=3$, which requires that α_m be of size $-3p/[2(p-3)]$, which is substantially weaker than the conditions of Newey and West.

Condition (V2) allows for covariance matrix estimation for weakly dependent, heterogeneous data without stochastic or deterministic trends. As mentioned in the introduction, some of the literature on unit roots and cointegration makes heavy use of covariance matrix estimation. In these cases, (V2) does not apply. We can instead use the following condition. Let $\{\delta_T\}$ be a sequence of nonsingular matrices.

(V3)(i) $V_t(\theta)=V_t-\left(\hat{\theta}-\theta_o\right)X_t$;

(ii) $$\sup_{t\leqslant T}\|\delta_T X_t\|=O_p(1);$$

(iii) $\sqrt{T}\left(\hat{\theta}-\theta_o\right)\delta_T^{-1}=O_p(1)$.

THEOREM 3: *Under* (K), (S), $(V1)$, *and* $(V3)$, $\hat{\Omega}\ \overrightarrow{_p}\ \Omega$ *as* $T\to\infty$.

Theorem 3 allows for trended regressors of general form in linear regression. For example, if $X_t=t$, then set $\delta_T=T^{-1}$, while if X_t is an $I(1)$ process, then set $\delta_T=T^{-1/2}$.

3. PROOFS

PROOF OF LEMMA 1: To show (a), by McLeish's α-mixing inequality (McLeish (1975, Lemma 2.1)), Minkowski's Inequality, Blackwell's Theorem, and Holder's Inequality,

$$\begin{aligned}
&\|E_{t-m}V_tV'_{t+j} - EV_tV'_{t+j}\|_\beta \\
&\quad \leqslant 6\alpha_m^{1/\beta-1/\gamma}\|V_tV'_{t+j} - EV_tV'_{t+j}\|_\gamma \\
&\quad \leqslant 6\alpha_m^{1/\beta-1/\gamma}\left(\|V_tV_{t+j}\|_\gamma + \|EV_tV_{t+j}\|_\gamma\right) \\
&\quad \leqslant 12\alpha_m^{1/\beta-1/\gamma}\|V_tV_{t+j}\|_\gamma \\
&\quad \leqslant 12\alpha_m^{1/\beta-1/\gamma}\|V_t\|_{2\gamma}\|V_{t+j}\|_{2\gamma}.
\end{aligned}$$

The argument for (b) is the same, expect that the first inequality is replaced by Serfling's inequality (Serfling (1968, Theorem 2.2)):

$$\|E_{t-m}V_tV'_{t+j} - EV_tV'_{t+j}\|_\beta \leqslant 2\varphi^{1-1/\gamma}\|V_tV'_{t+j} - EV_tV'_{t+j}\|_\gamma \qquad \text{Q.E.D.}$$

The proof of Lemma 2 will rely on the following result from Hansen (1991, 1992):

LEMMA A: *For a sequence of random variables $\{X_t\}_{t=1}^\infty$, set $\mathscr{F}_t = \sigma(X_i\colon 1 \leqslant i \leqslant t)$. If for some $\beta > 1$, all $t \geqslant 1$, and all $m \geqslant 1$, $\|E(X_t \mid \mathscr{F}_{t-m})\|_\beta \leqslant c_t\psi_m$, and $\Psi = \sum_{m=1}^\infty \psi_m < \infty$, then*

$$\left\|\sum_{t=1}^T X_t\right\|_\beta \leqslant 36\Psi[\beta/(\beta-1)]^{3/2}\left(\sum_{t=1}^T c_t^\alpha\right)^{1/\alpha}, \quad \text{where} \quad \alpha = \min(\beta, 2).$$

PROOF OF LEMMA 2: By Lemma 1, $\|E_{t-m}(V_tV'_{t+j} - EV_tV'_{t+j})\|_{r/2} \leqslant c_t\psi_m$ where $c_t = \|V_t\|_p\|V_{t+j}\|_p$ and $\psi_m = 12\alpha_m^{2(1/r-1/p)}$ or $4\varphi_m^{1-2/p}$. Thus by Lemmas 1 and A, and assumption (V1),

$$\begin{aligned}
\|\tilde{\Gamma}(j) - E\tilde{\Gamma}(j)\|_{r/2} &= \left\|\frac{1}{T}\sum_{t=1}^{T-j}(V_tV'_{t+j} - EV_tV'_{t+j})\right\|_{r/2} \\
&\leqslant \frac{1}{T}36\left(\sum_1^\infty \psi_m\right)[r/(r-2)]^{3/2}\left(\sum_1^{T-j}\|V_t\|_p^{r/2}\|V_{t+j}\|_p^{r/2}\right)^{2/r} \\
&\leqslant T^{-1+2/r}36A[r/(r-2)]^{3/2}C^2
\end{aligned}$$

which is the desired result. Q.E.D.

PROOF OF THEOREM 1: By Minkowski's Inequality and Lemma 2,

$$\begin{aligned}
\text{(A1)} \quad & S_T^{-1}T^{1-2/r}\|\tilde{\Omega} - E\tilde{\Omega}\|_{r/2} \\
&= S_T^{-1}T^{1-2/r}\left\|\sum_{j=-T}^T k(j/S_T)\left(\tilde{\Gamma}(j) - E\tilde{\Gamma}(j)\right)\right\|_{r/2} \\
&\leqslant S_T^{-1}\sum_{j=-T}^T |k(j/S_T)|T^{1-2/r}\|\tilde{\Gamma}(j) - E\tilde{\Gamma}(j)\|_{r/2} \\
&\leqslant \int_{\mathbb{R}}|k(x)|\,dx\,36A[r/(r-2)]^{3/2}C^2 < \infty,
\end{aligned}$$

uniformly in T. In addition

$$\text{(A2)}\qquad S_T T^{2/r-1} = S_T T^{-1/(2q+1)} T^{1/(2q+1)+2/r-1} = O(1) T^{2[1/r-1/(2+1/q)]} = o(1),$$

since $r > 2 + 1/q$. Thus $S_T^{-1} \cdot T^{1-2/r} \to \infty$. (A1) and (A2) yield $\|\tilde{\Omega} - E\tilde{\Omega}\|_{r/2} \to 0$, and hence $\tilde{\Omega} - E\tilde{\Omega} \to_p 0$ by Markov's inequality. Lemma 6.6 of Gallant and White (1988), shows that $E\tilde{\Omega} \to \Omega$, which completes the proof. *Q.E.D.*

PROOF OF THEOREM 2: The proof of Theorem 1(a) in Andrews (1991) only uses the conditions on $\{V_t\}$ to establish $\tilde{\Omega} \to_p \Omega$. We can instead use Theorem 1 to establish this result under our conditions. Condition (V2) is equivalent to Andrews' Assumption B, and the remainder of his proof carries over to the present case. *Q.E.D.*

PROOF OF THEOREM 3:

$$\hat{\Omega} - \tilde{\Omega} = \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t \left(\hat{V}_t \hat{V}_{t+j}' - V_t V_{t+j}' \right)$$

$$= \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t \left(-(\hat{\theta} - \theta_o) X_t V_{t+j}' - V_t X_{t+j}' (\hat{\theta} - \theta_o)' + (\hat{\theta} - \theta_o) X_t X_{t+j}' (\hat{\theta} - \theta_o)' \right).$$

By Minkowski's inequality,

$$\text{(A3)}\qquad \frac{\sqrt{T}}{S_T} \|\hat{\Omega} - \tilde{\Omega}\| \leqslant \frac{\sqrt{T}}{S_T} \left\| \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t (\hat{\theta} - \theta_o) X_t V_{t+j}' \right\| + \frac{\sqrt{T}}{S_T} \left\| \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t V_t X_{t+j}' (\hat{\theta} - \theta_o)' \right\| + \frac{\sqrt{T}}{S_T} \left\| \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t (\hat{\theta} - \theta_o) X_t X_{t+j}' (\hat{\theta} - \theta_o)' \right\|.$$

We will show that (A3) is $O_p(1)$. Since $\sqrt{T}/S_{\mathrm{T}} \to \infty$, it follows that $\hat{\Omega} - \tilde{\Omega} = o_p(1)$, and by Theorem 2, $\hat{\Omega} \to_p \Omega$.

First,

$$\text{(A4)}\qquad \frac{\sqrt{T}}{S_T} \left\| \sum_{j=-T}^{T} k(j/S_T) \frac{1}{T} \sum_t (\hat{\theta} - \theta_o) X_t V_{t+j}' \right\| \leqslant \frac{\sqrt{T}}{S_T} \sum_{j=-T}^{T} |k(j/S_T)| \left\| \frac{1}{T} \sum_t (\hat{\theta} - \theta_o) \delta_T^{-1} \delta_T X_t V_{t+j}' \right\| \leqslant \int_{\mathbb{R}} |k(x)|\, dx \|\sqrt{T} (\hat{\theta} - \theta_o) \delta_T^{-1}\| \left\| \frac{1}{T} \sum_t \delta_T X_t X_t' \delta_T' \right\| \left\| \frac{1}{T} \sum_t V_t V_t' \right\| = O_p(1).$$

Second,

$$\text{(A5)} \qquad \frac{\sqrt{T}}{S_T}\left\| \sum_{j=-T}^{T} k(j/S_T)\frac{1}{T}\sum_t V_t X_{t+j}'(\hat{\theta}-\theta_o)' \right\| = O_p(1).$$

Finally,

$$\text{(A6)} \qquad \frac{\sqrt{T}}{S_T}\left\| \sum_{j=-T}^{T} k(j/S_T)\frac{1}{T}\sum_t (\hat{\theta}-\theta_o)X_t X_{t+j}'(\hat{\theta}-\theta_o)' \right\|$$

$$\leqslant \frac{\sqrt{T}}{S_T}\sum_{j=-T}^{T}|k(j/S_T)|\left\| \frac{1}{T}\sum_t (\hat{\theta}-\theta_o)\delta_T^{-1}\delta_T X_t X_{t+j}'\delta_T'\delta_T'^{-1}(\hat{\theta}-\theta_o)' \right\|$$

$$\leqslant \int_{\mathbb{R}}|w(x)|\,dx \left\|\sqrt{T}(\hat{\theta}-\theta_o)\delta_T^{-1}\right\| \left\| \frac{1}{T}\sum_t \delta_T X_t X_t'\delta_T' \right\|^2 \left\|(\hat{\theta}-\theta_o)\delta_T^{-1}\right\|$$

$$= o_p(1).$$

(A4), (A5), and (A6) show that (A3) is $O_p(1)$, completing the proof. *Q.E.D.*

Dept. of Economics, University of Rochester, Rochester, N.Y. 14627, U.S.A.

Manuscript received December, 1990; final revision received December, 1991.

REFERENCES

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

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

ANDREWS, D. W. K., AND R. C. FAIR (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615–639.

GALLANT, A. R., AND H. WHITE (1988): *A Unified Theory of Estimation and Inference for Nonlinear Econometric Models*. New York: Basil Blackwell.

HANSEN, B. E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221; and (1992): "Errata," *Econometric Theory*, 8, (forthcoming).

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

KEENER, R. W., J. KMENTA, AND N. C. WEBER (1991): "Estimation of the Covariance Matrix of the Least-Squares Regression Coefficients when the Disturbance Covariance Matrix is of Unknown Form," *Econometric Theory*, 7, 22–45.

MCLEISH, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 5, 829–839.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

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.

PHILLIPS, P. C. B., AND S. OULIARIS (1990): "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.

PHILLIPS, P. C. B., AND P. PERRON (1987): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

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

SERFLING, R. J. (1968): "Contributions to Central Limit Theory for Dependent Variables," *Annals of Mathematical Statistics*, 39, 1158–1175.

WHITE, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.