Econometrica, Vol. 60, No. 4 (July, 1992), 953–966

AN IMPROVED HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATOR

BY DONALD W. K. ANDREWS AND J. CHRISTOPHER MONAHAN [1]

1. INTRODUCTION

THIS PAPER CONSIDERS A NEW CLASS of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators. The estimators considered are prewhitened kernel estimators with vector autoregressions (VARs) employed in the prewhitening stage. The paper establishes consistency of the estimators when a fixed or automatic bandwidth procedure is employed. Monte Carlo results show that prewhitening is very effective in reducing bias, improving confidence interval coverage probabilities, and reducing over-rejection of t statistics constructed using kernel HAC estimators. On the other hand, prewhitening is found to inflate the variance and MSE of the kernel estimators. Since confidence interval coverage probabilities and over-rejection of t statistics are often of primary concern, prewhitened kernel estimators provide a significant improvement over the standard kernel estimators.

Considerable attention has been paid in recent years to HAC covariance matrix estimation; see L. P. Hansen (1982), Levine (1983), White (1984), Gallant (1987), Newey and West (1987), Andrews (1991), Robinson (1991), Keener, Kmenta, and Webber (1991), Wooldridge (1991), and B. E. Hansen (1992). As shown in the Monte Carlo results of Andrews (1991), however, the kernel estimators considered in the above papers all perform quite poorly in certain contexts. In particular, kernel HAC covariance matrix estimators often yield confidence intervals whose coverage probabilities are too low (equivalently, test statistics that reject too often) and this phenomenon is not attributable to a particular choice of kernel or bandwidth parameter. The problem is especially severe when there is considerable temporal dependence in the data. This finding suggests that the standard class of kernel HAC estimators is too restrictive and that one needs to consider a larger class of estimators if an improved HAC estimator is to be found. In this paper, we consider such a class, viz. the class of VAR prewhitened kernel HAC estimators.

Prewhitening has a long history in the time series literature and dates from the work of Press and Tukey (1956). Additional references include Blackman and Tukey (1958) and Grenander and Rosenblatt (1957). The idea behind prewhitening is as follows: Suppose one is nonparametrically estimating a function $f(\lambda)$ at λ_0 by taking unbiased estimates of $f(\lambda)$ at a number of points λ in a neighborhood of λ_0 and averaging them. If the function $f(\lambda)$ is flat in this neighborhood, then this procedure yields an unbiased estimator of $f(\lambda_0)$. If $f(\lambda)$ is not flat in this neighborhood, however, then the procedure is biased and the magnitude of the bias depends on the degree of nonconstancy of $f(\lambda)$.

Suppose the data and the function $f(\lambda)$ can be transformed such that the transformed function $f^*(\lambda)$ is flatter in the neighborhood of λ_0 than is $f(\lambda)$. Then, using the transformed data, one can estimate $f^*(\lambda_0)$ by averaging unbiased estimates of $f^*(\lambda)$ at points λ in the neighborhood of λ_0. The bias incurred by doing so should be less than that incurred by estimating $f(\lambda)$ as described above, since $f^*(\lambda)$ is flatter than $f(\lambda)$. Finally, one can apply the inverse of the transformation from $f(\lambda)$ to $f^*(\lambda)$ to obtain an estimator of $f(\lambda)$ from the estimator of $f^*(\lambda)$.

[1] The authors thank Allan Gregory, Jim Nason, Masao Ogaki, Peter Phillips, two referees, and the co-editor for helpful comments. The first author gratefully acknowledges the research support of the National Science Foundation via Grant Nos. SES-8821021 and SES-9040137.

In the time series literature, the idea of prewhitening has been applied to nonparametric estimators of the spectral density function. In this case, one tries to transform (filter) the data in such a way that the transformed data is uncorrelated, since an uncorrelated sequence has a flat spectral density function. The estimand of interest in this paper is just the spectral density function at frequency zero in the special case where the observations are second order stationary and no parameters are estimated. Thus, it is natural to consider using a prewhitening procedure that attempts to transform the data into an uncorrelated sequence before applying a kernel estimator when constructing a HAC covariance matrix estimator.

The remainder of this note is organized as follows: Section 2 defines the estimand of interest, introduces the VAR prewhitening procedure, and presents the consistency results for the prewhitened HAC estimators. Section 3 describes a Monte Carlo experiment that is designed to assess the effectiveness of prewhitening. Throughout the paper, all limits are taken as $T \to \infty$.

2. VAR PREWHITENED HAC ESTIMATORS

First, we introduce the estimand of interest. Many parametric estimators $\hat{\theta}$ in nonlinear dynamic models satisfy

$$(2.1) \qquad (B_T J_T B_T')^{-1/2}\sqrt{T}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N(\mathbf{0}, I_\zeta), \qquad \text{where}$$

$$J_T = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EV_s(\theta_0)V_t(\theta_0)',$$

B_T is a nonrandom $\zeta \times p$ matrix, and $V_t(\theta)$ is a random p-vector for each $\theta \in \Theta \subset R^\zeta$. Usually it is easy to construct estimators $\hat{B}_T$ of B_T such that $\hat{B}_T - B_T \xrightarrow{p} 0$. The sample analogue of B_T with θ_0 replaced by $\hat{\theta}$ is usually sufficient. Thus, one can consistently estimate the "asymptotic variance" of $\sqrt{T}(\hat{\theta}-\theta_0)$, viz., $B_T J_T B_T'$, if one has a consistent estimator of J_T. It is the estimation of J_T that concerns us here.

A second scenario where estimands of the form J_T arise quite frequently is in the analysis of linear models with deterministic and/or stochastic trends, in particular, unit root and cointegration models; e.g., see Phillips (1987). The estimators that we consider here are also applicable in these scenarios.

We define a class of VAR prewhitened HAC estimators of J_T as follows: Suppose $\hat{\theta}$ is a $\sqrt{T}$-consistent estimator of θ_0. First, one estimates a bth order VAR model for $V_t(\hat{\theta})$:

$$(2.2) \qquad V_t(\hat{\theta}) = \sum_{r=1}^{b} \hat{A}_r V_{t-r}(\hat{\theta}) + V_t^*(\hat{\theta}) \qquad \text{for} \qquad t = b+1, \ldots, T,$$

where $\hat{A}_r$ for $r = 1, \ldots, b$ are $p \times p$ parameter estimates and $\{V_t^*(\hat{\theta})\colon t = b+1, \ldots, T\}$ are the corresponding residual vectors. For example, $\{\hat{A}_r\colon r = 1, \ldots, b\}$ could be the least squares (LS) estimators.[2] The estimated VAR model is not meant to be an estimate of a true model. It is used as a tool to "soak up" some of the temporal dependence in $\{V_t(\hat{\theta})\}$ and to leave one with residuals $\{V_t^*(\hat{\theta})\}$ that are closer to white noise than are the rv's $\{V_t(\hat{\theta})\}$.

[2] We suggest defining the estimators $\{\hat{A}_r\}$ in such a way as to ensure that $I_p - \sum_{r=1}^{b}\hat{A}_r$ is not too close to singularity. For example, for the Monte Carlo results of Section 3, we use an eigenvalue adjusted version of the LS estimator. See Section 3 for details.

Second, one computes a standard kernel HAC estimator, call it $\hat{J}_T^*(\hat{S}_T)$, based on the VAR residual vectors $\{V_t^*(\hat{\theta})\}$. Let

$$(2.3)\qquad \hat{J}_T^*(\hat{S}_T)=\frac{T}{T-\zeta}\sum_{j=-T+1}^{T-1}k\left(\frac{j}{\hat{S}_T}\right)\hat{\Gamma}^*(j),\qquad \text{where}$$

$$\hat{\Gamma}^*(j)=\begin{cases}\dfrac{1}{T}\displaystyle\sum_{t=j+1}^{T}\hat{V}_t^*\hat{V}_{t-j}^{*\prime} & \text{for } j\geqslant 0,\\ \dfrac{1}{T}\displaystyle\sum_{t=-j+1}^{T}\hat{V}_{t+j}^*\hat{V}_t^{*\prime} & \text{for } j<0,\end{cases}$$

$\hat{V}_t^*=V_t^*(\hat{\theta})$, $k(\cdot)$ is a real-valued kernel in the set $\mathscr{K}_1$ defined below, and $\hat{S}_T$ is a data-dependent bandwidth parameter. The factor $T/(T-\zeta)$ is a small sample degrees of freedom adjustment that is introduced to offset the effect of estimation of the ζ-vector θ_0.

Third, one recolors the estimator $\hat{J}_T^*(\hat{S}_T)$ to obtain the VAR prewhitened kernel estimator of J_T:

$$(2.4)\qquad \hat{J}_{Tpw}(\hat{S}_T)=\hat{D}\hat{J}_T^*(\hat{S}_T)\hat{D}',\qquad \text{where}\quad \hat{D}=\left(I_p-\sum_{r=1}^{b}\hat{A}_r\right)^{-1}.$$

Once a kernel k is chosen and a data-dependent bandwidth $\hat{S}_T$ is specified (e.g., as in Section 3 below), this yields an operational VAR prewhitened kernel estimator of J_T.

We mention two reasons why we have chosen a VAR model to do the prewhitening. First, VARs have been found in the econometrics literature to yield reasonable approximations of a wide variety of vector-valued time series processes. Second, it has been found in the statistical literature that autoregressive spectral density estimators provide reasonable estimators of the spectral density functions of more general stationary time series processes; see Parzen (1984) for references. Of course, if prior information suggests that a model different from a VAR model may give a better approximation in a given situation, then it may be preferable to use this model to do the prewhitening.

Next, we establish the consistency of prewhitened kernel HAC estimators. We consider the following classes of kernels:

$$(2.5)\qquad \mathscr{K}_1=\{k(\cdot)\colon R\to[-1,1]\,|\,k(0)=1,\ k(x)=k(-x)\,\forall x\in R,\ \int_{-\infty}^{\infty}|k(x)|\,dx<\infty,\ k(\cdot)\text{ is continuous at 0 and at all but a finite number of other points}\}\ \text{and}$$

$$\mathscr{K}_3=\left\{k(\cdot)\in\mathscr{K}_1\colon (i)\ |k(x)|\leqslant C_1|x|^{-B}\text{ for some } B>1+\frac{1}{q}\text{ and some } C_1<\infty,\text{ where } q\in(0,\infty)\text{ is such that } k_q\equiv\lim_{x\to 0}(1-k(x))/|x|^q\in(0,\infty),\ \text{and}(ii)|k(x)-k(y)|\leqslant C_2|x-y|\,\forall x,y\in R\text{ for some constant } C_2<\infty\right\}.$$ [3]

[3] The classes $\mathscr{K}_1$ and $\mathscr{K}_3$ are numbered to correspond to the numbering in Andrews (1991). The class $\mathscr{K}_2$ of Andrews (1991) is not considered here. Note that $\mathscr{K}_1$ in Andrews (1991) should be defined with $\int k^2(x)\,dx<\infty$ replaced by $\int|k(x)|\,dx<\infty$, as it is here.

$\mathscr{K}_1$ contains the quadratic spectral (QS), truncated, Bartlett, Parzen, and Tukey-Hanning kernels among others. $\mathscr{K}_3$ contains all of these kernels except the truncated kernel. For the QS, Parzen, and Tukey-Hanning kernels, q is 2. For the Bartlett kernel, q is 1. For fixed sequences of bandwidth parameters our consistency results hold for all kernels in $\mathscr{K}_1$. For data-dependent sequences $\{\hat{S}_T\}$, they hold for all kernels in $\mathscr{K}_3$.

Let $V_t = V_t(\theta_0)$. Let $\kappa_{abcd}(t, t+j, t+l, t+n)$ denote the fourth order cumulant of $(V_{at}, V_{bt+j}, V_{ct+l}, V_{dt+n})$, where V_{at} denotes the ath element of V_t. Let $f^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \sup_{t \geq 1} \|EV_t V'_{t+|j|}\|$. Let $\lambda_{\max}(A)$ denote the maximum eigenvalue of A. Let $\|A\|$ denote the Euclidean norm of a vector or matrix A (i.e., the square root of the sum of squares of its elements).

We now introduce a number of assumptions from Andrews (1991) plus one from B. E. Hansen (1992). See Andrews (1991) for discussion of these assumptions.

Assumption A: *$\{V_t\}$ is a mean zero sequence of rv's with $\sum_{j=0}^{\infty} \sup_{t \geq 1} \|EV_t V'_{t-j}\| < \infty$ and $\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \sum_{n=1}^{\infty} \sup_{t \geq 1} |\kappa_{abcd}(t, t+j, t+l, t+n)| < \infty$ $\forall a, b, c, d \leq p$.*

Assumption B: *Either (i) $\sqrt{T}(\hat{\theta} - \theta_0) = O_p(1)$ and $\sup_{t \geq 1} E \sup_{\theta \in \Theta} \|\partial/\partial\theta' V_t(\theta)\|^2 < \infty$, where $\Theta \subset R^{\zeta}$ is some neighborhood of θ_0, or (ii) $V_t(\theta) = V_t - (\hat{\theta} - \theta_0) X_t$, $\sup_{t \leq T} \|\Delta_T X_t\| = O_p(1)$, and $\sqrt{T}(\hat{\theta} - \theta_0)\Delta_T^{-1} = O_p(1)$, where $\{\Delta_T \colon T \geq 1\}$ is some sequence of nonrandom nonsingular matrices.*

Assumption C: *$\hat{S}_T = \hat{\alpha} T^{1/(2q+1)}$, where $\hat{\alpha}$ satisfies $\hat{\alpha} = O_p(1)$ and $1/\hat{\alpha} = O_p(1)$.*

Assumption D: *(i) $\sqrt{T}(\hat{A}_r - A_r) = O_p(1)$ for some $A_r \in R^{p \times p}$ $\forall r = 1, \ldots, b$. (ii) $I_p - \sum_{r=1}^{b} A_r$ is nonsingular.*

As shown in Andrews (1991, Lemma 1), Assumption A and $f^{(q)} < \infty$ are implied by α-mixing and moment conditions. Part (i) of Assumption B is used for nonlinear dynamic models without deterministic or stochastic trends. Part (ii) of Assumption B is used for linear models with trends and is due to B. E. Hansen (1992). The rate of growth specified in Assumption C of the automatic bandwidth parameter, viz. $T^{1/(2q+1)}$, is the optimal rate determined in Andrews (1991). A particular sequence of data-dependent bandwidth parameters that satisfies Assumption C is described in Section 3 below. Assumption D is the only assumption that is not used in the proof of consistency of standard kernel HAC estimators.

The main result of this section is the following:

Theorem 1: *Suppose $k \in \mathscr{K}_3$ with $q > 1/2$, $f^{(q)} < \infty$, and Assumptions A–D hold. Then $\hat{J}_{Tpw}(\hat{S}_T) - J_T \xrightarrow{p} 0$.*

Comment: Additional asymptotic properties of prewhitened kernel HAC estimators are established in Andrews and Monahan (1990). In particular, prewhitened kernel estimators are shown to converge to the estimand at the same rate as standard kernel estimators, their asymptotic truncated mean square error is established, and conditions are given under which prewhitening improves asymptotic truncated MSE.

3. MONTE CARLO RESULTS

In this section, Monte Carlo methods are used to evaluate the performance of the VAR prewhitened HAC estimator introduced above. We compare the prewhitened HAC estimator to the nonprewhitened HAC estimator and to a parametric estimator.

We consider several linear regression models, each with an intercept and four regressors, and the least squares (LS) estimators $\hat{\theta}$ for each of these models:

$$(3.1) \qquad Y_t = X_t'\theta_0 + U_t, \qquad t = 1,\ldots,T, \qquad \hat{\theta} = \left[\sum_1^T X_t X_t'\right]^{-1} \sum_1^T X_t Y_t, \qquad \text{and}$$

$$\operatorname{Var}\left(\sqrt{T}(\hat{\theta} - \theta_0)|X\right) = \left(\frac{1}{T}\sum_1^T X_t X_t'\right)^{-1} \frac{1}{T}\sum_{s=1}^T \sum_{t=1}^T E(U_s X_s U_t X_t'|X) \left(\frac{1}{T}\sum_1^T X_t X_t'\right)^{-1}.$$

The estimand of interest is the variance (conditional on $X = (X_1, \ldots, X_T)'$) of the LS estimator of the first nonconstant regressor. (That is, the estimand is the second diagonal element of $\operatorname{Var}(\sqrt{T}(\hat{\theta} - \theta_0)|X)$.)

Seven basic regression models are considered: AR(1)-HOMO, in which the errors and regressors are homoskedastic AR(1) processes; AR(1)-HET1 and AR(1)-HET2, in which the errors and regressors are AR(1) processes with multiplicative heteroskedasticity overlaid on the errors; MA(1)-HOMO, in which the errors and regressors are homoskedastic MA(1) processes; MA(1)-HET1 and MA(1)-HET2, in which the errors and regressors are MA(1) processes with multiplicative heteroskedasticity overlaid on the errors; and MA(m)-HOMO, in which the errors and regressors are homoskedastic MA(m) processes with linearly declining MA parameters. (Details are given below.) A range of different parameter values is considered for each model. Each parameter value corresponds to a different degree of autocorrelation.

Three variance estimators are considered. The first, denoted QS-PW, is a prewhitened kernel HAC estimator defined using the QS kernel (defined below), a first-order VAR prewhitening procedure ($b = 1$), and an automatic bandwidth procedure (defined below). The underlying rv's $\{V_t(\hat{\theta})\}$, upon which QS-PW is constructed (see (2.2)), are defined by $V_t(\hat{\theta}) = (Y_t - X_t'\hat{\theta})X_t$. The dimension, p, of $V_t(\hat{\theta})$ is 5. The $p \times p$ estimator $\hat{A}$ $(=\hat{A}_1)$ that is used to carry out the prewhitening is defined as follows: Let $\hat{A}_{LS}$ denote the LS estimator from the regression of $V_t(\hat{\theta})$ on $V_{t-1}(\hat{\theta})$ for $t = 2, \ldots, T$. The LS estimator $\hat{A}_{LS}$ is adjusted using its singular value decomposition to obtain an estimator $\hat{A}$ for which $I_p - \hat{A}$ is not too close to singularity. In particular, let $\hat{B}$ and $\hat{C}$ denote $p \times p$ orthogonal matrices whose columns are eigenvectors of $\hat{A}_{LS}\hat{A}_{LS}'$ and $\hat{A}_{LS}'\hat{A}_{LS}$, respectively. Let $\hat{\Delta}_{LS}$ be the diagonal $p \times p$ matrix defined by $\hat{\Delta}_{LS} = \hat{B}'\hat{A}_{LS}\hat{C}$. By construction, $\hat{A}_{LS} = \hat{B}\hat{\Delta}_{LS}\hat{C}'$. Let $\hat{\Delta}$ be the $p \times p$ diagonal matrix constructed from $\hat{\Delta}_{LS}$ by replacing any element of $\hat{\Delta}_{LS}$ that exceeds .97 by .97 and any element that is less than $-.97$ by $-.97$. Then, let $\hat{A} = \hat{B}\hat{\Delta}\hat{C}'$.[4] Note that this eigenvalue adjustment has no effect on the consistency of the estimator QS-PW. QS-PW is consistent even if the probability limit of $\hat{A}_{LS}$ has some eigenvalues closer to 1 than .97.

The QS kernel is defined by

$$(3.2) \qquad k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

The QS kernel yields an estimator $\hat{J}_{Tpw}(\hat{S}_T)$ that is necessarily positive semi-definite.

[4] This adjustment procedure guarantees that the eigenvalues of $I_p - \hat{A}$ are $\geqslant .03$ in absolute value. To see this, suppose λ is an eigenvalue of $\hat{A}$ with corresponding eigenvector x. Then $\hat{A}x = \lambda x$ and $\|x\| = 1$. Thus, we have $|\lambda| = \|\hat{A}x\| = (x'\hat{A}'\hat{A}x)^{1/2} = (x'\hat{C}'\hat{\Delta}^2\hat{C}x)^{1/2} = (y'\hat{\Delta}^2 y)^{1/2} = (\sum_{j=1}^p \hat{\delta}_j^2 y_j^2)^{1/2} \leqslant .97(y'y)^{1/2} = .97$, where $y = \hat{C}x$ and $y'y = x'x = 1$. The eigenvalues of $I_p - \hat{A}$ are of the form $1 - \lambda$ and, hence, are $\geqslant .03$ in absolute value.

This kernel possesses some large sample optimality properties; see Andrews (1991). It does not suffer from the drawbacks of the truncated kernel (advocated by White (1984, p. 152)) and the Bartlett kernel (advocated by Newey and West (1987)). (The former kernel does not necessarily generate positive semi-definite estimates and the latter yields an estimator with a slower rate of convergence, and hence lower asymptotic efficiency, than the QS kernel; see Andrews (1991).)

The bandwidth parameter that we use is a data-dependent plug-in estimate of an optimal value determined in Andrews (1991). The optimal value is

$$(3.3) \qquad S_T^* = \left(q k_q^2 \alpha^*(q) T \Big/ \int k^2(x)\,dx \right)^{1/(2q+1)},$$

where q, k_q, and $\int k^2(x)\,dx$ are known values that depend on the kernel k and $\alpha^*(q)$ is an unknown scalar quantity that depends on the covariances of the sequence $\{V_t^*\}$. For the QS kernel, $q = 2$, $k_q = 1.421223$, and $\int k^2(x)\,dx = 1$.

In brief, the data-dependent bandwidth parameter is defined as follows: First, we specify p univariate approximating parametric models for $\{V_{at}^*\}$ for $a = 1, \ldots, p$ (where $V_t^* = (V_{1t}^*, \ldots, V_{pt}^*)'$). Second, we estimate the parameters of the approximating parametric models by standard methods. Third, we substitute these estimates into a formula (given below and in Andrews (1991)) that expresses $\alpha^*(q)$ as a function of the parameters of the parametric models. This yields an estimate $\hat{\alpha}^*(q)$ of $\alpha^*(q)$. The estimate $\hat{\alpha}^*(q)$ is then substituted into the formula (3.3) for the optimal bandwidth parameter S_T^* to yield the data-dependent bandwidth parameter $\hat{S}_T^*$:

$$(3.4) \qquad \hat{S}_T^* = \left(q k_q^2 \hat{\alpha}^*(q) T \Big/ \int k^2(x)\,dx \right)^{1/(2q+1)}.$$

For the QS kernel, we have

$$(3.5) \qquad \hat{S}_T^* = 1.3221(\hat{\alpha}^*(2) T)^{1/5}.$$

The approximating parametric models we use are first order autoregressive (AR(1)) models for $\{V_{at}\}$, $a = 1, \ldots, p$ (with different parameters for each a). These models have advantages of parsimony and computational simplicity. Let (ρ_a, σ_a^2) denote the autoregressive and innovation variance parameters, respectively, for $a = 1, \ldots, p$. Let $\{(\hat{\rho}_a, \hat{\sigma}_a^2)$: $a = 1, \ldots, p\}$ denote the corresponding estimates. Then, for $q = 2$, we have

$$(3.6) \qquad \hat{\alpha}^*(2) = \sum_{a=1}^{p} w_a \frac{4\hat{\rho}_a^2 \hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8} \Bigg/ \sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4},$$

where $\{w_a$: $a = 1, \ldots, p\}$ are specified weights (which determine the weight attached to the estimation of each of the p diagonal elements of J_T). The choice of $(w_1, \ldots, w_5)$ used here is $(0, 1, \ldots, 1)$, which gives no weight to the intercept term. (Formulae analogous to (3.6), but for ARMA(1, 1), MA(m), and VAR(m) approximating parametric models are given in Andrews (1991, eqns. (6.4)–(6.9)).)

Plugging $\hat{\alpha}^*(2)$ from (3.6) into (3.5) completely determines $\hat{S}_T^*$. Our VAR prewhitened kernel HAC estimator is then defined using (2.3), (2.4), (3.5), and (3.6) to be

$$(3.7) \qquad \hat{J}_{Tpw}(\hat{S}_T^*) = \hat{D} \hat{J}_T^*(\hat{S}_T^*) \hat{D}'.$$

The second estimator used in the Monte Carlo experiments, denoted QS, is the nonprewhitened kernel HAC estimator that is defined exactly as is QS-PW except that $\hat{A}(=\hat{A}_1)=\mathbf{0}$.

The third estimator used in the experiments, denoted PARA, is a parametric estimator that is based on the assumption that the errors are homoskedastic AR(1) random variables. By definition,

$$(3.8)\qquad \text{PARA}=\left[\left(\frac{1}{T}\sum_1^T X_t X_t'\right)^{-1}\left(\frac{1}{T-5}\sum_1^T \hat{U}_t^2\right)\times\left(\frac{1}{T}\sum_{s=1}^T\sum_{t=1}^T \hat{\rho}^{|s-t|}X_s X_t'\right)\left(\frac{1}{T}\sum_1^T X_t X_t'\right)^{-1}\right]_{22},$$

where $\hat{U}_t = Y_t - X_t'\hat{\theta}$, $\hat{\rho}_{\text{LS}}$ is the LS regression parameter estimator from the regression of $\hat{U}_t$ on $\hat{U}_{t-1}$ for $t=2,\ldots,T$, $\hat{\rho}=\min(.97,\hat{\rho}_{LS})$, and $[\cdot]_{22}$ denotes the (2,2) element of $\cdot$. The QS and PARA estimators are the same estimators as in the Monte Carlo study reported in Andrews (1991).

For each variance estimator and each scenario, the following performance criteria are estimated by Monte Carlo simulation: (1) the exact bias, variance, and MSE of the variance estimator and (2) the true confidence levels of the nominal 99%, 95%, and 90% regression coefficient confidence intervals (CIs) based on the t statistic constructed using the LS coefficient estimator and the variance estimator.[5] The control variate method of Davidson and MacKinnon (1981) is used to estimate the true confidence levels in (2). The sample size is 128. One thousand repetitions are used for each scenario.

The distributions of all of the variance estimators considered here are invariant with respect to the regression coefficient vector θ_0 in the model. Hence, we set $\theta_0=\mathbf{0}$ in each model and do so without loss of generality.

Next we describe the models used in the Monte Carlo study. The AR(1)-HOMO model consists of mutually independent errors and regressors. The errors are mean zero, homoskedastic, AR(1), stationary (i.e., the initial error distribution is chosen to yield a stationary sequence), normal random variables with variance 1 and AR parameter ρ. The four regressors are generated by four independent draws from the same distribution as that of the errors, but then are transformed to achieve a diagonal $(1/T)\sum_{t=1}^T X_t X_t'$ matrix.[6] A new set of regressors is randomly drawn for each repetition of the experiment (to ensure that the results are not sensitive to the selection of a single, perhaps atypical, matrix of regressors). In consequence, the value of the estimand $\text{Var}(\sqrt{T}(\hat{\theta}\text{-}\theta_0)|X)$ (which is used in calculating the bias and mean squared error of the estimators considered) varies across repetitions. Its average value across the repetitions is reported in the tables. (This method was also used in Andrews (1991).) The values considered for the AR(1) parameter ρ are 0, .3, .5, .7, .9, .95, $-.3$, and $-.5$.

[5] The nominal $100(1-\alpha)\%$ CIs are based on an asymptotic normal approximation. For the PARA estimator, this normal approximation is valid asymptotically only in the AR(1)-HOMO model. Also, note that the bias, variance, and MSE of the estimators are well-defined since $E\|(X'X)^{-1}\|^2<\infty$ when $T=120$.

[6] The transformation used is described as follows. Let $\tilde{x}$ denote the $T\times 4$ matrix of pretransformed, randomly generated, AR(1) regressor variables. Let $\bar{x}$ denote $\tilde{x}$ with its column means subtracted off. Let $x=\bar{x}((1/T)\bar{x}'\bar{x})^{-1/2}$. Define the $T\times 5$ matrix of transformed regressors to be $X=[1_T \vdots x]$. By construction, $X'X=TI_5$.

Since $E\tilde{x}=\mathbf{0}$ and $E\tilde{x}'\tilde{x}=I_4$, this transformation should be close to the identity map. With this transformation, the estimand and the estimators simplify and the computational burden is reduced considerably.

TABLE I

BIAS, VARIANCE, AND MSE OF QS-PW, QS, AND PARA ESTIMATORS AND TRUE CONFIDENCE LEVELS OF NOMINAL 99%, 95%, AND 90% CONFIDENCE INTERVALS CONSTRUCTED USING THE QS-PW, QS, AND PARA ESTIMATORS FOR THE AR(1)-HOMO MODEL—$T = 128$

ρ	Average Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
		QS-PW	.005	.079	.079	98.5	93.9	88.1
0	1.00	QS	−.060	.044	.048	98.6	93.3	87.7
		PARA	−.002	.017	.017	98.8	94.8	89.1
		QS-PW	.010	.15	.15	98.8	93.1	88.3
.3	1.18	QS	−.14	.090	.11	98.0	91.7	86.5
		PARA	−.029	.043	.044	98.7	93.9	89.2
		QS-PW	−.040	.39	.39	97.7	93.4	88.1
.5	1.60	QS	−.35	.21	.33	96.9	90.6	84.0
		PARA	−.14	.11	.13	98.1	93.9	88.3
		QS-PW	−.21	1.84	1.89	97.1	91.3	84.4
.7	2.63	QS	−.90	.71	1.52	95.3	85.5	78.2
		PARA	−.48	.51	.74	98.2	91.0	83.5
		QS-PW	−1.93	29.4	33.1	90.4	83.0	75.3
.9	6.40	QS	−4.04	2.55	18.9	82.5	72.0	64.4
		PARA	−3.08	3.41	12.9	90.2	81.6	73.5
		QS-PW	−4.03	42.7	58.9	84.1	74.8	66.5
.95	8.75	QS	−6.69	3.00	47.8	72.9	60.6	53.1
		PARA	−5.75	5.11	38.2	82.1	71.8	63.6
		QS-PW	.030	.19	.19	98.4	94.1	88.6
−.3	1.19	QS	−.13	.11	.13	97.7	93.1	86.2
		PARA	−.008	.044	.044	98.8	95.2	89.4
		QS-PW	.018	.49	.49	98.2	93.1	88.0
−.5	1.63	QS	−.30	.28	.37	96.5	90.1	84.6
		PARA	−.038	.15	.16	98.7	93.9	89.3

The AR(1)-HET1 and AR(1)-HET2 models are constructed by introducing multiplicative heteroskedasticity to the errors of the AR(1)-HOMO model. Suppose $\{x_t, \tilde{U}_t\colon t = 1, \ldots, T\}$ are the nonconstant regressors and errors generated by the AR(1)-HOMO model (where $X_t = (1, x_t')'$). Let $U_t = |x_t'\omega| \times \tilde{U}_t$. Then, $\{x_t, U_t\colon t = 1, \ldots, T\}$ are the nonconstant regressors and errors for the AR(1)-HET1 and AR(1)-HET2 models when $\omega = (1,0,0,0)'$ and $\omega = (1/2, 1/2, 1/2, 1/2)'$ respectively. In the AR(1)-HET1 model, the heteroskedasticity is related only to the regressor whose coefficient estimator's variance is being estimated, whereas in the AR(1)-HET2 model, the heteroskedasticity is related to all of the regressors.[7] The same values of ρ are considered as in the AR(1)-HOMO model.

The MA(1)-HOMO, MA(1)-HET1, and MA(1)-HET2 models are exactly the same as the AR(1)-HOMO, AR(1)-HET1, and AR(1)-HET2 models, respectively, except that stationary MA(1) processes replace stationary AR(1) processes everywhere that the latter arise in the definitions above. The MA(1) processes have variance 1 and MA parameter

[7] When the regressor transformation map is the identity map, the errors in the AR(1)-HET1 and AR(1)-HET2 models are mean zero, variance one, AR(1) sequences with AR parameter ρ^2 and innovations that are uncorrelated (unconditionally and conditionally on $\{X_t\}$) but are not independent. Hence, the errors have an AR(1) correlation structure even after the introduction of heteroskedasticity.

TABLE II

BIAS, VARIANCE, AND MSE OF QS-PW, QS, AND PARA ESTIMATORS AND TRUE CONFIDENCE LEVELS OF NOMINAL 99%, 95%, AND 90% CONFIDENCE INTERVALS CONSTRUCTED USING THE QS-PW, QS, AND PARA ESTIMATORS FOR THE AR(1)-HET1 MODEL—$T = 128$

ρ	Average Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
		QS-PW	−.21	1.66	1.70	97.5	93.9	85.9
0	2.94	QS	−.25	51.31	1.37	97.6	93.8	85.8
		PARA	−1.94	.048	3.82	86.0	75.6	67.2
		QS-PW	−.90	1.92	2.74	96.6	89.9	83.1
.3	3.86	QS	−1.09	1.42	2.61	96.2	89.9	82.0
		PARA	−2.79	.066	7.83	82.9	68.1	60.4
		QS-PW	−1.58	4.85	7.34	95.2	88.8	81.6
.5	5.28	QS	−2.06	2.79	7.05	94.0	86.7	79.0
		PARA	−4.00	.14	16.1	80.0	68.7	59.6
		QS-PW	−3.50	17.0	29.3	92.7	83.1	77.3
.7	8.82	QS	−4.52	8.57	29.0	90.7	80.3	72.8
		PARA	−7.11	.44	50.9	74.8	61.0	51.4
		QS-PW	−14.7	240.	455.	81.4	70.0	60.9
.9	23.5	QS	−18.0	24.4	347.	75.1	62.5	53.4
		PARA	−20.9	2.77	441.	58.6	45.7	38.1
		QS-PW	−31.4	123.	1107.	70.6	57.5	50.5
.95	39.3	QS	−34.5	17.8	1208.	61.9	48.9	41.9
		PARA	−36.8	3.81	1356.	49.0	39.0	33.4
		QS-PW	.88	3.04	3.81	98.9	95.4	92.0
−.3	2.41	QS	.61	2.18	2.55	98.8	95.3	91.0
		PARA	−1.28	0.079	1.72	88.8	78.8	69.9
		QS-PW	2.23	6.02	11.0	99.3	96.5	92.8
−.5	1.89	QS	1.61	3.49	6.08	99.0	95.7	91.9
		PARA	−.49	.19	.43	94.7	87.4	80.1

ψ (and are parameterized as $\tilde{U}_t = \varepsilon_t + \psi\varepsilon_{t-1}$). The values of ψ that are considered are .3, .5, .7, .99, −.3, −.5, −.7, and −.99.

The MA(m)-HOMO model is exactly the same as the AR(1)-HOMO model except that the errors and the (pre-transformed) regressors are homoskedastic, stationary MA(m) random variables with variance 1 and MA parameters $\psi_1, \dots, \psi_m$ (where the MA(m) model is parameterized as $U_t = \varepsilon_t + \Sigma_{r=1}^{m} \psi_r \varepsilon_{t-r}$). The MA parameters are taken to be positive and to decline linearly to zero (i.e., $\psi_r = 1 - r/(m+1)$ for $r = 1, \dots, m$). The values of m that are considered are 3, 5, 7, 9, 12, and 15.

The Monte Carlo results for the parameter/model combinations discussed above are given in Tables I–V. For the MA(1) models, however, no results are reported for the negative ψ values, since they are very nearly the same as for the corresponding positive ψ values.[8] In addition, Monte Carlo results have been computed, but are not reported, for MA(m)-HET1 and MA(m)-HET2 models (defined analogously to the MA(m)-HOMO

[8] The reason for this is that for the nonintercept regressors $\{V_t\} = \{U_t X_t\}$ has autocorrelations given by the product of the autocorrelations of $\{U_t\}$ and $\{X_t\}$, and hence, these autocorrelations are independent of the sign of ψ. Thus, for the nonintercept regressors, the distribution of $\{\hat{V}_t\} = \{\hat{U}_t X_t\}$ depends on the sign of ψ only due to the effect of the sign of ψ on the distribution of the deviations $\hat{U}_t - U_t$.

TABLE III

BIAS, VARIANCE, AND MSE OF QS-PW, QS, AND PARA ESTIMATORS AND TRUE CONFIDENCE LEVELS OF NOMINAL 99%, 95%, AND 90% CONFIDENCE INTERVALS CONSTRUCTED USING THE QS-PW, QS, AND PARA ESTIMATORS FOR THE AR(1)-HET2 MODEL—$T = 128$

ρ	Average Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
		QS-PW	.0072	.66	.66	98.2	93.5	87.6
0	1.47	QS	−.14	.60	.63	98.4	92.0	85.8
		PARA	−.49	.044	.29	96.3	88.6	81.0
		QS-PW	.0080	.84	.84	98.1	92.7	86.3
.3	1.66	QS	−.24	.49	.55	97.8	91.0	84.2
		PARA	−.59	.070	.41	95.5	87.8	80.1
		QS-PW	−.098	1.78	1.79	97.9	92.2	85.8
.5	2.13	QS	−.54	.72	1.01	96.7	88.6	81.8
		PARA	−.87	.13	.89	94.5	85.3	78.2
		QS-PW	−.36	5.89	6.02	95.8	89.9	83.4
.7	3.29	QS	−1.20	1.80	3.23	94.3	86.0	78.4
		PARA	−1.60	.45	3.01	91.8	83.3	75.9
		QS-PW	−2.32	38.2	43.5	89.6	80.5	71.5
.9	7.15	QS	−4.45	4.93	24.8	84.4	72.5	63.4
		PARA	−4.64	2.33	23.9	84.5	73.0	64.7
		QS-PW	−3.53	246.	258.	83.9	74.1	66.9
.95	9.58	QS	−7.01	7.15	56.2	75.6	62.0	53.9
		PARA	−6.99	5.00	53.8	77.6	65.8	57.9
		QS-PW	.034	.89	.89	98.7	94.5	88.7
−.3	1.68	QS	−.24	.49	.55	98.2	93.1	86.4
		PARA	−.57	.080	.40	97.0	89.1	82.0
		QS-PW	.050	2.46	2.46	97.8	92.0	87.0
−.5	2.17	QS	−.48	.93	1.16	95.9	88.6	83.0
		PARA	−.79	.19	.81	94.7	87.9	80.5

model). These results are not reported because they are qualitatively quite similar to the AR(1)-HET1 and AR(1)-HET2 results.

Inspection of Tables I–V shows a number of clear patterns in the relative performance of the three estimators QS-PW, QS, and PARA. First, in almost all model/parameter cases, QS-PW has the smallest bias. In a number of cases, its bias is much less than that of the other two estimators. In the HOMO models (whether AR(1), MA(1), or MA(m)), PARA has the next smallest bias, while in the HET1 and HET2 models, QS has the next smallest bias. Second, PARA always has the smallest variance, often by a considerable margin. QS has the next smallest variance in each case. Third, in the HOMO and HET2 models, PARA has the smallest MSE, followed by QS. In the HET1 models, QS has the smallest MSE, followed by QS-PW. In sum, prewhitening has the desired effect on bias, but it inflates variance sufficiently that its MSE is always worse than that of the nonprewhitened estimator QS. The parametric estimator PARA performs well in terms of MSE in the homoskedastic models, but does poorly in the heteroskedastic models, especially the HET1 models.

Next we discuss the patterns in the confidence interval coverage probabilities exhibited in Tables I–V. In almost all cases, the true coverage probabilities are less than the nominal asymptotic coverage probabilities. In these cases, the best CI coverage probabilities are the largest ones. The estimator QS-PW yields the best CI coverage probabilities

TABLE IV

BIAS, VARIANCE, AND MSE OF QS-PW, QS, AND PARA ESTIMATORS AND TRUE CONFIDENCE LEVELS OF NOMINAL 99%, 95%, AND 90% CONFIDENCE INTERVALS CONSTRUCTED USING THE QS-PW, QS, AND PARA ESTIMATORS FOR THE MA(1)-HOMO AND MA(1)-HET1 MODELS—$T = 128$

Model	ψ	Average Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
MA(1)-HOMO	.3	1.14	QS-PW	.00019	.13	.13	98.1	93.3	88.0
			QS	−.178	.059	.091	97.3	91.1	85.5
			PARA	−.032	.034	.035	98.3	94.3	89.4
	.5	1.30	QS-PW	.042	.24	.25	98.1	94.1	88.8
			QS	−.25	.10	.17	97.0	91.1	85.5
			PARA	−.060	.050	.053	98.3	94.3	88.9
	.7	1.42	QS-PW	.055	.28	.28	98.8	93.4	89.1
			QS	−.27	.14	.22	97.8	90.9	85.1
			PARA	−.073	.074	.079	99.2	93.9	89.4
	.99	1.47	QS-PW	.082	.34	.35	98.5	93.9	88.5
			QS	−.27	.19	.26	97.1	91.0	84.3
			PARA	−.091	.072	.081	98.7	93.3	88.7
MA(1)-HET1	.3	3.29	QS-PW	−.44	1.74	1.93	97.6	91.4	85.4
			QS	−.66	1.16	1.59	97.5	90.6	84.4
			PARA	−2.25	.061	5.11	84.0	71.5	62.5
	.5	3.70	QS-PW	−.41	2.67	2.84	97.2	90.9	84.4
			QS	−.86	1.58	2.33	96.8	88.3	82.8
			PARA	−2.56	.082	6.61	84.2	71.2	63.5
	.7	4.00	QS-PW	−.49	3.16	3.40	97.5	91.8	84.7
			QS	−1.06	1.99	3.11	96.2	88.7	81.2
			PARA	−2.79	.098	7.91	83.2	71.5	62.5
	.99	4.19	QS-PW	−.37	16.6	16.7	97.2	92.5	84.8
			QS	−1.08	2.64	3.81	95.5	88.8	81.8
			PARA	−2.94	.11	8.76	83.2	72.2	61.9

in almost all cases except for the AR(1)-HOMO and MA(1)-HOMO models. In these models, QS-PW is just slightly worse than PARA. In many model/parameter combinations, QS-PW is better than QS by a considerable margin in terms of CI coverage probabilities. In addition, QS-PW is better than PARA in the HET1 and HET2 models by a considerable margin.

The good performance of QS-PW in terms of CI coverage probabilities is due to its relatively small bias. It is apparent from the tables that the magnitude of an estimator's bias is much more important than its variance in determining its corresponding CI coverage probabilities. In sum, QS-PW is clearly the best estimator of the three in terms of CI coverage probabilities. PARA does well in the homoskedastic models, but performs poorly in the heteroskedastic models.[9]

Based on the Monte Carlo results reported here, the choice between the QS-PW and QS estimators is evident. If one desires lower variance and MSE, then QS is preferable.

[9] We note that the QS-PW and QS estimators each provide a different tradeoff between bias and variance. Correspondingly, they provide different performance re CI coverage probabilities. Monte Carlo results using a wide grid of different fixed bandwidth parameters for the QS estimator show that the same tradeoff cannot be attained (or even approached) simply by using a different bandwidth parameter for the QS estimator.

TABLE V

BIAS, VARIANCE, AND MSE OF QS-PW, QS, AND PARA ESTIMATORS AND TRUE CONFIDENCE LEVELS OF NOMINAL 99%, 95%, AND 90% CONFIDENCE INTERVALS CONSTRUCTED USING THE QS-PW, QS, AND PARA ESTIMATORS FOR THE MA(1)-HET2 AND MA(m)-HOMO MODELS—$T = 128$

Model	ψ or m	Average Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
MA(1)-HET2	.3	1.62	QS-PW	.052	.96	.97	97.7	93.1	89.3
			QS	−.25	.39	.45	97.5	92.2	86.1
			PARA	−.56	.060	.38	94.9	87.8	81.4
	.5	1.82	QS-PW	.16	3.77	3.80	99.0	93.2	87.6
			QS	−.32	.66	.76	98.4	91.2	83.8
			PARA	−.66	.087	.53	96.6	86.7	81.1
	.7	1.95	QS-PW	.18	2.10	2.13	97.5	93.2	88.1
			QS	−.39	.71	.86	96.6	90.6	84.0
			PARA	−.73	.11	.65	95.7	88.5	80.4
	.99	2.00	QS-PW	.13	2.19	2.21	97.7	92.8	88.5
			QS	−.42	.92	1.10	96.2	90.2	83.6
			PARA	−.75	.12	.68	95.0	87.3	79.8
MA(m)-HOMO	3	2.11	QS-PW	.34	1.33	1.45	98.2	94.6	89.5
			QS	−.49	.43	.67	96.7	89.8	82.9
			PARA	−.18	.26	.29	98.5	93.1	87.5
	5	2.92	QS-PW	.54	4.01	4.30	98.2	94.3	88.9
			QS	−.92	1.00	1.86	95.0	86.8	79.3
			PARA	−.39	.69	.84	98.1	93.2	86.8
	7	3.68	QS-PW	.60	8.54	8.90	98.2	93.6	88.5
			QS	−1.38	1.73	3.63	93.8	84.6	77.4
			PARA	−.76	1.13	1.71	97.1	90.5	83.4
	9	4.45	QS-PW	.77	17.5	18.1	97.4	92.5	86.5
			QS	−1.83	2.42	5.79	90.9	82.6	74.8
			PARA	−1.00	2.10	3.11	96.3	90.5	82.4
	12	5.50	QS-PW	.073	27.4	27.4	94.2	89.1	83.2
			QS	−2.81	3.21	11.1	88.6	77.2	70.5
			PARA	−1.89	3.01	6.60	94.0	86.0	78.8
	15	6.46	QS-PW	−.28	32.2	32.3	94.1	87.8	80.9
			QS	−3.56	4.55	17.2	87.5	75.9	68.2
			PARA	−2.65	4.30	11.3	92.4	83.7	76.0

If one desires lower bias and better CI coverage probabilities, then QS-PW is preferable. In many cases, CI coverage probabilities and the rejection rates of t statistics are of primary concern, and hence, the prewhitened estimator QS-PW is preferred.

Cowles Foundation, Yale University, P.O. Box 2125 Yale Station, New Haven, CT 06520, U.S.A.

Manuscript received June, 1990; final revision received March, 1992.

APPENDIX

PROOF OF THEOREM 1: Let

$$J_T^* = \frac{1}{T} \sum_{s=b+1}^{T} \sum_{t=b+1}^{T} EV_s^* V_t^{*\prime} \quad \text{and} \quad V_t^* = V_t - \sum_{r=1}^{b} A_r V_{t-r}. \tag{A.1}$$

It is straightforward to show that

$$(A.2)\qquad J_T^* - D^{-1}J_T D^{-1\prime} = \sum_{r=0}^{b}\sum_{u=0}^{b}\tilde{A}_r\left(\frac{1}{T}\sum_{s=b+1}^{T}\sum_{t=b+1}^{T} EV_{s-r}V'_{t-u} - \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EV_sV'_t\right)\tilde{A}'_u \to 0$$

where $\tilde{A}_r = I_p$ for $r = 0$ and $\tilde{A}_r = -A_r$ for $r = 1, \dots, b$. Given this result and Assumption D, it suffices to show that

$$(A.3)\qquad \hat{J}_T^*(\hat{S}_T) - J_T^* \xrightarrow{p} \mathbf{0}.$$

To establish (A.3) under part (i) of Assumption B, we apply the results of Section 8 of Andrews (1991) to the estimator $\hat{J}_T^*(\hat{S}_T)$ of J_T^* with $\hat{\theta}$ and θ_0 elongated to include $(\hat{A}_1, \dots, \hat{A}_b)$ and $(A_1, \dots, A_b)$ respectively. To do this, Assumptions A and B(i) must hold with $\{V_t\}$ replaced by $\{V_t^*\}$ and with $\hat{\theta}$ and θ_0 elongated as above. It is not difficult to show that if A and B(i) hold as stated, then they also hold with $\{V_t\}$ replaced by $\{V_t^*\}$. Furthermore, if A and B(i) hold as stated, then they also hold with $\hat{\theta}$ and θ_0 elongated provided Assumption D(i) holds. Thus, the results of Section 8 of Andrews (1991) establish (A.3).

Next, to establish (A.3) under part (ii) of Assumption B, let $\tilde{J}_T^*(\hat{S}_T)$ denote $\hat{J}_T^*(\hat{S}_T)$ when $\hat{\theta}$ and $\hat{A}_r$ are replaced by θ_0 and A_r for $r = 1, \dots, b$ respectively. Then, $\tilde{J}_T^*(\hat{S}_T) - J_T^* \xrightarrow{p} 0$ by Theorem 1 using part (i) of Assumption B. Hence, it suffices to show that $\hat{J}_T^*(\hat{S}_T) - \tilde{J}_T^*(\hat{S}_T) \xrightarrow{p} 0$ or that $(\sqrt{T}/\hat{S}_T)(\hat{J}_T^*(\hat{S}_T) - \tilde{J}_T^*(\hat{S}_T)) = O_p(1)$ (since $\hat{S}_T^2/T \xrightarrow{p} 0$ under the assumptions). One can show the latter by writing out the left-hand side expression using the definitions of $\hat{J}_T(\hat{S}_T)$ and $\tilde{J}_T^*(\hat{S}_T)$ and bounding each of the resultant terms using the assumptions (especially Assumption B(ii)) and using the fact that

$$\sum_{j=-T}^{T} |k(j/\hat{S}_T)| \Big/ \hat{S}_T \xrightarrow{p} \int |k(x)|\,dx. \qquad \textit{Q.E.D.}$$

REFERENCES

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

ANDREWS, D. W. K., AND J. C. MONAHAN (1990): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Cowles Foundation Discussion Paper No. 942, Cowles Foundation, Yale University.

BLACKMAN, R. B., AND J. W. TUKEY (1958): *The Measurement of Power Spectra*. New York: Dover, 1959. Originally appeared in *Bell System Technical Journal*, 37, January 1958 and March 1958.

DAVIDSON, R., AND J. G. MACKINNON (1981): "Efficient Estimation of Tail-area Probabilities in Sampling Experiments," *Economic Letters*, 8, 73–77.

GALLANT, A. R. (1987): *Nonlinear Statistical Models*. New York: Wiley.

GRENANDER, U., AND M. ROSENBLATT (1957): *Statistical Analysis of Stationary Time Series*. New York: Chelsea.

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

LEVINE, D. (1983): "A Remark on Serial Correlation in Maximum Likelihood," *Journal of Econometrics*, 23, 337–342.

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

PARZEN, E. (1984): "Autoregressive Spectral Estimation," in *Handbook of Statistics*, Vol. 3, ed. by D. R. Brillinger and P. R. Krishnaiah. New York: North-Holland.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–302.

PRESS, H., AND J. W. TUKEY (1956): "Power Spectral Methods of Analysis and Their Application to Problems in Airplane Dynamics," *Flight Test Manual*, NATO, Advisory Group for Aeronautical Research and Development, Vol. IV-C, 1–41. Reprinted as Bell System Monograph No. 2606.

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

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

WOOLDRIDGE, J. M. (1991): "On the Application of Robust, Regression-based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47, 5–46.