COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 942

AN IMPROVED HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATOR

by

Donald W.K. Andrews
and
J. Christopher Monahan

March 1990

AN IMPROVED HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATOR

Donald W. K. Andrews[1]
and
J. Christopher Monahan

Cowles Foundation
Yale University

March 1990

ABSTRACT

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 employed in the prewhitening stage. The paper establishes consistency, rate of convergence, and asymptotic truncated mean squared error (MSE) results for the estimators when a fixed or automatic bandwidth procedure is employed. Conditions are obtained under which prewhitening improves asymptotic truncated MSE. 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 usually of primary concern, prewhitened kernel estimators provide a significant improvement over the standard non-prewhitened kernel estimators.

JEL Classification Number: 211.

Keywords: Asymptotics, autocorrelation, consistency, covariance matrix estimator, heteroskedasticity, kernel estimator, mean squared error, nonparametric estimator, prewhitening, spectral density function, vector autoregression.

1. INTRODUCTION

Considerable attention has been paid in recent years to the estimation of covariance matrices in the presence of heteroskedasticity and autocorrelation of unknown form, see Hansen (1982), Levine (1983), White (1984), Gallant (1987), Newey and West (1987), Keener, Kmenta, and Webber (1987), Robinson (1988), and Andrews (1990). As shown in the Monte Carlo results of Andrews (1990), however, the estimators considered in the above papers all perform quite poorly in certain contexts. This paper considers a new class of estimators that exhibit better performance in several important respects.

In particular, this paper is motivated by the following finding. Kernel heteroskedasticity and autocorrelation consistent (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 (see Andrews (1990)). The problem is especially severe when there is considerable temporal dependency 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 a class of prewhitened kernel HAC estimators that includes the class of standard kernel HAC estimators. The prewhitened kernel HAC estimators are shown to be consistent and to converge to the estimand at the same rate as non-prewhitened kernel HAC estimators. Their asymptotic truncated mean squared error (MSE) is established and conditions are given under which prewhitening improves asymptotic truncated MSE.

Monte Carlo results show that the prewhitened kernel HAC estimators have lower bias and considerably better confidence interval coverage probabilities than do standard kernel estimators for a wide variety of different distributions of the data. For this reason, prewhitened kernel HAC estimators are recommended over standard kernel estimators for use in constructing confidence intervals and test statistics. On the other hand, the Monte Carlo results show that prewhitening inflates the variance and MSE of kernel HAC estimators. In consequence, there may be cases where standard kernel HAC estimators are preferred to 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 non–constancy 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 $f^*(\lambda)$.

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.

In brief, the procedure we suggest is the following: Suppose the observations are $\{V_t(\hat{\theta}) : t = 1, \ldots, T\}$ and the estimand is $J_T = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EV_s(\theta_0)V_t(\theta_0)'$, where $\hat{\theta} \xrightarrow{p} \theta_0$ and $\{V_t(\theta_0) : t = 1, \ldots, T\}$ is a weakly dependent sequence of random vectors (rv's). For fixed b, one estimates a b–th order vector autoregressive (VAR) model for $\{V_t(\hat{\theta})\}$. Let $\hat{A}_1, \ldots, \hat{A}_b$ denote the estimated VAR matrices. Next, one applies a kernel estimator to the residuals from the VAR model. Denote this estimator by $\hat{J}^*_T$. Finally, one undoes the prewhitening transformation (a process referred to as *recoloring*) by taking the estimator $\hat{J}_{Tpw}$ of J_T to be

$$\hat{J}_{Tpw} = \left[I - \sum_{r=1}^{b} \hat{A}_r\right]^{-1} \hat{J}^*_T \left[I - \sum_{r=1}^{b} \hat{A}_r'\right]^{-1} . \tag{1.1}$$

This is the VAR prewhitened kernel HAC estimator.

There are several reasons for choosing 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. Third, VAR models are parsimonious, at least when b is taken to be small. Fourth, prewhitened kernel HAC estimators based on VARs are computationally quite simple. Fifth, the transformation of the data induced by prewhitening with a VAR model is linear, and hence, it is easy to determine the inverse transformation needed to undo the effects of prewhitening after a kernel estimator has been applied to the transformed data. With nonlinear transformations, the inverse transformation may be difficult to determine. 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.

The remainder of this paper is organized as follows: Section 2 defines the estimand of interest, introduces the VAR prewhitening procedure, defines the class of kernel HAC estimators, and describes a "plug-in" automatic bandwidth estimation procedure. Section 3 presents consistency, rate of convergence, and asymptotic truncated MSE results for prewhitened kernel HAC estimators. It also provides conditions under which prewhitening improves asymptotic truncated MSE. Section 4 describes a Monte Carlo experiment that is designed to assess the effectiveness of prewhitening.

2. VAR PREWHITENED HAC ESTIMATORS

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

$$(2.1) \qquad \begin{aligned} &(B_T J_T B_T')^{-1/2}\sqrt{T}(\hat{\theta}-\theta_0) \xrightarrow{d} N(\underset{\sim}{0}, I_\zeta) \text{ as } T\to\infty\,, \text{ where} \\ &J_T = \frac{1}{T}\Sigma_{s=1}^T \Sigma_{t=1}^T EV_s(\theta_0)V_t(\theta_0)'\,, \end{aligned}$$

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 $\hat{B}_T - B_T \xrightarrow{p} 0$ as $T\to\infty$. 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.

By change of variables, the estimand J_T can be rewritten as

$$(2.2) \qquad J_T = \sum_{j=-T+1}^{T-1} \Gamma_T(j)\,, \text{ where } \Gamma_T(j) = \begin{cases} \frac{1}{T}\sum_{t=1+j}^{T} Ev_t V_{t-j}' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} EV_{t+j} V_t' & \text{for } j < 0 \end{cases}$$

and $V_t = V_t(\theta_0)$. Since $\{V_t : t \geq 1\}$ is taken to be "asymptotically weakly dependent" (defined more precisely below), the lag j covariances die out as $j \to \infty$, i.e., $\overline{\lim}_{j\to\infty} \sup_{T\geq 1} \|\Gamma_T(j)\| = 0$, where $\|\cdot\|$ denotes the Euclidean norm. This property is exploited by kernel HAC estimators and by the prewhitened kernel HAC estimators introduced below.

We now introduce a class of VAR prewhitened HAC estimators. They are defined as follows: Suppose $\hat{\theta}$ is a $\sqrt{T}$–consistent estimator of θ_0. Estimate a b–th order VAR process for $V_t(\hat{\theta})$:

$$(2.3) \qquad V_t(\hat{\theta}) = \sum_{r=1}^{b} \hat{A}_r V_{t-r}(\hat{\theta}) + V_t^*(\hat{\theta}) \text{ for } t = b+1, \ldots, T,$$

where $\hat{A}_r$ for $r = 1, \ldots, b$ are $p \times p$ parameter estimates and $\{V_t^*(\hat{\theta}) : t = b+1, \ldots, T\}$ are the corresponding residual vectors. For example, $\{\hat{A}_r : r = 1, \ldots, b\}$ 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 $\{V_t(\hat{\theta})\}$.

Below we assume the VAR parameter estimates $\hat{A}_r$ satisfy

$$(2.4) \qquad \sqrt{T}(\hat{A}_r - A_r) = O_p(1) \text{ for some matrices } A_r,\ r = 1, \ldots, b.$$

Even though there is no true VAR model, this assumption is not restrictive.

The prewhitened HAC estimator is constructed by applying a kernel HAC estimator, call it $\hat{J}_T^*(S_T)$, to the sequence of VAR residual vectors $\{V_t^*(\hat{\theta})\}$ and then transforming this estimator into an estimator of J_T by taking account of the prewhitening procedure. Let

$$(2.5)\quad \hat{J}_T^*(S_T) = \frac{T}{T-\zeta}\sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right)\hat{\Gamma}^*(j)\ , \text{ where } \hat{\Gamma}^*(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{V}_t^*\hat{V}_{t-j}^{*\prime} & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{V}_{t+j}^*\hat{V}_t^{*\prime} & \text{for } j < 0 \end{cases},$$

$\hat{V}_t^* = V_t^*(\hat{\theta})$, $k(\cdot)$ is a real–valued kernel in the set $\mathcal{K}_1$ defined below, and S_T is a bandwidth parameter. A data–dependent choice of S_T is discussed below. 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.

Using results from Andrews (1990), we show that

$$(2.6)\quad \begin{aligned} &\hat{J}_T^*(S_T) - J_T^* \xrightarrow{p} \underset{\sim}{0}\ , \text{ where} \\ &J_T^* = \frac{1}{T}\sum_{s=b+1}^{T}\sum_{t=b+1}^{T} EV_s^*V_t^{*\prime} \text{ and } V_t^* = V_t - \sum_{r=1}^{b} A_r V_{t-r}\ , \end{aligned}$$

provided $S_T = o(T)$. In addition, it is straightforward to show that if

$$(2.7)\quad D = \left[I_p - \sum_{r=1}^{b} A_r\right]^{-1}$$

is well–defined, then

$$(2.8)\quad \begin{aligned} &J_T^* - D^{-1}J_T D^{-1\prime} \\ &= \sum_{r=0}^{b}\sum_{u=0}^{b} \tilde{A}_r\left[\frac{1}{T}\sum_{s=b+1}^{T}\sum_{t=b+1}^{T} EV_{s-r}V_{t-u}^{\prime} - \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EV_sV_t^{\prime}\right]\tilde{A}_u^{\prime} \to \underset{\sim}{0} \end{aligned}$$

under the assumptions given below, where $\tilde{A}_r = I_p$ for $r = 0$ and $\tilde{A}_r = -A_r$ for $r = 1, \ldots, b$.

Equations (2.6) and (2.8) indicate the appropriate method of recoloring, i.e., of transforming $\hat{J}_T^*(S_T)$ to obtain an estimate of J_T rather than of J_T^*. In particular, we define the following prewhitened kernel estimator of J_T:

$$(2.9) \qquad \hat{J}_{Tpw}(S_T) = \hat{D}\hat{J}^*_T(S_T)\hat{D}' \text{ , where } \hat{D} = \left[I_p - \sum_{r=1}^{b} \hat{A}_r\right]^{-1} .$$

Once a kernel k is chosen and a (data–dependent) bandwidth S_T is specified, this yields an operational VAR prewhitened kernel estimator of J_T .

The kernel we suggest using is the QS kernel defined by

$$(2.10) \qquad k_{QS}(x) = \frac{25}{12\pi^2x^2}\left[\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right] .$$

The QS kernel yields an estimator $\hat{J}_{Tpw}(S_T)$ that is necessarily positive semi–definite. This kernel possesses some large sample optimality properties, see Andrews (1990). 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 (1990).)

The bandwidth parameter that we suggest using is a data–dependent plug–in estimate of an optimal value determined in Andrews (1990). The optimal value is

$$(2.11) \qquad S^*_T = \left[qk_q^2\alpha^*(q)T/\int k^2(x)dx\right]^{1/(2q+1)} ,$$

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 $\int k^2(x)dx = 1$.

The data–dependent bandwidth parameters are defined as follows: First, one specifies p univariate approximating parametric models for $\{V^*_{at}\}$ for $a = 1, \ldots, p$ (where $V^*_t = (V^*_{1t}, \ldots, V^*_{pt})'$) or one specifies a single multivariate approximating parametric model for $\{V^*_t\}$. Second, one estimates the parameters of the approximating parametric model(s) by standard methods. Third, one substitutes these estimates into a formula (given below or in Andrews (1990)) that expresses $\alpha^*(q)$ as a function of the parameters of the parametric model(s). This yields an estimate $\hat{\alpha}^*(q)$ of $\alpha^*(q)$. The estimate $\hat{\alpha}^*(q)$ is then substituted into the formula (2.11) for the optimal bandwidth parameter S_T^* to yield the data–dependent bandwidth parameter $\hat{S}_T^*$:

$$\hat{S}_T^* = \left[qk_q^2\hat{\alpha}^*(q)T/\int k^2(x)dx\right]^{1/(2q+1)} . \tag{2.12}$$

For the QS kernel, we have

$$\hat{S}_T^* = 1.3221(\hat{\alpha}^*(2)T)^{1/5} . \tag{2.13}$$

For general purposes, the suggested approximating parametric models 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. If some other model(s) seem more appropriate for a particular problem, however, they should be used instead. 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

$$\hat{\alpha}^*(2) = \sum_{a=1}^{p} w_a \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8} \Big/ \sum_{a=1}^{p} w_a \frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4} , \tag{2.14}$$

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 usual choice for w_a is one for all a except that which corresponds to an intercept and zero for the latter or one for all a. Formulae analogous to (2.14), but for ARMA(1,1), MA(m), and VAR(1) approximating parametric models are given in Andrews (1990, eqns. (6.4)–(6.8)).

Plugging $\hat{\alpha}^*(2)$ from (2.14) into (2.13) completely determines $\hat{S}_T^*$. Our VAR prewhitened kernel HAC estimator is then defined using (2.5), (2.9), and (2.13) to be

$$\hat{J}_{Tpw} = \hat{J}_{Tpw}(\hat{S}_T^*) = \hat{D}\hat{J}_T^*(\hat{S}_T^*)\hat{D}' \,. \tag{2.15}$$

3. CONSISTENCY AND ASYMPTOTIC MSE

In this section we establish consistency, rate of convergence, and asymptotic truncated MSE properties of prewhitened HAC estimators. The latter results are used to determine those scenarios where prewhitening reduces asymptotic bias, variance, and/or MSE.

We consider the following classes of kernels:

$$\mathcal{K}_1 = \Big\{k(\cdot) : R \to [-1,1] \,|\, k(0) = 1,\ k(x) = k(-x)\ \forall x \in R,\ \int_{-\infty}^{\infty} k^2(x)dx < \infty,\ k(\cdot) \text{ is continuous at 0 and at all but a finite number of other points}\Big\} \text{ and}$$

$$\mathcal{K}_3 = \Big\{k(\cdot) \in \mathcal{K}_1 : \text{(i) } |k(x)| \le C_1|x|^{-B} \text{ for some } B > 1 + \tfrac{1}{q} \text{ and some } C_1 < \infty, \text{ where } q \in (0,\infty) \text{ is such that } k_q \in (0,\infty), \text{ and (ii) } |k(x) - k(y)| \le C_2|x-y|\ \forall x, y \in R \text{ for some constant } C_2 < \infty\Big\}\,. \tag{3.1}$$

$\mathcal{K}_1$ contains the QS, truncated, Bartlett, Parzen, and Tukey–Hanning kernels among others. $\mathcal{K}_3$ contains all of these kernels except the truncated kernel. For fixed sequences of bandwidth parameters our consistency results hold for all kernels in $\mathcal{K}_1$. For data–dependent sequences $\{\hat{S}_T^*\}$, they hold for all kernels in $\mathcal{K}_3$.

The asymptotic bias of kernel estimators depends on the smoothness of the kernel at zero. Following Parzen (1957), we define

$$k_q = \lim_{x \to 0} \frac{1 - k(x)}{|x|^q} \text{ for } q \in (0,\infty)\,. \tag{3.2}$$

If q is an even integer, then $k_q = -\frac{1}{q!}\frac{d^q k(x)}{dx^q}\Big|_{x=0}$ and $k_q < \infty$ if and only if $k(x)$ is q times differentiable at zero. For the QS kernel, $k_2 = 1.421223$, $k_q = 0$ for $q < 2$, and $k_q = \infty$ for $q > 2$.

We consider the following asymptotic truncated MSE criterion with arbitrarily high truncation point:

$$
\begin{aligned}
&\lim_{h\to\infty} \lim_{T\to\infty} \mathrm{MSE}_h(T^{2q/(2q+1)}, \hat{J}_T, J_T, W_T) \\
(3.3) \qquad &= \lim_{h\to\infty} \lim_{T\to\infty} E \min\left\{ | T^{2q/(2q+1)} \mathrm{vec}(\hat{J}_T - J_T)' W_T \, \mathrm{vec}(\hat{J}_T - J_T) |, h \right\},
\end{aligned}
$$

where $\hat{J}_T$ is some estimator of J_T, W_T is some (possibly random) $p^2 \times p^2$ weight matrix, and $\mathrm{vec}(\cdot)$ is the column by column vectorization operator. The squared error loss function in (3.3) is truncated at h to circumvent undue influence of $\hat{\theta}$ on the criterion of performance. For example, if $\hat{\theta}$ has infinite second moment, its use can have the undesirable consequence of dominating a non–truncated MSE criterion.

Below we give consistency and rate of convergence results that hold when $\{V_t\}$ is stationary or non–stationary (but not explosive). The asymptotic truncated MSE results only hold, however, when $\{V_t\}$ is eighth order stationary. This permits conditional heteroskedasticity, but precludes unconditional heteroskedasticity. For the latter case, bounds on the asymptotic MSE can be obtained via the method employed in Andrews (1988), but such bounds are not given here.

When $\{V_t\}$ is second order stationary, we define

$$
\begin{aligned}
&f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) e^{-ij\lambda}, \quad \Gamma(j) = EV_t V_{t-j}', \quad f = f(0), \\
(3.4) \qquad &f^*(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma^*(j) e^{-ij\lambda}, \quad \Gamma^*(j) = EV_t^* V_{t-j}^{*\prime}, \quad f^* = f^*(0), \\
&f^{(q)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \Gamma(j), \quad f_*^{(q)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \Gamma^*(j) \ \text{ for } q \in [0,\infty),
\end{aligned}
$$

where $i = \sqrt{-1}$. The spectral density matrices of $\{V_t\}$ and $\{V_t^*\}$ are $f(\lambda)$ and $f^*(\lambda)$ respectively. Their values at $\lambda = 0$ are denoted by f and f^*. The limit as $T \to \infty$ of the estimand J_T equals $2\pi f$ when $\{V_t\}$ is second order stationary. This fact motivates the use of spectral density estimators to estimate J_T. $f^{(q)}$ and $f_*^{(q)}$ index the smoothness of $f(\lambda)$ and $f^*(\lambda)$, respectively, at $\lambda = 0$. If q is even, then $f^{(q)} = (-1)^{q/2} \left.\frac{d^q f(\lambda)}{d\lambda^q}\right|_{\lambda=0}$ and $\|f^{(q)}\| < \infty$ if and only if $f(\lambda)$ is q times differentiable at $\lambda = 0$. Analogous results hold for $f_*^{(q)}$. The asymptotic biases of kernel and prewhitened kernel estimators depend on $f^{(q)}$ and $f_*^{(q)}$ respectively.

For the second order stationary case, it can be shown that

$$f(\lambda) = D(\lambda)f^*(\lambda)D(\lambda)^t \text{ and } f = Df^*D' \text{, where } D(\lambda) = \left[I_p - \sum_{r=1}^{b} A_r e^{-ir\lambda}\right]^{-1} \tag{3.5}$$

and $(\cdot)^t$ denotes the conjugate transpose operator.

Let $\kappa_{abcd}(t, t+j, t+\ell, t+n)$ denote the fourth order cumulant of $(V_{at}, V_{bt+j}, V_{ct+\ell}, V_{dt+n})$, where V_{at} denotes the a–th element of V_t. Let tr denote the trace function and $\otimes$ the tensor (or Kronnecker) product operator. Let K_{pp} denote the $p^2 \times p^2$ commutation matrix that transforms $\text{vec}(A)$ into $\text{vec}(A')$, i.e., $K_{pp} = \sum_{i=1}^{p} \sum_{j=1}^{p} e_i e_j' \otimes e_j e_i'$, where e_i is the i–th elementary p–vector. 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 (1990) that will be assumed to hold when stated. See Andrews (1990) for further discussion of these assumptions.

ASSUMPTION A: $\{V_t\}$ *is a mean zero, fourth order stationary sequence of rv's with* $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$ *and* $\sum_{j=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \kappa_{abcd}(0,j,\ell,n) < \infty$ $\forall a, b, c, d \leq p$.

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

As shown in Andrews (1990, Lemma 1), Assumptions A and A* (and Assumptions C(i), C*(i), G, and G* stated below) are implied by an α–mixing condition and a moment condition.

ASSUMPTION B: (i) $\sqrt{T}(\hat{\theta} - \theta_0) = O_p(1)$.

(ii) $\sup_{t\geq 1} E\|V_t\|^2 < \infty$.

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

(iv) $\int_{-\infty}^{\infty} k^2(x)dx < \infty$.

ASSUMPTION C: (i) *Assumption* A *holds with* V_t *replaced by* $\left[V_t', \text{vec}\left[\frac{\partial}{\partial\theta'}V_t(\theta_0) - E\frac{\partial}{\partial\theta'}V_t(\theta_0)\right]'\right]'$.

(ii) $\sup_{t\geq 1} E \sup_{\theta\in\Theta} \left\|\frac{\partial^2}{\partial\theta\partial\theta'} V_{at}(\theta)\right\|^2 < \infty \ \forall a = 1, \ldots, p$, *where* $V_t(\theta) = (V_{1t}(\theta), \ldots, V_{pt}(\theta))'$.

ASSUMPTION C*: *Assumption* C *holds but with reference to Assumption* A* *rather than Assumption* A *in part* (i).

ASSUMPTION D: (i) $\{V_t\}$ *is eighth order stationary with* $\sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_7=-\infty}^{\infty} \kappa_{a_1\cdots a_8}(0, j_1, \ldots, j_7) < \infty$.

(ii) $W_T \xrightarrow{p} W$.

ASSUMPTION E: $\hat{\alpha}^*(q) = O_p(1)$ *and* $1/\hat{\alpha}^*(q) = O_p(1)$.

For consistency of $\hat{J}_{Tpw}$, $\hat{\alpha}^*(q)$ need only satisfy Assumption E. For rate of convergence and asymptotic truncated MSE results, stronger conditions on $\hat{\alpha}^*(q)$ are needed.

Let $\hat{\xi}$ denote the estimator of the parameter of the approximating parametric model(s) for $\{V_{at}^*\}$, $a = 1, \ldots, p$, introduced in Section 2. For example, with univariate AR(1) approximating parametric models, $\hat{\xi} = (\hat{\rho}_1, \hat{\sigma}_1^2, \ldots, \hat{\rho}_p, \hat{\sigma}_p^2)'$. Let ξ denote the probability limit of $\hat{\xi}$. $\hat{\alpha}^*(q)$ is the value of $\alpha^*(q)$ that corresponds to $\hat{\xi}$. The probability limit of $\hat{\alpha}^*(q)$ depends on ξ and is denoted α_ξ^*.

ASSUMPTION F: $\sqrt{T}(\hat{\alpha}^*(q) - \alpha_\xi^*) = O_p(1)$ *for some* $\alpha_\xi^* \in (0,\infty)$.

ASSUMPTION G: $\lambda_{max}(\Gamma(j)) \le C_3 j^{-\tau}$ $\forall j \ge 0$ *for some* $C_3 < \infty$ *and some* $\tau > \max\{2, 1 + 2q/(q+2)\}$, *where* q *is as in* $\mathcal{K}_3$.

ASSUMPTION G*: *Assumption* G *holds with* $\lambda_{max}(\Gamma(j))$ *replaced by* $\sup_{t\geq 1} \lambda_{max}(EV_tV_{t+j}')$.

We add one assumption, regarding the behavior of the VAR parameter estimators $\{\hat{A}_r\}$, that is not considered in Andrews (1990).

ASSUMPTION H: (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) $D = I_p - \sum_{r=1}^{b} A_r$ *is nonsingular.*

Note that Assumption H(ii) does not require the matrices $\{A_r : r = 1, \ldots, b\}$ to correspond to a *stationary* VAR process. In practice, however, they usually would.

The main result of this section is the following:

THEOREM 1: *Suppose* $k \in \mathcal{K}_3$, q *is as in* $\mathcal{K}_3$, $\|f^{(q)}\| < \infty$, *and Assumption* H *holds.*
(a) *If Assumptions* A, B, *and* E *or Assumptions* A*, B, *and* E *hold and* $q > 1/2$, *then* $\hat{J}_{Tpw} - J_T \xrightarrow{p} 0$.
(b) *If Assumptions* B, C, F, *and* G *or Assumptions* B, C*, F, *and* G* *hold, then* $T^{q/(2q+1)}(\hat{J}_{Tpw} - J_T) = O_p(1)$.

(c) *If Assumptions* B–D, F, *and* G *hold, then*

$$\lim_{h\to\infty}\lim_{T\to\infty} \mathrm{MSE}_h(T^{2q/(2q+1)}, \hat{J}_{Tpw}, J_T, W_T)$$
$$= C_{k,q}\alpha_\xi^{*1/(2q+1)}\Big[(\mathrm{vec}\, f_*^{(q)})' W^* \,\mathrm{vec}\, f_*^{(q)}/(q\alpha_\xi^*) + \mathrm{tr}\, W^*(I_{p^2} + K_{pp})(f^* \otimes f^*)\Big] \quad \textit{and}$$
$$\mathrm{tr}\, W^*(I_{p^2} + K_{pp})(f^* \otimes f^*) = \mathrm{tr}\, W(I_{p^2} + K_{pp})(f \otimes f)\,,$$

where $C_{k,q} = 4\pi^2\Big[qk_q^2/\int k^2(x)dx\Big]^{1/(2q+1)}\int k^2(x)dx$ *and* $W^* = (D \otimes D)'W(D \otimes D)$.

COMMENTS: 1. Except for Assumption H, the assumptions of Theorem 1 are identical to those used in Andrews (1990) for non–prewhitened kernel estimators. In fact, Theorem 1 contains results for non–prewhitened kernel estimators by taking $\hat{A}_r = A_r = \underset{\sim}{0}$ $\forall r = 1, \ldots, b$.

2. For the QS kernel, $q = 2$ and the rate of convergence in Theorem 1(b) is $T^{-2/5}$.

3. When $q = 2$, as with the QS kernel, $f^{(q)}$ and $f_*^{(q)}$ are related as follows:

$$(3.6) \qquad f^{(q)} = -\left.\frac{d^2 f(\lambda)}{d\lambda^2}\right|_{\lambda=0} \quad \text{and} \quad f_*^{(q)} = -\left.\frac{d^2 f^*(\lambda)}{d\lambda^2}\right|_{\lambda=0} = -\left.\frac{d^2}{d\lambda^2}[D(\lambda)f(\lambda)D(\lambda)^t]\right|_{\lambda=0} .$$

These expressions and Theorem 3(c) allow one to compare the asymptotic truncated MSE of kernel HAC estimators and prewhitened kernel HAC estimators. To this comparison we now turn.

Any comparison between prewhitened and non–prewhitened kernel estimators depends upon the choice of bandwidth parameter for each estimator. Meaningful comparisons can be made only if a reasonable choice of bandwidth parameter is made for each estimator. In consequence, the most appropriate choice of bandwidth parameters for making such comparisons are the optimal bandwidth parameters for the two estimators.[3] Thus, for the prewhitened estimator, we consider the case where

$$(3.7)\qquad \alpha_\xi^* = \alpha^*(q)\ \left[= \frac{2(\text{vec } f_*^{(q)})' W^* \text{ vec } f_*^{(q)}}{\text{tr } W(I_{p^2} + K_{pp})(f \otimes f)} \right]$$

(where the latter equality uses Corollary 1 of Andrews (1990)). For the non-prewhitened estimator, we consider the case where

$$(3.8)\qquad \begin{aligned} &\hat{S}_T = \left[q k_q^2 \hat{\alpha}(q) T / \int k^2(x)dx \right]^{1/(2q+1)}, \ \hat{\alpha}(q) \xrightarrow{p} \alpha_\xi, \text{ and} \\ &\alpha_\xi = \alpha(q)\ \left[= \frac{2(\text{vec } f^{(q)})' W \text{ vec } f^{(q)}}{\text{tr } W(I_{p^2} + K_{pp})(f \otimes f)} \right]. \end{aligned}$$

Let $\hat{J}_T$ $(= \hat{J}_T(\hat{S}_T))$ denote the non–prewhitened estimator. (By definition, it is the same as a prewhitened estimator with $\hat{A}_r = \underset{\sim}{0}$ $\forall r$ and $\hat{S}_T$ as in (3.8).)

For the above choices of bandwidth parameters, we have

$$(3.9)\qquad \begin{aligned} &\lim_{h\to\infty} \lim_{T\to\infty} \text{MSE}_h(T^{2q/(2q+1)}, \hat{J}_{Tpw}, J_T, W_T) \\ &= \tilde{C}_{k,q}\left[(\text{vec } f_*^{(q)})' W^* \text{ vec } f_*^{(q)} \right]^{1/(2q+1)} [\text{tr } W(I_{p^2} + K_{pp})(f \otimes f)]^{2q/(2q+1)} \end{aligned}$$

and

$$(3.10)\qquad \begin{aligned} &\lim_{h\to\infty} \lim_{T\to\infty} \text{MSE}_h(T^{2q/(2q+1)}, \hat{J}_T, J_T, W_T) \\ &= \tilde{C}_{k,q}\left[(\text{vec } f^{(q)})' W \text{ vec } f^{(q)} \right]^{1/(2q+1)} [\text{tr } W(I_{p^2} + K_{pp})(f \otimes f)]^{2q/(2q+1)}, \end{aligned}$$

where $\tilde{C}_{k,q} = 4\pi^2 \left[\frac{2q+1}{2q}\right] \left[2qk_q^2 / \int k^2(x)dx \right]^{1/(2q+1)} \int k^2(x)dx$. It can be shown that the asymptotic squared bias component is $[1/(2q+1)]$–th of the MSE in (3.9) and (3.10). In consequence, the effect of prewhitening is the same for the asymptotic squared bias, variance, and MSE. Thus, it suffices to determine those cases where prewhitening improves asymptotic MSE.

For simplicity, suppose V_t is a scalar (i.e., $p = 1$). Let the weight W_T in the MSE criterion equal one for all T . Then, from (3.9) and (3.10), the asymptotic MSEs of $\hat{J}_{Tpw}$ and $\hat{J}_T$ for estimating J_T are

$$\tilde{C}_{k,q}\left[f_*^{(q)2}D^4\right]^{1/(2q+1)}\left[2f^2\right]^{2q/(2q+1)} \quad \text{and}$$

(3.11) $$\tilde{C}_{k,q}\left[f^{(q)2}\right]^{1/(2q+1)}\left[2f^2\right]^{2q/(2q+1)}$$

respectively. Thus, $\hat{J}_{Tpw}$ has smaller asymptotic MSE than $\hat{J}_T$ if and only if

(3.12) $$|f_*^{(q)}|D^2 < |f^{(q)}| .$$

Suppose the kernel under consideration has $q = 2$. This is true of the QS kernel and many other kernels. Then, by (3.6) and some algebra, equation (3.12) can be shown to be equivalent to

(3.13) $$D''^2 < D''|D|f''/f ,$$

where $f = f(0)$, $|D| = \left|1 - \sum_{r=1}^{b} A_r\right|$, $f'' = \frac{d^2}{d\lambda^2}f(\lambda)\Big|_{\lambda=0}$, $D'' = \frac{d^2}{d\lambda^2}|D(\lambda)|\Big|_{\lambda=0}$, and $|\cdot|$ denotes the modulus of a complex number. Note that f and $|D|$ are positive real numbers, while f'' and D'' may be any real numbers.

For the case of a VAR process of order $b \geq 1$, we have

(3.14) $$D'' = -|D|^3\left[D^{-1}\sum_{r=1}^{b} r^2A_r + \left[\sum_{r=1}^{b} rA_r\right]^2\right] .$$

This simplifies when $b = 1$ to

(3.15) $$D'' = -A/|1-A|^3 ,$$

where $A = A_r$ for $r = 1$. In consequence, for a first–order VAR prewhitening process, (3.13) can be written as

$$(3.16)\qquad \begin{aligned} &A/(1-A)^2 < -f''/f \ \text{ for } \ A/(1-A) > 0 \ \text{ and} \\ &A/(1-A)^2 > -f''/f \ \text{ for } \ A/(1-A) < 0 \,. \end{aligned}$$

To reiterate, (3.16) gives those scenarios where the asymptotic squared bias, variance, and MSE of the prewhitened estimator $\hat{J}_{Tpw}$ is less than that of the standard kernel estimator $\hat{J}_T$ for the case where V_t is a scalar $(p = 1)$, an AR(1) model is used for prewhitening $(b = 1)$ with estimated parameter $\hat{A}$ $(= \hat{A}_1)$ that satisfies $\hat{A} \to A$, $\{V_t\}$ is second order stationary with spectrum $f(\lambda)$, $f = f(0)$, $f'' = \frac{d^2}{d\lambda^2}f(\lambda)\Big|_{\lambda=0}$, and the kernel chosen has $q = 2$.

Now, suppose $\{V_t\}$ actually is a stationary AR(1) process with AR(1) coefficient $\eta \in (-1,1)$. Then, $-f''/f = 2\eta/(1-\eta)^2$ (see eqn. (6.4) of Andrews (1990)). This result, equation (3.16), and some algebra show that any of the following combinations of η and A lead to improved asymptotic MSE due to prewhitening:

$$(3.17)\qquad \begin{aligned} &\eta \in (-1, -.172] \ \text{ and } \ A \in (-1,0)\,, \\ &\eta \in (-.172, 0) \ \text{ and } \ A \in (h(\eta), 0)\,, \ \text{ and} \\ &\eta \in (0,1) \ \text{ and } \ A \in (0, h(\eta))\,, \ \text{ where} \\ &h(\eta) = \eta + \tfrac{1-\eta}{4\eta}[3\eta + 1 - (\eta^2 + 6\eta + 1)^{1/2}] \ \text{ for } \ \eta \neq 0\,. \end{aligned}$$

Table 1 provides the values of $h(\eta)$ for a number of values of η. For example, for $\eta = .5$, we have $h(\eta) = .610$. Thus, if $\{V_t\}$ is an AR(1) process with AR parameter $\eta = .5$, then asymptotic MSE is improved by prewhitening using an AR(1) model with AR parameter A for any $A \in (0, .610)$. Any combination of $\eta \in (-1,1)$ and $A \in (-1,1)$ that is not included in (3.17) corresponds to a case where prewhitening decreases (or leaves unchanged) the asymptotic MSE. (Note that (3.17) does not list combinations of $\eta \in (-1,1)$ and A with $|A| > 1$ where prewhitening improves asymptotic MSE, although such combinations do exist.) Equation (3.17) and Table 1 illustrate the margin for error in the estimation (or choice) of the autoregressive parameter used in the prewhitening procedure. If $A = \eta$ $(\neq 0)$, then prewhitening always improves asymptotic MSE. If A differs from η, there is still an interval of A values containing η for which MSE is improved by prewhitening.

Next, suppose $\{V_t\}$ is an MA(1) process with moving average parameter τ (i.e., $V_t = \epsilon_t + \tau\epsilon_{t-1}$ for $t \geq 1$). In this case, $-f''/f = 2\tau/(1+\tau)^2$ (see eqn. (6.5) of Andrews (1990)). The latter result, equation (3.16), and some algebra show that any of the following combinations of τ and A lead to improved asymptotic MSE due to prewhitening:

$$\tau \in (-1, -.101) \text{ and } A \in (-1,0),$$

(3.18)
$$\tau \in (-.101, 0) \text{ and } A \in (g(\tau), 0), \text{ and}$$
$$\tau \in (0,1) \text{ and } A \in (0, g(\tau)), \text{ where}$$
$$g(\tau) = -\frac{3\tau^2 - 10\tau - 1}{4\tau} - \frac{1}{4\tau}\left[\tau^4 + 12\tau^3 + 22\tau^2 + 12\tau + 1\right]^{1/2} \text{ for } \tau \neq 0.$$

Table 2 provides values of $g(\tau)$ for a number of values of τ. Table 2 also provides values of $A(\tau)$ for these values of τ, where $A(\tau)$ is defined to be the probability limit of the least squares estimator $\hat{A}$ of the AR(1) prewhitening model when $\{V_t\}$ is an MA(1) process with moving average parameter τ. (With this definition, $A(\tau) = \tau/(1 + \tau^2)$.) Any combination of $\tau \in (-1,1)$ and $A \in (-1,1)$ that is not included in (3.18) corresponds to a case where prewhitening decreases (or leaves unchanged) the asymptotic MSE.

Equation (3.18) illustrates the wide range of A values that improve asymptotic MSE when τ is negative. It and Table 2 also illustrate the smaller range of A values that improve MSE when τ is positive. Table 2 shows that for most positive values of τ the probability limit $A(\tau)$ of the least squares estimator $\hat{A}$ is too large to fit in the interval of A values for which the asymptotic MSE is improved by prewhitening. This is to be expected. Finite sample considerations, however, operate in favor of the prewhitening procedure, since the least squares estimator is biased downward for $\tau > 0$.

4. MONTE CARLO RESULTS

In this section, Monte Carlo methods are used to evaluate the performance of the VAR prewhitened HAC estimator introduced above. We are interested in comparing the prewhitened HAC estimator to the non–prewhitened 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:

$$Y_t = X_t'\theta_0 + U_t\,,\quad t = 1, \ldots, T\,,\quad \hat{\theta} = \left[\Sigma_1^T X_t X_t'\right]^{-1} \Sigma_1^T X_t Y_t\,,\ \text{and}$$

(4.1)
$$\mathrm{Var}(\sqrt{T}(\hat{\theta} - \theta_0)) = \left[\tfrac{1}{T}\Sigma_1^T X_t X_t'\right]^{-1} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} EU_s X_s U_t X_t' \left[\tfrac{1}{T}\Sigma_1^T X_t X_t'\right]^{-1}.$$

The estimand of interest is the variance of the LS estimator of the first non–constant regressor. (That is, the estimand is the second diagonal element of $\mathrm{Var}(\sqrt{T}(\hat{\theta} - \theta_0))$ in (4.1).)

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 are 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 the prewhitened kernel HAC estimator defined in (2.15) that uses the QS kernel, a first–order

VAR prewhitening procedure $(b = 1)$, and the automatic bandwidth procedure defined in (2.13) and (2.14) with weights $(w_1, \ldots, w_5) = (0, 1, \ldots, 1)$.[4] The underlying rv's $\{V_t(\hat{\theta})\}$, upon which QS–PW is constructed (see (2.3)), are defined by $V_t(\hat{\theta}) = (Y_t - X_t'\hat{\theta})X_t$. The second estimator, denoted QS, is the non–prewhitened kernel HAC estimator that is defined exactly as is QS–PW except that $\hat{A}\ (= \hat{A}_1) = \underset{\sim}{0}$. The third estimator, denoted PARA, is a parametric estimator that is based on the assumption that the errors are homoskedastic AR(1) random variables. By definition,

$$(4.2)\qquad \text{PARA} = \left[\left[\tfrac{1}{T}\Sigma_1^T X_t X_t'\right]^{-1}\left[\tfrac{1}{T-5}\Sigma_1^T \hat{U}_t^2\right]\left[\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{\rho}^{|s-t|}X_s X_t'\right]\left[\tfrac{1}{T}\Sigma_1^T X_t X_t'\right]^{-1}\right]_{22},$$

where $\hat{U}_t = Y_t - X_t'\hat{\theta}$, $\hat{\rho}_{LS}$ 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 Ca-lo study reported in Andrews (1990).

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 $\theta_0 = \underset{\sim}{0}$ 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, stationary, AR(1), 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 $\frac{1}{T}\sum_{t=1}^{T} X_t X_t'$ matrix.[6] The values considered for the AR(1) parameter ρ are 0, .3, .5, .7, .9, .95, −.3, and −.5.

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 $\{x_t, \tilde{U}_t : t = 1, \ldots, T\}$ are the non–constant 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, $\{x_t, U_t : t = 1, \ldots, T\}$ are the non–constant 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 ψ (and are parameterized as $\tilde{U}_t = \epsilon_t + \psi\epsilon_{t-1}$). 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 $\psi_1, \ldots, \psi_m$ (where the MA(m) model is parameterized as $U_t = \epsilon_t + \sum_{r=1}^{m} \psi_r \epsilon_{t-r}$). The MA parameters are taken to be positive and to decline linearly to zero (i.e., $\psi_r = 1 - r/(m+1)$ for $r = 1, \ldots, m$). 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 3–7. 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 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 3–7 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 non-prewhitened 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 3–7. 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 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. If one desires lower bias and better CI coverage probabilities, then QS–PW is preferable. In many cases, CI coverage probabilities and the corresponding rejection rates of t–statistics are of primary concern, and hence, the prewhitened estimator QS–PW is preferred.

Lastly, we note that the Monte Carlo results reported above are at odds in some respects with the asymptotic results of Section 3. The latter show that in large samples, when the bandwidth parameter is chosen appropriately, the effect of prewhitening is the same for the squared bias, variance, and MSE. That is, in any given model scenario prewhitening should reduce all three or increase all three. In the Monte Carlo experiments, however, we find that prewhitening has the effect of reducing bias and increasing variance and MSE in most model scenarios.

The explanation for this discrepancy is that the parameters $\{A_T\}$ of the prewhitening procedure are estimated rather than fixed. The asymptotics are the same for these two cases, but the finite sample properties differ. One would expect that the use of $\hat{A}_T$ rather than A_T would increase the variance of the prewhitened kernel estimator, but have little effect on its bias. In fact, simulations (not reported here) verify that this is exactly what happens. Furthermore, for the sample size considered here, the increase in variance of the prewhitened kernel estimator due to the estimation of A_r more than offsets the reduction in the estimator's variance due to prewhitening. In consequence, the prewhitened kernel estimator has higher variance and MSE than the standard kernel estimator, but lower bias. The lower bias is responsible for the prewhitened kernel estimator's improved CI coverage probabilities and improved t–statistic rejection rates.

APPENDIX

PROOF OF THEOREM 1: To establish part (a), we apply Theorem 3(a) of Andrews (1990) to the estimator $\hat{J}_T^*(\hat{S}_T)$ of J_T^* with $\hat{\theta}$ and θ_0 elongated to include $(\hat{A}_1, \ldots, \hat{A}_b)$ and $A_1, \ldots, A_b)$ respectively. To apply Theorem 3(a) in this fashion, Assumptions A, A*, B, and E must hold with $\{V_t\}$ replaced by $\{V_t^*\}$. It is not difficult to show, however, that if these assumptions hold as stated, then they also hold with $\{V_t\}$ replaced by $\{V_t^*\}$. Furthermore, if these assumptions hold as stated, then they also hold with $\hat{\theta}$ and θ_0 elongated as above provided Assumption H(i) holds. Thus, Theorem 3(a) and the results of Section 8 of Andrews (1990) yield $\hat{J}_T^*(\hat{S}_T^*) - J_T^* \xrightarrow{p} \underset{\sim}{0}$. This result, Assumption H, and equation (2.8) give the desired result

$$\text{(A.1)} \qquad \hat{J}_{Tpw} - J_T = \hat{D}\hat{J}_T^*(\hat{S}_T^*)\hat{D}' - DJ_T^*D' + o_p(1) = o_p(1) .$$

To establish part (b), we apply Theorem 3(b) and the corresponding results of Section 8 of Andrews (1990) in the same manner as above.

The second result of part (c) follows from the fact that $f = Df^*D'$, see equation (3.5), and the property of the commutation matrix K_{pp} that $(D \otimes D)K_{pp} = K_{pp}(D \otimes D)$, see Magnus and Neudecker (1979, Thm. 3.1(ix)).

To establish the first result of part (c), we apply Theorem 3(c) of Andrews (1990) in the same manner as above to obtain

$$\text{(A.2)} \qquad \begin{aligned} &\lim_{h\to\infty} \lim_{T\to\infty} MSE_h(T/S_{\xi T}^*, \hat{J}_T^*, J_T^*, W_T^*) \\ &= 4\pi^2\Big[k_q^2(\text{vec } f_*^{(q)})' W^* \text{ vec } f_*^{(q)}/\gamma_\xi^* + \textstyle\int k^2(x)dx \text{ tr } W^*(I_{p^2} + K_{pp})(f^*\otimes f^*)\Big] , \end{aligned}$$

where $\hat{J}_T^*$ denotes $\hat{J}_T^*(\hat{S}_T^*)$, $S_{\xi T}^* = \left[qk_q^2\alpha_\xi^* T/\int k^2(x)dx\right]^{1/(2q+1)}$, $W_T^* = (\hat{D} \otimes \hat{D})' W_T(\hat{D} \otimes \hat{D})$, and $\gamma_\xi^* = qk_q^2\alpha_\xi^*/\int k^2(x)dx$. Equation (A.2) and some algebra establish that $\lim_{h\to\infty}\lim_{T\to\infty} MSE_h(T^{2q/(2q+1)}, \hat{J}_T^*, J_T^*, W_T^*)$ equals the righthand side of the first result of part (c). Thus, it remains to show that

$$\text{(A.3)} \quad \begin{aligned} &\lim_{h\to\infty}\lim_{T\to\infty} MSE_h(T^{2q/(2q+1)}, \hat{J}_{Tpw}, J_T, W_T) \\ &= \lim_{h\to\infty}\lim_{T\to\infty} MSE_h(T^{2q/(2q+1)}, \hat{J}_T^*, J_T^*, W_T^*) \, . \end{aligned}$$

Let $J_T^{**} = \hat{D}^{-1}J_T(\hat{D}^{-1})'$. The lefthand side of (A.3) equals

$$\text{(A.4)} \quad \begin{aligned} &\lim_{h\to\infty}\lim_{T\to\infty} E\min\left\{T^{2q/(2q+1)}\text{vec}(\hat{D}(\hat{J}_T^* - J_T^{**})\hat{D}')' W_T \,\text{vec}(\hat{D}(\hat{J}_T^* - J_T^{**})\hat{D}'), h\right\} \\ &= \lim_{h\to\infty}\lim_{T\to\infty} E\min\left\{T^{2q/(2q+1)}\text{vec}(\hat{J}_T^* - J_T^{**})'(\hat{D}\otimes\hat{D})' W_T(D\otimes D)\text{vec}(\hat{J}_T^* - J_T^{**}), h\right\} \\ &\le \lim_{h\to\infty}\lim_{T\to\infty} E\min\left\{T^{2q/(2q+1)}\text{vec}(\hat{J}_T^* - J_T^*)' W_T^* \,\text{vec}(\hat{J}_T^* - J_T^*), h\right\} \\ &\quad + \lim_{h\to\infty}\lim_{T\to\infty} E\min\left\{T^{2q/(2q+1)}\text{vec}(J_T^* - J_T^{**})' W_T^* \,\text{vec}(J_T^* - J_T^{**}), h\right\} \\ &\quad + 2\lim_{h\to\infty}\lim_{T\to\infty} E\min\left\{T^{2q/(2q+1)}\left|\text{vec}(\hat{J}_T^* - J_T^*)' W_T^* \,\text{vec}(J_T^* - J_T^{**})\right|, h\right\} \\ &= G_1 + G_2 + G_3 \text{ (say).} \end{aligned}$$

Since G_1 equals the righthand side of (A.3), it suffices to show that $G_2 = G_3 = 0$. $G_2 = 0$ because $T^{q/(2q+1)}(J_T^* - J_T^{**}) = T^{q/(2q+1)}\left[D^{-1}J_T(D^{-1})' - \hat{D}^{-1}J_T(\hat{D}^{-1})'\right] = o_p(1)$, $W_T^* = O_p(1)$, and loss is truncated at h (using the fact that if $X_T \xrightarrow{p} 0$ and $\{X_T : T \ge 1\}$ is uniformly integrable, then $EX_T \to 0$). $G_3 = 0$ because $T^{q/(2q+1)}(\hat{J}_T^* - J_T^*) = O_p(1)$, $T^{q/(2q+1)}(J_T^* - J_T^{**}) = o_p(1)$, $W_T^* = O_p(1)$, and loss is truncated at h . □

FOOTNOTES

[1]The first author gratefully acknowledges the research support of the National Science Foundation via grant Nos. SES–8821021 and SES–9040137.

[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 4, where $b = 1$, we take $\hat{A}\ (= \hat{A}_1)$ to be an eigenvalue adjusted version of the LS estimator $\hat{A}_{LS}$. See below for details of the adjustment.

[3]Grenander and Rosenblatt (1957, pp. 273–274) carry out calculations that are similar to some of those given below. However, they compare the MSE of prewhitened and non-prewhitened kernel estimators when the same bandwidth sequence is used for each estimator. Such comparisons do not take account of the fact that one would want to use a different bandwidth sequence depending upon whether prewhitening has been done or not.

[4]The $p \times p$ estimator $\hat{A}$ that is used 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 $\hat{A} = \hat{B}\hat{\Delta}\hat{C}'$.

[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.

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

Since $E\tilde{x} = \underset{\sim}{0}$ 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.

[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.

[8]The reason for this is that for the non–intercept 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 non–intercept 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 $\hat{U}_t - U_t$.

[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.

REFERENCES

Andrews, D. W. K. (1988): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Cowles Foundation Discussion Paper No. 877, Yale University.

_______ (1990): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Cowles Foundation Discussion Paper No. 877R, Yale University.
Forthcoming in Econometrica.

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, 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 (1987): "Estimation of the Covariance Matrix of the Least Squares Regression Coefficients When the Disturbance Covariance Matrix is of Unknown Form," unpublished manuscript, Department of Statistics, University of Michigan.

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

Magnus, J. R. and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7, 381–394.

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. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

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

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. (1988): "Automatic Generalized Least Squares," unpublished manuscript, London School of Economics.

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

TABLE 1

Values of the Function $h(\eta)$

η	−.15	−.10	−.05	.1	.3	.5	.7	.9	.95
$h(\eta)$	−.533	−.264	−.106	.170	.417	.610	.777	.928	.964

TABLE 2

Values of the Functions $g(\tau)$ and $A(\tau)$

τ	−.1	−.05	.1	.2	.3	.5	.7	.99
$g(\tau)$	−.8	−.145	.126	.184	.217	.250	.263	.268
$A(\tau)$	−.099	−.050	.099	.192	.275	.400	.470	.500

TABLE 3

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

TABLE 4

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	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	.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

TABLE 5

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	Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
0	1.47	QS–PW	.0072	.66	.66	98.2	93.5	87.6
		QS	–.14	.60	.63	98.4	92.0	85.8
		PARA	–.49	.044	.29	96.3	88.6	81.0
.3	1.66	QS–PW	.0080	.84	.84	98.1	92.7	86.3
		QS	–.24	.49	.55	97.8	91.0	84.2
		PARA	–.59	.070	.41	95.5	87.8	80.1
.5	2.13	QS–PW	–.098	1.78	1.79	97.9	92.2	85.8
		QS	–.54	.72	1.01	96.7	88.6	81.8
		PARA	–.87	.13	.89	94.5	85.3	78.2
.7	3.29	QS–PW	–.36	5.89	6.02	95.8	89.9	83.4
		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
.9	7.15	QS–PW	–2.32	38.2	43.5	89.6	80.5	71.5
		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
.95	9.58	QS–PW	–3.53	246.	258.	83.9	74.1	66.9
		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
–.3	1.68	QS–PW	.034	.89	.89	98.7	94.5	88.7
		QS	–.24	.49	.55	98.2	93.1	86.4
		PARA	–.57	.080	.40	97.0	89.1	82.0
–.5	2.17	QS–PW	.050	2.46	2.46	97.8	92.0	87.0
		QS	–.48	.93	1.16	95.9	88.6	83.0
		PARA	–.79	.19	.81	94.7	87.9	80.5

TABLE 6

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	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

TABLE 7

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	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