HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATION

By

Donald W. K. Andrews

July 1988

COWLES FOUNDATION DISCUSSION PAPER NO. 877

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 877

HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX ESTIMATION

by

Donald W. K. Andrews

July 1988

ABSTRACT

This paper is concerned with the estimation of covariance matrices in the presence of heteroskedasticity and autocorrelation of unknown forms. Currently available estimators that are designed for this context depend upon the choice of a lag truncation parameter and a weighting scheme. Results in the literature provide a condition on the growth rate of the lag truncation parameter as $T \to \infty$ that is sufficient for consistency. No results are available, however, regarding the choice of lag truncation parameter for a fixed sample size, regarding data-dependent automatic lag truncation parameters, or regarding the choice of weighting scheme. In consequence, available estimators are not entirely operational and the relative merits of the estimators are unknown.

This paper addresses these problems. Upper and lower bounds on the asymptotic mean squared error of each estimator in a given class are determined and compared. Asymptotically optimal kernel/weighting scheme and bandwidth/lag truncation parameters are obtained using a minimax asymptotic mean squared error criterion. Higher order asymptotically optimal corrections to the first order optimal bandwidth/lag truncation parameters are introduced. Using these results, data-dependent automatic bandwidth/lag truncation parameters are defined and are shown to possess certain asymptotic optimality properties. Finite sample properties of the estimators are analyzed via Monte Carlo simulation.

1. INTRODUCTION

This paper considers heteroskedasticity and autocorrelation consistent (HAC) estimation of covariance matrices of parameter estimators in linear and nonlinear models. A prime example is the estimation of the covariance matrix of the least squares (LS) estimator in a linear regression model with heteroskedastic, temporally dependent errors of unknown forms. Other examples include covariance matrix estimation of LS estimators of nonlinear regression models and unit root models and of two and three stage least squares and generalized method of moments estimators of nonlinear simultaneous equations models. HAC estimators have found numerous applications recently in the macroeconomic, financial, and international financial literature, e.g., see Campbell and Clarida (1987), Mishkin (1987), and Hardouvelis (1988).

The paper has several objectives. The first is to analyze and compare the properties of several HAC estimators that have been proposed in the literature, see Levine (1983), White (1984, Ch. 6), White and Domowitz (1984), Gallant (1987, pp. 551, 573), Newey and West (1987), and Gallant and White (1988, Ch. 6). Currently only the consistency of such estimators has been established. In consequence, the relative merits of the estimators are unknown.

The second objective is to make existing estimators operational by determining suitable values for the lag truncation or bandwidth parameters that are used to define the estimators. At present, no guidance is available regarding the choice of these parameters for a given finite sample situation.[2] This is a serious problem, because the performance of these estimators depends greatly on the choice of lag truncation or bandwidth parameters. Given the lack of theoretical guidance available, it is not surprising that common choices for these parameters in applications are not appropriate. For example, in cases where the errors are known to be m-dependent, it is not uncommon to see the Newey-West estimator used with the lag length set equal to m or some multiple of m. This rule for choosing the lag length is inappropriate in general and actually yields an inconsistent estimator.

The third objective of the paper is to obtain an optimal estimator out of a class of kernel estimators that contains the HAC estimators that have been proposed in the literature. The optimal estimator, called a quadratic spectral (QS) estimator, is obtained using a minimax optimality criterion. Both fixed bandwidth and automatic bandwidth estimation procedures are considered.

The fourth objective of the paper is to investigate the finite sample performance of the asymptotically optimal kernel and bandwidth parameters. Monte Carlo simulation is used. Different kernels and bandwidth parameters are compared. In addition, kernel estimators are compared with standard covariance matrix estimators.

The class of kernel HAC estimators considered here includes estimators that give some weight to all T-1 lags of the sample autocovariance function. Such estimators have not been considered previously. As it turns out, the optimal estimator is of this form.

The consistency of kernel HAC estimators is established under much weaker conditions on the growth rate of the lag truncation/bandwidth parameter S_T than is available elsewhere. Instead of requiring $S_T = o(T^{1/4})$ or $O(T^{1/5})$, as in the papers referenced above, or $S_T = o(T^{1/2})$, as in Keener, Kmenta, and Weber (1987), we just require $S_T = o(T)$ as $T \to \infty$. For the Newey-West estimator these results are of particular interest because the optimal growth rate of S_T for this estimator is shown to be $T^{1/3}$, which is larger than previous consistency results allow. (Keener, Kmenta, and Weber's (1987) results do not apply to the

Newey-West estimator.) Our results also provide rates of convergence of the estimators to the estimand. These rates differ for different choices of kernel and bandwidth parameter S_T.

To achieve the objectives outlined above, the general approach taken in this paper is to exploit existing results in the literature on kernel density estimation--both spectral and probability--whenever possible. For this purpose, the following references are particularly pertinent: Parzen (1957), Priestley (1962), Epanechnikov (1969), Deheuvels (1977), Sacks and Ylvisacker (1981), Sheather (1986), Betrao and Bloomfield (1987), and Hall and Marron (1987b, c).

Robinson (1988) recently has considered the choice of bandwidth parameters for kernel covariance matrix estimators for linear regression models with errors that are conditionally and unconditionally homoskedastic. The estimators that he considers exploit the homoskedasticity of the errors, and hence, differ from the kernel HAC estimators considered here. Robinson's approach to the choice of bandwidth parameter also differs from that considered here, because we consider "plug-in" automatic bandwidths, whereas he considers cross-validated automatic bandwidths. Each approach has some advantages.

We now comment on the usefulness of HAC estimators in econometric practice. First, as is well known, positive autocorrelation of the errors in linear regression models dramatically increases the variance of the LS estimator in a way that is not captured by the standard LS covariance matrix estimator. In addition, heteroskedasticity can cause a substantial bias, positive or negative, for the standard LS covariance matrix estimator. The same problems with standard covariance matrix estimators arise with a host of other parameter estimators in linear and nonlinear models with single and multiple equations, see the references above.

To exemplify the severity of the problem with autocorrelation, consider a simple linear regression model with homoskedastic first order autoregressive (AR(1)) regressor and error, each with the same AR parameter ρ. In this model, the ratio of the true variance of the LS slope parameter estimator to the variance given by the standard formula for independent errors is approximately $(1 + \rho^2)/(1 - \rho^2)$. Hence, for $\rho \geq .3$ the standard formula yields a substantial downward bias. For example, for $\rho = .7$ and $.9$, the standard formula understates the true variance by factors of approximately three and ten, respectively.

Second, an alternative to the use of a HAC estimator is to specify a finite dimensional parametric model for heteroskedasticity and autocorrelation and to use the covariance matrix estimator that corresponds to this model. Although this approach is much preferred to ignoring heteroskedasticity and autocorrelation, it has several drawbacks. One usually has little information available regarding the form of heteroskedasticity and autocorrelation. In consequence, the choice of parametric model is difficult and subject to considerable error. Furthermore, if the finite dimensional parametric model is incorrect, the corresponding covariance matrix estimator generally is inconsistent. HAC estimators avoid these problems.

Third, if one does specify a finite dimensional parametric model for heteroskedasticity and autocorrelation, then a HAC estimator still can be useful. It can be used to capture any heteroskedasticity and autocorrelation remaining in the data due to misspecification of the parametric model.

The remainder of the paper is organized as follows: Section 2 describes the estimation problem of concern and introduces notation and definitions regarding the class of kernel HAC estimators under study. Section 3 presents upper and lower bounds on the asymptotic variance, bias, and mean squared error of kernel HAC estimators. Section 4 establishes the optimality of a particular kernel, viz., the QS kernel, under an asymptotic minimax criterion, using the results of Section 3. Section 5 determines first order and higher order asymptotically optimal bandwidth/lag truncation values under a minimax criterion using the results of Section 3. Section 6 extends the results to the standard case where parameters, such as regression parameters, are estimated. Section 7 introduces a method for data-dependent "automatic" determination of the bandwidth parameter. The estimators based on these automatic bandwidth parameters usually are the most appropriate for use in practice. Section 8 establishes the asymptotic mean squared error properties of these estimators and gives an asymptotic optimality property for them. Section 9 presents Monte Carlo results regarding the finite sample behavior of the estimators considered in earlier sections. Section 10 provides a summary of the results of the paper. An appendix contains proofs of results given in the paper.

Those interested primarily in the definition of the preferred HAC estimator should read Section 2 up to equation (2.8) and Section 7. The preferred estimator is a kernel estimator that uses the QS kernel and an automatic bandwidth procedure. If computational time is a binding constraint, the Parzen kernel can be used in place of the QS kernel.

2. A CLASS OF ESTIMATORS

To motivate the definition of the estimand given below, consider the linear regression model and LS estimator:

$$Y_t = X_t'\theta_0 + U_t,\ t = 1, \ldots, T,\ \hat{\theta} = \Big(\sum_{t=1}^{T} X_tX_t'\Big)^{-1} \sum_{t=1}^{T} X_tY_t\ , \quad \text{and}$$

(2.1)

$$\mathrm{Var}(\sqrt{T}(\hat{\theta}-\theta_0)) = \left(\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\right)^{-1} \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EU_sX_s(U_tX_t)'\left(\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\right)^{-1} .$$

Since X_t is observed, consistent estimation of $\mathrm{Var}(\sqrt{T}(\hat{\theta} - \theta_0))$ just requires a consistent estimator of $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EU_sX_s(U_tX_t)'$.

More generally, many parameter estimators $\hat{\theta}$ in nonlinear dynamic models satisfy

$$(B_TJ_{TP}B_T')^{-1/2}\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, I_r) \text{ as } T \to \infty, \text{ where}$$

(2.2)

$$J_{TP} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} E_PV_s(\theta_0)V_t(\theta_0)',$$

B_T is a nonrandom $r \times p$ matrix, $V_t(\theta)$ is a random p-vector for each $\theta \in \Theta \subset R^r$, P is the underlying distribution generating the data, and "obvious" estimators $\hat{B}_T$ exist such that $\hat{B}_T - B_T \xrightarrow{p} 0$ as $T \to \infty$. The estimators $\hat{B}_T$ usually are just sample analogues of B_T with θ_0 replaced by $\hat{\theta}$. See Hansen (1982), Gallant (1987, Ch. 7), Gallant and White (1988), and Andrews and Fair (1988) for the treatment of broad classes of parameter estimators and models that satisfy these conditions.[3] Since consistent estimators of B_T exist, one can estimate the "asymptotic variance" of $\sqrt{T}(\hat{\theta} - \theta_0)$, viz., $B_TJ_{TP}B_T'$, if one has a consistent estimator of J_{TP}. In consequence, we concentrate our attention on the estimation of J_{TP}.

The primary ingredient of J_{TP} is the vector $V_t(\theta)$. In the case of LS estimation of a linear regression model, $V_t(\theta) = (Y_t - X_t'\theta)X_t$. In the case of LS estimation of a nonlinear regression model, $V_t(\theta) = (Y_t - g(X_t,\theta))\frac{\partial}{\partial\theta}g(X_t,\theta)$, where $g(X_t,\theta)$ is the regression function. In the case of pseudo-ML estimation, $V_t(\theta)$ is the score function for the t-th observation. (That is, it is the derivative of the log likelihood function for the t-th observation conditional on the earlier observations.) In the case of instrumental variables estimation of a dynamic nonlinear simultaneous equations model, $V_t(\theta)$ is the Kronecker product of the vector of model equations evaluated at θ with the instrumental variables. In the case of unit root models, the LS estimator does not satisfy

(2.2). Nevertheless, one still needs to estimate the value of an expression that has the same form as J_{TP} with $V_t(\theta) = Y_t - Y_{t-1}$ or $V_t(\theta) = Y_t - \theta Y_{t-1}$, where $\{Y_t\}$ is the unit root process under investigation, see Phillips (1987).

By change of variables, the estimand J_{TP} can be rewritten as

$$J_{TP} = \sum_{j=-T+1}^{T-1} \Gamma_{TP}(j), \text{ where}$$

(2.3)

$$\Gamma_{TP}(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} E_P V_t V_{t-j}' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} E_P V_{t+j} V_t' & \text{for } j < 0 \end{cases}$$

and $V_t = V_t(\theta_0)$, $t = 1, \ldots, T$.

Define the <u>sample size T spectral density matrix</u> to be

$$\text{(2.4)} \qquad f_{TP}(\lambda) = \frac{1}{2\pi}\sum_{j=-T+1}^{T-1} \Gamma_{TP}(j)e^{-ij\lambda} \quad \text{for} \quad \lambda \in [-\pi, \pi],$$

where $i = \sqrt{-1}$. The estimand J_{TP} equals 2π times the sample size T spectral density matrix evaluated at $\lambda = 0$. If $\{V_t\}$ was a mean zero second order stationary sequence, then the limit as $T \to \infty$ of $f_{TP}(\lambda)$ would equal the spectral density matrix of $\{V_t\}$ as it is ordinarily defined. With the given definition of $f_{TP}(\lambda)$, however, $\{V_t\}$ need not be second order stationary. In fact, for many applications of interest, $\{V_t\}$ is not second order stationary. For example, in the linear regression model with estimation by LS, $V_t = U_t X_t$. Hence, if the regressors are fixed or the errors are unconditionally heteroskedastic, then $\{V_t\}$ is not second order stationary.

The fact that J_{TP} is proportional to the sample size T spectral density matrix at $\lambda = 0$ motivates the use of spectral density estimators to estimate J_{TP}. For the case of stationary $\{V_t(\theta)\}$ sequences, Hansen (1982, p. 1047) and

Phillips and Ouliaris (1988) already have suggested using the latter to estimate J_{TP}. Furthermore, it is easy to show that in the second order stationary context with known θ_0 the estimators of White (1984), Gallant (1987), and Newey and West (1987) correspond to the "truncated periodogram," Parzen, and modified Bartlett spectral density estimators, respectively, evaluated at $\lambda = 0$. The aforementioned authors have established consistency of their estimators, however, in the more general context in which $(V_t(\theta))$ is non-stationary and θ_0 is unknown.

The class of estimators we consider corresponds to Parzen's (1957) class of kernel estimators of the spectral density matrix. We consider estimators of the form

$$\hat{J}_T = \frac{T}{T-r} \sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\Gamma}(j), \text{ where}$$

$$\hat{\Gamma}(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} \hat{V}_t \hat{V}_{t-j} & \text{for } j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} \hat{V}_{t+j} \hat{V}'_t & \text{for } j < 0, \end{cases} \tag{2.5}$$

$\hat{V}_t = V_t(\hat{\theta})$, $k(\cdot)$ is a real-valued kernel in the set K_1 defined below, and S_T is a bandwidth parameter.[4] The factor $T/(T-r)$ is a small sample degrees of freedom adjustment that is introduced to offset the effect of estimation of the r-vector θ. In Sections 3-6, we consider estimators $\hat{J}_T$ for which S_T is a given non-random scalar. In Sections 7-9, however, we consider "automatic" estimators $\hat{J}_T$ for which S_T is a random function of the data.

The class of kernels K_1 is given by

$$K_1 = \Big\{k(\cdot) : R \to [-1,1] \Big| k(0) = 1,\ k(x) = k(-x)\ \forall x \in R,\ \int_{-\infty}^{\infty} k^2(x)dx = 1,\ k(\cdot) \text{ is continuous at 0 and at all but a finite number of other points}\Big\}. \tag{2.6}$$

The conditions $k(0) = 1$ and $k(\cdot)$ is continuous at 0 reflect the fact that for T large relative to j we want the weight given to $\hat{\Gamma}(j)$ to be close to one. The requirement that $k(\cdot)$ be an even function is natural because $b'\Gamma_{TP}(j)b$ is an even function of j for all $b \in R^p$. The condition $\int_{-\infty}^{\infty} k^2(x)dx = 1$ is a normalization condition that differs from normalizations conventionally used in the spectral density literature. This choice of normalization is quite natural, however, and is important for one result given below.[5]

Examples of kernels include the following:

$$\text{Truncated:} \quad \bar{k}_{TR}(x) = \begin{cases} 1 & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Bartlett:} \quad \bar{k}_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(2.7) \qquad \text{Parzen:} \quad \bar{k}_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \le |x| \le 1/2 \\ 2(1 - |x|)^3 & \text{for } 1/2 \le |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Tukey-Hanning}: \quad \bar{k}_{TH}(x) = \begin{cases} (1 + \cos(\pi x))/2 & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Quadratic Spectral:} \quad k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

The estimators $\hat{J}_T$ corresponding to the truncated, Bartlett, and Parzen kernels are the estimators proposed by White (1984), Newey and West (1987), and Gallant (1987), respectively. The Tukey-Hanning and QS kernels have not been considered in the literature concerning HAC estimation. The Tukey-Hanning kernel is popular in the spectral density estimation literature, however, and the QS kernel has been considered in the spectral and probability density estimation literature by Priestley (1962) and Epanechnikov (1969), respectively.

The QS kernel $k_{QS}(x)$ is in K_1. On the other hand, the kernels $\bar{k}_{TR}(x)$, $\bar{k}_{BT}(x)$, $\bar{k}_{PR}(x)$, and $\bar{k}_{TH}(x)$ are not in K_1 because they are normalized in the conventional way such that $k(x) = 0$ for $|x| > 1$. These kernels can be renormalized, however, to lie in K_1. The renormalized kernels are given by $k_{TR}(x) = \bar{k}_{TR}(c_{TR}x)$, $k_{BT}(x) = \bar{k}_{BT}(c_{BT}x)$, and $k_{PR}(x) = \bar{k}_{PR}(c_{PR}x)$, and $k_{TH}(x) = \bar{k}_{TH}(c_{TH}x)$, where $c_{TR} = \int \bar{k}^2_{TR}(x)dx = 2$, $c_{BT} = 2/3$, $c_{PR} = .539285$, and $c_{TH} = 3/4$. If M_T is the bandwidth parameter used with the normalization of (2.7), then $S_T = c_v M_T$ for v = TR, BT, PR, and TH is the corresponding bandwidth parameter for the kernel normalized to be in K_1. That is,

$$\hat{J}_T = \frac{T}{T-r}\sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\Gamma}(j) = \frac{T}{T-r}\sum_{j=-T+1}^{T-1} \bar{k}(j/M_T)\hat{\Gamma}(j) \ . \tag{2.8}$$

If $\bar{k}(x) = 0$ for $|x| > 0$ (and $\bar{k}(x) \neq 0$ for some $|x|$ arbitrarily to 1), then the parameter M_T is referred to as the <u>lag truncation parameter</u>, because lags of order $j > M_T$ receive zero weight.[6] Since some kernels in K_1 are non-zero for arbitrarily large values of x, it is not possible to renormalize all kernels in K_1 such that $k(x) = 0$ for $|x| > 1$. Thus, lag truncation parameters do not exist for all kernels in K_1. The QS kernel is an example.

Figure 1 graphs the five kernels of (2.7), but using the normalization of K_1. For a given value of S_T, the figure illustrates the different weights the kernels put on the lagged correlations. For example, if $S_T = 3$, then $k_{BT}(1/3)$, $k_{BT}(2/3)$, ... are the weights the Bartlett kernel puts on $\hat{\Gamma}(1)$, $\hat{\Gamma}(2)$, ... (The normalization of K_1 is appropriate for this figure, because the different kernels are most easily compared when any given S_T value is equally suitable for any kernel. With the normalization of K_1, the latter holds true because given any sequence $\{S_T\}$ the asymptotic variances of all five estimators are the same--only their asymptotic biases vary.)

For some results below, we consider certain subsets of $\mathcal{K}_1$. The first subset contains all kernels that are non-negative:

$$\text{(2.9)} \qquad \mathcal{K}_0 = \{k(\cdot) \in \mathcal{K}_1 | k(x) \geq 0 \ \forall x \in R\}.$$

The first four examples of (2.7) are in $\mathcal{K}_0$ when suitably normalized, but the QS kernel is not because it takes on negative values of small absolute magnitude for some values of x. There is no compelling reason for restricting consideration to kernels in $\mathcal{K}_0$, but some results given below require it.

A second subset of $\mathcal{K}_1$ that is of interest is

$$\text{(2.10)} \qquad \begin{aligned} &\mathcal{K}_2 = \{k(\cdot) \in \mathcal{K}_1 | K(\lambda) \geq 0 \ \forall \lambda \in R\}, \text{ where} \\ &K(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(x) e^{-ix\lambda} dx \quad \text{for} \quad \lambda \in R. \end{aligned}$$

The function $K(\lambda)$ is referred to as the <u>spectral window generator</u> corresponding to the kernel $k(\cdot)$. The set $\mathcal{K}_2$ contains all kernels in $\mathcal{K}_1$ that necessarily generate positive semi-definite (psd) estimates in finite samples. As emphasized by Newey and West (1987), this property usually is highly desirable. The set $\mathcal{K}_2$ contains the Bartlett, Parzen, and QS kernels, but not the truncated kernel (since $K_{TR}(\lambda) \propto \sin(\lambda)/\lambda$) or the Tukey-Hanning kernel.

The reason why kernels in $\mathcal{K}_2$ generate psd estimates is that estimators of the form (2.5) can be written as weighted averages of the periodogram matrix at different frequencies λ with weights given by $K(\lambda)$ (e.g., see Priestley (1981, pp. 580-581)). Since the periodogram is psd, so is an estimator $\hat{J}_T$ provided $K(\lambda) \geq 0$ for all λ.

The distribution P that generates $\{V_t\}$ is required to satisfy certain conditions. Let $\Gamma_P(t,t+j)$ denote $E_P V_t V_{t+j}'$ and let $\kappa_{Pb}(t,t+j,t+\ell,t+m)$ denote the fourth order cumulant of $(b'V_t, b'V_{t+j}, b'V_{t+\ell}, b'V_{t+m})$ under P for arbitrary

$b \in R^p$. That is, $\kappa_{Pb}(t,t+j,t+\ell,t+m) = E_P b'V_t b'V_{t+j} b'V_{t+\ell} b'V_{t+m}$ $- E_P b'\tilde{V}_t b'\tilde{V}_{t+j} b'\tilde{V}_{t+\ell} b'\tilde{V}_{t+m}$, where $\{\tilde{V}_t\}$ is a Gaussian sequence with the same mean and covariance structure as $\{V_t\}$. Let P_v for $v = 0, 1$ be distributions such that $\{V_t\}$ is a mean zero, second order stationary sequence under P_v with autocovariance function $\{\Gamma_v(j) : j = 0, \pm 1, \ldots\}$ and $\{b'V_t\}$ has spectral density function $f_{vb}(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} b'\Gamma_v(j) b e^{-ij\lambda}$ for all $b \in R^p$. Define

$$\mathcal{P}_1 = \Big\{P : -\Gamma_1(j) \le \Gamma_P(t,t+j) \le \Gamma_1(j) \text{ and } |\kappa_{Pb}(t,t+j,t+\ell,t+m)| \le \kappa_b(j,\ell,m)$$

$$\forall t \ge 1,\ \forall j,\ell,m \ge -t+1,\ \forall b \in R^p, \text{ for some function } \kappa_b(j,\ell,m) \tag{2.11}$$

$$\text{that satisfies } \sum_{j=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} \kappa_b(j,\ell,m) < \infty\Big\} \text{ and}$$

$$\mathcal{P}_0 = \Big\{P : \underset{\sim}{0} \le \Gamma_0(j) \le \Gamma_P(t,t+j)\ \forall t \ge 1,\ \forall j \ge -t+1 \text{ and } \kappa_{Pb}(t,t+j,t+\ell,t+m)$$

$$\text{satisfies the same condition as in } \mathcal{P}_1\Big\}. \tag{2.12}$$

Most of the results below hold for any true distribution $P \in \mathcal{P}_1$, while some hold for $P \in \mathcal{P}_0$, and others for $P \in \mathcal{P}_0 \cap \mathcal{P}_1$. The set $\mathcal{P}_1$ requires the autocovariances of $\{V_t\}$ under P to be dominated by those of some second-order stationary process whose spectral density exists. This allows considerable variability of $\Gamma_P(t,t+j)$ for fixed j and $t = 1, 2, \ldots$. In addition, it allows considerable temporal dependence of $\{V_t\}$. If $\{V_t\}$ is Gaussian, then the condition on $\kappa_{Pb}(t,t+j,t+\ell,t+m)$ is satisfied trivially since the latter is identically zero.

The set $\mathcal{P}_0$ requires the autocovariances of $\{V_t\}$ under P to be bounded below by the psd autocovariances of a second order stationary process. To establish the consistency of $\hat{J}_T$, we do not need to have $P \in \mathcal{P}_0$. To obtain a lower bound on its asymptotic MSE, however, we need to assume $P \in \mathcal{P}_0$.

3. BOUNDS ON ASYMPTOTIC VARIANCE, BIAS, AND MSE

In this section we determine the asymptotic MSE properties of kernel HAC estimators $\hat{J}_T$. These properties are used in subsequent sections to obtain a minimax optimal kernel and minimax optimal sequences of bandwidth parameters $\{S_T\}$. The proofs of the results use results of Parzen (1957) for spectral density estimators.

Let $\tilde{J}_T$ be the pseudo-estimator that is identical to $\hat{J}_T$ but is based on the unobserved sequence $\{V_t\} = \{V_t(\theta_0)\}$ rather than $\{\hat{V}_t\} = \{V_t(\hat{\theta})\}$ and is defined without the degrees of freedom correction $T/(T-r)$:

$$(3.1) \qquad \tilde{J}_T = \sum_{j=-T+1}^{T-1} k(j/S_T)\tilde{\Gamma}(j) \text{ and } \tilde{\Gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} V_t V_{t-j}' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} V_{t+j} V_t' & \text{for } j < 0. \end{cases}$$

In this section, we present results for the pseudo-estimator $\tilde{J}_T$. In Section 6, analogous results are shown to hold for the estimator $\hat{J}_T$. The latter are obtained by showing that the difference between the asymptotic MSEs of $\hat{J}_T$ and $\tilde{J}_T$, suitably normalized, goes to zero as $T \to \infty$.

Our first result gives upper and lower bounds on the asymptotic variance of $\tilde{J}_T$. These bounds depend on $f_{vb} = f_{vb}(0)$. Let $Var_P(\cdot)$ denote the variance of $\cdot$ when P is the underlying distribution.

LEMMA 1: *For all* $b \in R^p$, *if* $S_T \to \infty$ *and* $S_T/T \to 0$ *as* $T \to \infty$, *then*

(a) $\lim_{T\to\infty} \inf_{P\in\mathcal{P}_0} \frac{T}{S_T} Var_P(b'\tilde{J}_T b) = \lim_{T\to\infty} \frac{T}{S_T} Var_{P_0}(b'\tilde{J}_T b) = 8\pi^2 f_{0b}^2$ *for any* $k(\cdot) \in \mathcal{K}_0$

and

(b) $\lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} \frac{T}{S_T} Var_P(b'\tilde{J}_T b) = \lim_{T\to\infty} \frac{T}{S_T} Var_{P_1}(b'\tilde{J}_T b) = 8\pi^2 f_{1b}^2$ *for any* $k(\cdot) \in \mathcal{K}_1$.

COMMENTS: 1. The kernels $k(\cdot) \in K_1$ have been normalized (via $\int_{-\infty}^{\infty} k^2(x)dx = 1$) such that the upper and lower bounds on the asymptotic variance of $\tilde{J}_T$ are independent of the kernel. Only the asymptotic bias of $\tilde{J}_T$ depends on the kernel, see Lemma 2 below.

2. The rate of convergence to zero of the variance of $\tilde{J}_T$ depends on $\{S_T\}$. The rate is $O(S_T/T)$ and is slower than the standard rate obtained by parametric estimators, viz., $O(1/T)$. The slower is the rate of divergence to infinity of $\{S_T\}$ (or equivalently of the lag truncation parameters $\{M_T\}$ when they exist), the faster is the rate of convergence to zero of the variance of $\tilde{J}_T$.

Upper and lower bounds on the asymptotic bias of kernel estimators depend on the smoothness of the kernel at zero and on the smoothness of the spectral density matrix of $\{V_t\}$ at zero under P_1 and P_0. Following Parzen (1957), define

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

If q is an even integer, then $k_q = -\frac{1}{q!} \left.\frac{d^q k(x)}{dx^q}\right|_{x=0}$ and $k_q < \infty$ if and only if $k(x)$ is q times differentiable at zero. For a kernel $\bar{k}(\cdot)$ for which $\int \bar{k}^2(x)dx \neq 1$, let $\bar{k}_q$ be defined analogously to k_q with $k(\cdot)$ replaced by $\bar{k}(\cdot)$ in (3.2). If $k(\cdot)$ and $\bar{k}(\cdot)$ are the same kernel but with different normalizations, then $k_q = (\int \bar{k}^2(x)dx)^q \bar{k}_q$.

For the truncated kernel, $k_q = \bar{k}_q = 0$ for all $q < \infty$. For the Bartlett kernel, $k_1 = 2/3$, $\bar{k}_1 = 1$, $k_q = \bar{k}_q = 0$ for $q < 1$, and $k_q = \bar{k}_q = \infty$ for $q > 1$. For the Parzen kernel, $k_2 = 1.744975$, $\bar{k}_2 = 6$, $k_q = \bar{k}_q = 0$ for $q < 2$, and $k_q = \bar{k}_q = \infty$ for $q > 2$. For the Tukey-Hanning kernel, $k_2 = 9\pi^2/64$, $\bar{k}_2 = \pi^2/4$, $k_q = \bar{k}_q = 0$ for $q < 2$, and $k_q = \bar{k}_q = \infty$ for $q > 2$. For the QS kernel, $k_2 = 1.421223$, $k_q = 0$ for $q < 2$, and $k_q = \infty$ for $q > 2$.

The smoothness of $f_{vb}(\lambda)$ at $\lambda = 0$ is indexed by

(3.3) $$f_{vb}^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q b'\Gamma_v(j)b \quad \text{for} \quad q \in [0,\infty) \quad \text{and} \quad v = 0,\ 1.$$

If q is even, then $f_{vb}^{(q)} = (-1)^{q/2} \left.\frac{d^q f_{vb}(\lambda)}{d\lambda^q}\right|_{\lambda=0}$ and $f_{vb}^{(q)} < \infty$ if and only if $f_{vb}(\lambda)$ is q times differentiable at $\lambda = 0$. For example, if $\{b'V_t\}$ is a first-order autoregressive (AR(1)) process with autoregressive parameter $\rho \in [0,1)$ and innovation variance σ^2 under P_v, then $f_{vb}^{(q)} < \infty\ \forall q \in [0,\infty)$; $f_{vb}^{(1)} = \frac{\rho\sigma^2}{\pi(1-\rho)^3(1+\rho)}$, and $f_{vb}^{(2)} = \frac{\rho\sigma^2}{\pi(1-\rho)^4}$.

Upper and lower bounds on the asymptotic bias of $\tilde{J}_T$ are given in the following lemma.

LEMMA 2: *For all* $k(\cdot) \in K_1$, $q \in [0,\infty)$, *and* $b \in R^p$, *if* $S_T \to \infty$, $S_T^q/T \to 0$, *and* $S_T/T \to 0$ *as* $T \to \infty$, *then*

(a) $$\lim_{T\to\infty} \inf_{P\in P_0} S_T^q |E_P b'\tilde{J}_T b - b'J_{TP} b| = \lim_{T\to\infty} S_T^q |E_{P_0} b'\tilde{J}_T b - b'J_{TP_0} b| = 2\pi k_q f_{0b}^{(q)}$$

and

(b) $$\lim_{T\to\infty} \sup_{P\in P_1} S_T^q |E_P b'\tilde{J}_T b - b'J_{TP} b| = \lim_{T\to\infty} S_T^q |E_{P_1} b'\tilde{J}_T b - b'J_{TP_1} b| = 2\pi k_q f_{1b}^{(q)},$$

where $k_q f_{vb}^{(q)} = \infty$ *if* $k_q = \infty$ *and* $0 < f_{vb}^{(q)} \le \infty$ *or if* $f_{vb}^{(q)} = \infty$ *and* $0 < k_q \le \infty$ *and* $k_q f_{vb}^{(q)} = 0$ *by definition if* $k_q = \infty$ *and* $f_{vb}^{(q)} = 0$ *for* $v = 0,\ 1$. *The case where* $k_q = 0$ *and* $f_{vb}^{(q)} = \infty$ *is not covered by the Lemma.*

COMMENTS: 1. For all $k(\cdot) \in K_1$ we have $k_0 = 0$ and for all $P \in P_1$ we have $f_{1b}^{(0)} < \infty$. Hence, $\tilde{J}_T$ is an asymptotically unbiased estimator for J_{TP} for all $P \in P_1$ provided $S_T \to \infty$ and $S_T/T \to 0$ as $T \to \infty$. In addition, the variance of $\tilde{J}_T$ goes to zero as $T \to \infty$ under these conditions by Lemma 1(b). Hence, the MSE of $\tilde{J}_T$ goes to zero as $T \to \infty$ and $\tilde{J}_T$ is L^2- and weakly consistent for J_{TP} for all $P \in P_1$ and $k(\cdot) \in K_1$.

2. The maximal rate of convergence to zero of the bias of $\tilde{J}_T$ is obtained by taking q equal to the largest value such that $k_q < \infty$ and $f_{1b}^{(q)} < \infty$. For the

Bartlett, Parzen, Tukey-Hanning, and QS kernels, these values are 1, 2, 2, and 2, respectively. The corresponding maximal rates of convergence are $O(S_T^{-1})$, $O(S_T^{-2})$, $O(S_T^{-2})$, and $O(S_T^{-2})$, respectively. Note that the rate for the Bartlett kernel is slower than that for the other three kernels. For the truncated kernel, the largest value of q (if it exists) is determined by $f_{1b}^{(q)}$, since $k_q < \infty$ $\forall q < \infty$. If $f_{1b}^{(q)} < \infty$ for some $q > 2$, then the truncated kernel has a faster rate of convergence for each $P \in \mathcal{P}_1$ then do the other four kernels. For large q values, however, there is reason to believe that the rate of convergence for the truncated estimator is obtained only with relatively large sample sizes, see Section 9.

3. The only kernels for which $k_q < \infty$ for $q > 2$ are kernels that do not necessarily generate psd estimates. Thus, the maximal rate of convergence to zero of the bias for kernels in $\mathcal{K}_2$ is $O(S_T^{-2})$ or slower. To see this, note that $k_2 = \frac{1}{2}\int_{-\infty}^{\infty}\lambda^2 K(\lambda)d\lambda$. Thus, $k_q < \infty$ for $q > 2$ implies $k_2 = 0$ and $K(\lambda)$ must be negative for some $\lambda \in R$. The discussion following equation (2.9) now establishes the assertion.

Let $MSE_P(\cdot)$ denote the MSE of $\cdot$ when the underlying distribution is P. Lemmas 1 and 2 are used to obtain the following theorem regarding the asymptotic MSE of kernel estimators.

THEOREM 1: *For all* $b \in R^p$ *and* $q \in (0,\infty)$, *if* $S_T \to \infty$ *and* $S_T^{2q+1}/T \to \gamma \in [0,\infty]$, *then*

(a) $\lim_{T\to\infty} \frac{T}{S_T} \inf_{P\in\mathcal{P}_0} MSE_P(b'\tilde{J}_T b) = 4\pi^2(k_q^2(f_{0b}^{(q)})^2/\gamma + 2f_{0b}^2)$ *for any* $k(\cdot) \in \mathcal{K}_0$

and

(b) $\lim_{T\to\infty} \frac{T}{S_T} \sup_{P\in\mathcal{P}_1} MSE_P(b'\tilde{J}_T b) = 4\pi^2(k_q^2(f_{1b}^{(q)})^2/\gamma + 2f_{1b}^2)$ *for any* $k(\cdot) \in \mathcal{K}_1$,

where $k_q f_{vb}^{(q)}$ *is as defined in Lemma 2 in the case where* $k_q = \infty$ *or* $f_{vb}^{(q)} = \infty$ *for* $v = 0$ *or* 1. *The case where* $\gamma = \infty$ *also requires* $S_T/T \to 0$ *and* $S_T^q/T \to 0$ *as* $T \to \infty$

and $k_q^2(f_{1b}^{(q)})^2 < \infty$. *The case where* $\gamma = 0$ *also requires* $S_T \to \infty$ *as* $T \to \infty$ *and* $k_q^2(f_{0b}^{(q)})^2 > 0$. *The case where* $k_q = 0$ *and* $f_{vb}^{(q)} = \infty$ *for* $v = 0$ *or* 1 *is not covered by the Theorem.*

COMMENTS: 1. Theorem 1 yields optimal growth rates for the bandwidth parameters $\{S_T\}$ for kernels $k(\cdot)$ that have $k_q > 0$ for some q. These optimal rates are characterized by the value of q and by whether $\gamma = 0$, $\gamma \in (0,\infty)$, or $\gamma = \infty$. It is straightforward to show that for $\gamma \in (0,\infty)$ the rate of convergence of MSE is maximized by taking q as large as possible subject to $k_q < \infty$ and $f_{1b}^{(q)} < \infty$. For the Bartlett, Parzen, Tukey-Hanning, and QS kernels then, the optimal q values are 1, 2, 2, and 2, respectively, provided $f_{1b}^{(q)} < \infty$. For the truncated kernel no optimal value of q exists unless there is a number $q^* < \infty$ such that $f_{1b}^{(q)} < \infty$ for $q \le q^*$ and $f_{0b}^{(q)} = \infty$ for $q > q^*$.

Next, given q such that $k_q \in (0,\infty)$, the limit infimum of the MSE is infinity if $\gamma = 0$ by Theorem 1(a). In addition, it is straightforward to show that the MSE converges to zero at a faster rate if $\gamma \in (0,\infty)$ than if $\gamma = \infty$.

In sum, the optimal growth rate of $\{S_T\}$ requires $\gamma \in (0,\infty)$ and q such that $k_q \in (0,\infty)$. Thus, optimal S_T values are given by
$S_T = (\gamma T)^{1/(2q+1)} + o(T^{1/(2q+1)})$ as $T \to \infty$ for some $\gamma \in (0,\infty)$ and q as above. For the Bartlett, Parzen, Tukey-Hanning, and QS kernels this yields optimal growth rates of S_T equal to $T^{1/3}$, $T^{1/5}$, $T^{1/5}$, and $T^{1/5}$, respectively. The determination of a specific value for γ is considered in Theorem 3 below.

2. For the estimator based on the Bartlett kernel, the consistency results of Newey and West (1987) and Gallant and White (1988) require $S_T = o(T^{1/4})$ as $T \to \infty$. Thus, these results do not establish consistency for this estimator when the bandwidth/lag truncation parameter is allowed to grow at its asymptotically optimal rate $T^{1/3}$. The consistency results given here do cover this case, however, because they only require $S_T = o(T)$ as $T \to \infty$.

4. A MINIMAX OPTIMAL KERNEL

This section obtains a minimax optimal kernel from the class K_2 of kernels that generate psd estimators. The minimax criterion selects a kernel that minimizes, over kernels $k(\cdot) \in K_2$, the maximum asymptotic MSE over distributions P $\in P_1$. The optimal kernel is found to be the QS kernel. This kernel is optimal regardless of the choice of distribution P_1 that is used to define P_1 (provided P_1 satisfies the basic conditions outlined above (2.10)).

Let $Q\tilde{S}_T$ denote the pseudo-estimator $\tilde{J}_T$ when the latter is defined using the QS kernel. Let k_{2QS} denote the value of k_q for the QS kernel when $q = 2$.

THEOREM 2: *For any sequence of bandwidth parameters* $\{S_T\}$ *such that* $S_T \to \infty$, $S_T^2/T \to 0$, *and* $S_T^5/T \to \gamma \in [0,\infty]$ *as* $T \to \infty$ *and any "estimator"* $\tilde{J}_T$ *based on a kernel* $k(\cdot) \in K_2$, *the QS "estimator"* $Q\tilde{S}_T$ *satisfies:*

(a) $\forall b \in R^p$, $\underline{\lim}_{T\to\infty} r_T[\sup_{P \in P_1} MSE_P(b'\tilde{J}_T b) - \sup_{P \in P_1} MSE_P(b'Q\tilde{S}_T b)]$

$$= \begin{cases} 4\pi^2(f_{1b}^{(2)})^2(k_2^2 - k_{2QS}^2)/\gamma & \text{if } \gamma \in (0,\infty] \\ 4\pi^2(f_{1b}^{(2)})^2(k_2^2 - k_{2QS}^2) & \text{if } \gamma = 0 \end{cases}$$

$$\geq 0$$

provided $f_{1b}^{(2)} < \infty$, *where* $r_T = \begin{cases} T/S_T & \text{if } \gamma \in (0,\infty] \\ S_T^4 & \text{if } \gamma = 0 \end{cases}$. *If* $f_{1b}^{(2)} = 0$, *the limit infimum is zero even if* $k_2 = \infty$.

(b) *The inequality in part (a) holds strictly if* $k(x) \neq k_{QS}(x)$ *with positive Lebesgue measure,* $\gamma \neq \infty$, *and* $f_{1b}^{(2)} > 0$.

COMMENTS: 1. The normalization of kernels used in K_1 and K_2 is important for Theorem 2. This normalization is such that for any given sequence $\{S_T\}$ each kernel $k(\cdot) \in K_1$ has the same maximum asymptotic variance over P $\in P_1$. In consequence, it is appropriate to use the same sequence of bandwidth parameters

(S_T) when comparing different kernels. The Theorem has the highly desirable feature that the optimal kernel does not depend on the particular choice of bandwidth parameters.

2. As can be seen by the proof of Theorem 2, the asymptotic optimality of the QS kernel does not depend on the specific form of the MSE criterion. An analogous optimality property of the QS kernel holds for any criterion that is a non-decreasing function of the variance and the absolute value of the bias.

3. Optimality results for the QS kernel first were given by Priestley (1962, pp. 561-62) for kernel estimation of spectral density functions. Priestley considered a smaller class of kernels than those considered here. For the case of probability density function estimation, Epanechnikov (1969, pp. 155-56) and Sacks and Ylvisacker (1981) showed that the QS kernel is optimal in a more general class of kernels than that considered by Priestley. Their class of kernels is analogous to the class considered here.

5. OPTIMAL FIXED BANDWIDTH PARAMETERS

5.1. *First Order Asymptotically Optimal Bandwidth Parameters*

We now consider the first order asymptotically optimal choice of bandwidth parameter S_T for a given kernel $k(\cdot)$ for which $k_q \in (0,\infty)$ for some q. From Comment 1 to Theorem 1, it is evident that this problem reduces to the optimal choice of the constant γ in the formula $S_T = (\gamma T)^{1/(2q+1)} + o(T^{1/(2q+1)})$. Again a minimax optimality criterion is used. Unlike the result of Theorem 2, in which an optimal kernel was found that was the same for any "dominating" distribution P_1, the optimal constant γ depends on a scalar parameter α that is a function of P_1.

Let $\{w_\ell : \ell = 1, \ldots, r\}$ be a set of non-negative weights that sum to one. We consider a *weighted squared error* loss function

$$(5.1)\qquad L(\tilde{J}_T,\ J_{TP}) = \sum_{\ell=1}^{r} w_\ell(\tilde{J}_{T,\ell\ell} - J_{TP,\ell\ell})^2 ,$$

where the subscript $\ell\ell$ denotes the ℓ-th diagonal element of the corresponding matrix. The usual choice of w_ℓ is $1/r$ for $\ell = 1, \ldots, r$ or $1/(r-1)$ for all ℓ except that which corresponds to an intercept parameter.

Given a kernel $k(\cdot)$, suppose q is such that $k_q \in (0,\infty)$. Let b_ℓ be the r-vector with ℓ-th element one and all other elements zero. For a given dominating distribution P_1, define

$$(5.2)\qquad \alpha = \sum_{\ell=1}^{r} w_\ell (f_{1b_\ell}^{(q)})^2 \Big/ \sum_{\ell=1}^{r} w_\ell f_{1b_\ell}^2 .$$

The leading case is where P_1 is chosen such that $|f_{1b}^{(q)}/f_{1b}|$ is the same for all b. Then, $\alpha = (f_{1b}^{(q)}/f_{1b})^2$ and α does not depend on $\{w_\ell\}$ or on the variance of $b_\ell'V_t$ under P_1. In Sections 7 and 8 below, however, data-dependent bandwidth parameters are considered and in this case $|f_{1b_\ell}^{(q)}/f_{1b_\ell}|$ generally differs for different values of ℓ.

For example, suppose $b_\ell'V_t$ is an AR(1) process under P_1 with AR parameter ρ_ℓ and innovation variance σ_ℓ^2 for $\ell = 1, \ldots, r$. Then, for kernels with $q = 2$,

$$(5.3)\qquad \alpha = \sum_{\ell=1}^{r} w_\ell \frac{4\rho_\ell^2\sigma_\ell^4}{(1-\rho_\ell)^8} \Big/ \sum_{\ell=1}^{r} w_\ell \frac{\sigma_\ell^4}{(1-\rho_\ell)^4}$$

and for kernels with $q = 1$

$$(5.4)\qquad \alpha = \sum_{\ell=1}^{r} w_\ell \frac{4\rho_\ell^2\sigma_\ell^4}{(1-\rho_\ell)^6(1+\rho_\ell)^2} \Big/ \sum_{\ell=1}^{r} w_\ell \frac{\sigma_\ell^4}{(1-\rho_\ell)^4} .$$

If $\rho_\ell = \rho$ for all ℓ, then α in (5.3) and (5.4) simplifies to $\alpha = 4\rho^2/(1-\rho)^4$ and $\alpha = 4\rho^2/(1-\rho^2)^2$ respectively.

If $b_\ell'V_t$ is an ARMA(1,1) process under P_1 with ARMA parameters $(\rho_\ell,\ \psi_\ell)$ and innovation variance σ_ℓ^2 for $\ell = 1, \ldots, r,$[7] then for kernels with $q = 2$

$$(5.5) \qquad \alpha = \sum_{\ell=1}^{r} w_\ell \frac{4(1+\rho_\ell\psi_\ell)^2(\rho_\ell+\psi_\ell)^2\sigma_\ell^4}{(1-\rho_\ell)^8} \Big/ \sum_{\ell=1}^{r} w_\ell \frac{(1+\psi_\ell)^4\sigma_\ell^4}{(1-\rho_\ell)^4} ,$$

and for kernels with $q = 1$

$$(5.6) \qquad \alpha = \sum_{\ell=1}^{r} w_\ell \frac{4(1+\rho_\ell\psi_\ell)^2(\rho_\ell+\psi_\ell)^2\sigma_\ell^4}{(1-\rho_\ell)^6(1+\rho_\ell)^2} \Big/ \sum_{\ell=1}^{r} w_\ell \frac{(1+\psi_\ell)^4\sigma_\ell^4}{(1-\rho_\ell)^4} .$$

Alternatively, if $b_\ell' V_t$ is an MA(m) process under P_1 with MA parameters $\{\psi_{\ell u}, u = 1, \ldots, m\}$ and innovation variance σ_ℓ^2 for $\ell = 1, \ldots, r,$[8] then

$$(5.7) \quad \alpha = \sum_{\ell=1}^{r} w_\ell \Big(2 \sum_{j=1}^{m} j^q (\psi_{\ell j} + \sum_{u=1}^{m-j} \psi_{\ell u}\psi_{\ell u+j})\Big)^2 \sigma_\ell^4 \Big/ \sum_{\ell=1}^{r} w_\ell \Big(\sum_{j=-m}^{m} (\psi_{\ell |j|} + \sum_{u=1}^{m-|j|} \psi_{\ell u}\psi_{\ell u+|j|})\Big)^2 \sigma_\ell^4 .$$

In many macroeconomic, financial, and international financial applications, MA(m) processes (for known m) arise quite naturally, e.g., see Hansen and Hodrick (1980), Campbell and Clarida (1987), Mishkin (1987), and Hardouvelis (1988). They arise when V_t is partly comprised of an equation error that is a forecast error for some variable m time periods into the future. In such cases, one expects the errors, and hence $\{V_t\}$, to be m-dependent.

For any given $\alpha \in (0,\infty)$, let $\mathcal{P}_1(\alpha)$ denote some set $\mathcal{P}_1$ whose dominating distribution P_1 satisfies (5.2). For specificity, let $\tilde{J}_T(S_T)$ denote $\tilde{J}_T$ when the bandwidth parameter S_T is used to construct $\tilde{J}_T$.

The next result establishes first order asymptotically optimal bandwidth parameter values using a minimax criterion:

THEOREM 3: *For any given* $k(\cdot) \in \mathcal{K}_1$ *such that* $0 < k_q < \infty$ *for some* $q \in (0,\infty)$, *any given sequence* $\{S_T\}$ *of bandwidth parameters such that* $S_T \to \infty$ *and* $S_T^{2q+1}/T \to \gamma \in [0,\infty]$ *as* $T \to \infty$, *and any given* $\alpha \in (0,\infty)$, *the sequence* $\{S_T^*\}$ *defined by* $S_T^* = (qk_q^2\alpha)^{1/(2q+1)}T^{1/(2q+1)}$ *satisfies:*

$$\underline{\lim}_{T\to\infty} T^{2q/(2q+1)}\left(\sup_{P\in P_1(\alpha)} E_P L(\tilde{J}_T(S_T),\ J_{TP}) - \sup_{P\in P_1(\alpha)} E_P L(\tilde{J}_T(S_T^*),\ J_{TP})\right) \geq 0$$

provided $f_{1b_\ell} > 0$ *and* $f_{1b_\ell}^{(q)} > 0$ *for some* ℓ *for which* $w_\ell > 0$, *with equality if and only if* $S_T = S_T^* + o(T^{1/(2q+1)})$ *as* $T \to \infty$.

COMMENTS: 1. Suppose $\bar{k}(\cdot)$ is a kernel that satisfies all of the conditions of K_1 except the normalization condition $\int\bar{k}^2(x)dx = 1$. Then, the first order asymptotically optimal bandwidth parameter for this kernel is denoted M_T^* and is defined by

$$(5.8)\qquad M_T^* = (q\bar{k}_q^2\alpha/\int\bar{k}^2(x)dx)^{1/(2q+1)}T^{1/(2q+1)} .$$

For the normalizations of (2.7), we have

$$(5.9)\qquad \begin{aligned} &\text{Bartlett kernel:}\ \ M_T^* = 1.1447(\alpha T)^{1/3} ,\\ &\text{Parzen kernel:}\ \ M_T^* = 2.6614(\alpha T)^{1/5} ,\\ &\text{Tukey-Hanning kernel:}\ \ M_T^* = 1.7462(\alpha T)^{1/5} ,\\ &\text{Quadratic Spectral kernel:}\ \ S_T^* = 1.3221(\alpha T)^{1/5} . \end{aligned}$$

For illustrative purposes, Table 1 tabulates M_T^* for the Bartlett, Parzen, and Tukey-Hanning kernels and S_T^* for the QS kernel for a linear regression model in which the regressors and errors are mutually independent, homoskedastic, first order autoregressive (AR(1)) random variables each with autoregressive parameter ν. For this model each element of V_t (except that corresponding to the intercept) has correlation structure identical to that of an AR(1) process with parameter $\rho = \nu^2$. The weights are taken to be equal for the non-constant regressors and zero for the intercept.

2. The results of the Theorem can be used to assess Gallant's (1987, pp. 551, 573) suggestion to set the lag truncation parameter M_T at $[T^{1/5}] + 1$ with

the Parzen kernel. Suppose $b_\ell' V_t$ has AR(1) correlation structure with AR parameter $\rho = \nu^2$. As in comment 1, this occurs if the errors and regressors are mutually independent AR(1) random variables each with AR parameter ν. Then, for $T = 128$ Gallant's value of M_T is first order optimal when $\rho = .053$ and $\nu = .23$. When $T = 64$ it is optimal for $\rho = .073$ and $\nu = .27$. Hence, the formula $M_T = [T^{1/5}] + 1$ is appropriate only for a sequence $\{b_\ell' V_t\}$ that has relatively little autocorrelation.

3. When the bandwidth parameters are set equal to $\{S_T^*\}$, the limit of the maximum MSE over $P \in \mathcal{P}_1(\alpha)$ is such that the bias squared equals $1/(2q+1)$ of the total MSE. Thus, the bias of the Newey-West/Bartlett estimator accounts for a greater fraction of its MSE than do the biases of the QS, Gallant/Parzen, and Tukey-Hanning estimators.

4. When the first order optimal bandwidth parameters are used, the Gallant/Parzen and Tukey-Hanning estimators are 8.6% less and .9% more efficient asymptotically than the QS estimator, respectively, for any dominating distribution P_1. (Since the Tukey-Hanning kernel does not necessarily generate psd estimates, i.e., $k_{TH}(x) \notin \mathcal{K}_2$, the latter result does not violate Theorem 2.) Also, the Newey-West/Bartlett estimator is 100% less efficient asymptotically than the Gallant/Parzen, Tukey-Hanning, and QS estimators, since its MSE converges to zero at a slower rate than do the MSEs of the latter three. In particular finite sample situations, however, the Newey-West/Bartlett estimator may not perform nearly so poorly in relative terms, depending on the relative magnitudes of T, $f_{1b_\ell}^{(2)}$, $f_{1b_\ell}^{(1)}$, and f_{1b_ℓ}.

5. In order to utilize the formula for S_T^* one needs to specify a value of α. This is most easily accomplished by taking P_1 such that $(f_{1b_\ell}^{(q)}/f_{1b_\ell})^2$ is the same for all ℓ and such that $b_\ell' V_t$ is a well-known process, such as an AR(1), MA(1), or ARMA(1,1) process with parameter ρ, ψ, or (ρ,ψ), respectively. By

specifying a least favorable value of ρ, ψ, or (ρ,ψ), one obtains α and in turn S_T^*. Note that the resulting estimator is consistent not only for an AR(1), MA(1), or ARMA(1,1) process, but for a wide variety of processes. The AR(1), MA(1), or ARMA(1,1) process simply represents the least favorable process to which the bandwidth parameter is adjusted. For the common case in which it is difficult to specify a least favorable process, an automatic bandwidth estimator can be used. Such estimators are discussed in Sections 7 and 8 below. They make use of the formula for S_T^* given in the Theorem.

5.2. Higher Order Asymptotically Optimal Bandwidth Parameters

Next, we introduce corrections to the first order asymptotically optimal bandwidth parameter S_T^* that have higher order asymptotic justifications in general and in the Gaussian case are optimal for fixed sample sizes.

Let

(5.10)
$$B_{T\ell}(S) = -\sum_{j=-T+1}^{T-1}(1-k(j/S))(1-|j|/T)b_\ell'\Gamma_1(j)b_\ell \text{ and}$$

$$V_{T\ell}(S) = \sum_{j=-T+1}^{T-1}\sum_{m=-T+1}^{T-1}k(j/S)k(m/S)\left[\frac{1}{T^2}\sum_{s=|j|+1}^{T}\sum_{t=|m|+1}^{T}\cdot\left(b_\ell'\Gamma_1(t-s)b_\ell b_\ell'\Gamma_1(t-s+j-m)b_\ell + b_\ell'\Gamma_1(t-s-m)b_\ell b_\ell'\Gamma_1(t-s+j)b_\ell\right)\right].$$

$B_{T\ell}(S)$ gives the exact bias under P_1 for estimation of $b_\ell' J_{TP_1} b_\ell$ using the estimator $\bar{J}_T$ that has kernel $k(\cdot)$ and bandwidth parameter S. $V_{T\ell}(S)$ gives an approximation to this estimator's variance. This approximation differs from the exact variance only due to the omission of a fourth order cumulant term. In consequence, if $\{b_\ell' V_t\}$ is Gaussian under P_1, then $V_{T\ell}(S)$ gives the exact variance of the estimator under P_1.

Let S_T^{**} be any value that minimizes

$$(5.11)\qquad \sum_{\ell=1}^{r} w_\ell (B_{T\ell}^2(S) + V_{T\ell}(S))$$

over $S \in (0,\infty)$. S_T^{**} is a higher order asymptotically optimal bandwidth parameter. Such values usually exist. For example, they exist if $k(\cdot)$ is continuous. (In this case $B_{T\ell}^2(S) + V_{T\ell}(S)$ is continuous in S for all ℓ and its value diverges to infinity as S goes to zero or infinity.) Thus, optimal values S_T^{**} exist for Bartlett, Parzen, Tukey-Hanning, and QS kernels.

For example, if $\{b_\ell' V_t\}$ is an ARMA(1,1) process under P_1, then $B_{T\ell}(S)$ and $V_{T\ell}(S)$ are given by the expressions of (5.10) with

$$(5.12)\qquad \begin{gathered} b_\ell'\Gamma_1(0)b_\ell = \frac{1+\psi_\ell^2+2\rho_\ell\psi_\ell}{1-\rho_\ell^2}\sigma_\ell^2\ ,\ b_\ell'\Gamma_1(\pm1)b_\ell = \frac{(1+\rho_\ell\psi_\ell)(\rho_\ell+\psi_\ell)}{1-\rho_\ell^2}\sigma_\ell^2, \\ \text{and } b_\ell'\Gamma_1(j)b_\ell = \rho_\ell^{|j|-1} b_\ell'\Gamma_1(1)b_\ell \text{ for } j = \pm2,\ \pm3,\ \ldots\ . \end{gathered}$$

The same is true in the MA(m) case except that

$$(5.13)\qquad b_\ell'\Gamma_1(j)b_\ell = \begin{cases} \left(\psi_{\ell|j|} + \sum_{u=1}^{m-|j|} \psi_{\ell u}\psi_{\ell u+|j|}\right)\sigma_\ell^2 & \text{for } j = 0,\ \pm1,\ \ldots,\ \pm m \\ 0 & \text{otherwise.} \end{cases}$$

The bandwidth parameters $\{S_T^{**}\}$ satisfy the following optimality properties:

THEOREM 4: (a) *Let* $\{S_T\}$ *be any sequence of non-negative numbers. Then, for* $\{S_T^{**}\}$ *as defined above, we have*

$$\varliminf_{T\to\infty} \zeta_T\left[\sup_{P\in\mathcal{P}_1} E_P L(\tilde{J}_T(S_T),\ J_{TP}) - \sup_{P\in\mathcal{P}_1} E_P L(\tilde{J}_T(S_T^{**}),\ J_{TP})\right] \geq 0$$

for all sequences of constants $\{\zeta_T\}$ *such that* $\zeta_T = o(T^2)$ *as* $T\to\infty$.

(b) *If* $\{b_\ell' V_t\}$ *is Gaussian under* P_1 *for all* $\ell = 1,\ \ldots,\ r$, *then for all* S_T *in*

$[0,\infty)$ *and all* T *finite,*

$$E_{P_1} L(\tilde{J}_T(S_T),\ J_{TP_1}) \geq E_{P_1} L(\tilde{J}_T(S_T^{**}),\ J_{TP_1}).$$

COMMENT: 1. A standard optimization algorithm can be used to minimize $\sum_{\ell=1}^{r} w_\ell (B_{T\ell}^2(S) + V_{T\ell}(S))$ with respect to the positive scalar S. Analytic derivatives of the optimand can be employed, since they usually are easy to calculate. The first order asymptotically optimal bandwidth parameter S_T^* provides an excellent starting value for the optimization procedure.

6. TREATMENT OF ESTIMATED PARAMETERS

We now show that the results of Sections 2 through 5 apply not only to the pseudo-estimator $\tilde{J}_T$ but also to the actual estimator $\hat{J}_T$. We introduce two alternative assumptions regarding the estimator $\hat{\theta}$, the process $\{V_t(\theta)\}$, and the kernel $k(\cdot)$. The first assumption is easier to verify than the second and is sufficient to establish consistency of $\hat{J}_T$, but is not sufficient to show that $\tilde{J}_T$ and $\hat{J}_T$ have equivalent MSEs to the appropriate order as $T \to \infty$. The second assumption is sufficient for the latter purpose, as well as for establishing consistency. It is not, however, easy to verify. We utilize this assumption because more primitive assumptions would require a much lengthier presentation and would add little to the main points of interest in the paper.

ASSUMPTION A: *For all* $m = 1, \ldots, r$,

$$\overline{\lim_{T\to\infty}} \sup_{P\in\mathcal{P}_1} E_P\left(\frac{1}{T}\sum_{t=1}^{T} \sup_{\theta\in\Theta}\|V_t(\theta)\|^2\right)\left(\frac{1}{T}\sum_{t=1}^{T} \sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_m}V_t(\theta)\right\|^2\right)(\sqrt{T}|\hat{\theta}_m - \theta_{0m}|)^2 < \infty$$

and $\int_{-\infty}^{\infty}|k(x)|dx < \infty$.

This assumption can be verified by separately considering the three terms in the expectation using the Cauchy-Schwartz inequality. The condition on the kernel is not restrictive.

Define $Z_{1T} = \frac{1}{S_T}\sum_{j=-T+1}^{T-1}|k(j/S_T)|\cdot\left|\frac{1}{\sqrt{T}}\sum_{t=|j|+1}^{T}\frac{\partial}{\partial\theta}(b'V_t(\theta_0)b'V_{t-|j|}(\theta_0))\right|$ and

$Z_{2T} = \frac{1}{S_T}\sum_{j=-T+1}^{T-1}|k(j/S_T)|\cdot\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=|j|+1}^{T}\frac{\partial^2}{\partial\theta\partial\theta'}(b'V_t(\theta)b'V_{t-|j|}(\theta)\right|$. Let Z_{1Tm}, $\hat{\theta}_m$, and θ_{0m} denote the m-th elements of Z_{1T}, $\hat{\theta}$, and θ_0, respectively, for $m = 1, \ldots, r$. Let $|C|$ denote the vector or matrix of absolute values of the elements of C.

ASSUMPTION B: *For all* $m = 1, \ldots, r$,

$$\overline{\lim_{T\to\infty}}\sup_{P\in\mathcal{P}_1} E_P(Z_{1Tm}\sqrt{T}(\hat{\theta}_m - \theta_{0m}))^2 < \infty \text{ and}$$

$$\overline{\lim_{T\to\infty}}\sup_{P\in\mathcal{P}_1} E_P(\sqrt{T}|\hat{\theta} - \theta_0|'Z_{2T}\sqrt{T}(\hat{\theta} - \theta_0)|)^2 < \infty .$$

THEOREM 5: *Suppose* $k(\cdot) \in \mathcal{K}_1$, $S_T \to \infty$, *and* $S_T^{2q+1}/T \to \gamma \in (0,\infty)$ *as* $T \to \infty$ *for some* $q \in (0,\infty)$ *for which* $k_q < \infty$. *Then,*

$$\lim_{T\to\infty}\sup_{P\in\mathcal{P}_1}\left(\frac{T}{S_T}\right)^{\xi}|MSE_P(b'\hat{J}_Tb) - MSE(b'\tilde{J}_Tb)| = 0$$

for all $b \in R^p$ *such that* $f_{1b}^{(q)} < \infty$, *provided (a) assumption A holds,* $q > 1/2$, *and* $\xi = 0$ *or (b) assumption B holds and* $\xi = 1$.

COMMENTS: 1. Either assumption A or assumption B is sufficient for consistency of $\hat{J}_T$ but assumption A requires $S_T = o(T^{1/2})$ (since $q > 1/2$) whereas assumption B allows $S_T = o(T)$. The former is not restrictive, however, since optimal growth rates of S_T satisfy $S_T = o(T^{1/2})$ for efficient kernels.

2. Part (b) of the Theorem shows that $\tilde{J}_T$ and $\hat{J}_T$ have equivalent MSEs asymptotically, since T/S_T times the MSE of $\tilde{J}_T$ is bounded away from zero and infinity as $T \to \infty$ by Theorem 1. Thus, the optimal kernel and optimal bandwidth parameter results of Theorems 2 and 3 apply to $\hat{J}_T$ as well as to $\tilde{J}_T$.

7. AUTOMATIC BANDWIDTH ESTIMATORS

This section introduces automatic bandwidth HAC estimators of J_{TP}. These estimators are the same as the kernel estimators of Sections 2-6 except that the bandwidth parameter is a function of the data rather than a function of a least favorable distribution P_1. Since it often is difficult to specify a least favorable distribution, the automatic estimators introduced here are likely to be more useful in practice than the non-automatic estimators considered in Sections 3-6. Nevertheless, the results of these sections are quite useful because the formulae given there for the first order and higher order asymptotically optimal bandwidth parameters are used here to construct the automatic bandwidth parameters. In addition, the consistency and MSE results of this section depend on the results of the earlier sections.

In the density estimation literature, several automatic bandwidth methods have been developed. The two main types of methods are cross-validation methods, e.g., see Stone (1984), Hall and Marron (1987a, b), Beltrao and Bloomfield (1987), Robinson (1988), and references therein, and "plug-in" methods, see Deheuvels (1977), Sheather (1986), and Hall and Marron (1987c).

Cross-validation methods have been shown to possess certain optimality properties with respect to minimizing integrated or mean integrated squared error in probability density and spectral density estimation, e.g., see Stone (1984), Beltrao and Bloomfield (1987), and Robinson (1988). These procedures have two drawbacks, however, that limit their potential for use in the problem at hand. First, they are suitable only if one is interested in estimating a density over an interval, such as the real line, rather than estimating a density at a single point. Hence, they are not suitable for the problem at hand, since it is very closely related to that of estimating a density at a single point. Second, the rate of convergence to zero of the (percentage) deviation of the cross-validated bandwidth parameter from the (unknown) optimal bandwidth parameter is very slow--rate $T^{-1/10}$. Thus, the cross-validated bandwidth parameter may exhibit considerable variability in finite samples.

Plug-in methods of automatic bandwidth estimation are characterized by the use of an asymptotic formula for an optimal bandwidth parameter in which estimates are "plugged-in" in place of various unknowns in the formula. The estimates that are plugged-in may be parametric or nonparametric. The former have the advantage of yielding a less variable bandwidth parameter, because their variances decline to zero at a faster rate than do the variances of nonparametric estimates. The use of parametric estimates has the disadvantage, however, that parametric estimates introduce an asymptotic bias in the estimation of the optimal bandwidth parameter due to the approximate nature of the specified parametric model. (Note that this property has no affect on the consistency of the resultant density estimator, provided the automatic bandwidth parameter diverges to infinity in some appropriate sense as $T \to \infty$.)

The automatic bandwidth parameters considered here are of the plug-in type and use parametric estimates. They deviate from the finite sample optimal S_T values due to error introduced by estimation, the use of approximating parametric models, and the approximation inherent in the asymptotic formula employed. Good performance of a HAC estimator, however, only requires the automatic bandwidth parameter to be near the strictly optimal bandwidth value and not precisely equal to it. The reason is that the MSEs of kernel HAC estimators tend to be somewhat U-shaped functions of the bandwidth parameter S_T. This is illustrated in Figure 2, which shows the MSE of the QS estimator as a function of S_T for the AR(1)-HOMO model with $\rho = 0.0$, .3, .5, .7, .9, and .95. (As described in Section 9 below, the AR(1)-HOMO model is a linear regression model with regressors and errors that are homoskedastic, AR(1) rv's both with AR(1) coefficient ρ.) The automatic bandwidth parameters considered here are designed to produce bandwidth parameters that are on the flat part of the MSE function even if they are not at the point of minimum MSE.

The automatic bandwidth parameters are defined as follows: First, one specifies approximating parametric models for $\{b_\ell' V_t\}$ (where $b_\ell' V_t$ is the ℓ-th element of V_t) for $\ell = 1, \ldots, r$.[9] For general purposes, the suggested models are AR(1) models with different parameters for each $\ell = 1, \ldots, r$. These models are parsimonious. If some other parametric models seem appropriate for a particular problem, then they should be used in place of AR(1) models. For example, it may be necessary to use a parametric model that allows for seasonal patterns in the data. Alternatively, if $\{V_t\}$ is known to be m-dependent, as occurs in many macroeconomic, financial, and international financial applications, then it may be appropriate to use MA(m) models to approximate the distributions of $\{b_\ell' V_t\}$ for $\ell = 1, \ldots, r$. If m is large, however, then more parsimonious models, such as AR(1) models, usually will be more appropriate even when $\{V_t\}$ is known to be m-dependent.

Second, one estimates the parameters of the parametric models by standard methods. For the AR(1) case, let $\{(\hat\rho_\ell, \hat\sigma_\ell^2) : \ell = 1, \ldots, r\}$ denote estimates of the parameters $\{(\rho_\ell, \sigma_\ell^2) : \ell = 1, \ldots, r\}$. For the MA(m) case, let $\{(\hat\psi_{\ell 1}, \ldots, \hat\psi_{\ell m}, \hat\sigma_\ell^2) : \ell = 1, \ldots, r\}$ denote estimates of the parameters $\{(\psi_{\ell 1}, \ldots, \psi_{\ell m}, \sigma_\ell^2)\}$. (See footnote 8 for the parametrization of the MA(m) model.)

Third, one determines a random analogue $\hat\alpha$ of the parameter α of Section 5. For the AR(1) case, let $\hat\alpha$ be given by the value of α in equation (5.3), (5.4), or (5.2) with $\{(\rho_\ell, \sigma_\ell^2)\}$ replaced by $\{(\hat\rho_\ell, \hat\sigma_\ell^2)\}$. The standard choice for the weights $\{w_\ell : \ell = 1, \ldots, r\}$ is $w_\ell = 1/r$ for all ℓ or $w_\ell = 1/(r-1)$ for all ℓ except that which corresponds to an intercept term and zero for the latter. The appropriate equation--(5.3), (5.4), or (5.2)--depends on whether the kernel being used is such that $k_q \in (0,\infty)$ for $q = 2$, $q = 1$, or any positive q, respectively. Thus, equation (5.3) is used for the Parzen, Tukey-Hanning, and QS kernels and equation (5.4) is used for the Bartlett kernel. For the MA(m) case, let $\hat{\alpha}$ be given by the value of α in equation (5.7) with $\{(\psi_{\ell 1}, \ldots, \psi_{\ell m}, \sigma_\ell^2)\}$ replaced by $\{(\hat{\psi}_{\ell 1}, \ldots, \hat{\psi}_{\ell m}, \hat{\sigma}_\ell^2)$ and with $q = 2$ for the Parzen, Tukey-Hanning, and QS kernels and $q = 1$ for the Bartlett kernel. For parametric models other than AR(1), MA(m), or ARMA(1,1) models, equation (5.2) is used with the appropriate parametric formulae for $f_{1b_\ell}^{(q)}$ and f_{1b_ℓ} substituted in and evaluated at the parameter estimates.

The last step is to determine the automatic bandwidth parameter $\hat{S}_T$ by substituting $\hat{\alpha}$ for α in the formula for the first order asymptotically optimal bandwidth parameter S_T^* given in Theorem 3. Alternatively, for a kernel $\bar{k}(\cdot)$ that is not normalized to be in $\mathcal{K}_1$, one determines the automatic bandwidth parameter $\hat{M}_T$ by substituting $\hat{\alpha}$ for α in the expression (5.8) for M_T^*. That is,

$$\hat{S}_T = (qk_q^2)^{1/(2q+1)}(\hat{\alpha}T)^{1/(2q+1)} \quad \text{and}$$

(7.1)

$$\hat{M}_T = (q\bar{k}_q^2/\textstyle\int\bar{k}^2(x)dx)^{1/(2q+1)}(\hat{\alpha}T)^{1/(2q+1)} \ .$$

For the normalization of (2.7), we have

(7.2)

$$\text{Bartlett kernel:} \quad \hat{M}_T = 1.1447(\hat{\alpha}T)^{1/3} \ ,$$

$$\text{Parzen kernel:} \quad \hat{M}_T = 2.6614(\hat{\alpha}T)^{1/5} \ ,$$

$$\text{Tukey-Hanning kernel:} \quad \hat{M}_T = 1.7462(\hat{\alpha}T)^{1/5} \ ,$$

$$\text{Quadratic Spectral kernel:} \quad \hat{S}_T = 1.3221(\hat{\alpha}T)^{1/5} \ .$$

$\hat{S}_T$ and $\hat{M}_T$ will be referred to as *first order optimal automatic bandwidth parameters*, since they are based on the formulae for the first order optimal minimax bandwidth parameters.

A nice property of these automatic bandwidth procedures is that they generally yield scale invariant estimators in linear regression models, provided the weight w_ℓ corresponding to the intercept is set equal to zero. That is, if the dependent variable and the non-constant regressors in a linear regression model are multiplied by a constant c, then the least squares estimator of the regression parameters is unchanged and so is the HAC estimator of the covariance matrix of the least squares estimator. If the approximating models used are AR(1), scale invariance occurs provided the estimators $\hat{\rho}_\ell$ are scale invariant and the estimators $\hat{\sigma}_\ell$ are scale equivariant (i.e., $\hat{\sigma}_\ell(cY, cX) = c^2\hat{\sigma}(Y,X)$). The latter conditions are satisfied by standard estimation procedures.

The automatic bandwidth parameter $\hat{S}_T$ (or $\hat{M}_T$) can be improved by reducing the error attributable to the use of an asymptotic approximation. This can be done by replacing the first order asymptotically optimal formula used in (7.1) above by the higher order asymptotically optimal formula given in (5.11).

More specifically, one specifies and estimates approximating parametric models for $\{b_\ell' V_t\}$ as above. Then, one calculates a *higher order optimal automatic bandwidth parameter* $\dot{S}_T$ based on (5.11) using estimated analogues of $B^2_{T\ell}(S)$ and $V_{T\ell}(S)$. For the AR(1) case, let $b_\ell'\hat{\Gamma}_1(j)b_\ell$ be defined as $b_\ell'\Gamma_1(j)b_\ell$ is in equation (5.12) with $((\rho_\ell, \psi_\ell, \sigma^2_\ell))$ replaced by $((\hat{\rho}_\ell, 0, \hat{\sigma}^2_\ell))$. Then, let $\hat{B}_{T\ell}(S)$ and $\hat{V}_{T\ell}(S)$ be defined as $B_{T\ell}(S)$ and $V_{T\ell}(S)$ are in (5.10) with $b_\ell'\Gamma_1(j)b_\ell$ replaced by $b_\ell'\hat{\Gamma}_1(j)b_\ell$. For the MA(m) case, the definitions are the same except that (5.12) is replaced by (5.13) and $((\hat{\rho}_\ell, \hat{\psi}_\ell, \hat{\sigma}^2_\ell))$ is replaced by $((\hat{\psi}_{\ell 1}, \ldots, \hat{\psi}_{\ell m}, \hat{\sigma}^2_\ell))$. $\dot{S}_T$ is defined to be any value that minimizes

$$(7.3) \qquad \sum_{\ell=1}^{r} w_\ell(\hat{B}^2_{T\ell}(S) + \hat{V}_{T\ell}(S))$$

over $S \in [0,\infty)$. This minimization problem can be solved by standard methods using $\hat{S}_T$ as the starting value.

The simulation results of Section 9 show that the first order optimal bandwidth parameter $\hat{S}_T$ often is sufficiently close to the finite sample optimal value of S_T that the higher order optimal parameter $\dot{S}_T$ is not needed.

In practice, the value of a HAC estimator can be sensitive to the choice of the bandwidth parameter. Hence, it often is wise to calculate several bandwidth values centered about the automatic bandwidth value given by (7.1) or (7.3) in order to assess the degree of sensitivity of the estimator. These additional bandwidth values can be chosen by replacing the estimated parameters of the approximating parametric models used in (7.1) or (7.3) by the estimated parameters plus or minus one or two standard deviations of their values. For example, with AR(1) approximating models, one would replace $\hat{\rho}_\ell$ by $\hat{\rho}_\ell \pm 1/\sqrt{T}$ or $\hat{\rho}_\ell \pm 2/\sqrt{T}$.

To summarize this section, the automatic HAC estimators that are proposed here are the kernel estimators that use the QS kernel and either the first order or the higher order optimal automatic bandwidth parameters, $\hat{S}_T$ or $\dot{S}_T$.

8. PROPERTIES OF AUTOMATIC BANDWIDTH ESTIMATORS

We now present results concerning the asymptotic properties of kernel HAC estimators that are based on the automatic bandwidth parameters $\{\hat{S}_T\}$ or $\{\dot{S}_T\}$. In brief, the results show that these estimators have the same first order asymptotic MSE properties as "estimators" based on certain fixed bandwidth sequences $\{S^*_{TP}\}$ that depend on the unknown distribution P. In consequence, an asymptotic minimax optimality property can be stated for the automatic estimators.

The results of this section apply to kernels in the following class:

$$\mathcal{K}_3 = \Big\{k(\cdot) \in \mathcal{K}_1 : \text{(i) } |k(x)| \le C_1|x|^{-b} \text{ for some } b > 1 + 1/q \text{ and some } C_1 < \infty, \text{ where } q \in (0,\infty) \text{ is such that } k_q \in (0,\infty) \text{ if possible and otherwise is any value such that the preceding condition holds, and (ii) } |k(x) - k(y)| < C_2|x-y| \;\; \forall x, y \in R \text{ for some constant } C_2 < \infty\Big\} \tag{8.1}$$

This class contains the Bartlett, Parzen, Tukey-Hanning, and QS kernels, but not the truncated kernel, because the latter does not satisfy the Lipschitz condition (ii).

Next we introduce some notation. Let ξ be the finite dimensional parameter vector that indexes the approximating parametric models discussed above for $\{b_\ell' V_t\}$, $\ell = 1, \ldots, r$. Let Ξ be the parameter space of ξ. Let $\hat{\xi}$ be an estimator of ξ. Let ξ_P denote the probability limit of $\hat{\xi}$ under P. We only consider distributions P for which such a probability limit exists.

Let α_P be the value of α from equation (5.2) that is obtained when P_1 is given by the approximating parametric distribution with parameter ξ_P. Note that α_P is the probability limit of $\hat{\alpha}$ under P (provided the formula in (5.2) exhibits continuity in ξ), where $\hat{\alpha}$ is defined as above to be a primary ingredient of $\hat{S}_T$. For $\dot{S}_T$, or more generally, for any sequence of automatic bandwidth parameters S_T^0, we can define an analogue of $\hat{\alpha}$ (as it is defined for $\hat{S}_T$) via $\hat{\alpha} = (S_T^0)^{2q+1}/(qk_q^2 T)$. With this definition, the value $\hat{\alpha}$ that corresponds to $\dot{S}_T$ also converges in probability to α_P under P in general.

The class of distributions $\mathcal{P}_{11}$ that we consider is defined as follows: For some ε_* and ε^* such that $0 < \varepsilon_* \le \varepsilon^* < \infty$, let

$$
(8.2)\quad P_{11} = \Big\{ P \in P_1 : \text{(i) } \hat{\xi} \xrightarrow{p} \xi_P \text{ under P for some } \xi_P \in \Xi \text{ such that } \alpha_P \in [\varepsilon_*, \varepsilon^*],
\text{(ii) } |b'\Gamma_1(j)b| \le C_3|j|^{-m} \text{ for } j = 0, \pm1, \pm2, \ldots, \text{ for some } C_3 < \infty,
\text{for some } m > \max(2,\ 1 + 2q/(q+2)), \text{ for all } b \in R^r \text{ with } \|b\| = 1,
\text{where q is as in } K_3, \text{ (iii) } \sup_{j\ge1} \operatorname{Var}_{P_1}(b'\hat{\Gamma}(j)b) = O(1/T) \text{ as } T \to \infty,
\text{and (iv) } \overline{\lim_{T\to\infty}}\, E_P\left(\frac{1}{(S^*_{TP})^v} \sum_{j=1}^{[(S^*_{TP})^v]} \sqrt{T}|b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b|\right)^4 \le C_4 \text{ for some } C_4 < \infty
\text{and some v in the interval } \big(\max(1+1/(2b-2),\ q/(m-1)),\ 1 + q/2\big]\Big\}.
$$

Note that the assumption that $\varepsilon_* > 0$ eliminates any distribution P from P_{11} for which $\alpha_P = 0$. In particular, if $\{b_\ell' V_t\}$ is a white noise sequence under P, then for most approximating parametric models ξ_P is such that $\alpha_P = 0$. Such distributions are not covered by the results immediately below, but are discussed subsequently.

Let $S^*_{TP} = (qk_q^2\alpha_P)^{1/(2q+1)}T^{1/(2q+1)}$. The sequence $\{S^*_{TP}\}$ is the fixed sequence of bandwidth parameters that is closest to $\{\hat{S}_T\}$ and $\{\dot{S}_T\}$. In particular, $(\hat{S}_T - S^*_{TP})/S^*_{TP}$ and $(\dot{S}_T - S^*_{TP})/S^*_{TP}$ have probability limit zero under P. $\{S^*_{TP}\}$ is the sequence of first order asymptotically optimal bandwidth parameters for the case in which P_1 equals the approximating parametric model with parameter ξ_P.

Consider any sequence of automatic bandwidth parameters $\{S_T^0\}$ that satisfies

ASSUMPTION C: $\sup_{P\in P_{11}} E_P\left(\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{\hat{\alpha}^{1/(2q+1)}}\right)^4 = O(1)$ *as* $T \to \infty$, *where* q *is as defined in* K_3, *and* $\hat{\alpha} \le \alpha^* $ *for some constant* $\alpha^* < \infty$.

Most sequences $\{S_T^0\}$ for which $\hat{\alpha}$ is bounded and $\sqrt{T}$-consistent for α_P satisfy Assumption C. The primary examples of interest are $\{\hat{S}_T\}$ and $\{\dot{S}_T\}$.

The following result shows that for S_T^0 as above, $\hat{J}_T(S_T^0)$ has the same asymptotic MSE properties under P as the estimator $\hat{J}_T(S_{TP}^*)$, which is based on the sequence of fixed bandwidth parameters $\{S_{TP}^*\}$. Since the asymptotic MSE properties of estimators with fixed bandwidth parameters have been determined in Section 3, this result establishes the asymptotic MSE properties of $\hat{J}_T(S_T^0)$, including the consistency of $\hat{J}_T(S_T^0)$, and allows some asymptotic optimality properties of $\hat{J}_T(S_T^0)$ to be obtained.

THEOREM 6: *Consider any kernel* $k(\cdot) \in \mathcal{K}_3$. *Suppose* $\{S_T^0\}$ *satisfies Assumption C. Then,*

$$T^{2q/(2q+1)} \sup_{P \in \mathcal{P}_{11}} |MSE_P(b'\hat{J}_T(S_T^0)b) - MSE_P(b'\hat{J}_T(S_{TP}^*)b)| \to 0 \textit{ as } T \to \infty$$

for all $b \in R$.

COMMENTS: 1. Under the assumptions, Lemmas 1 and 2 and Theorems 5 and 6 combine to establish the consistency of $\hat{J}_T(\hat{S}_T)$ and $\hat{J}_T(\dot{S}_T)$ for $P \in \mathcal{P}_{11}$.

2. Theorems 1, 5, and 6 combine to establish upper bounds on the asymptotic MSE of $\hat{J}_T(\hat{S}_T)$ and $\hat{J}_T(\dot{S}_T)$ for any $P \in \mathcal{P}_{11}$ and lower bounds for any $P \in \mathcal{P}_0 \cap \mathcal{P}_{11}$.

Next we state some asymptotic minimax optimality properties for automatic bandwidth parameters $\{S_T^0\}$ that satisfy Assumption C. As above, $\{\hat{S}_T\}$ and $\{\dot{S}_T\}$ are the examples of primary interest. These results combine with Theorem 2 to yield asymptotically optimal kernel and automatic bandwidth parameters. These optimality properties are not definitive by any means, but they do suggest that the proposed procedures have some nice properties when the approximating parametric models are chosen appropriately.

For any $\alpha > 0$, let $\mathcal{P}_{11}(\alpha) = \mathcal{P}_{11}$ where $\mathcal{P}_{11}$ is defined with $\varepsilon_* = \varepsilon^* = \alpha$ and with P_1 given by an approximating parametric distribution with parameter ξ for some $\xi \in \Xi$ such that $\alpha_{P_1} = \alpha$ and f_{1b_ℓ}, $f^{(q)}_{1,b_\ell} > 0$ for some ℓ.

Let $\{S^1_T\}$ be any sequence of automatic bandwidth parameters such that for some sequence of fixed bandwidth parameters $\{S_T\}$, which satisfies $S_T \to \infty$ and $S_T^{2q+1}/T \to \gamma$ for some $\gamma \in [0,\infty]$, we have

$$(8.3) \qquad \lim_{T\to\infty} T^{2q/(2q+1)}\left|E_{P_1} L(\hat{J}_T(S^1_T),\ J_{TP_1}) - E_{P_1} L(\hat{J}_T(S_T),\ J_{TP_1})\right| = 0,$$

where P_1 is as defined for $P_{11}(\alpha)$.

THEOREM 7: *Let* $k(\cdot) \in K_3$ *and let* q *be as in* K_3. *Assume Assumptions A and B hold. Let* $\{S^1_T\}$ *be any sequence of automatic bandwidth parameters that satisfies (8.3) and let* $\{S^0_T\}$ *be any sequence that satisfies Assumption C. Then, for any* $\alpha > 0$, $\{S^0_T\}$ *is preferred to* $\{S^1_T\}$ *in the sense that*

$$\underline{\lim}_{T\to\infty} T^{2q/(2q+1)}\left[\sup_{P\in P_{11}(\alpha)} E_P L(\hat{J}_T(S^1_T),\ J_{TP}) - \sup_{P\in P_{11}(\alpha)} E_P L(\hat{J}_T(S^0_T),\ J_{TP})\right] \geq 0$$

with equality only if $S_T = S^*_{TP_1} + o(T^{1/(2q+1)})$ *as* $T \to \infty$.

To complete this section, we discuss the behavior of automatic HAC estimators when the true distribution P is such that $\alpha_P = 0$. The primary example is when $\{V_t\}$ is uncorrelated, i.e., $\Gamma_{TP}(j)$ equals zero for all $|j| \geq 1$ and all T. In this case, the optimal value of S_T is infinity, which yields $\hat{J}_T = \hat{\Gamma}(0)$. One could have $\alpha_P = 0$, however, even when $\{V_t\}$ is not uncorrelated. For example, suppose AR(1) parametric models are utilized. If the first order autocorrelations of $\{V_t\}$ are zero but higher order autocorrelations are non-zero, then α_P will equal zero even though the estimand J_{TP} includes terms $\Gamma_{TP}(j)$ that are non-zero for $j \neq 0$.

When $\alpha_P = 0$, $\hat{\alpha}$ generally is $O_p(T^{-1})$ for the automatic bandwidth parameters $\{\hat{S}_T\}$ and $\{\dot{S}_T\}$. (To see this, consider the formulae for α given in (5.2)-(5.7). Replace the parameters ρ_ℓ, ψ_ℓ, and/or $\psi_{\ell u}$ by estimators that are $O_p(T^{-1/2})$ to obtain $\hat{\alpha} = O_p(T^{-1})$.) In consequence, $\hat{S}_T$ and $\dot{S}_T$ are $O_p(1)$. Thus, if $\{V_t\}$ is uncorrelated, the automatic HAC estimator generally will be consistent. On the other hand, if $\alpha_p = 0$ and $\{V_t\}$ is correlated, then the automatic HAC estimator will be inconsistent in general, since $\hat{S}_T$ and $\dot{S}_T$ do not diverge to infinity. This case is quite special and may not be of great concern.

9. MONTE CARLO RESULTS

In this section, simulation methods are used to evaluate the asymptotic results obtained in Sections 3-8.[10] In particular we are interested in evaluating the results of Theorem 2 regarding the minimax optimal kernel, the results of Theorem 3 regarding first order asymptotically optimal bandwidth and lag truncation parameters, and the results of Theorems 6 and 7 regarding automatic bandwidth parameters.

The models we consider are linear regression models, each with an intercept and four regressors, see (2.1). 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 (2.1).) Four 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; and MA(1)-HOMO, in which the errors and regressors are homoskedastic MA(1) processes. (Details concerning the models are given below.) A range of six to eight parameter values are considered for each model. Each parameter value corresponds to a different degree of autocorrelation.

Estimators based on the five kernels of (2.7) are evaluated. They are: White/truncated (WH/TR), Newey-West/Bartlett (NW/BT), Gallant/Parzen (GAL/PR), Tukey-Hanning (TH), and quadratic spectral (QS). The performance of each kernel estimator is determined for a variety of different bandwidths. These bandwidths include the first order asymptotically optimal bandwidth given in (5.9), the automatic bandwidth given in (7.2) based on AR(1) approximating models with ρ_ℓ estimated by LS for each ℓ, and a grid of fixed bandwidths that are used to obtain the finite sample optimal bandwidth. For the former two bandwidths, the weights $\{w_\ell\}$ are taken to be zero for the intercept and one quarter for the others.

For comparative purposes, three estimators are considered in addition to the kernel estimators described above: the heteroskedasticity consistent estimator of Eicker (1967) and White (1980), denoted INID; the standard LS variance estimator for iid errors, denoted IID; and a parametric estimator that assumes that the errors are homoskedastic, AR(1) random variables, denoted PAR. More specifically,

$$\text{INID} = \left[\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{T-5}\sum_{t=1}^{T}\hat{U}_t^2 X_t X_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\right]_{22},$$

$$(9.1) \qquad \text{IID} = \left(\frac{1}{T-5}\sum_{t=1}^{T}\hat{U}_t^2\right)\left[\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right]_{22}, \quad \text{and}$$

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

where $\hat{\rho}_{LS}$ is the LS estimator of ρ 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 second diagonal element of the matrix $\cdot$.[11]

For each variance estimator and each scenario, the following performance criteria are estimated by Monte Carlo simulation: (1) the exact bias, variance, mean-squared error (MSE), and mean absolute error (MAE) 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. (The nominal 100(1−α)% CIs are based on an asymptotic normal approximation. For the INID, IID, and PAR estimators, this normal approximation is not valid asymptotically in all of the scenarios under consideration.) These two sets of criteria are considered because we are interested both in how well the kernel estimators estimate the LS coefficient estimator's variance and in how well they perform when used to form confidence intervals for the regression coefficient. The control variate method of Davidson and MacKinnon (1981) is used to estimate the true confidence levels in (2).

Sample sizes of 64, 128, and 256 are investigated. One thousand repetitions are used for each scenario.

Next we describe the four 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.[12] Since the expectation of $\frac{1}{T}\sum_{t=1}^{T} X_t X_t'$ is diagonal, the transformation used to obtain diagonality should be close to the identity map. With this transformation, the estimand and the estimators simplify and the computational burden is reduced considerably. The estimand becomes just the product of the second diagonal elements of the three 5×5 matrices multiplied together in (2.1). Two of these diagonal elements are known--only one has to be estimated, viz., the second diagonal element of the J_{TP} matrix. Without the transformation, one has to compute all twenty-five elements of the estimated J_{TP} matrix, rather than a single element, in order to compute the performance criteria described above.

The values considered for the AR(1) parameter ρ in the AR(1)-HOMO model are 0, .3, .5, .7, .9, .95, -.3, -.5. In each scenario, the same value of ρ is used for the errors and the regressors.

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

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 $U_t = |x_t'\xi| \times \tilde{U}_t$ for $t = 1, \ldots, T$. 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 $\xi = (1,0,0,0)'$ and $\xi = (1/2,1/2,1/2,1/2)'$, 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. The same values of ρ are considered as in the AR(1)-HOMO model.

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

The MA(1)-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(1) random variables with variance 1 and MA parameter ψ.[13] The values of ψ that are considered are: .3, .5, .7, .99, -.3, -.7.

The first table of simulation results, Table 2, provides a comparison of the five kernels of (2.7). The table presents ratios of the finite sample MSEs of the WH/TR, NW/BT, GAL/PR, and TH estimators to those of the QS estimator for each model scenario and T = 128. Each kernel estimator has its bandwidth/lag truncation parameter set equal to its non-random, finite sample optimal value (determined by grid search) to ensure comparability of the kernels.

The table shows that the QS estimator is slightly more efficient than the GAL/PR estimator and very slightly more efficient than the TH estimator in the scenarios considered. These results are basically consistent with the asymptotic results for kernel comparisons given in Theorem 2 and Theorem 3 Comment 4. The finite sample advantage of the QS kernel over the GAL/PR kernel, however, is clearly less than its asymptotic advantage. For these kernels, results corresponding to those of Table 2, but for sample sizes T = 64 and T = 256, are quite similar to those of Table 2.

In Table 2, the three estimators QS, GAL/PR, and TH consistently exhibit a distinct, but not huge, advantage over the NW/BT estimator. This advantage is predicted by the asymptotic results of Theorem 1 (also see Theorem 3 Comment 4), since the NW/BT estimator has MSE that converges to zero at a slower rate (viz., $O(1/T^{2/3})$) than the other three estimators ($O(1/T^{4/5})$). It is interesting to note that for sample size T = 256, the MSE advantage of the QS, GAL/PR, and TH estimators over the NW/BT estimator is more pronounced than in Table 2 where T = 128. This is expected given the asymptotic results.

For all of the estimators, the results of Table 2 are not changed much when the MSE criterion is replaced by the MAE criterion. The only change that occurs is that the differences between the estimators are somewhat less pronounced.

The WH/TR estimator exhibits wide fluctuations in its MSE relative efficiency with respect to the QS estimator and the other three estimators. In the AR(1)-HOMO model, it ranges from being 9% less efficient to 7% more efficient than the QS estimator. For most scenarios, however, it is more efficient than the QS estimator. This is what is suggested by the asymptotic results (see Lemma 2 Comment 2), since the bias of the WH/TR estimator declines at a faster rate than it does for the other estimators. Results corresponding to Table 2 but with sample sizes T = 64 and T = 256 show that the relative efficiency of the WH/TR estimator is increasing with T (i.e., the ratios of MSEs are declining) in most scenarios, but at a fairly slow rate.

Comparisons of the true confidence levels of the CIs constructed using the five different variance estimators are not given in the tables, because they are quite similar to the comparisons based on MSEs given in Table 2. In all cases, the CIs' true confidence levels fall short of their nominal confidence levels. Thus, the best CIs are the ones whose confidence levels are the largest. Of the NW/BT, GAL/PR, TH, and QS-based CIs, the QS-based CIs are fairly consistently the best, but only by a slight margin over the GAL/PR and TH-based CIs. The margin is larger with respect to the NW/BT-based CIs. There are two reasons why the NW/BT-based CIs do worse than the other CIs. First, the NW/BT variance estimator has greater MSEs than do the other estimators, and second, its squared bias/variance ratio is significantly larger than that of the other estimators in most cases. The latter property is to be expected given the asymptotic properties of the estimators (see Theorem 3 Comment 3).

The true confidence level results for the WH/TR-based CIs are similar to the WH/TR estimator's MSE results. In some scenarios they are the best CIs, and in some scenarios they are the worst. The scenarios in which they are best and worst are the same scenarios where the WH/TR estimator has lowest and highest MSEs, respectively, in Table 2.

One drawback of the WH/TR estimator (as well as the TH estimator) is that it does not necessarily generate non-negative variance estimates. In the Monte Carlo experiments, however, a significant number of negative estimates arise only when there is very heavy autocorrelation. For example, in the AR(1)-HOMO model with $\rho = .95$, the percentages of negative WH/TR estimates are 7.6, 1.2, and 0 for T = 64, 128, and 256, respectively (when the finite sample optimal lag truncation parameter was used). For smaller values of ρ, the percentages are all zero for all sample sizes considered.

A second drawback of the WH/TR estimator is that the formulae (5.8) and (7.1) for the first order asymptotically optimal fixed and automatic lag truncation parameters, M_T^* and $\hat{M}_T$ respectively, do not apply to it. Nevertheless, in the Monte Carlo results, it was found that these formulae work quite well for the WH/TR estimator if one treats it in the same fashion as the QS kernel. That is, one takes $q = 2$ and $k_2 = k_{2QS} = 1.4212$. The latter corresponds to $\bar{k}_2 = .3551$ for the normalization of the truncated kernel given in (2.7). This yields

$$\text{(9.2)} \qquad M_T^* = .6611(\alpha T)^{1/5} \quad \text{and} \quad \hat{M}_T = .6611(\hat{\alpha} T)^{1/5}$$

for the WH/TR estimator using the normalization of (2.7).

For brevity, we only discuss results for the QS estimator in the remainder of this section. For the most part, in the tables that follow, the relative performances of the other kernel estimators in comparison with the QS estimator follow patterns similar to those observed in Table 2.

Table 3 is designed to show how well the first order asymptotically optimal bandwidth/lag truncation parameter formulae (given in Theorem 3 and (5.9)) work. The table gives the ratio of the MSE of the QS estimator using S_T^* to its

MSE using the S_T value that minimizes its finite sample MSE. We will call the latter value the finite sample optimal S_T value. All model scenarios and sample sizes are reported in the table.

For the AR(1)-HOMO and MA(1)-HOMO models in Table 3, the value of α used in the formula for S_T^* is determined from (5.3) and (5.7), respectively, using the fact that $\{V_t\} = \{X_{2t}U_t\}$ has homoskedastic AR(1) and MA(1) correlation structures with AR(1) parameter ρ^2 and MA(1) parameter ψ^2 in these models. In the AR(1)-HET1 and AR(1)-HET2 models, $\{X_{2t}U_t\}$ does not have an AR(1) correlation structure. Its correlation structure is parametric, but complicated. The AR(1)-HOMO model is used to approximate the correlation structure in these models. Thus, the same values of α are used in the formula for S_T^* in the heteroskedastic AR(1) models as in the homoskedastic AR(1) model.

Table 3 shows that in general the S_T^* bandwidth values work very well. This is true in both the homoskedastic and heteroskedastic cases. The S_T^* values work better with positive serial dependence than with the less common case of negative serial dependence. They work better in the AR(1) models than in the MA(1) model. In addition, they work better with smaller values of ρ and ψ than with very large values. No clear improvement or deterioration of the MSE ratios occurs as T increases from 64 to 128 to 256.

The analogue of Table 3 that uses true confidence levels rather than MSEs as the performance criteria exhibits patterns similar to those of Table 3, and hence, is not reported here. The main differences are that the relatively poor performance in Table 3 for negative ρ and ψ values and for the MA(1)-HOMO model does not occur. In addition, the relatively poor performance in Table 3 when $\rho = .9$ or $.95$ and $T = 64$ or 128 is accentuated. For other parameter values and sample sizes, the QS-based CIs constructed using S_T^* are close to being as good as those constructed using the "best" fixed S_T value (i.e., the S_T value that minimizes the difference between the true and the nominal confidence levels).

In conclusion, in terms of both MSE and true confidence levels, the asymptotic formula for S_T^* works surprisingly well in selecting the bandwidth parameter for the QS estimator (at least in the models under consideration). For the other kernels, the results for the asymptotic formulae for the lag truncation parameters M_T^* are similar.

Table 4 assesses the performance of the automatic bandwidth procedure introduced in Section 7 (see (7.1)). In all scenarios, the approximating parametric models used by the automatic bandwidth procedure are AR(1) models.

The results of Table 4 are similar to those of Table 3 but in general the QS estimator does worse using $\hat{S}_T$ than using S_T^*, as one would expect. The difference in many cases is not large and in some cases $\hat{S}_T$ outperforms S_T^*. In particular, in each of the three AR(1) models, when $\rho = .9$ or $.95$ and $T = 64$, 128, or 256, $\hat{S}_T$ does better than S_T^*. The reason is that S_T^* exceeds the finite sample optimal S_T value in these scenarios and the downward bias of the AR(1) parameter estimators causes $\hat{S}_T$ to be less than S_T^* and closer to the finite sample optimal value on average. With the exception of these cases, the use of $\hat{S}_T$ rather than S_T^* generally incurs a penalty of a 0-10% increase in MSE.

The analogue of Table 4 that uses true confidence levels rather than MSEs as the performance criterion puts the automatic bandwidth parameter $\hat{S}_T$ in an even better light than does Table 4. In virtually every case, the use of $\hat{S}_T$ incurs only a small reduction in the true confidence levels from the true level obtained using the best fixed S_T value. (The latter confidence levels are always less than or equal to the nominal level, so a reduction always corresponds to an increase in the disparity between the true and nominal levels.) For example, in most scenarios, the reduction in the confidence level for the nominal 95% CIs is in the range of 0 to 1%.

In conclusion, the automatic bandwidth procedure $\hat{S}_T$ performs quite well in terms of MSE and true confidence levels in comparison with the optimal finite sample bandwidth (in the models considered).

Tables 5-7 aim to show how well kernel HAC estimators perform in comparison with other types of variance estimators, viz., INID, IID, and PAR. The kernel estimator used for all three tables is the QS estimator with the automatic bandwidth parameter $\hat{S}_T$ discussed above. The results for other kernels and other bandwidth choices (such as S_T^* and the finite sample optimal S_T value) can be deduced reasonably well from the comparative results given above.

Table 5 presents detailed results for the AR(1)-HOMO model with sample size $T = 128$. Table 6 presents analogous, but less detailed, results for a subset of parameter values in the AR(1)-HET1, AR(1)-HET2, and MA(1)-HOMO models with sample size $T = 128$. Table 7 presents a selected set of results for all four models with sample size $T = 256$.

The first feature of note in Tables 5-7 is that the QS estimator basically dominates INID, and PAR basically dominates IID, over all model scenarios. When ρ or ψ equals zero, INID and IID are at most slightly better than QS and PAR, respectively. When ρ or ψ is non-zero, QS and PAR usually are distinctly superior to INID and IID, respectively. Thus, when no autocorrelation is present, one pays a small price for using a HAC estimator with an automatic bandwidth parameter rather than a heteroskedasticity consistent estimator of the Eicker-White form. On the other hand, when autocorrelation is present, one stands to gain significantly from the use of a HAC estimator rather than an Eicker-White type estimator.

The next feature of note in Tables 5-7 is the very poor performance of all of the estimators in the AR(1) models when $\rho = .9$ or $.95$. This is expected for INID and IID, but it also is true for QS and PAR. For the QS estimator, this poor performance is not due to poor choices of S_T or to the choice of kernel--the results are improved little or none if $\hat{S}_T$ is replaced by the finite sample optimal S_T value or if the QS kernel is replaced any of the other four kernels.

A comparison of the QS and PAR estimators for sample size T = 128 (Tables 5 and 6) shows that PAR is better than QS in the AR(1)-HOMO and MA(1)-HOMO models in terms of MSE and true confidence levels. The differences in MSE are quite large for $\rho \leq .7$; the differences in true confidence levels are much smaller. In the AR(1)-HET1 model, the reverse is true. The QS estimator is much better than PAR in terms of both MSE and true confidence levels over the entire range of ρ values. In the AR(1)-HET2 model, neither QS nor PAR is dominant. PAR enjoys an edge in MSE, but QS is better in terms of true confidence levels.

In sum, for T = 128, PAR is the best all-round estimator if one ignores the AR(1)-HET1 model. Even PAR performs very poorly in each of the AR(1) models, however, when $\rho = .9$ or .95. If one includes the AR(1)-HET1 model, then the QS estimator is the best all-round estimator, since PAR does very poorly in this model. Nevertheless, the QS estimator pays a significant price for attaining its versatility, as the comparison with PAR in the AR(1)-HOMO model attests.

Next we discuss the changes that occur in the results when the sample size is increased from 128 to 256 (see Table 7). For the INID and IID estimators, there is not much change. When $\rho = 0$ or $\psi = 0$ there are improvements in their MSEs and some improvements in their true confidence levels. But, when $\rho > 0$ or $\psi > 0$, there is not much improvement in either their MSEs or their true confidence levels. In consequence, the dominance of QS over INID and PAR over IID is enhanced when the sample size is increased.

For the QS and PAR estimators, the increase in sample size from 128 to 256 causes a substantial improvement in their MSEs and true confidence levels in the AR(1)-HOMO model, especially for large values of ρ. The gap between the true confidence levels of the QS and PAR estimators is narrowed. In the AR(1)-HET1 and AR(1)-HET2 models the QS estimator exhibits similar improvements when the sample size is increased. The PAR estimator, however, shows no improvement in the AR(1)-HET1 model and only small improvements in the AR(1)-HET2 model. In consequence, the dominance of QS over PAR in the AR(1)-HET1 model is accentuated when $T = 256$, and the lack of dominance of either QS or PAR in the AR(1)-HET2 model when $T = 128$ is replaced by dominance of QS when $T = 256$. In the MA(1)-HOMO model, QS and PAR both improve in MSE with the sample size increase, QS improves in true confidence levels, but PAR does not.

In sum, the increase in sample size from $T = 128$ to $T = 256$ improves the overall performance of the QS estimator absolutely and relatively to the PAR, INID, and IID estimators. As when $T = 128$, QS has the best overall performance of the four estimators when $T = 256$ if one includes the AR(1)-HET1 model. PAR is the best estimator overall if this model is excluded. In the latter case, the preference for PAR over QS in terms of true confidence levels is much less when $T = 256$ than when $T = 128$.

10. CONCLUSION

The results of this paper are summarized as follows:

(i) The paper compares different kernel HAC estimators in the literature via asymptotic and simulation methods. The paper establishes an asymptotically optimal kernel, viz., the QS kernel, from the class of kernels that generate psd estimates. The latter includes the Bartlett and Parzen kernels. The Monte Carlo results substantiate the optimality of the QS kernel within this class, both in terms of MSE performance and in terms of true confidence level performance. The difference in performance between the kernels, however, is smaller than the asymptotic results suggest.

The asymptotics show the QS and Tukey-Hanning estimators to be nearly equal in terms of MSE. The Monte Carlo results bear this out. The asymptotics show that the White/truncated estimator has a bias that converges to zero at a faster rate than those of the other estimators. In the finite sample results, this sometimes corresponds to a smaller MSE for the Bartlett estimator than for the other estimators, but in some cases the Bartlett estimator has the largest MSE.

Overall, the Monte Carlo results indicate that the differences between the kernels are not large, with the exception of the Bartlett kernel. The asymptotics and finite sample results indicate that the Bartlett kernel is inferior to the other kernels considered.

(ii) The paper determines suitable fixed and automatic bandwidth/lag truncation parameters for use with HAC estimators. These parameters are determined using asymptotics and are evaluated via Monte Carlo simulation. The fixed and automatic parameters are found to perform very well in most cases in the simulations. This is true for all sample sizes considered and for both the MSE and the true confidence level criteria of performance. These results are particularly useful, because the simulations confirm that the performance of a HAC estimator is sensitive to the choice of bandwidth/lag truncation parameter.

(iii) The paper compares the performance of kernel HAC estimators with other types of covariance matrix estimators via Monte Carlo simulation. The other estimators considered are the Eicker-White heteroskedasticity consistent estimator, the standard LS covariance matrix estimator (IID), and a parametric estimator (PAR) that assumes that the errors are homoskedastic and AR(1). The QS HAC estimator more or less dominates the Eicker-White estimator, and the PAR estimator more or less dominates the IID estimator. The PAR estimator is better than the QS HAC estimator when the errors are homoskedastic but not when they are heteroskedastic. The performance of the QS HAC estimator relative to that of the PAR estimator improves as the sample size increases. In sum, the HAC estimator is the most versatile estimator of those considered. But, it pays a significant price for its versatility when the errors are homoskedasitc, as is illustrated by its performance relative to that of the IID or PAR estimators.

All of the estimators considered perform poorly in an absolute sense when the amount of autocorrelation is large--even the PAR estimator. For the HAC estimator, this is found to be true even if the finite sample optimal bandwidth/lag truncation parameter is used.

(iv) The paper establishes the consistency of kernel HAC estimators under conditions that are more general in some respects than other results in the literature. In particular, they are more general with respect to the class of kernels considered and with respect to the allowable rate of increase of the bandwidth/lag truncation parameters. In addition, the paper establishes rates of convergence to the estimand of HAC estimators. If suitable bandwidth/lag truncation parameters are used, these rates are fastest for the WH/TR estimator, next fastest for the QS, GAL/PR, and TH estimators, and slowest for the NW/BT estimator.

APPENDIX OF PROOFS

PROOF OF LEMMA 1: Let $Y_t = b'V_t$. For any $j = -T+1, \ldots, T-1$,

$$b'(\tilde{\Gamma}(j) - \Gamma_{TP}(j))b = \frac{1}{T}\sum_{t=|j|+1}^{T}(Y_t Y_{t-|j|} - E_P Y_t Y_{t-|j|}).$$

For any j, $m = -T+1, \ldots, T-1$, let j, m denote $|j|$, $|m|$. Then,

$$\sup_{P\in\mathcal{P}_1}|E_P b'(\tilde{\Gamma}(j) - \Gamma_{TP}(j))bb'(\tilde{\Gamma}(m) - \Gamma_{TP}(m))b|$$

$$= \sup_{P\in\mathcal{P}_1}\left|\frac{1}{T^2}\sum_{s=j+1}^{T}\sum_{t=m+1}^{T}(E_P Y_s Y_{s-j} Y_t Y_{t-m} - E_P Y_s Y_{s-j} E_P Y_t Y_{t-m})\right|$$

$$= \sup_{P\in\mathcal{P}_1}\left|\frac{1}{T^2}\sum_{s=j+1}^{T}\sum_{t=m+1}^{T}(E_P Y_s Y_{s-j} E_P Y_t Y_{t-m} + E_P Y_s Y_t E_P Y_{s-j} Y_{t-m}\right.$$

$$\text{(A.1)} \qquad \left. + E_P Y_s Y_{t-m} E_P Y_{s-j} Y_t + \kappa_{Pb}(s,s-j,t,t-m) - E_P Y_s Y_{s-j} E_P Y_t Y_{t-m})\right|$$

$$\leq \frac{1}{T^2}\sum_{s=j+1}^{T}\sum_{t=m+1}^{T}\Big(b'\Gamma_1(t-s)bb'\Gamma_1(t-s+j-m)b + b'\Gamma_1(t-s-m)bb'\Gamma_1(t-s+j)b$$

$$+ \kappa_b(-j,t-s,t-s-m)\Big)$$

$$\leq E_{P_1} b'(\tilde{\Gamma}(j)-\Gamma_{TP_1}(j))bb'(\tilde{\Gamma}(m)-\Gamma_{TP_1}(m))b + \frac{2}{T^2}\sum_{s=j+1}^{T}\sum_{t=m+1}^{T}\kappa_b(-j,t-s,t-s-m)$$

where the second equality holds by the definition of the fourth order cumulant $\kappa_{Pb}(s,s-j,t,t-m)$ of $\{Y_t\}$ under P, the first inequality holds by the definition of $\mathcal{P}_1$, and the second inequality holds by reversing the argument of the first and second equalities and the first inequality.

By a similar argument, we get

$$\inf_{P \in \mathcal{P}_0} E_P b'(\tilde{\Gamma}(j) - \Gamma_{TP}(j))bb'(\tilde{\Gamma}(m) - \Gamma_{TP}(m))b$$

(A.2)

$$\geq E_{P_0} b'(\tilde{\Gamma}(j) - \Gamma_{TP_0}(j))bb'(\tilde{\Gamma}(m) - \Gamma_{TP_0}(m))b - \frac{2}{T^2} \sum_{s=j+1}^{T} \sum_{t=m+1}^{T} \kappa_b(-j,t-s,t-s-m).$$

Let $\bar{J}_T$ denote the pseudo-estimator based on $|k(\cdot)|$, S_T, and $\{V_t\}$. Note that $k(\cdot) \in \mathcal{K}_1$ implies $|k(\cdot)| \in \mathcal{K}_1$.

For $k(\cdot) \in \mathcal{K}_1$, we have

$$\lim_{T\to\infty} \frac{T}{S_T} \mathrm{Var}_{P_1}(b'\tilde{J}_T b)$$

$$\leq \lim_{T\to\infty} \sup_{P \in \mathcal{P}_1} \frac{T}{S_T} \mathrm{Var}_P(b'\tilde{J}_T b)$$

$$= \lim_{T\to\infty} \sup_{P \in \mathcal{P}_1} \frac{T}{S_T} E_P \left(\sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) b'(\tilde{\Gamma}(j) - \Gamma_{TP}(j))b \right)^2$$

(A.3)

$$= \lim_{T\to\infty} \sup_{P \in \mathcal{P}_1} \frac{T}{S_T} \sum_{j=-T+1}^{T-1} \sum_{m=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) k\left(\frac{m}{S_T}\right) E_P b'(\tilde{\Gamma}(j)-\Gamma_{TP}(j))bb'(\tilde{\Gamma}(m)-\Gamma_{TP}(m))b$$

$$\leq \lim_{T\to\infty} \frac{T}{S_T} \sum_{j=-T+1}^{T-1} \sum_{m=-T+1}^{T-1} \left| k\left(\frac{j}{S_T}\right) k\left(\frac{m}{S_T}\right) \right| \cdot E_{P_1} b'(\tilde{\Gamma}(j)-\Gamma_{TP_1}(j))bb'(\tilde{\Gamma}(m)-\Gamma_{TP_1}(m))b$$

$$+ \lim_{T\to\infty} \frac{2}{S_T T} \sum_{j=-T+1}^{T-1} \sum_{m=-T+1}^{T-1} \sum_{u=-\infty}^{\infty} \kappa_b(-j,u,u-m)$$

$$= \lim_{T\to\infty} \frac{T}{S_T} \mathrm{Var}_{P_1}(b'\bar{J}_T b),$$

where the second inequality uses (A.1). If $k(\cdot) \in \mathcal{K}_1$, then $\tilde{J}_T = \bar{J}_T$ and (A.3) establishes the first equality of part (b) of the Lemma. For arbitrary $k(\cdot) \in \mathcal{K}_1$, however, we need to use the argument below to establish this equality.

For $k(\cdot) \in \mathcal{K}_0$, an argument analogous to that of (A.3) using (A.2) in place of (A.1) gives

$$\text{(A.4)}\qquad \begin{aligned} \lim_{T\to\infty} \frac{T}{S_T} \mathrm{Var}_{P_0}(b'\tilde{J}_T b) &\geq \lim_{T\to\infty} \inf_{P\in\mathcal{P}_0} \frac{T}{S_T} \mathrm{Var}_P(b'\tilde{J}_T b) \\ &\geq \lim_{T\to\infty} \frac{T}{S_T} \mathrm{Var}_{P_0}(b'\tilde{J}_T b). \end{aligned}$$

Under the assumptions, by Theorem 9 and its proof in Hannan (1970, pp. 280 and 313-316) which is based on Parzen's (1957) Theorem 5A, for all $k(\cdot) \in \mathcal{K}_1$ we have

$$\text{(A.5)}\qquad \lim_{T\to\infty} \mathrm{Var}_{P_v}(b'\tilde{J}_T b) = 8\pi^2 f_{vb}^2 \int_{-\infty}^{\infty} k^2(x)dx = 8\pi^2 f_{vb}^2 \text{ and}$$

$$\text{(A.6)}\qquad \lim_{T\to\infty} \mathrm{Var}_{P_v}(b'\bar{J}_T b) = 8\pi^2 f_{vb}^2 \int_{-\infty}^{\infty} |k(x)|^2 dx = 8\pi^2 f_{vb}^2 ,$$

for $v = 0, 1$.

Equations (A.3), (A.5), and (A.6) combine to establish part (b) of the Lemma. Equations (A.4) and (A.5) yield part (a). □

PROOF OF LEMMA 2: For part (b), we have

$$\text{(A.7)}\qquad \begin{aligned} Q &= \lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} S_T^q |E_P b'\tilde{J}_T b - b'J_{TP} b| \\ &= \lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} S_T^q \left| \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) b'\Gamma_{TP}(j)b - \sum_{j=-T+1}^{T-1} b'\Gamma_{TP}(j)b \right| \\ &= \lim_{T\to\infty} S_T^q \sum_{j=-T+1}^{T-1} \left(1 - k\left(\frac{j}{S_T}\right)\right)\left(1 - \frac{|j|}{T}\right) b'\Gamma_1(j)b \\ &= \lim_{T\to\infty} \sum_{\substack{j=-T+1 \\ j\neq 0}}^{T-1} \frac{1 - k(j/S_T)}{|j/S_T|^q} |j|^q b'\Gamma_1(j)b - \lim_{T\to\infty} \frac{S_T^q}{T} \sum_{j=-T+1}^{T-1} \left(1-k\left(\frac{j}{S_T}\right)\right) |j| b'\Gamma_1(j)b, \end{aligned}$$

where the third equality holds because $1 - k(j/S_T) \geq 0$ and

$|b'\Gamma_{TP}(j)b| \le b'\Gamma_{TP_1}(j)b = (1 - |j|/T)b'\Gamma_1(j)b,\ \forall j.$

If $f_{1b}^{(q)} < \infty$ and $q \ge 1$, then

$$\text{(A.8)}\qquad \begin{aligned} G &= \lim_{T\to\infty} \frac{S_T^q}{T} \sum_{j=-T+1}^{T-1} \left(1 - k\left(\frac{j}{S_T}\right)\right)|j|b'\Gamma_1(j)b \\ &\le \lim_{T\to\infty} \frac{2S_T^q}{T} \sum_{j=-\infty}^{\infty} |j|b'\Gamma_1(j)b = 0 \end{aligned}$$

since $S_T^q/T \to 0$ as $T \to \infty$. If $f_{1b}^{(q)} < \infty$ and $0 < q < 1$, then

$$\text{(A.9)}\qquad G \le \lim_{T\to\infty} 2S_T^q \sum_{j=-T+1}^{T-1} \left|\frac{j}{T}\right|^q b'\Gamma_1(j)b = 0$$

since $S_T/T \to 0$ as $T \to \infty$. If $f_{1b}^{(q)} < \infty$ and $q = 0$, then

$$\text{(A.10)}\qquad G \le \lim_{T\to\infty} 2\left[\sum_{j=-T+1}^{T-1} b'\Gamma_1(j)b - \sum_{j=-T+1}^{T-1}\left(1 - \frac{|j|}{T}\right)b'\Gamma_1(j)b\right] = 0,$$

where the equality holds by the standard result that the two sums converge to the same limit $2\sum_{j=-\infty}^{\infty} b'\Gamma_1(j)b$ (which is finite by definition of P_1).

If $f_{1b}^{(q)} < \infty$, then by (A.7), (A.8), (A.9), and (A.10),

$$\text{(A.11)}\qquad Q = \lim_{T\to\infty} \sum_{\substack{j=-T+1 \\ j\neq 0}}^{T-1} \frac{1 - k(j/S_T)}{|j/S_T|^q}|j|^q b'\Gamma_1(j)b.$$

Suppose $k_q < \infty$ and $f_{1b}^{(q)} < \infty$. Let $a(x) = (1 - k(x))/|x|^q$ for $x \neq 0$ and $a(x) = k_q$ for $x = 0$. By definition of k_q, $a(x) \to k_q$ as $x \to 0$. Under the assumptions on $k(\cdot)$, $a(\cdot)$ is non-negative and bounded. Hence, there exists a constant $M < \infty$ such that $a(x) \le M\ \forall x \in R$. Given $\varepsilon > 0$, choose a constant $T_0 < \infty$ such that $\sum_{j=T_0+1}^{\infty} |j|^q b'\Gamma_1(j)b < \varepsilon/(4M)$. Then, using (A.11) we get

$$|Q-2\pi k_q f_{1b}^{(q)}| \le \overline{\lim_{T\to\infty}} \sum_{j=-T_0}^{T_0} \left|a\left(\frac{j}{S_T}\right)-k_q\right| \cdot |j|^q b'\Gamma_1(j)b + \overline{\lim_{T\to\infty}}\, 2 \sum_{j=T_0+1}^{\infty} \left|a\left(\frac{j}{S_T}\right)-k_q\right| \cdot |j|^q b'\Gamma_1(j)b$$

(A.12)

$$\le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary the desired result follows.

Next, suppose $k_q = \infty$ and $0 < f_{1b}^{(q)} \le \infty$. Since $f_{1b}^{(q)} > 0$, $b'\Gamma_1(\ell)b > 0$ for some $\ell \ne 0$. Hence, using (A.7) we have

$$\text{(A.13)} \qquad Q \ge \overline{\lim_{T\to\infty}} \frac{1 - k(\ell/S_T)}{|\ell/S_T|^q} |\ell|^q \left(1 - \frac{|\ell|}{T}\right) b'\Gamma_1(j)b = \infty,$$

where the equality holds by the definition of k_q.

If $k_q = \infty$ and $f_{1b}^{(q)} = 0$, then $Q = 0$ because $f_{1b}^{(q)} = 0$ implies that $b'\Gamma_1(j)b = 0$ $\forall j \ne 0$ and this implies that the last line of (A.7) is zero.

Finally, suppose $f_{1b}^{(q)} = \infty$ and $0 < k_q < \infty$. Since $a(0) = k_q > 0$ and $a(x)$ is continuous in a neighborhood of zero, there exist constants $\varepsilon > 0$ and $c \in (0,1]$ such that $a(x) \ge \varepsilon > 0$ for $|x| \le c$. Hence, using (A.7) we have

$$Q = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} a(j/S_T)|j|^q(1 - |j|/T)b'\Gamma_1(j)b$$

(A.14)

$$\ge \lim_{T\to\infty} \sum_{j=-[cS_T]}^{[cS_T]} \varepsilon|j|^q(1 - [cS_T]/T)b'\Gamma_1(j)b = \infty,$$

since $[cS_T]/T \to 0$ as $T \to \infty$. This completes the proof of part (b).

The proof of part (a) is identical to that of part (b) above except that $\sup_{P\in\mathcal{P}_1}$, $\Gamma_1(j)$, $\Gamma_1(\ell)$, and $f_{1b}^{(q)}$ are replaced by $\inf_{P\in\mathcal{P}_0}$, $\Gamma_0(j)$, $\Gamma_0(\ell)$, and $f_{0b}^{(q)}$, respectively. □

PROOF OF THEOREM 1: For $\gamma \in (0,\infty)$ and $q \in (0,\infty)$, we have $S_T \to \infty$, $S_T^q/T \to 0$, and $S_T/T \to 0$ as $T \to \infty$. When $\gamma = 0$ or $\gamma = \infty$, these conditions hold by assumption. Hence, Lemmas 1 and 2 apply. When $\gamma \in (0,\infty)$, parts (a) and (b) of the Theorem follow from Lemmas 1(a) and 2(a) and Lemmas 1(b) and 2(b), respectively.

For $\gamma = \infty$, $(T/S_T)/S_T^{2q} \to 0$ and so $T/S_T = S_T^{2q} \times \eta_T$ where $\eta_T \to 0$ as $T \to \infty$. Thus, by Lemma 2 and the assumption that $k_q^2(f_{1b}^{(q)})^2 < \infty$ the squared bias terms in parts (a) and (b) of the Theorem converge to zero as $T \to \infty$ as desired. The variance terms in parts (a) and (b) of the Theorem converge to the appropriate limits as $T \to \infty$ by Lemma 1.

For $\gamma = 0$, $(T/S_T)/S_T^{2q} \to \infty$ and so $T/S_T = S_T^{2q} \times \xi_T$ where $\xi_T \to \infty$ as $T \to \infty$. Thus, by Lemma 2 and the assumption that $k_q^2(f_{0b}^{(q)})^2 > 0$ the squared bias terms in parts (a) and (b) of the Theorem diverge to infinity as $T \to \infty$ as desired. □

PROOF OF THEOREM 2: By Lemmas 1(b) and 2(b) with $q = 2$,

$$\text{(A.15)} \qquad \sup_{P\in\mathcal{P}_1} MSE_P(b'\tilde{J}_Tb) = S_T^{-4}4\pi^2k_2^2(f_{1b}^{(q)})^2 + \frac{S_T}{T}8\pi^2f_{1b}^2 + o(S_T^{-4}) + o(S_T/T)$$

as $T \to \infty$. The MSE of $b'Q\tilde{S}_Tb$ is analogous but with k_2^2 replaced by k_{2QS}^2. If $\gamma \in (0,\infty]$, then $r_T \times o(S_T^{-4}) = T/S_T \times o(S_T^{-4}) = o(1)$ and $r_T \times o(S_T/T) = o(1)$. If $\gamma = 0$, then $r_T \times o(S_T/T) = S_T^4 \times o(S_T/T) = o(1)$ and $r_T \times o(S_T^{-4}) = o(1)$. Hence,

$$\underline{\lim}_{T\to\infty} r_T[\sup_{P\in\mathcal{P}_1} MSE_P(b'\tilde{J}_Tb) - \sup_{P\in\mathcal{P}_1} MSE_P(b'Q\tilde{S}_Tb)]$$

$$= \underline{\lim}_{T\to\infty} r_TS_T^{-4}4\pi^2(f_{1b}^{(2)})^2(k_2^2 - k_{2QS}^2)$$

$$\text{(A.16)} \qquad = \begin{cases} 4\pi^2(f_{1b}^{(2)})^2(k_2^2 - k_{2QS}^2)/\gamma & \text{if } \gamma \in (0,\infty] \\ 4\pi^2(f_{1b}^{(2)})^2(k_2^2 - k_{2QS}^2) & \text{if } \gamma = 0 \end{cases}$$

$$\geq 0,$$

where the inequality follows from the result (proved below) that for all $k(\cdot) \in \mathcal{K}_2$, $k_2 \geq k_{2QS}$. The inequality of (A.16) is strict if $k_2 > k_{2QS}$, $\gamma \neq \infty$, and $f_{1b}^{(2)} > 0$. The former condition holds for $k(\cdot) \in \mathcal{K}_2$ if and only if $k(x) \neq k_{QS}(x)$ with positive Lebesgue measure (see below).

Since $k_2 = \int_{-\infty}^{\infty}\lambda^2 K(\lambda)d\lambda$, $k(0) = \int_{-\infty}^{\infty}K(\lambda)d\lambda$, and $\int_{-\infty}^{\infty}k^2(x)dx = \int_{-\infty}^{\infty}K^2(\lambda)d\lambda$, we have $k_2 \geq k_{2QS}$ for all $k(\cdot) \in \mathcal{K}_2$ if and only if $K_{QS}(\lambda)$ minimizes

$$\text{(A.17)} \qquad \int_{-\infty}^{\infty} \lambda^2 K(\lambda)d\lambda$$

subject to (a) $\int_{-\infty}^{\infty}K(\lambda)d\lambda = 1$, (b) $\int_{-\infty}^{\infty}K^2(\lambda)d\lambda = 1$, (c) $K(\lambda) \geq 0 \ \forall\lambda \in R$, and (d) $K(\lambda) = K(-\lambda) \ \forall\lambda \in R$, where $K_{QS}(\lambda) = \frac{5}{8\pi}(1 - \lambda^2/c^2)$ for $|\lambda| \leq c$ and $K_{QS}(\lambda) = 0$ otherwise for $c = 6\pi/5$. This minimization problem is solved using a calculus of variations argument.

For a function $K(\lambda)$ satisfying the conditions above, write $K(\lambda) = K_{QS}(\lambda) + \varepsilon(\lambda)$. Condition (a) implies: (1) $\int_{-\infty}^{\infty}\varepsilon(\lambda)d\lambda = 0$. Condition (b) implies: (2) $2\int_{-\infty}^{\infty}K_{QS}(\lambda)\varepsilon(\lambda)d\lambda + \int_{-\infty}^{\infty}\varepsilon^2(\lambda)d\lambda = 0$. Condition (c) implies: (3) $\varepsilon(\lambda) \geq 0$ for $|\lambda| \geq c$. Condition (d) implies: (4) $\varepsilon(\lambda) = \varepsilon(-\lambda) \ \forall\lambda \in R$. We need to show that $\int_{-\infty}^{\infty}\lambda^2 K(\lambda)d\lambda - \int_{-\infty}^{\infty}\lambda^2 K_{QS}(\lambda)d\lambda = \int_{-\infty}^{\infty}\lambda^2\varepsilon(\lambda)d\lambda \geq 0$.

By (1) and (3), $\int_{-\infty}^{\infty}\lambda^2\varepsilon(\lambda)d\lambda > -\infty$. If $\int_{-\infty}^{\infty}\lambda^2\varepsilon(\lambda)d\lambda = \infty$, the desired result holds. Hence, it suffices to consider the case where $\int_{-\infty}^{\infty}\lambda^2\varepsilon(\lambda)d\lambda < \infty$. By (2), (3), (4), and the definition of $K_{QS}(\lambda)$, we have

$$\text{(A.18)} \quad \begin{aligned} 0 &= \int_{-\infty}^{\infty}\varepsilon^2 + 4\int_0^c K_{QS}\varepsilon + 4\int_c^{\infty}K_{QS}\varepsilon = \int_{-\infty}^{\infty}\varepsilon^2 + \frac{5}{2\pi}\int_0^c(1 - \lambda^2/c^2)\varepsilon(\lambda)d\lambda \\ &= \int_{-\infty}^{\infty}\varepsilon^2 - \frac{5}{2\pi c^2}\int_0^{\infty}\lambda^2\varepsilon(\lambda)d\lambda - \frac{5}{2\pi}\int_c^{\infty}(1 - \lambda^2/c^2)\varepsilon(\lambda)d\lambda. \end{aligned}$$

Thus,

$$\text{(A.19)} \qquad \int_{-\infty}^{\infty}\lambda^2\varepsilon(\lambda)d\lambda = \frac{4\pi c^2}{5}\left[\int_{-\infty}^{\infty}\varepsilon^2 + \frac{5}{2\pi}\int_c^{\infty}\frac{\lambda^2 - c^2}{c^2}\varepsilon(\lambda)d\lambda\right] \geq 0.$$

This implies $k_2 \geq k_{2QS}$. The inequality is strict unless $\varepsilon(\lambda) = 0$ almost everywhere Lebesgue on $(-\infty,\infty)$. The latter holds only if $K(\lambda) = K_{QS}(\lambda)$ and $k(x) = k_{QS}(x)$ almost everywhere Lebesgue on $(-\infty,\infty)$. □

PROOF OF THEOREM 3: Suppose $S_T^{2q+1}/T \to \gamma = \infty$ as $T \to \infty$. Then, for any ℓ such that $f_{1b_\ell}^2 > 0$ and $w_\ell > 0$, Theorem 1(b) gives

$$\underline{\lim}_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_1(\alpha)} E_P L(\tilde{J}_T(S_T), J_{TP})$$

$$\geq \underline{\lim}_{T\to\infty} T^{2q/(2q+1)} w_\ell \sup_{P\in\mathcal{P}_1(\alpha)} MSE_P(b_\ell'\tilde{J}_T(S_T)b_\ell)$$

$$(A.20)\quad = \underline{\lim}_{T\to\infty}(S_T^{2q+1}/T)^{1/(2q+1)} w_\ell \frac{T}{S_T} \sup_{P\in\mathcal{P}_1(\alpha)} MSE_P(b_\ell'\tilde{J}_T(S_T)b_\ell)$$

$$= \gamma^{1/(2q+1)} w_\ell 8\pi^2 f_{1b_\ell}^2 = \infty\ .$$

Alternatively, suppose $\gamma = 0$. Then, for any ℓ such that $(f_{1b_\ell}^{(q)})^2 > 0$ and $w_\ell > 0$, Lemma 2(b) gives

$$\underline{\lim}_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_1(\alpha)} E_P L(\tilde{J}_T(S_T), J_{TP})$$

$$\geq \underline{\lim}_{T\to\infty} T^{2q/(2q+1)} w_\ell \sup_{P\in\mathcal{P}_1(\alpha)} MSE_P(b_\ell'\tilde{J}_T(S_T)b_\ell)$$

$$(A.21)\quad \geq \underline{\lim}_{T\to\infty}\left[1/(S_T^{2q+1}/T)^{2q/(2q+1)}\right] w_\ell S_T^{2q} \sup_{P\in\mathcal{P}_1(\alpha)} (E_P b_\ell'\tilde{J}_T(S_T)b_\ell - b_\ell' J_{TP} b_\ell)^2$$

$$= (1/\gamma^{2q/(2q+1)}) w_\ell (2\pi k_q f_{1b_\ell}^{(q)})^2 = \infty\ .$$

Next, suppose $\gamma \in (0,\infty)$. Then, by Theorem 1(b),

$$\varliminf_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_1(\alpha)} E_P L(\tilde{J}_T(S_T),\ J_{TP})$$

(A.22) $$= \varliminf_{T\to\infty}(S_T^{2q+1}/T)^{1/(2q+1)} \frac{T}{S_T} \sup_{P\in\mathcal{P}_1(\alpha)} \sum_{\ell=1}^{r} w_\ell MSE_P(b_\ell'\tilde{J}_T(S_T)b_\ell)$$

$$= \gamma^{1/(2q+1)}4\pi^2[\sum_{\ell=1}^{r} w_\ell(k_q^2(f_{1b_\ell}^{(q)})^2/\gamma + 2f_{1b_\ell}^2)] \ .$$

It is straightforward to show that the right-hand-side above is uniquely minimized over $\gamma \in (0,\infty)$ at $\gamma^* = qk_q^2\alpha$ and that a sequence $\{S_T\}$ satisfies $S_T^{2q+1}/T \to \gamma^*$ as $T \to \infty$ if and only if $S_T = S_T^* + o(T^{1/(2q+1)})$ as $T \to \infty$. □

PROOF OF THEOREM 4: First we establish part (a). By the proofs of Lemmas 1 and 2,

(A.23) $$\sup_{P\in\mathcal{P}_1} MSE_P(b_\ell'\tilde{J}_T(S_T)b_\ell) = MSE_{P_1}(b_\ell'\tilde{J}_T(S_T)b_\ell) + O(T^{-2})$$

as $T \to \infty$.

Also by the proofs of Lemmas 1 and 2,

(A.24) $$MSE_{P_1}(b_\ell'\tilde{J}_T(S_T)b_\ell) = B_{T\ell}^2(S_T) + V_{T\ell}(S_T) + O(T^{-2})$$

as $T \to \infty$, where the $O(T^{-2})$ term is the fourth order cumulant term.

Combining these results shows that the left hand side (LHS) of the expression in part (a) satisfies

(A.25) $$\text{LHS} = \varliminf_{T\to\infty} \zeta_T \sum_{\ell=1}^{r} w_\ell[B_{T\ell}^2(S_T) + V_{T\ell}(S_T) - B_{T\ell}^2(S_T^{**}) - V_{T\ell}(S_T^{**})] \geq 0 \ .$$

To establish part (b), note that in (A.24) above the term $O(T^{-2})$ is identically zero if $\{b_\ell'V_t\}$ is Gaussian under P_1. This holds because the fourth order cumulant of a Gaussian sequence is zero. In this case,

$MSE_{P_1}(b_\ell' \tilde{J}_T(S) b_\ell) = B_{T\ell}^2(S) + V_{T\ell}(S)$ and the result follows. □

PROOF OF THEOREM 5: Since the degrees of freedom correction T/(T-r) that appears in the definition of $\hat{J}_T$ has no effect on the asymptotic properties of $\hat{J}_T$, we proceed in this proof as though it does not appear in the definition.

Let $C_T = b'\hat{J}_T b - b'\tilde{J}_T b$. We have

$$\lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} |MSE_P(b'\hat{J}_T b) - MSE_P(b'\tilde{J}_T b)|$$

$$\text{(A.26)} \quad = \lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} |2E_P(b'\tilde{J}_T b - b'J_{TP} b)B_T + E_P C_T^2|$$

$$\leq 2 \lim_{T\to\infty}\left(\sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} MSE_P(b'\tilde{J}_T b)\right)^{1/2} \cdot \left(\sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} E_P C_T^2\right)^{1/2} + \lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} E_P C_T^2.$$

The right-hand-side above equals zero as desired if (i) $\lim_{T\to\infty} \sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} E_P C_T^2 = 0$ and (ii) $\overline{\lim}_{T\to\infty} \sup_{P\in\mathcal{P}_1} \left(\frac{T}{S_T}\right)^{\xi} MSE_P(b'\tilde{J}_T b) < \infty$. Result (ii) follows for $\xi = 0$ or 1 by Theorem 1(b) since $|f_{1b}| < \infty$, $k_q < \infty$, $f_{1b}^{(q)} < \infty$, and $\gamma > 0$.

For part (a) and $\xi = 0$, result (i) is obtained as follows. Let $\tilde{J}_T(\theta)$ denote the "estimator" calculated using $k(\cdot)$, S_T, and $\{V_t(\theta)\}$. A mean value expansion of $b'\hat{J}_T b = b'\tilde{J}_T(\hat{\theta})b$ about θ_0 yields $C_T = \left[\frac{\partial}{\partial\theta'} b'\tilde{J}_T(\bar{\theta})b\right](\hat{\theta} - \theta_0)$ for some $\bar{\theta}$ on the line segment joining $\hat{\theta}$ and θ_0. Straightforward manipulations yield

$$\text{(A.27)} \qquad W_{Tm} = \frac{\partial}{\partial\theta_m} b'\tilde{J}_T(\bar{\theta})b = \sum_{j=-T+1}^{T-1} k(j/S_T)\left(\frac{\partial}{\partial\theta_m} b'\hat{\Gamma}(j)b\right)\Bigg|_{\theta=\bar{\theta}},$$

$$\left|\frac{\partial}{\partial\theta_m}b'\hat{\Gamma}(j)b\right|\Bigg|_{\theta=\bar{\theta}} = \left|\frac{1}{T}\sum_{t=|j|+1}^{T}\left(b'V_t(\bar{\theta})b'\frac{\partial}{\partial\theta_m}V_{t-|j|}(\bar{\theta}) + b'V_{t-|j|}(\bar{\theta})b'\frac{\partial}{\partial\theta_m}V_t(\theta)\right)\right|$$

(A.28)

$$\leq 2\|b\|^2\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|V_t(\theta)\|^2\right)^{1/2}\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_m}V_t(\theta)\right\|^2\right)^{1/2},$$

$$\sup_{P\in P_1}E_PC_T^2 = \sup_{P\in P_1}\sum_{\ell=1}^{r}\sum_{m=1}^{r}E_PW_{T\ell}W_{Tm}(\hat{\theta}_\ell-\theta_{0\ell})(\hat{\theta}_m-\theta_{0m})$$

$$\leq r^2\max_{m\leq r}\sup_{P\in P_1}E_PW_{Tm}^2(\hat{\theta}_m-\theta_{0m})^2$$

(A.29)

$$\leq \frac{S_T^2}{T}r^2\left(\frac{1}{S_T}\sum_{j=-T+1}^{T-1}|k(j/S_T)|\right)^2$$

$$\times 4\|b\|^4\max_{m\leq r}\sup_{P\in P_1}E_P\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\|V_t(\theta)\|^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_m}V_t(\theta)\right\|^2\right)T(\hat{\theta}_m-\theta_{0m})^2.$$

$$= o(1) \text{ as } T\to\infty,$$

where the last equality uses assumption A and the fact that $S_T^2/T \to 0$ when $q \in (1/2,\infty)$ and $\frac{1}{S_T}\sum_{j=-T+1}^{T-1}|k(j/S_T)| \to \int_{-\infty}^{\infty}|k(x)|dx < \infty$ as $T\to\infty$. Thus, result (i) holds for part (a).

For part (b) and $\xi = 1$, a two term Taylor expansion gives

(A.30) $$C_T = \left[\frac{\partial}{\partial\theta'}b'\tilde{J}_T(\theta_0)b\right](\hat{\theta}-\theta_0) + \frac{1}{2}(\hat{\theta}-\theta_0)'\left[\frac{\partial^2}{\partial\theta\partial\theta'}b'\tilde{J}_T(\bar{\theta})b\right](\hat{\theta}-\theta_0) = L_{1P} + L_{2P},$$

where $\bar{\theta}$ lies on the line segment joining $\hat{\theta}$ and θ_0. Hence,

(A.31) $$E_PC_T^2 = E_PL_{1P}^2 + E_PL_{2P}^2 + 2E_PL_{1P}L_{2P}.$$

It is straightforward to show that $\left|\frac{\partial}{\partial\theta}b'\tilde{J}_T(\theta_0)b\right| \leq Z_{1Tm}S_T/\sqrt{T}$ and

$\left|\frac{\partial^2}{\partial\theta\partial\theta'}b'\tilde{J}_T(\bar{\theta})b\right| \le Z_{2T}S_T$. Thus,

$$\text{(A.32)}\quad \begin{aligned} \frac{T}{S_T}\sup_{P\in P_1} E_P L_{1P}^2 &\le \frac{T}{S_T} r^2 \max_{m\le r}\sup_{P\in P_1} E_P\left(\left[\frac{\partial}{\partial\theta_m}b'\tilde{J}_T(\theta_0)b\right](\hat{\theta}_m-\theta_{0m})\right)^2 \\ &\le \frac{S_T}{T} r^2 \max_{m\le r}\sup_{P\in P_1} E_P(Z_{1Tm}\sqrt{T}(\hat{\theta}_m-\theta_{0m}))^2 \to 0 \quad\text{and}\end{aligned}$$

$$\text{(A.33)}\quad \begin{aligned} \frac{T}{S_T}\sup_{P\in P_1} E_P L_{2P}^2 &\le \frac{T}{4S_T}\sup_{P\in P_1} E_P\left(|\hat{\theta}-\theta_0|'\cdot\left|\frac{\partial^2}{\partial\theta\partial\theta'}b'\tilde{J}_T(\bar{\theta})b\right|\cdot|\hat{\theta}-\theta_0|\right)^2 \\ &\le \frac{S_T}{4T}\sup_{P\in P_1} E_P\left(\sqrt{T}|\hat{\theta}-\theta_0|'Z_{2T}\sqrt{T}|\hat{\theta}-\theta_0|\right)^2 \to 0\end{aligned}$$

as $T\to\infty$. Equations (A.31) to (A.33) and the Cauchy-Schwartz inequality give result (i) for part (b). □

The proof of Theorem 6 uses the following Lemma:

LEMMA A1: *Let* $k(\cdot)$, $\{S_T^0\}$, $\{S_{TP}^*\}$, $\hat{\alpha}$, *and* q *be as in Theorem 6. Let* v *be as defined in* P_{11}. *Then,*

(a) $$T^{2q/(2q+1)}\sup_{P\in P_{11}} E_P\left(\sum_{j=[(S_{TP}^*)]^v+1}^{T-1} k(j/S_T^0)b'\hat{\Gamma}(j)b\right)^2 \to 0 \text{ as } T\to\infty \text{ for all } b\in R^r$$

and

(b) $$T^{2q/(2q+1)}\sup_{P\in P_{11}} E_P\left(\sum_{j=1}^{[(S_{TP}^*)^v]}(k(j/S_T^0)-k(j/S_{TP}^*))b'\hat{\Gamma}(j)b\right)^2 \to 0 \text{ as } T\to\infty$$

for all $b\in R^r$.

PROOF OF THEOREM 6: Let $\|\cdot\|_P = (E_P(\cdot)^2)^{1/2}$. For any constant J and any random variables $\hat{J}_1$ and $\hat{J}_2$, the triangle inequality gives

$$\text{(A.34)}\quad \|\hat{J}_1-\hat{J}_2\|_P \ge \big|\|\hat{J}_1-J\|_P - \|\hat{J}_2-J\|_P\big|.$$

Hence, it suffices to show that

$$\text{(A.35)} \qquad T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} \|b'\hat{J}_T(S_T^0)b - b'\hat{J}_T(S_{TP}^*)b\|_P^2 \to 0 \text{ as } T \to \infty.$$

The latter follows from Lemma A1 parts (a) and (b), since

$$\text{(A.36)} \qquad \begin{aligned} b'\hat{J}_T(S_T^0)b - b'\hat{J}_T(S_{TP}^*)b &= 2 \sum_{j=[(S_{TP}^*)^v]+1}^{T-1} (k(j/S_T^0) - k(j/S_{TP}^*))b'\hat{\Gamma}(j)b \\ &+ 2 \sum_{j=1}^{[(S_{TP}^*)^v]} k(j/S_T^0)b'\hat{\Gamma}(j)b - 2 \sum_{j=1}^{[(S_{TP}^*)^v]} k(j/S_{TP}^*)b'\hat{\Gamma}(j)b . \end{aligned}$$

For the third summand, we get $\lim_{T\to\infty} \sup_{P\in\mathcal{P}_{11}} T^{2q/(2q+1)} \left\| \sum_{j=1}^{[(S_{TP}^*)^v]} k(j/S_{TP}^*)b'\hat{\Gamma}(j)b \right\|_P^2 = 0$ by applying Lemma A1 part (a) with $\{S_T^0\}$ of the Lemma given by the fixed sequence $\{S_{TP}^*\}$. □

PROOF OF LEMMA A1: First we prove part (a). We have

$$\left(T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S_{TP}^*)^v]+1}^{T-1} k(j/S_T^0)b'\hat{\Gamma}(j)b\right)^2\right)^{1/2}$$

$$\text{(A.37)} \qquad \begin{aligned} &\le \left(T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S_{TP}^*)^v]+1}^{T-1} k(j/S_T^0)(b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b)\right)^2\right)^{1/2} \\ &+ \left(T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S_{TP}^*)^v]+1}^{T-1} k(j/S_T^0)b'\Gamma_{TP}(j)b\right)^2\right)^{1/2} \\ &= A_{1T} + A_{2T}. \end{aligned}$$

Since $|k(\cdot)| \le 1$ and $|b'\Gamma_{TP}(j)b| \le b'\Gamma_1(j)b$, we get

$$A_{2T} \leq \left(T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S^*_{TP})^v]+1}^{T-1} |k(j/S^0_T)|\cdot|b'\Gamma_{TP}(j)b|\right)^2\right)^{1/2}$$

$$\leq T^{q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} \sum_{j=[(S^*_{TP})^v]+1}^{T-1} b'\Gamma_1(j)b$$

(A.38)
$$\leq T^{q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} \sum_{j=[(S^*_{TP})^v]+1}^{T-1} C_3 j^{-m}$$

$$\leq T^{q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} C_3\int_{[(S^*_{TP})^v]}^{\infty} j^{-m}dj$$

$$\leq C_5 T^{[q-v(m-1)]/(2q+1)}$$

$$\to 0 \text{ as } T \to \infty,$$

for some constant $C_5 \in (0,\infty)$, using the fact that $\inf_{P\in\mathcal{P}_{11}} \alpha_P \geq \varepsilon_* > 0$ and $v > q/(m-1)$.

Next, we have

$$A^2_{1T} \leq T^{-1/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S^*_{TP})^v]+1}^{T-1} C_1(j/S^0_T)^{-b}\sqrt{T}|b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b|\right)^2$$

$$\leq T^{(2b-1)/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=[(S^*_{TP})^v]+1}^{T-1} C_1 j^{-b}\sqrt{T}|b'\hat{\Gamma}(j)b-b'\Gamma_{TP}(j)b|\right)^2\cdot(qk_q\hat{\alpha})^{2b/(2q+1)}$$

(A.39)
$$\leq C_6 T^{(2b-1)/(2q+1)} \sup_{P\in\mathcal{P}_{11}} \sum_{j=[(S^*_{TP})^v]+1}^{T-1}\ \sum_{m=[(S^*_{TP})^v]+1}^{T-1} j^{-b}j^{-m}T\ \mathrm{Var}_P(b'\hat{\Gamma}(j)b)\mathrm{Var}_P(b'\hat{\Gamma}(m)b)$$

$$\leq C_6 T^{(2b-1)/(2q+1)} \sup_{P\in\mathcal{P}_{11}} \left(\sum_{j=[(S^*_{TP})^v]+1}^{T-1} j^{-b}\right)^2\cdot T \sup_{j\geq 1} \mathrm{Var}_{P_1}(b'\hat{\Gamma}(j)b)$$

$$\leq C_7 T^{(2b-1)/(2q+1)} T^{-2(b-1)v/(2q+1)}\cdot T \sup_{j\geq 1} \mathrm{Var}_{P_1}(b'\hat{\Gamma}(j)b)$$

$$\to 0 \text{ as } T \to \infty$$

for some constants C_6 and C_7 in $(0,\infty)$, using the fact that $\hat{\alpha} \le \alpha^* < \infty$, $\inf_{P\in\mathcal{P}_{11}} \alpha_P \ge \varepsilon_* > 0$, and $v > 1 + 1/(2b-2)$.

Equations (A.37), (A.38), and (A.39) combine to establish part (a).

We now prove part (b). Let LHS denote the term that is claimed in part (b) to converge to zero as $T \to \infty$. Using the Lipschitz condition on $k(\cdot)$, we get

$$\text{(A.40)} \quad \begin{aligned} \text{LHS} &\le T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left(\sum_{j=1}^{[(S_{TP}^*)^v]} C_2(1/S_T^0 - 1/S_{TP}^*) jb'\hat{\Gamma}(j)b\right)^2 \\ &\le C_8 T^{-3/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{(\hat{\alpha}\alpha_P)^{1/(2q+1)}}\right]^2 \left[\sum_{j=1}^{[(S_{TP}^*)^v]} jb'\hat{\Gamma}(j)b\right]^2, \end{aligned}$$

for some constant $C_8 \in (0,\infty)$, and hence,

$$\text{(A.41)} \quad \begin{aligned} \text{LHS}^{1/2} &\le \left\{C_8 T^{-3/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{(\hat{\alpha}\alpha_P)^{1/(2q+1)}}\right]^2 \left[\sum_{j=1}^{[(S_{TP}^*)^v]} j(b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b)\right]^2\right\} \\ &\quad + \left\{C_8 T^{-3/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{(\hat{\alpha}\alpha_P)^{1/(2q+1)}}\right]^2 \left[\sum_{j=1}^{[(S_{TP}^*)^v]} jb'\Gamma_{TP}(j)b\right]^2\right\}^{1/2} \\ &= D_{1T} + D_{2T}. \end{aligned}$$

Using the assumption on (S_T^0), we obtain

$$D_{1T}^2 = C_8 T^{-3/(2q+1)+4v/(2q+1)-1} \sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)}-\alpha_P^{1/(2q+1)})}{(\hat{\alpha}\alpha_P)^{1/(2q+1)}}\right]^2$$

$$\times \left[\frac{1}{(S_{TP}^*)^v}\sum_{j=1}^{[(S_{TP}^*)^v]}\frac{j}{(S_{TP}^*)^v}\sqrt{T}(b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b)\right]^2 (qk_q^2\alpha_P)^{4v/(2q+1)}$$

(A.42)

$$\le C_9 T^{(-2q-4+4v)/(2q+1)}\left(\sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{\hat{\alpha}^{1/(2q+1)}}\right]^4\right)^{1/2}$$

$$\times \left(\sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{1}{(S_{TP}^*)^v}\sum_{j=1}^{[(S_{TP}^{*v})]}\sqrt{T}\left|b'\hat{\Gamma}(j)b - b'\Gamma_{TP}(j)b\right|\right]^4\right)^{1/2}$$

$$\to 0 \text{ as } T \to \infty$$

for some constant $C_9 \in (0,\infty)$, since $\sup_{P\in\mathcal{P}_{11}} \alpha_P < \infty$ and $v > 1 + q/2$.

In addition, we have

$$D_{2T}^2 \le T^{-3/(2q+1)} \sup_{P\in\mathcal{P}_{11}} E_P\left[\frac{\sqrt{T}(\hat{\alpha}^{1/(2q+1)} - \alpha_P^{1/(2q+1)})}{(\hat{\alpha}\alpha_P)^{1/(2q+1)}}\right]^2 \cdot \sup_{P\in\mathcal{P}_{11}}\left(\sum_{j=1}^{[(S_{TP}^*)^v]} j^{1-m}\right)^2$$

(A.43)

$$\to 0 \text{ as } T \to \infty,$$

since $m > 2$ implies that $\sum_{j=1}^{\infty} j^{1-m} < \infty$.

Equations (A.41), (A.42), and (A.43) combine to establish part (b). □

PROOF OF THEOREM 7: Under the assumptions on $\{S_T^1\}$,

$$H_1 = \underline{\lim}_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}(\alpha)} E_P L(\hat{J}_T(S_T^1),\ J_{TP})$$

$$\text{(A.44)} \qquad \geq \underline{\lim}_{T\to\infty} T^{2q/(2q+1)} E_{P_1} L(\hat{J}_T(S_T^1),\ J_{TP_1})$$

$$= \underline{\lim}_{T\to\infty} T^{2q/(2q+1)} E_{P_1} L(\hat{J}_T(S_T),\ J_{TP_1}) = H_{11}$$

and

$$H_2 = \overline{\lim}_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}(\alpha)} E_P L(\hat{J}_P(S_T^0),\ J_{TP})$$

$$\text{(A.45)} \qquad = \overline{\lim}_{T\to\infty} T^{2q/(2q+1)} \sup_{P\in\mathcal{P}_{11}(\alpha)} E_P L(\hat{J}_T(S_{TP}^*),\ J_{TP})$$

$$= \overline{\lim}_{T\to\infty} T^{2q/(2q+1)} E_{P_1} L(\hat{J}_T(S_{TP_1}^*),\ J_{TP_1}),$$

using Theorem 6 for the second equality of (A.45) and Theorems 1 and 5 and the fact that S_{TP}^* is the same for all $P \in \mathcal{P}_{11}(\alpha)$ for the last equality.

Equations (A.44) and (A.45) and Theorem 3 give the desired result $H_1 - H_2 \geq H_{11} - H_2 \geq 0$ with equality only if $S_T = S_{TP_1}^* + o(T^{1/(2q+1)})$ as $T \to \infty$.

□

FOOTNOTES

[1] I would like to thank Chris Monahan for excellent research assistance, and Jean-Marie Dufour, Whitney K. Newey, and Peter C. B. Phillips for helpful comments. In addition, I would like to thank the California Institute of Technology and the University of British Columbia for their hospitality while part of this research was carried out, and the Alfred P. Sloan Foundation and the National Science Foundation for research support provided through a Research Fellowship and grant No. SES-8618617 respectively.

[2] Consistency proofs require that the lag truncation parameters diverge to infinity at a bounded rate as the sample size T goes to infinity. This bounded rate typically is an artifact of the method of proof, however, and is not necessarily optimal in any respect. Furthermore, it is not sufficient to determine an optimal <u>rate</u> of divergence; one needs a particular sequence of lag truncation or bandwidth values.

[3] In particular, the estimand J_{TP} is given by $\mathrm{Var}\left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} w_n\right)$ in Hansen (1982), by $\bar{I}_n(\lambda_n^0)$ and $\bar{S}_n(\lambda_n^0)$ in Gallant (1987, pp. 549 and 570), by B_n^0 in Gallant and White (1988, p. 100), and by $\mathrm{Var}_P(\sqrt{T}\, \bar{m}_T(\theta_0, \tau_0))$ in Andrews and Fair (1988).

[4] Note that the parameter S_T indexes the "bandwidth" of the estimator in an inversely proportional fashion.

[5] One advantage of the normalization $\int_{-\infty}^{\infty} k^2(x)dx = 1$ is that the same normalization can be applied to all kernels of interest that satisfy the other requirements of K_1, since if $\int_{-\infty}^{\infty} k^2(x)dx = \infty$ the asymptotic variance of the corresponding estimator is infinite. In contrast, the convention in the spectral density literature is to normalize kernels for which $k(x) = 0$ for all $|x|$ sufficiently large such that $k(x) = 0$ for all $|x| > 1$, to normalize kernels for which the spectral window generator $K(\lambda)$ (defined in the text below) is 0 for all $|\lambda|$ sufficiently large such that $K(\lambda) = 0$ for all $|\lambda| > \pi$, and to leave the normalization of all other kernels unspecified.

[6] The lag truncation parameters of White (1984), White and Domowitz (1984), Newey and West (1987), and Gallant and White (1988), viz., ℓ, ℓ, m, and m_n respectively, are equal to M_T-1 in our notation when M_T is an integer. The aforementioned authors consider only integer valued lag truncation parameters, but there is no reason to restrict the estimators in this way and our formulae below for optimal M_T values yield real valued parameters.

For example, Newey and West (1987) define their weights as $1 - j/(m+1)$ for $j \le m$ and 0 otherwise, where m is an integer. In our notation, their weights are $1 - j/M_T$ for $j \le M_T$ and 0 otherwise, where M_T is real-valued. If M_T is an integer, then these weights are equivalent when $M_T = m+1$.

[7] The ARMA(1,1) model is parameterized as $b'V_t = \rho b'V_{t-1} + \varepsilon_t + \psi\varepsilon_{t-1}$.

[8]The MA(m) model is parametrized as $b_\ell' V_t = \sum_{u=0}^{m} \psi_{\ell u}\varepsilon_{\ell t-u}$, where $\psi_{\ell 0} = 1$ and $Var(\varepsilon_{\ell t}) = \sigma_\ell^2$.

[9]Several univariate approximating parametric models for $\{b_\ell' V_t\}$, $\ell = 1, \ldots, r$, are used here rather than a single approximate parametric model for the vectors $\{V_t\}$. This approach has the advantage that it reduces considerably the number of parameters to be estimated. It has the disadvantage that its use requires the loss function used in the determination of the optimal bandwidth parameter (see equation (5.1)) to be a function only of the loss involved in estimating the main diagonal elements of J_{TP} rather than all the elements of J_{TP}.

[10]The programming for the Monte Carlo results was done by Chris Monahan. The computations were carried out on the twenty-plus IBM-AT PC's at the Yale University Statistics Laboratory using the GAUSS normal random number generator.

[11]The truncated estimator $\hat{\rho}$, rather than $\hat{\rho}_{LS}$, is used to construct PAR because we do not want the performance of PAR to be dominated by a few observations for which $\hat{\rho}_{LS}$ is near or greater than one. Since $\hat{\rho}_{LS}$ has a large downward bias when ρ is large (say .9 or .95), the truncation at .97 does not affect many observations even when ρ is large.

[12]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, $X'X = TI_5$.

[13]Footnote 8 with $m = 1$ gives the parameterization of the MA(1) model used here.

REFERENCES

Andrews, D. W. K. and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," Review of Economic Studies, 55, forthcoming.

Betrao, K. I. and P. Bloomfield (1987): "Determining the Bandwidth of a Kernel Spectrum Estimate," Journal of Time Series Analysis, 8, 21-38.

Clarida, R. H. and J. Y. Campbell (1987): "The Term Structure of Euromarket Interest Rates: An Empirical Investigation," Journal of Monetary Economics, 19, 25-44.

Davidson, R. and J. G. MacKinnon (1981): "Efficient Estimation of Tail-area Probabilities in Sampling Experiments," Economics Letters, 8, 73-77.

Deheuvels, P. (1977): "Estimation Non-parametrique de la Densite par Histogrammes Generalises," Revue de la Statistique Appliquee, 25, 5-42.

Eicker, F. (1967): "Limit Theorems for Regressions with Unequal and Dependent Errors," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. Berkeley: University of California Press, pp. 59-82.

Epanechnikov, V. A. (1969): "Non-parametric Estimation of a Multivariate Probability Density," Theory of Probability and Its Applications, 14, 153-158.

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

Gallant, A. R. and H. White (1988): A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. New York: Basil Blackwell.

Hall, P. and J. S. Marron (1987a): "Extent to Which Least-squares Cross-validation Minimizes Integrated Square Error in Nonparametric Density Estimation," Probability Theory and Related Fields, 71, 567-581.

__________ (1987b): "On the Amount of Noice Inherent in Bandwidth Selection for a Kernel Density Estimator," Annals of Statistics, 15, 163-181.

__________ (1987c): "Estimation of Integrated Squared Density Derivatives," North Carolina Institute of Statistics, Mimeo Series #1720, Chapel Hill.

Hannan, E. J. (1970): Multiple Time Series. New York: Wiley.

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

Hansen, L. P. and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," Journal of Political Economy, 88, 829-853.

Hardouvelis, G. A. (1988): "The Predictive Power of the Term Structure During Recent Monetary Regimes," Journal of Finance, 43, forthcoming.

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.

Mishkin, F. S. (1987): "What Does the Term Structure Tell Us about Inflation?" unpublished manuscript, Graduate School of Business Administration, Columbia University, New York.

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.

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

Phillips, P. C. B. and S. Ouliaris (1988): "Testing for Cointegration Using Principal Components Methods," Journal of Economic Dynamics and Control, 12, forthcoming.

Priestley, M. B. (1962): "Basic Considerations in the Estimation of Spectra," Technometrics, 4, 551-564.

__________ (1981): Spectral Analysis and Time Series, Volumes I and II. New York: Academic Press.

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

Sacks, J. and D. Ylvisacker (1981): "Asymptotically Optimum Kernels for Density Estimation at a Point," Annals of Statistics, 9, 334-346.

Sheather, S. J. (1986): "An Improved Data-based Algorithm for Choosing the Window Width When Estimating the Density at a Point," Computational Statistics and Data Analysis, 4, 61-65.

Stone, C. J. (1984): "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," Annals of Statistics, 12, 1285-1297.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," Econometrica, 48, 817-838.

__________ (1984): Asymptotic Theory for Econometricians. New York: Academic Press.

White, H. and I. Domowitz (1984): "Nonlinear Regression with Dependent Observations," Econometrica, 52, 143-161.

TABLE 1

Asymptotically Optimal Lag Truncation/Bandwidth Values for the Newey-West/Bartlett, Gallant/Parzen, Tukey-Hanning, and QS Estimators for AR(1) $\{V_t\}$ Processes with Parameter ρ

	M_T^* for Newey-West/Bartlett Estimator						M_T^* for Gallant/Parzen Estimator					
T \ ν / ρ	.2 / .04	.3 / .09	.5 / .25	.7 / .49	.9 / .81	.95 / .90	.2 / .04	.3 / .09	.5 / .25	.7 / .49	.9 / .81	.95 / .90
32	.7	1.2	2.4	4.3	10.2	16.6	2.0	2.9	5.1	9.0	24.4	43.4
64	.9	1.5	3.0	5.4	12.9	20.9	2.3	3.3	5.8	10.4	28.0	49.9
128	1.1	1.8	3.8	6.8	16.2	26.3	2.6	3.8	6.7	11.9	32.2	57.3
256	1.4	2.3	4.8	8.6	20.4	33.1	3.0	4.4	7.7	13.7	36.9	65.8
512	1.7	2.9	6.0	10.9	25.7	41.7	3.5	5.0	8.8	15.8	42.4	75.6
1,024	2.1	3.7	7.6	13.7	32.4	52.6	4.0	5.8	10.2	18.1	48.7	86.8

	M_T^* for Tukey-Hanning Estimator						S_T^* for Quadratic Spectral Estimator					
T \ ν / ρ	.2 / .04	.3 / .09	.5 / .25	.7 / .49	.9 / .81	.95 / .90	.2 / .04	.3 / .09	.5 / .25	.7 / .49	.9 / .81	.95 / .90
32	1.3	1.9	3.3	5.9	16.0	28.5	1.0	1.4	2.5	4.5	12.1	21.6
64	1.5	2.2	3.8	6.8	18.4	32.7	1.1	1.6	2.9	5.2	13.9	24.8
128	1.7	2.5	4.4	7.8	21.1	37.6	1.3	1.9	3.3	5.9	16.0	28.5
256	2.0	2.9	5.0	9.0	24.2	43.2	1.5	2.2	3.8	6.8	18.4	32.7
512	2.3	3.3	5.8	10.3	27.8	49.6	1.7	2.5	4.4	7.8	21.1	37.5
1,024	2.6	3.8	6.7	11.9	32.0	57.0	2.0	2.9	5.0	9.0	24.2	43.1

TABLE 2

Ratio of MSE of White/truncated, Newey-West/Bartlett, Gallant/Parzen, and Tukey-Hanning Estimators to MSE of QS Estimator Using Finite Sample Optimal S_T Values - T = 128

Model	Estimator	ρ							
		.0	.3	.5	.7	.9	.95	-.3	-.5
AR(1)-HOMO	WH/TR	1.00	1.09	.93	.93	.93	.97	1.09	.94
	NW/BT	1.00	1.00	1.05	1.09	1.07	1.04	1.01	1.05
	GAL/PR	1.00	1.01	1.01	1.02	1.01	1.01	1.01	1.01
	TH	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.00
AR(1)-HET1	WH/TR	1.00	1.03	.98	.97	.97	.98	1.02	1.13
	NW/BT	1.00	1.00	1.02	1.04	1.03	1.02	1.02	1.13
	GAL/PR	1.00	1.00	1.01	1.01	1.01	1.01	1.02	1.13
	TH	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.13
AR(1)-HET2	WH/TR	1.00	1.00	1.07	.98	.96	.98	1.00	1.09
	NW/BT	1.00	1.00	1.00	1.03	1.04	1.03	1.00	1.00
	GAL/PR	1.00	1.00	1.00	1.00	1.01	1.01	1.00	1.00
	TH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Model	Estimator	ψ					
		.3	.5	.7	.99	-.3	-.7
MA(1)-HOMO	WH/TR	1.04	1.02	.99	.99	1.02	.98
	NW/BT	.99	.99	1.04	1.05	.99	1.02
	GAL/PR	.99	.99	1.01	1.02	.99	1.00
	TH	.99	.97	1.00	1.00	.99	.99

TABLE 3

Ratio of MSE of QS Estimator Using
First Order Asymptotically Optimal S_T Value, S_T^*, to
MSE of QS Estimator Using Finite Sample Optimal S_T Value

Model	T	ρ							
		.0	.3	.5	.7	.9	.95	-.3	-.5
AR(1)-HOMO	64	1.00	1.05	1.04	1.03	1.07	1.11	1.06	1.06
	128	1.00	1.05	1.05	1.02	1.04	1.07	1.07	1.03
	256	1.00	1.03	1.01	1.03	1.03	1.04	1.04	1.02
AR(1)-HET1	64	1.00	1.01	1.03	1.01	1.03	1.04	1.23	2.83
	128	1.00	1.01	1.00	1.00	1.02	1.03	1.42	3.97
	256	1.00	1.00	1.00	1.00	1.01	1.02	1.80	5.04
AR(1)-HET2	64	1.00	1.06	1.07	1.03	1.08	1.15	1.11	1.15
	128	1.00	1.12	1.13	1.03	1.08	1.08	1.09	1.14
	256	1.00	1.18	1.05	1.11	1.03	1.05	1.10	1.15
		ψ							
		.3	.5	.7	.99	-.3	-.7		
MA(1)-HOMO	64	1.01	1.01	1.06	1.10	1.01	1.08		
	128	1.01	1.10	1.17	1.16	1.00	1.12		
	256	1.06	1.22	1.24	1.23	1.05	1.29		

TABLE 4

Ratio of MSE of QS Estimator Using Automatic S_T Value, $\hat{S}_T$ to MSE of QS Estimator Using Finite Sample Optimal S_T Value

Model	T	ρ = .0	.3	.5	.7	.9	.95	-.3	-.5
AR(1)-HOMO	64 128 256	1.09 1.05 1.06	1.16 1.14 1.10	1.07 1.14 1.05	1.02 1.05 1.06	1.01 1.01 1.03	1.01 1.01 1.01	1.17 1.23 1.14	1.09 1.12 1.07
AR(1)-HET1	64 128 256	1.12 1.10 1.01	1.02 1.02 1.03	1.00 1.02 1.03	1.01 1.02 1.04	1.01 1.01 1.03	1.01 1.01 1.01	1.45 1.68 1.93	3.05 4.18 5.17
AR(1)-HET2	64 128 256	1.06 1.05 1.07	1.13 1.16 1.23	1.07 1.17 1.07	1.02 1.04 1.12	1.01 1.03 1.01	1.01 1.01 1.02	1.20 1.18 1.22	1.34 1.10 1.22

Model	T	ψ = .3	.5	.7	.99	-.3	-.7
MA(1)-HOMO	64 128 256	1.15 1.02 1.05	1.05 1.16 1.21	1.12 1.17 1.28	1.14 1.32 1.47	1.15 1.11 1.06	1.23 1.21 1.29

TABLE 5

Bias, Variance, and MSE of QS Estimator with Automatic S_T Value, $\hat{S}_T$, and True Confidence Levels for Nominal 99%, 95%, and 90% Confidence Intervals Constructed Using the QS Estimator with Automatic S_T Value for the AR(1)-HOMO Model - T = 128

ρ	Value of Estimand	Estimator	Bias	Variance	MSE	99%	95%	90%
0.0	1.00	QS	-.050	.045	.047	98.2	93.9	88.0
		INID	-.048	.043	.045	98.1	94.5	89.4
		IID	.0040	.016	.016	98.5	94.5	89.4
		PAR	.0045	.017	.017	98.5	94.5	89.5
.3	1.18	QS	-.16	.081	.11	98.1	92.9	85.5
		INID	-.26	.038	.10	97.6	90.8	84.6
		IID	-.20	.020	.059	98.2	92.7	85.8
		PAR	-.043	.038	.040	98.5	94.8	88.3
.5	1.59	QS	-.33	.22	.33	96.6	90.3	83.7
		INID	-.68	.048	.51	93.3	85.7	76.9
		IID	-.63	.025	.42	93.9	87.3	79.0
		PAR	-.11	.13	.14	98.8	93.2	87.1
.7	2.62	QS	-.87	.71	1.47	94.3	87.4	79.6
		INID	-1.78	.055	3.22	84.9	73.3	64.0
		IID	-1.70	.038	2.94	87.1	75.5	66.6
		PAR	-.42	.57	.74	97.4	92.3	84.8
.9	6.37	QS	-3.96	2.89	18.5	85.2	74.7	66.5
		INID	-5.75	.076	33.2	56.2	44.8	37.6
		IID	-5.68	.072	32.4	60.8	47.7	39.6
		PAR	-2.97	3.93	12.8	90.8	82.8	75.1
.95	8.67	QS	-6.60	2.90	46.5	75.5	63.6	55.6
		INID	-8.17	.049	67.9	45.6	34.4	28.1
		IID	-5.51	.055	66.8	49.4	37.0	30.5
		PAR	-5.51	5.26	35.6	86.3	73.0	63.7

TABLE 6

Bias and MSE of QS Estimator with Automatic S_T Value, $\hat{S}_T$, and
True Confidence Level of Nominal 95% Confidence Interval
Constructed Using the QS Estimator with Automatic S_T Value
for the AR(1)-HET1, AR(1)-HET2, and MA(1)-HOMO Models – T = 128

Model	Esti-mator	Bias	MSE	95%	Model	Esti-mator	Bias	MSE	95%
AR(1)-HET1	QS	-.32	1.35	92.8	AR(1)-HET1	QS	-.99	2.8	90.3
$\rho = 0.0$	INID	-.33	1.23	92.9	$\rho = .3$	INID	-1.2	2.7	89.3
(2.94)*	IID	-1.95	3.86	75.4	(3.87)	IID	-2.9	8.3	69.2
	PAR	-1.95	3.86	75.0		PAR	-2.8	7.8	71.0
AR(1)-HET1	QS	-2.1	7.1	85.3	AR(1)-HET1	QS	-18.	348.	60.5
$\rho = .5$	INID	-2.9	9.2	80.8	$\rho = .9$	INID	-22.	478.	38.8
(5.34)	IID	-4.4	19.4	60.4	(23.4)	IID	-23.	515.	27.7
	PAR	-4.1	16.8	67.1		PAR	-21.	442.	46.2
AR(1)-HET2	QS	-.15	.34	91.5	AR(1)-HET2	QS	-.23	.59	91.0
$\rho = 0.0$	INID	-.15	.32	91.6	$\rho = .3$	INID	-.32	.51	90.4
(1.47)	IID	-.49	.29	88.6	(1.67)	IID	-.70	.54	86.7
	PAR	-.49	.29	88.5		PAR	-.61	.44	87.3
AR(1)-HET2	QS	-.51	1.11	91.4	AR(1)-HET2	QS	-4.5	25.8	71.7
$\rho = .5$	INID	-.84	1.08	87.8	$\rho = .9$	INID	-6.3	40.2	47.7
(2.13)	IID	-1.18	1.47	82.1	(7.18)	IID	-6.5	41.9	45.5
	PAR	-.87	.94	87.6		PAR	-4.7	26.1	71.9
MA(1)-HOMO	QS	-.25	.17	91.6	MA(1)-HOMO	QS	-.24	.26	91.8
$\psi = .5$	INID	-.38	.19	90.0	$\psi = .99$	INID	-.56	.36	87.8
(1.30)	IID	-.33	.13	90.9	(1.48)	IID	-.50	.28	89.4
	PAR	-.06	.056	94.5		PAR	-.079	.078	94.4

*The numbers in parentheses in columns 1 and 6 of the table are the values of estimand.

TABLE 7

Bias and MSE of QS Estimator with Automatic S_T Value, $\hat{S}_T$, and True Confidence Levels of Nominal 95% Confidence Interval Constructed Using the QS Estimator with Automatic S_T Value for the AR(1)-HOMO, AR(1)-HET1, AR(1)-HET2, and MA(1)-HOMO Models - T = 256

Model	Estimator	Bias	MSE	95%	Model	Estimator	Bias	MSE	95%
AR(1)-HOMO ρ = 0.0 (1.00)*	QS INID IID PAR	-.036 -.034 -.0031 -.0032	.025 .024 .0098 .0098	93.7 93.7 94.5 94.5	AR(1)-HOMO ρ = .3 (1.19)	QS INID IID PAR	-.099 -.22 -.19 -.012	.069 .074 .046 .021	92.9 91.8 92.5 94.4
AR(1)-HOMO ρ = .5 (1.63)	QS INID IID PAR	-.25 -.70 -.65 -.071	.21 .50 .44 .076	92.0 86.1 86.6 94.7	AR(1)-HOMO ρ = .7 (2.78)	QS INID IID PAR	-.54 -1.84 -1.81 -.21	1.12 3.45 3.31 .44	90.6 74.0 74.4 94.1
AR(1)-HOMO ρ = .9 (7.88)	QS INID IID PAR	-3.5 -7.1 -7.0 -2.2	18.1 50.4 49.6 11.5	81.5 44.7 46.0 88.6	AR(1)-HOMO ρ = .95 (13.1)	QS INID IID PAR	-8.3 -12.5 -12.4 -6.3	80. 156. 155. 56.	71.8 33.1 34.6 80.6
AR(1)-HET1 ρ = .3 (3.92)	QS INID IID PAR	-.81 -1.11 -2.92 -2.80	1.73 1.89 8.54 7.91	92.1 90.4 71.9 74.3	AR(1)-HET1 ρ = .5 (5.44)	QS INID IID PAR	-1.7 -2.7 -4.5 -4.1	5.3 8.1 20.0 16.8	89.7 83.4 59.3 66.9
AR(1)-HET2 ρ = .3 (1.69)	QS INID IID PAR	-.17 -.28 -.71 -.60	.34 .31 .53 .40	92.5 91.6 85.3 87.7	AR(1)-HET2 ρ = .5 (2.21)	QS INID IID PAR	-.35 -.84 -1.24 -.86	.84 .91 1.56 .84	92.3 87.0 79.1 86.7
MA(1)-HOMO ψ = .5 (1.31)	QS INID IID PAR	-.19 -.35 -.33 -.030	.094 .15 .12 .031	92.6 90.3 90.8 93.7	MA(1)-HOMO ψ = .99 (1.49)	QS INID IID PAR	-.094 -.53 -.50 -.036	.15 .31 .26 .042	93.2 87.8 88.6 94.2

*The numbers in parentheses in columns 1 and 6 of the table are the values of estimand.

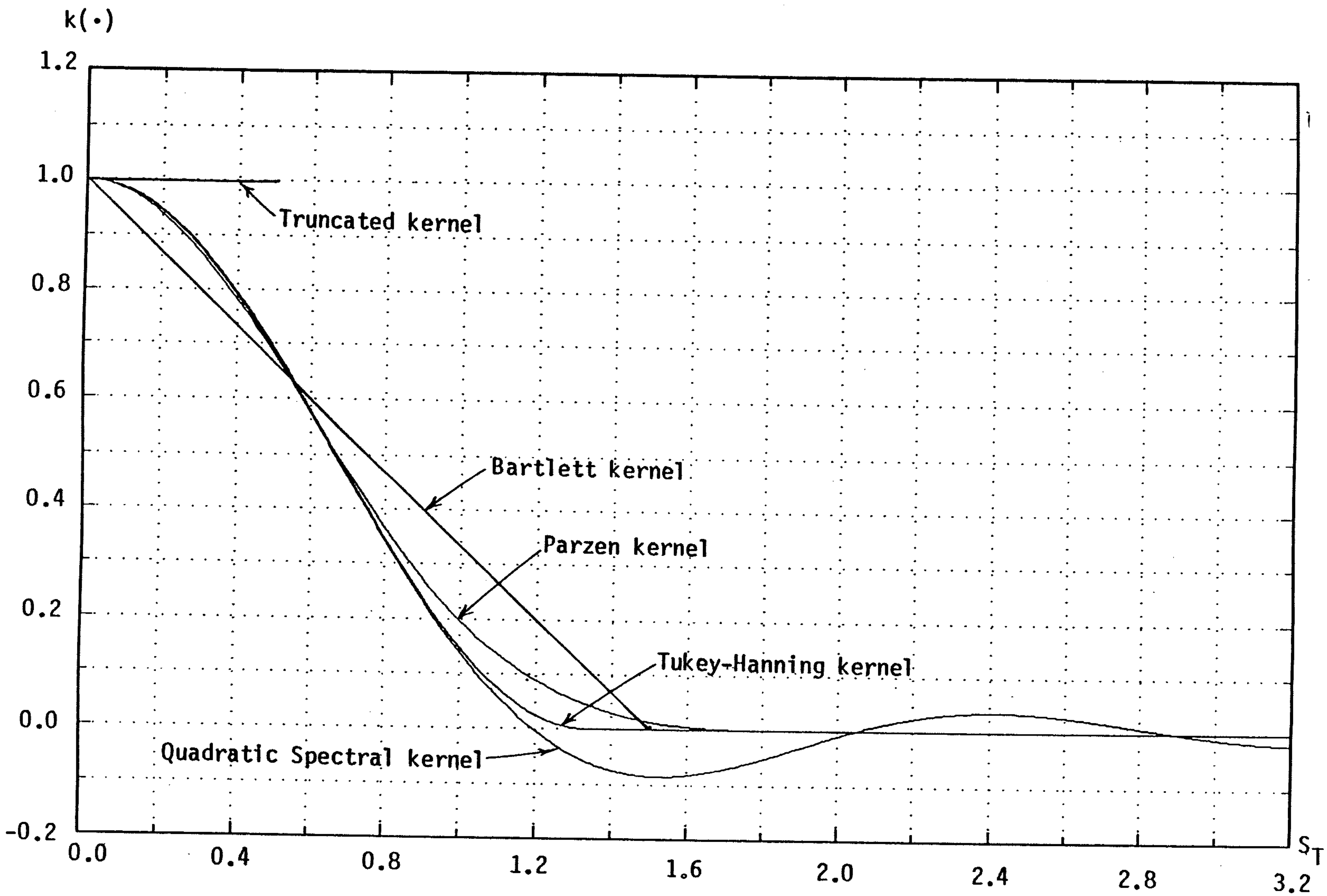

FIGURE 1 - Examples of Kernels.

FIG1

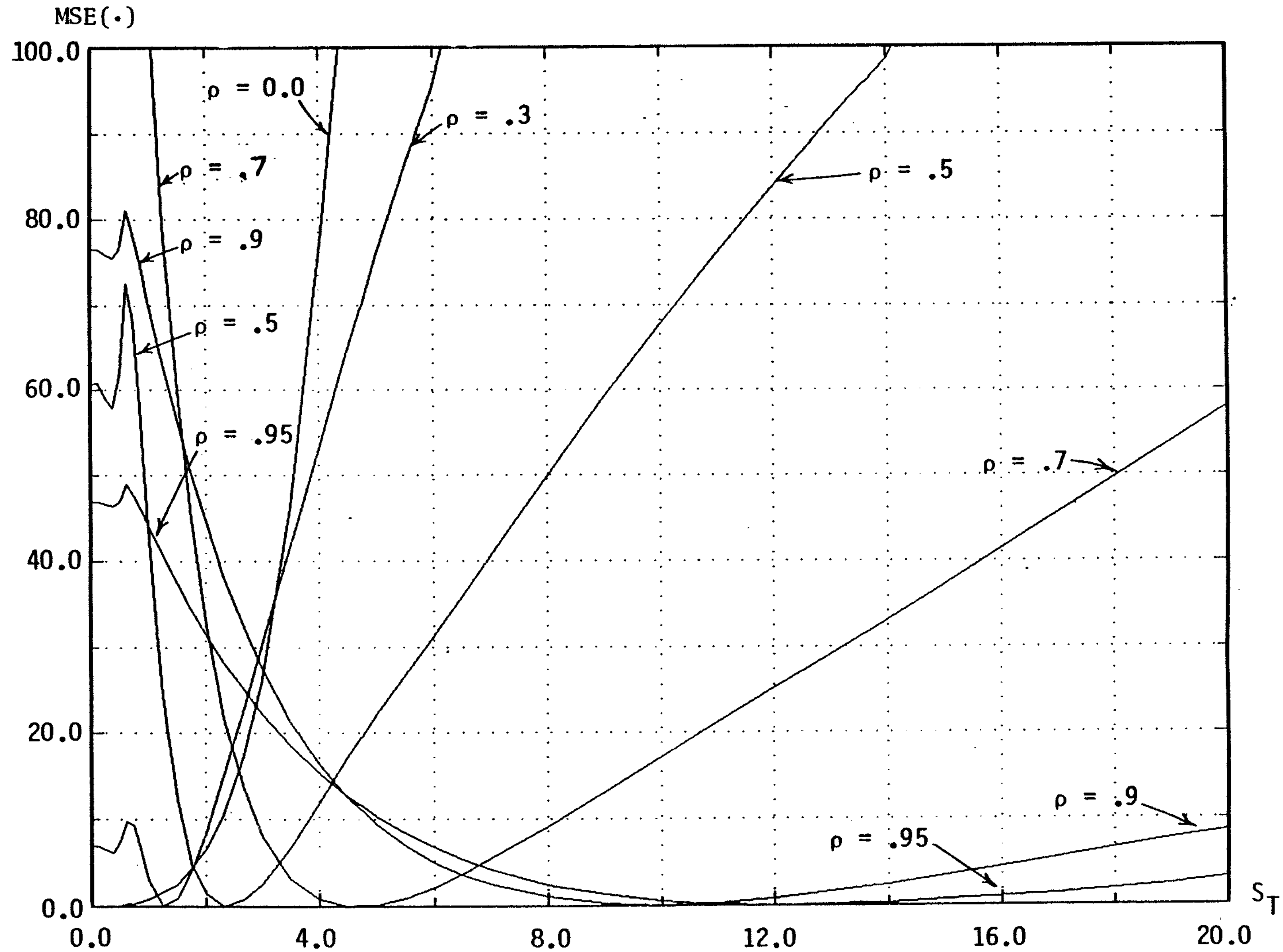

FIGURE 2 - Mean Squared Error as a Function of S_T for the QS Estimator in the AR(1)-HOMO Model with $\rho = 0.0 - .95$.