Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 504

March 1995

INFERENCES FROM PARAMETRIC AND NON-PARAMETRIC COVARIANCE MATRIX ESTIMATION PROCEDURES

Wouter J. Den Haan and Andrew T. Levin

ABSTRACT

We propose a parametric spectral estimation procedure for contructing heteroskedasticity and autocorrelation consistent (HAC) covariance matrices. We establish the consistency of this procedure under very general conditions similar to those considered in previous research. We also perform Monte Carlo simulations to evaluate the performance of this procedure in drawing reliable inferences from linear regression estimates. These simulations indicate that the parametric estimator matches, and in some cases greatly exceeds, the performance of the prewhitened kernel estimator proposed by Andrews and Monahan (1992). These simulations also illustrate the inherent limitations of non-parametric HAC covariance matrix estimation procedures, and highlight the advantages of explicitly modeling the temporal properties of the error terms.

Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures

Wouter J. Den Haan and Andrew T. Levin[1]

1. INTRODUCTION.

"The end of the story of the search for the perfect spectral estimator seems attainable if one does not think of spectral estimation as a non-parametric procedure which can be conducted independently of model identification."

Emmanuel Parzen (1983)

Over the past decade, the use of heteroskedasticity and autocorrelation consistent (HAC) covariance matrices has become relatively common in drawing inferences from parameter estimates, since in many structural economic or time-series models, the errors may have heteroskedasticity and temporal dependence of unknown form. The key step in constructing a HAC covariance matrix is to estimate the spectral density matrix at frequency zero of a vector of residual terms ux_t. In a least-squares context, for instance, ux_t consists of the vector of explanatory variables multiplied by the regression error. Thus, analytical results in the spectral density estimation literature (e.g. Parzen (1957) and Priestley (1982)) have contributed to the rapid development of new HAC covariance matrix estimation procedures (e.g., White (1984), Gallant (1987), Gallant and White (1988), Newey and West (1987, 1994), Andrews (1991) and Andrews and Monahan (1992)), but with an almost exclusive focus on non-parametric kernel-based spectral methods.[2] The AR(1)

[1]The first author is affiliated with the Economics Department at the University of California, San Diego; the second author is affiliated with the International Finance Division of the Federal Reserve Board of Governors. We appreciate comments and suggestions from Larry Christiano, Graham Elliott, Rob Engle, Neil Ericsson, Ron Gallant, Clive Granger, P.A.V.B. Swamy, Jim Stock, and Hal White. This paper represents the views of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

[2] Eichenbaum, Hansen and Singleton (1987) and West (1994) implemented covariance matrix estimators in which the residual terms follow a moving-average (MA) process of known finite order. Andrews (1991) and Andrews and Monahan (1992) briefly considered a first-order autoregressive spectral estimator, but the estimator did not correct for heteroskedasticity and did not perform very well in simulation experiments. Stock and Watson (1993) utilized AR(2) and AR(3) covariance matrix estimators in Monte Carlo experiments and in an empirical application.

prewhitening technique introduced by Andrews and Monahan (1992) has enhanced the accuracy of kernel-based procedures in a variety of Monte Carlo experiments, leading Newey and West (1994) to conclude that *'...a priority* [in econometric research] *is theoretical and empirical investigation of autoregressive or autoregressive-moving average spectral estimators'*.

In this paper we propose a parametric spectral estimation procedure for constructing HAC covariance matrices, and we establish the consistency of this procedure under very general conditions of heteroskedasticity and temporal dependence, similar to the conditions considered in previous research (e.g., Andrews (1991)). The parametric procedure utilizes standard estimation and model selection techniques in the time domain to construct a time series model for the vector of residual terms, ux_t. Next a simple transformation is used to obtain the spectral density matrix at frequency zero, and the resulting HAC covariance matrix is positive semi-definite by construction. In this paper, we focus on vector autoregressive models of ux_t, and we use the AIC criterion to select the best model. We will refer to the corresponding covariance matrix estimator as the VARHAC estimator. Our use of the AIC criterion was motivated by Shibata (1981), who showed that an asymptotically efficient estimate of the spectral density can be obtained by using AIC to select the lag order. The following arguments can be given to pursue a parametric approach.

(i) One important limitation of the kernel-based HAC covariance matrix estimators is that a single bandwidth must be used in calculating all of the elements of the spectral density matrix to ensure that the matrix is positive semi-definite. Furthermore, an arbitrary matrix must be specified to assign weights to the individual spectral matrix elements in implementing the automatic bandwidth formula. If some components of the vector of residuals, ux_t, have a relatively high degree of persistence while other components have low persistence, then imposing the same bandwidth for both sets of variables will tend to generate ill-behaved estimates of the spectral density matrix at frequency zero. Furthermore, the optimal bandwidth is not scale-independent in this case: if a particular regressor is rescaled, the persistence of that variable will receive higher weight in determining the optimal bandwidth. Thus, large measurement errors in a single explanatory variable or instrument can cause serious distortions in inference concerning the regression coefficient for all of the other variables. In contrast, the VARHAC

estimator is scale-free: a different lag order can be chosen for each component of the residual vector, because the parametric estimator of the spectral density matrix is positive semi-definite by construction.

(ii) The accuracy of kernel-based spectral estimators has been shown to be highly sensitive to the choice of bandwidth. This issue has been addressed in the literature, yielding some evidence on the comparative advantages of autoregressive and other parametric spectral density estimation methods (cf. Beamish and Priestley (1981), Kay and Marple (1981), and Parzen (1983)).[3] In particular, a larger bandwidth reduces the bias and increases the variance of the estimated spectral density. Using an asymptotic truncated mean squared error criterion, Andrews (1991) derived an optimal bandwidth analytically as a function of the true spectral density at frequency zero and of other parameters of the true data generating process. To obtain a feasible estimator of the "plug in" or "automatic" bandwidth, the practitioner must choose and estimate a time series model for the vector of residuals, ux_t. However, if the time series model is misspecified, the estimator may have relatively poor properties in finite samples. Of course, one could use a formal model selection procedure to calculate the optimal bandwidth. But if one follows this approach, one can simply use the spectral density at frequency zero implied by the best parametric model, rather than plugging this estimate into the automatic bandwidth formula to obtain a kernel-based spectral estimate.

(iii) Andrews and Monahan (1992) proposed that kernel-based methods be augmented by using an autoregressive filter of arbitrary order to prewhiten the regression residuals, and considered a first-order autoregressive filter in their Monte Carlo experiments. The AR(1) filter has provided improved inference properties in many Monte Carlo simulation experiments, some of which have considered data generating processes resembling actual economic time series (cf. Andrews and Monahan (1992), Newey and West (1994), Christiano and Den Haan (1994), and Burnside and Eichenbaum (1994)). Such first-order prewhitening is a special case of the VARHAC estimator, in which

[3] Autoregressive spectral estimation has also been used in formulating unit root test statistics (Fuller 1976), and simulation evidence suggests that this approach has superior size properties in comparison with unit root test statistics based on non-parametric spectral estimators (Schwert (1987)). The comparative advantage of the AR spectral approach in the unit root context has been analyzed recently by Perron and Ng (1994).

the autoregressive order is chosen by a data-dependent model selection criterion. Furthermore, when a sufficiently high autoregressive order is chosen to remove all serial correlation, no additional benefits can be obtained by applying a kernel-based technique to the prewhitened data.

(iv) We perform simulation experiments to evaluate the finite sample performance of the VARHAC estimator in generating accurate confidence intervals for linear regression coefficients. The Monte Carlo simulations indicate that the VARHAC estimator matches, and in some cases, greatly exceeds, the performance of the prewhitened quadratic-spectral (QS-PW) estimator proposed by Andrews and Monahan (1992). This may seem to contradict the Monte Carlo simulation results in Andrews (1991) and Andrews and Monahan (1992), in which their parametric estimator (PARA) performed poorly compared with non-parametric kernel-based estimation techniques. However, the PARA estimator imposed an AR(1) specification and failed to correct for heteroskedasticity. In contrast, the VARHAC estimator uses the AIC criterion to select the autoregressive order, and does correct for heteroskedasticity.

The remainder of this paper is organized as follows: Section 2 provides a step-by-step outline of the VARHAC covariance matrix estimation procedure, and establishes its asymptotic properties. Section 3 reviews the methodology behind kernel-based HAC estimators. Section 4 analyzes the inherent difficulties in implementing kernel-based HAC estimators, and outlines the relative advantages of the VARHAC procedure. Section 5 reports Monte Carlo simulation evidence comparing the performance of the VARHAC and prewhitened kernel-based HAC covariance matrix estimators. Section 6 concludes by outlining areas for further research.

2. A PARAMETRIC COVARIANCE MATRIX ESTIMATOR.

2.1 The VARHAC procedure.

In many estimation problems, a parameter estimate $\hat{\psi}_T$ for an $N\times 1$ parameter vector ψ_0 is obtained from the sample analog of a set of moment conditions, such as

$$E\, ux_t(\psi_0) = 0. \tag{2.1}$$

We will refer to ux_t as the vector of residual terms, although this terminology may not always be the best in each context. Under regularity conditions, the parameter $\hat{\psi}_T$ has the following limiting distribution

$$(2.2) \qquad [\ D\ S\ D\ ]^{-1/2}\ T^{1/2}\ (\hat{\psi}_T - \psi_0) \ \rightarrow\ N(0, I_N)$$

as the sample size $T \rightarrow \infty$, where

$$(2.3) \qquad S \ = \ \sum_{j=-\infty}^{\infty} \ C_j,$$

$$(2.4) \qquad C_j \ = \ E \ \ ux_t(\psi_0)\ ux_{t-j}(\psi_0)',$$

and

$$(2.5) \qquad D \ = \ E \left[\frac{\partial ux(\psi)}{\partial \psi'} \Big|_{\psi=\psi_0} \right].$$

Also, I_N is the $N \times N$ identity matrix. Usually D can be consistently estimated by its sample analog $\hat{D}_T(\hat{\psi}_T)$. The matrix S is the spectral density at frequency zero of the process $ux_t(\psi)$. As discussed in Section 3 below, non-parametric estimators of S construct a weighted average of the estimated autocovariances $\hat{C}_j$, where the weights and the number of autocorrelations depend on the particular kernel and bandwidth selection procedure.

The parametric procedure constructs the spectral estimator $\hat{S}_T(\hat{\psi}_T)$ by estimating a vector ARMA representation of $ux_t(\hat{\psi}_T)$ and uses an information criterion to select the optimal lag order for each equation in the ARMA representation. In this paper, we focus on vector autoregressive representations, and we use Akaike's (1973) information criterion to select the optimal lag order for each equation in the ARMA representation. One main advantage of AR representations is computational speed, which is important for Monte Carlo studies using vector processes. The AIC is easy to calculate, and Shibata (1981) demonstrated that the AIC yields an asymptotically efficient estimate of the spectral density. We will refer to the parametric estimator based on the vector autoregressive representation and the AIC model selection criterion as the VARHAC estimator.

Step 1. Lag order selection for each VAR equation. For the n^{th} element ux_{nt} of the vector $ux_t(\hat{\psi}_T)$ $(n = 1,\ldots,N)$ and for each lag order $\kappa = 1,\ldots,\bar{K}$, the following model is estimated by ordinary least squares:

$$(2.6) \qquad ux_{nt} = \sum_{j=1}^{N} \sum_{k=1}^{\kappa} \hat{\alpha}_{njk}(\kappa)\, ux_{n,t-k} + \hat{e}_{nt}(\kappa) \quad \text{for } t = \bar{K}+1,\ldots,T.$$

For lag order 0, we set $\hat{e}_{nt}(\kappa) \equiv ux_{nt}$. Below we will discuss the choice of the maximum lag order, $\bar{K}$, that one wants to consider. Equation (2.6) represents the regression of each component of ux_t on its own lags and the lags of the other components. Then the value of the AIC criterion is calculated for each lag order $\kappa = 0,\ldots,\bar{K}$.

$$(2.7) \qquad AIC(\kappa;n) = \log\left[(1/T) \sum_{t=\bar{K}+1}^{T} \hat{e}_{nt}^2(\kappa)\right] + 2\kappa N / T.$$

For each element of $ux_t(\hat{\psi}_T)$ the optimal lag order κ_n is chosen as the value of κ that minimizes $AIC(\kappa;n)$.

Step 2. Estimation of innovation covariance matrix. Using the results of Step 1, the restricted VAR can be expressed as:

$$(2.8) \qquad ux_t(\hat{\psi}_T) = \sum_{k=1}^{\bar{K}} A_k\, ux_{t-k}(\hat{\psi}_T) + \tilde{e}_t$$

where A_k is an $N\times N$ matrix that contains zeros and the estimated coefficients $\hat{\alpha}_{njk}(\kappa_n)$[4], and $\tilde{e}_t$ is an $n\times 1$ vector with typical element $\hat{e}_{nt}(\kappa_n)$. The innovation covariance matrix $\tilde{\Sigma}_T$ is estimated as follows:

$$(2.9) \qquad \tilde{\Sigma}_T = \frac{1}{T} \sum_{t=\bar{K}+1}^{T} \tilde{e}_t\, \tilde{e}_t'$$

Alternatively, seemingly unrelated regression (SUR) methods could be used to obtain joint estimates of the restricted VAR parameters and the innovation covariance matrix, which would yield more efficient parameter estimates if

[4] The (n,j) element of $\hat{A}_k$ is equal to zero if $k > \kappa_n$, and it is equal to $\hat{\alpha}_{njk}(\kappa_n)$ if $k \le \kappa_n$.

the innovation covariance matrix contains significant off-diagonal elements[5].

Step 3: Estimation of HAC covariance matrix. Using the results of Step 2, the spectral density matrix at frequency zero is estimated with

$$\hat{S}_T(\hat{\psi}_T) = [\, I_N - \sum_{k=1}^{\bar{K}} A_k \,]^{-1} \, \tilde{\Sigma}_T \, [\, I_N - \sum_{k=1}^{\bar{K}} A_k' \,]^{-1} \tag{2.10}$$

Finally, the VARHAC covariance matrix estimator is defined by

$$\hat{V}_T(\hat{\psi}_T) = \hat{D}_T(\hat{\psi}_T)\hat{S}_T(\hat{\psi}_T)\hat{D}_T(\hat{\psi}_T)' \tag{2.11}$$

In case one wants to include MA-terms in the regressions, then the spectral density matrix at frequency zero is estimated with

$$\begin{aligned} \hat{S}_T(\hat{\psi}_T) = {} & [I_N - \sum_{k=1}^{\bar{K}} \bar{A}_k]^{-1} \; [I_N + \sum_{k=1}^{\bar{K}} \bar{B}_k] \; \tilde{\Sigma}_T \\ & \times \; [I_N + \sum_{k=1}^{\bar{K}} \bar{B}_k'] \; [I_N - \sum_{k1}^{\bar{K}} \bar{A}_k']^{-1}, \end{aligned} \tag{2.12}$$

where $\bar{A}_k$, and $\bar{B}_K$ contain respectively the AR and the MA coefficients for the model chosen by the AIC criteria.

2.2 Relation to the prewhitening method.

Andrews and Monahan (1992) proposed the use of an AR(1) filter to prewhiten the vector $ux_t(\hat{\psi}_T)$, before using the non-parametric covariance matrix estimator. This suggestion has turned out to be very successful in improving the small sample behavior of covariance matrix estimators. However, the first-order prewhitening procedure does not use the data to select the lag-order κ_n but simply sets the lag-order equal to one for all elements of $ux_t(\hat{\psi}_T)$. Thus, this procedure may be viewed as a restrictive version of the VARHAC procedure. The AR components in the Monte Carlo experiments in Andrews and Monahan are at most first order, but in a

[5] Efficiency gains can also be achieved in small samples by reestimating the equations using observations before $\bar{K}$, whenever possible.

practical application the AR component is unknown *a priori* and frequently appears to be of a higher order.[6]

2.3 Relation to a previous parametric estimator.

In a linear regression framework the vector ux_t is equal to X_tU_t, where U_t is the regression residual and X_t is the vector of explanatory variables. The key step in calculating the standard errors is to estimate Z_t, where Z_t is defined as follows:

$$(2.13) \qquad Z_t = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E(U_s X_s U_t X_t' | X).$$

Under the assumption that the errors are homoskedastic, Z_t is equal to

$$(2.14) \qquad Z_t = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E(U_s U_t | X)\ E(X_s X_t' | X)\ .$$

Andrews (1991) and Andrews and Monahan (1992) constructed a simple parametric estimator by imposing homoskedasticity and making use of Equation (2.14). In contrast, the VARHAC estimator does not impose homoskedasticity, because this procedure analyses the properties of the product ux_t rather than its individual components.

2.4 Consistency of the VARHAC procedure.

The asymptotic properties of autoregressive spectral estimators have been analyzed extensively in the literature. Akaike (1969) proved the consistency of the AR spectral estimator under the assumptions that the true d.g.p. is an autoregressive process of known finite order and the innovations are i.i.d. with finite fourth moments. Berk (1974) extended this result to the case where the data are generated by an autoregressive process of unknown and possibly infinite order, by allowing the lag order to increase at the rate $o(T^{1/3})$ as the sample grows arbitrarily large. With the additional assumption of Gaussian innovations, Shibata (1981) demonstrated that the use

[6] See Christiano and Den Haan (1994) for detailed discussion of such examples.

of AIC to choose the actual lag order yields an asymptotically efficient estimate of the spectral density. Finally, Hannan and Kavalieris (1983, 1984) showed that these consistency and efficiency results could be derived under much more general assumptions, which we adopt in analyzing the asymptotic properties of the VARHAC estimator.

Consider a zero-mean, fourth-order stationary, purely non-deterministic, vector stochastic process $ux_t(\psi_0)$ of N elements. By the Wold representation theorem, we can express $ux_t(\psi_0)$ as follows:

(a1) $ux_t(\psi_0) = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$ where $\theta(0) \equiv 1$, $\sum_{j=0}^{\infty} \theta_j\theta_j' < \infty$,

$E\{\varepsilon_t\} = E\{\varepsilon_t|\mathcal{F}_{t-1}\} = 0$, $E\{\varepsilon_t\varepsilon_t'\} = \Sigma$, $E\{\varepsilon_t\varepsilon_s'\} = 0$ for $s \neq t$.

The ε_t are the linear innovations and hence are measurable $\mathcal{F}(t)$, where $\mathcal{F}(t)$ is the σ-algebra determined by $\{ux_s(\psi_0),\ s \leq t\}$.

The consistency of the autoregressive spectral estimator can be shown under the following additional conditions on the partial autocorrelations and innovations:

(a2) $\det \{ \sum_{j=0}^{\infty} \theta_j z^j \} \neq 0$ for all $|z| \leq 1$

(a3) $\sum_{j=0}^{\infty} j^{1/2} |\theta_j| < \infty$

(a4) $E\{ \varepsilon_{nt}^4 \} < \infty$ for all $n = 1,\ldots,N$

(a5) $E\{ \varepsilon_t\varepsilon_t' | \mathcal{F}(-\infty) \} = \Sigma$ a.s.

(a6) Conditions (a1) through (a5) hold with $ux_t(\psi_o)$ replaced by

$$(ux_t(\psi_0)', \mathrm{vec}(\frac{\partial}{\partial\psi'}ux_t(\psi_0) - E\ \frac{\partial}{\partial\psi'}ux_t(\psi_0)\ )')'.$$

(a7) $\sup_{t\geq 1} E\ \|ux_t(\psi_0)\|^2 < \infty$

(a8) $\sup_{t\geq 1} E \sup_{\psi\in\Psi} \|(\partial/\partial\theta')ux_t(\psi_0)\|^2 < \infty$

(a9) $\sup_{t\geq 1} E \sup_{\psi\in\Psi} \| (\partial^2/\partial\psi\partial\psi')ux_{nt}(\psi_0)\|^2 < \infty \ \forall\ n = 1,\ldots N.$

(a10) $T^{1/2}\ (\hat{\psi}_T - \psi_0) = O_p(1)$

In conditions (a8) and (a9), Ψ denotes some convex neighborhood of ψ_0. Condition (a2) implies that $ux_t(\psi_0)$ has a positive definite spectral density at frequency zero, and ensures that the Wold representation of $ux_t(\psi_0)$ can be inverted into an infinite-order autoregressive representation. Condition (a3) ensures that the correlation between $ux_t(\psi_0)$ and $ux_t(\psi_0)$ vanishes at a sufficiently rapid rate as the two observations become arbitrarily far apart. Condition (a4) requires the innovations to have finite kurtosis, which rules out probability distributions with excessively heavy tails. Condition (a5) indicates that the innovation variance is purely non-deterministic; this condition rules out a linear trend in the variance, but allows for conditional heteroskedasticity (e.g., ARCH processes of the type considered by Engle (1982) and subsequent authors). Condition (a6) is needed to obtain sharp convergence results. Suppose $ux_t(\psi)$ is of the form $ux(Z_t,\psi)$ for some random variable Z_t and some measurable function $ux(\cdot,\cdot)$. Then Andrews (1991) points out that condition (a6) can be verified under reasonable assumptions. Conditions (a7) and (a8) are common conditions to obtain asymptotic normality of $T^{1/2}\ (\hat{\psi}_T - \psi_0)$. Conditions (a8) and (a9) will generally be required to ensure the consistency of $D_T(\hat{\psi}_T)$ (cf. equations (2.5) and (2.11)). Condition (a10) follows from asymptotic normality of $T^{1/2}\ (\hat{\psi}_T - \psi_0)$. Andrews (1991) assumed conditions similar to (a1), (a3) through (a10) in analyzing the consistency of kernel-based spectral estimators.[7] The following theorem establishes consistency of the VARHAC estimator $\hat{S}_T$. The proof is given in the appendix.

[7] The assumption of fourth-order stationarity can also be relaxed, along the lines of Newey and West (1987), Gallant (1987), Gallant and White (1988), and Andrews (1991). With non-stationary data, the estimated spectral density at frequency zero would converge to the corresponding sequence of population moments rather than to a constant value. However, such convergence results have limited usefulness in the context of regression analysis or GMM, for which stationarity is generally required for consistency and asymptotic normality of parameter estimates.

Theorem 1: Suppose that $\{ux_t(\psi_0)\}$ is generated under conditions (a1) to (a9), and the estimator $\hat{\psi}_T$ satisfies condition (a10). If the maximum lag order $\bar{K}_T \to \infty$ such that $\bar{K}_T = o((T/\log(T))^{1/2})$, then $\hat{S}_T \to S$ almost surely.

Although $\hat{S}_T$ is defined as the VARHAC estimator where the order is chosen by AIC, other model selection criteria like the Schwartz information criterion also deliver consistency. If the AR representation is of finite order, then consistent estimation of the spectral density requires that the chosen lag order must be at least as large as the true order in the limit. The Schwartz information criterion (BIC) satisfies this condition because it chooses the true order with probability one in the limit. AIC chooses an order larger than the true order with positive probability, but selects orders less than the true order with probability zero.[8] However, for infinite order processes, BIC, is inefficient due to its relatively large penalty term. Shibata (1981) showed that for infinite order processes AIC is asymptotically efficient[9]. Despite these asymptotic differences, AIC and BIC yield nearly indistinguishable results in the Monte Carlo experiments reported in Section 5.

3. REVIEW OF KERNEL-BASED HAC ESTIMATORS.

In this section we give a short description of the non-parametric estimators of the spectral density at frequency zero used in Andrews (1991), Andrews and Monahan (1992) and Newey & West (1994). A more detailed description can be found in Christiano and Den Haan (1994).

Recall that we are interested in estimating the spectral density at frequency zero of the $N\times1$ vector $ux_t(\psi_0)$, defined by

$$S = \sum_{j=-\infty}^{\infty} C_j \tag{3.1}$$

and,

[8] See Shibata (1976). Also important is that Shibata (1976) showed that the probability of obtaining an order K' higher than the true order decreases fast with K' if the AIC criterion is used.

[9] Other model selection criteria that have the same penalty terms as AIC in the limit will be efficient as well.

(3.2) $$C_j = E\ ux_t(\psi_0)\ ux_{t-j}(\psi_0)'.$$

The non-parametric estimators described in this section have the following form:

(3.3) $$\hat{S}_T = \sum_{j=-T+1}^{T-1} \kappa(j)\ \hat{C}_j,$$

where $\kappa(\cdot)$ is a weighting function (kernel) and,

(3.4) $$\hat{C}_j = \frac{1}{T - N} \sum_{t=j+1}^{T} ux_t(\hat{\psi})\ ux_{t-j}(\hat{\psi})', \quad j = 0,\ldots,T-1,$$

and

$$\hat{C}_j = \hat{C}'_{-j}, \quad j = -1,-2,\ldots,-T+1.$$

Andrews (1991) proposes to use the QS kernel:

$$\kappa(j) = \kappa_{QS}(j/\xi),$$

where

(3.5) $$\kappa_{QS}(x) = \frac{25}{12\pi^2x^2}\left[\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5\right],$$

with $\kappa_{QS}(0) \equiv 1$, and this kernel guarantees a positive semi-definite estimator of S. Andrews (1991) refers to the parameter ξ as the bandwidth parameter. There are many other kernels being used in the literature. For HAC estimators, the QS kernel is not as asymptotically efficient as the simple truncated kernel proposed in White (1984).[10] This kernel is defined as

(3.6) $$\kappa_{TRUNC}(x) = 1 \text{ if } |x| \leq \xi$$
$$= 0 \text{ o.w.}$$

However, the truncated kernel does not guarantee that the estimated

[10] See Andrews (1991) proposition 1b and Theorem 1c.

variance-covariance matrix is positive semi-definite. In the class of kernels that guarantee positive semi-definite estimators, the QS-kernel is optimal in the sense that it minimizes an asymptotic truncated MSE criterion.[11] However, a common conclusion from many Monte Carlo experiments is that the choice of kernel in the class of kernels that guarantee a positive semi-definite covariance matrix is usually not very important for the small sample results.[12] For this reason we will only consider the QS kernel in this paper. Calculation of the spectral density is straightforward for a given choice of the bandwidth. Unfortunately, there is overwhelming evidence that the choice of the bandwidth is very important in small samples[13] and sometimes even in large samples.[14] For the kernels discussed in the previous subsection, consistency of $\hat{S}_T$, is guaranteed if $\xi_T \to \infty$ as $T \to \infty$, with $\xi_T/T^{1/2} \to 0$ (see Andrews (1991)). Andrews (1991) and Newey and West (1994) select $\{\xi_T\}$ from this class to minimize the asymptotic expectation of

$$vec(\hat{S}_T - S)' \, W \, vec(\hat{S}_T - S), \tag{3.7}$$

where $vec(\cdot)$ denotes the vectorization operator and W is an $N^2 \times N^2$ weighting matrix. Andrews (1991) shows that the optimal choice of ξ_T for the QS kernel is equal to

$$\xi_T^* = 1.322[\alpha(2)T]^{1/5}, \text{ with} \tag{3.8}$$

$$\alpha(q) = \frac{2 \, vec(S^{(q)})' W vec(S^{(q)})}{tr[W(I+K)(S \otimes S)]}, \tag{3.9}$$

and

$$S^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \, C_j. \tag{3.10}$$

[11] See Priestly (1982) and Andrews (1991).

[12] See Andrews (1991), Newey & West (1994), Christiano and Den Haan (1994), Burnside and Eichenbaum (1994).

[13] See Newey & West (1994), Andrews (1991), Andrews and Monahan (1992), and Christiano and Den Haan (1994).

[14] See for instance Christiano and Den Haan (1994).

where $tr(\cdot)$ denotes the trace operator, I is the $n^2 \times n^2$ identity matrix, and K is the $n^2 \times n^2$ commutation matrix defined by the property, $vec(A') = K\ vec(A)$. Recall that we are interested in ξ_T only because we want to get an estimate of S. To understand the automatic bandwidth procedures proposed by Andrews (1991) and Newey and West (1994), it is important to realize that the optimal bandwidth ξ_T^* depends on knowing the true value of S. Implementation of the automatic bandwidth selection procedure thus requires an initial estimate of S, as well as an estimate of $S^{(q)}$, and a choice for W.

Andrews (1991) proposed that ξ_T^* be estimated by fitting a parametric model for $ux_t(\hat{\psi}_T)$, and then using this model to obtain estimates of S and $S^{(q)}$ using equations like (2.12). In the Monte Carlo experiments reported by Andrews (1991) and Andrews and Monahan (1992), an AR(1) model was fitted to the n^{th} element of $ux_t(\hat{\psi}_T)$. Let (ρ_n, σ_n^2) denote the first-order autoregressive and innovations variance parameters. This choice for the *dgp* of $ux_t(\hat{\psi}_T)$ requires that the matrix W assigns non-zero weights only to the diagonal elements of $S_T - S$ in (3.7). In particular, with the exception of N terms, W is composed entirely of zeros. The exceptions are the $(n-1)\times N + 1^{th}$ diagonal elements of W, which we denote ω_n, for $n = 1,\ldots,N$. With this W matrix,

$$\hat{\alpha}(q) = \frac{\sum_{n=1}^{N} \omega_n \dfrac{4\hat{\rho}_n^2\, \hat{\sigma}_n^4}{(1-\hat{\rho}_n)^8}}{\sum_{n=1}^{N} \omega_n \dfrac{\hat{\sigma}_n^4}{(1-\hat{\rho}_n)^4}} . \tag{3.11}$$

In practice, Andrews (1991) and Andrews and Monahan (1992) used $\omega_n = 1$ for all elements except the one corresponding to the regression intercept.

Newey & West (1994) considered a weighting matrix which allows all terms in $S_T - S$ to enter (3.7), except the ones corresponding to the regression intercept. To determine the data dependent bandwidth, they obtained initial estimates S and $S^{(q)}$ using the truncated kernel, κ_{TRUNC}, and a truncation parameter $l = \beta(T/100)^{2/9}$ (cf. Equation (3.6)). In this case an appropriate

value for β is determined by trying alternative values and '*..then exercising some judgment about sensitivity of results*', in the hope that the final estimate of the covariance matrix is less sensitive to β than to ξ.

4. COMPARATIVE ANALYSIS OF PARAMETRIC AND KERNEL-BASED HAC ESTIMATORS.

The accuracy of kernel-based spectral estimators has been shown to be highly sensitive to the choice of bandwidth. For a particular finite sample, the optimal value of the bandwidth is influenced by two factors. On the one hand, raising the bandwidth tends to reduce the bias of the spectral estimator by incorporating higher-order sample autocorrelations. On the other hand, fewer observations are used to estimate each higher-order sample autocorrelation, so that a higher bandwidth tends to increase the variance of the spectral estimator[15]. The fragility of spectral estimation accuracy provided an important motivation for Andrews' (1991) derivation of the optimal rate at which to raise the bandwidth as the sample grows arbitrarily large.

Of course, parametric estimators also face difficulties in extracting information about the true model using a finite sample. However, the VARHAC estimator avoids several important limitations of the kernel-based estimator. These comparative advantages of the VARHAC estimator are confirmed by the Monte Carlo experiments reported in Section 5.

4.1 Required use of a single bandwidth.

To ensure that the variance-covariance matrix is positive semi-definite, kernel-based spectral estimators require that the same bandwidth be chosen for the entire vector ux_t, regardless of the procedure used to select the bandwidth. However, imposing a single bandwidth causes problems for the accuracy of the kernel-based estimator. Moreover, the automatic bandwidth parameter is sensitive to the measurement units of the individual variables. In contrast, the VARHAC estimator allows a different lag order to be chosen for each component of ux_t, and is not affected by rescaling of the

[15] For a given sample the mean-squared error of the spectral estimator exhibits an inverted-U shaped dependence on the bandwidth, with sharp increases in MSE as the bandwidth is moved away from its optimal value (cf Andrews (1991)).

variables. These issues are documented in the Monte Carlo experiments reported in Section 5.3.

(i) Loss of accuracy. In any particular application, the pattern of serial correlation may vary across the elements of ux_t. Thus, imposing the same bandwidth for all components of ux_t can lead to poor finite-sample performance of kernel-based estimators. For example, the procedures proposed by Andrews (1991) and Andrews and Monahan (1992) select the bandwidth based on the first-order autocorrelation and innovation variance of each component (cf. equation (3.11) above). Thus, if a single component of ux_t is highly persistent and the other components are serially uncorrelated, a relatively small bandwidth will be generated, leading to imprecise estimates of the standard errors that depend on the persistent component. For linear regression models, the coefficients and standard errors have a direct correspondence to the individual components of ux_t; in this case, one could consider calculating each standard error using a different estimate of the spectral density matrix, based on a different specification of the weighting matrix W. For more general applications of GMM, however, such a procedure is not possible, even in principle, because no direct correspondence exists between the regression parameters and error components. In contrast, the parametric estimator can easily deal with all components of $ux_t(\hat{\psi}_T)$ separately, allowing for the possibility of choosing a low-order process for one component of $ux_t(\hat{\psi}_T)$ and a high-order process for another component.

(ii) Sensitivity to measurement units. Kernel-based estimators have the unfortunate property that the estimated HAC covariance matrix is sensitive to a change in the scale used to measure any particular variable.[16] For example, suppose that the n^{th} instrument in an IV estimation procedure is multiplied by 10, thereby raising the innovation variance of the n^{th} component by a factor of 100. Then this component exerts more influence in the automatic bandwidth selection procedure, causing the bandwidth to shift toward a value which is optimal for the n^{th} component, but possibly less appropriate for the

[16] Of course, as Andrews (1991) indicated, the bandwidth selection procedure is not affected by an identical rescaling of all of the variables for the particular choice of weights in which the slope coefficients are assigned equal weight, and the regression intercept is assigned zero weight.

other components of ux_t. In general, it is unclear how the practitioner should choose a weighting matrix to offset such rescaling problems. Should the weights vary with the unconditional or the innovation variances of the components? By how much and in which direction? Such difficulties arose in a recent paper by Burnside and Eichenbaum (1994), in which very large bandwidths were implied by the data in a high-dimensional empirical application. The authors determined that zeroing out three of the ω_n weights caused a large reduction in the bandwidth and resulted in dramatically different inferences concerning the estimated coefficients.

In contrast, rescaling of any variable has no effect on the results of the parametric HAC covariance matrix estimation procedure.

4.2 Lack of finite sample justification.

For a given kernel, the optimal bandwidth formula only expresses the rate at which the bandwidth should grow as a function of the sample size. Thus, this formula cannot indicate the optimal bandwidth to use in any particular finite sample. More precisely, the bandwidth $\xi^{**} = \xi^* + M$ (for any fixed integer M) meets the same asymptotic optimality criterion as the bandwidth ξ^* defined in equation (3.8) above. Unfortunately, while ξ^* and ξ^{**} may yield dramatically different results in a particular finite sample, one has no basis *a priori* upon which to choose one bandwidth over the other. This non-uniqueness property may appear similar to other uses of asymptotic optimality criteria in the literature. For example, if the OLS estimator $\hat{\beta}$ is consistent, then so is $\hat{\beta} + M/T$, for any fixed value of M. The key difference is that the OLS estimator also satisfies a sensible finite-sample estimation criterion (namely, minimizing the sum of squared residuals of the regression model), whereas current bandwidth selection procedures do not satisfy any particular finite-sample criterion.

In contrast, parametric covariance matrix estimators can make use of explicit finite-sample model selection criteria. The VARHAC procedure selects the autoregressive lag order using AIC, which is an asymptotically efficient selection criterion under the conditions previously described in Section 2. Equally important, the AIC was designed to optimize parsimony and goodness-of-fit in finite samples, so that the VARHAC estimator may be expected to work relatively well for a wide variety of data generating processes and sample sizes.

4.3 Sensitivity to arbitrary parameterization.

For a given kernel, the implementation of the optimal bandwidth formula requires an initial estimate of the spectral density at frequency zero, as well as other properties of the true data generating process (cf. equations (3.8) to 3.10 above). One might hope that the kernel-based HAC covariance matrix estimate would be relatively insensitive to the method of calculating these initial estimates, including fairly arbitrary choices about the order of parametric approximations and/or the lag truncation point. In fact, however, Christiano and den Haan (1994) documented various cases in which the results remain highly sensitive to these choices, even in samples as large as 5000 observations.

Andrews (1991) proposed a parametric approach to calculate the initial spectral density estimate. In practice, Andrews (1991) and Andrews and Monahan (1992) obtained $\hat{\alpha}(2)$ by estimating a univariate AR(1) model for each component of ux_t, and subsequent research has also adopted this approach.[17] However, the next section provides several examples in which the final results are very sensitive to this particular specification.

Newey and West (1994) used a non-parametric estimator to obtain an initial estimate of the spectral density at frequency zero. However, the Monte Carlo experiments presented by Newey and West indicated that the results can be sensitive to the arbitrary choice of a truncation parameter used in calculating the initial spectral estimate.[18]

Thus, although these data-dependent bandwidth selection procedures are frequently referred to as "automatic," in practice one needs to make non-trivial decisions about various arbitrary parameters. Christiano and Den Haan (1994) investigated the consequences of some of these choices in a range of experiments, but in general there is little guidance in the literature on the consequences of alternative choices. An applied economist who needs to perform hypothesis tests would obviously prefer to avoid having to worry about the sensitivity of an estimated standard error with respect to prewhitening order, weighting matrix, or any of the other arbitrary parameters of the bandwidth selection procedure. In contrast, the conceptual background and implementation of the parametric estimator will be familiar to

[17] See Newey & West (1994), Christiano and Den Haan (1994), and Burnside and Eichenbaum (1994).

[18] See for instance Table IV, lines 7 and 8.

most practitioners who already have experience in estimating parametric time-series models and in choosing the best model. These issues are documented in the Monte Carlo experiments reported in Section 5.4.

4.4 Sensitivity to prewhitening order.

Andrews and Monahan (1992) proposed that kernel-based methods be augmented by using an autoregressive filter of arbitrary order to prewhiten the regression residuals, and in their Monte Carlo experiments they considered a first-order autoregressive filter. The AR(1) filter has provided improved inference properties in many Monte Carlo simulation experiments, some of which have considered data generating processes resembling actual economic time series (cf. Andrews and Monahan (1992), Newey and West (1994), Christiano and Den Haan (1994), and Burnside and Eichenbaum (1994)). Such first-order prewhitening is a special case of the VARHAC estimator, in which the autoregressive order is chosen by a data-dependent model selection criterion. Furthermore, when a sufficiently high autoregressive order is chosen to remove all serial correlation, no additional benefits can be obtained by applying a kernel-based technique to the prewhitened data. These issues are documented in the Monte Carlo experiments reported in Section 5.5.

5. MONTE CARLO EXPERIMENTS.

In this section we describe the Monte Carlo results. In Section 5.1 we give an overview of the four Monte Carlo experiments. The Monte Carlo experiments are described in Sections 5.2 through 5.5.

5.1 Overview.

In the following four subsections, we report the results of Monte Carlo experiments to compare the small-sample properties of two HAC covariance matrix estimators, and to evaluate the extent to which each estimator provides accurate inferences in two-tailed tests of the significance of the regression coefficients. The first estimator is the parametric VARHAC estimator described in Section 2 above, and the second is the non-parametric QS-PW estimator studied by Andrews and Monahan (1992), which uses the quadratic spectral kernel, first-order prewhitening, and univariate AR(1) models in the automatic bandwidth selection procedure. We do not use the

eigenvalue adjustment for the estimated AR(1) coefficient in the prewhitening regression; this adjustment appears rather arbitrary, and does not affect the results reported here.[19] For each Monte Carlo experiment, we simulate 10,000 replications of 128 observations each, and report the frequency for which the test statistics exceeds the 10%, 5% or 1% critical value.

In Section 5.2, we consider the data generating processes used in the Monte Carlo experiments of Andrews and Monahan (1992), and we find that the parametric VARHAC estimator matches the small-sample performances for the QS-PW estimator quite well. Next, section 5.3 documents the advantage of the VARHAC estimator in allowing different autoregressive orders for different components of the residual vector, in contrast to the rescaling problem faced by kernel-based estimators that impose a single bandwidth on the entire residual vector. Section 5.4 highlights the pitfalls associated with using an arbitrary parametric model in automatic bandwidth selection procedures. Finally, Section 5.5 verifies the benefits derived by the VARHAC estimator in using a model selection criterion to determine the autoregressive order, or equivalently, the degree of prewhitening.

5.2 The Andrews and Monahan (1992) experiments.

Andrews and Monahan (1992) use the following experiments to investigate the small sample properties of the covariance matrix estimator. They consider several linear regression models, each with an intercept and four regressors, and the least squares (LS) estimator $\hat{\theta}_T$ for each of these models:

$$(5.1) \qquad Y_t = X_t'\theta_0 + U_t, \quad t = 1,\ldots,T \qquad \hat{\theta}_T = [\ \textstyle\sum_1^T X_tX_t'\ ]^{-1} \sum_1^T X_tY_t.$$

and,

$$(5.2) \qquad VAR(T^{1/2}(\hat{\theta}_T - \theta_0)|X) = \left[\ (\ \frac{1}{T}\textstyle\sum_1^T X_tX_t'\ )^{-1}\right.$$

[19] Newey & West (1994) also find that the results show little sensitivity to this adjustment, and they do not recommend this adjustment.

$$\times \left(\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E(U_s X_s U_t X_t' | X) \left(\frac{1}{T} \sum_{1}^{T} X_t X_t' \right)^{-1} \right]$$

The estimand of interest is the variance (conditional on $X = (X_1, \ldots, X_T)'$) of the LS estimator of the first nonconstant regressor (i.e., the second diagonal element of $VAR((T^{1/2}(\hat{\theta}_T - \theta_0)|X)$. All elements of θ_0 are equal to zero.

Andrews and Monahan (1992) describe the experiment as follows. Seven basic regression models are considered: AR(1)-HOMO, in which the errors and regressors are homoskedastic AR(1) processes; AR(1)-HET1 and AR(1)-HET2, in which the errors and regressors are AR(1) processes with multiplicative heteroskedasticity overlaid on the errors; MA(1)-HOMO, in which the errors and regressors are homoskedastic MA(1) processes; MA(1)-HET1 and MA(1)-HET2, in which the errors and regressors are MA(1) processes with multiplicative heteroskedasticity overlaid on the errors; and MA(m)-HOMO, in which the errors and regressors are homoskedastic MA(m) processes with linearly declining MA parameters. A range of different parameter values is considered for each model. Each parameter value corresponds to a different degree of autocorrelation.

The AR(1)-HOMO model consists of mutually independent errors and regressors. The errors are mean zero, homoskedastic, AR(1), stationary, 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. A new set of regressors is randomly drawn for each repetition of the experiment (to ensure that the results are not sensitive to the selection of a single, perhaps atypical, matrix of regressors). The values considered for the AR(1) parameter ρ are 0., .3, .5, .7, .9, .95, -.3, and -.5.

The AR(1)-HET1 and AR(1)-HET2 models are constructed by introducing multiplicative heteroskedasticity to the errors of the AR(1)-HOMO model. Suppose $\{x_t, \bar{U}_t: t=1,\ldots,T\}$ are the nonconstant regressors and errors generated by the AR(1)-HOMO model (where $X_t' = (1, x_t')$). Let $U_t = |x_t'\omega| \times \bar{U}_t$. Then $\{x_t, U_t: t=1,\ldots,T\}$ are the nonconstant regressors and errors for the AR(1)-HET1 and AR(1)-HET2 models when $\omega=(1,0,0,0)'$ and $\omega=(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})'$

respectively. In the AR(1)-HET1 model, the heteroskedasticity is related only to the regressor whose coefficient is being tested, 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.

The MA(1)-HOMO, MA(1)-HET1, and MA(1)-HET2 models are exactly the same as the AR(1)-HOMO, AR(1)-HET1, and AR(1)-HET2 models respectively, except that stationary MA(1) processes replace stationary AR(1) processes everywhere that the latter arise in the definitions above. The MA(1) processes have variance 1 and MA parameter ϕ (and are parameterized as $\bar{U}_t=\varepsilon_t+\phi\varepsilon_{t-1}$). The values of ϕ that are considered are .3,.5,.7,.9,.99.

The MA(m)-HOMO model is exactly the same as the AR(1)-HOMO model except that the errors and the regressors are homoskedastic, stationary MA(m) random variables with variance 1 and MA parameters $\phi_1,\ldots,\phi_m$ (where the MA(m) model is parameterized as $U_t = \varepsilon_t + \sum_{r=1}^{m}\phi_r\varepsilon_{t-r}$). The MA parameters are taken to be positive and to decline linearly to zero (i.e. $\phi_r = 1 - r/(m+1)$ for $r=1,\ldots,m$). The values of m that are considered are 3, 5, 7, 9, 12, 15.

In the Monte Carlo experiments we compare the VARHAC estimator with the QS-PW estimator. The maximum lag-order $\bar{K}$ is equal to 4. A higher or sightly lower value for $\bar{K}$ did not change the results very much since higher-order lags were selected infrequently. The QS-PW estimator uses an AR(1) to prewhiten the vector of residuals. In this regression framework, $ux_t = X_tU_t$, and an AR(1) is used for each of the components of ux_t to estimate the optimal bandwidth. The weights required to calculate the optimal bandwidth are set equal to 1 for the nonconstant regressors and 0 for the constant regressors.[20]

Tables 1 through 5 report the coverage probabilities for the Monte Carlo experiments. The important conclusion that arises from the tables is that the inference accuracy of the VARHAC estimator matches that of the QS-PW

[20] Note that in many of the Monte Carlo experiments reported by Andrews and Monahan (1992), this particular choice of weights yields more accurate inferences about the slope coefficients. In particular, in the HOMO and HET(2) experiments, the components of ux_t corresponding to the four slope coefficients have the same serial correlation properties, but the first component of ux_t, corresponding to the regression intercept, has different serial correlation properties.

estimator quite well, despite the fact that these data generating processes might be expected to favor the QS-PW. In the AR models, for example, QS-PW imposes first-order prewhitening, while the VARHAC estimator chooses the lag-order with the AIC selection criterion. We observe the biggest difference, in table 4, for the MA(1)-HET1 model in which the QS-PW outperforms the VARHAC estimator to some extent. For instance, when $\phi = 0.5$, the 99%, 95% and 90% coverage probabilities are equal to 97.8%, 92.9% and 87.5% for QS-PW, and equal to 96.5%, 90.1% and 84.0% for VARHAC. That the QS-PW has some advantages for this model is no surprise, since the VARHAC estimator uses AR models to approximate the true MA(1) processes. Obviously, the parametric estimator could be improved by allowing MA terms. Since the results are so similar for the two estimators, we refer the reader to Andrews and Monahan (1992) for a more detailed discussion of the variation of the results across models and parameter values. The VARHAC estimator greatly outperforms the PARA estimator, as expected from the analysis in Section 2.3.

5.3 The limitations of a single bandwidth.

As discussed in Section 4.1, non-parametric kernel-based estimators require the bandwidth to be the same for all elements of the vector of residuals $(ux_t(\hat{\psi}_T))$ to guarantee a positive semi-definite covariance matrix. In this section, we document the impact of this restriction using the following Monte Carlo experiment. Consider the least squares estimator for the following scalar model:

$$y_t = \alpha + \beta Z_t + \varepsilon_t, \tag{5.3}$$

with

$$\begin{aligned} (1-\rho L)\,\varepsilon_t &= e1_t, \\ X_t &= e2_t, \\ Z_t &= \lambda X_t, \end{aligned}$$

where $\alpha = \beta = 0$, $e1_t$ and $e2_t$ are i.i.d. normally distributed random variables. The parameter λ scales the explanatory variable. The unconditional variance of ε_t and X_t is equal to 1. The two elements of ux_t are ε_t and $\lambda\varepsilon_t X_t$. Thus the first element is a first-order AR process, and

the second element is serially uncorrelated. Varying the scale coefficient λ is equivalent to expressing the explanatory variable in different measurement units.

To highlight the fundamental point, we do not use the prewhitening option for the QS-PW estimator, since first-order prewhitening would make both components close to white noise. For higher order processes for ε_t, the QS-PW estimator would encounter the same limitations as those discussed here. However, the discussion would be complicated by the misspecification bias of the AR(1) coefficient in the prewhitening regression. Therefore, for clarity, we focus on the QS estimator. The VARHAC estimator is exactly as defined in Section 5.3.

The choice of a smaller bandwidth in this experiment improves the small sample behavior of the standard error for the slope coefficient, while a larger bandwidth improves the small sample accuracy of inference concerning the regression intercept. Although one could vary the weights ω_n in equation (3.11), to trade off the accuracy between the two standard errors, it may not be clear how to do this in a practical application, especially because there is not always a direct link between the components of ux_t and the individual elements of the parameter vector.

The results of the Monte Carlo are given in table 6. The results for the VARHAC estimator do not depend on the value of λ, whereas the results for the QS are highly sensitive to the value of λ. For the QS procedure, choosing a larger value of λ raises the weight on the second element of ux_t, reduces the average bandwidth chosen, and reduces the accuracy of the estimated standard error of the regression intercept. The average bandwidth across Monte Carlo replications was equal to 23.4, 2.3, and 1.7 for values of λ equal to 1, 100 and 1000, respectively. Of course, by reducing the bandwidth, a larger value of λ improves the behavior of the estimated standard error for the slope coefficient.

5.4 The limitations of arbitrary parameterizations.

As discussed in Section 4.2, the QS-PW procedures require the specification of a time series process for $ux_t(\hat{\psi}_T)$. Andrews and Monahan (1992) used an AR(1) model for each of the components of $ux_t(\hat{\psi}_T)$ in all of their Monte Carlo experiments, and subsequent papers in the literature have generally followed the same procedure in implementing the QS-PW estimator.

This subsection highlights the consequences of adopting this AR(1) assumption, when the ux_t follows a different law of motion. The data are generated by the following time-series model:

$$Y_t = \nu\, \varepsilon_t - \mu\, \varepsilon_{t-q}, \qquad q = 2,3 \tag{5.4}$$

where ε_t is an i.i.d. normally distributed random variable with zero mean and unit variance. Again we are interested in estimating the mean of Y_t,

$$\hat{\psi}_T = \frac{1}{T} \sum_1^T Y_t, \tag{5.5}$$

The important aspect of this time-series model is that the MA(1) coefficient ν is equal to zero or small, but higher-order MA coefficients are not. If ν is equal to zero, then assumption *E* in Andrews (1991) is violated, and for this case the estimated optimal bandwidth under the AR(1) assumption converges to zero, and the QS-PW estimator is not consistent. Several empirical cases suggest that such a time series process for ux_t is not unrealistic. First, Fama and French (1988) documented that for stock returns, autocorrelations are small for short horizons, but relatively large for large horizons. For instance, the average first-order autocorrelation across industries is equal to -0.03 for one-year returns, but equal to -0.34 for four-year returns. Second, Christiano and Den Haan (1994) used a *dgp* resembling that of US quarterly GNP, and found that some prewhitened residuals had a very low first-order MA coefficient, but substantial higher-order serial correlation. Finally, this type of behavior might be important in data that are not seasonally adjusted.

As shown in table 7, the VARHAC estimator clearly outperforms the QS-PW estimator in this experiment. The small sample behavior of VARHAC is excellent even for 128 observations, despite the use of a vector autoregressive process to approximate a moving average process. Furthermore, the sign of the MA coefficient does not affect the inference accuracy of the VARHAC estimator, but has a large impact on the inference accuracy of the QS-PW estimator. If the third-order MA coefficient is positive then the QS-PW estimator underestimates the amount of volatility, and consequently rejects the null hypothesis too often. In contrast, for the negative

third-order MA coefficient, the QS-PW rejects the null hypothesis too infrequently.

5.5 The limitations of arbitrary prewhitening order.

An important motivation in developing the VARHAC estimator was the success of the prewhitening procedure proposed by Andrews and Monahan (1992). However, Andrews and Monahan (1992) only considered first-order prewhitening, whereas the VARHAC estimator uses a model selection criterion to choose the order of prewhitening. The advantages of the flexibility of the VARHAC estimator in choosing higher-order prewhitening were not apparent in the Monte Carlo experiments discussed in Section 5.2, since the AR component in the vector of residuals, ux_t, was at most of order one, and the QS-PW estimator imposes first-order prewhitening. In this section, we consider the following scalar AR(2) process, and estimate the mean.

$$(5.6) \qquad Y_t = \tfrac{1}{2}\phi Y_{t-1} + \tfrac{1}{2}\phi Y_{t-2} + \varepsilon_t, \qquad \hat{\psi}_T = \frac{1}{T}\sum_1^T Y_t,$$

where the ε_t is an i.i.d. N(0,1) process. The estimand of interest is the standard error of the mean. The values we consider for ϕ are .5, .7, .9, and .95. As seen in table 8, the VARHAC estimator clearly outperforms the QS-PW estimator, even for values of ϕ as low as .5. Given the success of first-order prewhitening, it is not surprising that higher-order prewhitening is also advantageous. It is important to note, however, that the VARHAC estimator does not impose the assumption that the residuals are generated by an AR(2) process. For this experiment, a lag order of two was chosen by AIC in 67%, 77%, 78% and 78% of all replications for parameter values equal to .5, .7, .9, and .95, respectively.

6. CONCLUDING COMMENTS.

Given the importance of accurate estimation of HAC covariance matrices in applied econometrics, it is essential to compare the properties of parametric and non-parametric estimators. This paper highlights several advantages of parametric estimators like the VARHAC estimator. First, the VARHAC estimator selects a different lag order for each element of ux_t to accommodate different serial correlation properties, whereas the QS-PW estimator imposes a single bandwidth. Second, the VARHAC estimator is not

sensitive to rescaling of any of the variables. Finally, the Monte Carlo experiments indicate that the VARHAC estimator matches, and in some cases greatly exceeds, the performance of the QS-PW estimator.

Some topics for future research remain.

(i) It would be interesting to investigate the performance of a parametric ARMA estimator. Although the VARHAC covariance matrix estimator performed well for the experiments in this paper, there may be cases where the inclusion of MA terms would be advantageous.

(ii) It may also be interesting to analyze different model selection criteria. Shibata (1981) provided asymptotic justification for using AIC, but other selection criteria could have superior performance in small samples. We have found that AIC and the Schwartz information criterion yield very similar results for the experiments reported here, but larger differences may become evident in other cases. We have also explored some model selection criteria that tested whether a reduction of the lag order caused a statistically significant change in the estimated spectral density at frequency zero. However, the chosen orders turned out to be very sensitive to the maximum lag order considered, and the small sample performance was worse than that of the VARHAC estimator.

(iii) In this paper, we have used the VARHAC covariance matrix estimator to draw inferences in a linear regression model. In an over-identified GMM framework, the estimated covariance matrix also influences the parameter estimates. It would be interesting to compare the VARHAC and the QS-PW procedures in this type of estimation problem.

REFERENCES

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in *Second International Symposium on Information Theory*, B.N. Petrov and F.Csaki, eds., Akademia Kiado (Budapest), pp. 267-281.

An Hong-Zhi, Chen Zhao-Guo and E.J. Hannan, 1982, Autocorrelation, autoregresion and autoregressive approximation, *The Annals of Statistics* 10(3), p. 926-936.

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, pp. 817-858.

Andrews, D.W.K. and H-Y Chen, 1994, Approximately median-unbiased estimation of autoregressive models, *Journal of Business and Economic Statistics* 12(2), pp. 187-204.

Andrews, D.W.K. and J.C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, pp. 953-966.

Beamish, N. and M.B. Priestly, 1981, A study of autoregressive and window spectral estimation, *Applied Statistics* 30, pp. 41-58.

Berk, K.N., 1974, Consistent autoregressive spectral estimates, *The Annals of Statistics Stat.*2, pp.489-502.

Burnside, C. and M. Eichenbaum, 1994, Small sample properties of generalized method of moments based Wald tests, manuscript, University of Pittsburgh and Northwestern University.

Christiano, L.J., and Wouter J. den Haan, 1994, Small sample properties of GMM for business cycle analysis, UCSD working paper no 94-17.

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, p. 987-1008.

Fama, E.F., and K.R. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96(2), pp. 246-273.

Gallant, A.R., 1987, Nonlinear statistical models, Wiley Press (New York).

Gallant, A.R., and H. White, 1988, A unified theory of estimation and inference for nonlinear dynamic models, Basil Blackwell (New York).

Hannan, E.J., and L. Kavalieris, 1983, The convergence of autocorrelations and autoregressions, *Australian Journal of Statistics* 25(2), pp. 287-297.

Hannan, E.J., and L. Kavalieris, 1986, Regression and autoregression models, *Journal of Time Series Analysis* 7, pp. 27-49.

Kay, S.M. and S.L. Marple, 1981, Spectrum analysis: a modern perspective. Proc. IEEE 69, pp. 1380-1419.

Newey, W.K. and K.D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, pp. 703-708.

Newey, W.K. and K.D. West, 1994, Automatic lag selection in covariance matrix estimation, manuscript, MIT and University of Wisconsin.

Parzen, E., 1957, On consistent estimates of the spectrum of a stationary time series, *Ann. Math. Stat* 28, pp. 329-348.

Parzen, E., 1983, Autoregressive spectral estimation, in *Handbook of Statistics* vol 3, D.R. Brillinger and P.R. Krishnaiah, eds., Elsevier Press, pp. 221-247.

Perron, P. and S. Ng., 1994, Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, manuscipt, Université de Montréal.

Pötscher, B.M., 1987, Estimation of autoregressive moving-average order given an infinite number of models and approximation of spectral densities, *Journal of Time Series Analysis* vol. 11(2), pp. 165-179.

Priestley, M.B., 1982, *Spectral analysis and time series*, Academic Press.

Shibata, R., 1976, Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika* 63(1), pp. 117-126.

Shibata, R., 1981, An optimal autoregressive spectral estimate, *The Annals of Statistics* 9, pp. 300-306.

Stock, J.H. and M.W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* 61, pp. 783-820.

West, K.D., 1994, Another heteroskedasticity and autocorrelation consistent covariance matrix estimator, manuscript, University of Wisconsion.

White, H., 1994, *Asymptotic theory for econometricians*, Academic Press.

Appendix: Proof of theorem 1.

Important for the proof of theorem 1 is the following lemma from Hannan and Kavalieris (1983). Define $\tilde{S}_T$ as the VARHAC estimator using lag order κ_T of the spectral density matrix at frequency zero of the series $\{ux_t(\psi_0)\}$.

Lemma 1: (Hannan and Kavalieris 1983, p. 293) Suppose that $\{ux_t(\psi_0)\}$ is generated under assumptions (a1) to (a5). If the lag order $\kappa_T \to \infty$ such that $\kappa_T = o((T/\log(T))^{1/2})$, then $\tilde{S}_T \to S$ almost surely.

We distinquish two cases. In the first case, the AR representation is of finite order. In that case, Shibata (1976) showed that the probability that the order chosen by AIC is less than the true order is zero in the limit. The spectral density is in this case a function of a finite number of parameters each of which are estimated consistently. The continuous mapping theorem then implies that the spectral density is also estimated consistently. In the second case, the AR representation is of infinite order. In this case, the order selected by AIC goes to infinity (cf. Pötscher (1987)). Therefore, lemma 1 also holds if the actual order is chosen by AIC, and the maximum lag order $\bar{K}_T = o (T/\log(T)^{1/2})$. The difference between $\tilde{S}_T$ and $\hat{S}_T$ is that $\tilde{S}_T$ uses $\{ux_t(\psi_0)\}$ instead of $\{ux_t(\hat{\psi}_T)\}$. Below we will show that $\hat{S}_T \to \tilde{S}_T$ in probability. This together with lemma 1 proves theorem 1.

Without loss of generality, we assume that ux_t is a scalar random variable, which simplifies the notation considerably. For a vector x, $\|x\|$ denotes the standard Euclidian norm. For a matrix A, we define the following matrix norm:

$$\|A\| = \sup_{\|x=1\|} \|Ax\|.$$

Let

$$\overline{ux}_t = ux_t(\bar{\psi}), \quad ux_{0t} = ux_t(\psi_0)$$

$$ux_{0t}^{(n)} = \left.\frac{\partial ux_t}{\partial \psi_n}\right|_{\psi=\psi_0} \qquad \overline{ux}_t^{(n)} = \left.\frac{\partial ux_t}{\partial \psi_n}\right|_{\psi=\bar{\psi}}$$

$$\gamma_j(\psi) = \frac{1}{T}\sum_{t=K+1}^{T} ux_t(\psi)\, ux_{t-j}(\psi) \qquad j = 0,\ldots,K$$

$$\bar{\gamma} = \gamma(\bar{\psi}), \quad \gamma_0 = \gamma(\psi_0)$$

$$\gamma_0^{(n)} = \left.\frac{\partial \gamma}{\partial \psi_n}\right|_{\psi=\psi_0} \qquad \bar{\gamma}^{(n)} = \left.\frac{\partial \gamma}{\partial \psi_n}\right|_{\psi=\bar{\psi}}$$

$$\Gamma_{ij}(\psi) = \frac{1}{T}\sum_{t=K+1}^{T} ux_{t-i}(\psi)\, ux_{t-j}(\psi) \qquad i,j = 1,\ldots,K$$

$$\Gamma_{ij}^{*} = E\; ux_{t-i}(\psi_0)\, ux_{t-j}(\psi_0) \qquad i,j = 1,\ldots,K$$

$$\bar{\Gamma} = \Gamma(\bar{\psi}), \quad \Gamma_0 = \Gamma(\psi_0)$$

$$\Gamma_0^{(n)} = \left.\frac{\partial \Gamma}{\partial \psi_n}\right|_{\psi=\psi_0} \qquad \bar{\Gamma}^{(n)} = \left.\frac{\partial \Gamma}{\partial \psi_n}\right|_{\psi=\bar{\psi}}$$

$$\alpha(\psi) = \Gamma^{-1}(\psi)\gamma(\psi)$$

$$\alpha_i(\psi) = \Gamma_i^{-1}(\psi)\gamma(\psi) \quad \text{where } \Gamma_i^{-1} \text{ is the } i^{th} \text{ row of } \Gamma^{-1}.$$

$$\hat{\alpha} = \alpha(\hat{\psi}), \quad \bar{\alpha} = \alpha(\bar{\psi}), \quad \alpha_0 = \alpha(\psi_0)$$

$$\bar{\alpha}^{(n)} = \left.\frac{\partial \alpha}{\partial \psi_n}\right|_{\psi=\bar{\psi}}$$

Lemma 2: If $\kappa_T = o(\,(T/\log(T))^{1/2})$, and under assumptions (a1) through (a10), then

$$\sup_n \sum_{i=1}^{\kappa_T} \bar{\gamma}_i^{(n)} = O_p(1), \quad \text{and} \quad \sup_{n,j} \sum_{i=1}^{\kappa_T} \bar{\Gamma}_{ij}^{(n)} = O_p(1).$$

proof: By mean value expansion about ψ_o,

$$(A.1) \qquad \bar{\gamma}_i^{(n)} = \gamma_{oi}^{(n)} + \sum_{m=1}^{N} \left. \frac{\partial^2 \gamma_i(\psi)}{\partial\psi_n \partial\psi_m} \right|_{\psi=\tilde{\psi}} (\bar{\psi}_m - \psi_{om})$$

for some $\tilde{\psi} \in \Psi$. Under assumptions (a9) and (a10),

$$\sup_{i,m} \left. \frac{\partial^2 \gamma_i(\psi)}{\partial\psi_n \partial\psi_m} \right|_{\psi=\tilde{\psi}} = O_p(1), \quad \text{and} \quad \sup_m (\bar{\psi}_m - \psi_{om}) = O_p(T^{-1/2}).$$

Since $k_T = o(T^{1/2})$ we have

$$(A.2) \qquad \sup_n \sum_{i=1}^{\kappa_T} \bar{\gamma}_i^{(n)} = \sup_n \sum_{i=1}^{\kappa_T} \gamma_{oi}^{(n)} + o_p(1),$$

where

$$(A.3) \qquad \gamma_{oi}^{(n)} = (1/T) \sum_{t=\kappa}^{T} ux_{ot} \; ux_{ot-i}^{(n)} + ux_{ot-i} \; ux_{ot}^{(n)}.$$

Using assumptions (a1) through (a6), the results of Hannan and Kavalieris (1983, p. 291, 293) ensure that $\sum_{i=1}^{\kappa_T} \gamma_{oi}^{(n)} = O_p(1)$. (Note that condition (a6) ensures that (a1) through (a5) hold for the vector including ux_{ot} and $ux_{ot}^{(n)}$). This result combined with (A.2) implies the first conclusion of the lemma. The proof of the second conclusion follows exactly the same arguments.

Lemma 3: (Hannan and Kavalieris 1983, p.293-296) If $\kappa_t = o((T/\log(T))^{1/2})$ and under assumptions (a1) through (a5), then

$$\sum_{i=1}^{\kappa_T} \alpha_{oi} = O_p(1).$$

Lemma 4: If $\kappa_T = o((T/\log(T))^{1/2})$, and under conditions (a1) through (a10), then

$$\sup_{i,j} | \bar{\Gamma}_{ij}^{-1} | = O_p(1).$$

proof: We have

$$(A.4) \qquad \bar{\Gamma} = \Gamma^* + (\Gamma_0 - \Gamma^*) + (\bar{\Gamma} - \Gamma_0).$$

The smallest eigenvalue of Γ^* is bounded away from zero uniformly in κ_T by assumption (a2). Under assumptions (a1) through (a5) Hannan and Kavalieris (1983, p.291,293) showed that $\|\Gamma_o - \Gamma^*_j\| = o_p(1)$. Finally, by a mean-value expansion, we have

$$(A.5) \qquad \bar{\Gamma} = \Gamma_0 + \sum_{n=1}^{N} \left.\frac{\partial \Gamma}{\partial \psi_n}\right|_{\psi=\tilde{\psi}} (\bar{\psi}_n - \psi_{on}),$$

for some $\tilde{\psi} \in \Psi$. Following arguments similar to (A.2) and (A.3) in Lemma 2, we have

$$(A.6) \qquad \sup_n \sum_{i=0}^{\kappa_T} \left.\frac{\partial \gamma_i}{\partial \psi_n}\right|_{\psi=\tilde{\psi}} = O_p(1), \text{ and } \sup_n (\bar{\psi}_n - \psi_{on}) = O_p(T^{-1/2}), \text{ so that}$$

$\| \bar{\Gamma} - \Gamma_0 \| = o_p(1)$. Thus, from (A.4), the smallest eigenvalue of $\bar{\Gamma}$ is uniformly bounded away from zero, and thus the largest eigenvalue of $\bar{\Gamma}^{-1}$ is bounded uniformly in κ_T.

Proof of Theorem 1: A mean value expansion of $\alpha_i(\hat{\psi})$ yields

$$(A.7) \qquad \alpha_i(\hat{\psi}) = \alpha_i(\psi_0) + \sum_{n=1}^{N} \left.\frac{\partial \alpha_i(\psi)}{\partial \psi_n}\right|_{\psi=\bar{\psi}} [\hat{\psi}_n - \psi_{0,n}]$$

for some value of $\bar{\psi}$ on the line segment joining $\hat{\psi}$ and ψ_0. Also

$$(A.8) \qquad \left.\frac{\partial \alpha_i(\psi)}{\partial \psi_n}\right|_{\psi=\bar{\psi}} = \bar{\Gamma}_i^{-1} [\, \bar{\gamma}^{(n)} - \bar{\Gamma}^{(n)} \bar{\alpha} \,].$$

Rearranging terms gives:

$$(A.9)\qquad \alpha_i(\hat{\psi}) - \alpha_i(\psi_0) = \sum_{n=1}^{N} \bar{\Gamma}_i^{-1} [\, \bar{\gamma}^{(n)} - \bar{\Gamma}^{(n)}\alpha_0] \,(\, \hat{\psi}_n - \psi_{on}) - \sum_{n=1}^{N} \bar{\Gamma}_i^{-1}\, \bar{\Gamma}^{(n)}\, (\bar{\alpha} - \alpha_0)\,\, (\, \hat{\psi}_n - \psi_{on}).$$

By Taylor's theorem $\|\bar{\alpha}-\alpha_0\|^2 \le \|\hat{\alpha}-\alpha_0\|^2$, so that $\|\bar{\alpha}-\alpha_0\|^2 \le \kappa_T \sup_i \|\hat{\alpha}_i-\alpha_{i0}\|^2$. Therefore,

$$(A.10)\qquad \sup_i \|\hat{\alpha}_i-\alpha_{i0}\|^2 \le [\, 1 + N\,\kappa_T \sup_{i,n}\|\bar{\Gamma}_i^{-1}\bar{\Gamma}^{(n)}\|^2 \,\|\hat{\psi}_n-\psi_{on}\|^2\,]^{-1} \{\, N \sup_{i,n} [\, \|\bar{\Gamma}_i^{-1}\bar{\gamma}^{(n)}\|^2 + \|\bar{\Gamma}_i^{-1}\bar{\Gamma}^{(n)}\|^2] \,\|\hat{\psi}_n-\psi_{on}\|^2 \,\}.$$

Now

$$(A.11)\qquad \sup_{i,\ell,n} |\textstyle\sum_{j=1}^{\kappa_T} \bar{\Gamma}_{ij}^{-1}\, \bar{\Gamma}_{j\ell}^{(n)}|$$

$$\le \sup_{ij} |\bar{\Gamma}_{ij}^{-1}| \sup_{\ell,n}|\textstyle\sum_{j=1}^{\kappa_T} \bar{\Gamma}_{j\ell}^{(n)}| = O_p(1) \text{ by}$$

lemmas 2 and 3.

Therefore,

$$(A.12)\qquad \sup_{i,n} \|\bar{\Gamma}_i^{-1}\bar{\Gamma}^{(n)}\|^2 = \sup_{i,n} [\textstyle\sum_{\ell=1}^{\kappa_T} |\sum_{j=1}^{\kappa_T} \bar{\Gamma}_{ij}^{-1}\, \bar{\Gamma}_{j\ell}^{(n)}|^2\,] = O_p(\kappa_T).$$

Similarly,

$$(A.13)\qquad \sup_{i,n} \|\bar{\Gamma}_{ij}^{-1}\, \bar{\gamma}^{(n)}\|^2 \le \sup_{ij} |\bar{\Gamma}_{ij}^{-1}|^2 \sup_n |\textstyle\sum_{j=1}^{\kappa_T} \bar{\gamma}_j^{(n)}|^2 = O_p(1).$$

Furthermore, by lemmas 2,3 and 4, we have:

$$(A.14) \qquad \sup_{i,n} \|\bar{\Gamma}^{-1}\bar{\Gamma}^{(n)}\alpha_0\|^2 = \sup_{i,n} |\sum_{j=1}^{K_T} \sum_{\ell=1}^{K_T} \bar{\Gamma}^{-1}_{ij}\bar{\Gamma}^{(n)}_{j\ell} \alpha_{o\ell}{}^2|$$

$$\leq (\sup_{i,j} |\bar{\Gamma}^{-1}_{ij}|^2)(\sup_{\ell,n}|\sum_{j=1}^{K_T} \bar{\Gamma}^{(n)}_{j\ell}|^2)(\sum_{\ell=1}^{K_T} \alpha_{o\ell})^2$$

$$= O_p(1).$$

Therefore, since N is fixed, $\kappa_T = o((T/\log(T))^{1/2})$ and $\sup_n\|\hat{\psi}_n-\psi_{on}\|^2 = O_p(T^{-1})$, we have from (A.10) that:

$$\sup_i\|\hat{\alpha}_i-\alpha_{io}\|^2 = O_p(T^{-1})$$

so that

$$\sum_{i=1}^{K_T} \hat{\alpha}_i = \sum_{i=1}^{K_T} \alpha_{io} + O_p[(\log T)^{-1/2}].$$

Using similar arguments, we have $\sigma^2_\varepsilon(\hat{\psi}) = \sigma^2_\varepsilon(\psi_0) + O_p[(\log T)^{-1/2}]$.

Then the conclusion from the theorem follows from lemma 1.

TABLE 1

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the AR(1)-HOMO model, $T = 128$, and 10,000 replications.

ρ	Estimator	99%	95%	90%
0	QS-PW	98.3	94.0	88.9
	VARHAC	98.4	94.1	89.2
.3	QS-PW	98.3	93.5	88.4
	VARHAC	97.9	92.6	86.6
.5	QS-PW	97.9	92.6	87.1
	VARHAC	96.6	90.5	84.4
.7	QS-PW	96.8	90.6	84.4
	VARHAC	96.2	89.6	83.4
.9	QS-PW	91.4	82.9	76.2
	VARHAC	90.6	81.8	74.8
.95	QS-PW	86.5	77.2	70.2
	VARHAC	85.5	76.1	69.1
-.3	QS-PW	98.3	93.5	88.4
	VARHAC	97.9	92.6	86.9
-.5	QS-PW	98.0	92.6	87.5
	VARHAC	96.8	90.7	84.8

TABLE 2

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the AR(1)-HET(1) model, T = 128, and 10,000 replications.

ρ	Estimator	99%	95%	90%
0	QS-PW	97.7	92.7	86.8
	VARHAC	97.8	92.6	86.7
.3	QS-PW	97.3	91.6	85.7
	VARHAC	96.4	89.5	83.1
.5	QS-PW	96.0	89.6	83.3
	VARHAC	94.5	87.0	80.2
.7	QS-PW	92.9	84.5	77.1
	VARHAC	92.1	83.6	76.3
.9	QS-PW	81.0	69.7	62.3
	VARHAC	79.6	68.9	61.3
.95	QS-PW	70.2	59.5	52.3
	VARHAC	69.0	58.3	51.0
-.3	QS-PW	97.6	92.1	86.2
	VARHAC	98.1	93.2	87.9
-.5	QS-PW	97.6	92.1	86.0
	VARHAC	97.7	92.6	87.3

TABLE 3

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the AR(1)-HET(2) model, T = 128, and 10,000 replications.

ρ	Estimator	99%	95%	90%
0	QS-PW	98.1	92.8	87.3
	VARHAC	98.0	93.1	87.4
.3	QS-PW	97.5	92.1	86.3
	VARHAC	97.3	91.2	84.9
.5	QS-PW	96.8	90.5	84.5
	VARHAC	95.7	88.5	81.7
.7	QS-PW	94.6	87.1	80.5
	VARHAC	93.1	85.4	78.3
.9	QS-PW	86.7	77.2	69.4
	VARHAC	85.6	75.2	67.7
.95	QS-PW	80.0	69.4	61.5
	VARHAC	78.8	67.3	60.0
-.3	QS-PW	97.9	93.1	87.2
	VARHAC	98.2	93.9	88.4
-.5	QS-PW	98.1	93.5	87.8
	VARHAC	98.3	93.8	88.5

TABLE 4

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the MA(1)-HOMO, MA(1)-HET(1) and MA(1)-HET(2) models, T = 128, and 10,000 replications.

model	ρ	Estimator	99%	95%	90%
MA(1)-HOMO	.3	QS-PW	98.3	93.7	88.7
		VARHAC	98.1	93.0	87.2
	.5	QS-PW	98.4	93.9	88.7
		VARHAC	97.4	91.7	86.0
	.7	QS-PW	98.4	94.0	89.1
		VARHAC	97.1	91.5	85.9
	.99	QS-PW	98.4	94.1	89.2
		VARHAC	97.2	91.6	85.9
MA(1)-HET1	.3	QS-PW	97.8	92.8	87.1
		VARHAC	97.0	90.7	84.7
	.5	QS-PW	97.8	92.9	87.5
		VARHAC	96.5	90.1	84.0
	.7	QS-PW	97.9	93.0	87.7
		VARHAC	96.4	89.9	83.6
	.99	QS-PW	97.9	93.0	87.7
		VARHAC	96.4	89.7	83.5
MA(2)-HET2	.3	QS-PW	97.8	93.0	87.4
		VARHAC	97.5	91.9	85.9
	.5	QS-PW	97.8	92.8	87.7
		VARHAC	97.1	91.0	85.0
	.7	QS-PW	97.7	92.8	87.7
		VARHAC	96.8	90.5	84.6
	.99	QS-PW	97.7	92.9	87.5
		VARHAC	96.6	90.2	84.2

TABLE 5

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the MA(m)-HOMO models, T = 128, and 10,000 replications.

m	Estimator	99%	95%	90%
3	QS-PW	97.8	93.5	88.7
	VARHAC	96.8	91.8	86.3
5	QS-PW	97.3	92.6	87.5
	VARHAC	96.6	90.8	85.1
7	QS-PW	96.7	91.5	85.8
	VARHAC	95.8	89.6	83.8
9	QS-PW	95.9	90.0	84.4
	VARHAC	94.9	88.2	82.1
12	QS-PW	94.2	88.0	82.1
	VARHAC	93.3	86.3	79.8
15	QS-PW	92.9	86.1	79.7
	VARHAC	91.9	84.2	77.6

TABLE 6: The limitations of a single bandwidth

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the following dgp:

$$Y_t = \varepsilon_t, \quad \varepsilon_t = .9\varepsilon_{t-1} + e_t,$$

ε_t and X_t are N(0,1)

$Z_t = \lambda X_t$, where λ_t is a scalar

regression: $Y_t = \alpha + \beta Z_t + u_t$,

T = 128, and 10,000 replications

	λ	Estimator	99%	95%	90%
	1	VARHAC	93.2	85.9	79.9
		QS	87.8	78.7	72.0
α	100	VARHAC	93.2	85.9	79.9
		QS	62.3	51.0	43.7
	1000	VARHAC	93.2	85.9	79.9
		QS	56.5	45.1	38.3
	1	VARHAC	98.7	94.2	89.3
		QS	92.7	84.3	76.7
β	100	VARHAC	98.7	94.2	89.3
		QS	98.6	94.0	88.2
	1000	VARHAC	98.7	94.2	89.3
		QS	98.6	94.2	88.7

TABLE 7: The limitations of arbitrary parameterizations

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the following dgp:

$$Y_t = \nu\, \varepsilon_t + \mu\, \varepsilon_{t-q}, \quad q = 2,3$$
$$\varepsilon_t \text{ is i.i.d. } N(0,1)$$

T = 128, and 10,000 replications

	Estimator	ν	μ	99%	95%	90%
q = 2	QS-PW	0	-.3	100.0	99.6	98.3
	VARHAC	0	-.3	98.7	94.9	90.3
	QS-PW	-.1	-.3	100.0	99.8	99.1
	VARHAC	-.1	-.3	99.0	95.7	91.4
	QS-PW	0	.3	95.1	87.4	80.3
	VARHAC	0	.3	97.9	92.9	87.8
	QS-PW	.1	.3	95.9	88.6	81.7
	VARHAC	.1	.3	97.7	92.8	87.8
q = 3	QS-PW	0	-.3	100.0	99.3	98.0
	VARHAC	0	-.3	99.1	96.0	91.9
	QS-PW	-.1	-.3	100.0	99.6	98.7
	VARHAC	-.1	-.3	99.2	96.7	93.3
	QS-PW	0	.3	95.5	87.5	80.9
	VARHAC	0	.3	97.8	92.8	88.4
	QS-PW	.1	.3	95.7	88.1	81.5
	VARHAC	.1	.3	97.8	92.8	88.1

TABLE 8: The limitations of an arbitrary prewhitening order

99%, 95%, and 90% confidence intervals constructed using QS-PW and PARA estimators for the following dgp:

$$Y_t = \tfrac{1}{2}\psi Y_{t-1} + \tfrac{1}{2}\psi Y_{t-2} + \varepsilon_t$$

ε_t is i.i.d. N(0,1)

T = 128, and 10,000 replications

ψ	Estimator	99%	95%	90%
.5	QS-PW	92.8	84.0	76.3
	VARHAC	96.8	91.1	85.7
.7	QS-PW	87.0	75.9	67.8
	VARHAC	96.0	89.9	84.5
.9	QS-PW	70.0	57.7	50.6
	VARHAC	90.4	82.5	76.4
.95	QS-PW	57.2	46.5	40.2
	VARHAC	83.1	74.2	67.8

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	1995	
504	Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures	Wouter J. Den Haan Andrew T. Levin
503	Exchange-Rate Based Inflation Stabilization: The Initial Real Effects of Credible Plans	Martin Uribe
502	Strategic Returns to International Diversification: An Application to the Equity Markets of Europe, Japan, and North America	John Ammer Jianping Mei
501	Real Exchange Rate Movements in High Inflation Countries	John H. Rogers Ping Wang
500	Political Competition, Casual Relations Between Taxes and Spending, and Their Influence on Government Size: Evidence From State-Level Data	Diane Lim Rogers John H. Rogers
499	International Stock Price Spillovers and Market Liberalization: Evidence From Korea, Japan, and the United States	Sang W. Kim John H. Rogers
498	How Wide is the Border?	Charles Engle John H. Rogers
497	Constrained Suboptimality in Economies with Limited Communication	David Bowman
496	Saving-Investment Associations and Capital Mobility On the Evidence from Japanese Regional Data	Robert Dekle
495	Convertibility Risk, Default Risk, and the Mexdollar Anomaly	John H. Rogers
494	Government Budget Deficits and Trade Deficits: Are Present-Value Constraints Satisfied in Long-Term Data?	Shaghil Ahmed John H. Rogers
493	Real Shocks and Real Exchange Rates in Really Long-Term Data	John H. Rogers

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.

International Finance Discussion Papers

IFDP Number	Titles	Author(s)
	1994	
492	Loss Aversion in a Consumption/Savings Model	David Bowman Deborah Minehart Matthew Rabin
491	Terms-of-Trade Uncertainty and Economic Growth: Are Risk Indicators Significant in Growth Regressions	Enrique G. Mendoza
490	Politics, Economics, and Investment: Explaining Plant and Equipment Spending by U.S. Direct Investors in Argentina, Brazil, and Mexico	Guy V.G. Stevens
489	On The Dynamic Properties of Asymmetric Models of Real GNP	Allan D. Brunner
488	A distributed block approach to solving near-block-diagonal systems with an application to a large macroeconometric model	Jon Faust Ralph Tryon
487	Conditional and Structural Error Correction Models	Neil R. Ericsson
486	Bank Positions and Forecasts of Exchange Rate Movements	Michael P. Leahy
485	Technological Progress and Endogenous Capital Depreciation: Evidence from the U.S. and Japan	Robert Dekle
484	Are Banks Market Timers or Market Makers? Explaining Foreign Exchange Trading Profits	John Ammer Allan D. Brunner
483	Constant Returns and Small Markups in U.S. Manufacturing	Susanto Basu John G. Fernald
482	The Real Exchange Rate and Fiscal Policy During the Gold Standard Period: Evidence from the United States and Great Britain	Graciela L. Kaminsky Michael Klein
481	The Debt Crisis: Lessons of the 1980s for the 1990s	Graciela L. Kaminsky Alfredo Pereira
480	Who Will Join EMU? Impact of the Maastricht Convergence Criteria on Economic Policy Choice and Performance	R. Sean Craig
479	Determinants of the 1991-93 Japanese Recession: Evidence from a Structural Model of the Japanese Economy	Allan D. Brunner Steven B. Kamin