VECTOR AUTOREGRESSIVE COVARIANCE MATRIX ESTIMATION

Wouter J. den Haan
Department of Economics, University of California, San Diego
and National Bureau of Economic Research

and

Andrew T. Levin
International Finance Division
Federal Reserve Board of Governors

First Draft: November 1994
This Version: February 1998

This paper proposes a vector autoregressive (VAR) spectral estimation procedure for constructing heteroskedasticity and autocorrelation consistent (HAC) covariance matrices. We establish the consistency of the VARHAC estimator under general conditions similar to those considered in previous research, and we demonstrate that this estimator converges at a faster rate than the kernel-based estimators proposed by Andrews and Monahan (1992) and Newey and West (1994). In finite samples, Monte Carlo simulation experiments indicate that the VARHAC estimator matches, and in some cases greatly exceeds, the performance of the prewhitened kernel estimator proposed
by Andrews and Monahan (1992). These simulation experiments also illustrate several important limitations of kernel-based HAC estimation procedures, and highlight the advantages of explicitly modeling the temporal properties of the error terms.

We appreciate comments and suggestions from Don Andrews, Larry Christiano, Graham Elliott, Rob Engle, Neil Ericsson, Ron Gallant, Clive Granger, Jim Hamilton, Adrian Pagan, Peter Phillips, Brian Poetscher, Peter Robinson, Jim Stock, P.A.V.B. Swamy, George Tauchen, Hal White, and three anonymous referees. This project is supported by NSF grant SBR-9514813. The views expressed in this paper do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or of any other members of its staff. GAUSS, RATS, and Fortran programs to calculate the VARHAC estimator can be found on the web-site http://weber.ucsd.edu/~wdenhaan.

1. INTRODUCTION

The recent literature on heteroscedasticity-and-autocorrelation-consistent (HAC) covariance matrices has mainly focused on kernel-based methods of estimating the spectral density matrix at frequency zero.[1] Nevertheless, Parzen (1969) identified several advantages of autoregressive (AR) spectral density estimation, and these advantages have been highlighted in a variety of simulation experiments (e.g., Beamish and Priestley 1981; Kay and Marple 1981; Parzen 1983). The consistency of the AR spectral estimator has been demonstrated for the case of weakly stationary data, under specific assumptions about the growth rate of the lag order h as a function of the sample length T (cf. Berk 1974; An et al. 1982; Hannan and Kavalieris 1983). However, the consistency of the AR spectral estimator has not been verified under more general mixing conditions or in the case of estimated residuals, and no results have been available regarding its convergence rate when the lag order is determined by a data-dependent model selection criterion.

This paper proposes a HAC covariance matrix estimator, referred to as the VARHAC estimator, in which the spectral density at frequency zero is constructed using vector autoregressive (VAR) spectral estimation, and Schwarz' (1978) Bayesian Information Criterion (BIC) is used to select the lag structure of the VAR model. We establish the consistency and convergence rate of the VARHAC estimator under general conditions of heteroskedasticity and temporal dependence, similar to the conditions used to analyze kernel-based estimators (e.g., Andrews 1991). In particular, the data generating process (*dgp*) need not be finite-order ARMA or even weakly stationary.

Under these conditions, we demonstrate that the VARHAC estimator efficiently captures the unconditional second moments of the data: even in the absence of covariance stationarity, this estimator achieves a faster convergence rate than any estimator in the class of positive semi-definite (PSD) kernels. The VAR spectral estimator can be expressed as an infinite sum of the autocovariances implied by the estimated VAR of order h; i.e., this estimator assigns weights of unity to all of the sample autocovariances up to order h, and ensures a PSD spectral density matrix by extrapolating higher-order autocovariances that vanish at an exponential rate. Thus, as originally conjectured by Parzen (1969), the bias of the VAR spectral estimator is asymptotically of the same order as the bias of the simple truncated kernel estimator (which assigns weights of unity to all sample autocovariances up to the truncation point h), and declines at a faster rate than the bias of any PSD kernel-based estimator (which assigns weights less than unity to the sample autocovariances, based on the value of the bandwidth parameter h). Moreover, as in Parzen (1969) and Berk (1974), we show that the asymptotic variance of the VAR spectral estimator is $\mathbf{O}(h / T)$, just as for kernel-based spectral estimators. Finally, we extend the lag order selection results of

[1] Eichenbaum et al. (1987) and West (1997) implemented covariance matrix estimators for the case of a vector moving-average (MA) process of known finite order. Andrews (1991) and Andrews and Monahan (1992) briefly considered a first-order AR spectral estimator, but the estimator did not correct for heteroskedasticity and performed poorly in simulation experiments. Stock and Watson (1993) utilized AR(2) and AR(3) covariance matrix estimators in simulation experiments and in an empirical application. Finally, Lee and Phillips (1994) have analyzed the properties of an ARMA-prewhitened HAC estimator for the case of a finite-order ARMA process with i.i.d. innovations.

Shibata (1980, 1981), Hannan and Kavalieris (1986), and Hannan and Deistler (1988) to allow for conditional or unconditional heteroskedasticity, and we show that BIC yields a lag order growth rate that asymptotically approaches the optimal rate (i.e., the rate that minimizes the asymptotic mean-squared error of the VAR spectral estimator).

In particular, by evaluating the goodness-of-fit relative to the degree of parsimony, BIC appropriately reflects the tradeoff between the asymptotic bias and asymptotic variance of the VAR spectral estimator.

We utilize Monte Carlo simulation experiments to evaluate the finite-sample performance of the VARHAC estimator in generating accurate confidence intervals for linear regression coefficients. In replicating the simulation experiments performed by Andrews and Monahan (1992), we find that the VARHAC estimator generally matches the performance of the first-order prewhitened quadratic-spectral (QS-PW(1)) estimator proposed by Andrews and Monahan (1992) for a wide variety of *dgp*s. To highlight the finite-sample advantages of the VARHAC estimator, we also describe several simulation experiments in which the accuracy of this estimator greatly exceeds that of the QS-PW estimator. In particular, the kernel-based estimator may yield very poor results when an arbitrarily specified time series process is used to determine the value of the bandwidth parameter, or when higher-order AR components are present in the data. Furthermore, to ensure a PSD covariance matrix, a kernel-based estimator must use the same bandwidth parameter for every element of the residual vector. As pointed out by Robinson (1996), this constraint can yield very low accuracy when the autocovariance structure varies substantially across elements. In contrast,

the VARHAC estimator permits the lag order to vary across equations in the VAR (and across the variables in each equation), since the resulting covariance matrix is PSD by construction.

The remainder of this paper is organized as follows: Section 2 provides a step-by-step description of the VARHAC covariance matrix estimation procedure. Section 3 establishes the consistency and rate of convergence of the VARHAC procedure. Section 4 compares the asymptotic and finite-sample properties of the VARHAC estimator with those of prewhitened kernel-based HAC covariance matrix estimators.

2. THE VARHAC PROCEDURE

In many estimation problems, a parameter estimate $\hat{\psi}_T$ for a $p\times1$ parameter vector ψ_0 is obtained for a sample of length T using the sample analog of a set of moment conditions, such as $\mathrm{E}\, V_t(\psi_0) = 0$, where $V_t(\psi_0)$ is an $N\times1$ vector of residual terms with $N \geq p$. This orthogonality condition is often used to motivate the following estimator of ψ_0:

$$\hat{\psi}_T = \operatorname{argmin}_{\psi} V'_T M_T V_T, \tag{2.1}$$

where $V_T = \sum_{t=1}^{T} V_t(\psi) / T$ is the vector of sample moments of $V_t(\psi)$, and M_T is an $N\times N$ (possibly random) symmetric weighting matrix (cf. Hansen 1982). Under certain regularity conditions, $T^{1/2}\left(\hat{\psi}_T - \psi_0\right)$ has a limiting normal distribution with mean 0 and covariance matrix $\Omega = 2\pi B' f(0) B$, where $f(0)$ denotes the limiting spectral density at frequency zero of the process $V_t(\psi_0)$. In particular, the $N\times N$ matrix $f(0) = \lim_{T\to\infty} S_T / 2\pi$;

the $N \times p$ matrix $B = \lim_{T\to\infty} M_T D_T (D_T' M_T D_T)^{-1}$, and S_T and D_T are defined as follows:

$$S_T = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathrm{E}\, V_s(\psi_0) V'_t(\psi_0) \tag{2.2}$$

$$D_T = \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left[\left.\frac{\partial V_t(\psi)}{\partial \psi'}\right|_{\psi=\psi_0}\right] \tag{2.3}$$

The matrix D_T is typically estimated by its sample analog $\hat{D}_T \equiv D_T(\hat{\psi}_T)$, and $\hat{D}_T - D_T \to 0$ in probability as $T \to \infty$. In this paper, we construct the spectral estimator $\hat{S}_T(\hat{\psi}_T)$ using a VAR representation of $V_t(\hat{\psi}_T)$, for which BIC is used to select the lag order for each equation in the VAR. The covariance matrix estimator based on $\hat{S}_T(\hat{\psi}_T)$ will be referred to as the VARHAC estimator.

Step 1. Lag order selection for each VAR equation. For the n^{th} element $\hat{V}_{nt}$ of the vector $V_t(\hat{\psi}_T)$ ($n = 1,...,N$) and for each lag order $h = 1,..., \overline{H}$, the following model is estimated by ordinary least squares (OLS):

$$V_{nt} = \sum_{j=1}^{N}\sum_{k=1}^{h} \hat{\alpha}_{njk}(h)\, V_{j,t-k} + \hat{e}_{nt}(h) \qquad \text{for} \quad t = \overline{H}+1, \cdots, T. \tag{2.5}$$

For lag order 0, we set $\hat{e}_{nt}(0) \equiv V_{nt}$. Below we will discuss the maximum lag order, $\overline{H}$, that one wants to consider. Then the value of the BIC criterion is calculated for each lag order $h = 0,..., \overline{H}$.

$$\mathrm{BIC}(h;n) = \log\left(\det\left(\frac{\sum_{t=\overline{H}+1}^{T} \hat{e}_{nt}(h)\hat{e}_{nt}(h)'}{T}\right)\right) + \frac{h\, N \log(T)}{T}. \tag{2.6}$$

For each element of $V_t(\hat{\psi}_T)$ the optimal lag order h_n is chosen as the value of h that minimizes BIC(h; n).

To minimize the computational requirements of the Monte Carlo simulation experiments, we only consider specifications in which all elements of V_t enter with the same number of lags in the regression equation for $\hat{V}_{nt}$, so that the model selection procedure involves estimating a total of $N(\overline{H}+1)$ equations. Allowing a different lag order for each variable in each equation would require a total of $N(\overline{H}+1)^N$ equations to be estimated, which is only computationally feasible if N and $\overline{H}$ are fairly small. On the other hand, one could further restrict the set of admissible VAR models by using a system criterion to select the same lag order for all elements of V_t. Section 3 demonstrates that the VARHAC estimator achieves a faster convergence rate than kernel-based estimators, even when a system criterion is used to determine the lag order. However, the experiments reported in Section 4.2 indicate that allowing the lag order to differ across equations can yield substantial benefits in finite samples.

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

$$\sum_{k=0}^{\overline{H}} \hat{A}_k\, V_{t-k}(\hat{\psi}_T) = \hat{e}_t, \tag{2.7}$$

where $\hat{e}_t$ is an $N\times1$ vector with typical element $\hat{e}_{nt}(h_n)$. The (n,j) element of $\hat{A}_k$ is equal to zero if $k > h_n$, and is equal to $-\hat{\alpha}_{njk}(h_n)$ if $0 < k \leq h_n$. $\hat{A}_0$ is the identity matrix. The innovation covariance matrix $\hat{\Sigma}_T$ is estimated as follows:

$$\hat{\Sigma}_T = \frac{\sum_{t=\overline{H}+1}^{T} \hat{e}_t \, \hat{e}'_t}{T-\overline{H}}. \tag{2.8}$$

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 the innovation covariance matrix contains significant off-diagonal elements.

Step 3: Estimation of HAC covariance matrix. Using the results of step 1 and 2, the VAR spectral estimator is constructed as follows:

$$\hat{S}_T(\hat{\psi}_T) = \left[\sum_{k=0}^{\overline{K}} \hat{A}_k\right]^{-1} \hat{\Sigma}_T \left[\sum_{k=0}^{\overline{K}} \hat{A}'_k\right]^{-1} \tag{2.9}$$

Finally, the VARHAC covariance matrix estimator is defined by:

$$\hat{V}_T(\hat{\psi}_T) = \left[\hat{D}'_T \, M_T \, \hat{D}_T\right]^{-1} \hat{D}'_T \, M_T \, \hat{S}_T(\hat{\psi}_T) \, M_T \, \hat{D}_T \left[\hat{D}'_T \, M_T \, \hat{D}_T\right]^{-1}. \tag{2.10}$$

3. ASYMPTOTIC PROPERTIES

In this section, we establish the consistency and convergence rate of the VAR spectral estimator under general conditions of heteroskedasticity and temporal dependence. (All proofs are given in the appendix.) Section 3.1 establishes the conditions under which the true autocovariance structure of the data can be represented by an infinite-order VAR. Section 3.2 evaluates the convergence rate of the VAR estimator of the spectral density of observed data. Finally, Section 3.3 extends these results to the VARHAC procedure, which is applied to estimated regression residuals.

3.1 VAR(∞) Representation of Autocovariance Structure.

Before analyzing the properties of AR approximation, it is important to establish the conditions under which the true autocovariance structure of a stochastic process can be represented by an infinite-order VAR. Consider a mean-zero sequence $\{V_t\}_{t=-\infty}^{\infty}$ of random N-vectors. For a given sample of length T, we define the average jth-order autocovariance matrix $\Gamma_T(j)$ as follows for $|j| < T$:

$$\begin{aligned} \Gamma_T(j) &= \frac{1}{T-j}\sum_{t=1}^{T-j} \mathrm{E}\left(V_t V'_{t+j}\right) \qquad \text{for } j \geq 0, \quad \text{and} \\ \Gamma_T(j) &= \frac{1}{T-|j|}\sum_{t=1-j}^{T} \mathrm{E}\left(V_t V'_{t+j}\right) \qquad \text{for } j < 0. \end{aligned} \tag{3.1}$$

We shall utilize the notation $|x| = \sup_j |x_j|$ to represent the supremum norm of a vector x. For an $L \times M$ matrix X, we utilize the matrix norm $|X| = \sup_{i=1,\ldots,L} \sum_{j=1}^{M} |X_{ij}|$ (cf. Hannan and Deistler 1988, p.266). Then the following condition is sufficient to ensure that $\{V_t\}$ meets Grenander's (1954) conditions for asymptotic stationarity (cf. Parzen 1961; Hannan 1970, p.77).

Condition A: $\{V_t\}_{t=-\infty}^{\infty}$ is a mean-zero sequence of random N-vectors, satisfying:

(1) $\sum_{j=0}^{\infty} \sup_{t \geq 1} \left| \mathrm{E}\left(V_t V'_{t+j}\right) \right| < \infty$

(2) $\exists \ lim_{T \to \infty} \Gamma_T(j) = \Gamma(j)$ for all j $\in (-\infty, +\infty)$

(3) $\Gamma(j) = \Gamma_T(j) + \mathbf{O}_p(T^{-1/2})$ for all j $\in (-\infty, +\infty)$

(4) $\Gamma(0)$ is positive definite

Condition A(1) is a standard assumption in the analysis of kernel-based spectral estimators (e.g., Andrews 1991); under this assumption, $\Gamma_T(j) = \mathbf{O}(1)$ for all j $\in (-\infty, +\infty)$. For each value of j, Conditions A(2) and A(3) rule out certain patterns of extreme temporal dependence in the sequence $\mathrm{E}(V_t V'_{t+j})$.[2] Condition A(4) excludes *dgp*s in which the components of V_t are asymptotically collinear. Under the assumptions of Condition A, the spectral density function, $f(\omega)$, can be defined as follows:

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) \exp(i\omega j) \quad \text{for} \quad \omega \in [-\pi, \pi]. \tag{3.3}$$

If $f(\omega)$ is positive definite almost everywhere in $[0, \pi]$, then the autocovariance sequence $\{\Gamma(j)\}_{j=-\infty}^{\infty}$ and the spectral density function $f(\omega)$ are identical to those of a vector MA(∞) process with i.i.d. Gaussian innovations, where the MA coefficients are square-summable (cf. Theorem IV.6.2 of Doob 1953, p. 160-161; Hannan 1970, pp.160-163; Priestley 1982, pp.730-733). In this paper, however, we focus primarily on the use of autoregressive approximation, in which case an additional assumption is required:

Condition B: The spectral density function $f(\omega)$ is positive definite over $[0,\pi]$.

This assumption ensures that the vector MA(∞) representation can be inverted into a VAR(∞) representation of the autocovariance sequence $\{\Gamma(j)\}_{j=-\infty}^{\infty}$. To formulate the VAR(∞) representation, it is useful to define the sequence G_{TM} of Toeplitz matrices, where the average autocovariance matrix $\Gamma_T(j-i)$ comprises the (i,j)th $N \times N$ block of G_{TM}, for $i, j = 1,\ldots, M$. The corresponding $MN \times MN$ matrices G_M, and the infinite-dimensional Hankel matrix G_∞, are each composed of the limiting autocovariance matrices. Thus, $\Gamma(j-i)$ comprises the (i, j)th block of G_M for $i, j = 1,\ldots, M$; and the (i,j)th block of G_∞ for $i, j = 1,2,\ldots$. In addition, let $-\Gamma(j)$ comprise the j-th $N \times N$ block of the $MN \times N$ matrix g_M for $j = 1,\ldots,M$; and the j-th $N \times N$ block of the matrix g_∞ for $j = 1,2,\ldots$. It should be noted that Condition B places an important restriction on the Toeplitz matrices G_M and G_∞. For univariate processes, the smallest eigenvalue of G_∞ is equal to the smallest value of $f(\omega)$ for ω in $[-\pi, \pi]$, and the smallest

[2] For example, Conditions A(2) and A(3) will hold if the moments $\mathrm{E}(V_t V'_{t+j})$ are themselves drawn from a random distribution with mean $\Gamma(j)$ and finite variance for each t and j, under mixing conditions sufficient to ensure that $\Gamma_T(j)$ follows a central limit theorem.

eigenvalue of G_M declines monotonically and converges to the smallest eigenvalue of G_∞ as $M \to \infty$ (cf. Grenander and Szegö 1958, pp.147-54; Hannan 1970, pp.148-50; Fuller 1996, p. 154). Thus, condition B is equivalent to the restriction that $\det(G_M) \neq 0$ for all $M \geq 1$ and that $\det(G_\infty) \neq 0$, thereby ruling out cases in which some linear combination of $\{ V_t \}_{t=-\infty}^{+\infty}$ has zero variance.

Under Conditions A and B, the following lemma indicates that the autocovariance sequence $\{ \Gamma(j) \}_{j=-\infty}^{\infty}$ has a VAR(∞) representation with absolutely summable coefficients. Furthermore, the VAR coefficients vanish at the same rate as the autocovariances, so that we can expect the rate of convergence of the VAR spectral estimator to be similar to that of the truncated kernel estimator.

Lemma 1: Under Conditions A and B, the limiting autocovariances $\Gamma(j)$ and spectral density function $f(\omega)$ are identical to those of a vector VAR(∞) process with i.i.d. Gaussian innovations:

$$f(\omega) = (1/2\pi) \{A[\exp(i\omega)]\}^{-1} \Sigma \{A^*[\exp(i\omega)]\}^{-1}$$

where Σ is a real symmetric positive-definite matrix; $A(z) = \sum_{j=0}^{\infty} A_j z^j$; $A^*(z)$ is the complex conjugate of $A(z)$; $A_0 = I_N$; $\sum_{j=0}^{\infty} | A_j |_\infty < +\infty$; and $\det(A(z)) \neq 0$ for $|z| \leq 1$. Furthermore, $A_\infty = G_\infty^{-1} g_\infty$ and $\Sigma = \Sigma_\infty = \Gamma(0) - g'_\infty A_\infty$. Finally, if $\sum_{j=0}^{\infty} j^\lambda |\Gamma_j| < \infty$ for $\lambda \geq 0$, then $\sum_{j=0}^{\infty} j^\lambda | A_j | < \infty$.

Even in the absence of Condition B, VAR approximation yields a consistent estimate of the spectral density at all frequencies as the lag order M increases to infinity (Fuller 1996, p. 165). Nevertheless, the convergence rate will generally be much slower than in cases where Condition B is satisfied (Grenander and Szegö 1958, p. 190).[3]

3.2. Convergence Rate of the VAR Spectral Estimator

For a sample of length T, we will consider VAR(h) approximations for $0 \leq h \leq H(T)$, where $H(T)$ indicates the maximum admissible lag order. Then the VAR(h) estimator of the spectral density at frequency zero can be expressed in terms of the sample autocovariances , defined as follows: $\tilde{\Gamma}_T(j) = (1/T) \sum_{t=H(T)+1}^{T} V_t V'_{t-j}$ for $j \geq 0$, and $\tilde{\Gamma}_T(j) = \tilde{\Gamma}'_T(-j)$ for $j < 0$. Now let $\tilde{\Gamma}_T(j-i)$ comprise the (i,j)th $N \times N$ block of the $hN \times hN$ Toeplitz matrix $\tilde{G}_{Th}$; let $-\tilde{\Gamma}_T(i)$ comprise the i-th $N \times N$ block of the $hN \times N$ matrix $\tilde{g}_{Th}$; and let the coefficient matrix $\tilde{A}'_{Th}(j)$ comprise the j-th $N \times N$ block of the $hN \times N$ matrix $\tilde{A}_{Th}$. Then $\tilde{A}_{Th}$ is determined by the OLS orthogonality conditions: $\tilde{A}_{Th} = (\tilde{G}_{Th})^{-1} \tilde{g}_{Th}$. The estimated innovation covariance matrix can be expressed as $\tilde{\Sigma}_{Th} = \sum_{j=0}^{h} \tilde{\Gamma}_T(j) \tilde{A}'_{Th}(j)$, where $\tilde{A}_{Th}(0)$ is the $N \times N$ identity matrix I_N . Finally, $f(0)$ is estimated by $\tilde{S}_{Th}^{ar} / 2\pi$, where $\tilde{S}_{Th}^{ar} = \left[\sum_{j=0}^{\infty} \tilde{A}_{Th}(j) \right]^{-1} \tilde{\Sigma}_{Th} \left[\sum_{j=0}^{\infty} \tilde{A}'_{Th}(j) \right]^{-1}$. We can simplify this notation to some extent by defining the $h \times 1$ vector q_h with all elements equal to unity, and setting the $hN \times N$ matrix $Q_h = q_h \otimes I_N$, where $\otimes$ represents the Kronecker product. Then $\tilde{S}_{Th}^{ar} = (I_N + Q'_h \tilde{A}_{Th})^{-1} \tilde{\Sigma}_{Th} (I_N + \tilde{A}'_{Th} Q_h)^{-1}$.

[3] Violation of Condition B implies that the vector MA(∞) representation given in Lemma 2 cannot be inverted into a VAR(∞) representation with absolutely summable coefficients, due to the singularity of the Toeplitz matrices G_M (for sufficiently large values of M) and of the infinite-dimensional Hankel matrix G_∞. However, the data can still be regressed on its own M lagged values using the generalized inverse of G_M (cf. Whittle 1983, pp.43-44).

To analyze the convergence rate of the VAR spectral estimator, it is useful to define VAR(h) coefficients and innovation covariance matrix that are constructed using the autocovariance sequence $\{\Gamma(j)\}_{j=-\infty}^{\infty}$. Since Condition B ensures that the smallest eigenvalue of the Toeplitz matrix G_h is bounded away from zero, we can define $A_h = G_h^{-1} g_h$, and $\Sigma_h = \sum_{j=0}^{h} \Gamma(j) A'_h(j)$, where $A'_h(j)$ comprises the j-th $N \times N$ block of A_h. Then the spectral density at frequency zero implied by the VAR(h) approximation is given by $S_h^{ar} / 2\pi$, where $S_h^{ar} = (I_N + Q'_h A_h)^{-1} \Sigma_h (I_N + A'_h Q_h)^{-1}$. Finally, using the results of Lemma 1, the spectral density at frequency zero may be expressed as $f(0) = (2\pi)^{-1}(I_N + Q'_\infty A_\infty)^{-1} \Sigma_\infty (I_N + A'_\infty Q_\infty)^{-1}$.

Now the convergence rate of $\widetilde{S}_{Th}^{ar}$ can be analysed using the following expression:

$$(3.2) \qquad |\widetilde{S}_{Th}^{ar} - 2\pi f(0)| \leq |S_h^{ar} - 2\pi f(0)| + |\widetilde{S}_{Th}^{ar} - S_h^{ar}|$$

The first term on the right-hand side of equation (3.2) will be referred to as the asymptotic bias, and only depends on the autocovariance sequence $\{\Gamma(j)\}_{j=-\infty}^{\infty}$. Since the second term measures the sampling variation in estimating the VAR(h) model, we will refer to $E|\widetilde{S}_{Th}^{ar} - S_h^{ar}|^2$ as the asymptotic variance. However, to avoid further notational complexity, this term also incorporates two asymptotically negligible sources of bias: (1) bias resulting from using the degrees of freedom correction factor $1/T$ instead of $1/(T - H(T))$ in constructing the sample autocovariances; and (2) bias resulting from the difference between $\Gamma_T(j)$ and $\Gamma(j)$. The first source of bias is asymptotically negligible when the VAR lag order $h = \mathbf{o}(T^{1/2})$, and the second source of bias is asymptotically negligible under Condition A.

By defining the asymptotic bias in terms of the autocovariance sequence $\{\Gamma(j)\}_{j=-\infty}^{\infty}$ (which corresponds to that of an AR(∞) process with i.i.d. Gaussian innovations), it is straightforward to evaluate this bias using earlier results in the literature on AR approximation and AR spectral estimation (cf. Baxter 1962; Berk 1974; Hannan and Deistler 1988). To facilitate this analysis, we explicitly consider three specific cases using the following condition:

Condition C: The autocovariances $\{\Gamma(j)\}_{j=-\infty}^{\infty}$ and spectral density $f(\omega)$ satisfy one of the following:

(i) **VAR(p) Representation**: $f(\omega) = (1/2\pi) \{A[\exp(i\omega)]\}^{-1} \Sigma \{A^*[\exp(i\omega)]\}^{-1}$, where $A(z) = \sum_{j=0}^{p} A_j z^j$ for some $0 \leq p < \infty$.

(ii) **VARMA(p,q) Representation**: $f(\omega) = (1/2\pi) \{A[\exp(i\omega)]\}^{-1} B[\exp(i\omega)] \Sigma B^*[\exp(i\omega)] \{A^*[\exp(i\omega)]\}^{-1}$, where $A(z) = \sum_{j=0}^{p} A_j z^j$ for some $0 \leq p < \infty$, $B(z) = \sum_{j=0}^{m} B_j z^j$ for some $0 < m < \infty$; the leading coefficient matrices A_p and B_m are not identically equal to zero; and ρ_0 is defined as the modulus of a zero of $B(z)$ nearest $|z| = 1$.

(iii) **Other VAR(∞) Representation**: The index r satisfies $0 < r < \infty$, where $r = \sup_{\bar{r}} \{\bar{r}: \sum_{j=1}^{\infty} j^{\bar{r}} |\Gamma(j)| < \infty\}$.

Lemma 2: Assume that the sequence $\{V_t\}$ satisfies Conditions A-C.

(a) Under Condition C(i), $|S_h^{ar} - 2\pi f(0)| = 0$ for $h \geq p$.

(b) Under Condition C(ii), $|S_h^{ar} - 2\pi f(0)| = \mathbf{O}[\rho_0^{-h}]$.

(c) Under Condition C(iii), $|S_h^{ar} - 2\pi f(0)| = \mathbf{O}[h^{-r}]$.

Next we consider the asymptotic variance of the VAR spectral estimator. Parzen (1969) conjectured that the asymptotic variance of the VAR spectral estimator is $\mathbf{O}(h/T)$, just as for kernel-based spectral estimators. This property was verified by Berk (1974) for the case of univariate AR(∞) processes with i.i.d. disturbances, under the condition that $h(T) = \mathbf{o}[T^{1/3}]$. Berk also required that the lag order grow faster than $\log(T) / 2\log\rho_0$ in the finite-order ARMA case, and faster than $T^{1/2r}$ when $r < \infty$; unfortunately, this restriction excludes the use of AIC or BIC, since these criteria yield a lag order that grows at the rate $\log(T) / 2\log\rho_0$ in the finite-order ARMA case, and at a rate approaching $T^{1/(2r+1)}$ when $r < \infty$.

To verify Parzen's conjecture under much more general conditions, we use Andrews' (1991)assumptions on the higher-order moments and strong-mixing coefficients of $\{V_t\}$; these assumptions are sufficient to ensure the absolute summability of the autocovariances (as in Condition A(1)) and the absolute summability of the fourth-order cumulants.

Condition D: The sequence $\{V_t\}_{t=-\infty}^{+\infty}$ is α-mixing. For some $\nu > 1$, the α-mixing coefficients are of size $-3\nu / (\nu - 1)$, and $\sup_t \mathbf{E}|V_t|^{4\nu} < \infty$.

As shown by Hansen (1992), Condition D implies that $\left|\widetilde{\Gamma}_T(j) - \Gamma_T(j)\right| = \mathbf{O}_p(T^{-1/2})$ uniformly in j for $0 \le j \le T-1$. Using this result, it is straightforward to determine the uniform convergence rate of the VAR coefficients:

Lemma 3: Let the sequence $\{V_t\}$ satisfy Conditions A-D. Then $\left|\widetilde{A}_{Th} - A_h\right| = \mathbf{O}_p(T^{-1/2})$ uniformly in $h(T)$ for $0 \le h(T) \le H(T) = \boldsymbol{O}(T^{1/3})$.

Lemma 3 generalizes the results of Lewis and Reinsel (1985), who considered VAR(∞) processes with i.i.d. innovations and showed that $\left|d_T'(\widetilde{A}_{Th} - A_h)\right| = \mathbf{O}_p(T^{-1/2})$ for any deterministic sequence d_T satisfying the constraint that $d_T' d_T \le c < \infty$ for all T. An et al. (1982), Hannan and Kavalieris (1983), Hannan and Deistler (1988), and Guo et al. (1990) obtained results parallel to those of Lemma 3 under different assumptions; these authors considered VAR(∞) processes with martingale-difference innovations, and showed that $|\widetilde{A}_{Th} - A_h| = \mathbf{O}[T/\log(T)]^{-1/2}$ uniformly in $h(T) \le H(T) = \mathbf{o}[T/\log(T)]^{1/2}$.

Since the variance of the VAR spectral estimator is dominated by the variance of the sum of VAR coefficients, an upper bound on the variance of $\widetilde{S}_{Th}^{ar}$ can be obtained directly from the uniform convergence rate of these coefficients. An et al. (1982) and Hannan and Kavalieris (1983) used this approach to show that $\left|\widetilde{S}_{Th}^{ar} - S_h^{ar}\right|^2 = \mathbf{O}[h^2(T)\log(T)/T]$; under the assumption that $h(T) \to \infty$ and $h(T) = \mathbf{o}[T/\log(T)]^{1/2}$, this result is sufficient to ensure the consistency of the VAR spectral estimator. However, to verify Parzen's conjecture concerning the asymptotic variance of the VAR spectral estimator, it is necessary to follow the approach of Berk (1974) in directly analysing the convergence rate of the sum of VAR coefficients.

Lemma 4: Let the sequence $\{V_t\}$ satisfy Conditions A-D. Then $E|\widetilde{S}_{Th}^{ar} - S_h^{ar}|^2 = \mathbf{O}[\, h(T) / T \,]$ uniformly in $h(T)$ for $0 \le h(T) \le H(T) = \boldsymbol{O}(T^{1/3})$.

From Lemmas 2 and 4, it can be seen that achieving the optimal convergence rate of the VAR spectral estimator involves a tradeoff in the choice of VAR lag order, since an increase in h raises the asymptotic variance and reduces the asymptotic bias (apart from the special case in which the true *dgp* is a finite-order VAR). Shibata (1979, 1980) analysed the AR lag order sequence generated by a model selection criterion like AIC or BIC when the true *dgp* is an AR(∞) process with i.i.d. Gaussian innovations, and the maximum lag order $H(T) = \mathbf{o}(T^{1/2})$. In this case, AIC yields a lag order growth rate that minimizes the mean-squared error (MSE) of all k-step-ahead forecasts and the MSE of the AR estimator of the integrated spectrum . For finite-order ARMA processes, BIC yields the same lag order growth rate as AIC; for $r < \infty$, the lag order chosen by BIC grows more slowly but asymptotically approaches the geometric rate generated by AIC. Hannan and Kavalieris (1986) obtained similar properties of AIC and BIC when the data are generated by a VAR(∞) process with conditionally homoskedastic martingale-difference innovations. [4] The following lemma indicates that these properties of BIC continue to hold in the absence of weak stationarity, because the penalty term, $h \log(T)/T$, is sufficiently large to dominate the sampling variation of the estimated innovation covariance matrix. It is likely that the lag order growth rate of AIC can be verified under more restrictive assumptions than those of Condition D (e.g., absolute summability of the sixteenth-order cumulants, as suggested by Shibata 1981, p.163), but we do not pursue this issue further here.

Lemma 5: Let the sequence $\{V_t\}$ satisfy Conditions A-D, and let $h_B(T)$ be the VAR lag order selected by BIC, where $0 \le h_B(T) \le H(T) = C_o\, T^{1/(2g+1)}$ for some $0 < C_o < \infty$ and $1 \le g < \infty$.

(a) Under Condition C(`i`), $h_B(T) = p + \mathbf{o}_p(1)$.

(b) Under Condition C(`ii`), $h_B(T) = (1 + \mathbf{o}_p(1)) \log(T) / 2 \log(\rho_0)$.

(c) Under Condition C(`iii`) with $g < r < \infty$, $h_B(T) = (1 + \mathbf{o}_p(1))\, C_1 (T / \log(T))^{1/(2r+1)}$ for some $0 < C_1 < \infty$.

(d) Under Condition C(`iii`) with $0 \le r \le g$, then $h_B(T) = (1 + \mathbf{o}_p(1))\, H(T)$.

Finally, using the results of Lemmas 1-5, the convergence properties of the VAR spectral estimator can be summarized as follows:

Theorem 1: Let the sequence $\{V_t\}$ satisfy Conditions A-D. If $h(T) \to \infty$ and $h(T) = \mathbf{O}(T^{1/3})$, then $|\, \widetilde{S}_{Th}^{ar} - 2\pi f(0) \,| = \mathbf{o}_p(1)$. If $h(T) = h_B(T)$, with $0 \le h_B(T) \le H(T) = C_o\, T^{1/3}$ for $0 < C_o < \infty$, then $\widetilde{S}_{Th}^{ar}$ has the following properties:

(a) Under Condition C(`i`), $|\, \widetilde{S}_{Th}^{ar} - 2\pi f(0) \,| = \mathbf{O}_p(T^{-1/2})$.

(b) Under Condition C(`ii`), $|\, \widetilde{S}_{Th}^{ar} - 2\pi f(0) \,| = \mathbf{O}_p[\, (T / \log(T))^{-1/2} \,]$.

---

[4] That is, $E(\varepsilon_t \varepsilon_t' \mid \mathcal{I}_{t-1}) = \Sigma$, where $\mathcal{I}_t$ is the σ-algebra of events determined by the innovations ε_s for $s \le t$. Further analysis of this case may be found in Hannan and Deistler (1988).

(c) Under Condition C(iii) with $1 < r < \infty$, $|\widetilde{S}^{ar}_{Th} - 2\pi f(0)| = \mathbf{O}_p[(T/\log(T))^{-r/(2r+1)}]$.

(d) Under Condition C(iii) with $0 \leq r \leq 1$, $|\widetilde{S}^{ar}_{Th} - 2\pi f(0)| = \mathbf{O}_p[T^{-r/3}]$.

3.3 Convergence Rate of the VARHAC Estimator.

The asymptotic properties of VAR spectral estimation given in Sections 3.1 and 3.2 can be readily extended to the case of HAC covariance matrix estimation, which typically involves the analysis of estimated regression residuals. In particular, $V_t(\psi)$ is a random N-vector for each $p\times 1$ vector of regression parameters ψ in the admissible region $\Psi \subset \Re^p$. To simplify notation in the following discussion, we will use V_t to refer to $V_t(\psi_o)$, the regression function evaluated at the true regression parameter vector ψ_o, and we will use $\hat{V}_t$ to refer to $V_t(\hat{\psi}_T)$, the regression function evaluated at the regression parameter estimate $\hat{\psi}_T$.

Thus, we continue to use $\Gamma(j)$ to refer to the limiting j-th order autocovariance evaluated at ψ_o, and $\Gamma_T(j)$ to refer to the average j-th order autocovariance matrix, as defined in equation (3.1). The matrices G_h, G_∞, g_h, g_∞, A_h, A_∞, Σ_h,, Σ_∞, S_h^{ar}, and $f(\omega)$ are as defined above, based on the limiting autocovariances evaluated at ψ_o. Similarly, $\widetilde{\Gamma}_T(j)$ refers to the sample j-th order autocovariance based on the true series $\{V_t\}$, and the matrices $\widetilde{G}_{Th}$, $\widetilde{g}_{Th}$, $\widetilde{C}_{Th}$, $\widetilde{A}_{Th}$, $\widetilde{\Sigma}_{Th}$, and $\widetilde{S}_{Th}^{ar}$ are as previously defined using the sample autocovariances $\widetilde{\Gamma}_T(j)$. Finally, $\hat{\Gamma}_T(j)$ refers to the sample j-th order autocovariance based on the estimated series $\{\hat{V}_t\}$, and the matrices $\hat{G}_{Th}$, $\hat{g}_{Th}$, $\hat{C}_{Th}$, $\hat{A}_{Th}$, and $\hat{\Sigma}_{Th}$ are constructed using the estimated autocovariances $\hat{\Gamma}_T(j)$. Then the VAR spectral estimator based on the estimated regression residuals can be expressed as

$$\hat{S}_{Th}^{ar} = [I_h + Q_h' \hat{A}_{Th}]^{-1} \hat{\Sigma}_{Th} [I_h + \hat{A}_{Th}' Q_h]^{-1}.$$

To analyze the rate at which the VAR spectral estimator $\hat{S}_{Th}^{ar}$ converges to $2\pi f(0)$, we use the following assumptions of Andrews (1991):

Condition E: The regression function $V_t(\psi)$ and the estimated regression parameter vector $\hat{\psi}_T$ satisfy:

(1) $sup_{t \geq 1} \left| E\left(V_t(\psi_o)\, V_t'(\psi_o)\right) \right| < \infty$

(2) $sup_{t \geq 1} \left| \mathrm{E}\, sup_{\psi \in \Psi} \left\{ \mathrm{vec}\left(\partial V_t(\psi)/\partial \psi'\right)\, \mathrm{vec}'\left(\partial V_t(\psi)/\partial \psi'\right) \right\} \right| < \infty$

(3) $sup_{t \geq 1} \left| \mathrm{E}\, sup_{\psi \in \Psi} \left\{ \partial\, \mathrm{vec}\left(\partial V_t(\psi)/\partial \psi'\right)/\partial \psi' \right\} \right| < \infty$

(4) Conditions A-D hold for the stochastic process $\left\{ V_t'(\psi_o),\ \mathrm{vec}\left(\partial V_t(\psi)/\partial \psi' - \mathrm{E}\left[\partial V_t(\psi)/\partial \psi'\right]\right) \right\}$

(5) $\sqrt{T}\left(\hat{\psi}_T - \psi_o\right) = O_p(1)$

This condition is sufficient to ensure that the use of estimated regression residuals does not affect the asymptotic properties of the VAR spectral estimator, as indicated by the following theorem.

Theorem 2: Let the sequence $\{\hat{V}_t\}$ satisfy Conditions A-E. If $h(T) \to \infty$ and $h(T) = \mathbf{O}(T^{1/3})$, then $|\hat{S}_{Th}^{ar} - 2\pi f(0)| = \mathbf{o}_p(1)$. If $h(T) = h_B(T)$, with $0 \leq h_B(T) \leq H(T) = C_o T^{1/3}$ for $0 < C_o < \infty$, then $\hat{S}_{Th}^{ar}$ has the following properties:

(a) Under Condition C(i), $|\hat{S}_{Th}^{ar} - 2\pi f(0)| = \mathbf{O}_p(T^{-1/2})$.

(b) Under Condition C(ii), $|\hat{S}_{Th}^{ar} - 2\pi f(0)| = \mathbf{O}_p[(T/\log(T))^{-1/2}]$.

(c) Under Condition C(iii) with $1 < r < \infty$, $|\hat{S}_{Th}^{ar} - 2\pi f(0)| = \mathbf{O}_p[(T/\log(T))^{-r/(2r+1)}]$.

(d) Under Condition C(iii) with $0 \leq r \leq 1$, $| \hat{S}_{Th}^{ar} - 2\pi f(0) | = \mathbf{O}_p[T^{-r/3}]$.

4. COMPARISON WITH KERNEL-BASED ESTIMATORS

This section compares the asymptotic and finite-sample properties of the VARHAC covariance matrix estimator with those of kernel-based estimators. In particular, Section 4.1 compares the convergence rates of these estimators, and Section 4.2 reports the results of various Monte Carlo simulation experiments.

4.1 Asymptotic Properties

For a given kernel $k(z)$: $\Re \rightarrow [-1, 1]$ and bandwidth parameter h, the corresponding spectral estimator may be represented as $S_T(k, h) = \sum_{j=1-T}^{T-1} k(j/h) \hat{\Gamma}_T(j)$.[5] It is useful to define the index $v = \min(q, r)$, where q indicates the smoothness of the kernel $k(z)$ at $z = 0$, and r indicates the smoothness of the spectral density $f(\omega)$ at $\omega = 0$. In particular, $q = \sup_\theta \{ \theta : \lim_{z \rightarrow 0} (1 - \kappa(z)) / |z|^\theta < \infty \}$, and r is defined by Condition C above. If the bandwidth parameter sequence satisfies the restriction that $h(T) \leq H(T) = C_k T^{1/(2g+1)}$ for some $0 < C_k < \infty$ and $g > 1/2$, then the asymptotic bias of the kernel-based spectral estimator is $\mathbf{O}[h(T)^{-v}]$, and the asymptotic variance is $\mathbf{O}[h(T)/T]$ (cf. Parzen 1957; Priestley 1982; Andrews 1991).[6] When $v \geq g$, the optimal bandwidth parameter sequence $h^*(T)$ grows at the rate $T^{1/(2v+1)}$, and the absolute error of the kernel-based spectral estimator is $\mathbf{O}_p[T^{-v/(2v+1)}]$; when $v < g$, then $h^*(T) = H(T)$, and the absolute error is $\mathbf{O}_p[T^{-v/(2g+1)}]$.

By comparing these properties with those of the VAR spectral estimator (cf. Theorems 1 and 2 above), it can be seen that kernel-based spectral estimators face two important limitations: the cost of ensuring a PSD spectral density matrix, and the difficulty of choosing an appropriate bandwidth parameter. The simple truncated kernel estimator assigns weight of unity to all sample autocovariances up to order h, so that the smoothness index $q = \infty$. Thus, as originally conjectured by Parzen (1969), the asymptotic bias and the asymptotic variance of the simple truncated kernel estimator are of the same order as for the VAR spectral estimator when $\nu < \infty$ (i.e., when the autocovariances do not correspond to a finite-order ARMA process). If one could choose a bandwidth parameter sequence $h(T)$ with a growth rate of $T^{1/(2r+1)}$, then the absolute error of the truncated kernel estimator would be $\mathbf{O}_p[T^{-r/(2r+1)}]$, the same order as that of the VAR spectral estimator. In practice, however, no data-dependent bandwidth selection procedure has been developed for the truncated kernel estimator (cf. Priestley 1982, pp.460-62; White 1984, p.159; Andrews 1991, p.834), whereas BIC yields a lag order sequence that approaches the optimal geometric growth rate for the VAR spectral estimator. Furthermore, the simple truncated kernel does not ensure a PSD estimate of the spectral density at frequency zero, whereas the VAR spectral estimator generates a PSD matrix by construction.

Within the class of kernels that ensure a PSD spectral density matrix, the smoothness index q cannot exceed 2, and $q = 2$ for the optimal kernel within this class, the quadratic spectral (QS) kernel. The analysis of

[5] In this case, the jth-order sample autocovariance is constructed by summing $V_t V_{t-j}'$ over the range $t = j+1$ to T instead of the range $t = H(T) + 1$ to T used in defining the VAR(h) estimator (cf. Section 3.2). However, under the condition $H(T) = \mathbf{O}(T^{1/3})$, the difference between these two definitions is $\mathbf{o}_p(T^{-1/2})$.

[6] The restriction that $g > 1/2$ is needed for the case of estimated residuals, as in HAC covariance matrix estimation; kernel-based spectral estimation with observed data only requires that $g > 0$ (cf. Andrews 1991).

Andrews (1991) and Newey and West (1994) focused on the case where $r > 2$, so that the optimal bandwidth parameter sequence grows at rate $T^{1/5}$, and the absolute error of the QS kernel estimator is $\mathbf{O}_p[T^{-2/5}]$. The VAR spectral estimator utilizes a slower lag order growth rate in this case: the lag order chosen by BIC approaches a geometric growth rate of $T^{1/(2r+1)}$, and the VAR spectral estimator converges in probability at a geometric rate approaching $T^{-r/(2r+1)}$. When $r < 2$, the QS kernel estimator could achieve the same convergence rate as the VAR spectral estimator if the bandwidth parameter sequence were optimally chosen (which would require a reasonable estimate of r). However, using the data-dependent bandwidth selection procedures proposed by Andrews (1991) or Newey and West (1994), the bandwidth parameter grows too slowly for $r < 2$, so that the absolute error of the QS kernel estimator is $\mathbf{O}_p[T^{-r/5}]$. In contrast, even for $r < 1$ (for which BIC yields the maximum lag order growth rate of $T^{1/3}$), the absolute error of the VAR spectral estimator is $\mathbf{O}_p[T^{-r/3}]$.

To understand the greater asymptotic efficiency of the VAR spectral estimator compared with PSD kernel-based estimators, it is useful to note that the VAR spectral estimator can be expressed as $\hat{S}_{Th}^{ar} = \sum_{j=-\infty}^{\infty} \hat{\Gamma}_{Th}^{*}(j)$, where the $\hat{\Gamma}_{Th}^{*}(j)$ are the autocovariances implied by the estimated VAR(h) model. The OLS orthogonality conditions ensure that $\hat{\Gamma}_{Th}^{*}(j) = \hat{\Gamma}_{h}(j)$ for $|j| \leq h$, while the implied higher-order autocovariances $\hat{\Gamma}_{Th}^{*}(j)$ decline exponentially toward zero as $j \to \infty$. Thus, the VAR spectral estimator can be viewed as a procedure that assigns weights of unity to all sample autocovariances up to order h (just as with the simple truncated kernel estimator), and ensures a PSD spectral density matrix by efficiently extrapolating the higher-order autocovariances. In contrast, kernel-based procedures guarantee a PSD estimate by assigning weights less than unity to the sample autocovariances, which incurs substantial cost in terms of asymptotic bias.

This bias differential between VAR and PSD kernel-based spectral estimators can be illustrated by considering the MA(1) process $y_t = \varepsilon_t + \phi\,\varepsilon_{t-1}$, where ε_t is i.i.d. with mean zero and variance $1/(1+\phi)^2$, so that the spectral density of y_t at frequency zero is equal to unity. Figure 1 depicts the absolute value of the bias of the VAR spectral estimator and the QS estimator as a function of h (i.e., the AR lag order or the bandwidth parameter, respectively). For $\phi = 0.7$ (Panel A), the absolute bias of the VARHAC estimator is initially greater than that of the QS estimator; however, since the AR bias shrinks more rapidly as a function of h, this bias becomes smaller than that of the QS estimator for $h \geq 14$. For $\phi = -0.7$ (Panel B), the absolute bias of the VARHAC estimator is substantially smaller than that of the QS estimator for all $h > 0$; for example, the bias differential between the two estimators
is about 1:3 when $h = 10$.

4.2. Finite-Sample Properties

In this section, Monte Carlo simulation experiments are used to compare the finite-sample properties of the VARHAC estimator with those of the QS and QS-PW(1) estimators considered by Andrews (1991) and Andrews and Monahan (1992). The QS and QS-PW(1) estimators use the quadratic spectral kernel, and the data-dependent bandwidth parameter is determined using a univariate AR(1) model for each residual; the QS-PW(1) estimator augments this procedure with first-order VAR prewhitening. We also consider a variant of the VARHAC estimator, referred to as VARHAC(AIC), in which the lag order of each VAR equation is chosen using AIC rather than BIC. In each simulation experiment, we analyse the extent to which the VARHAC and QS estimators provide accurate inferences in two-tailed *t*-tests of the significance of the estimated coefficients. In all experiments, the results are computed for sample length $T = 128$, using 10,000 replications. Additional simulation results and comparisons with other kernel-based estimators may be found in Den Haan and Levin (1997).

4.2.1 The Andrews-Monahan (1992) Experiments

The first simulation experiment utilizes the same design as in Andrews and Monahan (1992), who considered several linear regression models, each with an intercept and four regressors, and the OLS estimator $\hat{\psi}_T$ for each of these models:

$$(4.1) \qquad Y_t = X_t\psi_0 + u_t, \quad t = 1,\cdots,T \quad \text{and} \quad \hat{\psi}_T = \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}\left[\sum_{t=1}^{T} X_t Y_t'\right].$$

Andrews and Monahan (1992) considered regression models with five different types of *dgp*s for the regressors and errors: (a) homoskedastic AR(1) processes; (b) AR(1) processes with multiplicative heteroskedasticity overlaid on the errors; (c) homoskedastic MA(1) processes; (d) MA(1) processes with multiplicative heteroskedasticity overlaid on the errors; and (e) homoskedastic MA(*m*) processes with linearly declining MA parameters. A range of different parameter values is considered for each type of *dgp*. All elements of ψ_0 are equal to zero.

For each HAC covariance matrix, we perform a two-tailed *t*-test of the null hypothesis that the coefficient on the first non-constant regressor is equal to its true value. Figure 2 reports the true confidence level (at a nominal 90% confidence level) for the QS-PW(1) estimator (gray column), the VARHAC estimator (black column), and the VARHAC(AIC) estimator (white column). These results indicate that the inference accuracy of the VARHAC estimator generally matches that of the QS-PW(1) estimator, despite the fact that many of the *dgp*s in this simulation experiment might be expected to favor the latter. In the AR(1) models, for example, QS-PW(1) imposes first-order prewhitening, whereas the VARHAC and VARHAC(AIC) estimators use a model selection criterion to determine the lag order. For the MA models, the QS-PW(1) estimator consistently provides better coverage ratios than the VARHAC and ARHAC(AIC) estimators, which utilize VAR representations to approximate the true MA processes. It should be noted that the VARHAC and VARHAC(AIC) estimators yield similar coverage ratios for most of the *dgp*s under consideration, and neither estimator consistently outperforms the other.

4.2.2 MA(1) Processes with Negative Autocorrelation

Although the Andrews-Monahan (1992) experiments only considered MA(1) processes with positive autocorrelation, it is also useful to compare the performance of the VARHAC and QS estimators for MA(1) processes with negative autocorrelation. In particular, consider the following *dgp*:

(4.2) $Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$.

where $\mu = 0$, the random variable ε_t is an i.i.d. standard normal process, and the parameter θ varies from -0.1 to -0.9. For each HAC estimator, we perform a two-tailed *t*-test of the significance of the sample mean.

For each value of the MA(1) parameter θ, Table 1A reports the true confidence level (at a nominal 90% significance level) for each estimator, while Table 1B reports the average bandwidth parameter chosen by QS and QS-PW(1) and the average AR lag order chosen by VARHAC and VARHAC(AIC). The VARHAC estimators consistently provide more accurate coverage ratios than the QS or QS-PW(1) estimators, even when the MA representation is close to being non-invertible. Since the average bandwidth parameters chosen by the QS and QS-PW(1) procedures are roughly similar to the average lag orders chosen by AIC and BIC, this difference in coverage ratios can be mainly attributed to the lower bias of AR approximation compared with the QS kernel, as shown in Figure 1B for the case where $\theta = -0.7$.

This experiment also highlights the limitations of using the estimated coefficients of an arbitrary parametric model to construct the data-dependent bandwidth parameter. As shown in Table 1B, the average bandwidth parameters used by QS and QS-PW(1) are much too small in comparison with the optimal value constructed using the population moments of the true *dgp*. In particular, for a scalar MA(1) process, the bandwidth parameter sequence that minimizes the asymptotic MSE of the QS spectral estimator can be expressed as an increasing function of $\left| f''(0)/f(0) \right| = |\theta| / (1+\theta)^2 = |\rho_1| / (1 + 2\rho_1)$, where ρ_1 indicates the first-order autocorrelation of Y_t.

Thus, the optimal bandwidth parameter for an MA(1) process is fairly small for all $\rho_1 > 0$, and grows arbitrarily large as $\rho_1 \rightarrow -1/2$. In contrast, for the QS and QS-PW(1) estimators, the bandwidth parameter is determined by the estimated coefficients of an AR(1) model, for which $\left| f''(0)/f(0) \right| = 2|\rho_1| / (1-\rho_1)^2$. As a result, the data-dependent bandwidth parameter implied by the AR(1) model remains small even when the estimated value of ρ_1 approaches -1/2.

These considerations suggest that the performance of the QS estimator can be improved substantially if the data-dependent bandwidth parameter is constructed using the correct specification of the true *dgp*. For this experiment, Table 1B reports that the average value of the data-dependent bandwidth parameter is fairly close to the optimal value if the bandwidth parameter is calculated under the assumption that the *dgp* is an MA(1) rather than an AR(1), and Table 1A shows that the resulting estimator, referred to as the QS-MA(1) estimator, yields much more accurate coverage ratios than either the QS or QS-PW(1) estimators. Nevertheless, the analysis in the previous paragraph indicates that the QS-MA(1) estimator may perform very poorly if the true *dgp* is an AR(1), because

the data-dependent bandwidth parameter associated with the QS-MA(1) estimator will be much smaller than the optimal bandwidth parameter for $\rho_1 > 0$ and much larger for $\rho_1 < 0$. Since the true *dgp* is unknown in practice, one might consider employing a model selection criterion to choose an appropriate parametric model, which would then be used in constructing the data-dependent bandwidth parameter of the kernel-based spectral estimator. However, it seems equally reasonable to estimate the spectral density directly from the estimated parameters of the chosen model.

Finally, while VARHAC(AIC) yields noticeably better performance than VARHAC in this experiment, it should be noted that both AIC and BIC appear to be too conservative in choosing the AR lag order. The final column of Table 1B indicates the lag order at which the VARHAC estimator yields a confidence level closest to the nominal 90% level; this "ideal" lag order is consistently higher than that chosen by either AIC or BIC. Hall (1994) and Ng and Perron (1995) have obtained similar results in experimental studies concerning the choice of AR lag order for the augmented Dickey-Fuller unit root test. Thus, the development of alternative model selection criteria for VAR spectral estimation may be a fruitful topic for further research.

4.2.3 Higher-order Autoregressive Components

The VARHAC estimator may be viewed as generalizing the Andrews-Monahan (1992) approach, such that the order of VAR prewhitening is determined by a model selection criterion rather than being fixed *a priori*, and no kernel is applied since the prewhitened residuals are approximately uncorrelated. The advantages of model selection-based VAR prewhitening were not evident in the experiments reported in section 4.2.1: in those experiments, the AR component in the vector of residuals, V_t, was at most of order one, and of course the QS-PW(1) estimator imposes first-order prewhitening. Now we consider the following scalar AR(2) process:

$$Y_t = \mu + \frac{\phi}{2}(Y_{t-1} + Y_{t-2}) + \varepsilon_t \tag{4.3}$$

where $\mu = 0$, ε_t is an i.i.d. standard normal process, and ϕ varies from 0.3 to 0.9.

Table 2 reports the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the sample mean of Y_t. The VARHAC and VARHAC(AIC) estimators clearly outperform the QS-PW(1) estimator, even for values of ϕ as low as 0.5. Given the success of first-order prewhitening in the Andrews-Monahan (1992) experiments, it is not surprising that higher-order VAR 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 BIC (AIC) in 14% (36%), 60% (67%), 90% (77%), and 96% (78%) of all replications for parameter values equal to 0.3, 0.5, 0.7, and 0.9, respectively.

4.2.4 Multivariate Applications with Heterogeneous Components

The final set of simulation experiments document the advantages of using the VARHAC procedure when the autocovariance structure differs substantially across components of the residual vector, $V_t(\hat{\psi}_T)$. In such cases,

the VARHAC estimator permits the lag order to vary across equations in the VAR (and across the variables in each equation), whereas a kernel-based estimator must use the same bandwidth parameter for every element of $V_t(\hat{\psi}_T)$ to ensure a PSD covariance matrix (cf. Robinson 1996).

To illustrate this issue, we consider the OLS estimator for the following scalar model:

$$Y_t = \alpha X_t + \beta Z_t + u_t \tag{4.4}$$

where the random variable $X_t = 0.95\, X_{t-1} + \varepsilon_t$, the random variable ε_t is an i.i.d. standard normal process, the random variable Z_t is normally distributed with zero mean and variance equal to the variance of X_t, and $\alpha = \beta = 0$. We consider two alternative *dgp*s for the random variable u_t: (1) $u_t = v_t + \theta\, v_t$, with the parameter θ varying from -0.1 to -0.9; and (2) $u_t = \phi\, u_{t-1} + v_t$, with ϕ varying from 0.1 to 0.9; in both cases, the random variable v_t is an i.i.d. standard normal process. Thus, V_t consists of two components, one of which $(Z_t u_t)$ is serially uncorrelated, while the other $(X_t u_t)$ has the same autocovariance structure as either an MA(1) process with negative autocorrelation or an AR(1) process with positive autocorrelation. Since X_t and Z_t are independent and the two components of V_t are mutually uncorrelated at all leads and lags, both the spectral density matrix $f(0)$ and the asymptotic covariance matrix Ω are diagonal. Thus, to a first approximation, the HAC standard error of $\hat{\alpha}$ will depend on the estimated spectral density of the persistent component $(X_t u_t)$, whereas the HAC standard error of $\hat{\beta}$ will depend on the estimated spectral density of the idiosyncratic component $(Z_t u_t)$.

We use each HAC covariance matrix estimator to conduct inferences concerning the significance of α and β. To highlight the fundamental issue, we focus on the QS estimator and on the QS-MA(1) estimator defined in section 4.2.2; in the presence of higher-order autocorrelation, similar considerations would apply to the QS-PW(1) estimator. For each kernel-based estimator, the specified parametric model is used to construct the data-dependent bandwidth parameter, with equal weights on both components of V_t. For VARHAC and VARHAC(AIC), the specified model selection criterion is used to determine a separate lag order for each equation in the VAR.

Table 3 provides results for the case where u_t is an MA(1) process, and Table 4 reports the corresponding results for the case where u_t is an AR(1) process. In both cases, the VARHAC and VARHAC(AIC) estimators consistently provide more accurate coverage ratios than either the QS or QS-MA(1) estimators. Panels A and B of each table indicate the true confidence levels (at the nominal 90% confidence level) of two-tailed t-tests of α and β, respectively, while Panel C reports the average bandwidth parameter of each kernel-based estimator and the average lag order chosen by AIC and BIC for each VAR equation.

In light of the analysis in Section 4.2.2, it is not surprising that the kernel-based estimators yield inaccurate inferences concerning α when the bandwidth parameter is constructed based on an incorrect specification for the *dgp* of this component. Thus, Table 3A shows that the QS estimator generates inaccurate inferences about α when the persistent component is generated by an MA(1) process with negative autocorrelation, while Table 3B shows that the QS-MA(1) estimator yields very poor inferences about α when the persistent component is generated by an AR(1) process with positive autocorrelation.

However, even when the *dgp* is correctly specified, the performance of the kernel-based estimators is adversely affected by the constraint that the same bandwidth parameter must be used in analysing both components of V_t. When the persistent component is generated by an MA(1) process, Table 3 shows that the average bandwidth parameter chosen by QS-MA(1) is generally much smaller than the average value reported in Table 1B for the corresponding scalar process. Thus, the QS-MA(1) estimator maintains reasonably accurate coverage ratios for β (which is related to the idiosyncratic component), whereas the coverage ratios for α are substantially worse than those reported for scalar MA(1) processes in Table 1A. When the persistent component is generated by an AR(1) process, Table 4 indicates that the average bandwidth parameter chosen by the QS estimator increases sharply with the value of ϕ, leading to deteriorating accuracy of inferences concerning β.

These results are directly attributable to the construction of the data-dependent bandwidth parameter, which can be expressed as an increasing function of $| f_p''(0) | / (f_p(0) + f_i(0))$, where the subscripts *p* and *i* refer to the persistent and idiosyncratic components, respectively. Since both components have equal variance by construction, $f_p(0)$ is much smaller than $f_i(0)$ when the persistent component is generated by an MA(1) process with negative autocorrelation. In this case, the data-dependent bandwidth parameter will tend to be close to the optimal value

for the idiosyncratic component, and will be much smaller than the optimal value for the idiosyncratic component. On the other hand, when the persistent component is generated by an AR(1) process with positive autocorrelation, $f_p(0)$ is much larger than $f_i(0)$, so that the data-dependent bandwidth parameter will be much closer to the optimal value for the persistent component, and will be much larger than the optimal value for the idiosyncratic component.

In contrast, the VARHAC and VARHAC(AIC) estimators can use a different AR lag order in modelling each component of V_t. Thus, as shown in Tables 3 and 4, both BIC and AIC consistently select a low lag order for the idiosyncratic component and a substantially higher lag order for the persistent component. As a result, both estimators yield reasonably accurate inferences concerning β, while the coverage ratios for α are similar to those shown in Figure 1A (for the corresponding scalar AR(1) processes) and in Table 1A (for the corresponding scalar MA(1) processes).

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), pp.926-936.

Anderson, T.W., 1971, *The Statistical Analysis of Time Series*, Wiley Press.

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, 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.

Baxter, G. 1962, An asymptotic result for the finite predictor, *Mathematica Scandinavica* 10, pp.137-144.

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* 2, pp.489-502.

Brillinger, D.R., 1981, *Time-Series: Data Analysis and Theory* (expanded edition), Holden Day.

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, 1996, Small sample properties of GMM for business cycle analysis, *Journal of Business and Economic Statistics* 14, pp.309-327.

Davidson, J., 1995, *Stochastic Limit Theory*, Oxford University Press.

Den Haan, W.J., and A. Levin, 1997, A practitioner's guide to robust covariance matrix estimation, in *Handbook of Statistics* 15, G.S. Maddala and C.R. Rao, eds., Elsevier (Amsterdam), pp.299-342.

Eichenbaum, M.S., L.P. Hansen, and K.J. Singleton, 1988, A time series analysis of representative agent models of consumption and leisure choice under uncertainty, *Quarterly Journal of Economics* CIII, pp.51-78.

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

Fuller, W.A., 1996, *Introduction to Statistical Time Series,* 2nd edition., Wiley Press.

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).

Grenander, U., 1954, On the estimation of regression coefficients in the case of an autocorrelated disturbance, *Annals of Mathematical Statistics* 25, pp.252-272.

Grenander, U., and G. Szegö, 1958, *Toeplitz Forms and Their Applications*, Univ. of California Press.

Guo, L., D.W. Huang, and E.J. Hannan, 1990, On ARX( ) Approximation, *Journal of Multivariate Analysis* 32, pp.17-47.

Hall, A.R., 1994, Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection, *Journal of Business and Economic Statistics*, 12, pages 461-70.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton Univ. Press.

Hannan, E.J., 1970, *Multiple Time Series*, Wiley Press.

Hannan, E.J., and M. Deistler, 1988, *The Statistical Theory of Linear Systems*, Wiley Press.

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, 1984, Multivariate linear time series models, *Advances in Applied Probability* 16, pp.492-561.

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

Hannan, E.J. 1987, Rational transfer function approximation, *Statistical Science* 2, pp.135-161.

Hansen, B. E. 1992, Consistent covariance matrix estimation for dependent heterogeneous processes, *Econometrica* 60, pp.967-972.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, pp.1029-2054.

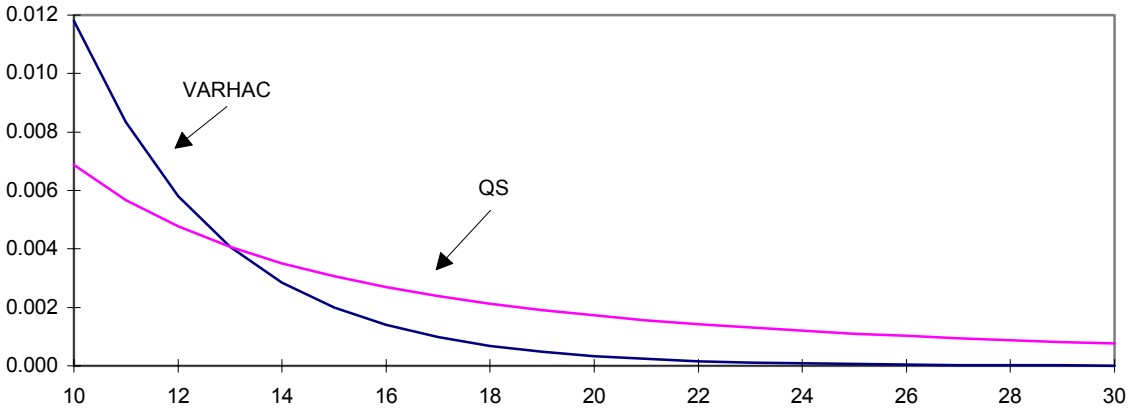
Hansen, L.P. and T.J. Sargent, 1981, Formulating and estimating dynamic linear rational expectations models, in R.E. Lucas, Jr. and T.J. Sargent, eds., *Rational Expectations and Econometric Practice*, vol. 1, University of Minnesota Press.

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

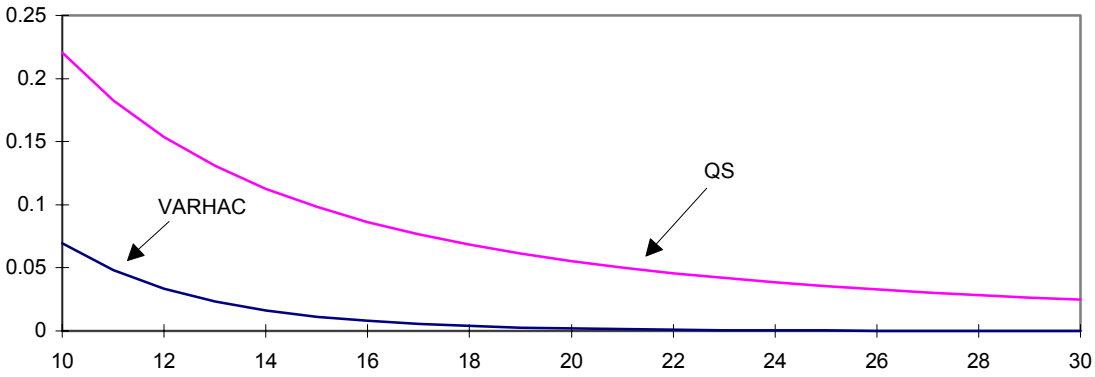
Kolmogorov, A., 1941, Interpolation und extrapolation von stationären zufälligen folgen, *Bull. Acad. Sci (Nauk), U.S.S.R., Ser. Math.* 5, pp.3-14.
Lee, C.C., and P.C.B. Phillips, 1994, An ARMA prewhitened long-run variance estimator, manuscript, Yale University.
Lewis, R., and G.C. Reinsel, 1985, Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* 16, pp.393-411.
Lütkepohl, H., 1992, *Introduction to Multiple Time Series Analysis*, Springer-Verlag.
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, *Review of Economic Studies* 61, pp.631-653.
Ng, S., and P. Perron, 1995, Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association*, 90, pages 268-81.
Nsiri, S., and R. Roy, 1993, On the invertibility of multivariate linear processes, *Journal of Time Series Analysis* 14, pp.305-316.
Parzen, E., 1957, On consistent estimates of the spectrum of a stationary time series, *Annals of Mathematical Statistics* 28, pp.329-348.
Parzen, E., 1963, Spectral analysis of asymptotically stationary time series, *Bulletin of the International Statistical Institute*, 33rd session, pp.162-179.
Parzen, E., 1983, Autoregressive spectral estimation, in *Handbook of Statistics* 3, D.R. Brillinger and P.R. Krishnaiah, eds., Elsevier Press, pp.221-247.
Priestley, M.B., 1982, *Spectral analysis and time series*, Academic Press.
Robinson, P.M., 1991, Automatic frequency domain inference on semiparametric and nonparametric models, *Econometrica* 59, pp.1329-1364.
Robinson, P.M., 1995, Inference-without-smoothing in the presence of nonparametric autocorrelation, manuscript, London School of Economics.
Robinson, P.M., and C. Velasco, 1995, Autocorrelation-robust inference, manuscript, London School of Economics.
Sargent, T.J. 1987, *Macroeconomic Theory*, 2nd edition., 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., 1980, Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Annals of Statistics* 8, pp.147-164.
Shibata, R., 1981, An optimal autoregressive spectral estimate, *Annals of Statistics* 9, pp.300-306.
Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics*, 6, pp.461-464.
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., 1997, Another heteroskedasticity and autocorrelation consistent covariance matrix estimator, manuscript, *Journal of Econometrics* 76(1-2), pp.171-191.
Wiener, N., 1949, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, Wiley Press.
White, H., 1994, *Asymptotic Theory for Econometricians*, Academic Press.
Wold, H., 1938, *The Statistical Analysis of Stationary Time Series*, Uppsala: Almquist and Wicksell Press.

Figure 1
Comparison of VARHAC vs. QS Bias for MA(1) Processes

A. MA(1) Parameter = 0.7



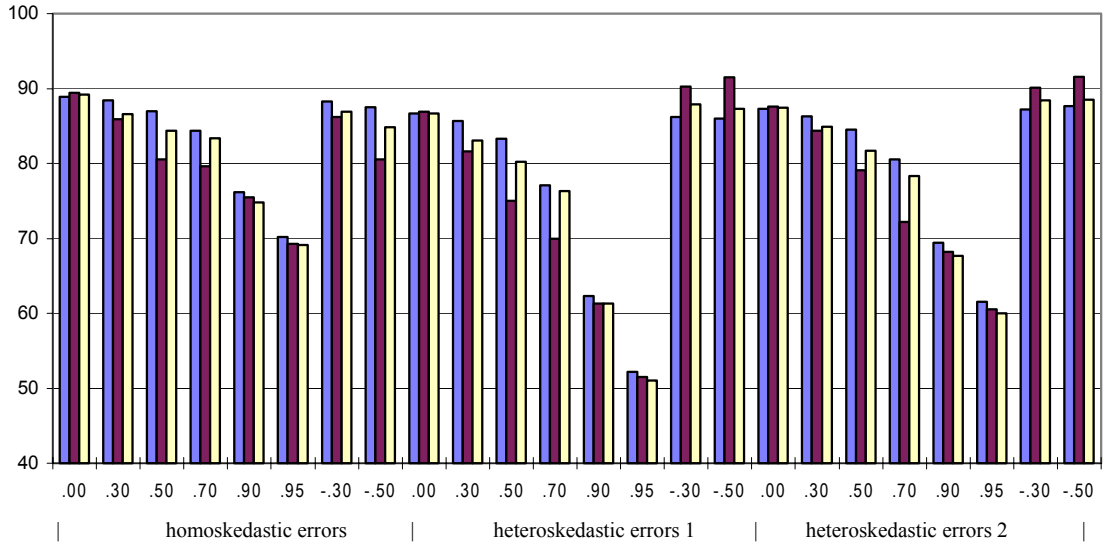
B. MA(1) Coefficient = -0.7



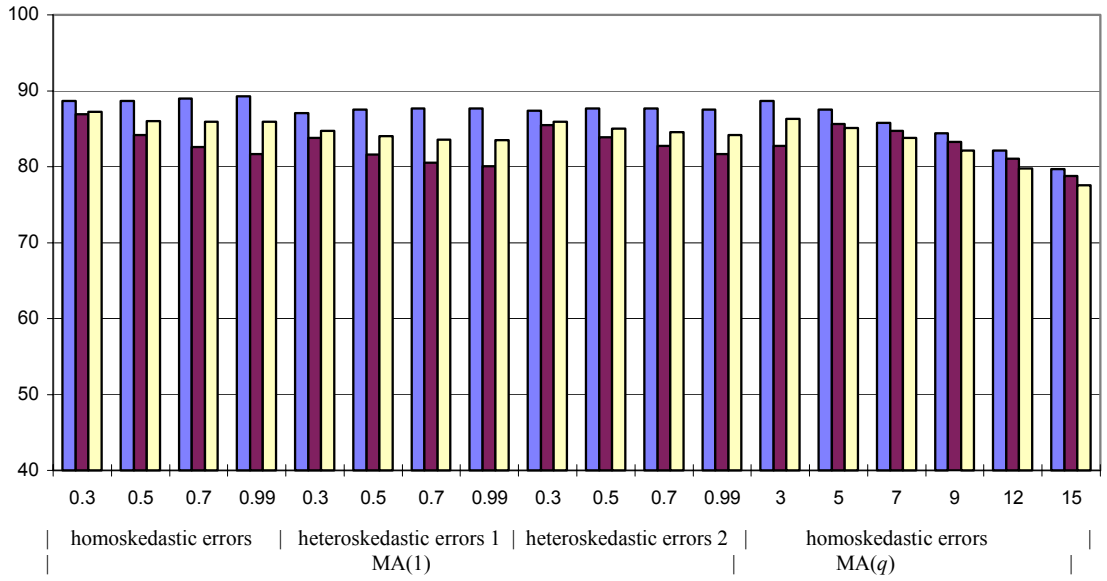
Note: Each panel indicates the absolute value of the bias (relative to the true spectral density) for the QS-estimator and the VARHAC estimator as functions of the bandwidth parameter and AR lag order, respectively. These biases are computed using the population moments of the univariate MA(1) process $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, where ε_t is an i.i.d. normal process with mean zero and variance $1/(1+\theta)^2$; thus, the true spectral density at frequency zero is equal to unity.

Figure 2
The Andrews and Monahan (1992) experiments
(True confidence level of the nominal 90% confidence interval)

Panel A: AR(1) examples



Panel B: MA examples



Note: For each *dgp* and each HAC covariance matrix, this figure reports the frequency that a two-tailed *t*-test at the nominal 90% confidence level does not reject the hypothesis that the coefficient corresponding to the first non-constant regressor is equal to its true value. This frequency is reported for the QS-PW(1) estimator (gray column); the VARHAC estimator (black column); and a variant of the VARHAC estimator that uses AIC rather than BIC (white column). The VARHAC estimator estimators are computed using a maximum lag order of 4. Panel A indicates the results for experiments in which the regressors and the error term are generated by AR(1) processes; for each experiment, the value of the AR(1) coefficient is indicated below the *x*-axis. Panel B reports the results for experiments in which the regressors and the error term are generated by MA processes; for each experiment, either the MA(1) coefficient or the order q of the MA process is indicated below the *x*-axis. The sample length $T = 128$, and the results are computed using 10,000 replications.

Table 1
MA(1) Processes with Negative Autocorrelation

A. Inferences about the significance of the mean

True confidence level of the nominal 90% confidence interval

θ	QS	QS-PW(1)	QS-MA(1)	VARHAC	VARHAC (AIC)
-0.1	90.7	89.7	90.0	89.4	88.7
-0.3	93.2	92.9	90.4	91.9	89.8
-0.5	97.0	97.3	89.7	94.1	90.9
-0.7	99.8	99.9	90.1	97.2	95.6
-0.9	100.	100.	96.9	99.9	99.9

B. Bandwidth Parameter and AR Lag Order Selection

	Average Bandwidth Parameter				Average AR Lag Order		
θ	QS	QS-PW(1)	QS-MA(1)	Optimal	BIC	AIC	"Ideal"
-0.1	1.7	0.7	2.1	2.0	1.1	1.4	1
-0.3	2.2	1.0	4.0	3.3	1.2	1.7	2
-0.5	2.4	1.5	7.3	6.1	1.7	2.5	4
-0.7	2.5	1.8	10.8	10.5	2.6	3.4	7
-0.9	2.5	2.0	11.7	27.8	3.3	3.8	12

Note: The data are generated as follows: $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, where ε_t is an i.i.d. standard normal process. The VARHAC estimator selects the AR lag order using Schwarz' Bayesian Information Criterion (BIC), while VARHAC (AIC) uses Akaike's Information Criterion; in both cases, the maximum AR lag order is equal to 4. For the QS and QS-PW(1) estimators, the data-dependent bandwidth parameter is determined by an AR(1) parametrization; for the QS-MA(1) estimator, the bandwidth parameter is determined by an MA(1) parametrization. The QS-PW(1) estimator uses AR(1) prewhitening, whereas no prewhitening is performed for the QS or QS-MA(1) estimators. The sample length $T = 128$, and the results are computed using 10,000 replications. For each value of θ and each HAC estimator, panel A shows the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the mean of Y_t. Panel B indicates the average bandwidth parameter or AR lag order chosen by each estimator. The optimal bandwidth parameter for QS is calculated using population moments and the correct specification of the *dgp*. The "ideal" lag order is the one at which the true confidence level of the VARHAC estimator is closest to 90%.

Table 2
Higher-Order Autoregressive Components

True confidence level of the nominal 90% confidence interval

ϕ	QS-PW(1)	VARHAC	VARHAC (AIC)
0.3	82.8	81.8	83.8
0.5	76.3	83.8	85.7
0.7	67.8	84.6	84.5
0.9	50.6	76.8	76.4

Note: The data are generated by $Y_t = 0.5\phi\, Y_{t-1} + 0.5\phi\, Y_{t-2} + \varepsilon_t$, where ε_t is an i.i.d standard normal random variable. The QS-PW(1) and VARHAC estimators are described at the end of Table 1. The sample length $T = 128$, and the results are computed using 10,000 replications. For each value of ϕ and each HAC estimator, this table indicates the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the mean of Y_t.

Table 3
Bivariate Applications with an MA(1) Component

A. Inferences about the significance of α

True confidence level of the nominal 90% confidence interval

θ	QS	QS-MA(1)	VARHAC	VARHAC (AIC)
-0.1	89.9	89.1	90.8	87.4
-0.3	93.5	90.9	92.4	88.2
-0.5	97.5	93.5	94.3	90.0
-0.7	99.5	96.7	96.8	93.4
-0.9	99.9	98.6	98.4	96.2

B. Inference about the significance of β

True confidence level of the nominal 90% confidence interval

θ	QS	QS-MA(1)	VARHAC	VARHAC (AIC)
-0.1	88.8	88.5	89.2	88.6
-0.3	88.7	88.1	89.1	88.5
-0.5	88.8	87.9	89.1	88.6
-0.7	88.7	88.1	89.0	88.5
-0.9	88.7	87.9	88.9	88.5

C. Bandwidth parameter and AR lag order selection

	Average Bandwidth Parameter		Average AR Lag Order			
			BIC		AIC	
θ	QS	QS-MA(1)	$X_t u_t$	$Z_t u_t$	$X_t u_t$	$Z_t u_t$
-0.1	1.7	2.0	0.2	0.01	1.1	0.4
-0.3	1.8	2.6	0.6	0.01	1.7	0.4
-0.5	1.9	3.1	1.2	0.02	2.2	0.5
-0.7	1.9	3.3	1.6	0.03	2.7	0.5
-0.9	1.9	3.4	1.9	0.03	3.0	0.5

Note: The data are generated as follows: $Y_t = \alpha X_t + \beta Z_t + u_t$, where $X_t = 0.95\, X_{t-1} + \varepsilon_t$; $u_t = v_t + \theta v_{t-1}$; $\alpha = \beta = 0$; the random variables ε_t and v_t are generated by i.i.d. standard normal processes; and the random variable Z_t is generated by a normal process with zero mean and variance equal to the variance of X_t. The QS and VARHAC estimators are described at the end of Table 1. For each VARHAC estimator, the specified model selection criterion is used to determine a separate AR lag order for each equation. The sample length T=128, and the results are computed using 10,000 replications. For each value of θ and each HAC estimator, panels A and B indicate the true confidence levels (at a nominal 90% significance level) of two-tailed t-tests of the significance of α and β, respectively. For each value of θ, panel C shows the average bandwidth parameter or AR lag order chosen by each estimator.

Table 4
Bivariate Applications with an AR(1) Component

A. Inferences about the significance of α

True confidence level of the nominal 90% confidence interval

ϕ	QS	QS-MA(1)	VARHAC	VARHAC (AIC)
0.1	87.0	86.9	85.9	85.6
0.3	84.5	83.7	83.4	84.4
0.5	82.0	79.4	84.7	83.7
0.7	78.1	70.3	83.1	81.7
0.9	69.9	52.1	76.6	75.7

B. Inference about the significance of β

True confidence level of the nominal 90% confidence interval

ϕ	QS	QS-MA(1)	VARHAC	VARHAC (AIC)
0.1	88.7	88.7	89.1	88.7
0.3	88.3	88.7	89.2	88.6
0.5	87.6	88.5	89.0	88.3
0.7	86.4	88.0	89.5	88.8
0.9	82.4	87.7	89.4	88.3

C. Bandwidth parameter and AR lag order selection

	Average Bandwidth Parameter		Average AR Lag Order			
			BIC		AIC	
ϕ	QS	QS-MA(1)	$X_t u_t$	$Z_t u_t$	$X_t u_t$	$Z_t u_t$
0.1	2.0	1.7	0.1	0.02	1.0	0.4
0.3	3.2	2.1	0.5	0.02	1.5	0.4
0.5	5.0	2.4	1.0	0.04	1.7	0.5
0.7	8.0	2.6	1.1	0.08	1.7	0.7
0.9	15.5	2.7	1.1	0.14	1.6	1.1

Note: The data are generated as follows: $Y_t = \alpha X_t + \beta Z_t + u_t$, where $X_t = 0.95\, X_{t-1} + \varepsilon_t$; $u_t = \phi\, u_{t-1} + \nu_t$; $\alpha = \beta = 0$; the random variables ε_t and ν_t are generated by i.i.d. standard normal processes; and the random variable Z_t is generated by a normal process with zero mean and variance equal to the variance of X_t. The QS and VARHAC estimators are described at the end of Table 1. For each VARHAC estimator, the specified model selection criterion is used to determine a separate AR lag order for each equation. The sample length T=128, and the results are computed using 10,000 replications. For each value of ϕ and each HAC estimator, panels A and B indicate the true confidence levels (at a nominal 90% significance level) of two-tailed t-tests of the significance of α and β, respectively. For each value of θ, panel C shows the average bandwidth parameter or AR lag order chosen by each estimator.

Proof of Lemma 1:

Condition A is sufficient to ensure that $\{V_t\}$ meets Grenander's (1954) conditions for asymptotic stationarity (cf. Hannan 1970, p.77). Condition A(2) ensures that the average *j*th-order autocovariance converges to a limiting value for each integer *j*. Condition A(4) ensures the "persistence of excitation;" i.e., the process $\{V_t\}$ has positive variance infinitely often, so that the sum of individual variances diverges to infinity. Finally, Conditions A(1) and A(2) ensure asymptotic negligibility; i.e., as the sample length *T* grows arbitrarily large, the limiting autocovariances are not affected by the exclusion of a finite number

of observations. Finally, equation (3.1) implies the symmetry condition $\Gamma_T(j) = \Gamma_T'(j)$ for all $|j| < T$.

Thus, the limiting autocovariances form a positive semi-definite sequence; i.e., $\det(G_M) \geq 0$ for all $M \geq 1$ (cf. Hannan 1970, p.77). Furthermore, the limiting autocovariances are identical to those of a weakly stationary Gaussian process (Doob 1953, Theorem X.3.1, p. 473; Ibragimov and Linnik 1971, p. 311). The absolute summability of the limiting autocovariances follows from Condition A(1), ensuring that $\Gamma(j) \rightarrow 0$ as $j \rightarrow \infty$, so that the corresponding weakly stationary process contains no purely deterministic harmonic components (cf. Priestley 1982, p.230). Given the absolute summability of the limiting autocovariances, the Riesz-Fischer Theorem indicates that the function $f(\omega) \in L_2[-\pi, \pi]$ and that $\Gamma(j) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega j)\, d\omega$ (cf. Sargent 1987, p. 249). Since the limiting autocovariances form a positive semi-definite sequence, $f(\omega)$ is a Hermitian positive semi-definite matrix function, by Theorem II.11 of Hannan (1970, p.78). Furthermore, there exists a weakly stationary Gaussian process with spectral density $f(\omega)$ (cf. Ibragimov and Linnik 1971, p. 311).

Under Condition B, the spectral density $f(\omega)$ can be factorized into a vector MA(∞) representation (cf. Wold 1938; Theorem IV.6.2 of Doob 1953, pp.160-161; Hannan 1970, pp.157-163). Condition B also implies that the MA coefficients are absolutely summable (cf. Theorems 3.8.2 and 3.8.3 of Brillinger 1981, pp.76-78), and that all roots of $\Theta(z)$ are outside the unit circle (cf. Nsiri and Roy 1993). Finally, Condition B ensures that the vector MA(∞) representation can be inverted to obtain a vector AR(∞) representation with absolutely summable VAR coefficients (cf. Nsiri and Roy (1993). Similar results may also be found in Fuller (1996, Theorems 2.8.2 and 4.4.1, pp.78-180), among many other references. The final statement of Lemma 1 follows directly from Theorems 3.8.2 and 3.8.3 of Brillinger (1981, pp.76-78).

Finally, the validity of the infinite-order Yule-Walker equations can be confirmed by noting that $\frac{1}{2\pi}\int e^{i\omega L} d\omega = 1$ for $L = 0$ and $\frac{1}{2\pi}\int e^{i\omega L} d\omega = 0$ for $L \neq 0$. Since $\Theta^*(z) = [A^*(z)]^{-1}$, we have $A(e^{i\omega}) f(\omega) = \Sigma\, \Theta^*(e^{i\omega})$, i.e.,

$$\text{(A1)} \qquad \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} A(j)\Gamma(k) e^{i\omega(j-k)} = \Sigma \sum_{L=0}^{\infty} \Theta'_L\, e^{-i\omega L}\,.$$

By multiplying both sides of (A1) by $e^{-i\omega m}$ for $m > 0$, and then integrating over $\omega \in [-\pi, \pi]$ we obtain $\sum_{j=0}^{\infty} A(j)\Gamma(j-m) = 0$, i.e. $\sum_{j=1}^{\infty} \Gamma'(j-m) A'(j) = -\Gamma'(-m) = -\Gamma(m)$. Collecting these equations

together for all $m \geq 1$, we obtain $G_\infty A_\infty = g_\infty$. By integrating both sides of (A1) over $\omega \in [-\pi, \pi]$, and dividing both sides by 2π, we obtain $\sum_{j=0}^{\infty} A(j)\Gamma(j) = \Sigma$. Since Σ is symmetric, transposing both sides yields $\Sigma = \sum_{j=0}^{\infty} \Gamma'(j)A'(j) = \Gamma(0) - g'_\infty A_\infty$.

Proof of Lemma 2:

Let $M_h = [A'_h Q_h]^{-1}$ and $M_\infty = [A'_\infty Q_\infty]^{-1}$. Then $S_h^{ar} - f(0) = M_h \Sigma_h M'_h - M_\infty \Sigma_\infty M'_\infty$

$= (M_h - M_\infty)\Sigma_h M'_h + M_\infty \Sigma_h (M'_h - M'_\infty) + M_\infty (\Sigma_h - \Sigma_\infty) M'_\infty$. Lemma 1 ensures that $\det[A'_h Q_h] \neq 0$ and that $\det[A'_\infty Q_\infty] \neq 0$, so that $|M_h| = \mathbf{O}(1)$ and $|M_\infty| = \mathbf{O}(1)$. Furthermore, the inverse function is continous at $A'_\infty Q_\infty$, so that $|M_h - M_\infty| = \mathbf{O}(|A'_h Q_h - A'_\infty Q_\infty|)$.

Thus, $|S_h^{ar} - f(0)| = \max\left(\left| A'_h Q_h - A'_\infty Q_\infty \right|, \left| \Sigma_h - \Sigma_\infty \right| \right)$.

Using the definitions of g_h, g_∞, G_h, and G_∞ given below Condition B, it is useful to define the following partitions: $G_\infty = \begin{bmatrix} G_h & G_{h+} \\ G'_{h+} & G_{h++} \end{bmatrix}, g_\infty = \begin{bmatrix} g_h \\ g_{h+} \end{bmatrix}$, and the corresponding partitions $A_\infty = \begin{bmatrix} A_{\infty h} \\ A_{\infty h+} \end{bmatrix}$ and $Q_\infty = \begin{bmatrix} Q_h \\ Q_{h+} \end{bmatrix}$. Under Conditions A and B, Hannan and Deistler (1988, p. 271) showed that $|\Sigma_h - \Sigma_\infty| = \mathbf{O}(|A'_{\infty h+}|^2)$. We will show below that $|A'_h Q_h - A'_\infty Q_\infty| = \mathbf{O}(|A'_{\infty h+}|)$, so that $|S_h^{ar} - f(0)| = \mathbf{O}(|A'_{\infty h+}|)$. This result directly implies the result given in Lemma 2(a). Lemma 2(b) follows from the property that $|A'_{\infty h+}| = \mathbf{O}(\rho_o^{-h})$ (cf. Hannan and Deistler 1988, pp.259-260). Lemma 2(c) follows directly from Lemma 1, which indicates that $|A'_{\infty h+}| = \mathbf{O}(h^{-r})$.

Since $|Q_h| = |Q_{h+}| = 1$, it should be noted that $|A'_h Q_h - A'_\infty Q_\infty| \leq |A'_h - A'_{\infty h}| + |A'_{\infty h+}|$. It can be seen that $G_h A_{\infty h} + G_{h+} A_{\infty h+} = g_h$, , so that $A'_h = g'_h G_h^{-1} = A'_{\infty h} + A'_{\infty h+} G'_{h+} G_h^{-1}$. Thus, $|A'_h - A'_{\infty h}| \leq |A'_{\infty h+}| \, |G'_{h+}| \, |G_h^{-1}|$. Condition A ensures that $|G'_{h+}| \leq \sum_{j=-\infty}^{+\infty} |\Gamma(j)|_\infty < +\infty$ for all $h > 0$. Finally, we verify that $|G_h^{-1}| = \mathbf{O}(1)$ for all $h > 0$. This result was given in Theorem 6.6.11 of Hannan and Deistler (1988, pp. 267-268) for a weakly stationary, purely non-deterministic process under the restriction that $q > 1/2$, but it is straightforward to obtain the result under more general assumptions, based on the properties of the limiting autocovariances and spectral density function.

The symmetry of G_h ensures that $G_h = U_h \Lambda_h U'_h$, where U_h is orthonormal and Λ_h is the diagonal matrix of eigenvalues of G_h. Conditions A and B ensure that the eigenvalues of G_h are uniformly bounded away from zero and infinity (cf. Theorems 9.2(a) and 9.6(a) of Grenander and Szegö 1958, pp.147-154). Thus, $|G_h^{-1}| \leq |U_h| \, |\Lambda_h^{-1}| \, |U'_h| < +\infty$ for all finite h. Thus, it only remains to verify that G_h^{-1} remains bounded in the uniform norm as $h \to \infty$. Under Condition A, the Hankel matrix $G_\infty = U_\infty \Lambda_\infty U'_\infty$, where the elements of Λ_∞ are given by the eigenvalues of $f(\omega)$ for $\omega \in [0, \pi]$. The eigenvector matrix U_∞ is determined by the values of $\sin(\omega)$ and $\cos(\omega)$ for $\omega \in [0, \pi]$, and is identical for all weakly stationary matrices with absolutely summable autocovariances (cf. Theorem

4.2.1 of Fuller 1996, p. 154). Now Theorem 1(a) indicates that $f(\omega)$ can be expressed as $f(\omega) = [A(e^{i\omega})]^{-1} \Sigma [A^*(e^{i\omega})]^{-1}$, where $\det(\Sigma) > 0$, and the vector AR(∞) coefficients $\{A(j)\}$ are absolutely summable. Thus, under Conditions A and B, G_∞^{-1} is the Hankel matrix of a weakly stationary, purely non-deterministic process with autocovariances $C(j)$ and spectral density function $g(\omega) = f^{-1}(\omega) = A^*(e^{i\omega})\ \Omega\ A(e^{i\omega})$, where $\Omega = \Sigma^{-1}$; in other words, $g(\omega)$ is the spectral density of a vector MA(∞) process, with MA coefficients $B(j) = A'(j)$. Thus, the result follows from the absolute summability of the autocovariances $C(j)$, which is implied by the absolute summability of the MA coefficients:

$$\text{(A2)} \qquad \left| G_\infty^{-1} \right| = \sum_{j=-\infty}^{\infty} \left| C(j) \right| \le \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left| B(j) \right\| \Omega \left\| B'(k) \right| < +\infty$$

Proof of Lemma 3:

Using the least-squares orthogonality conditions $\widetilde{G}_{Th} \widetilde{A}_{Th} = \widetilde{g}_{Th}$ and $G_h A_h = g_h$, it is evident that $\left| \widetilde{A}_{Th} - A_h \right| = \left| \widetilde{G}_{Th}^{-1} \right| \left| \widetilde{g}_{Th} - g_h \right| + \left| \widetilde{G}_{Th}^{-1} \right| \left| \left(\widetilde{G}_{Th} - G_h \right) A_h \right|$. Under Conditions A and D, Lemma 1 of Hansen (1992) indicates that $\mathbf{E}(\widetilde{\Gamma}_T(L) - \Gamma(L))^2 = \phi_o^2 \, T^{-1}$ uniformly in $0 \le L \le T-1$ for some $0 < \phi_o < \infty$, so that $\max_{L \in [-h,h]} |\widetilde{\Gamma}_T(L) - \Gamma(L)| = \mathbf{O}_p(T^{-1/2})$ for $h \le H(T) = \mathbf{o}(T^{1/2})$. Thus, $\left| \widetilde{g}_{Th} - g_h \right| = \mathbf{O}_p(T^{-1/2})$. Furthermore, $\left| \left(\widetilde{G}_{Th} - G_h \right) A_h \right| \le \max_k \sum_{j=1}^{h} \left| \widetilde{\Gamma}_T(j-k) - \Gamma(j-k) \right| |A_h(j)|$

$\le \max_{L \in [-h,h]} |\widetilde{\Gamma}_T(L) - \Gamma(L)| \ \sum_{j=0}^{h} |A_h(j)| = \mathbf{O}_p(T^{-1/2})$, since the proof of Lemma 2 indicates that $|A_h| = \mathbf{O}(1)$. Finally, we verify that $\left| \widetilde{G}_{Th}^{-1} \right| = \mathbf{O}_p(1)$, following essentially the same approach as in Hannan and Deistler (1988, pp.268-9). Note that $|G_h^{-1}| = \mathbf{O}(1)$ by the proof of Lemma 2.

Now $|\widetilde{G}_{Th}^{-1}| \le |G_h^{-1}| + |\widetilde{G}_{Th}^{-1} - G_h^{-1}| \le |G_h^{-1}| + |\widetilde{G}_{Th}^{-1}||\widetilde{G}_{Th} - G_h||G_h^{-1}|$. Thus, we find that

$$\left| \widetilde{G}_{Th}^{-1} \right| \le \frac{|G_h^{-1}|}{1 - |\widetilde{G}_{Th} - G_h||G_h^{-1}|} = \mathbf{O}_p(1).$$

Proof of Lemma 4:

Following the same approach used at the beginning of the proof of Lemma 2 , it can be seen that:

$$\text{(A3)} \qquad |\widetilde{S}_h^{ar} - S_{\boldsymbol{h}}^{\boldsymbol{ar}}| = \max\left(\left| \left(\widetilde{A}_{Th}' - A_h' \right) Q_h \right|, \left| \widetilde{\Sigma}_{Th} - \Sigma_h \right| \right).$$

Now it is useful to define $\widetilde{C}_{Th} = \widetilde{G}_{Th}(\widetilde{A}_{Th} - A_h)$, so that:

$$\text{(A4)} \qquad \left(\widetilde{A}_{Th}' - A_h' \right) Q_h = \widetilde{C}_{Th}' \left[\widetilde{G}_{Th}^{-1} - G_h^{-1} \right] Q_h + \widetilde{C}_{Th}' \, G_h^{-1} \, Q_h.$$

Thus, since $|Q_h| = 1$, we find that:

(A5) $\left|(\widetilde{A}'_{Th} - A'_h)Q_h\right| \le \left|\widetilde{C}'_{Th}\right|\left|\widetilde{G}_{Th}^{-1} - G_h^{-1}\right| + \left|\widetilde{C}'_{Th}\, G_h^{-1}\, Q_h\right|$

Lemma A1 shows that $\left|\widetilde{C}'_{Th}\right| = \mathbf{O}_p(hT^{-1/2})$, while Lemma A2 shows that $\left|\widetilde{G}_{Th}^{-1} - G_h^{-1}\right| = \mathbf{O}_p(hT^{-1/2})$. Thus, the product of these terms is $\mathbf{O}_p(h^2T^{-1}) = \mathbf{o}_p[(h/T)^{1/2}]$ for $h = \mathbf{o}(T^{1/3})$. Lemma A3 shows that $\left|\widetilde{C}'_{Th}\, G_h^{-1}\right| = \mathbf{O}_p[(h/T)^{1/2}]$. Thus, using equation (A5), we find that $\left|(\widetilde{A}'_{Th} - A'_h)Q_h\right| = \mathbf{O}_p[(h/T)^{1/2}]$. Finally, Lemma A4 shows that $\left|\widetilde{\Sigma}_{Th} - \Sigma_h\right| = \mathbf{O}_p[T^{-1/2}]$. Substituting these results into equation (A3) completes the proof.

Lemma A1: Under Conditions A - D, $\left|\widetilde{C}'_{Th}\right| = \mathbf{O}_p(hT^{-1/2})$ uniformly in $0 \le h(T) \le H(T) = \mathbf{O}(T^{1/3})$.

Proof: Note that $\widetilde{c}_{Th}(k) = \sum_{j=0}^{h}(\Gamma(j-k) - \widetilde{\Gamma}_T(j-k))A'_h(j)$ comprises the k-th $N\times N$ block of $\widetilde{C}_{Th}$. Just as in the proof of Lemma 3, we find that $\max_k|\widetilde{c}_{Th}(k)| = \mathbf{O}_p(T^{-1/2})$. Thus, $|\widetilde{C}'_{Th}| = \sum_{k=1}^{h}|\hat{c}_{Th}(k)| = \mathbf{O}_p(hT^{-1/2})$ uniformly in $h \le H(T) = \mathbf{O}(T^{1/3})$.

Lemma A2: Under Conditions A - D, $\left|\widetilde{G}_{Th}^{-1} - G_h^{-1}\right| = \mathbf{O}_p(hT^{-1/2})$ uniformly in $0 \le h(T) \le H(T) = \mathbf{O}(T^{1/3})$.

Proof: First, note that $\left|\widetilde{G}_{Th} - G_h\right| \le \max_k \sum_{j=1}^{h}\left|\widetilde{\Gamma}_T(j-k) - \Gamma(j-k)\right| = \mathbf{O}_p(hT^{-1/2})$; furthermore, $|G_h^{-1}| = \mathbf{O}(1)$ by the proof of Lemma 2, and $\left|\widetilde{G}_{Th}^{-1}\right| = \mathbf{O}_p(1)$ by the proof of Lemma 3. Thus, since $\widetilde{G}_{Th}^{-1} - G_h^{-1} = -\widetilde{G}_h^{-1}(\widetilde{G}_{Th} - G_h)G_h^{-1}$, we find $|\widetilde{G}_{Th}^{-1} - G_h^{-1}| = |\widetilde{G}_h^{-1}||\widetilde{G}_{Th} - G_h||G_h^{-1}| = \mathbf{O}_p[hT^{-1/2}]$.

Lemma A3: Under Conditions A - D, $\left|\widetilde{C}'_{Th}\, G_h^{-1}\, Q_h\right| = \mathbf{O}_p[(h/T)^{1/2}]$ uniformly in $0 \le h(T) \le H(T) = \mathbf{O}(T^{1/3})$.

Proof: We begin with a detailed proof for the scalar case, in which the intuition is not obscured by the additional notation required in the multivariate case. In the scalar case, $Q_h = q_h$, the $h\times 1$ vector with all elements equal to unity. Let $F_h(i,j)$ denote the (i,j)th element of G_h^{-1}, and let $z_h = G_h^{-1}q_h$, so that $z_h(j) = \sum_{k=1}^{h}F_h(j,k)$. Then $\left|z_h(j)\right| < +\infty$ for all k=1, $\cdots$, h, because $|G_h^{-1}|_\infty < +\infty$ (i.e., the rows and columns of G_h^{-1} are absolutely summable, as indicated in Lemma 2(a) above). Now we have:

$$\text{(A6)}\quad \widetilde{c}'_{Th}\, z_h = \sum_{j=1}^{h} z_h(j)\widetilde{c}_{Th}(j) = \sum_{j=1}^{h}\sum_{k=1}^{h} z_h(j)\left(\widetilde{\gamma}_T(k-j) - \gamma(k-j)\right)\alpha_h(k)$$

$$= \sum_{L=1-h}^{h}\left(\widetilde{\gamma}_T(L) - \gamma(L)\right)\left[\sum_{j=\max(0,\,1-L)}^{\min(h,\,h-L)} z_h(j)\alpha_h(j+L)\right]$$

$$= \sum_{L=1-h}^{h} b_L\left(\widetilde{\gamma}_T(L) - \gamma(L)\right)$$

where $|b_L| < +\infty$, since $|z_{hj}| < +\infty$ and $\sum_{j=0}^{\infty}|\alpha_h(j)| < +\infty$. Using the appropriate extension

of Bartlett's (1946) result (cf. Priestley 1982, p. 326; Andrews 1991), we find that:

$$\text{(A7)} \qquad \mathbf{E}\left(\widetilde{c}'_{Th}\, z_h \right)^2 = \sum_{k=1-h}^{h} \sum_{L=1-h}^{h} b_k \; b_L \; \mathbf{Cov}\left(\widetilde{\gamma}_T(k), \widetilde{\gamma}_T(L) \right)$$

$$\leq \sum_{k=1-h}^{h} \sum_{L=1-h}^{h} b_k \; b_L \left[\frac{1}{T} \sum_{m=-\infty}^{\infty} \left(\gamma(m)\,\gamma(m+L-k) + \gamma(m+L)\,\gamma(m-k) + \sup_{t \geq 1} K_4(t,k,L,m) \right) \right]$$

$$\leq \mathbf{O}\left(\frac{1}{T}\right) \sum_{k=-h}^{h} \sum_{L=-h}^{h} \left(\max_j \left| \gamma(j+L-k) \right| + \max_j \left| \gamma(j+L) \right| \right) \left[\sum_{m=-\infty}^{\infty} \left| \gamma(m) \right| \right]$$

$$+ \; \mathbf{O}\left(\frac{1}{T}\right) \sum_{K=-\infty}^{\infty} \sum_{L=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sup_{t \geq 1} \left| K_4(t,k,L,m) \right|$$

where K_4 is the fourth-order cumulant. Now Condition A ensures that the autocovariances are absolutely summable, and Condition D ensures that the fourth-order cumulants are absolutely summable (cf. Lemma 1 of Andrews 1991). Thus, we find that $\mathbf{E}\left(\widetilde{c}'_{Th}\, z_h \right)^2 = \mathbf{O}(h/T)$, so that $\left| \widetilde{c}'_{Th}\, G_h^{-1}\, q_h \right| = \mathbf{O}_p[\,(h/T)^{1/2}\,]$ in the scalar case.

In the multivariate case, let the $N \times hN$ matrix $z_h = G_h^{-1} Q_h$ with the k^{th} $N \times N$ block denoted by $z_h(k)$. As above, $|\, z_h(k) \,|_\infty < +\infty$, because $|\, G_h^{-1} \,|_\infty < +\infty$. Now we have:

$$\text{(A8)} \qquad \widetilde{C}'_{Th}\, G_h^{-1}\, Q_h = \sum_{k=1}^{h} \sum_{j=0}^{h} z_h(k) \left(\Gamma(j-k) - \widetilde{\Gamma}_T(j-k) \right) A'_h(j)$$

$$= \sum_{L=1-h}^{h} \sum_{j=\max(0,1-L)}^{\min(h,h-l)} z_h(j+L) \left(\Gamma(L) - \widetilde{\Gamma}_T(L) \right) A'_h(j)$$

By the properties of the *vec* operator (cf. Lütkepohl 1992, p.464), we find that:

$$\text{(A9)} \qquad vec\left(\widetilde{C}'_{Th}\, G_h^{-1}\, Q_h \right) = \sum_{L=1-h}^{h} \left[\sum_{j=\max(0,1-L)}^{\min(h,h-L)} A_h(j) \otimes z_h(j+L) \right] vec\left(\Gamma(L) - \widetilde{\Gamma}_T(L) \right)$$

$$= \sum_{L=1-h}^{h} b_L \; vec\left(\Gamma(L) - \widetilde{\Gamma}_T(L) \right),$$

where each $N^2 \times N^2$ matrix b_L satisfies $|\, b_L \,|_\infty \leq \max_k |\, z_h(k) \,|_\infty \sum_{j=0}^{h} |\, A_h(j) \,|_\infty < +\infty$.

Let $w_h(j)$ be the j-th element of $vec\left(\widetilde{C}'_{Th}\, G_h^{-1}\, Q_h \right)$, and let $b_L(j)$ denote the j-th row of b_L, so that $w_h(j) = \sum_{L=1-h}^{h} b_L(j)\, vec(\Gamma(L) - \widetilde{\Gamma}_T(L))$. Then we obtain:

$$\text{(A10)} \qquad \mathbf{E}\left(w_h^2(j) \right) = \sum_{K=1-h}^{h} \sum_{L=1-h}^{h} b'_k(j)\, \mathbf{E}\left[vec\left(\Gamma(K) - \widetilde{\Gamma}_T(K) \right) vec\left(\Gamma(L) - \widetilde{\Gamma}_T(L) \right)' \right] b_L(j).$$

Now this formula can be expressed in terms of the autocovariances and fourth-order cumulants, using equation (3.3) of Hannan (1970, p. 209). Then, given the absolute summability of the autocovariances and fourth-order cumulants (cf. Lemma 1 of Andrews 1991), we find that $\mathbf{E}[\, w_h(j)^2 \,] = \mathbf{O}[\, (h/T)^{1/2} \,]$ for all $j = 1, \cdots, N^2$, so that $\left| \widetilde{C}'_{Th}\, G_h^{-1}\, Q_h \right| = \mathbf{O}_p[\, (h/T)^{1/2} \,]$.

Lemma A4: Under Conditions A - D, $\left| \widetilde{\Sigma}_{Th} - \Sigma_h \right| = \mathbf{O}_p(T^{-1/2})$ uniformly in $0 \le h(T) \le H(T) = \mathbf{O}(T^{1/3})$.

From the definitions given in the text, it can be seen that $\widetilde{\Sigma}_{Th} = \widetilde{\Gamma}_T(0) - \widetilde{A}'_{Th}\widetilde{g}_{Th}$ and that $\Sigma_h = \Gamma(0) - A_h' g_h$. These identities, together with the orthogonality conditions $\widetilde{G}_{Th}\widetilde{A}_{Th} = \widetilde{g}_{Th}$ and $G_h A_h = g_h$ imply that:

(A11) $$\widetilde{\Sigma}_{Th} - \Sigma_h = -(\widetilde{A}'_{Th} - A'_h)\widetilde{G}_{Th}(\widetilde{A}_{Th} - A_h) + \left\{ \left(\widetilde{\Gamma}_T(0) - \Gamma(0)\right) - 2\, A'_h(\widetilde{g}_{Th} - g_h) + A'_h(\widetilde{G}_{Th} - G_h)A_h \right\}$$

In analysing the right-hand side of equation (A11), we will show that the first term vanishes at rate $\mathbf{O}_p(h/T)$; thus, this term is $\mathbf{o}_p(T^{-1/2})$ under the restriction that $h \le H(T) = \mathbf{O}(T^{1/3})$. We will also show that the terms enclosed in braces are $\mathbf{O}_p(T^{-1/2})$. In particular, these terms can be expressed as the sum of two components: the first component vanishes at rate $\mathbf{O}_p(T^{-1/2} |\, \Sigma_h - \Sigma \,|^{1/2})$, while the second component vanishes at rate $\mathbf{O}_p(T^{-1/2})$ and does not depend on h. The particular convergence rates of these terms will be important in verifying the properties of BIC in Lemma 5. Since all of these results are uniform in h for $0 \le h \le H(T) = \mathbf{O}(T^{1/3})$, the conclusion of the lemma follows.

Lemma 3 shows that $\left| \widetilde{A}_{Th} - A_h \right| = \mathbf{O}_p(T^{-1/2})$; using this result, we also find that $\left| \widetilde{A}'_{Th} - A'_h \right| = \sum_{j=1}^{h} \left| \widetilde{A}_{Th}(j) - A_h(j) \right| = \mathbf{O}_p(hT^{-1/2})$. Finally, $\left| \widetilde{G}_{Th} \right| \le \left| \widetilde{G}_{Th} - G_h \right| + \left| G_h \right| = \mathbf{O}_p(1)$, since Condition A ensures that $\left| G_h \right| = \mathbf{O}(1)$, and the proof of Lemma A2 indicates that $\left| \widetilde{G}_{Th} - G_h \right| = \mathbf{O}_p(hT^{-1/2}) = \mathbf{O}_p(1)$ under the restriction that $h \le H(T) = \mathbf{O}(T^{1/3})$. Thus, the first term on the right-hand-side of equation (A11) is $\mathbf{O}_p(h/T)$ as previously indicated.

Now we analyse the terms of equation (A11) that are enclosed in braces. As indicated in the proof of Lemma 3, Conditions A and D ensure that $max_{j=1,\ldots,T} |\widetilde{\Gamma}_T(L) - \Gamma(L)| = \phi_o T^{-1/2}$ for some $0 < \phi_o < \infty$. Thus, $\left| \widetilde{\Gamma}_T(0) - \Gamma(0) \right| = \mathbf{O}_p(T^{-1/2})$. In considering the second term enclosed in braces, we see that $\left| A'_h(\widetilde{g}_{Th} - g_h) \right| \le \phi_o T^{-1/2} \sum_{j=1}^{h} \left| A_h(j) \right|$. Thus, we find that:

(A12) $$\left| A'_h(\widetilde{g}_{Th} - g_h) \right| \le \phi_o T^{-1/2} \sum_{j=1}^{\infty} \left| A(j) \right| + \phi_o T^{-1/2} \left\{ \sum_{j=1}^{h} \left| A_h(j) - A(j) \right| - \sum_{j=h+1}^{\infty} \left| A(j) \right| \right\}$$

Using the results of Lemma 1, we find that the first term in equation (A12) is $\mathbf{O}_p(T^{-1/2})$, and does not depend on h. Using the proof of Lemma 2, we find that the second term in equation (A12) converges at rate $\mathbf{O}_p(T^{-1/2} |\, \Sigma_h - \Sigma \,|^{1/2})$, uniformly in h for $0 \le h \le H(T) = \mathbf{O}(T^{1/3})$. Following the same approach, it is straightforward to show that $\left| A'_h(\widetilde{G}_{Th} - G_h)A_h \right|$ can be expressed in terms of two components, where the first component is $\mathbf{O}_p(T^{-1/2})$ and does not depend on h, and the second component is

$\mathbf{O}_p(T^{-1/2} \mid \Sigma_h - \Sigma \mid^{1/2})$, uniformly in h for $0 \le h \le H(T) = \mathbf{O}(T^{1/3})$.

Proof of Lemma 5: Using the results of Lemmas 2 and 4 concerning the behavior of $\left|\Sigma_h - \Sigma\right|$ and $\left|\widetilde{\Sigma}_{Th} - \Sigma_h\right|$, the proof of Lemma 5 esentially follows the approach of Shibata (1980) and Hannan and Deistler (1988, pp.333-4). In particular, let $h_B(T)$ be the VAR lag order that minimizes $BIC(h; T) = log(det(\widetilde{\Sigma}_{Th})) + hN^2 log(T)/T$ subject to the constraint that $h_B(T) \le H(T) = C_o T^{1/(2g+1)}$ for some $0 < C_o < \infty$ and $1 \le g < \infty$. Based on the proof of Lemma 4, we see that $BIC(h;T) = log(det(\Sigma_h)) + hN^2 log(T)/T + \mathbf{O}_p(T^{-1/2})$. Let $\lambda_h = log(det(\Sigma_h)) - log(det(\Sigma_{h+1}))$. If $\lambda_h > 0$, then there exists some T_h such that $BIC(h+1; T) < BIC(h;T)$ for all $T \ge T_h$. Thus, if the limiting autocovariances correspond to a VAR of order p, as in Lemma 5(a), then $\text{limsup}(h_B(T)) \ge p$; and if the limiting autocovariances correspond to a VAR(∞), as in Lemma 5(b-d), then $h_B(T) \to \infty$ as $T \to \infty$. Therefore, it is clear that $\mid \Sigma_h - \Sigma \mid = \mathbf{o}(1)$.

Now we show that the behavior of $h_B(T)$ is effectively determined by $h^*(T)$, where $h^*(T)$ minimizes $L_{BIC}(h; T) = trace(\Sigma^{-1}(\Sigma_h - \Sigma)) + N^2 h\, log(T)/T$. We follow Hannan (1970, p.158) in defining the matrix functions $exp(\cdot)$ and $log(\cdot)$: for any Hermitian positive semi-definite matrix B, there is a unique Hermitian matrix C such that $B = exp(C)$, where $exp(C)$ is defined via the exponential series $\sum_{j=0}^{\infty} B^j / j!$, and $C = log(B)$. Using these definitions, it can be shown that $log(det(C)) = trace(log(C))$ for any positive semi-definite matrix C (cf. Hannan 1970, p159), and that $log(I + C) \to C$ as $C \to 0$ (cf. Hannan and Kavalieris 1986).

Now the proof of Lemma 4 indicates that:

$$\text{(A13)} \quad \widetilde{\Sigma}_{Th} = \Sigma\left[I + \Sigma^{-1}(\Sigma_h - \Sigma) + \mathbf{O}_p(h/T) + \mathbf{O}_p(T^{-1/2}\left|\Sigma_h - \Sigma\right|^{1/2}) + \mathbf{O}_p(T^{-1/2}) \right]$$

where I is the NxN identity matrix, and the last term in equation (A13) does not depend on the lag order h. (Conditions A and B ensure the invertibility of Σ.) Thus, using the properties of the matrix functions described above, we find that:

$$\text{(A14)} \quad BIC(h; T) = L_{BIC}(h; T) + log(det(\Sigma)) + \mathbf{O}_p(h/T) + \mathbf{O}_p(T^{-1/2}\left|\Sigma_h - \Sigma\right|^{1/2}) + \mathbf{O}_p(T^{-1/2})$$

The second and last terms on the right-hand-side of equation (A14) do not depend on h. The third term is clearly $\mathbf{o}_p[\, h\, log(T)/T \,]$.

Finally, we verify that the fourth term is $\mathbf{o}_p[L_{BIC}(h; T)]$ uniformly in h for $0 \le h \le H(T) = \mathbf{O}(T^{1/3})$. If the limiting autocovariances correspond to a VAR of order p, as in Lemma 5(a), this term vanishes, since $\Sigma_h = \Sigma$ for $h \ge p$. If the limiting autocovariances correspond to a finite-order VARMA, as in Lemma 5(b), then $\mid \Sigma_h - \Sigma \mid^{1/2} = \mathbf{O}[\, \rho_o^{-h} \,] = \mathbf{O}[\, exp\{- log(\rho_o)\, h\,\}]$. If $h = log(T) / [2(1+\varepsilon)\, log(\rho_o)]$ for some $\varepsilon > 0$, then $T^{-1/2} = exp\{(1+\varepsilon)\, log(\rho_o)\, h\}$, and the fourth term is $\mathbf{O}_p[\, exp\{-(2+\varepsilon)\, log(\rho_o)\, h\}] = \mathbf{O}_p[\left|\Sigma_h - \Sigma\right| \rho_o^{-\varepsilon h}]$
$= \mathbf{o}_p[\, L_{BIC}(h; T)\,]$. If $h = log(T) / [2(1-\varepsilon)\, log(\rho_o)]$ for some $\varepsilon \ge 0$, then $exp\{- log(\rho_o)\, h\} = T^{-1/2(1-\varepsilon)}$ $= \mathbf{O}[T^{-1/2}]$, and the fourth term is $\mathbf{O}_p[\, T^{-1} \,] = \mathbf{o}_p[\, h\, log(T)/T \,] = \mathbf{o}_p[\, L_{BIC}(h; T)\,]$. If the limiting autocovariances do not correspond to a finite-order VARMA, as in Lemma 5(c-d), then $\mid \Sigma_h - \Sigma \mid^{1/2}$ $= \mathbf{O}[\, h^{-r} \,]$. If $h = T^{1/(2r+2\varepsilon)}$ for $\varepsilon > 1/2$, then $T^{-1/2} = h^{-r-\varepsilon}$, and the fourth term is $\mathbf{O}_p[\, h^{-2r-\varepsilon} \,]$

$= \mathbf{o}_p[\, L_{BIC}(\, h;\, T\,)\,]$. If $h = T^{1/(2r+2\varepsilon)}$ for $\varepsilon \leq 1/2$, then $h^{-(r+1)} = T^{-(r+1)/(2r+2\varepsilon)} = \mathbf{O}[\, T^{-1/2}\,]$, and the fourth term is $\mathbf{O}_p[\, h / T\,] = \mathbf{o}_p[\, h\, log(T)/\, T] = \mathbf{o}_p[\, L_{BIC}(\, h;\, T\,)\,]$.

Thus, BIC($h;\, T$) can be expressed as follows, uniformly in h for $0 \leq h \leq H(T) = \mathbf{O}(T^{1/3})$:

(A15) $\quad BIC(h;\, T) = L_{BIC}(\, h;\, T\,) + \mathbf{o}_p[L_{BIC}(\, h;\, T\,)] + C_o + \mathbf{O}_p[\, T^{-1/2}\,]$

where the last two terms in equation (A15) do not depend on h. Using this equation, the remainder of the proof is identical to that of Hannan and Deistler (1988, pp.333-4).

Proof of Theorem 2: Using equation (3.2), the results follow directly from Lemmas 2 through 5.

Proof of Theorem 3: The results of Lemmas 1 and 2 remain unchanged. Now let the j-th sample autocovariance deviation $d_T(j) = vec(\hat{\Gamma}_T(j) - \tilde{\Gamma}(j))$, and let the weighted sum of sample autocovariance deviations $D_{Th} = \sum_{j=-h}^{h} b'(j)\, d_T(j)$, where the elements of each $N^2 \times 1$ non-random weighting vector $b(j)$ are uniformly bounded for all $j \in [-\infty, +\infty]$. Under Conditions A - E, Andrews (1991) has shown that each element of $d_T(j)$ converges to zero at rate $O_p(T^{-1/2})$, and that D_{Th} converges to zero at rate $\mathbf{o}_p[(h(T)/T)^{1/2}]$ for $h(T) = \mathbf{o}(T^{1/2})$. Since the proofs of Lemmas 3 through 5 have been expressed in terms of weighted sums of sample autocovariance deviations from the limiting autocovariances, these proofs can be immediately extended to the case of estimated regression residuals, in which the extra terms are asymptotically negligible, leading directly to the conclusions of Theorem 3.