TECHNICAL WORKING PAPER SERIES

INFERENCES FROM PARAMETRIC AND NON-PARAMETRIC COVARIANCE MATRIX ESTIMATION PROCEDURES

Wouter J. den Haan
Andrew Levin

Technical Working Paper 195

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1996

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, Jim Stock, P.A.V.B. Swamy, George Tauchen, Hal White, and participants at the December 1994 meetings of the European Conference on Quantitative Methods in Economics and Econometrics. This project is supported by NSF grant SBR-9514813. GAUSS, RATS, and Fortran programs to calculate the VARHAC estimator can be found on the web-site: http://weber.ucsd.edu/~wdenhaan. This paper is part of NBER's research program in Economic Fluctuations and Growth. Any opinions expressed are those of the authors and not those of the Board of Governors of the Federal Reserve System, any other members of its staff, or the National Bureau of Economic Research.

NBER Technical Working Paper 195
May 1996

INFERENCES FROM PARAMETRIC AND NON-PARAMETRIC COVARIANCE MATRIX ESTIMATION PROCEDURES

ABSTRACT

In this paper, we propose a parametric spectral estimation procedure for constructing heteroskedasticity and autocorrelation consistent (HAC) covariance matrices. We establish the consistency of this procedure under very general conditions similar to those considered in previous research, and we demonstrate that the parametric 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, our Monte Carlo experiments indicate that the parametric estimator matches, and in some cases greatly exceeds, the performance of the prewhitened kernel estimator proposed by Andrews and Monahan (1992). These simulation experiments illustrate several important limitations of non-parametric HAC estimation procedures, and highlight the advantages of explicitly modeling the temporal properties of the error terms.

Wouter J. den Haan
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
and NBER

Andrew Levin
International Finance Division
Federal Reserve Board of Governors
20th Street and Constitution Avenue
Washington, DC 20551

1. INTRODUCTION.

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

Emmanuel Parzen (1983)

Over the past decade, the use of heteroskedasticity and autocorrelation consistent (HAC) covariance matrices has become relatively common in drawing inferences from parameter estimates, since in many structural economic or time-series models, the errors may have heteroskedasticity and temporal dependence of unknown form. The key step in constructing a HAC covariance matrix is to estimate the spectral density matrix at frequency zero of a vector of residual terms. Thus, analytical results in the spectral density estimation literature (e.g. Parzen 1957; Priestley 1982) have contributed to the rapid development of new HAC covariance matrix estimation procedures (e.g., White 1984; Gallant 1987; Newey and West 1987, 1994; Gallant and White 1988; Andrews 1991; Robinson 1991; and Andrews and Monahan 1992), but with an almost exclusive focus on non-parametric kernel-based spectral methods.[1] However, autoregressive spectral estimation has been utilized in formulating unit root test statistics (Fuller 1976), and simulation evidence suggests that this approach yields superior size properties compared with unit root test statistics based on non-parametric spectral estimators (cf. Schwert 1987).[2] Furthermore, the AR(1) prewhitening technique introduced by Andrews and Monahan (1992) has enhanced the accuracy of kernel-based procedures in a variety of Monte Carlo experiments, leading Newey and West (1994) to conclude that '*...a priority [in econometric research] is theoretical and empirical investigation of autoregressive or autoregressive-moving average spectral estimators.*'

In this paper, we propose a parametric spectral estimation procedure for constructing HAC covariance matrices. This procedure utilizes standard estimation and model selection techniques in the time domain to construct a time series model for the vector of residual terms. In particular, we focus on vector autoregressive (VAR) models, and we use Schwarz' (1978) Bayesian Information Criterion (BIC) to select the appropriate lag structure. Then a simple transformation is used to obtain the spectral density matrix at frequency zero, and the resulting HAC covariance matrix is positive semi-definite by construction. Henceforth, we will refer to this covariance matrix estimator as the VARHAC estimator.

We establish the consistency and asymptotic mean-squared error (MSE) of the VARHAC estimator under very general conditions of heteroskedasticity and temporal dependence, similar to the conditions considered in previous research (e.g., Andrews 1991). For example, the stochastic process

---

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

[2] The comparative advantage of AR spectral methods in the unit root context has been analyzed recently by Perron and Ng (1994).

need not be finite-order ARMA or even covariance stationary. Under these conditions, we demonstrate that the VARHAC estimator achieves a higher convergence rate than any estimator in the class of positive semi-definite kernels. As the sample grows large, we show that the VAR spectral estimator is able to efficiently capture the unconditional second moments of the data, and that BIC provides an effective method of lag order selection by evaluating the goodness-of-fit relative to the degree of parsimony of the model.

The VARHAC estimator faces a tradeoff in MSE associated with the choice of VAR lag order, similar to the MSE tradeoff associated with the choice of bandwidth parameter for kernel-based estimators. In particular, for a given sample length, using a higher VAR lag order reduces the bias from neglected high-order autocovariances, but raises the estimation variance as a result of the loss in degrees of freedom. Our analysis demonstrates that the growth rate of the lag order chosen by BIC asymptotically approaches arbitrarily closely to the geometric rate that minimizes the asymptotic MSE of the VAR spectral estimator.

In principle, the truncated (unweighted) kernel estimator (cf. White 1984) could achieve the same asymptotic MSE as the VARHAC estimator. In practice, however, no method is currently available for determining the optimal lag truncation point for the truncated kernel. Furthermore, the truncated kernel does not ensure a positive semi-definite covariance matrix.

Other kernel-based estimators, including those proposed by Andrews and Monahan (1992) and Newey and West (1994), impose a substantial bias to ensure a positive semi-definite covariance matrix. These kernels assign weights less than unity to the higher-order autocovariances, with a bandwidth parameter indicating the rate at which the weights decline to zero. Under fairly general conditions, the resulting bias dominates all other sources of bias (e.g., the contribution of excluded autocovariances to the spectral density function), so that the optimal growth rate of the bandwidth parameter can be expressed as a function of the sample length and of the true spectral density function (cf. Andrews 1991). Nevertheless, a relatively high bandwidth parameter growth rate must be chosen in order to offset the bias induced by the kernel. Thus, even at the optimal growth rate of the bandwidth parameter, positive definite kernel-based estimators exhibit higher asymptotic bias and asymptotic variance compared with the VARHAC estimator.

To augment these asymptotic results, 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 experiments performed by Andrews and Monahan (1992), we find that the VARHAC estimator generally matches the performance of the prewhitened quadratic-spectral (QS-PW) estimator proposed by Andrews and Monahan (1992), even for data generating processes (*dgp*s) that do not correspond to a finite-order autoregression. In addition, we report simulation experiments which illustrate three important limitations of non-parametric HAC covariance estimation procedures and highlight the advantages of explicitly modeling the temporal properties of the error terms.

First, to ensure that the HAC covariance matrix is positive semi-definite, kernel-based estimation procedures must utilize a single bandwidth parameter in calculating all elements of the spectral density matrix. If some components of the vector of residuals have a relatively high degree of persistence, while other components have low persistence, then imposing the same bandwidth parameter for both sets of variables will tend to generate ill-behaved estimates of the spectral density matrix at frequency zero.[3] In contrast, the VARHAC estimator can utilize a different lag order for each component of the residual vector, because the parametric estimator of the spectral density matrix is positive semi-definite by construction.

Second, the accuracy of kernel-based spectral estimators is sensitive to the procedure used to determine the bandwidth parameter, especially in small samples. As noted above, the optimal growth rate of the bandwidth parameter depends on the true spectral density function and its derivatives at frequency zero. To obtain a feasible "plug-in" bandwidth parameter, Andrews (1991) proposed the use of a low-order parametric model to provide an initial spectral density estimate, which is then utilized to select the bandwidth parameter, which in turn determines the final spectral density estimate. However, our Monte Carlo simulations indicate that if the low-order parametric model is misspecified, the final spectral density estimate may have relatively poor properties in finite samples. Alternatively, Newey and West (1994) proposed the use of the truncated kernel to provide an initial estimate of the spectral density to plug into the bandwidth parameter formula, but their simulations indicate that this procedure can be sensitive to the initial lag truncation point. Finally, one could use a formal model selection procedure to determine the appropriate parametric model, and then use the implied spectral density to calculate the optimal bandwidth parameter. But in this case, one might simply choose to utilize the spectral density associated with the best parametric model, rather than plugging this estimate into the optimal bandwidth formula to obtain a kernel-based spectral estimate.

Finally, kernel-based procedures are sensitive to the prewhitening method. Andrews and Monahan (1992) proposed that kernel-based methods be augmented by using an autoregressive filter of arbitrary order to prewhiten the regression residuals, and considered a first-order autoregressive filter in their Monte Carlo experiments. The AR(1) filter has also yielded improvements in inference properties in subsequent simulation experiments, some of which have considered *dgps* resembling actual economic time series (cf. Newey and West 1994; Christiano and Den Haan 1994; Burnside and Eichenbaum 1994). Such first-order prewhitening is a special case of the VARHAC estimator, in which the AR order is chosen by a data-dependent model selection criterion. Furthermore, as demonstrated by our Monte Carlo simulations, it is not difficult to find *dgps* for which first-order prewhitening is insufficient to remove most of the persistence from the data, so that AR(1) prewhitened kernel estimators perform very poorly in comparison with the VARHAC estimator.

Lee and Phillips (1994) analyzed the properties of an ARMA-prewhitened HAC estimator,

[3] This limitation has also been noted by Robinson (1995).

in which a kernel-based spectral estimator is applied to the prewhitened residuals. In the case where the true *dgp* is a finite-order ARMA process with i.i.d. innovations, Lee and Phillips demonstrated that the ARMA-prewhitened kernel estimator is asymptotically more efficient than any AR(1)-prewhitened positive semi-definite kernel estimator, and is nearly as efficient as a parametric ARMA spectral estimator.[4] Our analysis yields similar results for the VARHAC estimator under much more general conditions of heteroskedasticity and temporal dependence: as the sample length increases, the data becomes truly prewhitened by the VARHAC procedure, so that no additional benefits can be derived from applying a kernel-based procedure to the prewhitened data. In small samples, of course, the VARHAC procedure does not always prewhiten the data completely, so that applying a kernel estimator to the VARHAC residuals may provide improved inferences under certain conditions. In future research, this possibility should be explored using Monte Carlo simulation experiments.

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 asymptotic properties of the VARHAC procedure. Section 4 reports the results of simulation experiments designed to compare the performance of the VARHAC and prewhitened kernel-based HAC covariance matrix estimators.

2. A PARAMETRIC COVARIANCE MATRIX ESTIMATOR.

In many estimation problems, a parameter estimate $\hat{\psi}_T$ for a $p\times 1$ parameter vector ψ_0 is obtained from the sample analog of a set of moment conditions, such as E $V_t(\psi_0) = 0$, where $V_t(\psi_0)$ is an $N\times 1$ vector of residual terms with $N \geq p$. This orthogonality condition is often used to motivate the following estimator of ψ_0:

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

where $V_T = \sum_{t=1}^{T} V_t(\psi)/T$ is the vector of sample moments of $V_t(\psi)$ and F_T is an $N\times N$ (possibly random) symmetric weighting matrix (cf. Hansen 1982). When $N = p$, the choice of the weighting matrix F_T does not matter. Under certain regularity conditions, the parameter $\hat{\psi}_T$ has the following limiting distribution as the sample length T grows arbitrarily large:

$$\left[D_T^{-1} S_T D_T'^{-1}\right]^{-1/2} T^{1/2}\left(\hat{\psi}_T - \psi_0\right) \rightarrow \text{N}(0, I_N) \tag{2.2}$$

$$S_T = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} \text{E} V_s(\psi_0) V'_t(\psi_0), \tag{2.3}$$

---

[4] BIC asymptotically selects the true ARMA(p,q) model with finite p and q, so that the parametric procedure converges at rate $O_p[T^{-1/2}]$. Since the first-order sample autocorrelation of the ARMA-prewhitened residuals converges to zero, Andrews' (1991) plug-in method yields a bandwidth parameter that grows very slowly, at $O[\log T]$. Thus, applying the kernel to the prewhitened residuals causes a slightly higher asymptotic variance, and reduces the convergence rate to $O_p[(T/\log T)^{-1/2}]$. However, this convergence rate still exceeds the rate of $O_p[T^{-2/5}]$ attained by the QS kernel with fixed prewhitening order $\bar{p} < p$.

$$(2.4) \qquad D_T = \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left[\left.\frac{\partial V_t(\psi)}{\partial \psi'}\right|_{\psi=\psi_0}\right],$$

where D_T is an $N\times p$ matrix, and I_N is the $N\times N$ identity matrix.[5] Usually, D_T is estimated by its sample analog $D_T(\hat{\psi}_T)$ and $D_T(\hat{\psi}_T)$-$D_T \to 0$ in probability as $T\to\infty$. If $V_t(\psi_0)$ is a stationary process, then $\lim_{T\to\infty} S_T$ is the spectral density at frequency zero of the process $V_t(\psi_0)$. In Section 3.2, we show that S_T also converges to a well-defined limit if $V_t(\psi_0)$ is not weakly stationary but satisfies certain moment and temporal dependence conditions.

In this paper, we construct the spectral estimator $\hat{S}_T(\hat{\psi}_T)$, using a vector autoregressive representation (VAR) of $V_t(\hat{\psi}_T)$, and either Akaike's (1973) information criterion (AIC) or Schwarz' (1978) Bayesian information criterion (BIC) to select the lag order for each equation in the VAR representation. One main advantage of VAR representations is computational speed, which is important for Monte Carlo studies using vector processes. We will refer to the parametric estimator based on the VAR representation and the BIC model selection criterion as the VARHAC estimator.

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

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

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

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

$$(2.7) \qquad \mathrm{AIC}(\kappa; n) = \log\left(\det\left(\frac{\sum_{t=\overline{K}+1}^{T} \hat{e}_{nt}(\kappa)\hat{e}_{nt}(\kappa)'}{T}\right)\right) + \frac{2\kappa N}{T}.$$

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

---

[5] When $N > p$, the asymptotic covariance matrix can be expressed as the limit of $(D'_T F_T D_T)^{-1} D'_T F_T S_T F_T D_T (D'_T F_T D_T)^{-1}$.

Note that to minimize the computational requirements of the Monte Carlo simulation experiments, we only consider specifications in which all elements of $V_t(\hat{\psi}_T)$ 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 $N(\bar{K}+1)$ equations. Allowing a different lag order for each variable in each equation would require a total of $N[(N(\bar{K}+1))!]$ equations to be estimated, which is computationally feasible if N and $\bar{K}$ 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 simulation experiments reported in Section 4.3 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:

$$V_t(\hat{\psi}_T) = \sum_{k=1}^{\bar{K}} \hat{A}_k V_{t-k}(\hat{\psi}_T) + \hat{e}_t, \tag{2.8}$$

where $\hat{e}_t$ is an $N\times 1$ vector with typical element $\hat{e}_{nt}(\kappa_n)$. The (n,j) element of $\hat{A}_k$ is equal to zero if $k > \kappa_n$ and it is equal to $\hat{\alpha}_{njk}(\kappa_n)$ if $k \le \kappa_n$. The innovation covariance matrix $\hat{\Sigma}_T$ is estimated as follows:

$$\hat{\Sigma}_T = \frac{\sum_{t=\bar{K}+1}^{T} \hat{e}_t \hat{e}'_t}{T}. \tag{2.9}$$

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.[6]

Step 3: Estimation of HAC covariance matrix. Using the results of step 1 and 2, the spectral density matrix at frequency zero is estimated with

$$\hat{S}_T(\hat{\psi}_T) = \left[I_N - \sum_{k=1}^{\bar{K}} \hat{A}_k \right]^{-1} \hat{\Sigma}_T \left[I_N - \sum_{k=1}^{\bar{K}} \hat{A}'_k \right]^{-1} \tag{2.10}$$

When $N = p$, the VARHAC covariance matrix estimator is defined by[7]

$$\hat{V}_T(\hat{\psi}_T) = \left[\hat{D}_T(\hat{\psi}_T)\right]^{-1} \hat{S}_T(\hat{\psi}_T) \left[\hat{D}_T(\hat{\psi}_T)'\right]^{-1}. \tag{2.11}$$

---

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

[7] Using the formula in footnote 6, the estimator can easily be changed for the general case in which $N \ge p$.

3. ASYMPTOTIC THEORY.

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 briefly reviews the existing literature related to AR spectral estimation. Section 3.2 establishes the conditions under which the true autocovariance structure of the data can be represented by an infinite-order VAR. Section 3.3 evaluates the convergence rate of the VAR estimator of the spectral density of observed data. Finally, Section 3.4 extends these results to the VARHAC procedure, which is applied to estimated regression residuals.

3.1 Review of Previous Literature.

The existing literature has extensively analyzed the consistency of the autoregressive spectral estimator for weakly stationary data, but has not considered its properties under more general mixing conditions or in the case of estimated residuals. In particular, Akaike (1969) proved the consistency of the AR spectral estimator when the true *dgp* is an AR process of known finite order and the innovations are i.i.d. with finite fourth moments. Berk (1974) extended this consistency result to the case where the data are generated by an autoregressive process of unknown and possibly infinite order, by allowing the lag order K to increase with the sample size at the rate $o[T^{1/3}]$. Berk also showed that the AR spectral estimator has variance $O[K/T]$ if the lag order grows at a sufficiently rapid rate. Finally, Ahn, Chen, and Hannan (1982) and Hannan and Kavalieris (1984) demonstrated the strong consistency of the VAR spectral estimator, allowing for a martingale-difference sequence of innovations with unconditional homoskedasticity, and permitting the lag order to increase at nearly $o[T^{1/2}]$. However, since the work of Berk (1974), no further results have been available on the convergence rate of the VAR spectral estimator under more general conditions.

Nevertheless, the more general literature on autoregressive models is both vast and invaluable in analyzing the asymptotic properties of the VAR spectral estimator. For example, Grenander and Szegö (1958) and Baxter (1962) evaluated the mean-squared prediction error of finite-order AR approximations to an infinite-order linear process, extending the seminal work of Wold (1938), Wiener (1949), and Kolmogorov (1941). Shibata (1976,1980, 1981) analyzed the asymptotic properties of alternative model selection criteria in fitting linear processes with i.i.d. innovations, and demonstrated the asymptotic efficiency of AIC in minimizing the mean-squared error of the integrated spectrum for infinite-order processes. Lewis and Reinsel (1985) extended several of Berk's (1974) results to the case of VARs with i.i.d. innovations, using the same upper and lower bounds on the growth rate of the lag order; Lütkepohl (1992) provides a useful overview of these and other related results. Hannan et al. (1984, 1986, 1987, 1988) analyzed the algebraic and topological structure of linear systems, and developed asymptotic results for general linear processes with martingale-difference innovations and unconditional homoskedasticity. Finally, Guo, Huang, and Hannan (1990) extended these results on VAR approximation to allow for martingale-difference innovations with unconditional heteroskedasticity.

3.2 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. These conditions are well-understood for weakly stationary processes: if a time series is linearly non-deterministic, then the process has an MA(∞) representation with white-noise (homoscedastic and orthogonal) innovations; if no linear combination of the process $\{V_t\}_{t=-\infty}^{\infty}$ has zero variance, then the process also has an AR(∞) representation. In the absence of weak stationarity, the stochastic process itself does not have an MA(∞) or AR(∞) representation with white-noise innovations. Nevertheless, under the conditions concerning temporal dependence and existence of moments that have been utilized in previous studies of HAC estimation (e.g., White 1984; Andrews 1991; Hansen 1992), we show that the limiting population autocovariances have an MA(∞) representation. If we assume further that no linear combination of the process $\{V_t\}_{t=-\infty}^{\infty}$ has zero variance (a condition typically used to verify the presence of strong mixing), then the limiting autocovariances also have an AR(∞) representation.

Condition A indicates the relevant conditions in the weakly stationary case. Condition A* relaxes these conditions to allow for more general heteroskedasticity and temporal dependence:

Condition A: $\{V_t\}_{t=-\infty}^{\infty}$ is a mean-zero, covariance-stationary, sequence of random N-vectors with absolutely summable autocovariances: i.e. $\sum_{j=0}^{\infty} \left| E\left(V_t V'_{t+j}\right) \right|_{\infty} < +\infty$.

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

(a) $\sup_{t \geq 1} \mathrm{E}(V_t V'_t) < +\infty$,

(b) $\Gamma(0) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\left(V_t V'_t\right)$ is positive definite, and

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

Henceforth we will refer to Condition A/A* when either Condition A or Condition A* is sufficient for the purpose at hand. In these conditions, and throughout the paper, we utilize the notation $|x|_{\infty} = \sup_j |x_j|$ to represent the supremum norm of a vector x. For an $L \times M$ matrix A, we utilize the matrix norm $|A|_{\infty} = \sup_{i=1,\ldots,L} \sum_{j=1}^{M} |A_{ij}|$ (cf. Hannan and Deistler 1988, p.266). Condition A* (a) controls the variance of the observations; Condition A* (b) rules out sequences in which an infinite number of observations have zero variance; and Condition A* (c) controls the degree of dependence across observations. It should be noted that Condition A implies Condition A*.

Condition A* is sufficient to ensure that $\{V_t\}$ meets Grenander's (1954) conditions for asymptotic stationarity (cf. Hannan 1970, p.77). First, Condition A* (b) 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. Second, Conditions A* (a) and A* (b) ensure asymptotic negligibility; i.e., as the sample length T grows arbitrarily large, the variance of each observation makes a negligible contribution to the sum of the variances. Third, Conditions A* (a) and A* (c) ensure that the average jth-order autocovariance converges to a limiting value for each integer j, and Condition A* (b) ensures that the limiting autocovariances are not affected by the exclusion of a finite number of observations.

In particular, for a given sample of length T, the average jth-order autocovariance matrix $\Gamma_T(j)$ is defined as follows for $|j| < T$:

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

It is also 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 $i, j=1, \cdots, M$. 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 $i, j=1, \cdots, M$; and the (i,j)th block of G_∞ for $i,j=1,2, \cdots$. Finally, let $-\Gamma(j)$ comprise the j-th $N\times N$ block of the $MN\times N$ matrix g_M for $j=1, \cdots, M$; and the j-th $N\times N$ block of the matrix g_∞ for $j=1,2, \cdots$.

Now for a given sample of length T, and a given lag truncation point h, we define the truncated spectral density function, $f_{Th}(\omega)$, as follows:

$$f_{Th}(\omega) = \frac{1}{2\pi}\sum_{j=-h}^{h} \Gamma_T(j)\exp(i\omega j) \quad \text{for } \omega \in [-\pi, \pi]. \tag{3.2}$$

The limiting spectral density function, $f(\omega)$, is defined in terms of the limiting autocovariances:

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

Lemma 1 indicates that the limiting autocovariances and limiting spectral density function have all of the properties associated with a weakly stationary Gaussian process with no purely deterministic harmonic components. This lemma also indicates the convergence rates of the average autocovariances and of the truncated spectral density to the corresponding limiting matrices. We use $O(\cdot)$ to denote almost sure convergence, and $O_p(\cdot)$ to denote convergence in probability.

Lemma 1: Under Condition A/A*, the limiting autocovariances and limiting spectral density function are identical to those of a weakly stationary Gaussian process with no purely deterministic harmonic components:

(a) The limiting autocovariances satisfy $\Gamma(j) = \Gamma'(-j)$ for every integer j; $\det(G_M) \geq 0$ for every $M \geq 1$; and $\sum_{j=-\infty}^{\infty} |\Gamma(j)|_\infty < +\infty$.

(b) The limiting spectral density function, $f(\omega)$, is a continuous, Hermitian, positive semi-definite harmonic function of periodicity 2π, $f(\omega) \in L_2[-\pi,\pi]$ (i.e., $|f(\omega)|^2$ is Lebesgue-integrable over $[-\pi,\pi]$); and $\Gamma(j) = \int_{-\pi}^{\pi} f(\omega) \exp(i\omega j)\, d\omega$.

(c) $\Gamma_T(j) = \Gamma(j) + O(1/T)$ for each integer j.

(d) $|f_{Th}(\omega) - f(\omega)|_\infty \leq \frac{1}{2\pi} \sum_{|j|>h} |\Gamma_T(j)|_\infty + O(\frac{h}{T})$.

The intuition behind Lemma 1 can be summarized as follows. Since $\Gamma_T(j) = \Gamma'_T(-j)$ for $|j| < T$, the limiting autocovariances also satisfy $\Gamma(j) = \Gamma'(-j)$. Since Grenander's conditions for asymptotic stationarity are satisfied under Condition A*, the limiting autocovariances $\Gamma(j)$ form a positive semi-definite sequence; i.e., $\det(G_M) \geq 0$ for every $M \geq 1$ (cf. Hannan 1970, p.78). Thus, the limiting autocovariances are identical to those of a weakly stationary Gaussian process, and the integrated spectrum has the properties of a weakly stationary Gaussian process, i.e., is Lebesque-integrable and monotone non-decreasing (cf. Theorems X.3.1 and X.3.2 of Doob 1953, p.474-475; Ibragimov and Linnik 1971, p. 311). The absolute summability of the limiting autocovariances ensures that the data contain no purely deterministic harmonic components, since $\Gamma_T(j)$ would not converge to zero with increasing j in the presence of such components. Thus, the integrated spectrum is absolutely continuous (i.e., with no jumps) and differentiable, with continuous spectral density $f(\omega)$ (cf. Priestley 1982, pp. 226-230). Finally, Condition A/A* ensures that $E(V_t V_{t+j})$ is bounded for all t and j, yielding the convergence rates given in parts (c) and (d) of Lemma 1. Thus, if the lag truncation point increases with sample length in such a way that $h = o(T)$, then the truncated spectral density converges a.s. to the limiting spectral density (cf. White 1984; Andrews 1991).

If the limiting spectral density function $f(\omega)$ is positive definite almost everywhere in $[0, \pi]$ (i.e., $f(\omega)$ is only singular at a countable number of points), then the limiting autocovariances and limiting spectral density function 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 a stronger assumption is required:

Condition B: The limiting 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 structure. Condition B also places an important restriction on the Toeplitz matrices G_M and G_∞. For univariate processes, Theorems 9.2(a) and 9.6(a) of Grenander and Szegö (1958, pp. 147-154) demonstrated that the smallest eigenvalue of G_∞ is equal to the smallest value of $f(\omega)$ for ω in $[-\pi, \pi]$, and that the smallest eigenvalue of G_M declines monotonically and converges to the smallest eigenvalue of G_∞ as $M \to \infty$. (Closely related results may also be found in Hannan 1970, pp. 148-150, and in Theorem 4.2.1 of 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.

Condition B also plays an important role in analyzing the convergence properties of kernel-based spectral estimators, which have mainly been derived under assumptions of strong or uniform mixing (e.g., White 1984; Newey and West 1987; Hansen 1992). These convergence properties can be derived using assumptions on fourth-order cumulants, but such assumptions are difficult to verify without the use of mixing conditions (cf. Andrews 1991, pp. 823-824). Strong mixing is directly implied if the data are q-mixing; i.e., $E(V_t V_{t+K}) = 0$ for all $k > q$ and all $t = 1, \cdots, T$. (cf. Davidson 1995, p. 215). For more general processes, however, all of the available sets of sufficiency conditions for strong mixing require the spectral density function to be strictly positive definite (Ibragimov and Linnik 1971, Theorem 17.3.3, p. 313; Chandha 1974; Gorodetskii 1977; Pham and Tran 1985; Athreya and Pantula 1986; Davidson 1995, Theorem 14.9, pp. 219-225).

Condition B may be expected to hold for essentially all untransformed data, since measurement error generates a positive definite spectral density at all frequencies (Priestley 1982, pp. 390-391). The most important exception to Condition B occurs when stationary data is "over-differenced," thereby removing all variation at frequency zero. However, the case of over-differencing is rather problematic for all HAC covariance estimation, because the parameter estimates generally converge at $O_p(T^{-1})$ instead of $O_p(T^{-1/2})$, and the true HAC covariance matrix becomes singular, thereby generating confidence intervals of measure zero.

Under Conditions A/A* and B, Theorem 1indicates that the limiting autocovariance structure has a VAR(∞) representation with absolutely summable coefficients. Furthermore, the VAR coefficients $\{A_j\}$ satisfy the infinite-order Yule-Walker equations $G_\infty A_\infty = g_\infty$, where the transpose of the $N \times N$ matrix A_j forms the j-th block of A_∞. Since $\det(G_\infty) \neq 0$, its inverse is well-defined, so that $A_\infty = G_\infty^{-1} g_\infty$. Finally, under these assumptions, the VAR coefficients vanish at the same rate as the limiting 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.

Theorem 1: Conditions A/A* and B imply the following:

(a) The limiting autocovariances $\Gamma(j)$ and spectral density $f(\omega)$ are identical to those of a vector MA(∞) process with i.i.d. Gaussian innovations: $f(\omega) = (1/2\pi)\ \Theta[\exp(i\omega)]\ \Sigma\ \Theta^*[\exp(i\omega)]$, where Σ is a real symmetric positive-definite matrix; $\Theta(z) = \sum_{j=0}^{\infty} z^j \Theta_j$; $\Theta^*(z)$ is the complex conjugate of $\Theta(z)$; $\Theta_0 = I_N$; $\sum_{j=0}^{\infty} |\Theta_j|_\infty < +\infty$; and $\det(\Theta(z)) \neq 0$ for $|z| \leq 1$.

(b) The limiting autocovariances $\Gamma(j)$ and spectral density $f(\omega)$ 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 $|z| \leq 1$.

(c) $A_\infty = G_\infty^{-1} g_\infty$ and $\Sigma = \Sigma_\infty = \Gamma(0) - g'_\infty A_\infty$.

(d) If $\sum_{j=0}^{\infty} j^\lambda |\Gamma_j|_\infty < +\infty$ for $\lambda \geq 0$, then $\sum_{j=0}^{\infty} j^\lambda |\Theta_j|_\infty < +\infty$ and $\sum_{j=0}^{\infty} j^\lambda |A_j|_\infty < +\infty$.

As indicated by Theorem 1(a), Conditions A/A* and B are sufficient to ensure that the limiting autocovariances have a vector MA(∞) representation $\Theta(z)$ with absolutely summable coefficients,[8] and that $\Theta(z)$ has no roots on or inside the unit circle; i.e., $\det(\Theta(z)) \neq 0$ for $|z| \leq 1$.[9] Thus, $A(z)=[\Theta(z)]^{-1}$ is well-defined for all $|z| \leq 1$, so that the limiting autocovariances also have a VAR(∞) representation, as indicated in Theorem 1(b). Then Theorem 1(c) follows from the orthogonality properties of the trigonometric polynomials $\exp(i\omega j)$. Finally, Theorem 1(d) follows directly from Theorem 3.8.2 and 3.8.2 of Brillinger (1981, pp. 76-78).

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 (cf. Fuller 1996, Theorem 4.3.4, 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). 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).

3.3. Convergence Rate of the VAR Spectral Estimator

Thus far we have been considering the structure of the true (limiting) autocovariances $\{\Gamma(j)\}$, where $\Gamma(j) = E(V_t V'_{t-j})$ for covariance stationary data, and $\Gamma(j) = \lim_{T\to\infty} \sum_{t=1}^{T-j} E(V_t V'_{t-j})$ for asymptotically stationary data. The true autocovariances are generally unobserved in practice, but can be estimated by the sample autocovariances $\tilde{\Gamma}_T(j) = (1/T)\sum_{t=1}^{T-j} V_t V'_{t-j}$ for $j \geq 0$, and $\tilde{\Gamma}_T(j) = \tilde{\Gamma}'_T(-j)$

[8] The autocovariance structure is not restricted by the absence of roots of $\Theta(z)$ inside the unit circle, since an MA representation with one or more roots inside the unit circle and no roots on the unit circle can always be re-normalized as an MA representation with all roots outside the unit circle (Hansen and Sargent 1981, p. 102; Hamilton 1995, pp. 64-68).

[9] Under Condition A/A*, weaker assumptions than Condition B can be used to ensure that the limiting autocovariances have an MA(∞) representation with square-summable coefficients. This would permit somewhat greater temporal dependence than the absolutely summable coefficients implied by Conditions A/A* and B (cf. Hamilton 1994, p. 69).

for $j < 0$. The sample autocovariances are biased by a factor $|j|/T$ relative to the true (limiting) autocovariances, but this bias is asymptotically negligible. A kernel-based estimator of the spectral density at frequency zero can be represented as a weighted sum of the sample autocovariances: i.e., $\widetilde{S}^{np}_{T,\xi} = \sum_{j=1-T}^{T-1} \kappa(j/\xi)\,\widetilde{\Gamma}_T(j)$, where the kernel $\kappa(\cdot)$ is a bounded function of the lag order j and bandwidth parameter ξ.

The VAR spectral estimator can also be expressed in terms of the sample autocovariances. For the VAR(h) estimator, let $\hat{\Gamma}_T(j-i)$ comprise the (i,j)th $N\times N$ block of the $hN\times hN$ Toeplitz matrix $\widetilde{G}_{Th}$; let $-\widetilde{\Gamma}_T(i)$ comprise the i-th $N \times N$ block of the $hN\times N$ matrix $\widetilde{g}_{Th}$; and let the coefficient matrix $\widetilde{A}'_{Thj}$ comprise the j-th $N\times N$ block of the $hN\times N$ matrix $\hat{A}_{Th}$. Then $\widetilde{A}_{Th}$ is determined by the sample Yule-Walker equations as a non-linear function of the sample autocovariances: $\widetilde{A}_{Th} = (\widetilde{G}_{Th})^{-1}\,\widetilde{g}_{Th}$. The estimated innovation covariance matrix can be expressed as $\widetilde{\Sigma}_{Th} = \sum_{j=0}^{h} \widetilde{\Gamma}_T(j)\,\widetilde{A}'_{Thj}$, where $\widetilde{A}_{Th0}$ is the $N\times N$ identity matrix I_N. . Finally, the VAR estimator of the spectral density at frequency zero is given by $\widetilde{S}^{ar}_{Th} = \left[\sum_{j=0}^{\infty}(\widetilde{A}_{Thj})^{-1}\right]\widetilde{\Sigma}_{Th}\left[\sum_{j=0}^{\infty}(\widetilde{A}'_{Thj})^{-1}\right]$. 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 $\widetilde{S}^{ar}_{Th} = (I_N + Q'_h\,\widetilde{A}_{Th})^{-1}\,\widetilde{\Sigma}_{Th}\,(I_N + \widetilde{A}'_{Th}\,Q_h)^{-1}$.

To analyze the asymptotic mean-squared error of the VAR spectral estimator, it is useful to construct the VAR(h) coefficients and innovation covariance matrix based on the set of true (limiting) autocovariances $\{\Gamma(j), j = -h,\cdots,h\}$. As noted above, condition B ensures that the smallest eigenvalue of the Toeplitz matrix G_h is bounded away from zero, so that $\det(G_h) > 0$ for all h. Thus, we can define $A_h = G_h^{-1}\,g_h$, and $\Sigma_h = \sum_{j=0}^{h}\Gamma(j)\,A'_{hj}$, where A'_{hj} comprises the j-th $N\times N$ block of A_h. The spectral density at frequency zero implied by the true VAR(h) approximation is given by $S^{ar}_h = (I_N + Q'_h\,A_h)^{-1}\,\Sigma_h\,(I_N + A'_h\,Q_h)^{-1}$. As shown in the previous section, the true (limiting) spectral density at frequency zero may be expressed as $f(0) = (I_N + Q'_\infty\,A_\infty)^{-1}\,\Sigma_\infty\,(I_N + A_\infty Q_\infty)^{-1}$, using the VAR($\infty$) representation of the true (limiting) autocovariance structure of the data.

Now the asymptotic root mean-squared error (RMSE) of $\widetilde{S}^{ar}_{Th}$ can be expressed as $|\mathrm{vec}(\widetilde{S}^{ar}_{Th} - f(0))|_\infty$. Given the triangle inequality, the asymptotic RMSE can be expressed in terms of two components as follows:

$$(3.4)\qquad |\mathrm{vec}(\widetilde{S}^{ar}_{Th} - f(0)|_\infty \;\le\; |\mathrm{vec}(S^{ar}_h - f(0)|_\infty \;+\; |\mathrm{vec}(\widetilde{S}^{ar}_{Th} - S^{ar}_h|_\infty .$$

The first component, henceforth referred to as the asymptotic bias, represents the absolute difference between the spectral densities implied by the VAR(h) approximation and by the true (limiting) VAR(∞) representation, where both spectral densities are based on the true (limiting) autocovariances.

The second component, henceforth referred to as the asymptotic standard deviation, measures the sampling variation in estimating the VAR(h) model. To avoid further notational complexity, this term also includes two asymptotically negligible sources of bias related to the sample autocovariances: first, the degrees of freedom correction factor $|j|/T$ mentioned above; and second, in the case of asymptotically

stationary data, the difference between each average autocovariance $\Gamma_T(j) = (1/T)\sum_{t=1}^{T-j} \mathrm{E}(V_t V'_{t+j})$, and the corresponding limiting autocovariance $\Gamma(j)$. Of course, $\Gamma_T(j) = \Gamma(j)$ for all j and T for weakly stationary data.

3.3.1. Evaluation of Asymptotic Bias.

To analyze the asymptotic bias of the VAR estimator, we use either Condition C or Condition C* to quantify the rate at which the autocovariances decline to zero. We will refer to Condition C/C* when either condition is sufficient for the purpose at hand.

Condition C: $q = \sup_{\bar{q}} \left\{\bar{q}: \sum_{j=1}^{\infty} j^{\bar{q}} \left|\mathrm{vec}\,\Gamma(j)\right|_{\infty} < \infty\right\} \geq 0.$

Condition C*: $q = \sup_{\bar{q}} \left\{\bar{q}: \sum_{j=1}^{\infty} j^{\bar{q}} \sup_{t \geq 1} \left|\mathrm{vec}\left(E\left\{V_t V'_{t+j}\right\}\right)\right|_{\infty} < \infty\right\} \geq 0.$

Condition C/C* indicates that the autocovariances decline toward zero at least geometrically as a function of the lag order. The condition $q > 0$ ensures the absolute summability of the autocovariances, as specified in Condition A/A*.[10] Thus, the upper tail sum $\sum_{j=h}^{\infty} \left|\Gamma(j)\right|_{\infty}$ converges to zero at the rate $O(h^{-q})$ (cf. Davidson 1994, p. 32). In the special case where the autocovariances correspond to those of a finite-order vector ARMA process, the autocovariances decline at an exponential rate, corresponding to an arbitrarily large value of q.

Condition C/C* can also be interpreted as indicating the degree of smoothness of the spectral density function at frequency zero. For even values of r, the r-th derivative of $f(\omega)$ at $\omega = 0$ is given by $\sum_{j=-\infty}^{\infty} j^r \Gamma(j)$ (Priestley 1982, p. 459). Under Condition C/C*, it is clear that the absolute value of the r-th derivative is bounded for even values of $r \leq q$. For a finite-order vector ARMA process, the spectral density is a rational function of $\exp(i\omega)$, and is therefore infinitely differentiable at frequency zero (Priestley 1982, pp. 283-284). On the other hand, if $q < 1$, the spectral density will be continuous but not differentiable, exhibiting a cusp at frequency zero.

The asymptotic bias of the VAR spectral estimator depends on the rate at which the sum of VAR(h) coefficients ($Q_h' A_h$) converges to the sum of VAR(∞) coefficients ($Q'_{\infty} A_{\infty}$), and the rate at which the VAR(h) innovation covariance matrix Σ_h converges to Σ_{∞}. To analyze these terms, it is useful to partition A_{∞} into two submatrices: $A_{\infty,h}$ contains the first h $N \times N$ blocks, and $A_{\infty,h+}$ contains the remainder of A_{∞}. Thus, the bias from the sum of VAR coefficients can be represented by two components: $Q'_h (A_h - A_{\infty,h})$ and $Q'_{\infty} A_{\infty,h+}$. From Theorem 1(c) and Condition C/C*, we know that $\left| Q'_{\infty} A_{\infty,h+} \right|_{\infty} = O(h^{-q})$. Now using an approach similar to that of Baxter (1962) and Theorem 6.6.12 of Hannan and Deistler (1986, p. 269-271), we can demonstrate the following lemma:

[10] Berk (1973), Lewis and Reinsel (1986) and Hannan and Deistler (1988) all imposed the slightly stronger condition that $q > 1/2$.

Lemma 2: Under Conditions A/A*, B, and C/C*, the VAR(h) estimator has the following properties:

(a) $\left|\text{vec}\left(Q'_h A_h - A_{\infty,h}\right)\right|_\infty = O(h^{-q})$,

(b) $\left|\text{vec}\left(\Sigma_h - \Sigma_\infty\right)\right|_\infty = O(h^{-2q})$,

(c) $\left|\text{vec}\left(S_h^{ar} - f(0)\right)\right|_\infty = O(h^{-q})$.

It is useful to compare the order rate for the asymptotic bias of the VAR spectral estimator with that of the truncated kernel, $S_h^{tk} = \sum_{j=-h}^{h}\Gamma(j)$. The bias of the truncated kernel is simply the sum of excluded autocovariances, so that the absolute bias is bounded by $\sum_{|j|>h}\Gamma(j) = O(h^{-q})$. Thus, the truncated kernel and the VAR spectral estimator have asymptotic biases that vanish at the same rate. The truncated kernel, however, does not necessarily yield a positive definite spectral density matrix at frequency zero, whereas the VAR(h) spectral estimator is ensured to be positive definite.

To understand this result further, it is useful to note that the VAR spectral estimator can be expressed as $S_h^{ar} = \sum_{j=-\infty}^{\infty}\Gamma_h^*(j)$, where $\Gamma_h^*(j)$ are the autocovariances implied by the VAR(h) model. Since the VAR(h) coefficients are determined by the Yule-Walker equations, $\Gamma_h^*(j) = \Gamma_h(j)$ for $|j| \le h$. Thus, the difference between the VAR(h) and truncated spectral estimators can be expressed as $D_h^{ar} = \sum_{|j|>h}\Gamma_h^*(j)$. Furthermore, as with any stationary finite-order VAR process, the implied higher-order autocovariances $\Gamma_h^*(j)$ decline exponentially toward zero as $j \to \infty$ (cf. Hannan 1995, p. 256). This implies that D_h^{ar} vanishes at the same rate as the leading term $\Gamma_{h+1} = O(h^{-q-1})$. Thus, by including these implied higher-order autocovariances, the VAR(h) estimator ensures a positive definite spectral density matrix with negligible effects on the asymptotic bias relative to the truncated kernel estimator.

It is also useful to compare these results with those of the class of positive definite kernel-based spectral estimators. Any kernel estimator based on T autocovariances may be represented as $S_{T,\xi(T)}^{ke} = \sum_{1-T}^{T-1}\kappa\left(j/\xi(T)\right)\Gamma(j)$, where the kernel $\kappa(z)$ is continuous at zero and a.e. for $z \neq 0$; $\kappa(0) = 1$; $\kappa(z) = \kappa(-z)$; and $\int_{-\infty}^{\infty}\kappa^2(z)\,dz < \infty$. To ensure a positive definite spectral density, the kernel $\kappa(z)$ must satisfy the restriction that $|\kappa(z)| < 1$ for all $z \neq 0$; in other words, the autocovariances $\Gamma(j)$ for $|j| > 0$ receive lower weight compared with the truncated kernel. (The sufficient conditions for a positive definite kernel may be found in Andrews 1991, p. 822).

The smoothness of the kernel $\kappa(z)$ at $z = 0$ indicates the extent to which the weights on low-order autocovariances differ from unity. In fact, it can be shown that no positive definite kernel can be differentiable of order higher than 2 at $z = 0$ (Priestley 1982, p. 568). If the spectral density is sufficiently smooth such that $q > 2$ in Condition C/C*, then the bias from assigning weights less than unity to the low-order autocovariances (i.e. the weights on $\Gamma(j)$ for $0 < |j| \le T$ dominates the bias due to the neglected autocovariances ($\Gamma(j)$ for $|j| > T$). If the second derivative of the kernel is finite at $z = 0$ (as with the QS and Parzen kernels), the absolute bias is of order $O(h^{-2})$. If only the first derivative of the kernel is finite at $z = 0$ (as with the Bartlett kernel), the absolute bias is of order $O(h^{-1})$. Thus, the asymptotic bias of positive definite kernels shrinks relatively slowly as a function of the bandwidth

parameter, compared with the rate at which the asymptotic bias of the VAR spectral estimator vanishes as a function of the lag order.

3.3.2. Evaluation of Asymptotic Variance.

The asymptotic variance of kernel-based and VAR spectral estimators depend on the second moments of the sample autocovariances. For the case of fourth-order stationary univariate data, Bartlett (1946) demonstrated the following result, which is exact in finite samples:

$$(3.5) \qquad \mathrm{Cov}(\tilde{\Gamma}(i),\tilde{\Gamma}(j)) = \frac{1}{T}\sum_{s=1+i-T}^{T-j-1}\left(1-\frac{\eta(i,j,k)}{T}\right)\left\{\Gamma(k)\Gamma(k+j-i)+\Gamma(k+j)\Gamma(k-i)+K_4(0,i,j,k)\right\},$$

where $|\eta(i,j,k)| \leq 1$ (cf. Priestley 1982, pp. 325-326). The fourth-order cumulants $K_4(0,i,j,k)$ measure the extent to which V_t displays excess kurtosis relative to the fourth-order moments implied by a normally distributed process, $\breve{V}_t$, with identical autocovariances (cf. Hannan 1970, p. 23; Priestley 1982, p. 58).

$$(3.6) \qquad \begin{aligned} K_4(t,t+j,t+m,t+n) &= \mathrm{E}(V_t - \mathrm{E}V_t)(V_{t+j} - \mathrm{E}V_{t+j})(V_{t+m} - \mathrm{E}V_{t+m})(V_{t+n} - \mathrm{E}V_{t+n}) \\ &- \mathrm{E}(\breve{V}_t - \mathrm{E}\breve{V}_t)(\breve{V}_{t+j} - \mathrm{E}\breve{V}_{t+j})(\breve{V}_{t+m} - \mathrm{E}\breve{V}_{t+m})(\breve{V}_{t+n} - \mathrm{E}\breve{V}_{t+n}) \end{aligned}$$

The result in equation (3.5) was extended to multivariate fourth-order stationary processes by Hannan (1970, p. 209), and to more general mixing conditions processes by Andrews (1988, 1991).

If V_t has absolutely summable autocovariances and absolutely summable fourth-order cumulants, equation (3.5) indicates that each sample autocovariance $\tilde{\Gamma}_T(j)$ converges to the true autocovariance $\Gamma(j)$ at rate $O_p(T^{-1/2})$. Each sample autocovariance has a bias factor $|j|/T$ as indicated at the beginning of Section 3.3; however, this bias is $o(T^{-1/2})$ for $|j| \leq h = o(T^{1/2})$. Finally, in the case of asymptotically stationary data, each sample autocovariance $\tilde{\Gamma}_T(j)$ has asymptotically negligible bias of $O(T^{-1})$ corresponding to the difference between the average autocovariance $\Gamma_T(j)$ and the limiting autocovariance $\Gamma(j)$.

Now consider any weighted sum of $h(T)$ sample autocovariances with uniformly bounded weights, and its deviation from the corresponding weighted sum of true autocovariances: e.g., $Z_{T,h(T)} = \sum_{j=0}^{h(T)}\lambda(T,j)(\tilde{\Gamma}_T(j) - \Gamma(j))$, where $|\lambda(T,j)| \leq M < +\infty$ for all j and T. Then we find that:

$$(3.7) \qquad \mathrm{E}\left(Z_{T,h(T)}^2\right) = \sum_{i=0}^{h(T)}\sum_{j=0}^{h(T)}\lambda(T,i)\lambda(T,j)\,\mathrm{Cov}\left(\tilde{\Gamma}_T(i),\tilde{\Gamma}_T(j)\right) = O(h(T)/T)$$

using equation (3.5) under the conditions of absolutely summable autocovariances and fourth-order cumulants. Intuitively, the result follows from the asymptotic independence of sample autocovariances at different leads and lags (Priestley 1982, pp. 327 and 425). This implies that $Z_{T,h(T)}$ converges in mean squared to zero at the rate $O_p\left((h(T)/T)^{1/2}\right)$. Again, this result also holds for asymptotically stationary

data, where $\sum_{j=0}^{h(T)}\lambda(T,j)\left(\widetilde{\Gamma}_T(j)-\Gamma_T(j)\right) = O_p\left((h(T)/T)^{1/2}\right)$, and $\sum_{j=0}^{h(T)}\lambda(T,j)\left(\Gamma_T(j)-\Gamma(j)\right)$ $= O_p\left(h(T)/T\right) = o_p\left((h(T)/T)^{1/2}\right)$ for $h(T) = o(T^{1/2})$.

Since many kernel-based spectral estimators (e.g., those which utilize the truncated, Bartlett, and Parzen kernels) are a weighted average of the first $h(T)$ sample autocovariances, these results immediately indicate that such estimators will have asymptotic variance of $O_p(h(T)/T)$ if the lag truncation parameter $h(T) = o(T)$. This order result for the asymptotic variance can also be obtained for spectral estimators based on more general kernels, such as the QS kernel (cf. Priestley 1982,p. 457; Andrews 1988,1991).

To apply these results to the asymptotic variance of the VAR spectral estimator, we follow Andrews (1991, p. 839) in assuming that the fourth-order cumulants of V_t are absolutely summable, whether the data are weakly stationary (Condition D) or asymptotically stationary (Condition D*). We will refer to Condition D/D* when either condition is sufficient for the purpose at hand.

Condition D: $\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\left|\mathrm{vec}(K_4(0,i,j,k))\right|_\infty < +\infty$

Condition D*: $\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sup_{t\geq 1}\left|\mathrm{vec}(K_4(t,i,j,k))\right|_\infty < +\infty$

Lemma 1 of Andrews (1991, p. 824) provides conditions on the degree of strong mixing and existence of higher moments which are sufficient to verify Condition D/D*. Hansen (1992) utilized mixingale bounds to demonstrate the consistency of kernel-based spectral estimators under somewhat weaker conditions. It is likely that Hansen's (1992) conditions are also applicable to the VAR spectral estimator; however, to avoid additional complexities, we do not pursue this approach here.

The asymptotic variance of the VAR spectral estimator is determined by the sampling variation in estimating the sum of VAR(h) coefficients and of the covariance matrix of the estimated VAR(h) residuals; i.e. $|\mathrm{vec}(\widetilde{S}_{Th}^{ar} - S_h^{ar})|_\infty$ is determined by $|\mathrm{vec}(Q'(h)(\widetilde{A}_{Th} - A_h))|_\infty$ and $|\mathrm{vec}(\widetilde{\Sigma}_{Th} - \Sigma_h)|_\infty$. Since the lag order h generally will increase as a function of the sample length T, the convergence rate of the asymptotic standard deviation must hold uniformly in h up to some maximum lag order H_T that also increases monotonically with the sample length.

Now it is useful to define the $hN\times N$ matrix $\widetilde{C}_{Th}$ as follows:

$$\widetilde{C}_{Th} = \widetilde{G}_{Th}\left(\widetilde{A}_{Th} - A_h\right) = \left(G_h - \widetilde{G}_{Th}\right)A_h + \left(\widetilde{g}_{Th} - g_h\right). \tag{3.8}$$

Now we have:

$$\left(\widetilde{A}'_{Th} - A'_h\right)Q_h = \widetilde{C}'_{Th}\widetilde{G}_{Th}^{-1}Q_h = \widetilde{C}'_{Th}\left(\widetilde{G}_{Th}^{-1} - G_h^{-1}\right)Q_h + \widetilde{C}'_{Th}G_h^{-1}Q_h \tag{3.9}$$

The k-th block of $\widetilde{C}_{Th}$ is equal to $\sum_{j=0}^{h}\left(\Gamma(k-j) - \widetilde{\Gamma}_T(k-j)\right)A'_{hj}$. Thus, each of the elements of $\widetilde{C}'_{Th}$ is a weighted sum of deviations between the sample and true (limiting) autocovariances, so that $|\widetilde{C}_h|_\infty = O_p\left((h(T)/T)^{1/2}\right)$. It can also be shown that each element of $\widetilde{G}_{Th}^{-1}$ converges at rate $O_p(T^{-1/2})$ to the

corresponding element of G_h^{-1}, so that $|\widetilde{G}_{Th}^{-1} - G_h^{-1}|_\infty = O_p((h(T)/T)^{1/2})$. Finally, $|Q_h|_\infty = 1$, and $|G_h^{-1}|_\infty = O(1)$. Thus, each element of the first term on the right-hand side of equation (3.10) converges at rate $O_p(h(T)^{3/2}/T)$, and each element of the second term converges at rate $O_p((h(T)/T)^{1/2})$, which dominates the order of the first term as long as $h(T) = O(T^{1/3})$. Similar results can be shown for the elements of $\widetilde{\Sigma}_{Th} - \Sigma_h$. Thus, we have the following lemma:

Lemma 3: Under Conditions A/A*, B, C/C*, and D/D*, the VAR(h) estimator has the following properties uniformly in $h(T)$ for $0 \le h(T) \le H(T) = O(T^{1/3})$:

(a) $$\left|\mathrm{vec}\left(Q_h'\left(\widetilde{A}_{Th} - A_h\right)\right)\right|_\infty = O_p((h(T)/T)^{1/2}),$$

(b) $$\left|\mathrm{vec}\left(\widetilde{\Sigma}_{Th} - \Sigma_h\right)\right|_\infty = O_p(\max\{h(T)/T, T^{-1/2}\}),$$

(c) $$\left|\mathrm{vec}\left(\widetilde{S}_{Th}^{ar} - S_h\right)\right|_\infty = O_p((h(T)/T)^{1/2}).$$

Thus, the ratio of lag order to sample length affects the asymptotic variance of the VAR spectral estimator in the same way that the ratio of bandwidth parameter to sample length ratio affects the asymptotic variance of kernel-based spectral estimators.[11]

3.3.3 Evaluation of Asymptotic Mean-Squared Error.

As indicated in equation (3.4), the asymptotic RMSE of the VAR spectral estimator can be decomposed into two terms which represent the asymptotic bias and the asymptotic standard deviation. Thus, from Lemma 2(c) and Lemma 3(c), it is clear that the asymptotic RMSE of $\widetilde{S}_{Th}^{ar}$ will be $O(h^{-q})$ + $O_p((h/T)^{1/2})$, uniformly in $h = h(T) \le H(T) = O(T^{1/3})$. This result reveals a tradeoff in the choice of lag order h; namely, a higher lag order reduces the asymptotic bias and increases the asymptotic variance. Since the optimal growth rate of the lag order depends on the smoothness of the spectral density at frequency zero, one might suppose that these convergence rates are unattainable in practice. In fact, however, we can approach arbitrarily closely to the optimal geometric growth rate by using Schwarz' (1978) Bayesian Information Criterion (BIC) to select the lag order. In this case, the lag order is chosen by minimizing $\mathrm{BIC}_T(h) = \log(\det(\widetilde{\Sigma}_{Th})) + h N^2 \log(T)/T$ over $h = 0, \cdots, H(T)$, where $H(T)$ is specified as the highest lag order to be considered for a given sample of length T. Alternatively, we may consider using Akaike's (1973) Information Criterion (AIC), where the lag order is chosen by minimizing AIC_T (h)

---

[11] Results similar to those of Lemma 4(b) were previously obtained by Shibata (1979) for the univariate case of a general linear process with i.i.d. Gaussian innovations, with a maximum lag order growing at rate $o[T^{1/2}]$, slightly faster than the maximum lag order specified in Lemma 4. Shibata's results were later extended by Hannan and Kavalieris (1986) to the multivariate case with conditionally homoscedastic martingale difference innovations, and with a maximum lag order growth rate of $o[(T/\log(T))^{1/2}]$.
Both Shibata (1979) and Hannan and Kavalieris (1986) also demonstrated a slightly faster convergence rate of o_p instead of O_p.
Unfortunately, it remains unclear whether the faster convergence rate and the weaker maximum lag order restriction obtained by Shibata (1979) and Hannan and Kavalieris (1986) remain valid in the presence of conditional or unconditional heteroskedasticity.

$= \log(\det(\widetilde{\Sigma}_{Th})) + 2\,h\,N^2/T$ over $h = 0, \cdots, H(T)$. However, Lemma 4 indicates that BIC may be expected to have better sampling properties than AIC, at least in the presence of heteroskedastic innovations.

Lemma 4: Under Conditions A/A*, B, C/C*, and D/D*, the following properties hold uniformly in $0 \le h \le H(T) = O(T^{1/3})$:

(a) $AIC_T(h) = \text{constant} + \text{trace}(\Sigma_\infty^{-1}(\Sigma_h - \Sigma_\infty)) + O_p(h/T)$,

(b) $BIC_T(h) = \text{constant} + \text{trace}(\Sigma_\infty^{-1}(\Sigma_h - \Sigma_\infty)) + h\,N^2 \log(T)/T + O_p(h/T)$.

Lemma 4 follows directly from Lemma 2(b), Lemma 3(b), and the properties that $\log(\det(A)) = \text{trace}(\log(A))$ for any positive definite matrix A, and that $x - \log(1+x)$ approaches zero as x approaches zero (cf. Hannan 1970, p.159; Hannan and Kavalieris 1986; Hannan and Deistler 1988, p. 343).

Lemma 4(a) indicates that the penalty term h/T used in AIC may not be sufficiently large to dominate the sampling variation in estimating $\widetilde{\Sigma}_{Th}$. Shibata (1979) and Hannan and Kavalieris (1986) demonstrated a slightly faster convergence rate of $\widetilde{\Sigma}_{Th}$ to Σ_h, so that the AIC penalty term was just large enough to dominate the sampling variation for AR(∞) processes with conditionally homoscedastic innovations. Nevertheless, Shibata (1976) showed that in the case of finite-order AR models, the AIC penalty term is too small to dominate the sampling variation, so that the lag order chosen by AIC does not converge to the true lag order with probability one. Thus, it is not surprising that sampling variation appears to cause problems for AIC in the presence of conditional or unconditional heteroskedasticity.

On the other hand, Lemma 4(b) indicates that BIC provides an effective method of lag order selection under very general conditions. The BIC penalty term, $h\log(T)/T$, is sufficiently large to dominate the sampling variation of the estimated innovation covariance matrix, so that $\widetilde{\Sigma}_{Th}$ can be used as a proxy for Σ_h, the covariance matrix implied by the true AR(h) approximation. Furthermore, Σ_h converges at rate $O(h^{-2q})$ to Σ_∞, the innovation covariance matrix implied by the AR(∞) representation. Thus, BIC provides a means of evaluating the tradeoff between asymptotic bias (by measuring the extent to which additional lags improve the goodness-of-fit) and asymptotic variance (by penalizing the use of additional parameters).

In fact, by differentiating the formula given in Lemma 4(b) with respect to h and setting the result to zero, we can determine the growth rate of the lag order chosen by BIC, and the corresponding convergence rate of the VAR spectral estimator. If the spectral density is differentiable at frequency zero (i.e., $q \ge 1$), the lag order chosen by BIC converges to $(T/\log(T))^{1/(2q+1)}$, so that the VAR spectral estimator converges in probability at a geometric rate arbitrarily close to $T^{(-q/(2q+1))}$.[12] If the true (limiting) autocovariances correspond to those of a finite-order ARMA process, then the spectral density is infinitely

[12] The results of Berk (1973) and Lewis and Reinsel (1986) were obtained using a lower bound of $o(T^{1/2q})$ on the lag order growth rate, which ensures the asymptotic normality of the AR coefficients, but excludes the use of model selection criteria which yield a lag order growth rate approaching the optimal rate $T^{1/(2q+1)}$.

differentiable at frequency zero (i.e., $q \to +\infty$). In this case, the lag order chosen by BIC grows at a logarithmic rate, and the VAR spectral estimator converges in probability at a geometric rate arbitrarily close to $T^{-1/2}$. Finally, in the case where the spectral density is not differentiable at frequency zero (i.e., $0 < q < 1$), the lag order chosen by BIC approaches the maximum rate $H(T) = T^{1/3}$, and the VAR spectral estimator converges in probability at the rate $T^{-q/3}$.

Thus, the main results for VAR spectral estimation can be summarized as follows:

Theorem 2: Assume Conditions A/A*, B, C/C*, and D/ D*. Then the VAR spectral estimator has the following properties:

(a) $|\text{vec}(\widetilde{S}_{Th}^{ar} - f(0))|_\infty = \max(O(h^{-q}), O_p[(h/T)^{1/2}])$ uniformly in $0 \le h < H(T) = O(T^{1/3})$,

(b) If $h(T) \to \infty$ and $h(T) = O(T^{1/3})$, then $\widetilde{S}_{Th}^{ar} - f(0) \to_p 0$,

(c) If $h(T) = \text{argmin}(\text{BIC}_T(h))$, and the maximum lag order $H(T) = O(T^{1/3})$, then $|\text{vec}(S_{Th}^{ar} - f(0))|_\infty = O(T^{-q/(2q+1)})$ for $q \ge 1$, and $|\text{vec}(\widetilde{S}_{Th}^{ar} - f(0))|_\infty = O(T^{-q/3})$ for $0 < q < 1$.

Since it is straightforward to extend Theorem 2 to incorporate the analysis of estimated regression residuals, the implications of Theorem 2 will be discussed as part of the discussion of Theorem 3 in the next section.

3.4 Convergence Rate of the VARHAC Estimator.

The asymptotic properties of VAR spectral estimation given in Sections 3.2 and 3.3 can be readily extended to the case of HAC covariance matrix estimation, which typically involves the analysis of estimated regression residuals. In this case, $V_t(\psi)$ is a random N-vector for each $p \times 1$ vector of regression parameters ψ in the admissible region $\Psi \subset \Re^p$. For example, in OLS estimation of a linear regression model, $V_t(\psi) = (Y_t - X'_t \psi)X_t$. For pseudo-ML estimation, $V_t(\psi)$ is the score function for the t-th observation. For IV estimation of a dynamic non-linear simultaneous equation model, $V_t(\psi)$ is the Kronecker product of the instrument vector with the vector of model equations evaluated at ψ. In all of these cases, the asymptotic HAC covariance matrix can be expressed as a function of the $N \times N$ matrix $f(0) = \lim_{T \to \infty} S_T(\psi_0)$, where $S_T(\psi_0) = (1/T)\sum_{s=1}^{T}\sum_{t=1}^{T} \text{E}\{V_s(\psi_0)V'_t(\psi_0)\}$ and ψ_0 is the true vector of regression parameters.

To simplify notation in the following discussion, we will use V_t to refer to $V_t(\psi_0)$, the regression function evaluated at the true regression parameter vector ψ_0, 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 true (limiting) j-th order autocovariance evaluated at ψ_0, 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 true (limiting) autocovariances evaluated at ψ_0. Similarly, $\widetilde{\Gamma}_T(j)$ refers to the sample j-th order autocovariance based

on the true series $\{V_t\}$, and the matrices $\tilde{G}_{Th}$, $\tilde{g}_{Th}$, $\tilde{C}_{Th}$, $\tilde{A}_{Th}$, $\tilde{\Sigma}_{Th}$, and $\tilde{S}_{Th}^{ar}$ are as previously defined using the sample autocovariances $\tilde{\Gamma}_T(j)$. 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 $S_T(\psi_o)$, we will utilize the following inequality:

$$
\begin{aligned}
|\mathrm{vec}(\hat{S}_{Th}^{ar} - S_T(\psi_o))|_\infty \le\ & |\mathrm{vec}(\hat{S}_{Th}^{ar} - \tilde{S}_{Th}^{ar})|_\infty + |\mathrm{vec}(\tilde{S}_{Th}^{ar} - S_h^{ar})|_\infty \\
& + |\mathrm{vec}(S_h^{ar} - f(0))|_\infty + |\mathrm{vec}(f(0) - S_T(\psi_o))|_\infty
\end{aligned} \tag{3.10}
$$

The convergence rates of the second and third terms of equation (3.10) have already been evaluated in Section 3.3, while the convergence rates of the first and last terms remain to be determined.

To analyze the first term in equation (3.8), we follow Andrews (1991, pp. 825-826) in making the following assumptions concerning the regression function $V_t(\psi)$ and the estimated regression parameter vector $\hat{\psi}_T$:

Condition E:

(a) $\sup_{t \ge 1} \left| \mathrm{vec}\left(E\left(V_t(\psi_o)\, V_t'(\psi_o) \right) \right) \right|_\infty < +\infty$,

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

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

(d) Conditions A/A*, B, C, and D/D* hold for the following stochastic process:

$$\left\{ V_t'(\psi_o),\ \mathrm{vec}\left(\partial V_t(\psi) / \partial \psi' - \mathrm{E}\left[\partial V_t(\psi) / \partial \psi' \right] \right) \right\},$$

(e) $\sqrt{T}\left(\hat{\psi}_T - \psi_o \right) = O_p\left(T^{-1/2} \right)$.

Parts (a) and (b) of Condition E are commonly utilized to demonstrate the asymptotic normality of the estimated regression parameter vector, $\hat{\psi}_T$. Parts (b) and (c) are typically required to demonstrate the consistency of $D_T(\hat{\psi}_T)$ (cf. equation (2.4) above). Part (d) can be verified under reasonable assumptions if $V_t(\psi)$ has the form $V(X_t, \psi)$ for some measurable function $V(\cdot,\cdot)$ and some random variable X_t (cf. Andrews 1991, p. 826). Part (e) follows from the asymptotic normality of $\hat{\psi}_T$.

Under Conditions A/A*, C/C*, D/D*, and E, Andrews (1991, p. 852) has shown that $\hat{\Gamma}_T(j) = \tilde{\Gamma}_T(j) + O_p(T^{-1/2})$. Thus, as in Section 3.3.2, we find that $|\hat{G}_{Th}^{-1} - \tilde{G}_{Th}^{-1}|_\infty = O_p((h(T)/T)^{1/2})$. Now consider a weighted sum of $h(T)$ sample autocovariances with uniformly bounded weights, evaluated at $\hat{\psi}_T$, and consider its deviation from the corresponding weighted sum of sample autocovariances evaluated at ψ_o . Under the same conditions, Andrews (1991, pp. 852-853)

demonstrated that any such weighted sum of deviations vanishes at rate $O_p((h(T)/T)^{1/2})$. This result applies directly to $|\hat{C}_{Th} - \tilde{C}_{Th}|_\infty$, so that following the same steps as in Section 3.3.2, we find that $|\text{vec}(\hat{S}^{ar}_{Th} - \tilde{S}^{ar}_{Th})|_\infty = O_p((h(T)/T)^{1/2})$ for $h(T) = O(T^{1/3})$. Thus, the first term of equation (3.8) is asymptotically negligible compared with the second term; i.e., under Condition E, the use of estimated residuals does not affect the asymptotic variance of the VAR spectral estimator.

Now consider the last term in equation (3.11), namely, the rate at which $S_T(\psi_o)$ converges to $f(0)$, the true (limiting) spectral density matrix of V_t at frequency zero. Using the definition of $S_T(\psi_o)$ and rearranging terms, it can be seen that $S_T(\psi_o) = \sum_{j=1-T}^{T-1}(1-\frac{|j|}{T})\Gamma_T(j)$. Thus, the difference between $S_T(\psi_o)$ and $f(0)$ can be expressed as follows:

$$(3.11)\quad S_T(\psi_o) - f(\omega) = \sum_{|J|\le T^b} (\Gamma_T(j) - \Gamma(j)) + \sum_{T^b<|j|<T} (\Gamma_T(j) - \Gamma(j)) - \sum_{|J|<T} \frac{|j|}{T}\Gamma_T(j) - \sum_{T\le|J|} \Gamma(j)$$

where we have set $b = 1/(q+1)$. The first and second terms in equation (3.11) are identically equal to zero in the weakly stationary case. In the asymptotically stationary case, the first term vanishes at rate $O(T^{b-1}) = O(T^{-q/(q+1)})$, since each average autocovariance converges at rate $O(T^{-1})$ to the corresponding limiting autocovariance. Under Conditions A* and C*, the second term of equation (3.11) vanishes at the rate $O(T^{-bq}) = O(T^{-q/(q+1)})$. Under Conditions A/A* and C/C*, the third term vanishes at rate $O(T^{-1})$ for $q \ge 1$, and at rate $O(T^{-q})$ for $q < 1$, and the fourth term vanishes at rate $O(T^{-q}) = o(T^{-q/(q+1)})$.

Since the difference between $S_T(\psi_o)$ and $f(0)$ is independent of the VAR lag order h(T), the use of BIC continues to yield a VAR lag order growth rate that asymptotically approaches $O(T^{1/(2q+1)})$, so that the VAR spectral estimator converges to $f(0)$ at a rate approaching $O_p(T^{-q/(2q+1)})$, as in Section 3.3.3. Since the difference between $S_T(\psi_o)$ and $f(0)$ vanishes at the faster rate $O(T^{-q/(q+1)})$, this difference is asymptotically negligible compared with the sources of bias and variance already analyzed in Section 3.3. Thus, all the results of Theorem 2 continue to hold for the VARHAC estimator:

Theorem 3: Assume Conditions A/A*, B, C/C*, D/D*, and E. Then the VARHAC spectral estimator has the following properties:

(a) $|\text{vec}(\hat{S}^{ar}_{Th} - J_T(\psi_o))|_\infty = \max(O(h^{-q}), O_p[(h/T)^{1/2}])$ uniformly in $0 \le h < H(T) = O(T^{1/3})$.

(b) If $h(T) \to \infty$ and $h(T) = O(T^{1/3})$, then $\hat{S}^{ar}_{Th} - J_T(\psi_o) \to_p 0$

(c) If $h(T) = \text{argmin}(\text{BIC}_T(h))$, and the maximum lag order $H(T) = O(T^{1/3})$, then $|\text{vec}(\hat{S}^{ar}_{Th} - J_T(\psi_o))|_\infty = O(T^{-q/(2q+1)})$ for $q \ge 1$, and $|\text{vec}(\hat{S}^{ar}_{Th} - J_T(\psi_o))|_\infty = O(T^{-q/3})$ for $0 < q < 1$.

It is useful to compare the convergence rate given in Theorems 2 and 3 for the VAR spectral estimator with the convergence rates of kernel-based spectral estimators. As discussed in Sections 3.3.1 and 3.3.2, the truncated kernel estimator also has asymptotic bias of $O(h^{-q})$ and asymptotic variance of

$O_p\,(h/T)^{1/2}$. Thus, in principle, the truncated kernel estimator could converge at rate $O_p\ (T^{-q/(2q+1)})$ if the truncation parameter $h(T)$ could be chosen to grow at the optimal rate $T^{1/(2q+1)}$. In practice, however, a data-dependent bandwidth selection procedure has not been developed for the truncated kernel estimator (cf. Priestley 1982, pp. 460-462; White 1984, p. 159; Andrews 1991, p. 834).

As discussed in Section 3.3.1, positive definite kernel-based spectral estimators have asymptotic bias of $O(\max(h^{-q}, h^{r})$, where r measures the smoothness of the kernel $k(z)$ at $z = 0$, and $r \leq 2$ for all positive definite kernel-based estimators. Thus, if the true spectral density function is smoother than the kernel function, then the optimal growth rate of the bandwidth parameter can be determined analytically as $O(T^{1/(2r+1)})$, without any knowledge of the specific value of q. Under the assumption that $q > r$, Andrews (1991) proposed a bandwidth selection procedure in which $h(T) = O(T^{1/3})$ for the Bartlett kernel, and
$h(T) = O(T^{1/5})$ for the QS and Parzen kernels.

Based on these considerations, it is clear that the VAR spectral estimator will converge at a faster rate than any positive definite kernel-based estimator for almost all autocovariance structures. If $q > r$, the positive definite kernel estimators lose efficiency by placing weight less than unity on the low-order autocovariances. The extreme case is one in which the autocovariances have the structure of a finite-order ARMA process, so that q is arbitrarily large. In this case, the VAR spectral estimator converges at a rate approaching $O_p\ (T^{-1/2})$, whereas spectral estimators based on either the Parzen or QS kernel converge at the rate $O_p\,(T^{-2/5})$, and the spectral estimator based on the Bartlett kernel converges at the rate $O_p\,(T^{-1/3})$.

For $q < r$, positive definite kernel estimators with $r = 2$ are also less efficient than the VAR spectral estimator, because the bandwidth parameter specified by Andrews'(1991) formula grows too slowly. For example, in the case where $q = 1/2$, BIC will asymptotically select the maximum lag order $O(T^{1/3})$. In this case, Theorem 2(c) indicates that the VAR estimator converges at rate $O_p\ (T^{-1/6})$. In contrast, the spectral estimators which are based on either the Parzen or QS kernel, and which utilize Andrews' (1991) bandwidth selection procedure, will converge at rate $O_p\ (T^{-1/10})$. Thus, the VAR estimator converges at a faster rate than the QS or Parzen kernels except in the special case where q is exactly equal to 2.

4. MONTE CARLO EXPERIMENTS.

In this section, we report the results of Monte Carlo experiments to compare the small-sample properties of two HAC covariance matrix estimators, and to evaluate the extent to which each estimator provides accurate inferences in two-tailed tests of the significance of the estimated coefficients. The first estimator is the parametric VARHAC estimator described in Section 2 above, and the second is the non-parametric QS-PW estimator studied by Andrews and Monahan (1992), which uses the quadratic spectral kernel, first-order prewhitening, and univariate AR(1) models in the automatic bandwidth selection procedure. This estimator is briefly reviewed in Section 4.1.

In Section 4.2, we consider the data generating processes used in the Monte Carlo experiments of Andrews and Monahan (1992), and we find that the parametric VARHAC estimator matches the small-sample performances for the QS-PW estimator quite well. Next, Section 4.3 documents the advantage of the VARHAC estimator in allowing different autoregressive orders for different components of the residual vector. Section 4.4 highlights the pitfalls associated with using an arbitrary parametric model in automatic bandwidth selection procedures. Finally, Section 4.5 verifies the benefits derived by the VARHAC estimator in using a model selection criterion to determine the autoregressive order, or equivalently, the degree of prewhitening. All experiments set $T = 128$.

4.1 Review of kernel-based HAC estimators.

In this section, we give a short description of the non-parametric (kernel-based) estimators of the spectral density at frequency zero used in Andrews (1991) and Andrews and Monahan (1992). A more general description of kernel-based HAC estimators can be found in Christiano and Den Haan (1996), Den Haan and Levin (1996), and Robinson and Velasco (1995). The non-parametric estimators described in this section have the following form:

$$\hat{S}_T = \sum_{j=-T+1}^{T-1} \kappa\left(\frac{j}{\xi_T}\right) \hat{\Gamma}_j \,, \tag{4.1}$$

where $\kappa(\cdot)$ is a weighting function (kernel) and ξ is a bandwidth parameter. Also,

$$\hat{\Gamma}_j = \frac{1}{T-N} \sum_{t=1}^{T-j} V_t(\hat{\psi}_T) V'_{t+j}(\hat{\psi}_T) \,, \; j = 0, \cdots, T\text{-}1, \tag{4.2}$$

and

$$\hat{\Gamma}_j = \hat{\Gamma}'_{-j}, j = -1,-2, \cdots, -T+1.$$

Andrews and Monahan (1992) adjusted this procedure by prewhitening the elements of $V_t(\hat{\psi}_T)$ with an AR(1) prewhitening filter. Let $\bar{\alpha}_n$ be the autoregressive coefficient of the n-th element of $V_t(\hat{\psi}_T)$. Then the n-th prewhitened residual is given by $\bar{e}_{nt} = V_t(\hat{\psi}_T) - \bar{\alpha}_n V_{t-1}(\hat{\psi}_T)$. Andrews and Monahan (1992) included an eigenvalue adjustement for the estimated AR(1) coefficient in the

prewhitening regression. We do not use this adjustment, which appears somewhat arbitrary, and does not affect the results reported here. Also, let $\bar{A}$ be the $N\times N$ diagonal matrix with typical element $\bar{\alpha}_n$ and $\bar{e}_t$ be the $N\times 1$ vector with typical element $\bar{e}_{nt}$. The QS-PW estimator is given by

$$\hat{S}_T^{QS-PW}(\hat{\psi}_T) = \left[I_N - \bar{A}\right]^{-1} \bar{\Sigma}_T \left[I_N - \bar{A}'\right]^{-1}, \tag{4.3}$$

where $\bar{\Sigma}_T$ is an estimate of the spectral density at frequency zero of $\bar{e}_t$. To estimate $\bar{\Sigma}_T$, the kernel-based procedure from Andrews (1991) is used. As documented in equation (4.1) this requires a choice of kernel and a choice of bandwidth parameter. A common conclusion from many Monte Carlo experiments is that the choice of kernel in the class of kernels that guarantee a positive semi-definite covariance matrix is usually not very important for the small sample results.[10] For this reason we follow Andrews and Monahan (1992) and consider only the QS kernel in this paper.

Unfortunately, there is overwhelming evidence that the choice of the bandwidth parameter is very important in small samples[11] and sometimes even in large samples.[12] For a given sample, the mean-squared error (MSE) of the spectral estimator exhibits an U-shaped dependence on the bandwidth parameter, with sharp increases in MSE as the bandwidth parameter is moved away from its optimal value (cf. Andrews 1991). The fragility of spectral estimation accuracy provided an important motivation for Andrews' (1991) derivation of the optimal rate at which to raise the bandwidth parameter as the sample grows arbitrarily large. This method chooses ξ_T to minimize the asymptotic expectation of

$$\text{vec}(\hat{S}_T - S)' W \,\text{vec}(\hat{S}_T - S), \tag{4.4}$$

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

$$\xi_T^* = 1.322\left(\alpha(2)\,T\right)^{1/5}, \text{ with} \tag{4.5}$$

$$\alpha(2) = \frac{2\,\text{vec}(S^{(2)})' W \,\text{vec}(S^{(2)})}{\text{tr}\left(W\left(I+K\right)\left(S\otimes S\right)\right)}, \tag{4.6}$$

and

$$S^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \,\bar{\Gamma}_j \,. \tag{4.7}$$

Here $\bar{\Gamma}_j$ is the autocovariance of $\bar{e}_t$, tr(·) denotes the trace operator, I is the $N^2\times N^2$ identity matrix, and K is the $N^2\times N^2$ commutation matrix defined by the property, vec(A') = K vec(A). To understand the

---

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

[11] This sensitivity was originally identified in simulation experiments which compared the finite-sample properties of AR and kernel-based spectral density estimation procedures (cf. Beamish and Priestley 1981; Kay and Marple 1981; and Parzen 1983). The sensitivity is documented for the case of estimated residuals in Andrews (1991), Andrews and Monahan (1992), Christiano and Den Haan (1994), and Newey and West (1994).

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

automatic bandwidth procedures proposed by Andrews (1991), it is important to realize that the optimal bandwidth parameter ξ_T^* depends on the true value of S. Implementation of the automatic bandwidth selection procedure thus requires an initial estimate of S, as well as an estimate of $S^{(q)}$, and a choice for W.

Andrews and Monahan (1992) followed Andrews (1991) by fitting a parametric model for $\bar{e}_t$ and then using this model to obtain estimates of S and $S^{(q)}$ using equations like (2.10). In particular, an AR(1) model was fitted to the n^{th} element of $\bar{e}_t$. Moreover, only the elements of W that correspond to the diagonal elements of $(\hat{S}_T - S)$ are given a positive weight. In particular, the elements corresponding to the slope coefficient are given a weight equal to one and the elements corresponding to the constant are given a weight equal to zero. Newey and West (1994) proposed the use of the truncated kernel to obtain initial estimates of S and $S^{(q)}$ in the formula for the optimal bandwidth parameter. Robinson (1991) also considered data- dependent bandwidth procedures to determine the bandwidth parameter, and proposed a form of cross validation to determine the bandwidth parameter used in estimating the spectral density over a range of frequencies.

4.2 The Andrews and Monahan (1992) experiments.

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

$$(4.8) \qquad Y_t = X_t\psi_0 + u_t, \quad t=1,\cdots,T \quad \text{and} \quad \hat{\psi}_T = \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}\left[\sum_{t=1}^{T} X_t Y_t'\right].$$

The estimand of interest is the variance (conditional on $X = (X_1, \cdots, X_T)'$) of the LS estimator of the first nonconstant regressor (i.e., the second diagonal element of $\text{Var}(T^{1/2}(\hat{\psi}_T - \psi_0)|X)$. All elements of ψ_0 are equal to zero.

Andrews and Monahan (1992) considered seven basic regression models: AR(1)-HOMO, in which the errors and regressors are homoskedastic AR(1) processes; AR(1)-HET1 and AR(1)-HET2, in which the errors and regressors are AR(1) processes with multiplicative heteroskedasticity overlaid on the errors; MA(1)-HOMO, in which the errors and regressors are homoskedastic MA(1) processes; MA(1)-HET1 and MA(1)-HET2, in which the errors and regressors are MA(1) processes with multiplicative heteroskedasticity overlaid on the errors; and MA(m)-HOMO, in which the errors and regressors are homoskedastic MA(m) processes with linearly declining MA parameters. A range of different parameter values is considered for each model, with each parameter value corresponding to a different degree of autocorrelation. For a more detailed description of these experiments, see Andrews and Monahan (1992).

Figures 1 and 2 report the coverage probabilities of the t-statistic that tests whether the (first) least-squares slope coefficient is equal to its true value. In Figure 1 (2), we report the outcomes for the

experiments in which the errors and regressors are AR(1) (MA) processes. The columns report the frequency that the t-statistic is higher than the 10% critical value for three different HAC estimators. The gray column correspond to the results of the QS-PW estimator, the black column correspond to the results of the VARHAC estimator using BIC, and the white column corresponds to the results of the VARHAC estimator using AIC.

The important conclusion that arises from the figures is that the inference accuracy of the VARHAC estimator matches that of the QS-PW estimator quite well, despite the fact that these *dgp*s might be expected to favor the QS-PW. In the AR(1) models, for example, QS-PW imposes first-order prewhitening, while the VARHAC estimators chooses the lag order with a model selection criterion. We observe the biggest difference for the MA(1)-HET1 model, for which the QS-PW outperforms the VARHAC estimators to some extent. For instance, when the MA coefficient is equal to 0.5, the 10% coverage probability is equal to 12.5% for QS-PW, equal to 18.4% for VARHAC using BIC and equal to 16.0% for VARHAC using AIC. That the QS-PW has some advantages for this model is no surprise, since the VARHAC estimator uses AR models to approximate the true MA(1) processes. It should be noted that neither AIC nor BIC dominates the other in all cases.

4.3 The limitations of a single bandwidth.

As discussed in the introduction, non-parametric kernel-based estimators require the bandwidth to be the same for all elements of the vector of residuals ($V_t(\hat{\psi}_T)$) to guarantee a positive semi-definite covariance matrix. Therefore, the optimal bandwidth is a compromise of the values that would be chosen for an element by element analysis. In this section, we document the impact of this restriction using the following Monte Carlo experiment. Consider the least-squares estimator for the following scalar model:

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \quad \text{and} \quad (1-\rho L)\,\varepsilon_t = e_t. \tag{4.9}$$

Here $\alpha = \beta = 0$, X_t and e_t are i.i.d. normally distributed random variables. The two elements of V_t are ε_t and $\varepsilon_t X_t$. Thus, the first element is an AR(1) process, and the second element is serially uncorrelated. To highlight the fundamental point, we do not use the prewhitening option for the QS-PW estimator, since first-order prewhitening would make both components close to white noise. For higher-order processes

for ε_t, the QS-PW estimator would encounter the same limitations as those discussed here.

The choice of the weighting matrix W is crucial for determining the optimal bandwidth parameter in equations (4.5) and (4.6) when V_t is a vector. We consider three alternatives. The first alternative uses the true unconditional covariance matrix of V_t, while the second alternative uses the true spectral density at frequency zero of V_t. To focus on our main points we use population values; in practice, these weights would need to be estimated.[13] The third alternative uses a different weighting

[13] Moreover, the automatic bandwidth procedure of Andrews and Monahan (1992) that estimates univariate AR(1) models for each of the elements of the residual vector cannot be used when the weighting matrix is not diagonal. In this example, the population values of the innovation matrix and the spectral density are diagonal.

matrix, and thus a different bandwidth and spectral density, for the calculation of each standard error. In particular, to calculate the standard error of the constant term, zero weight is given to the second element of V_t ; and to calculate the standard error of the slope coefficient, zero weight is given to the first element of V_t . This choice is motivated by the fact that for linear regression models, the coefficients and standard errors have a direct correspondence to the individual components of V_t. For more general applications, however, this procedure is not possible, even in principle, because no direct correspondence exists between the regression parameters and error components. This procedure can also not be used to test restrictions across coefficients of the linear regression model.

The results of the Monte Carlo experiment are presented in Table 1. First, consider the behavior of the test-statistic for the constant regressor, that corresponds to the persistent element of V_t. As documented in the table, the confidence intervals are very similar for the three choices of the weighting matrix. For the first two scaling methods, the optimal bandwidth parameter is heavily influenced by the persistent component of V_t. For example, when ρ equals 0.9, then the optimal bandwidth parameter equals 23.24 and 19.48, respectively, for the first two choices of W. The optimal bandwidth parameter is equal to 23.24 for the third choice of W, which assigns no weight to the second, serially uncorrelated, element.

The counterpart of these results is that the choice of the weighting matrix is important for the behavior of the test-statistic of the slope coefficient. The slope coefficient is directly related to the second element of V_t that is serially uncorrelated. Using the first two choices of W, however, large values for the bandwidth parameters are chosen for both elements of $V_t(\hat{\psi}_T)$. This means that the calculated standard error depends on a large number of sample autocovariances with a population value equal to zero. The third choice of W only uses the sample covariance of the second element of $V_t(\hat{\psi}_T)$ to determine the bandwidth parameter and consequently chooses on average small bandwidth parameters. To see how important these choices of the bandwidth parameter can be in small samples, consider the accuracy of the spectral estimate for conducting inference when $\rho = 0.9$. In this case, the frequency the test-statistic is in the 90% confidence interval is equal to 77.5, 79.4 and 88.6 for the first, second and third choice of the weighting matrix, respectively.

It should be emphasized that the third choice of the weighting matrix is only possible when there is a direct relation between the test-statistic and an element of V_t. When there is such a relation, then the results from the Monte Carlo suggest that it is worthwhile to use this procedure. In general applications, however, the chosen bandwidth parameter will be a compromise of the serial correlation properties of the elements of V_t.

In contrast, the VARHAC procedure can utilize a different lag order in modelling each of the elements of V_t. As indicated in the panel C of Table 1, both AIC and BIC almost never choose a zero lag length for the persistent component of V_t . However, for the component of V_t that is serially uncorrelated, BIC (AIC) chooses a zero lag length 98% (76%), 96% (70%), 93% (61%), and 88% (50%)

for the four experiments with different values for the autoregressive coefficients. Consequently, the VARHAC procedure is able to accurately calculate the standard errors for the slope coefficient. Although AIC and BIC do not always select the same lag order, both criteria yield empirical sizes that are quite close to the nominal 10% size.

It should be emphasized that the weighting matrix should be chosen to ensure that the data-dependent bandwidth parameter is not sensitive to a rescaling of the variables. Suppose, for example, that the explanatory variable X_t is measured in smaller units. Without an adjustment of the weighting matrix, this rescaling increases the relative variance of the serially uncorrelated component of V_t, and thereby reduces the bandwidth parameter. As a result, inferences concerning more persistent components of V_t can become severely distorted in finite samples (cf. Den Haan and Levin 1996).

4.4 The limitations of arbitrary parameterizations.

As discussed in Section 4.1, the QS-PW procedures require the specification of a time series process for $V_t(\hat{\psi}_T)$. Andrews and Monahan (1992) used an AR(1) model for each of the components of $V_t(\hat{\psi}_T)$ in all of their Monte Carlo experiments, and subsequent papers in the literature have generally followed the same procedure in implementing the QS-PW estimator. This subsection highlights the consequences of adopting this AR(1) assumption, when V_t follows a different law of motion. We consider the following scalar process, and estimate the mean.

$$(4.10) \qquad Y_t = \varepsilon_t + \nu\,\varepsilon_{t-1} + \mu\,\varepsilon_{t-q},\ q \in \{2,3\}, \quad \text{and} \qquad \hat{\psi}_T = \frac{\sum_{t=1}^{T} Y_t}{T}$$

where ε_t is an i.i.d. normally distributed random variable with zero mean and unit variance. The parameters are chosen in such a way that the first-order autocorrelation coefficient of the prewhitened series is equal to zero or small, but higher-order autocorrelation coefficients are not. In this case, using an AR(1) specification leads to estimates of the optimal bandwidth parameter that are downward biased. Several empirical cases suggest that such a time series process for V_t is not unrealistic. First, Fama and French (1988) documented that for stock returns, autocorrelations are small for short horizons, but relatively large for large horizons. For instance, the average first-order autocorrelation across industries is equal to -0.03 for one-year returns, but equal to -0.34 for four-year returns. Second, Christiano and Den Haan (1994) used a *dgp* resembling that of US quarterly GNP, and found that some prewhitened residuals had a very low first-order MA coefficient, but substantial higher-order serial correlation.

As shown in Table 2, the VARHAC estimators clearly outperforms the QS-PW estimator in this experiment. In this example, the VARHAC estimator that uses AIC has better small sample properties than the VARHAC estimator that uses BIC. In fact, the small sample behavior of VARHAC using AIC is excellent even for 128 observations, despite the use of a vector autoregressive process to approximate a moving average process. In this experiment, the higher-order autoregressive processes chosen by AIC are

better to capture the MA structure of the data than the lower-order autoregressive processes chosen by BIC. Furthermore, the sign of the MA coefficients does not affect the inference accuracy of this estimator. Note that the sign of the MA coefficients has a large impact on the inference accuracy of the QS-PW estimator. If the MA coefficients are positive then the QS-PW estimator underestimates the amount of volatility, and consequently rejects the null hypothesis too often. In contrast, for the negative MA coefficients, the QS-PW rejects the null hypothesis too infrequently. Den Haan and Levin (1996) reported simulation experiments which indicate that the kernel-based procedure of Newey and West (1994) performs better than the QS-PW procedure, but does not outperform the VARHAC estimator.

4.5 The limitations of arbitrary prewhitening order.

An important motivation in developing the VARHAC estimator was the success of the prewhitening procedure proposed by Andrews and Monahan (1992). However, Andrews and Monahan (1992) only considered first-order prewhitening, whereas the VARHAC estimator uses a model selection criterion to choose the order of prewhitening. The advantages of the flexibility of the VARHAC estimator in choosing higher-order prewhitening were not apparent in the Monte Carlo experiments discussed in Section 4.1, since the AR component in the vector of residuals, V_t, was at most of order one, and the QS-PW estimator imposes first-order prewhitening. In this section, we consider the following scalar AR(2) process, and estimate the mean.

$$Y_t = \frac{1}{2}\phi\ Y_{t-1} + \frac{1}{2}\phi\ Y_{t-2} + \varepsilon_t, \qquad \hat{\psi}_T = \frac{\sum_{t=1}^{T} Y_t}{T}, \tag{4.4}$$

and ε_t is an i.i.d. N(0,1) process. The estimand of interest is the standard error of the mean. The values we consider for ϕ are .3, .5, .7, and .9. As seen in Table 3, the VARHAC estimators clearly outperform the QS-PW estimator, even for values of ϕ as low as .5. Given the success of first-order prewhitening, it is not surprising that higher-order prewhitening is also advantageous. It is important to note, however, that the VARHAC estimator does not impose the assumption that the residuals are generated by an AR(2) process. For this experiment, a lag order of two was chosen by BIC (AIC) in 14.3%(35.8%), 60% (67%), 90% (77%), and 96% (78%) of all replications for parameter values equal to .3, .5, .7, and .9, respectively.

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.

Athreya, K.B., and S.G. Pantula, 1986, A note on strong mixing of ARMA processes, *Statistics and Probability Letters* 4, pp. 187-190.

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.

Chandha, K.C., 1974, Strong mixing properties of linear stochastic processes, *Journal of Applied Probability* 11, pp. 401-408.

Christiano, L.J., and Wouter J. den Haan, 1996, Small sample properties of GMM for business cycle analysis, *Journal of Business and Economic Statistics*, forthcoming.

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

Den Haan, W.J., and A. Levin, 1996, A practitioner's guide to robust covariance matrix estimation, manuscript, UCSD and Federal Reserve Board.

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).

Gorodetskii, V.V., 1977, On the strong mixing property for linear sequences, *Theory of Probability and its Applications* 22, pp. 411-413.

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.

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.

Ibragimov, I.A., and Y.V. Linnik, 1971, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen.

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, Unit root tests in ARMA models with data dependent methods for the selection of the truncated lag, manuscript, University of Montreal.

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

Perron, P. and S. Ng., 1994, Useful modifications to some unit root tests with dependent errors and their local asymptotic properties, manuscript, University of Montreal.

Pham, T.D., and L.T. Tran, 1985, Some mixing properties of time series models, *Stochastic Processes and their Applications* 19, pp. 297-303.

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

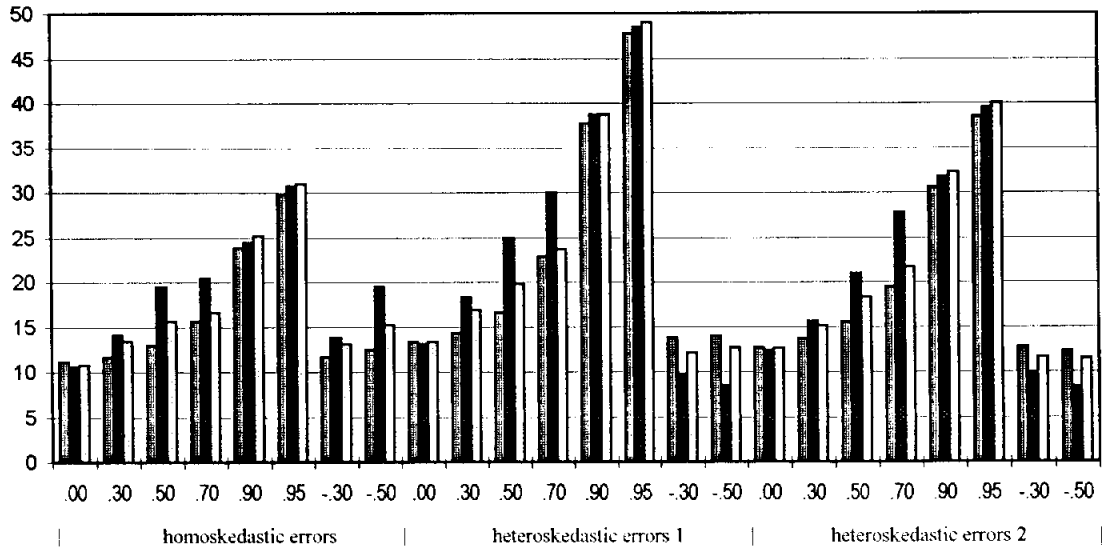
Wiener, N., 1949, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, Wiley Press.

White, H., 1994, *Asymptotic Theory for Econometricians*, Academic Press.

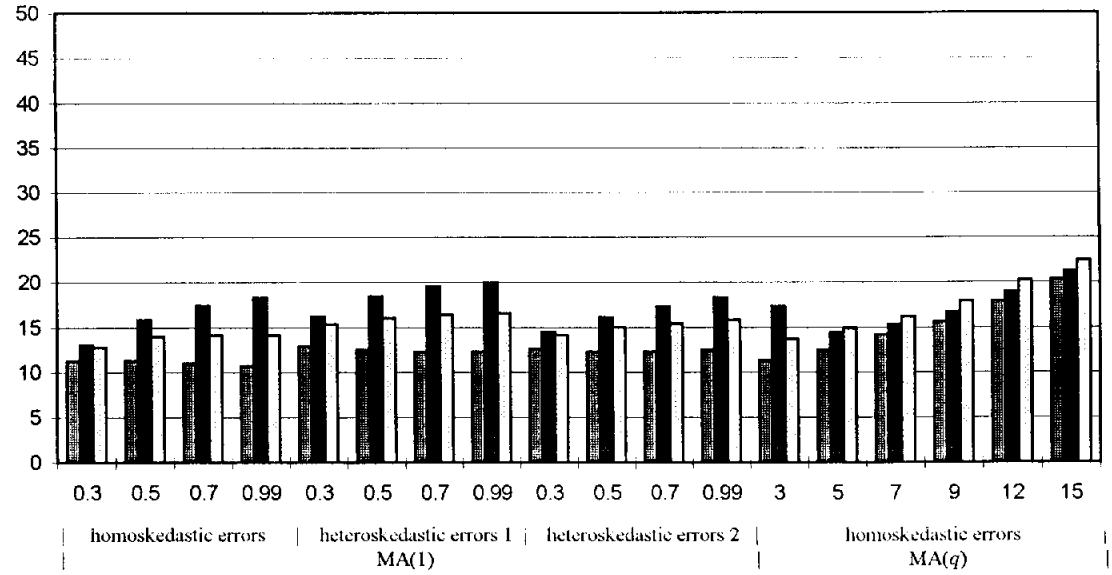
Wold, H., 1938, *The Statistical Analysis of Stationary Time Series*, Uppsala: Almquist and Wicksell Press.

Figure 1: Andrews and Monahan (1992) experiments.

A: AR(1)examples.



B: MA examples.



Note: This graph reports the coverage probabilities of the t-statistic that tests whether the (first) least-squares slope coefficient is equal to its true value. The columns report the frequency the t-statistic is higher than the 10% critical value. The gray column correspond to the results of the QS-PW estimator, the black column correspond to the results of the VARHAC estimator using BIC, and the white column corresponds to the results of the VARHAC using AIC. Part A reports the results for the experiments in which both the independent variables and the error term are an AR(1) process. Part B reports the results for the experiments in which both the independent variables and the error term are an MA processes. For the AR(1) (MA(1)) examples the x-axis reports the AR(1) (MA(1)) coefficient. For the MA(q) examples, the x-axis reports the order of the MA process. $T = 128$ and the results are based on 10,000 replications.

Table 1: The limitations of a single bandwidth.

A: 90% confidence intervals for QS and VARHAC.

ρ	QS(I)		QS(II)		QS(III)		VARHAC (BIC)		VARHAC (AIC)	
	constant	slope	constant	slope	constant	slope	constant	slope	constant	slope
0.3	86.5	88.6	86.4	88.1	86.7	88.7	85.3	89.0	87.9	88.5
0.5	85.3	87.1	84.9	87.3	85.4	88.8	88.3	89.0	87.8	88.5
0.7	82.7	84.9	82.3	85.5	82.6	88.8	86.9	89.3	88.6	88.3
0.9	71.5	77.5	71.4	79.3	71.5	88.6	80.5	90.1	80.4	89.3

B: Average bandwidth parameters for the QS estimator.

ρ	QS(I)	QS(II)	QS(III)	
			constant	slope
0.3	3.48	3.23	3.64	1.48
0.5	5.71	5.09	5.85	1.54
0.7	9.77	8.49	9.85	1.62
0.9	23.24	19.48	23.26	1.70

C: Bandwidths chosen by BIC and AIC (frequencies).

	VARHAC (BIC)						VARHAC (AIC)									
ρ	constant			slope			constant					slope				
	0	1	≥2	0	1	≥2	0	1	2	3	4	0	1	2	3	4
0.3	.39	.61	.01	.98	.02	.00	.07	.74	.11	.05	.03	.76	.14	.05	.03	.02
0.5	.01	.98	.00	.96	.04	.00	.00	.80	.12	.05	.03	.70	.19	.06	.03	.02
0.7	.00	.99	.01	.93	.06	.00	.00	.80	.12	.05	.03	.61	.23	.09	.05	.03
0.9	.00	.99	.02	.88	.10	.01	.00	.78	.13	.05	.04	.50	.25	.12	.07	.06

Note: Panel A reports the 90% confidence intervals constructed using QS and VARHAC estimators of the *t*-statistic that tests whether the estimated least-squares coefficient is equal to its true value in the regression equation $Y_t = \alpha + \beta X_t + u_t$. The following *dgp* is used to generate the data: $Y_t = \alpha + \beta X_t + \varepsilon_t$, $\varepsilon_t = \rho\, \varepsilon_{t-1} + e_t$, where X_t and e_t are i.i.d. standard normal random variables. The roman numerals indicate the weighting matrix used to calculate the optimal bandwidth parameter in the QS procedure. The weighting matrix for QS(I) is the matrix of the (true) unconditional (co)variances. The weighting matrix for QS(II) is the (true) spectral density at frequency zero. For the QS(III) procedure we use a different weighting matrix to calculate the standard error for the two parameter estimates. In particular, when the *i*-th parameter is being estimated a weight of one is given to the *i*-th error term and a weight of zero to the other error term. $T = 128$ and the results are based on 10,000 replications. The maximum lag-order considered for VARHAC is equal to 4. Frequencies in panel C may not add up to 1 due to rounding errors.

Table 2: The limitations of arbitrary parametrizations.

			QS-PW			VARHAC (BIC)			VARHAC (AIC)		
q	ν	μ	99%	95%	90%	99%	95%	90%	99%	95%	90%
2	0.0	-0.3	100.0	99.6	98.3	99.4	96.9	93.8	98.7	94.9	90.3
2	-0.1	-0.3	100.0	99.8	99.1	99.6	97.7	95.1	99.0	95.7	91.4
2	0.0	0.3	95.1	87.4	80.3	97.4	91.9	86.5	97.9	92.9	87.8
2	0.1	0.3	95.9	88.6	81.7	96.9	91.7	86.2	97.7	92.8	87.8
3	0.0	-0.3	100.0	99.3	98.0	99.5	97.9	95.2	99.1	96.0	91.9
3	-0.1	-0.3	100.0	99.6	98.7	99.6	98.4	96.4	99.2	96.7	93.3
3	0.0	0.3	95.5	87.5	80.9	96.5	90.1	84.2	97.8	92.8	88.4
3	0.1	0.3	95.7	88.1	81.5	96.0	89.0	83.3	97.8	92.8	88.1

Note: This table reports the 99%, 95%, and 90% confidence intervals constructed using the QS-PW and the VARHAC estimators for the t-statistic that tests whether sample mean of Y_t is equal to its true value. The following *dgp* is used to generate the data: $Y_t = \varepsilon_t + \nu \varepsilon_{t-1} + \mu \varepsilon_{t-q}$, $q = 2,3$, where ε_t is an i.i.d standard normal random variable. $T = 128$ and the results are based on 10,000 replications.

Table 3: The limitations of a fixed prewhitening order.

	QS-PW			VARHAC (BIC)			VARHAC (AIC)		
ϕ	99%	95%	90%	99%	95%	90%	99%	95%	90%
0.3	96.2	89.2	82.8	95.4	88.4	81.8	95.8	89.8	83.8
0.5	92.8	84.0	76.3	95.8	89.8	83.8	96.8	91.1	85.7
0.7	87.0	75.9	67.8	96.1	89.8	84.6	96.0	89.9	84.5
0.9	70.0	57.7	50.6	90.7	82.9	76.8	90.4	82.5	76.4

Note: This table reports the 99%, 95%, and 90% confidence intervals constructed using the QS-PW and the VARHAC estimators for t-statistic that tests whether the sample mean of Y_t is equal to its true value. The following *dgp* is used to generate the data: $Y_t = 0.5\phi Y_{t-1} + 0.5\phi Y_{t-2} + \varepsilon_t$, where ε_t is an i.i.d standard normal random variable. $T = 128$ and the results are based on 10,000 replications.

Proof of Lemma 1(a): The first statement follows from the fact that $\Gamma_T(j) = \Gamma'_T(j)$ for all $|j| < T$. Parts (a) and (c) of Condition A* ensure that the process $\{V_t\}$ meets Grenander's (1954) conditions for asymptotic stationarity, so that 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). Thus, 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 autocovariances follows from parts (a) and (c) of Condition A*. Furthermore, since the limiting autocovariances are absolutely summable, we find $\Gamma(j) \to 0$ as $j \to \infty$, which ensures the absence of purely deterministic harmonic components from the corresponding weakly stationary process (Priestley 1982, p.230).

Proof of Lemma 1(b): Given the absolute summability of the limiting autocovariances $\Gamma(j)$, the Riesz-Fischer Theorem indicates that $f(\omega) \in L_2[-\pi, \pi]$ and that $\Gamma(j) = \int_{-\pi}^{\pi} f(\omega)\exp(i\omega j)\,d\omega$ (cf. Sargent 1987, p. 249). Since the limiting autocovariances $\Gamma(j)$ 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). The periodicity of $f(\omega)$ follows from its definition in equation (3.3) as a weighted sum of harmonic functions whose periodicity is some integer multiple of 2π. The continuity of $f(\omega)$ follows from the corresponding property of $\exp(i\omega j)$ and from the absolute summability of the limiting autocovariances (Priestley 1982, p. 416).

Proof of Lemma 1(c): Part (a) of Condition A* ensures that $\Gamma_T(j) = \Gamma(j) + O(T^{-1})$, since $E(V_t V'_{t+j})$ is uniformly bounded for all t and j.

Proof of Lemma 1(d): This result follows directly from the definitions in equations (3.2) and (3.3) and from Lemma 1(c).

Proof of Theorem 1(a): Condition B and the results of Lemma 1 ensure that $f(\omega)$ 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 ensures 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).

Proof of Theorem 1(b): Given Condition B and the results of Theorem 1(a), these results follow directly from 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.

Proof of Theorem 1(c): First, note 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 Theorem 1(d): Under Conditions A/A* and B, this property follows directly from Theorems 3.8.2 and 3.8.3 of Brillinger (1981, pp.76-78).

Proof of Lemma 2(a):

We shall 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 $A_\infty = \begin{bmatrix} A_{\infty h} \\ A_{\infty h+} \end{bmatrix}$.

It can be seen that $G_h A_{\infty h} + G_{h+} A_{\infty h+} = g_h$, , so that $A_h = G_h^{-1} g_h = A_{\infty h} + G_h^{-1} G_{h+} A_{\infty h+}$. Thus,

$$\text{(A2)} \qquad \left|\left(A'_h - A'_{\infty h}\right) Q_h\right|_\infty \;\leq\; \left|A_{\infty h+}\right|_\infty \left|G_{h+}\right|_\infty \left|G_h^{-1}\right|_\infty \left|Q_h\right|_\infty .$$

Then Lemma 3(a) follows from evaluating the terms on the right-hand-side of equation (A2). Theorem 1(c) and Condition C ensure that $\left|A'_{\infty h+}\right|_\infty \leq \sum_{j=h+1}^{\infty}|A_j|_\infty = O(h^{-q})$. Condition A/A* ensures that $|G'_{h+}|_\infty \leq C_1 = \sum_{j=-\infty}^{+\infty}|\Gamma(j)|_\infty < +\infty$ for all $h > 0$. From the definition of Q_h, $\left|Q_h\right|_\infty = 1$ for all $h > 0$.

Now we verify that $|G_h^{-1}|_\infty \leq C_1 < +\infty$ 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/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, $\left|G_h^{-1}\right|_\infty \leq \left|U_h\right|_\infty \left|\Lambda_h^{-1}\right|_\infty \left|U_h'\right|_\infty < +\infty$ for all finite h.

Finally, we verify that G_h^{-1} remains bounded in the uniform norm as $h \to \infty$. Under Condition A/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(ω) and cos(ω) 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 the (limiting) spectral density $f(\omega)$ 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/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:

$$\text{(A3)} \qquad \left| G_\infty^{-1} \right|_\infty = \sum_{j=-\infty}^{\infty} \left| C(j) \right|_\infty \le \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left| B(j) \right|_\infty |\Omega|_\infty \left| B'(k) \right|_\infty < +\infty$$

Proof of Lemma 2(b):

This result was shown in the frequency domain by Hannan and Deistler (1988, p. 271). However, the proof is more straightforward in the time domain. First, note that $x' \Sigma_h x = x' \sum_{j=0}^{h} A_{hj} \Gamma_j x \le x' \sum_{j=0}^{h} A_{\infty j} \Gamma_j x$ for all $x \in \Re^h$, since A_h minimizes the innovation covariance. Now $\sum_{j=0}^{h} A_{\infty j} \Gamma_j = \Sigma_\infty - \sum_{j=h+1}^{h} A_{\infty j} \Gamma_j$. Thus, $|\Sigma_h - \Sigma_\infty|_\infty \le |\sum_{j=h+1}^{\infty} A_{\infty j} \Gamma_j|_\infty$)

$\le \sup_{K>h} |A_{\infty K}|_\infty \sum_{j=h+1}^{\infty} |\Gamma_j|_\infty = O(h^{-2q})$ by Theorem 1(a) and Condition C.

Proof of Lemma 2(c):

Let $M_h = [Q'_h A_h]^{-1}$ and $M_\infty = [Q'_\infty A_\infty]^{-1}$. Then $S_h^{qr} - f(0) = M_h \Sigma_h M'_h - M_\infty \Sigma_\infty M'_\infty$

$= (M_h - M_\infty) \Sigma_h M'_h + M_\infty \Sigma_h (M'_h - M'_\infty) + M_\infty (\Sigma_h - \Sigma_\infty) M'_\infty$. Since $\det[Q'_\infty A_\infty] \neq 0$ by Theorem 1(a), the inverse function is continous at $Q'_\infty A_\infty$, so that $|M_h - M_\infty| = O(h^{-q})$ from Lemma 2(a). Furthermore, $|\Sigma_h - \Sigma_\infty| = O(h^{-2q})$ by Lemma 2(b). Thus, $S_h^{qr} - f(0) = \max\left(|M_h - M_\infty|_\infty, |\Sigma_h - \Sigma_\infty|_\infty \right) = O(h^{-q})$.

Proof of Lemma 3(a):

Using the definition of $\tilde{C}_{Th}$ given in Equation (3.8), we find that

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

Now using standard properties of the vec operator (Lutkepohl 1992, p. 464), and the property that $|A \otimes I|_\infty = |A|_\infty$ for any matrix A, we find that:

$$\text{(A5)} \quad \begin{aligned} \left|vec(Q_h'\left(\widetilde{A}_{Th} - A_h\right))\right|_\infty &\le \left|\left[\widetilde{C}'_{Th}\left(\widetilde{G}_{Th}^{-1} - G_h^{-1}\right)\right]\otimes I\right|_\infty \left|vec(Q_h)\right|_\infty + \left|vec(Q'_h\, G_h^{-1}\widetilde{C}_{Th}\right|_\infty \\ &\le \left|\widetilde{C}'_{Th}\right|_\infty \left|\widetilde{G}_{Th}^{-1} - G_h^{-1}\right|_\infty + \left|vec(Q'_h\, G_h^{-1}\widetilde{C}_{Th}\right|_\infty \end{aligned}$$

since $|\text{vec}(Q_h)|_\infty = 1$. In Lemma A1, we show that $|\widetilde{C}'_{Th}|_\infty = O_p(hT^{-1/2})$. In Lemma A2, we show that $|\widetilde{G}_{Th}^{-1} - G_h^{-1}|_\infty = O_p(hT^{-1/2})$. Thus, the product of these terms is $O_p(h^2T^{-1}) = o_p((h/T)^{1/2})$ for $h = o(T^{1/3})$. Finally, in Lemma A3, we show that $\left|vec(Q'_h\, G_h^{-1}\widetilde{C}_{Th}\right|_\infty$ is $O_p((h/T)^{1/2})$, which completes the proof.

Lemma A1: $\left|\widetilde{C}'_{Th}\right|_\infty = O_p(hT^{-1/2})$ under Conditions A/A*, B, and D/D*.

Proof: Let $\widetilde{c}_{Th,k}$ comprise the k-th $N\times N$ block of $\widetilde{C}_{Th}$. Thus, $\widetilde{c}_{Th,k} = \sum_{j=0}^{h}(\Gamma(j-k) - \widetilde{\Gamma}_T(j-k))A'_{hj}$. Now $|\widetilde{C}'_{Th}|_\infty = \sum_{k=1}^{h}|\hat{c}_{Th,k}|_\infty \le h\max_k|\widetilde{c}_{Th,k}|_\infty \le h\max_{L\in[-h,h]}|\widetilde{\Gamma}_T(L) - \Gamma(L)|_\infty \sum_{j=0}^{h}|A'_{hj}|_\infty$. Then the conclusion of the lemma follows from $|\widetilde{\Gamma}_T(L) - \Gamma(L)|_\infty = O_p(T^{-1/2})$ under Conditions A/A* and D/D* (cf. Andrews 1991), and because $\sum_{j=0}^{h}|A'_{hj}|_\infty$ is uniformly bounded in h as indicated by Lemma 2(a).

Lemma A2: $\left|\widetilde{G}_{Th}^{-1} - G_h^{-1}\right|_\infty = O_p(hT^{-1/2})$ under Conditions A/A*, B, and D/D*.

Proof: Condition A/A* ensures that $|G_h|_\infty = \max_k\sum_{j=1}^{h}|\Gamma(j-k)|_\infty \le \sum_{j=-\infty}^{\infty}|\Gamma(j)|_\infty < +\infty$. Lemma 2(a) indicates that $|G_h^{-1}|_\infty < +\infty$. Conditions A/A* and D/D* ensure that $|\text{vec}(\widetilde{\Gamma}_T(j) - \Gamma(j))|_\infty = O_p(T^{-1/2})$ for $j \le h = o(T)$ (cf. Andrews 1991). Thus, $|\widetilde{G}_{Th} - G_h|_\infty = \max_k\sum_{j=1}^{h}|\widetilde{\Gamma}_T(j-k) - \Gamma(j-k)|_\infty = O_p(hT^{-1/2})$. Now we find that $|\widetilde{G}_{Th}^{-1}|_\infty \le |G_h^{-1}|_\infty + |\widetilde{G}_{Th}^{-1} - G_h^{-1}|_\infty \le |G_h^{-1}|_\infty + |\widetilde{G}_{Th}^{-1}|_\infty |\widetilde{G}_{Th} - G_h|_\infty |G_h^{-1}|_\infty$ so that $|\widetilde{G}_{Th}^{-1}|_\infty \le \dfrac{|G_h^{-1}|_\infty}{1 - |\widetilde{G}_{Th} - G_h|_\infty |G_h^{-1}|_\infty} < \infty$. Finally, since $\widetilde{G}_{Th}^{-1} - G_h^{-1} = -\widetilde{G}_h^{-1}(\widetilde{G}_{Th} - G_h)G_h^{-1}$, we find that $|\widetilde{G}_{Th}^{-1} - G_h^{-1}|_\infty = |\widetilde{G}_h^{-1}|_\infty |\widetilde{G}_{Th} - G_h|_\infty |G_h^{-1}|_\infty = O_p(hT^{-1/2})$

Lemma A3: $\left|vec\left(Q'_h\, G_h^{-1}\,\widetilde{C}_{Th}\right)\right|_\infty = O_p[(h/T)^{1/2}]$ under Conditions A/A*, B, and D/D*.

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, $Q_h = q_{h\cdot}$, 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 $z'_h = q'_h\, G_h^{-1}$, so that $z_{hk} = \sum_{j=1}^{h}F_h(j,k)$. Then $|z_{hk}| < +\infty$ 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:

$$\text{(A6)} \qquad z'_h \tilde{c}_{Th} = \sum_{j=1}^{h} z_{hj} \tilde{c}_{Thj} = \sum_{j=1}^{h}\sum_{k=1}^{h} z_{hj} \left(\tilde{\gamma}_{T,k-j} - \gamma_{k-j}\right)\alpha_{hk}$$

$$= \sum_{L=1-h}^{h} \left(\tilde{\gamma}_{T,L} - \gamma_L\right)\left[\sum_{j=\max(0,\,1-L)}^{\min(h,\,h-L)} z_{hj}\alpha_{h,j+L}\right]$$

$$= \sum_{L=1-h}^{h} b_L \left(\tilde{\gamma}_{T,L} - \gamma_L\right)$$

where $|\, b_L \,| < +\infty$, since $|\, z_{hj} \,| < +\infty$ and $\sum_{j=0}^{\infty} |\alpha_{hj}| < +\infty$. Using the appropriate extension of Bartlett's (1946) result (cf. Priestley 1982, p. 326; Andrews 1991), we find that:

$$\text{(A7)} \qquad E\left(z'_h \tilde{c}_{Th}\right)^2 = \sum_{K=1-h}^{h}\sum_{L=1-h}^{h} b_K\, b_L\, \mathrm{Cov}\left(\tilde{\gamma}_{T,K}, \tilde{\gamma}_{T,L}\right)$$

$$\leq \sum_{K=1-h}^{h}\sum_{L=1-h}^{h} b_K\, b_L \left[\frac{1}{T}\sum_{m=-\infty}^{\infty}\left(\gamma_m\gamma_{m+L-K} + \gamma_{m+L}\gamma_{m-K} + \sup_{t\,\geq 1} K_4(t,K,L,m)\right)\right]$$

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

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

where K_4 is the fourth-order cumulant. Now Condition A/A* ensures that the autocovariances are absolutely summable, and Condition D/D* ensures that the fourth-order cumulants are absolutely summable. Thus, we find that $E\left(z'_h \tilde{c}_{Th}\right)^2 = O(h/T)$, so that $\left|q'_h G_h^{-1} \tilde{c}_{Th}\right| = O[\,(h/T)^{1/2}\,]$ in the scalar case. In the multivariate case, let the $N\times hN$ matrix $z'_h = Q'_h G_h^{-1}$ with the k^{th} $N\times N$ block denoted by z_{hk}. As above, $|\, z_{hk} \,|_\infty < +\infty$, because $|\, G_h^{-1} \,|_\infty < +\infty$. Now we have:

$$\text{(A8)} \qquad Q'_h G_h^{-1}\tilde{C}_{Th} = \sum_{k=1}^{h}\sum_{j=0}^{h} z_{hk}\left(\Gamma(j-k) - \tilde{\Gamma}_T(j-k)\right)A'_{hj}$$

$$= \sum_{L=1-h}^{h}\sum_{j=\max(0,1-L)}^{\min(h,h-L)} z_{h,j+L}\left(\Gamma(L) - \tilde{\Gamma}_T(L)\right)A'_{hj}$$

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

$$\text{(A9)} \qquad vec\left(Q'_h G_h^{-1}\tilde{C}_{Th}\right) = \sum_{L=1-h}^{h}\left[\sum_{j=\max(0,1-L)}^{\min(h,h-L)} A_{hj}\otimes z_{h,j+L}\right] vec\left(\Gamma(L) - \tilde{\Gamma}_T(L)\right)$$

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

where each $N^2 \times N^2$ matrix b_L satisfies $|b_L|_\infty \leq \max_k |z_{hk}|_\infty \sum_{j=0}^{h} |A_{hj}|_\infty < +\infty$. Let w_{hj} be the j-th element of $vec\left(Q'_h \, G_h^{-1} \widetilde{C}_{Th} \right)$, and let b_{Lj} denote the j-th row of b_L, so that $w_{hj} = \sum_{L=1-h}^{h} b_{Lj} \; \text{vec}(\Gamma(L) - \widetilde{\Gamma}_T(L))$. Then we obtain:

$$\text{(A10)} \quad Ew_{hj}^2 = \sum_{K=1-h}^{h} \sum_{L=1-h}^{h} b'_{kj} \, E\left[\text{vec}\left(\Gamma(K) - \widetilde{\Gamma}_T(K)\right) \text{vec}\left(\Gamma(L) - \widetilde{\Gamma}_T(L)\right)' \right] b_{Lj}.$$

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, we find that $Ew_{hj}^2 = O[(h/T)^{1/2}]$ for all $j = 1, \cdots, N^2$, so that $\left| vec\left(Q'_h \, G_h^{-1} \widetilde{C}_{Th} \right) \right|_\infty = O[(h/T)^{1/2}]$.

Proof of Lemma 3(b):

The set of estimated VAR(h) residuals are defined as $\widetilde{\varepsilon}_{ht} = \sum_{j=0}^{h} \widetilde{A}_{Th,j} \, y_{t-j}$ for $t=h+1, \cdots, T$. These residuals satisfy the standard OLS orthogonality conditions: $\sum_{t=h+1}^{T} \widetilde{\varepsilon}_{ht} \, y_{t-j} = 0$ for $j=1, \cdots, h$. Thus, the estimated innovation covariance matrix $\widetilde{\Sigma}_{Th} = \frac{1}{T} \sum_{t=h+1}^{T} \widetilde{\varepsilon}_{ht} \, \widetilde{\varepsilon}'_{ht} = \sum_{j=0}^{h} \widetilde{A}_{Th,j} \, \widetilde{\Gamma}_T(j)$. As noted in the proof of Lemma 2, the innovation covariance matrix of the VAR(h) approximation using the true (limiting) autocovariances can be expressed as $\Sigma_h = \sum_{j=0}^{h} A_{h,j} \Gamma(j)$. Therefore, we have;

$$\text{(A11)} \quad \widetilde{\Sigma}_{Th} - \Sigma_h = \sum_{j=0}^{h} \left(\widetilde{A}_{Th,j} - A_{h,j}\right)\left(\widetilde{\Gamma}_T(j) - \Gamma(j)\right) + \sum_{j=0}^{h} A_{h,j} \left(\widetilde{\Gamma}_T(j) - \Gamma(j)\right)$$

$$+ \sum_{j=0}^{h} \left(\widetilde{A}_{Th,j} - A_{h,j}\right)\Gamma(j)$$

Using the standard properties of the vec operator, we find that:

$$\text{(A12)} \quad \left| vec\left(\widetilde{\Sigma}_{Th} - \Sigma_h\right) \right|_\infty \leq \max_{k=1,\ldots,h} \left| \widetilde{A}_{Th,k} - A_{h,k} \right|_\infty \sum_{j=0}^{h} \left| vec\left(\widetilde{\Gamma}_T(j) - \Gamma(j)\right) \right|_\infty$$

$$+ \sum_{j=0}^{h} \left| A_{h,j} \right|_\infty \max_{k=1,\ldots,h} \left| vec\left(\widetilde{\Gamma}_T(k) - \Gamma(k) \right) \right|_\infty + \max_{k=1,\ldots,h} \left| \widetilde{A}_{Th,k} - A_{h,k} \right|_\infty \sum_{j=0}^{h} \left| vec\left(\Gamma(j)\right) \right|_\infty$$

Now we note that $\max_{k=1,\ldots,h} \left| \widetilde{A}_{Th,k} - A_{h,k} \right|_\infty = \left| \widetilde{A}_{Th} - A_h \right|_\infty$, using the definition of the matrix norm (i.e., the maximum row sum). Then using the matrix $\widetilde{C}_{Th}$ defined in equation (3.7), we find that $|\widetilde{A}_{Th} - A_h|_\infty \leq |\widetilde{G}_{Th}^{-1}|_\infty \, |\widetilde{C}_{Th}|_\infty$. Now $|\widetilde{G}_{Th}^{-1}|_\infty = O_p(1)$ from the proof of Lemma A3. Furthermore, $|\widetilde{C}_{Th}|_\infty = \max_{k=1,\ldots,h} |\widetilde{c}_{Th,k}|_\infty = O_p(T^{-1/2})$ from the proof of Lemma A2. (It is important to note that $|\widetilde{C}_{Th}|_\infty \neq |\widetilde{C}'_{Th}|_\infty$; the rows of $\widetilde{C}_{Th}$ are of length N, whereas the rows of $\widetilde{C}_{Th}$' are of length $h(T)N$.

Thus, we find that $\max_{k=1,\ldots,h} |\widetilde{A}_{Th,k} - A_{h,k}|_\infty = O_p(T^{-1/2})$. Under Conditions A/A* and D/D*, each element of $vec(\widetilde{\Gamma}_T(j) - \Gamma(j))$ also converges to zero at rate $O_p(T^{-1/2})$ (cf. Andrews 1991). Thus, the first term of equation (A12) converges at rate $O_p(h/T)$. Under Conditions A/A* and B, Lemma 2 indicates that the AR(h) coefficients are absolutely summable for all h, so that the second term of equation (A12) vanishes at the rate $O_p(T^{-1/2})$. Finally, Condition A/A* ensures that the autocovariances are absolutely summable, so that the third term of equation (A12) also vanishes at rate $O_p(T^{-1/2})$. Combining these results yields the conclusion of Lemma 3(b).

Proof of Lemma 3(c): This follows directly from parts (a) and (b) of Lemma 3, using the same approach utilized to prove Lemma 2(c).

Proof of Lemma 4: Now following Hannan (1970, p.158), we define the matrix functions exp(·) and log(·) as follows: for any Hermitian positive semi-definite matrix A, there is a unique Hermitian matrix B such that $A = \exp(B)$, where $\exp(B)$ is defined via the exponential series $\sum_{j=0}^{\infty} B^j / j!$, and $B = \log(A)$. Using these definitions, it can be shown that $\log(\det(A))=\text{trace}(\log(A))$ for any positive semi-definite matrix A (cf. Hannan 1970, p159), and that $\log(I + A) \to A$ as $A \to 0$ (cf. Hannan and Kavalieris 1986).

Now using the results of Lemma 3, we find that $\Sigma^{-1}\hat{\Sigma}_h = \Sigma^{-1}\left[\Sigma_h + O_p(h/T)\right]$.

Thus, we obtain $\log\left(\det\left(\hat{\Sigma}_h\right)\right) = \log\left(\det(\Sigma)\right) + \log\left(\det\left(\Sigma^{-1}\left[\Sigma_h + O_p(h/T)\right]\right)\right) = \log\left(\det(\Sigma)\right)$

$+ \text{trace}\left[\Sigma^{-1}\left(\Sigma_h - \Sigma + O_p(h/T)\right)\right] = \log\left(\det(\Sigma)\right) + \text{trace}\left[\Sigma^{-1}\left(\Sigma_h - \Sigma\right)\right] + O_p(h/T)$

since Σ_h approaches Σ as $h \to \infty$, and Σ is a positive definite NxN matrix. Finally, the conclusions of Lemma 4 follow from the definitions of AIC and BIC.

Proof of Theorem 2: Using equation (3.5), parts (a) and (b) follow directly from Lemma 2(c) and Lemma 3(c). Part (c) follows immediately from part (a) and Lemma 4(b).

Proof of Theorem 3: The results of Theorem 1 and Lemma 2 remain unchanged. Now let the j-th sample autocovariance deviation $d_T(j) = vec(\hat{\Gamma}_T(j) - \widetilde{\Gamma}(j))$, 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 $b(j)$ is uniformly bounded for all $j \in [-\infty, +\infty]$. Under Conditions A/A*, D/D*, and 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_T converges to zero at rate $o_p[(h(T)/T)^{1/2}]$ for $h(T) = o(T^{1/2})$. Since the proofs of Lemmas 3 and 4 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.