TECHNICAL WORKING PAPER SERIES

ROBUST COVARIANCE MATRIX ESTIMATION WITH DATA-DEPENDENT VAR PREWHITENING ORDER

Wouter J. den Haan
Andrew T. Levin

Technical Working Paper 255
http://www.nber.org/papers/T0255

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2000

We appreciate comments and suggestions from Yongmiao Hong, Tim Vogelsang, and Jonathan Wright. The views expressed in this paper do not necessarily reflect the views of the National Bureau of Economic Research or the Board of Governors of the Federal Reserve System or of any other members of its staff. GAUSS, RATS, and Fortran programs to calculate the VARHAC estimator can be found on the web-site http://weber.ucsd.edu/~wdenhaan.

.

Robust Covariance Matrix Estimation with Data-Dependent
VAR Prewhitening Order
Wouter J. den Haan and Andrew T. Levin
NBER Technical Working Paper No. 255
June 2000
JEL No. C12, C22

ABSTRACT

This paper analyzes the performance of heteroskedasticity-and-autocorrelation-consistent (HAC) covariance matrix estimators in which the residuals are prewhitened using a vector autoregressive (VAR) filter. We highlight the pitfalls of using an arbitrarily fixed lag order for the VAR filter, and we demonstrate the benefits of using a model selection criterion (either AIC or BIC) to determine its lag structure. Furthermore, once data-dependent VAR prewhitening has been utilized, we find negligible or even counter-productive effects of applying standard kernel-based methods to the prewhitened residuals; that is, the performance of the prewhitened kernel estimator is virtually indistinguishable from that of the VARHAC estimator.

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

Andrew T. Levin
International Finance Division
Federal Reserve Board of Governors
Washington, DC 20551-0001
levina@frb.gov

1. Introduction

Existing procedures for constructing heteroscedasticity-and-autocorrelation-consistent (HAC) covariance matrices have largely focused on kernel-based methods of estimating the spectral density matrix at frequency zero. Given that these methods tend to yield relatively poor inference properties in the presence of strong temporal dependence (cf. Andrews 1991), Andrews and Monahan (1992) proposed a class of kernel-based HAC estimators that incorporate a fixed order of vector autoregressive (VAR) prewhitening. As originally suggested by Press and Tukey (1956), the approach of prewhitening is intended to flatten the relevant portion of the spectral density function, thereby reducing the bias of the kernel estimator and hence permitting the use of a smaller bandwidth parameter when the kernel is applied to the prewhitened residuals (cf. Priestley 1981, pp.556-7). In practice, the simulation experiments of Andrews and Monahan (1992) utilized first-order VAR prewhitening, and this specification has generally been followed in subsequent research (e.g., Newey and West 1994).

In this paper, we highlight the pitfalls of using a fixed order of VAR prewhitening, and then we analyze the benefits of using data-dependent VAR prewhitening in constructing HAC covariance matrix estimators. In particular, we consider the use of either Akaike's (1973) Information Criterion (AIC) or Schwarz' (1978) Bayesian Information Criterion (BIC) to select the lag structure of the VAR model. We evaluate the use of data-dependent VAR prewhitening in conjunction with the kernel-based methods studied by Andrews (1991) and Andrews and Monahan (1992). We also consider the VARHAC procedure proposed by Den Haan and Levin (1996, 1997), in which the spectral density matrix at frequency zero is constructed directly from the VAR model, rather than applying a kernel-based method to the prewhitened residuals.

We utilize Monte Carlo simulation experiments to evaluate the finite-sample performance of these procedures in generating accurate confidence intervals for linear regression coefficients. In comparing these procedures, we find that dramatic improvements in inference accuracy can be obtained by permitting the lag order to vary across the equations of the VAR, and indeed, to vary across the explanatory variables within each equation. This flexibility of VAR prewhitening is in striking contrast to one of the fundamental restrictions of kernel-based methods, namely, that the entire spectral density matrix must be constructed using a single bandwidth parameter to ensure a positive semi-definite HAC covariance matrix (cf. Robinson 1996).

Furthermore, once data-dependent VAR prewhitening has been utilized, we find relatively little benefit from applying a kernel-based method to the prewhitened residuals. The bandwidth selection method used in the simulation studies of Andrews (1991) and

Andrews and Monhana (1992) focuses mainly on the degree of low-order autocorrelation of the prewhitened residuals; hence, this method tends to yield a low value of the bandwidth parameter, even if the residuals are not truly "white noise." Thus, the inferences obtained using the VAR-prewhitened kernel estimator tend to be virtually indistinguishable from those obtained using the VARHAC estimator, at least for all of the data-generating processes (*dgp*s) considered here.

The remainder of this paper is organized as follows: Section 2 discusses the steps used in constructing each of the various HAC covariance matrix estimators. Section 3 reports on the Monte Carlo simulation results. Section 4 summarizes our conclusions and highlights some issues for further research.

2. HAC Covariance Matrix Estimation

This section describes the procedures used in constructing HAC covariance matrix estimators. As shown in section 2.1, the principal challenge is to estimate the spectral density matrix at frequency zero. Section 2.2 describes the VARHAC procedure proposed by Den Haan and Levin (1996, 1997). Section 2.3 highlights several important pitfalls associated with kernel-based estimation procedures. Section 2.4 reviews the VAR-prewhitened kernel procedure of Andrews and Monahan (1992), and indicates how this method can be generalized to allow for data-dependent VAR prewhitening order.

2.1 The General Problem

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

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

where $V_T = \sum_{t=1}^{T} V_t(\psi)/T$ is the vector of sample moments of $V_t(\psi)$, and the $N\times N$ symmetric weighting matrix M_T may be deterministic or stochastic (cf. Hansen 1982).

Under certain regularity conditions, $\sqrt{T}\left(\hat{\psi}_T - \psi_0\right)$ has a limiting normal distribution with mean vector 0 and covariance matrix $\Omega = 2\pi\, B'\, f(0)\, B$, where $f(0)$ denotes the limiting spectral density at frequency zero of the process

$V_t(\psi_0)$. In particular, the $N \times N$ matrix $f(0)=\lim_{T\to\infty} S_T / 2\pi$; the $N \times p$ matrix $B = \lim_{T\to\infty} M_T D_T (D_T' M_T D_T)^{-1}$, and the matrices S_T and D_T are defined as follows:

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

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

The matrix D_T is typically estimated by its sample analog $\hat{D}_T \equiv D_T(\hat{\psi}_T)$, where $\hat{D}_T - D_T \to 0$ in probability as $T \to \infty$. Thus, the key challenge in estimating the asymptotic HAC covariance matrix Ω is to estimate the spectral density matrix, $f(0)$.

2.2 The VARHAC procedure

By specifying a parametric time series model for the $N \times 1$ vector process $V_t(\hat{\psi}_T)$, (henceforth referred to simply as $\hat{V}_t$), an estimate of the spectral density matrix at frequency zero can be constructed directly from the estimated parameters of the model. This approach has a long history in the statistics literature (cf. Press and Tukey 1956; Blackman and Tukey 1958; Grenander and Rosenblatt 1956). In particular, Parzen (1969) identified several advantages of estimating the spectral density using an autoregressive (AR) model rather than a non-parametric kernel, and these advantages were subsequently highlighted in a variety of simulation experiments (e.g., Beamish and Priestley 1981; Kay and Marple 1981; Parzen 1983). Asymptotic results concerning the properties of AR spectral estimation may be found in Berk (1974), An, Chen and Hannan (1982), Hannan and Kavalieris (1983), and Den Haan and Levin (1998).

Nevertheless, parametric spectral estimation has not been widely used to date in HAC covariance matrix estimation. Andrews (1991) and Andrews and Monahan (1992) briefly considered a first-order AR spectral estimator, but the estimator did not correct for heteroskedasticity and performed poorly in simulation experiments. Stock and Watson (1993) utilized second and third-order VAR spectral estimation in simulation experiments and in an empirical application. Finally, Eichenbaum et al. (1988) and West (1997) proposed a class of HAC covariance matrix estimators for the case of

a vector moving-average (VMA) process of known finite order.

Here we consider the VARHAC procedure proposed by Den Haan and Levin (1996, 1997), which uses a model selection criterion to select the best VAR representation of V_t. (For comparison, VAR estimators with fixed lag order will also be considered in the simulation experiments in Section 3.)

Step 1. Lag order selection for each VAR equation. For each equation $n = 1,..,N$ of the VAR, we permit the number of lagged values h_{n1} of the dependent variable $\hat{V}_{nt}$ (that is, the n^{th} element of $\hat{V}_t$) to differ from the number of lagged values h_{n2} of the other *N-1* residuals $\hat{V}_{mt}$ for $m \neq n$. (Note that the restriction that the lag order h_{n2} is identical for all $m \neq n$ drastically reduces the computational burden when N is large, but need not be imposed when N is very small.)

Thus, for every combination of the lag orders $h_{n1} = 0,..., \overline{H}$ and $h_{n2} = 0,..., \overline{H}$, the following model is estimated by ordinary least squares (OLS):

$$\text{(2.4)} \qquad \hat{V}_{nt} = \sum_{k=1}^{h_{n1}} \hat{\alpha}_{nnk}(h_{n1}, h_{n2}) \hat{V}_{n,t-k} + \sum_{m \neq n}^{N} \sum_{k=1}^{h_{n2}} \hat{\alpha}_{nmk}(h_{n1}, h_{n2}) \hat{V}_{m,t-k} + \hat{e}_{nt}(h_{n1}, h_{n2})$$

for $t = \overline{H} + 1, ..., T$. Of course, when $h_{n1} = h_{n2} = 0$, no estimation is required, and $\hat{e}_{nt}(0,0) \equiv \hat{V}_{nt}$. Note that the maximum lag length $\overline{H}$ should increase with the sample size. (For example, the asymptotic analysis in Den Haan and Levin (1998) assumes that $\overline{H}(T) = \boldsymbol{O}[\, T^{1/3}\,]$). However, for notational simplicity, we suppress the dependence of $\overline{H}$ on T.

Using these OLS results, the value of the model selection criterion (either AIC or BIC) is calculated for each combination of lag orders h_{n1} and h_{n2}:

$$\text{(2.5)} \qquad \text{BIC}(h_{n1}, h_{n2}) = \log\left(\sum_{t=\overline{H}+1}^{T} \hat{e}_{nt}^2(h_{n1}, h_{n2}) \right) + \frac{(h_{n1} + h_{n2}(N-1)) \log(T - \overline{H})}{T - \overline{H}}$$

$$\text{(2.6)} \qquad \text{AIC}(h_{n1}, h_{n2}) = \log\left(\sum_{t=\overline{H}+1}^{T} \hat{e}_{nt}^2(h_{n1}, h_{n2}) \right) + \frac{2(h_{n1} + h_{n2}(N-1))}{T - \overline{H}}$$

Then the optimal combination of lag orders h_{n1}^* and h_{n2}^* for the n^{th} VAR equation

are chosen as the values of h_{n1} and h_{n2} that minimize the model selection criterion. Note that when h_{n1}^* and h_{n2}^* are both strictly less than $\overline{H}$, the n^{th} VAR equation can be reestimated using a slightly longer sample, namely $t = \max(h_{n1}^*, h_{n2}^*),...,T$; this approach is followed in the simulation experiments reported in Section 3.

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

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

where $\hat{e}_t$ is an $N\times 1$ vector with typical element $\hat{e}_{nt}(h_{n1}^*, h_{n2}^*)$, and $\hat{A}_0$ is the $N\times N$ identity matrix. The (n,n) element of $\hat{A}_k$ equals $-\hat{\alpha}_{nnk}(h_{n1}^*, h_{n2}^*)$ if $k \le h_{n1}^*$, and equals zero if $k > h_{n1}^*$. For $m \neq n$, the (n, m) element of $\hat{A}_k$ equals $-\hat{\alpha}_{nmk}(h_{n1}^*, h_{n2}^*)$ if $0 < k \le h_{n2}^*$, and equals zero if $k > h_{n2}^*$. Finally, the innovation covariance matrix $\hat{\Sigma}_T$ is estimated as follows:

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

As an alternative to the OLS regression methods used here, seemingly unrelated regression (SUR) methods could be used to obtain joint estimates of the restricted VAR parameters and the innovation covariance matrix; SUR estimation would yield more efficient parameter estimates if the innovation covariance matrix has non-negligible off-diagonal elements.

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

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

Finally, the VARHAC covariance matrix estimator is defined by:

$$\hat{\Omega}_T = \left[\hat{D}_T' M_T \hat{D}_T\right]^{-1} \hat{D}_T' M_T \hat{S}_T M_T \hat{D}_T \left[\hat{D}_T' M_T \hat{D}_T\right]^{-1} \tag{2.10}$$

2.3 Kernel-based methods

The true spectral density matrix at frequency zero, *f(0)*, is an unweighted sum of the autocovariances of the process V_t; this property provides the basic motivation for kernel-based spectral estimators, which have the following form:

$$\hat{S}_T(\kappa,\xi_T) = \sum_{j=-T+1}^{T-1} \kappa(\frac{j}{\xi_T})\,\hat{\Gamma}_j\,, \tag{2.11}$$

where $\kappa(\cdot)$ is a weighting function referred to as the kernel; ξ_T is referred to as the bandwidth parameter; and $\hat{\Gamma}_j$ is the j^{th}-order sample autocovariance matrix of the process $\hat{V}_t$ (with $\hat{\Gamma}_j = \hat{\Gamma}'_{-j}$ for $j < 0$). The remainder of this section highlights several important issues that arise in implementing kernel-based methods.

2.3.1 The Choice of Kernel

An efficient choice for $\kappa(\cdot)$ is the simple truncated kernel, where $\kappa(x) = 1$ for $|x| \leq 1$ and $\kappa(x) = 0$ for $|x| > 1$. For infinite-order ARMA processes, Den Haan and Levin (1998) demonstrated that the asymptotic bias and asymptotic variance are of the same order for both the simple truncated kernel and the VAR spectral estimator (a result that was originally conjectured by Parzen 1969). Unfortunately, the simple truncated kernel does not necessarily generate a positive semi-definite (PSD) spectral density estimator $\hat{S}_T(\kappa,\xi_T)$, so that the resulting covariance matrix $\hat{\Omega}_T$ may not be invertible.

Within the class of kernels that ensure a PSD spectral density matrix, a simple choice is the Barlett kernel (cf. Newey and West 1987, 1994), which is defined as $\kappa(x) = 1 - |x|$ for $|x| \leq 1$ and $\kappa(x) = 0$ for $|x| > 1$. However, small efficiency gains may be obtained in some cases by using the quadratic spectral (QS) kernel, which is optimal in the class of PSD kernels (cf. Priestley 1962; Andrews 1991):

$$\kappa_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right). \tag{2.12}$$

Nevertheless, it should be emphasized that using a PSD kernel is associated with non-negligible efficiency costs compared with parametric methods (such as VAR spectral estimation), for which the estimated spectral density matrix (and hence the HAC covariance matrix) is PSD by construction. In finite samples, this efficiency cost is particularly apparent when the true *dgp* is known to be a MA process of order *m*: the optimal value of $\xi_T = m$ for the simple truncated kernel, whereas this choice of

ξ_T yields very poor results for PSD kernel-based spectral estimators (cf. Ogaki 1993). Asymptotically, Den Haan and Levin (1998) have shown under very general conditions (e.g., unconditional heteroskedasticity or even the absence of covariance stationarity) that the mean-squared error (MSE) of the VARHAC estimator shrinks at a faster rate than the MSE of any PSD kernel-based estimator.

2.3.2 **Constructing a Data-Dependent Bandwidth Parameter**

Under certain regularity conditions, it is possible to derive the asymptotically optimal sequence of bandwidth parameters for a particular PSD kernel; that is, the sequence which minimizes the asymptotic MSE of the spectral estimator (cf. Priestley 1981; Andrews 1991). The optimal sequence, ξ_T^*, depends on the smoothness of the kernel (as indicated by the characteristic exponent q), as well as the true spectral density and its generalized q^{th} derivative at frequency zero. In settings in which V_t contains multiple elements, it is also necessary to specify an $N^2 \times N^2$ weighting matrix W. Then the asymptotically optimal bandwidth parameter sequence for the QS kernel is given by:

$$\xi_T^* = 1.3221\left[\frac{2\ \mathrm{vec}(S^{(2)})' W\ \mathrm{vec}(S^{(2)})}{\mathrm{trace}\left(W(I+K)(S\otimes S)\right)}\ T\right]^{1/5} \tag{2.13}$$

where K is the $N^2 \times N^2$ commutation matrix that transforms vec(B) into vec(B'), and where $S^{(2)}$ indicates the generalized second derivative of the spectral density at frequency zero:

$$S^{(2)} = \sum_{j=-\infty}^{\infty} j^2\, C_j \tag{2.14}$$

Thus, to estimate the value of the optimal bandwidth parameter ξ_T^* using equation (2.13), we need initial estimates of the spectral density matrix S and its generalized second derivative $S^{(2)}$. Andrews (1991) proposed that parametric methods be used to provide these initial estimates; in the simulation experiments performed by Andrews (1991) and Andrews and Monahan (1992), univariate AR(1) representations were used for this purpose.

Nevertheless, it is important to recognize that a poor choice of parametric model will generate relatively poor initial estimates of S and $S^{(2)}$, and hence the actual data-dependent bandwidth parameter ξ_T may differ substantially from its optimal value ξ_T^*. Thus, given that the accuracy of kernel-based spectral estimation can be very sensitive to the value of the bandwidth parameter (cf. Andrews 1991), a poor choice for the initial parametric model may lead to substantial distortion of the final spectral estimator

$\hat{S}_T(\kappa, \xi_T)$. Of course, since the true *dgp* is unknown in practice, one might want to consider employing a model selection criterion to choose an appropriate parametric model, which would then be used in constructing the data-dependent bandwidth parameter for the kernel-based spectral estimator. However, it would seem equally reasonable simply to estimate the spectral density directly from the estimated parameters of the chosen model (as in the VARHAC procedure).

2.3.3 **Choosing the Weighting Matrix**

When V_t contains multiple elements, the data-dependent bandwidth parameter given by equation (2.13) depends on the particular choice of the weighting matrix W. The simulation experiments of Andrews (1991) and Andrews and Monahan (1992) utilized a diagonal weighting matrix; in this case, the value of ξ_T^* does not depend on any cross-products between different elements of S (or $S^{(2)}$). These authors also assigned zero weights to all off-diagonal elements of S and $S^{(2)}$; for this specification of W, the value of ξ_T^* does not depend on the cross spectrum of the data. Finally, in the context of a least-squares estimation problem, Andrews and Monahan (1992) assigned zero weight to the diagonal element that corresponded to the regression intercept, and assigned unit weight to each of the other $N - 1$ diagonal elements (corresponding to the slope coefficients). Unfortunately, this choice of weights makes the bandwidth parameter sensitive to the scaling of the data, which can lead to highly unsatisfactory results; further analysis of this issue may be found in Den Haan and Levin (1997).

Finally, it should be emphasized that the entire spectral density matrix must be constructed using the same bandwidth parameter in order to ensure a PSD spectral density estimate (and hence an invertible HAC covariance matrix). As pointed out by Robinson (1996), this constraint can yield very low accuracy when the autocovariance structure varies substantially across elements. In contrast, the VARHAC estimator can permit the lag order to vary across equations in the VAR and across the variables in each equation, since the resulting covariance matrix is PSD by construction.

2.3.4 **Prewhitening**

It has long been understood that kernel-based spectral estimation performs very poorly in the presence of strong temporal dependence. As seen in equation (2.14), the optimal bandwidth parameter ξ_T^* is large (and hence the kernel-based spectral estimate has high sampling variance) when the spectral density exhibits strong curvature at frequency zero (as in the case of a dgp with a large AR root). Thus, as originally proposed by Press and Tukey (1956), the approach of prewhitening is intended to flatten

the relevant portion of the spectral density function, thereby reducing the bias of the kernel estimator and hence permitting the use of a smaller bandwidth parameter when the kernel is applied to the prewhitened residuals (cf. Priestley 1981, pp.556-7).

Two prewhitening kernel-based estimators are considered in this paper. First, the prewhitening kernel-based estimator from Andrews and Monahan (1992) applies a prewhitening VAR filter of order b and then applies the kernel-based procedure of Andrews (1991) to the "prewhitened" residuals. (Thus, when b is set equal to zero, this estimator is identical to that of Andrews 1991.) The second estimator uses a model selection criterion to determine the lag structure of the VAR prewhitening filter, using exactly the same procedures described in section 2.2, and then applies the kernel-based procedure of Andrews (1991) to the residuals from this procedure.

Andrews and Monahan (1992) only consider fixed values for b. In their Monte Carlo experiments, b is set equal to zero or one for each element of $V_t(\hat{\psi}_T)$. Note that we have placed the term "prewhitened" in quotation marks, because no correction for serial correlation would be needed if the residuals were truly prewhitened. The spectral density at frequency zero of the "prewhitened" residuals is given by:

$$\hat{\Sigma}_T^{QS-PW} = \sum_{j=1-T}^{T-1} \kappa(\frac{j}{\hat{\xi}_T})\hat{\Gamma}_T(j), \quad \text{where}$$

$$\hat{\Gamma}_T(j) = \frac{1}{T}\sum_{t=1}^{T-j} \hat{e}_t \hat{e}'_{t+j} \quad \text{for } j \geq 0 \quad \text{and,} \tag{2.23}$$

$$\hat{\Gamma}_T(j) = \hat{\Gamma}'_T(-j) \quad \text{for } j < 0.$$

The estimate of the spectral density at frequency zero is given by

$$\hat{S}_T^{QS-PW}(\hat{\psi}_T) = \left[I_N - \sum_{k=1}^{b} \hat{A}_k\right]^{-1} \hat{\Sigma}_T^{QS-PW} \left[I_N - \sum_{k=1}^{b} \hat{A}'_k\right]^{-1} \tag{2.24}$$

Finally, it should be noted that using a prewhitening filter does not circumvent any of the inherent problems noted above. The obvious exception is when the data are truly prewhitened, but in that case, there is no rationale for applying kernel-based spectral estimation to the prewhitened residuals.

It should also be noted that Phillips and Lee (1994) have analyzed the properties of an ARMA-prewhitened HAC estimator for the case of a finite-order ARMA process with i.i.d. innovations.

3. Finite-Sample Properties

In this section, Monte Carlo simulation experiments are used to compare the finite-sample properties of the autoregressive parametric estimators with the prewhitening kernel-based estimators. In all experiments, the results are computed for sample length $T = 128$ using 10,000 replications. Additional simulation results and comparisons with other kernel-based estimators may be found in Den Haan and Levin (1996, 1997).

3.1 Scalar Autoregressive Processes

In our first set of experiments, we perform inferences concerning the mean of the following scalar process:

$$(3.1) \qquad Y_t = \mu + V_t \text{ where } V_t = \frac{0.5}{p} \sum_{k=1}^{p} V_{t-k} + \varepsilon_t$$

where the AR order $p \in \{1\ 2\ 3\ 4\}$, the disturbance ε_t is an i.i.d. standard normal process, and we normalize the true mean $\mu = 0$. In this case, the HAC variance of the sample mean is obtained by estimating the spectral density at frequency zero of the estimated residual $\hat{V}_t = Y_t - \hat{\mu}$. For each HAC variance estimator, we compute the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the sample mean of Y_t.

3.1.1 Autoregressive HAC Estimators

We first consider the performance of HAC variance estimators that are computed solely from an AR model for $\hat{V}_t$. The AR order is either fixed *a priori* at a value between 1 and 4, or is determined by minimizing either AIC or BIC subject to the maximum lag order $\overline{H} = 4$; however, the results are virtually identical for $\overline{H} = 7$. (The specification $\overline{H} = 4$ is used throughout the remainder of this paper in implementing AIC and BIC) We also consider HAC variance estimators in which the QS kernel is used to compute the spectral density at frequency zero of the AR-prewhitened residuals. In each case, the bandwidth parameter is determined using the same method as in the simulation experiments of Andrews and Monahan (1992), namely, formula (2.13) is evaluated using the OLS parameter estimates of an AR(1) model of the prewhitened residuals.

For each value of the true lag order p, Panel A of Table 1 indicates the true confidence level obtained using each autoregressive HAC estimator, as well as the mean

values of the lag orders chosen by AIC and BIC. It is immediately apparent that the largest inference distortions occur when the specified AR order is substantially smaller than the true AR order. As noted in Section 2, the prewhitening order h was fixed at unity in the Monte Carlo experiments of Andrews and Monahan (1992), and this specification has been followed in numerous subsequent applications. Nevertheless, when the true AR order $p = 4$ (which is a plausible value for quarterly data), using a fixed value of $h = 1$ yields a true confidence level of only 70 percent.

Not surprisingly, the most accurate significance level is always obtained when the AR order h is fixed at its true value p (although even in this case some distortion results from downward bias of the estimated AR coefficient(s)). It is interesting to note that relatively little cost is associated with using a slightly higher AR order, presumably because 128 observations are sufficient to ensure that the extra AR coefficients have estimated values near zero (and hence don't have much effect on the distribution of the test statistic).

While both data-dependent methods of choosing the AR order provide more accurate inferences than simply fixing $h = 1$, it should be noted that AIC performs much better than BIC in this example. (Hall (1994) and Ng and Perron (1995) obtained similar results in the context of data-dependent AR lag order selection for unit root tests.) The penalty function of BIC grows at a sufficiently rapid rate to ensure that the true AR lag order is chosen consistently, whereas AIC tends to overestimate the lag order penalty even when the sample size becomes arbitrarily large (cf. Shibata 1976). Nevertheless, these are not necessarily the relevant considerations in choosing the appropriate AR lag order for HAC covariance matrix estimation. In small samples, the lag order chosen by BIC is typically *smaller* than the true lag order, leading to substantial bias in the estimated spectral density at frequency zero, whereas the somewhat higher lag order chosen by AIC reduces this bias at relatively little cost in terms of additional variance.

3.1.2 Prewhitened Kernel HAC Estimators

Panel B of Table 1 shows the effects of applying the QS kernel to the prewhitened AR residuals. When the fixed or data-dependent AR lag order h is at least as large as the true lag order p, no benefits are gained by applying the QS kernel to the estimated AR residuals. In these cases, the estimated residuals are truly prewhitened in the sense that relatively little serial correlation remains. Thus, the automatic bandwidth selection procedure tends to choose a bandwidth parameter very close to zero, and hence the

prewhitened kernel estimate is virtually indistinguishable from the AR estimate of the HAC variance.

One might have hoped that applying a kernel to the estimated residuals would yield substantial benefits in the cases where $h < p$, since these are the cases for which the "prewhitened" residuals are furthest from being white noise. Unfortunately, applying the kernel actually reduces the accuracy of the confidence interval in these cases. For example, when the true AR order $p = 2$ and the prewhitening order $h = 1$, the AR estimator yields a true confidence level of 78.1 percent whereas the prewhitened kernel estimator yields a true confidence level of 76.0 percent. The basic problem is that the data-dependent bandwidth parameter is typically fairly small, so that the kernel assigns substantial weight to the first-order autocovariance (which is negative) and assigns much smaller weights to the higher-order autocovariances (which are positive and decline slowly toward zero). Consequently, applying the kernel exacerbates the downward bias of the spectral density estimate and hence reduces the accuracy of the HAC variance estimate.

3.2 Scalar Moving-Average Processes

Now we analyze inferences concerning the mean of the following scalar process:

$$(3.2) \qquad Y_t = \mu + V_t \quad \text{where} \quad V_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

where ε_t is an i.i.d. standard normal process. We consider values of the MA(1) parameter θ between -0.1 and -0.9, and we normalize the true mean $\mu = 0$. As in the previous set of experiments, the HAC variance of the sample mean of Y_t is obtained by estimating the spectral density at frequency zero of the estimated residual $\hat{V}_t = Y_t - \hat{\mu}$. For each HAC variance estimator, we compute the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the sample mean.

As in the previous experiment, prewhitening is performed using scalar AR models for which the lag order is either fixed at unity (as in the simulation experiments of Andrews and Monahan 1992) or determined by either AIC or BIC up to a maximum lag order of 4. When the QS kernel is applied to the prewhitened residuals, we determine the bandwidth parameter using one of three alternative parametric models for the prewhitened residuals: an AR(1) model estimated by OLS; an MA(1) model estimated by non-linear least squares (NLLS); and an MA(2) model estimated by NLLS. For the latter two methods, we use the IMSL subroutine DNSLSE with no backcasting.

3.2.1 Autoregressive HAC Estimators

Panel A of Table 2 provides inference results (together with the mean lag order chosen by each model selection criterion) when the HAC variance is computed solely from the estimated AR model. When the parameter θ is relatively large in absolute value, a relatively high-order AR representation is required to approximate this autocovariance structure to a given level of accuracy. Thus, as shown in panel A, the HAC estimator with a fixed lag order $h = 1$ performs relatively poorly compared with estimators in which the lag order is determined by either AIC or BIC.

As in the previous set of experiments, we find that the less conservative penalty function of AIC tends to yield a higher lag order which in turn facilitates more accurate estimates of the spectral density and hence more accurate confidence levels. In fact, Shibata (1980, 1981) demonstrated that AIC is asymptotically efficient in choosing the AR lag order when the true AR order is infinite; that is, AIC minimizes the prediction error variance and the MSE of the integrated spectrum for infinite-order AR processes.

Finally, as θ approaches the limiting value of -1, the MA representation of Y_t cannot be inverted into an infinite-order AR representation; thus, we are not surprised to find that all of the AR estimators perform relatively poorly when θ = -0.7 or -0.9.

3.2.2 Kernel HAC Estimators

Before evaluating the performance of HAC estimators in which the QS kernel is applied to the AR-prewhitened residuals, it is useful to consider kernel HAC estimators with no prewhitening at all (that is, estimators in which the kernel is applied directly to $\hat{V}_t$). For each method of determining the bandwidth parameter, Panel B of Table 2 provides inference results as well as the mean value of the bandwidth parameter. When the bandwidth parameter is determined using an AR(1) model (as in the simulation experiments of Andrews and Monahan 1992 and in many subsequent applications), the kernel HAC estimator performs very poorly, especially when $\theta = -0.7$ or -0.9. In contrast, when the correctly specified MA(1) model is used in determining the bandwidth parameter, the kernel HAC estimator yields fairly accurate inferences for all five values of θ. Finally, inference accuracy deteriorates somewhat when the over-parameterized MA(2) model is used to determine the bandwidth parameter.

These differences in inference accuracy can be largely understood by considering the characteristics of the optimal bandwidth parameter for the true MA(1) process V_t. With a larger bandwidth parameter, the QS kernel assigns weight closer to unity on the

first-order sample autocovariance (thereby reducing the bias of the spectral density estimate), and also assigns larger weights to the higher-order sample autocovariances (thereby raising the variance of the spectral density estimate, because the true value of these autocovariances is zero). Given this bias-variance tradeoff, the asymptotic MSE is minimized by the bandwidth parameter sequence in formula (2.13). In particular, the true spectral density at frequency zero $S = (1 + \theta)^2$ and its generalized second-derivative $S^{(2)} = 2\theta$. Thus, as θ approaches -1, this MA(1) process is associated with very strong curvature of the spectral density at frequency zero; that is, as the ratio $S^{(2)} / S$ becomes large, the optimal bandwidth parameter implied by formula (2.13) also becomes large and very sensitive to the exact value of θ. In fact, when the bandwidth parameter is determined using the correctly specified MA(1) model, the mean bandwidth parameter value increases from 2.1 for $\theta = -0.1$ to 12.3 for $\theta = -0.7$, and to 65.4 for $\theta = -0.9$.

In contrast, an AR(1) model with negative first-order autocorrelation is associated with very low curvature of the spectral density at frequency zero and hence a relatively small value of the optimal bandwidth parameter. Thus, when the data actually follow an MA(1) process but the bandwidth parameter is obtained from formula (2.13) using parameter estimates from a misspecified AR(1) model, the bandwidth parameter will generally be much smaller than the optimal value. In this particular experiment, the mean bandwidth parameter value obtained using the AR(1) parametric model reaches a plateau of only 2.5 as the true MA(1) parameter θ approaches -1.

These results indicate the importance of discriminating between AR and MA processes in order to obtain an appropriate value of the bandwidth parameter. More generally, one might want to consider employing a model selection criterion to choose an appropriate parametric model for the prewhitened residuals. As noted in Section 2.3.2, however, it would seem equally reasonable to estimate the spectral density directly from the estimated parameters of the chosen model (as in the VARHAC procedure).

3.2.3 Prewhitened Kernel HAC Estimators

Now we consider the performance of HAC variance estimators in which the QS kernel is applied to AR-prewhitened residuals. For each method of determining the bandwidth parameter, Panel C of Table 2 reports the true confidence level and mean bandwidth parameter value when the AR prewhitening order is fixed at unity. Panels D and E provide corresponding results when the prewhitening order is chosen by minimizing AIC and BIC, respectively. For ease of reference, the top row of each panel indicates the inference results from Panel A for the corresponding autoregressive

HAC estimator, or equivalently, those obtained when the bandwidth parameter is fixed at zero.

When the bandwidth parameter is determinedusing either an AR(1) or MA(1) model, it is evident in all three panels that applying the kernel to the prewhitened residuals has negligible effects on inference (relative to the autoregressive HAC estimator). The underlying problem is that both the AR(1)and MA(1) models yield very low mean values of the bandwith parameter, because the first-order autocorrelation of the prewhitened residuals is very close to zero.

For example, consider the properties of the AR(1)-prewhitened residuals as the sample becomes arbitrarily large. Since V_t has first-order autocorrelation $\rho = \theta / (1 + \theta^2)$, the AR(1)-prewhitened series $\breve{V}_t = V_t - \rho V_{t-1}$ is described by the following MA(2) process:

$$(3.3) \quad \breve{V}_t = \varepsilon_t + \frac{\theta^3}{1+\theta^2}\varepsilon_{t-1} - \frac{\theta^2}{1+\theta^2}\varepsilon_{t-2}$$

The first-order autocorrelation of $\breve{V}_t$ is $\theta^3 / (1 + 2\theta^2 + 2\theta^4 + \theta^6)$, which approaches a value of only $-1/6$ as θ approaches -1. Thus, when either an AR(1) or MA(1) model is fit to $\breve{V}_t$, the estimated process has low curvature of the spectral density at frequency zero, and hence formula (2.13) yields a relatively low bandwidth parameter. (The MA(1) model yields a somewhat higher mean bandwidth parameter than the AR(1) model, for the reasons discussed in the previous subsection.)

In light of this analysis, it is useful to consider the properties of the AR(1)-prewhitened kernel estimator when the bandwidth parameter is determined using the correctly-specified MA(2) model. As seen in the final row of Panel C, this method yields relatively accurate inferences for values of θ in the range of -0.1 to -0.5. Unfortunately, the performance of this estimator begins to deteriorate substantially as θ approaches -1, apparently because sampling error in estimating the MA(2) parameters leads to wider variation in the bandwidth parameter and hence to lower accuracy of the kernel estimate of the spectral density of the prewhitened residuals.

In contrast, when the AR prewhitening order is determined by AIC, Panel D of Table 2 shows that the kernel has relatively little effect on inference accuracy even when the bandwidth parameter is determined using an MA(2) model for the prewhitened residuals. When $|\theta| > 0.1$, AIC frequently chooses a lag order greater than unity, and hence the prewhitened residuals tend to exhibit very small first-order and second-order autocorrelation. Thus, to get substantial improvements in performance from applying

the kernel, one would need to estimate an even higher-order MA process for the prewhitened residuals, which might be subject to more severe sampling error problems like those noted above.

Finally, Panel E indicates the performance of the prewhitened kernel estimator when the AR prewhitening order is chosen by minimizing BIC. As in the previous two panels, applying the kernel has negligible effects when the bandwidth parameter is determined using an AR(1) or MA(1) model, because the prewhitened residuals exhibit very little first-order autocorrelation. However, in contrast to the results for AIC, noticeable improvements in inference accuracy are obtained when an MA(2) model is used to determine the bandwidth parameter. Given its more conservative penalty function, BIC frequently selects first-order prewhitening even when θ is relatively large in absolute value, and hence an MA(2) model is able to capture the remaining serial correlation for the reasons discussed above.

3.3 Bivariate Processes with Heterogeneous Components

In this set of simulation experiments, we perform inferences concerning the significance of the OLS regression estimates of α and β in the following scalar model:

$$Y_t = \alpha X_t + \beta Z_t + u_t, \tag{3.3}$$

The random variable $X_t = 0.95\, X_{t-1} + \varepsilon_t$, where ε_t is i.i.d. $\mathbf{N}(0,1)$. The random variable Z_t is i.i.d. $\mathbf{N}(0, 1/(1-0.95^2))$, so that Z_t has the same variance as X_t. The disturbance $u_t = v_t + \theta v_t$, where v_t is standard normal white noise, and the parameter θ varies from -0.1 to -0.9. Finally, we normalize the true regression coefficients as $\alpha = \beta = 0$.

For this specification, the vector V_t consists of the two components $V_{1t} = u_t X_t$ and $V_{2t} = u_t Z_t$. The first component has the same autocovariance structure as an MA(1) process with parameter 0.95θ, whereas the second component has no serial correlation. Furthermore, because X_t and Z_t are independent, these two components are mutually uncorrelated at all leads and lags, and hence the spectral density $f(0)$ and the asymptotic covariance Ω are diagonal matrices. Thus, to a first approximation, the HAC standard error of $\hat{\alpha}$ is determined by the estimated spectral density of the persistent component $\hat{V}_{1t} = \hat{u}_t X_t$, while the HAC standard error of $\hat{\beta}$ is determined by the estimated spectral density of the idiosyncratic component $\hat{V}_{2t} = \hat{u}_t Z_t$.

3.3.1 VARHAC Estimators

In constructing a bivariate VAR model for $\hat{V}_t$, three alternative approaches are used to determine the lag structure. The first approach simply fixes the lag order at unity for the entire VAR, as in the simulation experiments of Andrews and Monahan (1992). The other two methods use a model selection criterion (either AIC or BIC) to determine the lag order of each variable in each equation (cf. Step 1 of Section 2.2). For each method of determining the VAR lag structure, we analyze the accuracy of inferences when the HAC covariance matrix is calculated directly from the VAR model, and for HAC estimators in which the QS kernel is applied to the VAR residuals.

For each VAR estimator of the HAC covariance matrix, Panel A of Table 3 reports the true confidence levels (at a nominal 90% significance level) of two-tailed t-tests of the significance of $\hat{\alpha}$ and $\hat{\beta}$, and also reports the mean lag orders chosen by AIC and BIC. These results confirm that the inference accuracy of each VARHAC estimator is not sensitive to the degree of heterogeneity across components of $\hat{V}_t$, because the lag order is permitted to vary across equations (and across variables within each equation).

As noted above, the robust standard error of $\hat{\alpha}$ (and hence inferences concerning its statistical significance) primarily depends on the estimated spectral density of $\hat{V}_{1t}$. This time series has the same autocovariance structure as a univariate MA(1) process; that is, the true component V_{1t} is related to its own lagged values but not to lagged values of V_{2t}. It is evident from Panel A that both AIC and BIC detect this pattern of temporal dependence: for increasing values of θ, each model selection criterion chooses an increasing mean value of $\hat{h}_{11}$, while continuing to choose a very small lag order $\hat{h}_{12}$. Furthermore, for both AIC and BIC, the mean value of $\hat{h}_{11}$ is similar to that obtained for the comparable univariate process in Section 3.2. (The mean value is slightly smaller here because the relevant MA parameter is only 0.95θ instead of θ.) Finally, as in the previous two sets of experiments, AIC is less conservative than BIC in choosing each lag order; hence, BIC yields somewhat more accurate inferences when θ equals -0.1 or -0.3, while AIC yields more accurate inferences when θ equals -0.7 or -0.9.

The accuracy of inferences concerning the statistical significance of $\hat{\beta}$ primarily depends on the estimated spectral density of $\hat{V}_{2t}$. Since the true component V_{2t} is serially uncorrelated, both model selection criteria generate very low mean values of both $\hat{h}_{21}$ and $\hat{h}_{22}$, and hence yield true confidence levels that are quite close to the nominal 90% level. In this case, the greater parsimony of BIC results in uniformly better inference accuracy than that obtained using AIC.

3.3.2 VAR-Prewhitened Kernel Estimators

Now we consider the performance of HAC covariance matrix estimators in which the QS kernel is used to estimate the spectral density matrix of the VAR-prewhitened residuals. As noted in Section 2.3.3, a single bandwidth parameter must be used in constructing the entire spectral density matrix in order to ensure that the resulting HAC covariance matrix is PSD. As in the simulation experiments of Andrews and Monahan (1992), the bandwidth parameter is constructed using a diagonal weighting matrix W with values (1, 0, 0, 1) along its diagonal, so that formula (2.13) can be implemented by estimating a univariate parametric model for each VAR-prewhitened residual. We consider the same univariate models as in Section 2.2, namely, an AR(1) model estimated by OLS; an MA(1) model estimated by non-linear least squares (NLLS); and an MA(2) model estimated by NLLS.

For each method of determining the bandwidth parameter, Panel B of Table 3 reports the true confidence level and mean bandwidth parameter value when the VAR prewhitening order is fixed at unity. Panels C and D provide corresponding results when the prewhitening order is determined using AIC and BIC, respectively. For ease of reference, each panel also repeats the inference results from Panel A for the corresponding VARHAC estimator, or equivalently, the results obtained when the bandwidth parameter is fixed at zero.

Panels B, C, and D indicate that applying the kernel to the VAR-prewhitened residuals yields negligible improvements and frequently has counterproductive effects on inference accuracy. The underlying problem is evident from part 3 of each panel: namely, the mean value of the bandwidth parameter is very small regardless of the value of θ. Moreover, in contrast to the results obtained in Section 3.2, this problem is not resolved by using univariate MA(2) models (rather than AR(1) or MA(1) models) to determine the data-dependent bandwidth parameter. Rather, the problem is simply that a single bandwidth parameter must be imposed in constructing the entire spectral density matrix.

If one were only concerned with estimating the spectral density of the first VAR residual (which is obtained by filtering the persistent component V_{1t}), one would prefer to use a relatively large bandwidth parameter; that is, the mean value would be roughly similar to those reported in Panel B of Table 2. In particular, the spectral density at frequency zero S_{11} and its generalized second-derivative $S_{11}^{(2)}$ each decline toward zero as θ approaches -1, but the ratio $S_{11}^{(2)} / S_{11}$ becomes very large. Thus, formula (2.13) indicates that the optimal bandwidth parameter for this component is also very large (that is, in the context of univariate estimation).

On the other hand, if one were only concerned with estimating the spectral density of the second VAR residual (which is obtained by filtering the idiosyncratic component, V_{2t}), one would prefer to to use a bandwidth parameter near zero. For this component, the generalized second derivative $S_{22}^{(2)}$ equals zero, while the spectral density S_{22} is a non-zero constant (namely, the variance of this component); thus, the optimal bandwidth parameter for this component is equal to zero in the univariate context.

Thus, kernel estimation faces an unpleasant compromise, because no single bandwidth parameter will be ideal for analyzing both residuals. When formula (2.13) is used for this particular bivariate system (with equal weight on both components), the bandwidth parameter depends on the ratio $S_{11}^{(2)} / (S_{11} + S_{22})$, which approaches zero as θ approaches –1. Thus, in the limiting case, the kernel fails to capture any of the temporal dependence that remains in the VAR-prewhitened residuals.

In fact, as in Section 3.1, applying the kernel often has counterproductive effects on inference accuracy. Because the bandwidth parameter is generally too small, the kernel assigns substantial weight to the first-order autocovariance (which is negative) and assigns much smaller weights to the higher-order autocovariances (which are positive and decline slowly toward zero). Consequently, applying the kernel exacerbates the downward bias of the spectral density estimate and hence reduces the accuracy of the HAC variance estimate.

More generally, one would expect that applying a kernel to the VAR-prewhitened residuals will tend to yield negligible or even counterproductive effects on inference accuracy whenever the residuals exhibit substantial heterogeneity with respect to the degree of temporal dependence. In such cases, any given value of the bandwidth parameter will be much smaller than the optimal value for a highly persistent component (and hence induce substantial bias in estimating its spectral density) and/or be much larger than the optimal value for a relatively idiosyncratic component (and hence induce excessive variance in estimating its spectral density). And such problems will tend to be exacerbated in estimation problems in which V_t contains a larger number of variables.

3.4 Multivariate Processes

This set of simulation experiments utlizes the same design as in Andrews and Monahan (1992), who considered several linear models in which the coefficient vector ψ_0 is estimated by OLS:

$$Y_t = \psi_0 X_t + u_t \tag{3.4}$$

where the vector of explanatory variables X_t includes a constant term and four random variables. All elements of ψ_0 are normalized to zero. Andrews and Monahan (1992) considered five different sets of assumptions regarding the *dgp*s of the regressors X_t and the disturbance u_t: (a) homoskedastic AR(1) processes; (b) AR(1) processes with multiplicative heteroskedasticity overlaid on the errors; (c) homoskedastic MA(1) processes; (d) MA(1) processes with multiplicative heteroskedasticity overlaid on the errors; and (e) homoskedastic MA(q) processes with linearly declining MA parameters. In each case, a range of different parameter values is considered, so that this set of experiments involves a total of 24 different *dgp*s for the AR(1) examples and 18 different *dgp*s for the MA examples. For each *dgp* and each choice of HAC covariance matrix estimator, we perform a two-tailed t-test of the null hypothesis that the coefficient on the first non-constant regressor is equal to its true value.

3.4.1 VARHAC Estimators

For each *dgp*, Figure 1 reports the true confidence level (at a nominal 90% confidence level) for three alternative methods of determining the VAR lag structure. In particular, the grey column depicts the result obtained when the lag order is fixed at unity for every variable in every VAR equation. It should be noted that this VAR estimator is not the same as the PARA estimator considered by Andrews and Monahan (1992); the PARA estimator was constructed under the assumption of homoskedasticity, and performed very poorly for all of the *dgp*s involving conditional heteroskedasticity. The black column depicts the result when BIC is used to select a separate lag order for the lagged dependent variable and for the other explanatory variables in each VAR equation (as described in Step 1 of Section 2.2). Finally, the white column depicts the corresponding result obtained using AIC.

For each of the AR(1) examples depicted in Panel A, the data-dependent lag order selection procedures yield confidence levels very close to those obtained when the VAR lag order is fixed at unity. All three VAR estimators perform quite poorly when the AR(1) coefficient is equal to 0.9 or 0.95: when the *dgp* is nearly non-stationary, inferences are severely distorted by the strong downward bias of the sample autocovariances. In such cases, one would expect to obtain more accurate inferences by using median-unbiased estimation methods similar to those considered by Andrews and Phillips (1987) and Andrews and Chen (1994).

For almost all of the MA examples depicted in Panel B, AIC yields more accurate confidence levels than BIC. As in the univariate examples considered in Sections 3.1 and 3.2, AIC tends to select a somewhat higher lag order, which substantially reduces the

spectral estimation bias at relatively little cost in terms of additional variance. Finally, for the higher-order MA(q) processes on the right side of Panel B, either data-dependent lag order selection method performs noticeably better than simply fixing the VAR lag order at unity.

As noted previously, the VAR lag structure can be determined with a considerable degree of flexibility, because the VARHAC covariance matrix estimator is PSD by construction. For the results depicted in Figure 1, either AIC or BIC selects a separate lag order for the lagged dependent and for the other variables in each VAR equation. It is useful to compare this approach (henceforth referred to as "asymmetric" lag order selection) with the more restrictive approach of selecting a single lag order for all of the variables in each VAR equation (henceforth referred to as "symmetric" lag order selection).

For each model selection criterion, Figure 2 indicates that asymmetric lag order selection yields more accurate inferences than symmetric lag order selection for virtually all of the *dgp*s under consideration; that is, inference accuracy is generally enhanced by allowing greater flexibility in determining the VAR lag structure. Of course, in practical applications, one could consider the even more flexible approach of choosing a separate lag order for every variable in every equation; unfortunately, this approach is not computationally feasible for the present simulation experiment, because too many alternative VAR models would have to be considered at each Monte Carlo replication.

3.4.2 VAR-Prewhitened Kernel Estimators

For each of the three methods of VAR lag order determination considered in Figure 1, we consider the effects of using the QS kernel to estimate the spectral density matrix of the VAR-prewhitened residuals. As in the simulation experiments of Andrews and Monahan (1992), the bandwidth parameter is determined by estimating a univariate AR(1) model for each VAR residual.

For each *dgp* and each VAR-prewhitened kernel HAC covariance matrix estimator, Figure 3 depicts the change in true confidence level (in percentage points) relative to the confidence level obtained using the corresponding VARHAC estimator (that is, the confidence level shown in Figure 1). In particular, the grey column indicates the effects of applying the QS kernel to the VAR residuals when the VAR lag order is fixed at unity. The black column indicates the effects of applying the kernel when BIC is used to determine the VAR lag structure according to the "asymmetric" approach described above. Finally, the white column indicates the corresponding results obtained using AIC.

Since the true confidence level is always below the nominal 90 percent confidence level, a positive value in Figure 3 indicates that more accurate inferences are obtained by applying the kernel QS rather than simply using the VARHAC estimator. Thus, the use of kernel estimation leads to improved inference accuracy for some of the AR(1) *dgp*s in Panel A and for all of the MA(1) *dgp*s in Panel B; these improvements are particularly noticeable when the VAR lag structure is determined using BIC. On the other hand, when either AIC or BIC is used to determine the VAR lag structure, kernel estimation leads to lower inference accuracy for all of the MA(q) examples in Panel B, and for many of the AR(1) examples in Panel A. These mixed results are symptomatic of the same problems identified in Sections 3.1 through 3.3, namely, kernel estimation may have counterproductive effects on inference accuracy when the data-dependent bandwidth parameter is far from optimal and/or the degree of temporal dependence varies substantially across the prewhitened VAR residuals.

Nevertheless, perhaps the most striking result in Figure 3 is that kernel estimation *never* leads to very large changes in the true confidence level compared with simply using the VARHAC covariance matrix estimator. This result was not apparent in the simulation study of Andrews and Monahan (1992), in which the PARA estimator (which is not robust to heteroskedasticity) was the only parametric procedure considered. In contrast, regardless of the method of determining the VAR lag structure, we find that the VAR-prewhitened residuals have relatively low first-order autocorrelation. Thus, formula (2.13) yields a relatively small bandwidth parameter when the initial estimates of the spectral density matrix and its generalized second-derivative are obtained using univariate AR(1) models as in the simulation experiments of Andrews (1991) and Andrews and Monahan (1992). And as previously noted, when the bandwidth parameter is close to zero, the kernel estimate of the spectral density matrix is virtually identical to the sample innovation covariance matrix used in constructing the VARHAC estimator.

4. Conclusions

The simulation experiments reported in this paper highlight the potential pitfalls of using an arbitrarily fixed order of VAR prewhitening in constructing HAC covariance matrices. One might have hoped that such pitfalls would be addressed by using a kernel to estimate the spectral density of the VAR-prewhitened residuals. Unfortunately, we find that this hope may not be fulfilled for any of several reasons. First, the kernel performs very poorly if the data contain an AR component of order higher than that of the prewhitening filter (cf. Section 3.1). (In fact, this problem was noted by Andrews (1991) for AR(1) processes, and motivated the first-order VAR prewhitening used

in the simulation experiments of Andrews and Monahan (1992).) Second, applying the kernel estimator to the VAR-prewhitened residuals tends to have negligible or even counterproductive effects unless the bandwidth parameter is determined using the correctly-specified parametric model (cf. Section 3.2). Furthermore, prewhitening frequently results in low-order autocovariances with opposite sign from the higher-order autocorrelations, so that applying a kernel to the prewhitened residuals may actually have counterproductive effects on performance when the bandwidth parameter is too low. Finally, in any application involving more than a single variable, a single bandwidth parameter must be used to construct the entire spectral density matrix of the prewhitened residuals; hence the kernel estimator may perform very poorly when the autocovariance structure differs substantially across elements (cf. Section 3.3).

Given these pitfalls, it seems clear that a model selection criterion should be used to determine the VAR prewhitening order of the HAC covariance estimator (unless, of course, one has *a priori* information about the true VAR order of the data). Furthermore, once data-dependent VAR prewhitening has been utilized, we find that applying a kernel-based method to the prewhitened residuals has negligible or even counterproductive effects unless the bandwidth parameter is chosen using the correct parametric model. In practice, of course, the true dgp is unknown. Thus, one might consider employing a model selection criterion to choose the appropriate parametric model (e.g., from the class of vector MA or ARMA models), which would then be used in determining the bandwidth parameter for the kernel estimator. However, an equally reasonable alternative would be to simply estimate the spectral density matrix directly from the estimated parameters of the chosen model (as in the VARHAC procedure considered here).

Beyond these general principles, several additional lessons are apparent from the simulation experiments considered in this paper. First, the VARHAC estimator can permit the lag order to vary across equations in the VAR and across variables within each equation; we find that taking advantage of this flexibility yields substantial improvements in inference accuracy in multivariate applications (cf. Sections 3.3.1 and 3.4.1). Second, we confirm the well-known result that AR spectral estimation performs very poorly when the largest AR root is close to unity (cf. Section 3.4). In future work, it will be useful to investigate the extent to which this problem can be resolved using median-unbiased estimation methods similar to those considered by Andrews and Phillips (1987) and Andrews and Chen (1994). Third, we find that AIC performs noticeably better than BIC for the higher-order AR processes in Section 3.1

as well as most of the MA processes in Sections 3.2 through 3.4. Given this finding, asymptotically consistent selection of the true (finite) AR order is evidently not the relevant consideration in constructing a HAC covariance matrix that will yield accurate inferences. In fact, both AIC and BIC tend to be much too conservative for higher-order AR processes and for nearly non-invertible MA processes. Thus, investigation of alternative model selection criteria should be a priority in future research on VARHAC covariance matrix estimation.

Finally, it should be noted that this paper has focused on the problem of constructing an asymptotically consistent HAC covariance matrix estimate, which is then used to obtain inferences concerning a regression estimator with a limiting normal distribution. An alternative approach, proposed by Keifer, Vogelsang and Bunzel (2000), standardizes the regression estimator using a kernel estimator of the spectral density, with bandwidth parameter equal to the sample size. The resulting test statistic has a limiting distribution which is invariant to the time-series properties of the residuals; this limiting distribution is non-standard, but Monte Carlo simulation is used to obtain the relevant critical values. In future work, it will be useful to investigate the extent to which the performance of this estimator can be enhanced by using data-dependent VAR prewhitening in conjunction with the kernel estimator.

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, H.Z., Z.G. Chen, and E.J. Hannan, 1982, *Annals of Statistics* 10, pp.926-936.

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

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

Andrews, D.W.K. and P.C.B. Phillips, 1987, Best median-unbiased estimation in linear regression with bounded asymmetric loss functions, *Journal of the American Statistical Association* 82, pp.886-93.

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.

Blackman, R.B. and J.W. Tukey, 1958, *The Measurement of Power Spectra*, Dover Press. Originally appeared in *Bell System Technical Journal* 37.

Den Haan, W.J. and A. Levin, 1996, Inferences from parametric and non-parametric covariance matrix estimation procedures, *National Bureau of Economic Research Technical Working Paper* #195.

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

Den Haan, W.J. and A. Levin, 1998, Vector autoregressive covariance matrix estimation, University of California, San Diego, manuscript.

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.

Grenander, U. and M. Rosenblatt, 1957, *Statistical Analysis of Stationary Time Series*, Chelsea Press.

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

Hannan, E.J., and L. Kavalieris, 1983, *Australian Journal of Statistics* 25, pp.287-297.

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

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

Kiefer, N., T. Vogelsang, and H. Bunzel, 2000, Simple robust testing of regression hypotheses, *Econometrica*, forthcoming.

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

Newey, W.K., and K.D. West, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, pp.631-653.

Ng, S., and P. Perron, 1995, Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association*, 90, pages 268-81.

Ogaki, M., 1993, Generalized method of moments: econometric applications, in *Handbook of Statistics* 11, G.S. Maddala, C.R. Rao, and H.D. Vinod eds., Elsevier Press, pp. 455-482.

Parzen, E., 1969, Multiple time series modelling, in *Multivariate Analysis II*, P.R. Krishnaiah, ed., Academic Press, pp.389-409.

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

Phillips, P.C.B. and C.C. Lee, 1994, An ARMA prewhitened long-run variance estimator, manuscript, Yale University.

Press, H. and J.W. Tukey, 1956, Power spectral methods of analysis and their application to problems in airplane dynamics, *Flight Test Manual IV-C*, NATO Advisory Group for Aeronautical Research and Development, pp.1-41. Reprinted as *Bell System Monograph* #2606.

Priestley, M.B., 1962, Basic considerations in the estimation of spectra, *Technometrics* 4, pp.551-564.

Priestley, M.B., 1981, *Spectral analysis and time series*, Academic Press.

Robinson, P.M., 1996, Inference-without-smoothing in the presence of nonparametric autocorrelation, *Econometrica* 66, pp.1163-1182.

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

Shibata, R., 1980, Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Annals of Statistics* 8, pp.147-164.

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

Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics* 6, pp.461-464.

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

West, K.D., 1997, Another heteroskedasticity and autocorrelation consistent covariance matrix estimator, manuscript, *Journal of Econometrics* 76(1-2), pp.171-191.

Table 1
Scalar Autoregressive Processes

Panel A: Autoregressive Estimators

True confidence level of the nominal 90% confidence interval

	True AR Order			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$
Fixed AR Order				
$h = 1$	88.1	78.1	72.8	69.8
$h = 2$	87.4	86.9	80.6	76.1
$h = 3$	87.0	86.6	86.2	81.6
$h = 4$	86.4	85.9	85.5	85.1
Data-Dependent AR Order				
AIC	87.1	85.7	83.9	80.7
(mean of h_{AIC})	*(1.4)*	*(2.2)*	*(2.5)*	*(2.4)*
BIC	87.8	83.8	77.3	72.2
(mean of h_{BIC})	*(1.0)*	*(1.6)*	*(1.3)*	*(1.0)*

Panel B: Applying the QS Kernel to the "Prewhitened" Residuals

True confidence level of the nominal 90% confidence interval

	True AR Order			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$
Fixed AR Order				
$h = 1$	88.3	76.0	71.7	69.6
$h = 2$	87.4	86.9	80.3	76.0
$h = 3$	87.0	86.3	86.3	81.4
$h = 4$	86.7	86.1	85.7	85.3
Data-Dependent AR Order				
AIC	87.2	85.6	83.7	80.4
BIC	88.0	83.3	76.5	71.1

Note: The data are generated by $Y_t = V_t$ where $V_t = (0.5/p)(V_{t-1} + \cdots + V_{t-p}) + \varepsilon_t$, where p is the order of the autoregressive process and ε_t is an i.i.d standard normal random variable. The sample length $T = 128$, and the results are computed using 10,000 replications. For each *dgp* and each HAC variance estimator, this table reports the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the sample mean of Y_t. Panel A provides results for AR estimators of the HAC variance; a maximum lag order of 4 is used when the AR order is chosen by AIC or BIC. Panel B provides results for HAC estimators in which the QS kernel is applied to a particular set of AR residuals, and the bandwidth parameter is determined using an AR(1) parametric model for these residuals.

Table 2

Scalar Moving-Average Processes

Panel A: Autoregressive HAC Estimators (No Kernel applied to AR Residuals)

True confidence level of the nominal 90% confidence interval

Lag Order Determination	MA(1) Parameter θ -0.1	-0.3	-0.5	-0.7	-0.9
Fixed $h = 1$	**89.7**	**93.1**	**97.9**	**100.0**	**100.0**
AIC	**89.6**	**89.8**	**90.9**	**95.6**	**99.9**
(mean of $\hat{h}$)	***(0.8)***	***(1.7)***	***(2.5)***	***(3.4)***	***(3.8)***
BIC	**91.6**	**92.3**	**94.1**	**97.2**	**99.9**
(mean of $\hat{h}$)	***(0.2)***	***(1.0)***	***(1.7)***	***(2.6)***	***(3.3)***

Panel B: Kernel HAC Estimators (No AR Prewhitening)

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ -0.1	-0.3	-0.5	-0.7	-0.9
AR(1) model	**90.7**	**93.2**	**97.0**	**99.8**	**100.0**
(mean of $\hat{\xi}_T$)	***(1.7)***	***(2.2)***	***(2.4)***	***(2.5)***	***(2.5)***
MA(1) model	**89.9**	**90.0**	**89.6**	**88.1**	**88.1**
(mean of $\hat{\xi}_T$)	***(2.1)***	***(3.9)***	***(6.5)***	***(12.3)***	***(65.4)***
MA(2) model	**88.8**	**89.1**	**88.5**	**86.6**	**85.5**
(mean of $\hat{\xi}_T$)	***(3.1)***	***(4.3)***	***(6.8)***	***(14.1)***	***(42.9)***

Note: This table reports inferences concerning the mean of the process $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$, where ε_t is i.i.d. $N(0, 1)$, and we normalize $\mu = 0$. The sample length $T = 128$, and the results are computed using 10,000 replications. For each value of θ and each HAC variance estimator, this table reports the true confidence level (at a nominal 90% significance level) of a two-tailed t-test of the significance of the sample mean. Panel A provides results for AR estimators of the HAC variance; a maximum lag order of 4 is used when the AR order is chosen by AIC or BIC. Panel B provides results for kernel estimators, with the bandwidth parameter determined by estimating the specified parametric model. Panels C, D, and E provide results for HAC estimators in which the QS kernel is applied to a particular set of AR residuals, using the same methods of determining the bandwidth parameter; for ease of reference, the relevant results from Panel A are repeated in the top row of each panel.

Table 2 (contd.)
Scalar Moving-Average Processes

Panel C: Prewhitened Kernel Estimators
Prewhitening Order Fixed at Unity

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ -0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	89.7	93.1	97.9	100.0	100.0
AR(1) model	89.7	92.9	97.3	99.9	100.0
(mean of $\hat{\xi}_T$)	*(0.7)*	*(1.0)*	*(1.5)*	*(1.8)*	*(2.0)*
MA(1) model	89.6	92.6	95.9	98.1	99.9
(mean of $\hat{\xi}_T$)	*(0.8)*	*(1.2)*	*(2.1)*	*(3.3)*	*(4.7)*
MA(2) model	88.3	88.7	88.4	86.1	82.1
(mean of $\hat{\xi}_T$)	*(2.8)*	*(3.8)*	*(6.5)*	*(12.3)*	*(41.5)*

Panel D: Prewhitened Kernel Estimators
Prewhitening Order Determined by AIC

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ -0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	89.6	89.8	90.9	95.6	99.9
AR(1) model	89.1	89.8	90.9	95.6	99.9
(mean of $\hat{\xi}_T$)	*(1.1)*	*(0.8)*	*(0.8)*	*(0.9)*	*(1.1)*
MA(1) model	89.0	89.7	90.8	95.5	99.9
(mean of $\hat{\xi}_T$)	*(1.2)*	*(0.9)*	*(0.9)*	*(1.1)*	*(1.7)*
MA(2) model	88.3	88.9	90.1	94.6	97.8
(mean of $\hat{\xi}_T$)	*(2.4)*	*(2.3)*	*(2.2)*	*(2.6)*	*(5.7)*

Table 2 (contd.)
Scalar Moving-Average Processes

Panel E: Prewhitened Kernel Estimators
Prewhitening Order Determined by BIC

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	91.6	92.3	94.1	97.2	99.9
AR(1) model	90.2	92.0	93.9	97.2	99.9
(mean of $\hat{\xi}_T$)	*(1.4)*	*(1.1)*	*(1.1)*	*(1.1)*	*(1.2)*
MA(1) model	89.8	91.7	93.7	97.0	99.9
(mean of $\hat{\xi}_T$)	*(1.7)*	*(1.3)*	*(1.4)*	*(1.6)*	*(2.2)*
MA(2) model	88.5	88.9	90.2	93.8	96.0
(mean of $\hat{\xi}_T$)	*(3.0)*	*(3.4)*	*(3.8)*	*(4.1)*	*(8.9)*

Table 3
Bivariate Processes with Heterogeneous Components

Panel A: VARHAC Estimators
(No Kernel Applied to VAR Residuals)

1. Inferences about the significance of $\hat{\alpha}$

True confidence level of the nominal 90% confidence interval

	MA(1) Parameter θ				
Lag Order Determination	**-0.1**	**-0.3**	**-0.5**	**-0.7**	**-0.9**
Fixed h = 1	**88.5**	**91.5**	**96.6**	**99.2**	**99.8**
AIC	**86.2**	**86.9**	**88.9**	**92.6**	**95.7**
(mean of $\hat{h}_{11}$)	***(1.5)***	***(2.0)***	***(2.5)***	***(3.0)***	***(3.2)***
(mean of $\hat{h}_{12}$)	***(0.6)***	***(0.6)***	***(0.6)***	***(0.6)***	***(0.6)***
BIC	**89.1**	**90.0**	**92.1**	**94.8**	**97.1**
(mean of $\hat{h}_{21}$)	***(0.5)***	***(1.1)***	***(1.6)***	***(2.2)***	***(2.5)***
(mean of $\hat{h}_{22}$)	***(0.1)***	***(0.1)***	***(0.1)***	***(0.1)***	***(0.1)***

2. Inferences about the significance of $\hat{\beta}$

True confidence level of the nominal 90% confidence interval

	MA(1) Parameter θ				
Lag Order Determination	**-0.1**	**-0.3**	**-0.5**	**-0.7**	**-0.9**
Fixed h = 1	**88.7**	**88.6**	**88.6**	**88.9**	**88.7**
AIC	**87.7**	**87.8**	**87.4**	**87.5**	**87.5**
(mean of $\hat{h}_{11}$)	***(0.5)***	***(0.5)***	***(0.6)***	***(0.6)***	***(0.6)***
(mean of $\hat{h}_{12}$)	***(0.6)***	***(0.6)***	***(0.6)***	***(0.7)***	***(0.7)***
BIC	**88.9**	**88.9**	**88.3**	**88.5**	**88.3**
(mean of $\hat{h}_{21}$)	***(0.0)***	***(0.1)***	***(0.1)***	***(0.1)***	***(0.1)***
(mean of $\hat{h}_{22}$)	***(0.1)***	***(0.1)***	***(0.1)***	***(0.1)***	***(0.1)***

Note: This table reports inferences concerning the OLS regression estimates of α and β for the model $Y_t = \alpha X_t + \beta Z_t + u_t$, . The variable $X_t = 0.95\, X_{t-1} + \varepsilon_t$, where ε_t is i.i.d. $\mathbf{N}(0, 1)$. The variable Z_t is i.i.d. $\mathbf{N}(0, 1/(1\text{-}0.95^2))$. The disturbance $u_t = v_t + \theta v_t$, where v_t is i.i.d. $\mathbf{N}(0, 1)$. The true coefficients are normalized as $\alpha = \beta = 0$. The sample length $T = 128$, and the results are computed using 10,000 replications. For each value of θ and each HAC covariance matrix estimator, this table reports the true confidence level (at a nominal 90% significance level) of two-tailed t-tests of the significance of $\hat{\alpha}$ and $\hat{\beta}$. A maximum lag order of 4 is imposed when AIC or BIC is used to determine the lag order of each variable in each VAR equation. Panel A provides results for VAR estimators of the HAC covariance matrix; that is, no kernel is applied to the VAR residuals. Panels B, C, and D provide results for HAC estimators in which the QS kernel is applied to a particular set of AR residuals, with the bandwidth parameter determined by estimating the specified parametric model; for ease of reference, the relevant results from Panel A are repeated in the top row of each panel.

Table 3 (contd.)
Bivariate Processes with Heterogeneous Components

Panel B: VAR-Prewhitened Kernel Estimators
VAR Prewhitening Order Fixed at Unity

1. Inferences about the significance of $\hat{\alpha}$

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	88.5	91.6	96.6	99.2	99.8
AR(1) model	91.1	96.0	99.0	99.8	100.0
MA(1) model	89.1	92.7	97.2	99.2	99.7
MA(2) model	88.9	92.3	94.6	96.5	97.9

2. Inferences about the significance of $\hat{\beta}$

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	88.7	88.6	88.6	88.9	88.7
AR(1) model	88.7	88.6	88.5	88.3	88.0
MA(1) model	88.9	88.9	88.9	88.9	88.9
MA(2) model	87.6	86.9	86.6	86.3	86.1

3. Mean Bandwidth Parameter

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
AR(1) model	0.8	1.0	1.2	1.3	1.4
MA(1) model	0.8	1.0	1.4	1.8	2.0
MA(2) model	2.7	3.0	3.8	4.6	5.1

Table 3 (contd.)
Bivariate Processes with Heterogeneous Components

Panel C: VAR-Prewhitened Kernel Estimators
Prewhitening Order Determined Using AIC

1. Inferences about the significance of $\hat{\alpha}$

True confidence level ofthe nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	86.2	86.9	88.9	92.6	95.7
AR(1) model	89.2	93.3	95.5	97.6	98.8
MA(1) model	87.3	89.7	93.2	96.5	98.3
MA(2) model	87.6	91.7	94.2	96.8	98.3

2. Inferences about the significance of $\hat{\beta}$

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	87.7	87.8	87.4	87.5	87.5
AR(1) model	87.9	87.8	87.8	87.7	87.4
MA(1) model	88.1	88.0	88.0	88.1	88.0
MA(2) model	86.7	86.6	86.0	85.9	85.5

3. Mean Bandwidth Parameter

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
AR(1) model	1.2	1.2	1.2	1.2	1.2
MA(1) model	1.0	1.0	1.0	1.0	1.0
MA(2) model	2.1	2.1	2.1	2.1	2.2

Table 3 (contd.)
Bivariate Processes with Heterogeneous Components

Panel D: VAR-Prewhitened Kernel Estimators
Prewhitening Order Determined Using BIC

1. Inferences about the significance of $\hat{\alpha}$

True confidence level of the nominal 90% confidence interval

Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	89.1	90.0	92.1	94.8	97.1
AR(1) model	90.2	94.6	97.1	98.4	99.2
MA(1) model	89.0	91.7	95.0	97.5	98.8
MA(2) model	88.2	92.1	95.0	97.2	98.5

2. Inferences about the significance of $\hat{\beta}$

True confidence level of the nominal 90% confidence interval

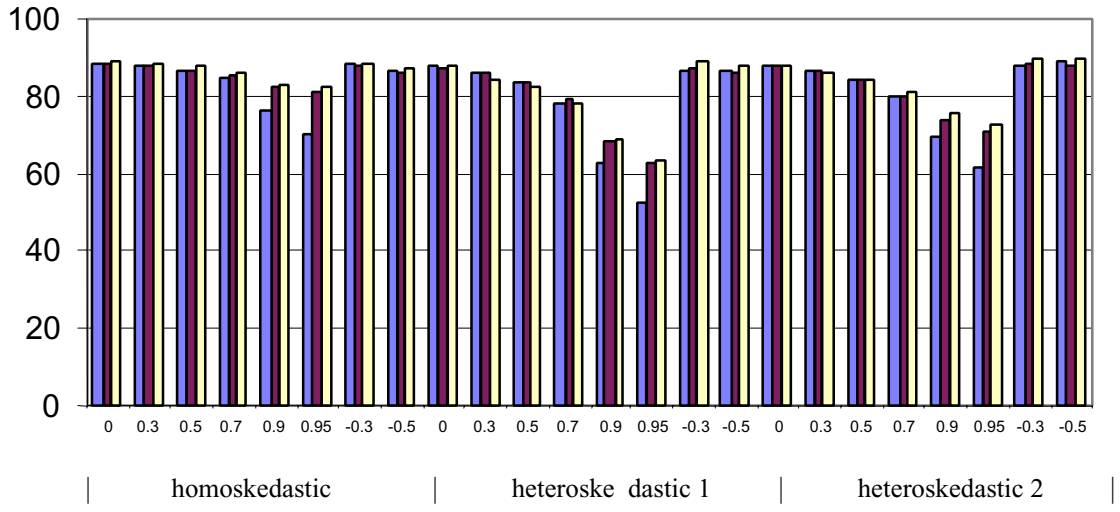
Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
None ($\xi \equiv 0$)	88.9	88.9	88.3	88.5	88.3
AR(1) model	88.7	88.6	88.5	88.4	88.1
MA(1) model	88.6	88.4	88.4	88.4	88.4
MA(2) model	87.8	87.0	86.6	86.4	85.8

3. Mean Bandwidth Parameter

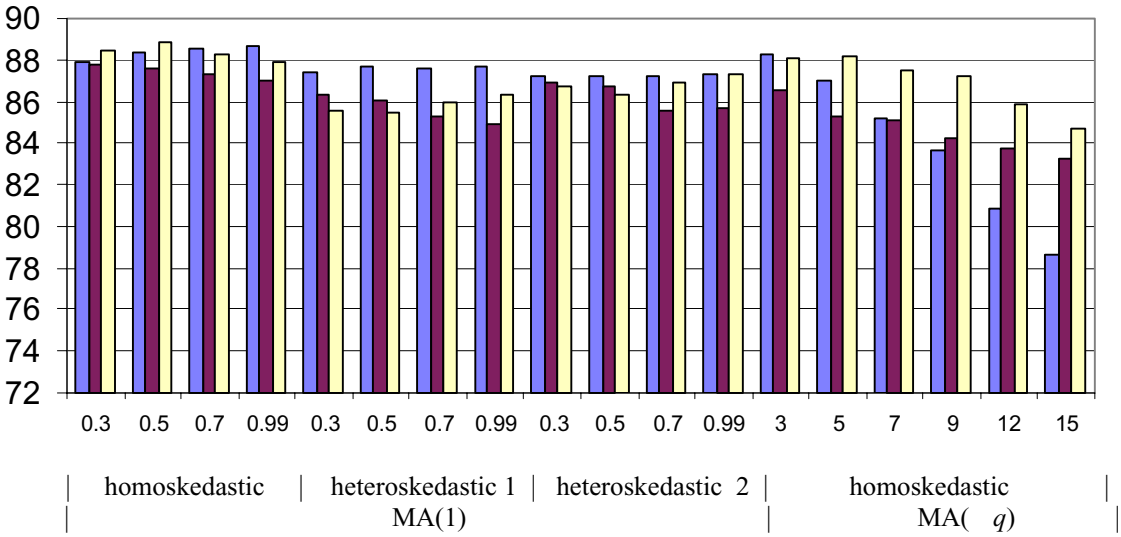
Bandwidth Parameter Determined By:	MA(1) Parameter θ				
	-0.1	-0.3	-0.5	-0.7	-0.9
AR(1) model	1.5	1.4	1.4	1.4	1.4
MA(1) model	1.3	1.3	1.3	1.3	1.3
MA(2) model	2.6	2.6	2.7	2.8	2.9

Figure 1
VAR Estimators for Multivariate Processes
(True confidence level of the nominal 90% confidence interval)

Panel A: AR(1) examples



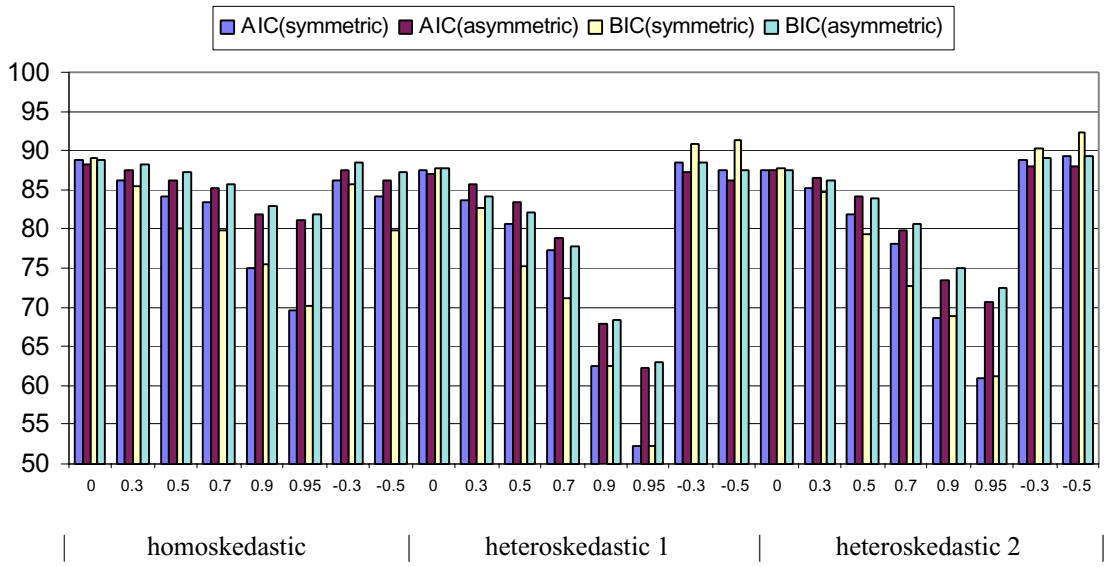
Panel B: MA Examples



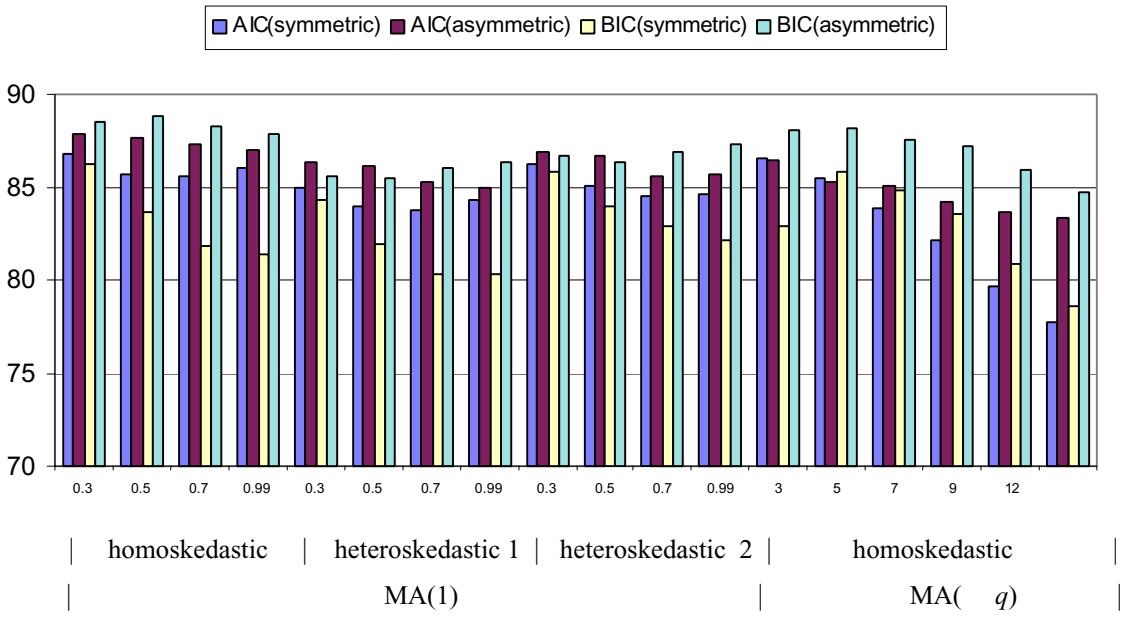
Note: For each *dgp* and for each method of determining the VAR lag structure, this figure indicates the frequency (in percentage points) that a two-tailed *t*-test at the nominal 90% confidence level does not reject the hypothesis that the coefficient corresponding to the first non-constant regressor is equal to its true value. The gray column indicates the results when the VAR lag order is fixed at unity. The black column indicates the results when BIC is used to select a separate lag order (up to a maximum of 4 lags) for the lagged dependent variable and for the other explanatory variables in each VAR equation, and the white column indicates the corresponding results obtained using AIC. Panel A indicates the results for experiments in which the regressors and the error term are generated by AR(1) processes; for each experiment, the value of the AR(1) coefficient is indicated below the *x*-axis. Panel B reports the results for experiments in which the regressors and the error term are generated by MA processes; for each experiment, either the MA(1) coefficient or the order q of the MA process is indicated below the *x*-axis. The sample length $T = 128$, and the results are computed using 10,000 replications.

Figure 2
Symmetric vs. Asymmetric Lag Order Selection for Multivariate Processes
(True confidence level of the nominal 90% confidence interval)

Panel A: AR(1) examples



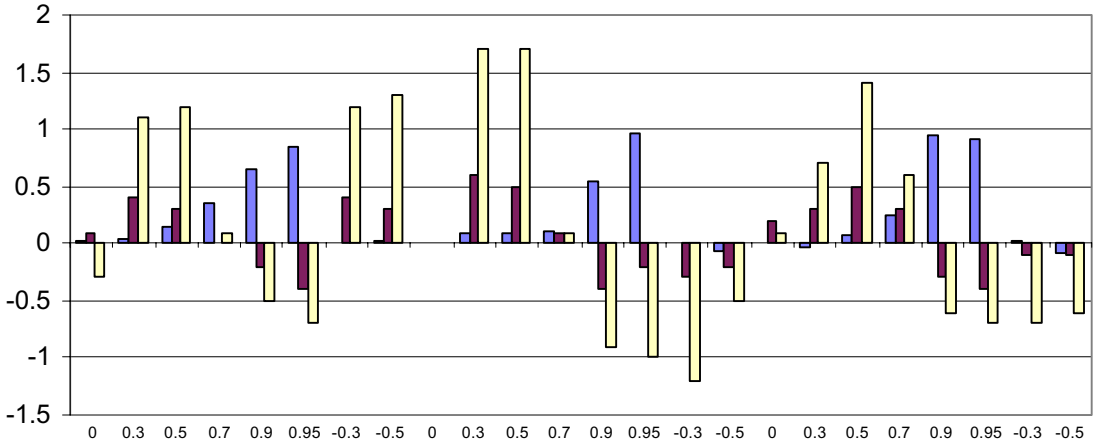
Panel B: MA Examples



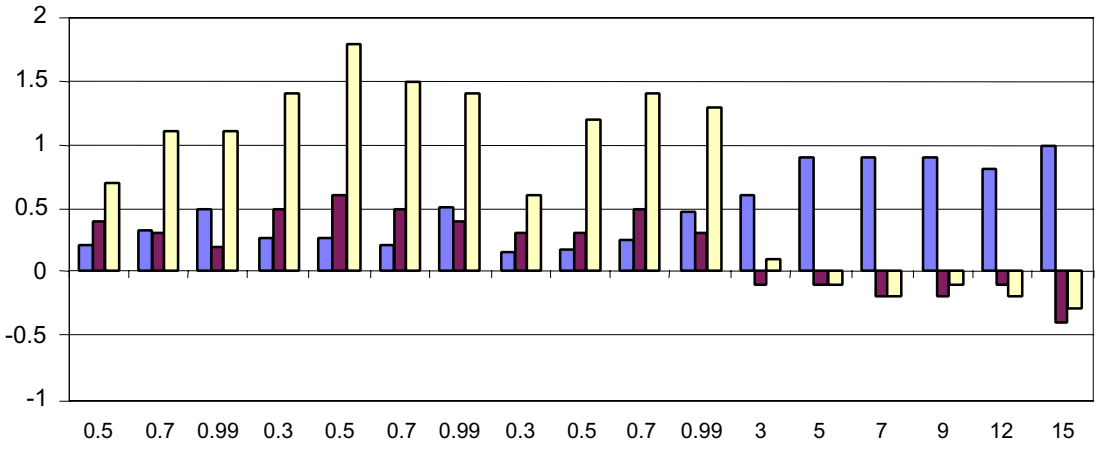
Note: For each *dgp* described in Figure 1 and each VARHAC covariance matrix estimator, this figure indicates the frequency that a two-tailed *t*-test at the nominal 90% confidence level does not reject the hypothesis that the coefficient corresponding to the first non-constant regressor is equal to its true value. The "symmetric" VARHAC estimators impose the constraint that the lagged dependent variable in each VAR equation has the same lag order as the other explanatory variables in that equation; the "asymmetric" VARHAC estimators do not impose this constraint. The VARHAC estimatorsare computed using a maximum lag order of 4. The sample length $T = 128$, and the results are computed using 10,000 replications.

Figure 3
Prewhitened Kernel Estimators for Multivariate Processes
(Difference from True Confidence Level shown in Figure 1)

Panel A: AR(1) Examples



Panel B: MA Examples



Note: For each *dgp* described in Figure 1 and for each method of determining the VAR lag structure, this figure indicates the change in the true confidence level (in percentage points) of the *t*-test described in Figure 1 when the QS kernel is used to estimate the spectral density matrix at frequency zero of the VAR-prewhitened residuals. The gray column indicates the results when the VAR lag order is fixed at unity. The black column indicates the results when BIC is used to select a separate lag order for the lagged dependent variable and for the other explanatory variables in each VAR equation, and the white column indicates the corresponding results obtained using AIC. The sample length $T = 128$, and the results are computed using 10,000 replications.