heteroskedasticity and autocorrelation corrections

Heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation refers to calculation of covariance matrices that account for conditional heteroskedasticity of regression disturbances and serial correlation of cross products of instruments and regression disturbances. The heteroskedasticity and serial correlation may be of unknown form. HAC estimation is integral to empirical research using generalized method of moments (GMM) estimation (Hansen, 1982). In this article I summarize results relating to HAC estimation, with emphasis on practical rather than theoretical aspects.

The central issue is consistent and efficient estimation of what is called a 'long-run variance', subject to the constraint that the estimator is positive semidefinite in finite samples. Positive semidefiniteness is desirable since the estimator will be used to compute standard errors and test statistics. To fix notation, let h_t be a $q \times 1$ stationary mean zero random vector. Let Γ_j denote the $q \times q$ autocovariance of h_t at lag q, $\Gamma_j \equiv Eh_t h_{t-j}'$; of course, $\Gamma_j = \Gamma_{-j}'$. The long run variance of h_t is the $q \times q$ matrix

$$S = \sum_{j=-\infty}^{n} \Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j'). \qquad (1)$$

Apart from a factor of 2π, the symmetric matrix S, which I assume to be positive definite, is the spectral density of h_t at frequency zero. As discussed below, techniques for spectral density estimation are central to HAC estimation. (For an arbitrary stationary process, the sum in the right-hand side of (1) may not converge, and may not be positive definite even if it does converge. But here and throughout I assume unstated regularity conditions. As well, I use formulas that allow for relatively simple notation, for example assuming covariance stationarity even when that assumption can be relaxed. The cited papers

may be referenced for generalizations and for technical conditions.)

To illustrate how estimation of S figures into covariance matrix estimation, consider the following simple example. As in Hansen and Hodrick (1980), let us suppose that we wish to test the 'rationality' of a scalar variable x_t as an n period ahead predictor of a variable y_{t+n+1}, for $n \geq 0$: the null hypothesis is $E_t y_{t+n+1} = x_t$, where E_t denotes expectations conditional on the information set used by market participants. The variable x_t might be the expectation of y_{t+n+1} reported by a survey, or it might be a market determined forward rate. Let u_t denote the expectational error: $u_t = y_{t+n+1} - E_t y_{t+n+1} = y_{t+n+1} - x_t$. (The expectational error u_t, which is not realized until period $t+n+1$, is dated t to simplify notation.)

One can test one implication of the hypothesis that x_t is the expectation of y_{t+n+1} by regressing y_{t+n+1} on a constant and x_t, and checking whether the coefficient on the constant term is zero and that on x_t is 1:

$$y_{t+n+1} = \beta_0 + \beta_1 x_t + u_t \equiv X_t'\beta + u_t;$$
$$H_0 : \beta = (0,1)'. \tag{2}$$

Under the null, $EX_t u_t = 0$, so least squares is a consistent estimator. As well, $X_t u_t$ follows a moving average process of order n. Thus the asymptotic variance of the least squares estimator of β is $(EX_t X_t')^{-1} S (EX_t X_t')^{-1}$, where $S = \Gamma_0 + \sum_{j=1}^{n}(\Gamma_j + \Gamma_j')$, $\Gamma_j \equiv EX_t u_t (X_{t-j} u_{t-j})'$. This example maps into the notation used in (1) with $h_t = X_t u_t$, $q = 2$ and a known upper bound to the number of non-zero autocovariances of h_t. Clearly one needs to estimate $EX_t X_t'$ and S to conduct inference. A sample average of $X_t X_t'$ can be used to estimate $EX_t X_t'$. If $n = 0$, so that h_t is serially uncorrelated, $S = EX_t u_t (X_t u_t)'$ and estimation of S is equally straightforward; White's (1980) heteroskedasticity consistent estimator can be used. The subject at hand considers ways to estimate S when h_t is serially correlated. I note in passing that one cannot sidestep estimation of S by applying generalized least squares. In this example and more generally, generalized least squares is inconsistent. See Hansen and West (2002).

To discuss estimation of S, let us describe a more general set-up. In GMM estimation, h_t is a $q \times 1$ orthogonality condition used to identify a k-dimensional parameter vector β. The orthogonality condition takes the form

$$h_t = Z_t u_t \tag{3}$$

for a $q \times \ell$ matrix of instruments Z_t and an $\ell \times 1$ vector of unobservable regression disturbances u_t. The vector of regression disturbances depends on observable data through β, $u_t = u_t(\beta)$. In the example just given, $q = 2$, $\ell = 1$, $Z_t = X_t$, $u_t(\beta) = y_{t+n+1} - X_t'\beta$. The example just given is overly simple in that the list of instruments typically will not be identical to right-hand side variables, and the model may be nonlinear. For a suitable $k \times q$ matrix D, the asymptotic variance of the GMM estimator of β takes the form DSD' (for example, $D = (EX_t X_t')^{-1}$ in the example just given). In an overidentified model (that is, in models in which the dimension of the orthogonality condition q is greater than the number of parameters k) the form D takes depends on a certain weighting matrix. Let $h_{t\beta}$ be the $q \times k$ matrix $\partial h_t / \partial \beta$. When the weighting matrix is chosen optimally, $D = (Eh_{t\beta}' S^{-1} Eh_{t\beta})^{-1} Eh_{t\beta}' S^{-1}$ and the asymptotic variance DSD' simplifies to $(Eh_{t\beta}' S^{-1} Eh_{t\beta})^{-1}$. The optimal weighting matrix is one that converges in probability to S, and thus the results about to be presented are relevant to efficient estimation as well as to hypothesis testing. In any event, the matrix $Eh_{t\beta}$ typically is straightforward to estimate; the question is how to estimate S. This will be the focus of the remainder of the discussion.

We have sample of size T and sample counterparts to u_t and h_t, call them $\hat{u}_t = u_t(\hat{\beta})$ and $\hat{h}_t = h_t(\hat{\beta})$. Here, $\hat{\beta}$ is a consistent estimate of β. In the least squares example given above, $\hat{u}_t$ is the least squares residual, $\hat{u}_t = y_{t+n+1} - X_t'\hat{\beta}$, and $\hat{h}_t = X_t \hat{u}_t = X_t (y_{t+n-1} - X_t'\hat{\beta})$. One path to consistent estimation of S involves consistent estimation of the autocovariances of h_t. The natural estimator is a sample average,

$$\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{h}_t \hat{h}_{t-j}' \quad \text{for} \quad j \geq 0. \tag{4}$$

For given j, (4) is a consistent ($T \to \infty$) estimator of Γ_j.

I now discuss in turn several possible estimators, or classes of estimators, of S: (1) the truncated estimator; (2) estimators applicable only when h_t follows a moving average (MA) process of known order; (3) an autoregressive spectral estimator; (4) estimators that smooth autocovariances; (5) some recent work, on estimators that might be described as extensions or modifications of ones the estimators described in (4).

1 The truncated estimator

Suppose first that it is known a priori that the autocovariances of h_t are zero after lag n, as is the case in the empirical example above. A natural estimator of S is one that replaces population objects in (1) with sample analogues. This is the *truncated* estimator:

$$\hat{S}_{TR} = \hat{\Gamma}_0 + \sum_{j=1}^{n} (\hat{\Gamma}_j + \hat{\Gamma}_j'). \tag{5}$$

In the more general case in which $\Gamma_j \neq 0$ for all j, the truncated estimator is consistent if the truncation point $n \to \infty$ at a suitable rate. Depending on exact technical conditions, the rate may be $n/T^{1/2} \to 0$ or $n/T^{1/4} \to 0$ (Newey and West, 1987). The truncated estimator need not, however, yield a positive semidefinite estimate. With

certain plausible data generating processes, simulations indicate that it will not be p.s.d. in a large fraction of samples (West, 1997). Hence this estimator is not used much in practice.

2 Estimators applicable only when h_t follows an MA process of known order n

Such a process for h_t holds in studies of rationality (as illustrated above) and in the first order conditions from many rational expectations models (for example, Hansen and Singleton, 1982).

Write the Wold representation of h_t as $h_t = e_t + \Theta_1 e_{t-1} + \cdots + \Theta_n e_{t-n}$. Here, e_t is the $q \times 1$ innovation in h_t. Let Ω denote the $q \times q$ variance covariance matrix of e_t. Then it is well known (for example, Hamilton, 1994, p. 276) that

$$S = (I + \Theta_1 + \cdots + \Theta_n)\Omega(I + \Theta_1 + \cdots + \Theta_n)' \tag{6}$$

Suppose that one fits an MA(n) process to $\hat{h}_t$, and plugs the resulting estimates of the Θ_i and Ω into the formula for S. Clearly the resulting estimator is $T^{1/2}$ consistent and positive semidefinite. Nevertheless, to my knowledge this estimator has not been used, presumably because of numerical difficulties in estimating multivariate moving average processes.

Two related estimators have been proposed that impose a smaller computational burden. Hodrick (1992) and West (1997) suggest an estimator that requires fitting an MA(n) to the vector of regression residuals $\hat{u}_t$ (or, in Hodrick's, 1992, application, using MA coefficients that are known a priori). The computational burden of such MA estimation will typically be considerably less than that of MA estimation of the h_t process, because the dimension of u_t is usually much smaller than that of h_t. For example, $\hat{u}_t$ will be a scalar in a single equation application, regardless of the number of orthogonality conditions captured in h_t. Write the estimated MA process for $\hat{u}_t$ as $\hat{u}_t = \hat{\epsilon}_t + \hat{\psi}_1\hat{\epsilon}_{t-1} + \cdots + \hat{\psi}_n\hat{\epsilon}_{t-n}$, where the $\hat{\psi}_j$ are $\ell \times \ell$. (Note that ϵ_t, the $\ell \times 1$ innovation in u_t, is not the same as e_t, the $q \times 1$ innovation in h_t.) Then a $T^{1/2}$ consistent and positive semidefinite estimator of S is

$$\begin{aligned}\hat{S}_{MA-\ell} &= T^{-1}\sum_{t=1}^{T-n}\hat{d}_{t+n}\hat{d}_{t+n}'\hat{d}_{t+n} \\ &= (Z_t + Z_{t+1}\hat{\psi}_1 + \cdots + Z_{t+n}\hat{\psi}_n)\hat{\epsilon}_t,\end{aligned} \tag{7}$$

where, again, Z_t is the $q \times \ell$ matrix of instruments (see eq. (3)).

Eichenbaum, Hansen and Singleton (1988) and Cumby, Huizanga and Obstfeld (1983) propose a different strategy that avoids the need to estimate a moving average process for either u_t or h_t. They suggest estimating the parameters of $\hat{h}_t$'s autoregressive representation, and inverting the autoregressive weights to obtain moving average weights. Call the results $\hat{\Theta}_1, \ldots, \hat{\Theta}_n$, with $\hat{\Omega}$ the estimate of the innovation variance–covariance matrix. The resulting estimator $\hat{S} = (I + \hat{\Theta}_1 + \cdots + \hat{\Theta}_n)\hat{\Omega}(I + \hat{\Theta}_1 + \cdots + \hat{\Theta}_n)'$ is positive semidefinite by construction. The rate at which it converges to S depends on the rate at which the order of the autoregression is increased.

3 Autoregressive estimators

Den Haan and Levin (1997) propose and evaluate an autoregressive spectral estimator. Suppose that h_t follows a (possibly) infinite-order vector autoregression (VAR)

$$h_t = \sum_{j=1}^{\infty}\Phi_j h_{t-j} + e_t, \quad Ee_te_t' = \Omega. \tag{8}$$

Then (Hamilton, 1994, p. 237)

$$S = \left(I - \sum_{j=1}^{\infty}\Phi_j\right)^{-1}\Omega\left(I - \sum_{j=1}^{\infty}\Phi_j\right)^{-1'}. \tag{9}$$

The idea is to approximate this quantity via estimates from a finite-order VAR in $\hat{h}_t$. Write the estimate of a VAR in $\hat{h}_t$ of order p as

$$\begin{aligned}\hat{h}_t &= \hat{\Phi}_1\hat{h}_{t-1} + \cdots + \hat{\Phi}_p\hat{h}_{t-p} + \hat{e}_t, \\ \hat{\Omega} &= T^{-1}\sum_{t=p+1}^{T}\hat{e}_t\hat{e}_t'.\end{aligned} \tag{10}$$

Then the estimator of $\hat{S}$ is

$$\hat{S}_{AR} = \left(I - \sum_{j=1}^{p}\hat{\Phi}_j\right)^{-1}\hat{\Omega}\left(I - \sum_{j=1}^{p}\hat{\Phi}_j\right)^{-1'}. \tag{11}$$

Den Haan and Levin (1997, Section 3.5) conclude that if p is chosen by BIC, and some other technical conditions hold, then this estimator converges at a rate very near $T^{1/2}$ (the exact rate depends on certain characteristics of the data). A possible problem in practice with this estimator (as well as with the estimator described in the final paragraph of Section 2, which also requires estimates of a VAR in $\hat{h}_t$) is that it may require estimation of many parameters and inversion of a large matrix. Den Haan and Levin therefore suggest judiciously parametrizing the autoregressive process, for example by using the BIC criterion equation-by-equation for each of the q elements of $\hat{h}_t$.

4 Estimators that smooth autocovariances

In practice, the most widely used class of estimators is one that relies on smoothing of autocovariances. Andrews (1991), building on the literature on estimation of spectral densities, established a general framework for

analysis. Andrews considers estimators that can be written

$$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{T-1} k_j(\hat{\Gamma}_j + \hat{\Gamma}_j') \qquad (12)$$

for a series of kernel weights $\{k_j\}$ that obey certain properties. For example, to obtain a consistent estimator, we need k_j near zero (or perhaps identically zero) for values of j near $T-1$, since autocovariances at large lags are estimated imprecisely, while $k_j \to 1$ for each j is desirable for consistency. We would also like the choice of k_j to ensure positive definiteness.

The two most commonly used formulas for the kernel weights are:

$$\text{Bartlett : for some } m \geq 0 : k_j = 1 - [j/(m+1)] \text{ for } j \leq m,\ k_j = 0 \text{ for } j > m. \qquad (13a)$$

$$\text{Quadratic spectral (QS) : for some } m > 0, \text{ and with } x_j = j/m : k_j = [25/12\pi^2 x_j^2] \times \{[\sin(6\pi x_j/5)/(6\pi x_j/5)] - \cos(6\pi x_j/5)\}. \qquad (13b)$$

If we let $z_j = 6\pi x_j/5$, the QS formula for k_j can be written in more compact form as $(3/z_j^2)\{[\sin(z_j)/z_j] - \cos(z_j)\}$. Call the resulting estimators $\hat{S}_{BT}$ and $\hat{S}_{QS}$. For example,

$$\hat{S}_{BT} = \hat{\Gamma}_0 + \sum_{j=1}^{m} [1 - j/(m+1)](\hat{\Gamma}_j + \hat{\Gamma}_j'). \qquad (14)$$

The vast literature on spectral density estimation suggests many other possible kernel weights. For conciseness, I consider only the Bartlett and QS kernels.

To operationalize these estimators, one needs to choose the lag truncation parameter or bandwidth m. I note that for both kernels, consistency requires $m \to \infty$ as $T \to \infty$, even if h_t follows an MA process of known finite order, as in the example given above. Thus one should not set m to be the number of non-zero autocovariances. Subject to possible problems with positive definiteness, setting $m = n$ is fine for the truncated estimator (5) but not for estimators that use nontrivial weights $\{k_j\}$.

Andrews shows that maximizing the rate at which $\hat{S}$ converges to S requires that m increase as a suitable function of sample size, with the 'suitable function' varying with kernel. For the Bartlett and QS, the maximal rates of convergence are realized when

$$\text{Bartlett : } m = \gamma T^{1/3} (\text{or } m = (\text{integer part of } \gamma T^{1/3})) \text{ for some } \gamma \neq 0,$$
$$\text{QS : } m = \gamma T^{1/5} \text{ for some } \gamma \neq 0, \qquad (15)$$

in which case $\hat{S}_{BT}$ converges to S at rate $T^{1/3}$ and the mean squared error in estimation of S goes to zero at rate $T^{2/3}$; the comparable figures for QS are $T^{2/5}$ and $T^{4/5}$. Since both estimators are nonparametric, they converge at rates slower than $T^{1/2}$; since faster convergence is better, the QS rate is preferable to that of the Bartlett. Indeed, Andrews (1991), drawing on Priestley (1981), shows that for a certain class of kernel weights $\{k_j\}$, the mean squared error of QS rate is optimal in the following sense: a $T^{4/5}$ rate on the asymptotic mean squared error is the fastest that can be achieved if one wants to ensure a positive definite $\hat{S}$, and within the class of kernels that achieve the $T^{4/5}$ rate, the QS has the smallest possible asymptotic mean squared error.

As a practical matter, the formulas in (15) have merely pushed the question of choice of m to one of choice of γ; putting arbitrary γ in (15) yields convergence that is as fast as possible, but different choices of γ lead to different asymptotic mean squared errors. The choice of γ that is optimal from the point of view of asymptotic mean squared error is a function of the data (Hannan, 1970, p. 286). Let $S^{(0)} = \sum_{j=-\infty}^{\infty} \Omega_j (= S)$; $S^{(1)} = \sum_{j=-\infty}^{\infty} |j| \Omega_j$; $S^{(2)} = \sum_{j=-\infty}^{\infty} j^2 \Omega_j$. For scalar ($q$=1) S optimal choices are:

$$\text{Bartlett : } \gamma = 1.1447[S^{(1)}/S^{(0)}]^{2/3};$$
$$\text{QS : } \gamma = 1.3221[S^{(2)}/S^{(0)}]^{2/5}. \qquad (16)$$

(See Andrews, 1991, for the derivation of these formulas.)

Andrews (1991), Andrews and Monahan (1992) and Newey and West (1994) proposed feasible data dependent to procedures to estimate γ, for vector as well as scalar h_t. Rather than exposit the general case, I will describe two 'cookbook' procedures that have been offered as reasonable starting points in empirical work. One procedure relies on Andrews (1991) and Andrews and Monahan (1992), and assumes the QS kernel and estimation of γ via parametric models. The second relies on Newey and West (1994), and assumes a Bartlett kernel and nonparametric estimation of γ. I emphasize that both papers present more general results than are presented here; both allow the researcher to (for example) use any one of a wide range of kernels.

Let there be a $q \times 1$ vector of weights $w = (w_1, w_2, \ldots, w_q)'$ whose elements tells us how to weight the various elements of S with respect to mean squared error. The weights might be sample dependent, and den Haan and Levin (1997) argue that there are benefits to certain sample-dependent weights, but a simple choice proposed by both papers is: w_i=0 if the corresponding element of h_t is a cross product of a constant term and a regression disturbance, otherwise $w_i = 1$. Andrews's loss function is the normalized expectation of $\sum_{i=1}^{q} w_i (S_{ii} - \hat{S}_{ii})^2$, while Newey and West's loss function is the normalized expectation of $[w'(\hat{S} - S)w]^2$; the normalization is $T^{4/5}$ for QS and $T^{2/3}$ for Bartlett.

Both procedures begin with using a vector autoregression to prewhiten, and end with re-colouring. The basic justification for prewhitening and re-colouring is that simulation evidence indicates that this improves finite sample performance.

1. Prewhitening: Estimate a vector autoregression in $\hat{h}_t$, most likely of order 1. Call the residuals $\hat{h}_t^\dagger$

$$\hat{h}_t = \hat{A}\hat{h}_{t-1} + \hat{h}_t^\dagger, \quad \hat{A} = \sum_{t=2}^{T} \hat{h}_t \hat{h}_{t-1}' \left(\sum_{t=2}^{T} \hat{h}_{t-1}\hat{h}_{t-1}' \right)^{-1}. \tag{17}$$

2. Let $\hat{\Gamma}_j^\dagger$ denote the jth autocovariance of the VAR residual $\hat{h}_t^\dagger$, $\hat{\Gamma}_j^\dagger = (T-1)^{-1}\sum_{t=2+j}^{T} \hat{h}_t^\dagger \hat{h}_{t-j}^{\dagger\prime}$. Using $\{\hat{\Gamma}_j^\dagger\}$ (rather than $\{\hat{\Gamma}_j\}$ [the autocovariances of $\hat{h}_t$]), and choosing m optimally as described in steps 2a or 2b below, construct an estimate of the long run variance of the residual of the VAR just estimated. Call the result $\hat{S}^\dagger$.
2a. Andrews and Monahan (1992): Fit a univariate AR(1) to each of the q elements of $\hat{h}_t^\dagger$. Call the resulting estimate of the AR coefficient and variance of the residual $\hat{\rho}_i$ and $\hat{\sigma}_i^2$. Compute

$$\begin{aligned} \hat{s}_2 &= \sum_{i=1}^{q} w_i (4\hat{\rho}_i^2 \hat{\sigma}_i^4)/(1-\hat{\rho}_i)^8, \quad \hat{s}_0 \\ &= \sum_{i=1}^{q} w_i \hat{\sigma}_i^4/(1-\hat{\rho}_i)^4, \hat{\gamma}_{QS} \\ &= 1.3221[\hat{s}_2/\hat{s}_0]^{1/5}, \quad \hat{m}_{QS} = \hat{\gamma}_{QS} T^{1/5}. \end{aligned} \tag{18}$$

Then plug $\hat{m}_{QS}$ into formula (13b). Call the result $\hat{k}_j$. Compute $\hat{S}^\dagger = \hat{\Gamma}_0^\dagger + \sum_{j=1}^{T-1} \hat{k}_j (\hat{\Gamma}_j^\dagger + \hat{\Gamma}_j^{\dagger\prime})$.
2b. Newey and West (1994): Set n = integer part of $12(T/100)^{2/9}$. Compute

$$\begin{aligned} \hat{s}^{(1)} &= w' \hat{\Gamma}_0 w + 2\sum_{i=1}^{n} i w' \Gamma_i^\dagger w, \\ \hat{s}^{(0)} &= 2\sum_{i=1}^{n} w' \hat{\Gamma}_i^\dagger w, \quad \hat{\gamma}_{BT} = 1.1447[\hat{s}^{(1)}/\hat{s}^{(0)}]^{2/3}, \\ \hat{m}_{BT} &= \text{integer part of } \hat{\gamma}_{BT} T^{1/3}. \end{aligned} \tag{19}$$

Then compute $\hat{S}^\dagger$ according to (14), using $\hat{m}_{BT}$.
3. Re-colouring: compute $\hat{S} = (I-\hat{A})^{-1}\hat{S}^\dagger (I-\hat{A})^{-1'}$.
These two recipes for estimates of S can serve as a starting point for experimentation for alternative choices of m and alternative kernels.

What is the simulation evidence on behaviour of these and other proposed estimators? In answering this question, I focus on sizing of test statistics and accuracy of confidence interval coverage: accuracy in estimation of S is desirable mainly insofar as it leads to accuracy of inference using the relevant variance–covariance matrix. The simulations in papers cited in this article suggest the following. First, no one estimator dominates others. This means in particular that the rate of convergence is not a sufficient statistic for performance in finite samples. The truncated estimator often and the autoregressive estimator sometimes perform more poorly than the slower converging QS estimator, which in turn sometimes performs more poorly than the still slower converging Bartlett estimator. Second, given that one decides to use QS or Bartlett, performance generally though not always is improved if one prewhitens and uses a data-dependent bandwidth as described in the recipes above. Third, the QS and Bartlett estimators tend to reject too much in the presence of positive serial correlation in h_t, and have what I read as a DGP dependent rejection rate (sometimes over-reject, sometimes under-reject) in the presence of negative serial correlation in h_t. The truncated estimator is much likelier to fail to be positive semidefinite in the presence of negative than positive serial correlation. Finally, the performance of all estimators leaves much to be desired. Plausible data-generating processes and sample sizes can lead to serious mis-sizing of any given estimator. Nominal 0.05 tests can have empirical size as low as 0.01 and higher than 0.25.

5 Some recent work

Because simulation studies have yielded disappointing performance, ongoing research aims to develop better estimators. I close by summarizing a few of many recently published papers.

1. I motivated my topic by observing that consistent estimation of S is a natural element of consistent estimation of the variance–covariance matrix of a GMM estimator. Typically we estimate the variance–covariance matrix because we wish to construct confidence intervals or conduct hypothesis tests. A recent literature has evaluated inconsistent estimators that lead to well-defined test statistics, albeit statistics with non-standard critical values. These estimators set lag truncation (or bandwidth) equal to sample size. For example, for the Bartlett estimator, these estimators set $m = T-1$ (see Kiefer, Vogelsang and Bunzel, 2000; Kiefer and Vogelsang, 2002). Simulation evidence indicates that the non-standard statistics may be better behaved than standard statistics. Jansson (2004) provides a theoretical rationale for improved performance in a special case, with more general results in Kiefer and Vogelsang (2005). Phillips, Sun and Jin (2006; 2007) propose a related approach, which under some assumptions will yield statistics with standard critical values.

2. Politis and Romano (1995) propose what they call a 'trapezoidal' kernel. A trapezoidal kernel is a combination of the truncated and Bartlett kernels. For given truncation lag m, let $x_j = j/(m+1)$. Then for some c,

$0<c<1$, the trapezoidal weights satisfy: $k_j=1$ if $0 \leq x_j \leq c$, $k_j=(x_j-1)/(c-1)$ for $c<x_j \leq 1$. Thus for $0 \leq j \leq c(m+1)$, the autocovariances receive equal weight, as in the truncated kernel; for $c(m+1)<j \leq m+1$, the weights on the autocovariances decline linearly to zero, as in the Bartlett kernel. Such kernels have the advantage that, like the truncated kernel, their convergence is rapid (near $T^{1/2}$). They share with the truncated kernel the possibility of not being positive semidefinite. The authors argue, however, that these kernels are better behaved in finite samples than is the truncated kernel.

3. Xiao and Linton (2002) propose 'twicing' kernels. Operationally, one first computes an estimate such as one of those described in Section 4. One also constructs a multiplicative bias correction by smoothing periodogram ordinates via a 'twiced' kernel. For a properly chosen bandwidth and kernel, the mean squared error of the estimator is of order $T^{8/9}$ (versus $T^{4/5}$ for the QS and $T^{2/3}$ for the Bartlett, absent any corrections). As well, Hirukawa's (2006) version of the Xiao and Linton estimator is positive semidefinite by construction. (The rate results for this estimator and that described in the previous paragraph do not contradict Andrews's, 1991, optimality result for the QS kernel, because these procedures fall outside the class considered by Andrews.)

KENNETH D. WEST

See also **rational expectations models, estimation of; Euler equations; generalized method of moments estimation; spectral analysis; time series analysis.**

I thank Steven Durlauf, Masayuki Hirukawa and Tim Vogelsgang for helpful comments, and the National Science Foundation for financial support.

Bibliography

Andrews, D.W.K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–58.

Andrews, D.W.K. and Monahan, J.C. 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 60, 953–66.

Cumby, R.E., Huizanga, J. and Obstfeld, M. 1983. Two step, two-stage least squares estimation in models with rational expectations. *Journal of Econometrics* 21, 333–55.

den Haan, W.J. and Levin, A.T. 1997. A practitioner's guide to robust covariance matrix estimation. In *Handbook of Statistics: Robust Inference*, vol. 15, eds. G. Maddala and C. Rao. New York: Elsevier.

Eichenbaum, M.S., Hansen, L.P. and Singleton, K.J. 1988. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* 103, 51–78.

Hamilton, J. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

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

Hansen, B.E. and West, K.D. 2002. Generalized method of moments and macroeconomics. *Journal of Business and Economic Statistics* 20, 460–9.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–54.

Hansen, L.P. and Hodrick, R.J. 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy* 96, 829–53.

Hansen, L.P. and Singleton, K.J. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–86.

Hirukawa, M. 2006. A modified nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis* 27, 441–76.

Hodrick, R.J. 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357–86.

Jansson, M. 2004. The error in rejection probability of simple autocorrelation robust tests. *Econometrica* 72, 937–46.

Kiefer, N.M. and Vogelsang, T.J. 2002. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica* 70, 2093–5.

Kiefer, N.M. and Vogelsang, T.J. 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–64.

Kiefer, N.M., Vogelsang, T.J. and Bunzel, H. 2000. Simple robust testing of regression hypotheses. *Econometrica* 68, 695–714.

Newey, W.K. and West, K.D. 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–08.

Newey, W.K. and West, K.D. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–54.

Phillips, P.C.B., Sun, Y. and Jin, S. 2006. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review* 47, 837–94.

Phillips, P.C.B., Sun, Y. and Jin, S. 2007. Long run variance estimation and robust regression testing using sharp origin kernels with no truncation. *Journal of Statistical Planning and Inference* 137, 985–1023.

Politis, D.N. and Romano, J.P. 1995. Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* 16, 67–103.

Priestley, M.B. 1981. *Spectral Analysis and Time Series*. New York: Academic Press.

West, K.D. 1997. Another heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Journal of Econometrics* 76, 171–91.

White, H. 1980. A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–38.

Xiao, Z. and Linton, O. 2002. A nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis* 23, 215–50.