Econometric Theory, **21**, 2005, 158–170. Printed in the United States of America.
DOI: 10.1017/S0266466605050103

AUTOMATIC POSITIVE SEMIDEFINITE HAC COVARIANCE MATRIX AND GMM ESTIMATION

RICHARD J. SMITH
cemmap, U.C.L. and I.F.S.
and
University of Warwick

This paper proposes a new class of heteroskedastic and autocorrelation consistent (HAC) covariance matrix estimators. The standard HAC estimation method reweights estimators of the autocovariances. Here we initially smooth the data observations themselves using kernel function–based weights. The resultant HAC covariance matrix estimator is the normalized outer product of the smoothed random vectors and is therefore automatically positive semidefinite. A corresponding efficient GMM criterion may also be defined as a quadratic form in the smoothed moment indicators whose normalized minimand provides a test statistic for the overidentifying moment conditions.

1. INTRODUCTION

Consider a random vector process that may be parameter dependent and may display serial dependence and conditional heteroskedasticity. Heteroskedastic and autocorrelation consistent (HAC) estimation of the long-run covariance matrix of such processes has received considerable attention in the econometrics literature over the last two decades. The standard estimation method employs lag kernel smoothing whereby autocovariance estimators are weighted by some suitably chosen kernel function and that also incorporates a bandwidth parameter. Seminal contributions in the statistics literature to the theoretical study of such HAC estimators include Parzen (1957) and Priestley (1962). More recently, Andrews (1991) analyzes the properties of a number of HAC estimators and prescribes suitable choices of the bandwidth parameter given a particular choice

This paper is based on research originally presented at the Prospectus Workshop, Cowles Foundation for Research in Economics, Yale University. This revision has benefited much from constructive and useful comments by the editor, P.C.B. Phillips, and two referees on a revision and an earlier draft that was prepared during a visit to CREST-INSEE, Paris, whose hospitality is gratefully acknowledged. I am also grateful to P. Guggenberger and seminar participants for their helpful comments. Financial support for this research was provided by a 2002 Leverhulme Major Research Fellowship. Address correspondence to Richard J. Smith, Department of Economics, University of Warwick, Coventry CV4 7AL, U.K.; e-mail: r.j.smith@warwick.ac.uk.

  0266-4666/05 $12.00

of kernel function. In addition, the quadratic spectral kernel is shown to be optimal according to an asymptotic truncated mean squared error criterion. The bibliography in Andrews (1991) also provides some previous contributions to the econometrics literature on HAC covariance matrix estimation. In particular, the method of Newey and West (1987a) based on the Bartlett kernel function is now commonly adopted in many econometrics packages. A related lag kernel–based approach is discussed in Andrews and Monahan (1992) that initially prewhitens the random vector process.

This paper suggests a novel alternative class of HAC covariance matrix estimators. Rather than weight the estimated autocovariances as in the standard lag kernel method, we initially smooth the data observations on the random vector process itself using an appropriately chosen kernel function as weights. The HAC covariance matrix estimator is then defined as the normalized outer product of the smoothed random vectors. The resultant class of HAC covariance estimators belongs to the general class of quadratic estimators described in Grenander and Rosenblatt (1984, Sect. 4.1). Standard lag kernel estimators are also quadratic estimators, but their weight matrix has Toeplitz form. As shown by Grenander and Rosenblatt (1984, Sect. 4.2), for any linear random vector process, a standard lag kernel estimator may be found that has asymptotic mean squared error no larger than that of any given asymptotically unbiased quadratic estimator with non-Toeplitz weight matrix. Interestingly, however, the weight matrices for members of the class of HAC covariance matrix estimators proposed here asymptotically take Toeplitz form. Therefore, each HAC estimator corresponds implicitly to an asymptotically equivalent lag kernel estimator. Although this paper does not formally address their limiting distributional and asymptotic mean squared error properties, one might therefore suppose that members of this class of HAC covariance matrix estimators might inherit the properties of the corresponding lag kernel estimator.

The reweighting scheme adopted here may be viewed as a form of multitaper. Recall that the long-run covariance matrix of a stationary vector process is its (second-order) spectrum at frequency zero. Brillinger (1981, Theorem 5.6.4, p. 150) details the large-sample properties of a lag kernel estimator for the spectrum where the data observations have initially been smoothed using a single taper that has bandwidth parameter equal to sample size. Brillinger (1981, p. 151) notes that lag kernel estimators with tapered data may have desirable asymptotic bias properties relative to standard lag kernel estimators; see also Brillinger (1981, Sects. 3.3 and 4.6).[1] Thomson (1982) proposes a spectrum estimator using multitapers that is the average of a sequence of periodogram estimators each of which uses a different taper. In a more recent development, Walden (2000) defines a general class of multitaper spectrum estimators employing orthogonal tapers that may regarded as being defined in terms of the eigenvectors and eigenvalues of the weight matrix of some quadratic estimator. The number of tapers comprising the average is allowed to increase with sample size but generally at a slower rate; see Walden (2000, Sect. 4.9, p. 785). Our approach, however, displays a number of important differences from these multitaper methods. In particular, the estimator is a standard outer product, and, thus, the number of tapers equals the sample size. Also a bandwidth parameter is incorporated that is defined similarly to that used in the standard lag kernel method.

Another recent innovation is that of Phillips (2005), which proposes a new class of HAC covariance matrix estimators obtained as the explained sum of squares from a regression of the data observations on a sequence of basis trend functions. Estimators in this class are members of the general quadratic class where the weight matrix is the orthogonal projection matrix formed from the trend function sequence. These estimators may also be regarded as multitaper estimators where the tapers are the eigenvectors of the trend function projection matrix. Like the class proposed in this paper, Phillips's estimators also correspond implicitly to an asymptotically equivalent lag kernel estimator. Interestingly, this approach has the advantage of avoiding a choice of kernel function and bandwidth (which equals the sample size) but does, however, require a choice of the number and the sequence of basis trend functions to be included in the regression. Phillips (2005, eqn. (8)) provides an automated rule, obtained via an asymptotic mean squared error analysis, for the determination of the number of trend function terms to include.

The class of HAC covariance estimators proposed here is automatically positive semidefinite as are the estimators proposed in Phillips (2005). This property is a particular advantage if a consistent estimator is required for the asymptotic covariance matrix of the limiting normal distribution of some parameter estimator. For example, the asymptotic covariance matrix estimator may then be used in the computation of t- or F-type test statistics based on the parameter estimator. Furthermore, the inverse of the HAC covariance estimator may be employed as an estimator for the efficient metric in generalized method of moments (GMM) estimation.

The standard construction of t- and F-type statistics incorporating a HAC covariance matrix estimator has been subject to some severe criticism in the literature because of their poor finite-sample properties relative to the nominal normal or chi-square asymptotic reference distributions. In the regression context, Kiefer, Vogelsang, and Bunzel (2000) suggest an alternative approach to inference that completely avoids the use of a HAC estimator and appears to possess better properties in small samples than that based on standard t- and F-type statistics. For a more recent contribution, see Phillips, Sun, and Jin (2003). A major disadvantage of these approaches, however, from which t- and F-type statistics do not suffer, is that they seem to be restricted to just-identified models only in the nonlinear GMM context; see Kiefer et al. (2000, Sect. 4, p. 702).

An efficient GMM criterion may also be formulated as a quadratic form in the smoothed moment indicators whose normalized minimand provides a test statistic for the overidentifying moment conditions similar in structure to that of Hansen (1982). Being based solely on the smoothed moment indicators this GMM criterion is similar in structure to that for observations obtained from a random sample and thus does not require separate estimation of the efficient metric for GMM estimation, which would normally be the case; merely evaluation of the outer product of the smoothed moment indicators at an initially consistent estimator for the parameters of interest is necessary. A continuous updating estimator (Hansen, Heaton, and Yaron, 1996) may also be defined based on the revised GMM criterion.

Section 2 introduces the time series setup and briefly discusses the standard method of HAC covariance matrix estimation. The class of HAC covariance matrix estimators that is the subject of this paper is then defined. Consistency for this class of covariance matrix estimators is demonstrated. An alternative GMM criterion appropriate for serially dependent and conditionally heteroskedastic time series moment conditions is given in Section 3. Consistency, asymptotic normality, and efficiency of the GMM estimator are shown together with the limiting distribution of the normalized minimand. Section 4 concludes. Proofs of the results are given in the Appendix.

The following abbreviations are used throughout the paper: w.p.a.1: with probability approaching one; $\xrightarrow{p}$: converges in probability to; $\xrightarrow{d}$: converges in distribution to; $\|.\|$ is the matrix norm defined by $\|A\| = \sqrt{\lambda_{\max}(A'A)}$ where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$; p.d.: positive definite; p.s.d.: positive semidefinite.

2. HAC COVARIANCE MATRIX ESTIMATION

Let z_t, $(t = 1,\dots,T)$, denote observations on a finite-dimensional stationary and strongly mixing process $\{z_t\}_{t=1}^{\infty}$. The particular focus is the random vector $g(z_t,\beta)$, an m-vector of known functions of the data observation z_t and the p-vector β of unknown parameters, where it is assumed that $m \geq p$.

Let $g_t(\beta) \equiv g(z_t,\beta)$, $(t = 1,\dots,T)$, and $\hat{g}(\beta) \equiv \sum_{t=1}^{T} g_t(\beta)/T$.

We further assume that there exists a true value β_0 of the parameter vector β at which the vector $g_t(\beta)$ has unconditional mean zero, that is, $E[g_t(\beta_0)] = 0$. In the GMM estimation context (Hansen, 1982), $g_t(\beta)$ would denote a vector of moment indicators, and β_0 would be of some inferential interest. In many circumstances, the moment restrictions $E[g_t(\beta_0)] = 0$ will arise from a conditional moment restriction. For such cases, z_t would also need to include lagged endogenous and current and lagged values of exogenous variables.

The following assumption describes the basic properties of the observation process $\{z_t\}_{t=1}^{\infty}$.

Assumption 2.1. The observation process $\{z_t\}_{t=1}^{\infty}$ is a stationary and α-mixing sequence such that $\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$.

Our next assumption details some restrictions on the random process $\{g(z_t,\beta)\}_{t=1}^{\infty}$.

Assumption 2.2. (a) $E[g(z_t,\beta_0)] = 0$; (b) $E[\sup_{\beta\in\mathcal{B}}\|g(z_t,\beta)\|^{4\nu}] < \infty$.

These assumptions are quite standard; see Kitamura and Stutzer (1997), albeit somewhat stronger than Assumption A, p. 823, and the hypotheses in Lemma 1 of Andrews (1991, p. 824).

2.1. Some Preliminaries

The particular concern of this section is HAC estimation of the long-run covariance matrix of the random process $\{g(z_t,\beta_0)\}_{t=1}^{\infty}$ that is defined by

$$\Omega \equiv \sum_{s=-\infty}^{\infty} \Gamma(s), \tag{2.1}$$

where $\Gamma(s) = E[g_{t+s}(\beta_0)g_t(\beta_0)']$, $\Gamma(-s) = \Gamma(s)'$, is the sth autocovariance of the process $\{g(z_t,\beta_0)\}_{t=1}^{\infty}$, $(s = 0, \pm 1, \dots)$.

A HAC covariance matrix estimator is often required when estimating the asymptotic covariance matrix of the limiting normal distribution of a root-T consistent estimator for β_0. Furthermore, in a time series setting with unknown serial dependence and conditional heteroskedasticity, the inverse of a HAC estimator for Ω provides a consistent estimator for the metric required for the implementation of efficient GMM. In both of these examples, the particular requirement would be a consistent estimator for the limiting variance matrix $\lim_{T\to\infty}\mathrm{var}[T^{1/2}\hat{g}(\beta_0)]$ of the normalized sample average $T^{1/2}\hat{g}(\beta_0)$, that is, the long-run covariance matrix Ω as $\mathrm{var}[T^{1/2}\hat{g}(\beta_0)] = \sum_{s=1-T}^{T-1}\Gamma_T(s)$, where $\Gamma_T(s) \equiv \sum_{t=\max[1,1-s]}^{\min[T,T-s]} E[g_{t+s}(\beta_0)g_t(\beta_0)']/T$, $\Gamma_T(-s) = \Gamma_T(s)'$, $(s = 0, \pm 1, \dots)$. See (2.2) and (2.3) of Andrews (1991, pp. 819–820).

The standard method for estimation of Ω (2.1) is based on smoothing consistent sample autocovariance estimators $\hat{C}_T(s) \equiv T^{-1}\sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_{t+s}(\tilde{\beta}) \times g_t(\tilde{\beta})'$, $\hat{C}_T(-s) = \hat{C}_T(s)'$, $(s = 1 - T, \dots, T - 1)$, where $\tilde{\beta}$ is a preliminary consistent estimator for β_0. Let $k^*(\cdot)$ be some real-valued kernel function belonging to the class of symmetric kernels $\mathcal{K}_1$ defined by

$$\mathcal{K}_1 = \Big\{k^*(\cdot) : \mathcal{R} \to [-1,1] \,|\, k^*(0) = 1, k^*(-x) = k^*(x) \forall x \in \mathcal{R},$$
$$\int_{[0,\infty)} \bar{k}^*(x)\,dx < \infty,$$
$$k^*(\cdot) \text{ continuous at 0 and almost everywhere}\Big\}, \tag{2.2}$$

where $\bar{k}^*(x) = \sup_{y\geq x}|k^*(y)|$; see, for example, Andrews (1991) and Andrews and Monahan (1992).[2] The standard class of feasible HAC estimators for the limiting covariance matrix Ω (eqn. (2.1)) is then given by

$$\hat{\Omega}(\tilde{\beta}) \equiv \frac{T}{T-p} \sum_{s=1-T}^{T-1} k^*\left(\frac{s}{S_T}\right) \hat{C}_T(s) \tag{2.3}$$

(Andrews, 1991, (2.5), p. 820), where S_T is a bandwidth parameter and $T/(T-p)$ is a finite-sample adjustment that takes into account estimation of β_0.

2.2. Positive Semidefinite HAC Covariance Matrix Estimation

In contradistinction to the standard approach, we initially reweight the vectors $g_t(\beta)$, $(t = 1,\ldots,T)$, themselves to yield their smoothed counterparts

$$g_{tT}(\beta) \equiv \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) g_{t-s}(\beta), \qquad (t = 1,\ldots,T), \tag{2.4}$$

where, as before, $k(\cdot)$ is some kernel and S_T a bandwidth parameter, both of whose properties are defined in Assumption 2.3, which follows. The redefinition (2.4) of the random vectors $\{g(z_t,\beta)\}_{t=1}^{\infty}$ was suggested in Smith (2001) as a means of achieving an asymptotic first-order equivalence between generalized empirical likelihood and efficient GMM estimators in the moment condition framework with serially dependent and conditionally heteroskedastic moment indicators $\{g(z_t,\beta)\}_{t=1}^{\infty}$. See also Smith (1997) and Kitamura and Stutzer (1997), which employ related and special cases of the class of kernels $k(\cdot)$ and particular choices for the bandwidth parameter S_T.

The class of HAC covariance matrix estimators for Ω (eqn. (2.1)) proposed in this paper is formed directly as the normalized outer product of the smoothed random vectors $g_{tT}(\beta)$, $(t = 1,\ldots,T)$ (eqn. (2.4)), also evaluated at a preliminary consistent estimator $\tilde{\beta}$, namely,

$$\hat{\Omega}_T(\tilde{\beta}) \equiv \left(\sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right)^2\right)^{-1} \sum_{t=1}^{T} g_{tT}(\tilde{\beta}) g_{tT}(\tilde{\beta})'/T, \tag{2.5}$$

with the divisor $\sum_{s=1-T}^{T-1} k(s/S_T)^2$ as a necessary normalisation factor. Clearly $\hat{\Omega}_T(\tilde{\beta})$ (eqn. (2.5)) is p.s.d. and is a member of the general class of quadratic estimators (Grenander and Rosenblatt, 1984, Sect. 4.1). Restriction to consideration of p.s.d. HAC covariance estimators is particularly desirable as the estimator $\hat{\Omega}_T(\tilde{\beta})$ (eqn. (2.5)) would often form a component of a consistent estimator for the covariance matrix of the limiting normal distribution of a parameter estimator that may then be required for the construction of t- or F-type test statistics. Moreover, the property that a HAC covariance matrix estimator is p.s.d. is important if its inverse is to be used as the metric for efficient GMM estimation; see Section 3.

The next assumption introduces standard conditions on the kernel $k(\cdot)$ and bandwidth parameter S_T. Let $k_j \equiv \int_{-\infty}^{\infty} k(a)^j \, da$, $j = 1,2$, and

$$\bar{k}(x) = \begin{cases} \sup_{y \geq x} |k(y)| & \text{if } x \geq 0 \\ \sup_{y \leq x} |k(y)| & \text{if } x < 0. \end{cases}$$

Assumption 2.3. (a) $S_T \to \infty$ and $S_T/T^2 \to 0$; (b) $k(\cdot): \mathcal{R} \to [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, and is continuous at 0 and almost everywhere; (c) $\int_{-\infty}^{\infty} \bar{k}(x)\, dx < \infty$.

The bandwidth parameter therefore obeys the conditions described in Andrews (1991, Theorem 1(a), p. 827). Assumption 2.3(c) is required to ensure that certain normalized sums defined in terms of the kernel $k(\cdot)$ converge appropriately to their integral representation counterparts; see Jansson (2002) and note 2 in the Notes section.

To gain some intuition about the suitability of $\hat{\Omega}_T(\tilde{\beta})$ (eqn. (2.5)) as an estimator for Ω (eqn. (2.1)), consider the infeasible HAC covariance estimator $\hat{\Omega}_T(\beta_0)$. In the generalized empirical likelihood context, Smith (2001) discusses the asymptotic equivalence of $\hat{\Omega}_T(\beta_0)$ with the estimator $\Omega_T(\beta_0) = \sum_{s=1-T}^{T-1} k_T^*(s/S_T) C_T(s)$, where the infeasible sample covariances $C_T(s) \equiv T^{-1} \sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_{t+s}(\beta_0) g_t(\beta_0)'$, $C_T(-s) = C_T(s)'$, $(s = 1 - T, \ldots, T - 1)$, and the implicit kernel $k_T^*(\cdot)$ is given by $k_T^*(s/S_T) \equiv \sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k[(t-s)/S_T] k(t/S_T) / \sum_{t=1-T}^{T-1} k(t/S_T)^2$. The estimator $\Omega_T(\beta_0)$ also belongs to the general quadratic class but has Toeplitz weight matrix. Therefore, we might expect that the estimator $\hat{\Omega}_T(\beta_0)$ would inherit the desirable asymptotic mean squared error properties of standard lag kernel estimators (Grenander and Rosenblatt, 1984, Sect. 4.2). The implicit kernel $k_T^*(\cdot)$ approximates the kernel $k^*(\cdot)$ defined by $k^*(a) = \int_{-\infty}^{\infty} k(b-a) k(b)\, db/k_2$. Smith (2001) establishes that if Assumptions 2.3(b) and (c) hold then the induced $k^*(\cdot)$ belongs to the p.s.d. class $\mathcal{K}_2$ defined in Andrews (1991, p. 822), that is,

$$\mathcal{K}_2 = \{k^*(\cdot) \in \mathcal{K}_1 : K^*(\lambda) \geq 0 \text{ for all } \lambda \in \mathcal{R}\}, \tag{2.6}$$

where the class $\mathcal{K}_1$ is given in (2.2) and $K^*(\lambda) = (2\pi)^{-1} \int k^*(x) \exp(-\iota x \lambda)\, dx$ is the spectral window generator of the kernel $k^*(\cdot)$.

Given a choice of $k^*(\cdot)$, the corresponding kernel $k(\cdot)$ may be obtained from the relation $K^*(\lambda) = 2\pi |K(\lambda)|^2$, where $K(\cdot)$ is the spectral window generator of $k(\cdot)$. Smith (2001) provides examples of kernels $k(\cdot)$ that satisfy Assumption 2.3 and the consequent implicit kernels $k^*(\cdot)$.

Initially we consider the infeasible covariance matrix estimator $\hat{\Omega}_T(\beta_0)$ and state a preliminary result.[3]

LEMMA 2.1. (Consistency of $\hat{\Omega}_T(\beta_0)$). *If Assumptions 2.1–2.3 hold, then* $\hat{\Omega}_T(\beta_0) - \Omega \xrightarrow{p} 0$.

Lemma 2.1 parallels Proposition 1(a) of Andrews (1991, p. 825).

Let $\mathcal{N}$ denote some convex neighborhood of β_0. The next assumption states the root-T consistency of the preliminary estimator $\tilde{\beta}$ for β_0 and bounds the expectation of the derivative matrix of the vector $g(z_t, \beta)$ on $\mathcal{N}$.

Assumption 2.4. (a) $\sqrt{T}(\tilde{\beta} - \beta_0) = O_p(1)$; (b) $E[\sup_{\beta\in\mathcal{N}}\|\partial g(z_t,\beta)/\partial\beta'\|^2] < \infty$.

Assumption 2.4(a) mimics Assumption B(i) of Andrews (1991), and Assumption 2.4(b) is Assumption B(iii) of Andrews (1991, p. 825) rewritten for our context.

The consistency of the feasible HAC covariance matrix estimator $\hat{\Omega}_T(\tilde{\beta})$ then follows. Cf. Theorem 1(a) of Andrews (1991, p. 827).

THEOREM 2.1. (Consistency of $\hat{\Omega}_T(\tilde{\beta})$ for Ω). *If Assumptions 2.1–2.4 are satisfied, then* $\hat{\Omega}_T(\tilde{\beta}) - \Omega \xrightarrow{p} 0$.

3. EFFICIENT GMM ESTIMATION

The next assumption is standard and states regularity conditions for the consistency of GMM estimators.

Assumption 3.1. (a) $\beta_0 \in \mathcal{B}$ is the unique solution to $E[g(z_t,\beta)] = 0$; (b) $\mathcal{B}$ is compact; (c) $g(z_t,\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one; (d) Ω is p.d.

Let $\hat{g}_T(\beta) \equiv \sum_{t=1}^{T} g_{tT}(\beta)/T$. Then, if Assumptions 2.1–2.3 and 3.1 hold, from Smith (2001, Lemma A.1), $\hat{g}_T(\beta)/(S_T k_1) \xrightarrow{p} E[g(z_t,\beta)]$ uniformly $\beta \in \mathcal{B}$. Therefore, given a preliminary consistent estimator $\tilde{\beta}$, a HAC covariance estimator $\hat{\Omega}_T(\tilde{\beta})$ may be defined by either (2.5) as in Section 2 or as the centered version $(\sum_{s=1-T}^{T-1} k(s/S_T)^2)^{-1} \sum_{t=1}^{T}[g_{tT}(\tilde{\beta}) - \hat{g}_T(\tilde{\beta})][g_{tT}(\tilde{\beta}) - \hat{g}_T(\tilde{\beta})]'/T$ as $\hat{g}_T(\tilde{\beta}) \xrightarrow{p} 0$. Furthermore, from Assumption 3.1(d), by Theorem 2.1, w.p.a.1 $\hat{\Omega}_T(\tilde{\beta})$ is p.d.

Consider a GMM criterion based on the smoothed random vectors $g_{tT}(\beta)$, $(t = 1,\ldots,T)$, defined in (2.4), with $\hat{\Omega}_T(\tilde{\beta})^{-1}$ as efficient metric; that is, $\hat{Q}_T(\beta) \equiv \hat{g}_T(\beta)'\hat{\Omega}_T(\tilde{\beta})^{-1}\hat{g}_T(\beta)$. The GMM estimator $\hat{\beta}$ is then defined as

$$\hat{\beta} = \arg\min_{\beta\in\mathcal{B}} \hat{Q}_T(\beta). \tag{3.1}$$

The consistency of $\hat{\beta}$ for β_0 follows.

THEOREM 3.1. (Consistency of $\hat{\beta}$). *If Assumptions 2.1–2.4 and 3.1 are satisfied, then* $\hat{\beta} \xrightarrow{p} \beta_0$.

Asymptotic normality requires additional regularity conditions. Let $G \equiv E[\partial g(z_t,\beta_0)/\partial\beta']$.

Assumption 3.2. (a) $\beta_0 \in \text{int}(\mathcal{B})$; (b) $rk(G) = p$.

Let $\Sigma \equiv (G'\Omega^{-1}G)^{-1}$.

THEOREM 3.2. (Asymptotic normality of $\hat{\beta}$.) *If Assumptions 2.1–2.4, 3.1, and 3.2 are satisfied, then*

$$T^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,\Sigma), \qquad (T/S_T^2)\hat{Q}_T(\hat{\beta})/(k_1)^2 \xrightarrow{d} \chi^2_{m-p}.$$

The GMM estimator $\hat{\beta}$ ((3.1)) shares the standard properties of efficient GMM estimators in the class of GMM estimators that minimize the quadratic form GMM criterion $\hat{Q}^W(\beta) = \hat{g}(\beta)'W_n\hat{g}(\beta)$, $W_n = W + o_p(1)$, W p.d. Moreover, it is asymptotically first-order equivalent to the efficient GMM estimator that minimizes $\hat{Q}^W(\beta)$ with $W_n = \hat{\Omega}(\tilde{\beta})^{-1}$, where $\hat{\Omega}(\tilde{\beta})$ is defined in (2.3). The optimized criterion function statistic is likewise first-order asymptotically equivalent to the usual Hansen (1982) test statistic for overidentifying moment restrictions.

To estimate Σ consistently, consistent estimators of G and Ω are required. The former matrix may be estimated consistently by $\hat{G}_T(\hat{\beta})/(S_Tk_1)$, where $\hat{G}_T(\beta) \equiv \partial\hat{g}_T(\beta)/\partial\beta'$, which is an immediate by-product from the first-order conditions defining $\hat{\beta}$, and the latter matrix by $\hat{\Omega}_T(\tilde{\beta})$ (eqn. (2.5)) considered in Section 2 or as stated previously.

Although not pursued in this paper, a continuous updating GMM estimator for β_0 that will share the asymptotic properties of $\hat{\beta}$ given in Theorems 3.1 and 3.2 may be defined using the criterion $\hat{g}_T(\beta)'\hat{\Omega}_T(\beta)^-\hat{g}_T(\beta)$, where $\hat{\Omega}_T(\beta)^-$ is a generalized inverse for $\hat{\Omega}_T(\beta)$. The optimized criterion $(T/S_T^2)\hat{g}_T(\hat{\beta}_{CUE})'\hat{\Omega}_T(\hat{\beta}_{CUE})^-\hat{g}_T(\hat{\beta}_{CUE})/(k_1)^2 \xrightarrow{d} \chi^2_{m-p}$. An asymptotically equivalent statistic may be computed as $((k_2)/S_T^2(k_1)^2)$ times the (uncentered) explained sum of squares or $(k_2T)/(S_Tk_1)^2$ times the (uncentered) $\mathcal{R}^2$ from a least squares regression of 1 on $g_{tT}(\hat{\beta}_{CUE})$, $(t = 1,\ldots,T)$. The factors k_j may be replaced by $\sum_{t=1-T}^{T-1} k(t/S_T)^j/S_T$, $j = 1,2$. See Smith (2001) for further consideration of the continuous updating estimator $\hat{\beta}_{CUE}$ for β_0 that is a special case of the generalized empirical likelihood class of estimators considered there.

Test statistics for overidentifying moment restrictions and parametric restrictions on β_0 may be constructed in a similar fashion to those proposed in Newey (1985) and Newey and West (1987b), respectively. See also Smith (2001) for related test statistics.

4. SUMMARY AND CONCLUSIONS

A new class of HAC covariance matrix estimators is proposed. The point of departure for these estimators is that rather than smoothing estimated sample autocovariances the random vectors themselves are smoothed using kernel function weights. Consistency of the class is shown. A corresponding GMM criterion based on the smoothed random vectors is also defined. The resultant GMM estimator is first-order asymptotically equivalent to the efficient GMM estimator (Hansen, 1982) and thus shares the same large-sample properties. The normalized GMM criterion function is also asymptotically equivalent to the standard GMM criterion function statistic for overidentifying moment restrictions.

An analysis of the higher order properties of this class of HAC covariance matrix estimators similar to that in Andrews (1991) for the standard class would seem apposite to detail the optimal choice of kernel and bandwidth and to describe automatic bandwidth estimators. It also remains to examine the finite-sample properties of the various estimators and statistics suggested in this paper.

NOTES

1. An example of the application of tapers in econometrics is the nonparametric cointegration analysis of Bierens (1997).

2. Jansson (2002) notes that neither the square integrability condition $\int_{-\infty}^{\infty} k^*(x)^2\,dx < \infty$ in Andrews (1991, (2.6), p. 821) nor the stronger absolute integrability condition $\int_{-\infty}^{\infty} |k^*(x)|\,dx < \infty$ in Andrews and Monahan (1992, (2.5), p. 955) is sufficient for the consistency results claimed in those papers. The condition $\int_{[0,\infty)} \bar{k}^*(x)\,dx < \infty$ is required to rule out certain pathological cases and to ensure that particular summations used in those papers converge appropriately; see Lemma 1 of Jansson (2002).

3. A referee pointed out that the original proof of Lemma 2.1 abstracted from that of Lemma A.3 in Smith (2001) was incomplete. The revision of Smith (2001), currently in preparation, will address this deficiency.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

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

Bierens, H.J. (1997) Nonparametric cointegration analysis. *Journal of Econometrics* 77, 379–404.

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

Grenander, U. & M. Rosenblatt (1984) *Statistical Analysis of Stationary Time Series*. Chelsea.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P., J. Heaton, & A. Yaron (1996) Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14, 262–280.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

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

Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65, 861–874.

Newey, W.K. (1985) Generalized method of moments specification testing. *Journal of Econometrics* 29, 229–256.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, 2111–2245. North-Holland.

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

Newey, W.K. & K.D. West (1987b) Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777–787.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.

Phillips, P.C.B. (2005) HAC estimation by automated regression. *Econometric Theory* (this issue).

Phillips, P.C.B., Y. Sun, & S. Jin (2003) Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation. Cowles Foundation Discussion paper 1407, Yale University.
Priestley, M.B. (1962) Basic considerations in the estimation of spectra. *Technometrics* 4, 551–564.
Smith, R.J. (1997) Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal* 107, 503–519.
Smith, R.J. (2001) GEL Criteria for Moment Condition Models. Working paper, University of Bristol.
Thomson, D.J. (1982) Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70, 1055–1096.
Walden, A.T. (2000) A unified view of multitaper multivariate spectral estimation. *Biometrika* 87, 767–788.
White, H. (1984) *Asymptotic Theory for Econometricians*. Academic Press.

APPENDIX: PROOFS OF RESULTS

Proof of Theorem 2.1. Given Lemma 2.1, as in the proof of Theorem 1(a) in Andrews (1991, p. 852), it is only necessary to prove that the difference $\hat{\Omega}_T(\tilde{\beta}) - \hat{\Omega}_T(\beta_0) \xrightarrow{p} 0$. Without loss of generality let $\{g(z_t, \beta_0)\}_{t=1}^{\infty}$ be a scalar process.

Using a mean value expansion of $\hat{\Omega}_T(\tilde{\beta})$ about β_0

$$\begin{aligned}\hat{\Omega}_T(\tilde{\beta}) - \hat{\Omega}_T(\beta_0) &= 2\left[\sum_{t=1}^{T}\sum_{r=t-T}^{t-1} k\left(\frac{r}{S_T}\right) g_{t-r}(\bar{\beta}) \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) \partial g_{t-s}(\bar{\beta})/\partial\beta'/T\right] \\ &\quad\times\left[\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2\right]^{-1}(\tilde{\beta}-\beta_0) \\ &= 2\sum_{s=1-T}^{T-1}\sum_{r=\max[1,1-s]}^{\min[T,T-s]} g_{r+s}(\bar{\beta})\partial g_r(\bar{\beta})/\partial\beta'/T \\ &\quad\times\left[\sum_{t=1-r}^{T-r} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\Big/\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2\right](\tilde{\beta}-\beta_0),\end{aligned} \tag{A.1}$$

where $\bar{\beta}$ is on the line segment joining $\tilde{\beta}$ and β_0. Similarly to Andrews (1991, equation (A.10), p. 852), by Cauchy–Schwartz, w.p.a.1,

$$\begin{aligned}&\sup_{|s|\geq 1}\left\|\sum_{r=\max[1,1-s]}^{\min[T,T-s]} g_{r+s}(\bar{\beta})\partial g_r(\bar{\beta})/\partial\beta'/T\right\| \\ &\quad\leq\left(\sum_{r=1}^{T}\sup_{\beta\in\mathcal{N}} g_r(\beta)^2/T\right)^{1/2}\left(\sum_{r=1}^{T}\sup_{\beta\in\mathcal{N}}\|\partial g_r(\beta)/\partial\beta'\|^2/T\right)^{1/2} \\ &\quad= O_p(1),\end{aligned} \tag{A.2}$$

using Assumptions 2.2 and 2.4(b). Therefore, from equations (A.1) and (A.2),

$$\begin{aligned}\frac{\sqrt{T}}{S_T}|\hat{\Omega}_T(\tilde{\beta}) - \hat{\Omega}_T(\beta_0)| &\leq O_p(1)\left|\sum_{s=1-T}^{T-1}\sum_{t=1-r}^{T-r} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)/S_T^2\right| \\ &\quad\times\sqrt{T}\|\tilde{\beta}-\beta_0\|\Big/\left(\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2\Big/S_T\right).\end{aligned} \tag{A.3}$$

Now

$$\frac{1}{S_T^2}\left|\sum_{s=1-T}^{T-1}\sum_{t=1-r}^{T-r} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right| \le \frac{1}{S_T}\sum_{s=1-T}^{T-1}\frac{1}{S_T}\sum_{t=1-T}^{T-1}\left|k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right|$$
$$= \frac{1}{S_T}\sum_{t=1-T}^{T-1}\left(\frac{1}{S_T}\sum_{s=1-T}^{T-1}\left|k\left(\frac{t-s}{S_T}\right)\right|\right)\left|k\left(\frac{t}{S_T}\right)\right|. \tag{A.4}$$

Let $k_T(a) = k((s-1)/S_T)$, $(s-1)/S_T \le a < s/S_T$, if $s \le 0$, $k(s/S_T)$, $(s-1)/S_T < a \le s/S_T$, if $s > 0$. Using the change of variables $t = [S_T b]$ and $s = [S_T a]$, where $[\cdot]$ denotes the integer part of $\cdot$,

$$\frac{1}{S_T}\sum_{t=1-T}^{T-1}\left|k\left(\frac{t-s}{S_T}\right)\right| \le \lim_{T\to\infty}\frac{1}{S_T}\sum_{t=1-T}^{T-1}\left|k\left(\frac{t-s}{S_T}\right)\right|$$
$$= \lim_{T\to\infty}\int_{(1-T)/S_T}^{(T-1)/S_T} |k_T(b-a)|\,db + \lim_{T\to\infty}\frac{1}{S_T}|k(0)|$$
$$\le \lim_{T\to\infty}\int_{(1-T)/S_T}^{(T-1)/S_T} \bar{k}(b-a)\,db + o(1)$$
$$= \int_{-\infty}^{\infty}\bar{k}(b)\,db + o(1)$$

uniformly s. Hence, from equation (A.4), by Assumption 2.3(c),

$$\frac{1}{S_T^2}\sum_{s=1-T}^{T-1}\sum_{t=1-T}^{T-1}\left|k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right| \le \left(\int_{-\infty}^{\infty}\bar{k}(a)\,da + o(1)\right)^2$$
$$= O(1). \tag{A.5}$$

Similarly, using the change of variable $s = [S_T a]$, by the dominated convergence theorem, using Assumption 2.3(c),

$$\lim_{T\to\infty}\sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right)^2\Big/ S_T = \lim_{T\to\infty}\int_{(1-T)/S_T}^{(T-1)/S_T} k_T(a)^2\,da + \frac{1}{S_T}k(0)^2$$
$$= \int_{-\infty}^{\infty} k(a)^2\,da + o(1) > 0. \tag{A.6}$$

Therefore, substituting equations (A.5) and (A.6) into equation (A.3), by Assumption 2.4(a),

$$\frac{\sqrt{T}}{S_T}\,|\hat{\Omega}_T(\tilde{\beta}) - \hat{\Omega}_T(\beta_0)| = O_p(1). \tag{A.7}$$

The result then follows from equation (A.7) using Assumption 2.3(a). ■

Proof of Theorem 3.1. As Ω is p.d. from Assumption 3.1(d), $\hat{\Omega}_T(\tilde{\beta})$ is p.d. and invertible w.p.a.1. Let $g(\beta) \equiv E[g(z_t,\beta)]$. Under Assumption 2.1, $\{g(z_t,\beta_0)\}_{t=1}^{\infty}$ is a stationary and α-mixing sequence (White, 1984, Theorem 3.49, p. 47) and, thus, ergodic (White, 1984, Proposition 3.44, p. 46). By a uniform weak law of numbers (Smith, 2001, Lemma A.1), if Assumptions 2.1–2.3 and 3.1 hold, $\sup_{\beta\in\mathcal{B}}\|S_T^{-1}\hat{g}_T(\beta) - k_1 g(\beta)\| = o_p(1)$ and $g(\beta)$ is continuous by the strictly stationary and ergodic version of Lemma 2.4 in Newey and McFadden (1994, p. 2129). Let $Q(\beta) = g(\beta)'\Omega^{-1}g(\beta)$. Then, by Assumption 3.1(a), $Q(\beta)$ is uniquely minimized at β_0 and is continuous in $\beta \in \mathcal{B}$. Therefore, as $\lambda_{\min}[\hat{\Omega}_T(\tilde{\beta})] > 0$ w.p.a.1 where $\lambda_{\min}[\hat{\Omega}_T(\tilde{\beta})]$ is the smallest eigenvalue of $\hat{\Omega}_T(\tilde{\beta})$,

$$
\begin{aligned}
|S_T^{-2}\hat{Q}_T(\beta) - (k_1)^2 Q(\beta)| \le{}& \|S_T^{-1}\hat{g}_T(\beta) - k_1 g(\beta)\|^2 \lambda_{\min}[\hat{\Omega}_T(\tilde{\beta})]^{-1} \\
&+ 2\|g(\beta)\|\,\|S_T^{-1}\hat{g}_T(\beta) - k_1 g(\beta)\|\lambda_{\min}[\hat{\Omega}_T(\tilde{\beta})]^{-1} \\
&+ \|g(\beta)\|^2\,\|\hat{\Omega}_T(\tilde{\beta})^{-1} - \Omega^{-1}\| \\
={}& o_p(1)
\end{aligned}
$$

uniformly $\beta \in \mathcal{B}$. The result follows by Theorem 2.1 in Newey and McFadden (1994, p. 2121). ■

Proof of Theorem 3.2. As $\hat{\beta} \xrightarrow{p} \beta_0$ by Theorem 3.1, $\hat{\beta} \in \text{int}(\mathcal{B})$ w.p.a.1. Therefore the first-order conditions $\hat{G}_T(\hat{\beta})'\hat{\Omega}_T(\tilde{\beta})^{-1}\hat{g}_T(\hat{\beta}) = 0$ w.p.a.1, where $\hat{G}_T(\beta) \equiv \partial\hat{g}_T(\beta)/\partial\beta'$. By the mean value theorem, $\hat{g}_T(\hat{\beta}) = \hat{g}_T(\beta_0) + \hat{G}_T(\bar{\beta})(\hat{\beta} - \beta_0)$ where $\bar{\beta}$ lies on the line segment joining $\hat{\beta}$ and β_0 and may differ from row to row. An application of the uniform weak law of large numbers (Smith, 2001, Lemma A.1) to $S_T^{-1}\hat{G}_T(\bar{\beta})$ shows that $S_T^{-1}\hat{G}_T(\bar{\beta}) = k_1 G + o_p(1)$. Therefore, $[(k_1)^2\Sigma^{-1} + o_p(1)]T^{1/2}(\hat{\beta} - \beta_0) = -[k_1 G'\Omega^{-1} + o_p(1)](T^{1/2}/S_T)\hat{g}_T(\beta_0)$. Now, by a central limit theorem (Smith, 2001, Lemma A.2), $(T^{1/2}/S_T)\hat{g}_T(\beta_0) \xrightarrow{d} N(0,(k_1)^2\Omega)$, and the first conclusion follows.

As $T^{1/2}(\hat{\beta} - \beta_0) = -(k_1)^{-1}\Sigma G'\Omega^{-1}(T^{1/2}/S_T)\hat{g}_T(\beta_0) + o_p(1)$, $(T^{1/2}/S_T)\hat{g}_T(\hat{\beta}) = [I_m - G\Sigma G'\Omega^{-1}](T^{1/2}/S_T)\hat{g}_T(\beta_0) + o_p(1)$. Therefore, $(T/S_T^2)\hat{Q}_T(\hat{\beta})/(k_1)^2 = (T/(S_T k_1)^2)\hat{g}_T(\beta_0)'P\hat{g}_T(\beta_0) + o_p(1)$, where $P \equiv \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$. The second conclusion follows from Lemma A.2 of Smith (2001), $P\Omega P = P$, and $rk(P) = m - p$. ■