Econometrica, Vol. 72, No. 3 (May, 2004), 937–946

THE ERROR IN REJECTION PROBABILITY OF SIMPLE AUTOCORRELATION ROBUST TESTS

BY MICHAEL JANSSON[1]

A new class of autocorrelation robust test statistics is introduced. The class of tests generalizes the Kiefer, Vogelsang, and Bunzel (2000) test in a manner analogous to Anderson and Darling's (1952) generalization of the Cramér–von Mises goodness of fit test. In a Gaussian location model, the error in rejection probability of the new tests is found to be $O(T^{-1}\log T)$, where T denotes the sample size.

KEYWORDS: Asymptotic expansion, autocorrelation robust inference.

1. INTRODUCTION

IN MANY APPLICATIONS in time series econometrics, estimators that enjoy optimality properties in cross-sectional environments remain asymptotically normally distributed, albeit with a covariance matrix that depends on the autocovariance function of the data. A leading example is the OLS estimator in a linear regression model with exogenous regressors and an autocorrelated error term. In such cases, autocorrelation robust inference is typically based on Wald-type test statistics constructed by employing a standardization involving a consistent estimator of the asymptotic covariance matrix of an estimator of the parameter of interest (e.g., Robinson and Velasco (1997) and Wooldridge (1994)). While this approach often delivers inference procedures with certain asymptotic optimality properties, the finite sample null rejection probabilities of these procedures have been found to be somewhat less than satisfactory in many cases (e.g., den Haan and Levin (1997)).

Kiefer, Vogelsang, and Bunzel (2000, hereafter denoted KVB) demonstrate that the properties of Wald-type test statistics can be ameliorated if an inconsistent covariance matrix "estimator" is used and the critical values are adjusted to accommodate the randomness of the matrix employed in the standardization. Using higher-order asymptotic theory, the present paper provides an analytical explanation of the encouraging performance of the KVB procedure. In a Gaussian location model, the error in rejection probability (ERP) of the KVB test is found to be $O(T^{-1}\log T)$, where T denotes the sample size. The ERP of conventional procedures is no better than $O(T^{-1/2})$ under similar circumstances (Velasco and Robinson (2001)). In spite of the restrictive nature of the assumptions under which the higher-order asymptotic result is obtained, the rate $O(T^{-1}\log T)$ is remarkable in view of the fact that existing results on the performance of bootstrap-based autocorrelation robust inference procedures indicate that most of these procedures fail to achieve the same rate of convergence in the presence of nonparametric autocorrelation (Härdle, Horowitz, and Kreiss (2003)).

The bound on the ERP of the KVB test is established as a special case of a result characterizing the ERP of tests belonging to a new class of autocorrelation robust in-

[1]A previous version of this paper was entitled "Autocorrelation Robust Tests with Good Size and Power." The author thanks an editor, two referees, Guido Imbens, Jim Powell, Tom Rothenberg, Paul Ruud, Tim Vogelsang, and seminar participants at Aarhus, Brown, Cornell, Harvard/MIT, Houston/Rice, the 2002 NBER/NSF Time Series Conference, UC Berkeley, and USC for helpful comments.

ference procedures. Analogously to Anderson and Darling's (1952) generalization of the Cramér–von Mises goodness of fit test, the class of autocorrelation robust tests introduced here generalizes the KVB test by accommodating a weight function in the definition of the covariance matrix "estimator" used in the construction of the test statistic. All members of the new class of tests have an ERP exhibiting the same rate of decay as the ERP of the KVB test.

Section 2 introduces the model, states the assumptions under which formal results will be developed and introduces the new class of inference procedures. Section 3 studies the ERP of these procedures in a Gaussian location model, while mathematical derivations appear in Section 4.

2. PRELIMINARIES

Consider the location model

$$y_t = \beta + u_t \qquad (t = 1, \ldots, T), \tag{1}$$

where u_t is an unobserved error term satisfying the following high-level assumption.

ASSUMPTION A1: $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_t \to_d \omega W(\cdot)$, *where* $\omega^2 > 0$, $W(\cdot)$ *is a Wiener process, and* $\lfloor\cdot\rfloor$ *denotes the integer part of the argument.*

Assumption A1 is satisfied under a wide range of primitive moment and memory conditions on u_t (e.g., Phillips and Solo (1992)). An important implication of Assumption A1 is that the OLS estimator $\hat\beta = T^{-1}\sum_{t=1}^{T} y_t$ of β is root-T consistent and asymptotically normal:

$$\sqrt{T}(\hat\beta - \beta) \xrightarrow{d} \mathcal{N}(0, \omega^2).$$

Suppose the objective is to conduct a two-sided test of the simple null hypothesis $H_0 : \beta = \beta_0$, where β_0 is some constant. The standard approach is to base inference on a test statistic of the form

$$F_{HAC} = \frac{T(\hat\beta - \beta_0)^2}{\hat\omega^2_{HAC}},$$

where $\hat\omega^2_{HAC}$ is a consistent estimator of ω^2, the long-run variance of u_t. Although consistent estimators of ω^2 are available under conditions resembling Assumption A1 (e.g., Andrews (1991), Andrews and Monahan (1992), Hansen (1992), Jansson (2002), de Jong and Davidson (2000), Newey and West (1987, 1994), and Robinson (1991)), the ERP of tests based on F_{HAC} can be quite unsatisfactory in finite samples (e.g., den Haan and Levin (1997)).[2]

To the extent that the poor finite-sample properties of F_{HAC} are likely to be due to the fact that distributional approximations based on conventional asymptotic theory

[2]Kiefer and Vogelsang (2003) find that this problem can be mitigated by modeling the bandwidth (employed in the construction of $\hat\omega^2_{HAC}$) as a fixed proportion of the sample size when developing first-order asymptotic theory for F_{HAC}.

fail to capture the finite-sample variability in $\hat{\omega}^2_{HAC}$, it seems plausible that tests with better ERP can be obtained by employing an "estimator" of ω^2 whose limiting distribution is nondegenerate. Corroboration of this conjecture has been provided by KVB, who proposed the test statistic

$$F_{KVB} = \frac{T(\hat{\beta}-\beta_0)^2}{\hat{\omega}^2_{KVB}},$$

where $\hat{\omega}^2_{KVB} = T^{-2}\sum_{t=1}^{T-1}\hat{S}_t^2$ and $\hat{S}_t = \sum_{s=1}^{t}(y_s - \hat{\beta})$.

Unlike $\hat{\omega}^2_{HAC}$, $\hat{\omega}^2_{KVB}$ is not a consistent estimator of ω^2. Nevertheless, F_{KVB} is asymptotically pivotal under H_0 and the associated test has good finite-sample ERP and respectable (finite-sample and local asymptotic) power properties. Recently, Kiefer and Vogelsang (2002a) have shown that $\hat{\omega}^2_{KVB}$ equals one half times the kernel estimator of ω^2 computed using the Bartlett kernel with the bandwidth parameter equal to the sample size, while Kiefer and Vogelsang (2002b) have shown that the Bartlett kernel dominates other popular kernels in terms of the local asymptotic power of tests based on kernel estimators implemented with the bandwidth parameter equal to the sample size.

The present paper studies test statistics of the form

$$F_\kappa = \frac{T(\hat{\beta}-\beta_0)^2}{\hat{\omega}^2_\kappa},$$

where $\hat{\omega}^2_\kappa = T^{-2}\sum_{t=1}^{T-1}\kappa(t/T)\hat{S}_t^2$ and $\kappa(\cdot):(0,1)\to[0,\infty)$ is a (nonzero) weight function satisfying the following smoothness condition.

ASSUMPTION A2: *For some* $C_\kappa < \infty$, $|\sqrt{\kappa(r)} - \sqrt{\kappa(s)}| \le C_\kappa|r-s|$ *for all* $0 \le r \le s \le 1$.

When the weight function $\kappa(\cdot)$ is constant, the statistic F_κ is equivalent to F_{KVB}. On the other hand, nonconstant weight functions give rise to test statistics that are not covered by the results of Kiefer and Vogelsang (2002b). The statistic F_κ generalizes F_{KVB} in a manner analogous to Anderson and Darling (1952) generalization of the Cramér–von Mises goodness of fit test. Specifically, the limiting distribution of $\hat{\omega}^2_{KVB}$ is of the Cramér–von Mises variety, whereas the limiting representation of $\hat{\omega}^2_\kappa$ turns out to be the same as the statistic W^2 appearing in equation (4.5) of Anderson and Darling (1952).

3. ERP IN A GAUSSIAN LOCATION MODEL

In Monte Carlo experiments, KVB and Bunzel, Kiefer, and Vogelsang (2001) have found the ERP of F_{KVB} to be much smaller than that of F_{HAC}. A heuristic explanation of these findings can be found in Bunzel, Kiefer, and Vogelsang (2001, p. 1093), but to the best of this author's knowledge no previous paper has attempted to use higher-order asymptotic theory to provide an analytical explanation of the encouraging performance of the KVB procedure. As a first step in that direction, this paper derives the rate of convergence of F_κ to its (non-normal) limiting null distribution under the following strengthening of Assumption A1.

ASSUMPTION A1*: *$u_t = \psi(L)\eta_t$, where $\eta_t \sim$ i.i.d.$\mathcal{N}(0,1)$ and $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$ is a lag polynomial with $\psi(1) \neq 0$ and $\sum_{i=1}^{\infty} i|\psi_i| < \infty$.*

Employing a model similar to the one studied here, Velasco and Robinson (2001) derive Edgeworth expansions of the distribution of F_{HAC} under the assumption that $\hat{\omega}^2_{HAC}$ belongs to a certain class of kernel estimators. The leading term in the asymptotic expansion of the distribution function of F_{HAC} is no smaller than $O(T^{-1/2})$ when the bandwidth expansion rate is such that the order of the asymptotic mean squared error of $\hat{\omega}^2_{HAC}$ is minimized (Velasco and Robinson (2001, Section 4)).[3] In contrast, Theorem 1 shows that the ERP associated with F_κ is $O(T^{-1}\log T)$.

THEOREM 1: *If y_t is generated by* (1) *and Assumptions* A1* *and* A2 *hold, then*

$$\sup_{c\in\mathbb{R}}\left|\Pr\left(\frac{T(\hat{\beta}-\beta)^2}{\hat{\omega}^2_\kappa} \leq c\right) - \Pr\left(\frac{\mathcal{Z}^2}{\int_0^1 \kappa(r)B(r)^2\,dr} \leq c\right)\right| = O(T^{-1}\log T),$$

where $B(\cdot)$ is a Brownian bridge and $\mathcal{Z} \sim \mathcal{N}(0,1)$ is independent of $B(\cdot)$.

Normality plays an important simplifying role in the proof of Theorem 1, greatly facilitating the construction of a good coupling between the test statistic and its (non-normal) asymptotic representation. It is plausible that an extension of Theorem 1 to non-Gaussian time series can be based on Götze and Tikhomirov (2001), but an investigation along those lines is beyond the scope of this paper.

In view of Theorem 1, the fact that F_{KVB} dominates F_{HAC} in terms of finite-sample ERP is consistent with the predictions of higher-order asymptotic theory. Theorem 1 therefore complements the Monte Carlo results of KVB and Bunzel, Kiefer, and Vogelsang (2001) and sheds new light on these. Heuristically, the fast rate of decay of the ERP of F_κ is achieved by employing a standardization factor $\hat{\omega}^2_\kappa$ whose finite-sample distribution is well approximated by its asymptotic counterpart. Indeed, the contribution of the stochastic difference between $\hat{\omega}^2_\kappa$ and its limiting representation to the asymptotic expansion reported in Theorem 1 is (apart from a slowly varying factor) of the same order of magnitude as the contribution due to the stochastic difference between $T^{1/2}(\hat{\beta}-\beta)$ and its limiting normal distribution. In contrast, only the discrepancy between $\hat{\omega}^2_{HAC}$ and its probability limit is reflected in the leading term in Velasco and Robinson's (2001) asymptotic expansion of the ERP of F_{HAC}.

The best currently available results on the performance of bootstrap-based autocorrelation robust symmetrical testing procedures in the presence of nonparametric

[3]Specifically, it follows from equation (4) of Velasco and Robinson (2001) that

$$\Pr\left(\frac{T^{1/2}(\hat{\beta}-\beta_0)}{\hat{\omega}_{HAC}} \leq z\right) - \Phi(z) = p(z)T^{-\gamma} + o(T^{-\gamma})$$

for any $z \in \mathbb{R}$, where $t_{HAC} = T^{1/2}(\hat{\beta}-\beta_0)/\hat{\omega}_{HAC}$, $\Phi(\cdot)$ is the standard normal cdf, $p(\cdot)$ is an odd function, and $\gamma < 1/2$ is a constant. By implication, the leading term in the asymptotic expansion of $\Pr(F_{HAC} \leq c)$ is of the form $2p(\sqrt{c})T^{-\gamma} = O(T^{-\gamma})$ for any $c > 0$.

autocorrelation would appear to be those of Inoue and Shintani (2003).[4] Under somewhat weaker distributional assumptions than those of Theorem 1, Inoue and Shintani (2003) find that the rate of decay of the ERP can be made arbitrarily close to T^{-1} if the block bootstrap is applied to F_{HAC} and $\hat{\omega}^2_{HAC}$ is constructed using the truncated kernel.[5] Kitamura (1997) gives conditions under which the ERP of a Bartlett corrected blockwise empirical likelihood ratio test is $O(T^{-5/6})$.[6]

REMARKS: (i) One-sided tests of $H_0: \beta = \beta_0$ can be based on $t_\kappa = T^{1/2}(\hat{\beta} - \beta_0)/\hat{\omega}_\kappa$. The null distribution of this statistic is symmetric under the assumptions of Theorem 1. As a consequence, $\Pr(t_\kappa \le c) = \frac{1}{2} + \frac{1}{2}\Pr(F_\kappa \le c^2)$ for any $c > 0$ (with an analogous result holding for $c < 0$) and it follows that the ERP of one-sided tests is $O(T^{-1}\log T)$ under the assumptions of Theorem 1.

(ii) Bootstrap-based autocorrelation robust one-sided testing procedures have been studied by Choi and Hall (2000) and Götze and Künsch (1996). Choi and Hall (2000) find that the sieve bootstrap delivers an ERP with a polynomial rate of decay arbitrarily close to $O(T^{-1})$ under conditions similar to those of Theorem 1. Under weaker conditions, Götze and Künsch (1996) show that the ERP is no better than $O(T^{-3/4})$ when the block bootstrap is applied to F_{HAC} and $\hat{\omega}^2_{HAC}$ is constructed using the truncated kernel. (The ERP is no better than $O(T^{-2/3})$ when $\hat{\omega}^2_{HAC}$ is constructed using a kernel guaranteed to yield positive semidefinite estimates.)

(iii) It is an open question whether the rate reported in Theorem 1 can be improved. Shorack (2000, p. 261) claims that

$$\sup_{c\in\mathbb{R}} |\Pr(W + \Delta \le c) - \Pr(W \le c)| \le 4E|W\Delta| + 4E|\Delta|$$

for *any* random variables W and Δ. If that claim was correct, the factor $\log T$ could be omitted in Theorem 1. An easy counterexample to the displayed inequality is the following. Let W and Δ satisfy $\Pr(W = 0) = \Pr(\Delta = -\delta) = 1$ for some $\delta > 0$. Then

$$\sup_{c\in\mathbb{R}} |\Pr(W + \Delta \le c) - \Pr(W \le c)| = |\Pr(W + \Delta \le 0) - \Pr(\xi \le 0)| = 1,$$

whereas $4E|W\Delta| + 4E|\Delta| = 4\delta$ can be made arbitrarily close to zero.

(iv) As do conventional testing procedures, the test based on F_κ has nontrivial power against alternatives of the form $\beta = \beta_0 + O(T^{-1/2})$. Specifically, suppose y_t is generated by (1) and suppose Assumptions A1 and A2 hold. If $b = T^{1/2}\omega^{-1}(\beta - \beta_0)$ is fixed as T increases without bound, then

$$F_\kappa \xrightarrow{d} \frac{(\mathcal{Z} + b)^2}{\int_0^1 \kappa(r)B(r)^2\,dr}, \tag{2}$$

[4]Contemporary reviews of bootstrap methods for time series can be found in Bühlmann (2002), Härdle, Horowitz, and Kreiss (2003), and Politis (2003).

[5]The ERP is no better than $O(T^{-2/3})$ when $\hat{\omega}^2_{HAC}$ is constructed using a kernel guaranteed to yield positive semidefinite estimates.

[6]The ERP of the blockwise empirical likelihood ratio test is $O(T^{-2/3})$ if no Bartlett correction is employed.

where $B(\cdot)$ is a Brownian bridge and $\mathcal{Z} \sim \mathcal{N}(0,1)$ is independent of $B(\cdot)$. It seems plausible that a testing procedure associated with a suitably chosen weight function $\kappa(\cdot)$ will match the KVB procedure in terms of ERP and dominate it in terms of local asymptotic power. Research on this question is currently under way.

4. PROOF OF THEOREM 1

The result is obvious when $c \le 0$, so suppose $c > 0$. Let $X_N = \psi(1)^{-1}T^{1/2}(\hat\beta - \beta)$ and $X_D = \psi(1)^{-1}T^{-1}(\sqrt{\kappa(1/T)}\hat S_1, \ldots, \sqrt{\kappa((T-1)/T)}\hat S_{T-1})'$. By assumption, $X = (X_N, X_D')' \overset{d}{=} \Sigma^{1/2}\mathcal{Z}_T$, where "$\overset{d}{=}$" signifies equality in distribution,

$$\Sigma = \begin{pmatrix} \sigma_{NN} & \sigma_{DN}' \\ \sigma_{DN} & \Sigma_{DD} \end{pmatrix}$$

is the covariance matrix of X (partitioned in the obvious way), and $\mathcal{Z}_T \sim \mathcal{N}(0, I_T)$. Now, $\Pr[T(\hat\beta-\beta)^2/\hat\omega_\kappa^2 \le c] = \Pr[X_N^2 - cX_D'X_D \le 0] = \Pr[\mathcal{Z}_T' Y_T^c \mathcal{Z}_T \le 0]$, where

$$Y_T^c = \begin{pmatrix} \sigma_{NN} - c\sigma_{NN}^{-1}\sigma_{DN}'\sigma_{DN} & -c\sigma_{NN}^{-1/2}\sigma_{DN}'\Sigma_{DD.N}^{1/2} \\ -c\sigma_{NN}^{-1/2}\Sigma_{DD.N}^{1/2\prime}\sigma_{DN} & -c\Sigma_{DD.N} \end{pmatrix},$$

and $\Sigma_{DD.N} = \Sigma_{DD} - \sigma_{NN}^{-1}\sigma_{DN}\sigma_{DN}'$. Let

$$\dot Y_T^c = \begin{pmatrix} 1 & 0 \\ 0 & -c\dot\Sigma_{DD,T} \end{pmatrix},$$

where $\dot\Sigma_{DD,T}$ is the $(T-1)\times(T-1)$ matrix whose (i,j)th element is given by

$$\dot\Sigma_{DD,T}(i,j) = T^{-2}\sqrt{\kappa(i/T)}\sqrt{\kappa(j/T)}(\min(i,j) - ij/T).$$

Because $\Pr(\mathcal{Z}^2 - c\int_0^1 \kappa(r)B(r)^2\,dr \le 0) = \lim_{i\to\infty}\Pr(\mathcal{Z}_{2^iT}'\dot Y_{2^iT}^c\mathcal{Z}_{2^iT} \le 0)$, it follows from the triangle inequality that

$$\left|\Pr\left(\frac{T(\hat\beta-\beta)^2}{\hat\omega_\kappa^2} \le c\right) - \Pr\left(\frac{\mathcal{Z}^2}{\int_0^1 \kappa(r)B(r)^2\,dr} \le c\right)\right| \le \Delta_{1,T}^c + \Delta_{2,T}^c, \quad \text{where}$$

$$\Delta_{1,T}^c = |\Pr(\mathcal{Z}_T'Y_T^c\mathcal{Z}_T \le 0) - \Pr(\mathcal{Z}_T'\dot Y_T^c\mathcal{Z}_T \le 0)| \quad \text{and}$$

$$\Delta_{2,T}^c = \sum_{i=1}^\infty \left|\Pr\left(\mathcal{Z}_{2^iT}'\dot Y_{2^iT}^c\mathcal{Z}_{2^iT} \le 0\right) - \Pr\left(\mathcal{Z}_{2^{i-1}T}'\dot Y_{2^{i-1}T}^c\mathcal{Z}_{2^{i-1}T} \le 0\right)\right|.$$

The proof will be completed by using inequalities (3) and (4), developed next, to show that $\sup_{c>0}\Delta_{1,T}^c = O(T^{-1}\log T)$ and $\sup_{c>0}\Delta_{2,T}^c = O(T^{-1}\log T)$.

Now, $\mathcal{Z}_T'\dot Y_T^c\mathcal{Z}_T =_d \chi_0^2(1) - c\sum_{j=1}^{T-1}\lambda_{T,j}\chi_j^2(1)$, where $\chi_j^2(1) \sim$ i.i.d. $\chi^2(1)$ and $\lambda_{T,1} \ge \cdots \ge \lambda_{T,T-1} \ge 0$ are the eigenvalues of $\dot\Sigma_{DD,T}$. It follows from the convolution formula that $\sup_{r>0} r^{-1}\Pr(|\mathcal{Z}_T'\dot{Y}_T^c\mathcal{Z}_T| < r) < 2c^{-1}\lambda_{T,2}^{-1}$. By Mercer's theorem (e.g., Shorack and Wellner (1986)),

$$\lim_{T\to\infty} T\dot{\Sigma}_{DD,T}\left(\frac{r}{T},\frac{s}{T}\right) = \sqrt{\kappa(r)}\sqrt{\kappa(s)}(\min(r,s) - rs)$$

can be represented as $\sum_{j=1}^{\infty}\mu_j f_j(r)f_j(s)$, where the functions $\{f_j\}$ are orthonormal and continuous, and $\mu_1 \geq \mu_2 \geq \cdots \geq 0$. Because the right-hand side in the preceding display cannot be represented as $\mu_1 f_1(r)f_1(s)$, it must be the case that $\mu_2 > 0$. Therefore, $\underline{\lim}_{T\to\infty}\lambda_{T,2} \geq \mu_2 > 0$ and no generality is lost by assuming that

$$\Pr(|\mathcal{Z}_T'\dot{Y}_T^c\mathcal{Z}_T| < r) \leq c^{-1}M_\kappa r, \quad r > 0, \tag{3}$$

for all T and some finite constant M_κ (depending on $\kappa(\cdot)$).

Let Q_T be any symmetric matrix. Because the characteristic function of $\mathcal{Z}_T'Q_T\mathcal{Z}_T$ is given by $\varphi_T(t) = |I - 2itQ_T|^{-1/2}$, it can be shown that

$$E|\mathcal{Z}_T'Q_T\mathcal{Z}_T|^k = (-1)^{k/2}\frac{d^k}{dt^k}\varphi_T(t)\bigg|_{t=0} \leq \|Q_T\|^k\frac{(2k)!}{2^k k!} \leq 2\|Q_T\|^k(2k)^k\exp(-k)$$

for any even number k, where $\|Q_T\| = \max(|\operatorname{tr}(Q_T)|, \sqrt{\operatorname{tr}(Q_T^2)})$ and the second inequality follows from Stirling's formula. (The first inequality is sharp in the sense that equality holds whenever $Q_T = \operatorname{diag}(\|Q_T\|, 0, \ldots, 0)$.) For any $r \in [1,\infty)$,

$$E|\mathcal{Z}_T'Q_T\mathcal{Z}_T|^r \leq (E|\mathcal{Z}_T'Q_T\mathcal{Z}_T|^{k_r})^{r/k_r} \leq 2\|Q_T\|^r[2(r+1)]^r\exp(-r),$$

where k_r denotes the smallest even number exceeding r, the first inequality follows from the Hölder inequality, and the second inequality uses $2k_r \leq 2(r+1)$. By Markov's inequality,

$$\Pr[|\mathcal{Z}_T'Q_T\mathcal{Z}_T| > 2\|Q_T\|(r+1)] \leq 2\exp(-r), \quad r \in [1,\infty). \tag{4}$$

Elementary manipulations can be used to show that $\|Y_T^c - \dot{Y}_T^c\| \leq cM_{\kappa,\psi}T^{-1}$ for some finite constant $M_{\kappa,\psi}$ (depending on $\kappa(\cdot)$ and $\{\psi_j\}$). As a consequence,

$$\begin{aligned}
\Delta_{1,T}^c &\leq \Pr\big[|\mathcal{Z}_T'(Y_T^c - \dot{Y}_T^c)\mathcal{Z}_T| > 2\|Y_T^c - \dot{Y}_T^c\|(\log T + 1)\big] \\
&\quad + \Pr\big[|\mathcal{Z}_T'\dot{Y}_T^c\mathcal{Z}_T| < 2\|Y_T^c - \dot{Y}_T^c\|(\log T + 1)\big] \\
&\leq 2T^{-1} + 2c^{-1}M_\kappa\|Y_T^c - \dot{Y}_T^c\|(\log T + 1) \\
&\leq 2T^{-1} + 2M_\kappa M_{\kappa,\psi}T^{-1}(\log T + 1) \\
&= O(T^{-1}\log T),
\end{aligned}$$

where the first inequality uses Sargan and Mikhail (1971, Theorem 1) and the second inequality uses (3) and (4).

For any $i, T \geq 1$, it can be shown that $\mathcal{Z}'_{2^{i-1}T}\dot{Y}^c_{2^{i-1}T}\mathcal{Z}_{2^{i-1}T} =_d \mathcal{Z}'_{2^iT}\ddot{Y}^c_{2^iT}\mathcal{Z}_{2^iT}$ and $\mathcal{Z}'_{2^iT}\dot{Y}^c_{2^iT}\mathcal{Z}_{2^iT} =_d \mathcal{Z}'_{2^iT}\dddot{Y}^c_{2^iT}\mathcal{Z}_{2^iT}$, where

$$\ddot{Y}^c_{2^iT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -c\frac{1}{2}\dot{\Sigma}_{DD,2^{i-1}T} & -c\frac{1}{2}\dot{\Sigma}_{DD,2^{i-1}T} \\ 0 & 0 & -c\frac{1}{2}\dot{\Sigma}_{DD,2^{i-1}T} & -c\frac{1}{2}\dot{\Sigma}_{DD,2^{i-1}T} \end{pmatrix},$$

$$\dddot{Y}^c_{2^iT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -c\,\dddot{\sigma}_{00,2^iT} & -c\,\dddot{\sigma}'_{10,2^iT} & -c\,\dddot{\sigma}'_{20,2^iT} \\ 0 & -c\,\dddot{\sigma}_{10,2^iT} & -c\,\dddot{\Sigma}_{11,2^iT} & -c\,\dddot{\Sigma}'_{12,2^iT} \\ 0 & -c\,\dddot{\sigma}_{20,2^iT} & -c\,\dddot{\Sigma}_{21,2^iT} & -c\,\dddot{\Sigma}_{22,2^iT} \end{pmatrix},$$

$\dddot{\sigma}_{00,2^iT} = \dot{\Sigma}_{2^iT}(1,1)$, $\dddot{\sigma}_{10,2^iT}$ and $\dddot{\sigma}_{20,2^iT}$ are $(2^{i-1}T-1)$-vectors whose jth elements are $\dot{\Sigma}_{2^iT}(2j+1,1)$ and $\dot{\Sigma}_{2^iT}(2j,1)$, respectively, while $\dddot{\Sigma}_{11,2^iT}$, $\dddot{\Sigma}_{21,2^iT}$, and $\dddot{\Sigma}_{22,2^iT}$ are $(2^{i-1}T-1)\times(2^{i-1}T-1)$ matrices whose (j,k)th elements are $\dot{\Sigma}_{DD,2^iT}(2j+1,2k+1)$, $\dot{\Sigma}_{DD,2^iT}(2j,2k+1)$, and $\dot{\Sigma}_{DD,2^iT}(2j,2k)$, respectively.

Using the relation $\dot{\Sigma}_{DD,2^iT}(2j,2k) = \frac{1}{2}\dot{\Sigma}_{DD,2^{i-1}T}(j,k)$, it can be shown that each element of $\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}$ is bounded in absolute value by $cM^*_\kappa(2^iT)^{-2}$ for some finite constant M^*_κ (depending on $\kappa(\cdot)$). In particular, $\|\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}\| < cM^*_\kappa 2^{-i}T^{-1}$. As a consequence,

$$\begin{aligned}
\Delta^c_{2,T} &= \sum_{i=1}^{\infty}\left|\Pr\left(\mathcal{Z}'_{2^iT}\dddot{Y}^c_{2^iT}\mathcal{Z}_{2^iT} \leq 0\right) - \Pr\left(\mathcal{Z}'_{2^iT}\ddot{Y}^c_{2^iT}\mathcal{Z}_{2^iT} \leq 0\right)\right| \\
&\leq \sum_{i=1}^{\infty}\Pr\left[\left|\mathcal{Z}'_{2^iT}\left(\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}\right)\mathcal{Z}_{2^iT}\right| > 2\left\|\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}\right\|(\log(2^iT)+1)\right] \\
&\quad + \sum_{i=1}^{\infty}\Pr\left[\left|\mathcal{Z}'_{2^iT}\ddot{Y}^c_{2^iT}\mathcal{Z}_{2^iT}\right| < 2\left\|\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}\right\|(\log(2^iT)+1)\right] \\
&\leq 2\sum_{i=1}^{\infty}2^{-i}T^{-1} + \sum_{i=1}^{\infty}c^{-1}M_\kappa 2\left\|\dddot{Y}^c_{2^iT} - \ddot{Y}^c_{2^iT}\right\|(i\log 2 + \log T + 1) \\
&\leq 2T^{-1} + 2M_\kappa M^*_\kappa T^{-1}(2\log 2 + \log T + 1) \\
&= O(T^{-1}\log T),
\end{aligned}$$

where the first inequality uses Sargan and Mikhail (1971, Theorem 1) and the second inequality uses (3) and (4).

Dept. of Economics, UC Berkeley, CA 94720, U.S.A.; mjansson@econ.berkeley.edu.

Manuscript received August, 2002; final revision received October, 2003.

REFERENCES

ANDERSON, T. W., AND D. A. DARLING (1952): "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

BÜHLMANN, P. (2002): "Bootstraps for Time Series," *Statistical Science*, 17, 52–72.

BUNZEL, H., N. M. KIEFER, AND T. J. VOGELSANG (2001): "Simple Robust Testing of Hypotheses in Nonlinear Models," *Journal of the American Statistical Association*, 96, 1088–1096.

CHOI, E., AND P. HALL (2000): "Bootstrap Confidence Regions Computed from Autoregressions of Arbitrary Order," *Journal of the Royal Statistical Society*, *Series B*, 62, 461–477.

DE JONG, R. M., AND J. DAVIDSON (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407–423.

DEN HAAN, W. J., AND A. T. LEVIN (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics*, Vol. 15, ed. by G. S. Maddala and C. R. Rao. New York: North Holland, 299–342.

GÖTZE, F., AND H. R. KÜNSCH (1996): "Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics*, 24, 1914–1933.

GÖTZE, F., AND A. N. TIKHOMIROV (2001): "Asymptotic Expansions in Non-Central Limit Theorems for Quadratic Forms," Working Paper, University of Bielefeld.

HANSEN, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60, 967–972.

HÄRDLE, W., J. HOROWITZ, AND J.-P. KREISS (2003): "Bootstrap Methods for Time Series," *International Statistical Review*, 71, 435–459.

INOUE, A., AND M. SHINTANI (2003): "Bootstrapping GMM Estimators for Time Series," *Journal of Econometrics*, forthcoming.

JANSSON, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18, 1449–1459.

KIEFER, N. M., AND T. J. VOGELSANG (2002a): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093–2095.

——— (2002b): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

——— (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Working Paper, Cornell University.

KIEFER, N. M., T. J. VOGELSANG, AND H. BUNZEL (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

KITAMURA, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Data," *Annals of Statistics*, 25, 2084–2102.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

POLITIS, D. N. (2003): "The Impact of Bootstrap Methods on Time Series Analysis," *Statistical Science*, 18, 219–230.

ROBINSON, P. M. (1991): "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica*, 59, 1329–1363.

ROBINSON, P. M., AND C. VELASCO (1997): "Autocorrelation-Robust Inference," in *Handbook of Statistics*, Vol. 15, ed. by G. S. Maddala and C. R. Rao. New York: North-Holland, 267–298.

SARGAN, J. D., AND W. M. MIKHAIL (1971): "A General Approximation to the Distribution of Instrumental Variables Estimates," *Econometrica*, 39, 131–169.

SHORACK, G. R. (2000): *Probability for Statisticians*. New York: Springer-Verlag.

SHORACK, G. R., AND J. A. WELLNER (1986): *Empirical Processes with Applications to Statistics*. New York: John Willey & Sons.

VELASCO, C., AND P. M. ROBINSON (2001): "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean," *Econometric Theory*, 17, 497–539.

WOOLDRIDGE, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle and D. L. McFadden. New York: North-Holland, 2639–2738.