Econometrica, Vol. 68, No. 3 (May, 2000), 695–714

SIMPLE ROBUST TESTING OF REGRESSION HYPOTHESES

BY NICHOLAS M. KIEFER, TIMOTHY J. VOGELSANG, AND HELLE BUNZEL[1]

1. INTRODUCTION

IN THIS PAPER WE CONSIDER the problem of hypothesis testing in models with errors that have serial correlation or heteroskedasticity of unknown form. This situation is often encountered in regression models applied to economic time series data. It is a classic textbook result that while ordinary least squares (OLS) estimates of regression parameters remain consistent and asymptotically normal when errors are heteroskedastic or autocorrelated (provided usual regularity conditions hold and no lagged dependent variables are in the model), standard tests are no longer valid. If the true form of serial correlation/heteroskedasticity were known, then generalized least squares (GLS) provides efficient estimates and standard inference can be conducted on the GLS transformed model. But, in practice the form of serial correlation/heteroskedasticity is often unknown, and this has led to the development of techniques that permit valid asymptotic inference without having to specify a model of the serial correlation or heteroskedasticity. The most common approach is to estimate the variance-covariance matrix of the OLS estimates nonparametrically using spectral methods (heteroskedasticity and autocorrelation consistent (HAC) estimators) and construct standard tests using the asymptotic normality of the OLS estimates. HAC estimators have been extensively analyzed in the econometrics literature and important contributions are given by Andrews (1991), Andrews and Monahan (1992), Gallant (1987), Hansen (1992), Newey and West (1987), Robinson (1991, 1998), and White (1984) among others. The benefit of HAC estimator tests is asymptotically valid inference that is robust to general forms of serial correlation/heteroskedasticity in the errors.

We propose an alternative method of constructing robust test statistics. We apply a nonsingular data dependent stochastic transformation to the OLS estimates. The asymptotic distribution of the transformed estimates does not depend on nuisance parameters. Then, test statistics that are asymptotically invariant to nuisance parameters (asymptotic pivotal statistics) are constructed. The resulting test statistics have nonstandard asymptotic distributions that only depend on the number of restrictions being tested, and critical values are easy to simulate using standard techniques. The main advantage of our approach compared to standard approaches is that estimates of the variance-covariance matrix are not explicitly required to construct the tests. This is potentially important for two reasons. First, with the exception of the estimator proposed by Robinson (1998),[2] consistent nonparametric estimates of variance-covariance matrices in models with serial

[1] We thank Jean-Marie Dufour, Pierre Perron, an editor, and three anonymous referees for helpful comments and suggestions on an earlier version of the paper. We thank seminar participants at Boston College, Boston University, University of Aarhus, SMU, Yale University, Queen's University, Erasmus University, and the 1998 Summer Meetings of the Econometric Society in Montréal. We also thank the Center for Analytic Economics at Cornell University.

[2] Robinson (1998) showed that in certain models with serial correlation, asymptotic variance-covariance matrices of estimators can be consistently estimated using spectral methods but without any truncation or smoothing.

correlation require the choice of a truncation lag (bandwidth). Even if a data dependent method is used to choose the truncation lag, arbitrary choices must be made in practice, and these choices can affect inference. Our tests completely avoid these choices. Second, simulation studies have shown that sampling variability of HAC estimators in finite samples can lead to tests that have substantial size distortions (e.g. Andrews (1991), Andrews and Monahan (1992), and Den Haan and Levin (1997)). We report results from finite sample simulations that show that our new tests have better finite sample size properties than HAC estimator tests (including prewhitening).

The remainder of the paper is organized as follows. In the next section we lay out the model and review some well known OLS results. We show how the OLS estimates can be transformed so that their joint distribution becomes asymptotically invariant to serial correlation/heteroskedasticity nuisance parameters. Natural by-products of this transformation are t type statistics for testing hypotheses about individual parameters. In Section 3 we show how to construct tests of general linear hypotheses. Limiting null distributions are obtained, and asymptotic critical values are tabulated. The tests developed in these sections are natural extensions to regression models of the univariate trend function tests proposed by Vogelsang (1998). The tests in Vogelsang (1998) share the property that serial correlation parameters need not be estimated to carry out valid asymptotic inference. In Section 4 we show how our approach easily extends to GLS and instrumental variables (IV) estimation. We report results on local asymptotic power of the new tests compared to HAC estimator tests in Section 5. We show that the new tests have nontrivial local asymptotic power that is comparable but slightly below that of HAC estimator tests. We note that local asymptotic power calculations for HAC estimator tests are the same as those for tests with known variance-covariance parameters, while our statistic implicitly corrects for unknown variance-covariance parameters. In Sections 6 and 7 we report results on the finite sample behavior of the tests. Because the local asymptotic power approximation does not capture the influence of sampling variability of HAC estimators on finite sample power, we provide cases based on an empirical example where the power of the new tests dominates power of HAC estimator tests. Since our tests can be more powerful and they dominate HAC estimator tests in the accuracy of the asymptotic null approximation, our tests are very competitive in practice. Section 8 concludes and proofs are given in an Appendix.

2. THE MODEL AND SOME ASYMPTOTIC RESULTS

Consider the regression model given by

$$y_t = X_t'\beta + u_t \qquad (t = 1, 2, \ldots, T), \tag{1}$$

where β is a $(k \times 1)$ vector of regression parameters, X_t is a $(k \times 1)$ vector of regressors that may include a constant, and $\{u_t\}$ is a mean zero (conditional on X_t) random process. It is assumed that u_t does not have a unit root, but u_t may be serially correlated and have conditional heteroskedasticity. The following notation is used throughout the paper. Let $v_t = X_t u_t$ and define $\Omega = \Lambda\Lambda' = \Gamma_0 + \Sigma_{j=1}^{\infty}(\Gamma_j + \Gamma_j')$ where $\Gamma_j = E(v_t v_{t-j}')$ and Λ is a lower triangular matrix based on the Cholesky decomposition of Ω. Note that Ω is equal to 2π times the spectral density matrix of v_t evaluated at frequency zero. Define $S_t = \Sigma_{j=1}^{t} v_j$, which are the partial sums of $\{v_t\}$. Let $W_k(r)$ denote a k-vector of independent standard Wiener processes, and let $[rT]$ denote the integer part of rT where $r \in [0, 1]$. We use $\Rightarrow$ to denote weak convergence.

The following two assumptions regarding X_t and u_t are sufficient for us to obtain our main results.

ASSUMPTION A1: $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \Lambda W_k(r)$ *for all* $r \in [0,1]$,

ASSUMPTION A2: $\text{plim}(T^{-1}\sum_{t=1}^{[rT]} X_t X_t') = rQ$ *for all* $r \in [0,1]$ *and* Q^{-1} *exists*.

Assumption A1 holds under a variety of regularity conditions. One set of conditions is given by Phillips and Durlauf (1986) that require that v_t be weakly stationary, that the elements of v_t have a finite moment greater than two, and that v_t satisfy well known α-mixing conditions. These conditions permit conditional heteroskedasticity in $\{v_t\}$ but rule out most forms of unconditional heteroskedasticity. Andrews (1991) showed that consistent HAC estimators can be obtained under the stronger assumption that $\{v_t\}$ is fourth order stationary and α-mixing (see his Lemma 1). Assumption 1 is also satisfied by stationary and invertible ARMA processes with innovation with finite fourth moments (see Hall and Heyde (1980)). Assumption A1 rules out unit roots in $\{X_t\}$ and $\{u_t\}$.

Assumption A2 holds, for example, when X_t is a weakly (second order) stationary vector process and rules out trends in the regressors. However, the asymptotic results remain valid for certain hypotheses if the regressors are trend stationary. To be more precise suppose the regression model is $y_t = \mu + \gamma t + X_t'\beta + u_t$ and $X_t = \mu_x + \gamma_x t + \zeta_t$ where μ_x and γ_x are $(k \times 1)$ vectors, and $\{\zeta_t\}$ and $\{\zeta_t u_t\}$ satisfy Assumptions A1 and A2. In the Appendix we show that the new statistic proposed in this paper is invariant to projections of subsets of regressors. Therefore, hypotheses involving β can be tested using the regression $\tilde{y}_t = \tilde{X}_t\beta + \tilde{u}_t$ where $\{\tilde{y}_t\}$ and $\{\tilde{X}_t\}$ are residuals from the regression of $\{y_t\}$ and $\{X_t\}$ on $\{1,t\}$. This detrended regression satisfies Assumptions A1 and A2 because $\tilde{X}_t = \tilde{\zeta}_t$ and it is easy to show that $T^{-1/2}\sum_{t=1}^{[rT]}\tilde{\zeta}_t u_t = T^{-1/2}\sum_{t=1}^{[rT]}\zeta_t u_t + o_p(1)$ and $T^{-1}\sum_{t=1}^{[rT]}\tilde{\zeta}_t\tilde{\zeta}_t' = T^{-1}\sum_{t=1}^{[rT]}\zeta_t\zeta_t' = o_p(1)$. Once $\{t\}$ is included in the regression, the asymptotic results we obtain for tests of the β parameters do not apply to tests that involve the parameters μ (the intercept) or γ in which case the asymptotic distributions depend on the specific deterministic trends included in the regression.

Suppose regression (1) is estimated by OLS to obtain $\hat{\beta}$, the OLS estimate. The limiting distribution of $T^{1/2}(\hat{\beta}-\beta)$ follows directly from Assumptions A1 and A2 as

$$
\begin{aligned}
(2) \qquad T^{1/2}(\hat{\beta}-\beta) &= \left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T} X_t u_t \\
&= \left(T^{-1}\sum_{t=1}^{T} X_t X_t'\right)^{-1} T^{-1/2} S_T \\
&\Rightarrow Q^{-1}\Lambda W_k(1) \sim N(0, Q^{-1}\Lambda\Lambda' Q^{-1}) = N(0, Q^{-1}\Omega Q^{-1}) = N(0,V).
\end{aligned}
$$

This asymptotic normality result can be used to test hypotheses about β. To construct standard tests that are asymptotically invariant to nuisance parameters, an estimate of $V = Q^{-1}\Omega Q^{-1}$ is required. A natural estimator of Q^{-1} is $(T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}$. A HAC estimator of Ω can be constructed from $\hat{v}_t = X_t\hat{u}_t$ where $\hat{u}_t$ are the OLS residuals.

Consider the estimator $\hat{V} = (T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}\hat{\Omega}(T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}$ where $\hat{\Omega}$ is a HAC estimator of Ω. If we transform $T^{1/2}(\hat{\beta}-\beta)$ using $\hat{V}^{-1/2} = (T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}\hat{\Lambda}$ where $\hat{\Lambda}$ is obtained from the Cholesky decomposition of $\hat{\Omega}$, we have

$$
(3) \qquad \hat{V}^{-1/2} T^{1/2}(\hat{\beta}-\beta) \Rightarrow N(0, I_k).
$$

Using (3), hypotheses about individual β's can be tested using t statistics in the usual way with standard errors given by square roots of the diagonal elements of the matrix $\hat{V}/T$. The asymptotic theory does not explicitly account for the effects of sampling variation in $\hat{V}$, and this variation is potentially important in finite samples.

We take a different approach to testing that is similar in spirit to the transformation in (3) except that we transform $T^{1/2}(\hat{\beta}-\beta)$ using a moment matrix constructed from the data that does not require an estimate of Ω. Define $\hat{S}_t = \sum_{j=1}^{t} X_j \hat{u}_j = \sum_{j=1}^{t} \hat{v}_j$. Using Assumptions A1 and A2, the limiting behavior of $T^{-1/2}\hat{S}_{[rT]}$ as $T \to \infty$ is

$$
\begin{aligned}
(4)\qquad T^{-1/2}\hat{S}_{[rT]} &= T^{-1/2}\sum_{t=1}^{[rT]} X_t \hat{u}_t = T^{-1/2}\sum_{t=1}^{[rT]} \{X_t u_t - X_t X_t'(\hat{\beta}-\beta)\} \\
&= T^{-1/2}\sum_{t=1}^{[rT]} v_t - \left(T^{-1}\sum_{t=1}^{[rT]} X_t X_t'\right) T^{1/2}(\hat{\beta}-\beta) \\
&= T^{-1/2} S_{[rT]} - \left(T^{-1}\sum_{t=1}^{[rT]} X_t X_t'\right) T^{1/2}(\hat{\beta}-\beta) \\
&\Rightarrow \Lambda W_k(r) - rQQ^{-1}\Lambda W_k(1) = \Lambda(W_k(r) - rW_k(1)).
\end{aligned}
$$

Consider $\hat{C} = T^{-2}\sum_{t=1}^{T} \hat{S}_t \hat{S}_t'$. From (4) and the continuous mapping theorem we have

$$
(5)\qquad \hat{C} \Rightarrow \Lambda\left[\int_0^1 (W_k(r) - rW_k(1))(W_k(r) - rW_k(1))'\,dr\right]\Lambda'.
$$

To simplify later developments let $P_k = \int_0^1 (W_k(r) - rW_k(1))(W_k(r) - rW_k(1))'\,dr$, which is the integral of the outer product of a k-dimensional multivariate Brownian bridge. In the univariate case P_1 is the limiting distribution of the Cramér-von Mises statistic and is related to the Anderson-Darling statistic. Because P_k is positive definite by construction, we can use a Cholesky decomposition to write $P_k = Z_k Z_k'$ or equivalently $P_k^{-1} = (Z_k')^{-1} Z_k^{-1}$ where Z_k is lower triangular.

Now consider $\hat{B} = (T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}\hat{C}(T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}$. Define $\hat{M} = (T^{-1}\sum_{t=1}^{T} X_t X_t')^{-1}\hat{C}^{1/2}$ with $\hat{C}^{1/2}$ lower triangular and the Cholesky decomposition of C. Consider the transformation $\hat{M}^{-1}T^{1/2}(\hat{\beta}-\beta)$. It follows directly from (2) and (5):

$$
(6)\qquad \hat{M}^{-1}T^{1/2}(\hat{\beta}-\beta) \Rightarrow Z_k^{-1}W_k(1).
$$

This transformation results in a limiting distribution that does not depend on the nuisance parameters Q and Ω. The distribution of $Z_k^{-1}W_k(1)$ is nonstandard. Because $W_k(1)$ and $W_k(r) - rW_k(1)$ are Gaussian and $E[W_k(1)(W_k(r) - rW_k(1))] = 0$ for all $r \in [0,1]$, they are independent, and it follows that Z_k and $W_k(1)$ are independent as well. Therefore, conditional on Z_k, $Z_k^{-1}W_k(1) \sim N(0, P_k^{-1})$. If we let $p(P_k)$ denote the distribution function of P_k, we can write the unconditional distribution of $Z_k^{-1}W_k(1)$ as $\int_0^1 N(0, P_k^{-1})p(P_k)\,dP_k$, which is a mixture of normals. Thus, the distribution of $Z_k^{-1}W_k(1)$ is symmetric with thicker tails than a normal distribution. This result is analogous to Fisher's classic development of the t statistic. After using a data dependent stochastic transformation (dividing by a moment of the data proportional to the error variance), Fisher obtained a finite sample distribution free of nuisance parameters with fatter tails than a normal distribution, a t distribution. This analogy is not exact as we obtain a distribution free of nuisance parameters only asymptotically, and the distribution of $Z_k^{-1}W_k(1)$ is not equivalent to a multivariate t distribution. But, the analogy is accurate as a nuisance parameter is eliminated and this results in increased dispersion of the null limiting distribution.

Hypotheses about individual β's can be tested using t type statistics, which we label t^*, that are constructed in the same way as usual t statistics with the usual standard errors replaced with square roots of the diagonal elements of the $\hat{B}/T$ matrix. Because the t^* statistics are invariant to the ordering of the regressors, the limiting distribution of any t^* is given by the first element in the vector $Z_k^{-1}W_k(1)$. Using the fact that Cholesky decompositions are lower triangular, it is easy to show that the first element of $Z_k^{-1}W_k(1)$ has the same distribution as $W_1(1)/[\int_0^1(W_1(r)-rW_1(1))^2\,dr]^{1/2}$. Therefore, as $T \to \infty$

$$t^* \Rightarrow W_1(1)/\left[\int_0^1 (W_1(r)-rW_1(1))^2\,dr\right]^{1/2}. \tag{7}$$

Critical values of (7) were computed using simulations and are tabulated in Table I. The Wiener process, $W_1(r)$, was approximated by normalized sums of i.i.d. $N(0,1)$ pseudo random deviates using 1,000 steps and 50,000 replications. The simulations were written in the GAUSS programming language using an initial seed of 1,000 for the random number generator. We also computed the density of (7) and the density of (7) with variance normalized to one by smoothing the 50,000 realizations of (7) using standard kernel techniques.[3] These densities are plotted in Figure 1 along with the density of a standard normal random variable. The asymptotic distribution of the normalized t^* has tails slightly fatter than a standard normal random variable.

3. TESTS FOR GENERAL LINEAR HYPOTHESES

Suppose we are interested in testing more general linear hypotheses of the form

$$H_0\colon R\beta = r, \qquad H_1\colon R\beta \neq r,$$

where R is a $(q\times k)$ matrix with rank q and r is a $(q\times 1)$ vector. When the null hypothesis is true, we have $R\hat{\beta}-r=R(\hat{\beta}-\beta)$. To motivate the new statistic consider $T^{1/2}R(\hat{\beta}-\beta)$. From (2) it follows that $T^{1/2}R(\hat{\beta}-\beta)\Rightarrow RQ^{-1}\Lambda W_k(1)$. Because $W_k(1)$ is a vector of independent Wiener processes and is Gaussian, $RQ^{1}\Lambda W_k(1)$ is equivalent in distribution to $\Lambda^* W_q^*(1)$ where $W_q^*(1)$ is a $(q\times 1)$ vector of independent Wiener processes and Λ^* is the $(q\times q)$ matrix square root of $RQ^{-1}\Lambda\Lambda'Q^{-1}R'$. Λ^* exists and is invertible because the matrix $RQ^{-1}\Lambda\Lambda'Q^{-1}R'$ has full rank of q. Now consider the matrix $R\hat{B}R'$. It is simple to show that $R\hat{B}R'\Rightarrow RQ^{-1}\Lambda P_k\Lambda'Q^{-1}R$, which is equivalent in distribution to $\Lambda^*P_q^*\Lambda^{*\prime}$ (see the Appendix). Let $\hat{M}^*$ denote the matrix square root of $R\hat{B}R'$ and note that $\hat{M}^*\Rightarrow[\Lambda^*P_q^*\Lambda^{*\prime}]^{1/2}=\Lambda^*Z_q^*$. Suppose we transform $T^{1/2}R(\hat{\beta}-\beta)$ using $\hat{M}^{*-1}$ giving $\hat{M}^{*-1}T^{1/2}R(\hat{\beta}-\beta)$. It evidently follows that $\hat{M}^{*-1}T^{1/2}R(\hat{\beta}-$

TABLE I

ASYMPTOTIC CRITICAL VALUES OF t^*

1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
−8.544	−6.811	−5.374	−3.890	0.000	3.890	5.374	6.811	8.544

[3] We computed the variance of (7) to be 10.893 using the sample variance of the 50,000 replications.

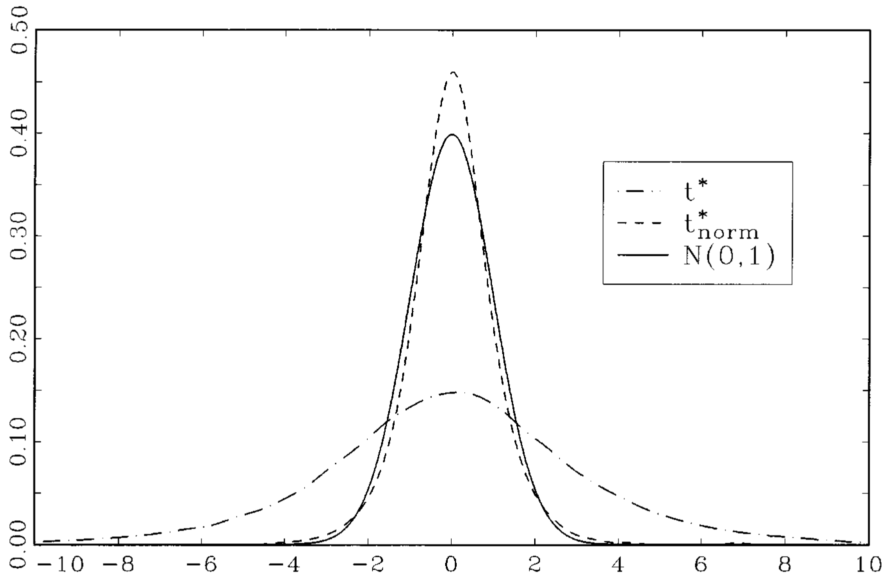

FIGURE 1.—Densities of t^*, t^*_{norm}, and $N(0,1)$.

$\beta) \Rightarrow Z_q^{*-1} W_q^*(1)$, which is free of nuisance parameters. Forming the usual quadratic form using $\hat{M}^{*-1} T^{1/2} R(\hat{\beta}-\beta)$ gives

$$(8) \qquad [\hat{M}^{*-1} T^{1/2} R(\hat{\beta}-\beta)]'[\hat{M}^{*-1} T^{1/2} R(\hat{\beta}-\beta)]$$
$$= T(R(\hat{\beta}-\beta))'[R\hat{B}R']^{-1} R(\hat{\beta}-\beta).$$

The quadratic form (8) suggests the following statistic for testing H_0 against H_1:

$$F^* = T(R\hat{\beta}-r)'[R\hat{B}R']^{-1}(R\hat{\beta}-r)/q.$$

Notice that F^* is the classic F test except that $\hat{B}$ replaces $\hat{V}$. (If $T^{1/2}R(\hat{\beta}-\beta)$ were transformed using the matrix square root of $R\hat{V}R'$, the quadratic form would lead to the construction of the classic F test based on a HAC estimate of V.) We prove in the Appendix the following asymptotic result:

THEOREM 1: *Suppose that Assumptions* $A1$ *and* $A2$ *hold. Then under the null hypothesis* H_0: $R\beta = r$, $F^* \Rightarrow W_q(1)' P_q^{-1} W_q(1)/q$ *as* $T \to \infty$.

The limiting distribution of F^* is free of nuisance parameters and only depends on q. The distribution is nonstandard, but critical values can easily be simulated because the distribution is a function of independent standard Wiener processes. By approximating each Wiener process in the vector $W_q(r)$ using the same techniques that were used to simulate (7), critical values of $W_q(1)' P_q^{-1} W_q(1)/q$ were computed for $q = 1, 2, \ldots, 29, 30$ and are tabulated in Table II. Because the distribution depends only on q, using Table II is no more difficult in practice than using a chi-square distribution table. In addition,

TABLE II

ASYMPTOTIC CRITICAL VALUES OF F^*

%	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$	$q=7$	$q=8$	$q=9$	$q=10$
90.0	28.88	35.68	42.39	48.79	55.02	61.18	67.37	73.10	78.52	83.84
95.0	46.39	51.41	58.17	65.33	71.69	78.70	84.63	90.89	96.38	101.8
97.5	65.94	69.76	76.07	83.35	89.65	96.53	102.7	109.8	114.2	120.0
99.0	101.2	96.82	100.7	108.4	114.2	121.2	126.9	134.4	139.6	144.9

%	$q=11$	$q=12$	$q=13$	$q=14$	$q=15$	$q=16$	$q=17$	$q=18$	$q=19$	$q=20$
90.0	89.39	94.47	100.1	105.3	110.3	115.5	121.2	126.6	131.5	136.5
95.0	107.7	113.6	119.9	125.5	131.5	136.6	141.4	147.1	152.9	158.0
97.5	127.2	132.9	138.8	145.2	151.0	155.9	161.1	167.6	174.0	179.8
99.0	152.6	157.8	163.8	169.7	174.7	181.6	188.8	194.8	203.2	208.5

%	$q=21$	$q=22$	$q=23$	$q=24$	$q=25$	$q=26$	$q=27$	$q=28$	$q=29$	$q=30$
90.0	141.9	146.6	152.1	157.0	161.8	167.2	171.6	177.0	181.6	187.0
95.0	163.6	169.3	174.7	180.3	184.9	190.7	196.0	201.5	206.4	211.4
97.5	186.0	191.2	197.0	202.3	207.5	213.3	218.9	224.4	229.1	236.0
99.0	214.0	219.3	224.6	230.1	236.3	242.4	246.9	252.9	259.8	266.3

Note: q is the number of restrictions being tested.

asymptotic critical values for a wide range of percentage points for $q = 1, 2, \ldots 40$ have been computed by MacKinnon (1999) using response surface methods.[4]

Construction of the F^* statistic amounts to replacing the HAC estimator, $\hat{\Omega}$, with $\hat{C}$ and using the scaling matrix $\hat{B}$ in place of the usual scaling matrix $\hat{V}$. The scaling matrix $\hat{B}$ converges to a random matrix rather than the fixed variance-covariance matrix. Viewed in this way, our approach creates a new class of statistics that are robust to serial correlation/heteroskedasticity in the errors and are asymptotically pivotal. In general, an asymptotically pivotal statistic can be obtained by replacing $\hat{\Omega}$ with any moment matrix of the data that has a limiting distribution of the form $\Lambda f(W_k(r))\Lambda'$ where $f(W_k(r))$ is a random matrix that is a functional of $W_k(r)$. Therefore, our particular choice of $\hat{C}$ is somewhat arbitrary and to some degree ad hoc, but it yields an elegant distribution theory with asymptotic distributions that do not depend on R, r, or k. Other choices of $\hat{C}$ might not satisfy this property. In addition, we prove in the Appendix that our choice of $\hat{C}$ ensures that F^* is invariant to projecting out subsets of regressors, i.e. F^* satisfies the Frisch-Waugh-Lovell (FWL) Theorem (see Davidson and MacKinnon (1993)).[5] Also, F^* has the important practical property of invariance to rescaling of the regressors (i.e. invariance to units of measurement).[6] This discussion raises the natural question as to

[4] Consult the personal web page of James MacKinnon, http://www.econ.queensu.ca/pub/faculty/mackinnon for a computer program that calculates asymptotic P values and critical values.

[5] HAC based tests satisfy the FWL Theorem only if a fixed truncation lag is used without prewhitening. Therefore, if an automatic truncation lag and/or prewhitening is used, different test statistics can result when one, say, first detrends regressors before estimating a regression as opposed to directly including a trend in the regression. See Section 7 for an example.

whether a theory of optimality can be created to help guide the choice of $\hat{C}$. We leave this interesting and challenging problem as an open research topic.

4. EXTENSIONS TO GLS AND IV ESTIMATION

In this section we briefly discuss how the F^* statistic can be applied to more general regression models that include GLS and IV estimation. Stack y_t, X_t, and u_t into matrices y, X, and u and consider a transformation of regression (1),

$$y^* = X^*\beta + u^*, \tag{9}$$

where $y^* = \Psi y$, $X^* = \Psi X$, $u^* = \Psi u$, and Ψ is a $(T \times T)$ transformation matrix. Estimating (9) by OLS is equivalent to minimizing $(y^* - X^*\beta)'\Psi'\Psi(y^* - X^*\beta)$. When $\Psi = \Sigma^{1/2}$ where $\Sigma = E(uu')$ we obtain the GLS estimate of β. When $\Psi = Z(Z'Z)^{-1}Z'$ where Z is a $(T \times m)$ vector of instruments with $m \geq k$ and $E(Z_t u_t) = 0$, we obtain the IV estimate of β. Provided that $v_t^* = X_t^* u_t^*$ and $T^{-1}\sum_{t=1}^{[rT]} X_t^* X_t^{*\prime}$ satisfy Assumptions A1 and A2, Theorem 1 still applies to F^* constructed from regression (9). In the case of IV estimation, sufficient conditions are stationary Z_t and $\text{plim}(T^{-1}\sum_{t=1}^{T} Z_t X_t') \neq 0$.

The natural extension beyond regression (9) is to consider a generalized methods of moments (GMM) framework that would include OLS, GLS, and IV as special cases. This would be an important extension as GMM models are widely used in empirical macroeconomics. We conjecture that Theorem 1 generalizes to GMM models, but it is not clear whether standard GMM regularity conditions will be sufficient to obtain such a result. Furthermore, in overidentified GMM models, it is not obvious whether F^* should be constructed using sample analogs of the original moment conditions or sample analogs of the moment conditions implied by the weighted GMM minimization problem. It is unclear how the choice of weighting matrix will affect the asymptotic properties of F^*. An extension to GMM models is nontrivial and is beyond the scope of this paper.

5. LOCAL ASYMPTOTIC POWER $k = 1$ CASE

In this section we contrast the power properties of the t^* and HAC estimator t statistic using a local asymptotic framework. Of course, both tests have unit power against nonlocal alternatives. We restrict attention to the single regressor case ($k = 1$) as this special case sufficiently illustrates local asymptotic power comparisons. With $k = 1$ the regression becomes

$$y_t = \beta x_t + u_t, \tag{10}$$

where β and x_t are scalars and x_t is mean zero (this assumption has no effect on the local power results). We consider testing the null hypothesis H_0: $\beta \leq \beta_0$ against the alternative H_1: $\beta > \beta_0 + cT^{-1/2}$. Under the alternative, we model β as local to β_0 such that β converges to β_0 at rate $T^{-1/2}$ with local alternative parameter c. Let $\sigma_x^2 = E(x_t^2)$, and let $\sigma^2 = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$ where $\gamma_j = E(v_t v_{t-j})$ with $v_t = x_t u_t$. The parameter σ_x^2 is the variance of x_t, and the parameter σ^2 is equal to 2π times the spectral density of v_t evaluated at frequency zero. Define $\hat{\sigma}_x^2 = T^{-1}\sum_{t=1}^{T} x_t^2$ and let $\hat{\sigma}^2$ be a HAC estimator of

[6] As a referee pointed out, the F^* statistic is not invariant to the ordering of the observation (as is White's HC estimator). We anticipate F^* being used in time series settings where there is a natural ordering of the data. Should F^* be used in a pure cross section situation, ordering of the data becomes an issue.

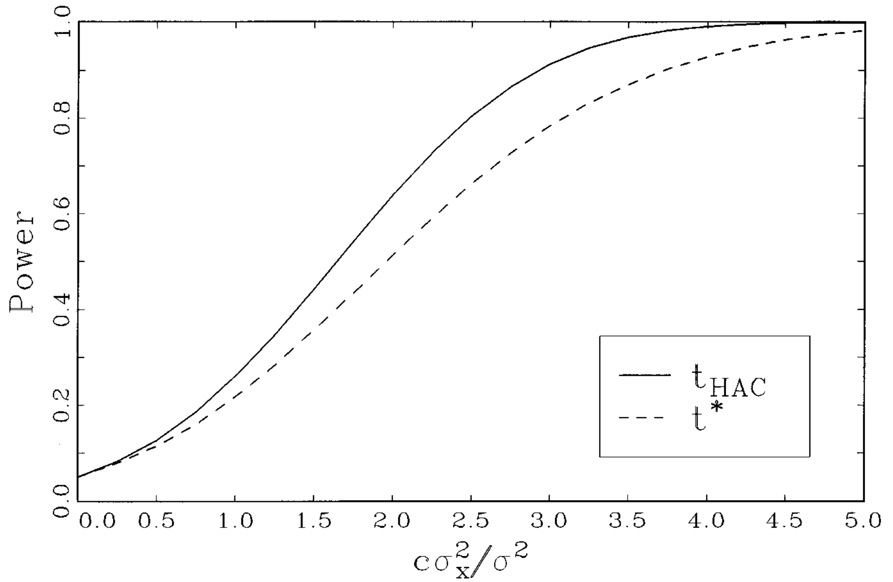

FIGURE 2.—Local asymptotic power, 5% nominal size, $k=1$, $y_t = \beta x_t + u_t$, H_0: $\beta \leqq \beta_0$, H_1: $\beta > \beta_0 + cT^{-1/2}$.

σ^2 based on $\hat{v}_t = x_t \hat{u}_t$ where $\{\hat{u}_t\}$ are the OLS residuals from (10). Let $\hat{S}_t = \Sigma_{j=1}^{t} \hat{v}_j$ and define $\hat{C} = T^{-2}\Sigma_{t=1}^{T}\hat{S}_t^2$.

Using this notation, the HAC estimator t test, t_{HAC}, and t^* can be calculated as

$$t_{HAC} = T^{1/2}(\hat{\beta} - B_0)/\left(\hat{\sigma}_x^{-2}\hat{\sigma}^2\hat{\sigma}_x^{-2}\right)^{1/2},$$

$$t^* = T^{1/2}(\hat{\beta} - \beta_0)/\left(\hat{\sigma}_x^{-2}\hat{C}\hat{\sigma}_x^{-2}\right)^{1/2}.$$

In the Appendix we show under the local alternative and Assumption A1 and A2, as $T \to \infty$,

$$\text{(11)} \qquad t_{HAC} \Rightarrow c\sigma_x^2/\sigma + W_1(1) \sim N(c\sigma_x^2/\sigma, 1),$$

$$\text{(12)} \qquad t^* \Rightarrow (c\sigma_x^2/\sigma + W_1(1))/\left[\int_0^1 (W_1(r) - rW_1(1))^2\, dr\right]^{1/2}.$$

Results (11) and (12) show that local asymptotic power of both statistics depends on $c\sigma_x^2/\sigma$. Naturally, as c increases, power increases. As σ_x^2 increases, power also increases, which follows from the standard regression result that more variability in the regressors leads to more efficient estimates. As σ^2 increases, power decreases, which follows since variability in $\{u_t\}$ is higher.

The local asymptotic distributions were used to compute asymptotic power, which is plotted in Figure 2. The power of t_{HAC} was computed analytically. The power of t^* was simulated using methods similar to those used to simulate the asymptotic critical values. Power was computed using the asymptotic 5% critical values.[7] The power of t^* is nontrivial and is comparable to, but slightly below, that of t_{HAC}. In finite samples power of the tests is likely to be closer since the asymptotic power of t_{HAC} does not reflect the finite sample variability in $\hat{\sigma}^2$. In fact, we give examples in Section 7 where power of the t^* (F^*) statistic exceeds the power of HAC estimator tests.

6. FINITE SAMPLE SIZE

In this section we report the results of an extensive simulation experiment with the purpose of comparing the finite sample size of HAC estimator tests and the F^* test. We designed the simulations so that they replicate the data generating processes (DGPs) and estimators used by Andrews (1991) and Andrews and Monahan (1992).

We consider a regression model with a constant and four stochastic regressors so that $k = 5$. We use the HAC estimator recommended by Andrews (1991) that uses the quadratic spectral kernel. The truncation lag (or bandwidth) was chosen using the automatic data-dependent procedure proposed by Andrews (1991) using the plug-in method based on univariate AR(1) models fit to the individual elements of $\hat{v}_t$. Tests based on this estimator are labeled *QS*. Consult Andrews (1991) for additional details. Following Andrews and Monahan (1992), we also computed HAC estimator tests using pre-whitening based on a VAR(1) parametric model of $\hat{v}_t$. We also employed the eigenvalue adjustment procedure used by Andrews and Monahan (1992) when fitting the VAR to $\hat{v}_t$. The pre-whitening tests are labeled *QS-PW*. Consult Andrews and Monahan (1992, p. 957) for additional details. Note that we are comparing our test with tests based on optimal HAC estimators.

We report results for six of the seven DGPs used by Andrews (1991) and Andrews and Monahan (1992). The models are: AR(1)−HOMO, where the errors and stochastic regressors are AR(1) homoskedastic processes; AR(1)−HET1 and AR(1)−HET2 where the DGPs are the same as the AR(1)−HOMO DGP except the error has multiplicative heteroskedasticity; MA(1)−HOMO, where the errors and stochastic regressors are MA(1) homoskedastic processes; and MA(1)−HET1 and MA(1)−HET2 where the DGPs are the same as the MA(1)−HOMO DGP except the error has multiplicative heteroskedasticity. In all cases, the regressors and errors were drawn independently of each other. In the AR(1) models, the stochastic regressors and errors were generated according to the model $\eta_t = \rho\eta_{t-1} + e_t$ where e_t is drawn from i.i.d. $N(0, 1-\rho^2)$ random variables which results in η_t having unit variance. The initial condition was drawn from the stationary distribution of the AR(1) model. In each replication a new set of regressors was randomly drawn. We transformed the regressor matrix so that $T^{-1}\sum_{t=1}^{T} X_t X_t'$ is an identity matrix following Andrews and Monahan (1992, p. 959). For the HET1 and HET2 models, the errors were first drawn from the AR(1) process and then multiplied by $|X_{2t}|$ and $|\frac{1}{2}\sum_{i=2}^{5} X_{it}|$ respectively. We report results for $\rho = -0.5, -0.3, 0.0, 0.3, 0.5, 0.7, 0.9, 0.95$.

The MA(1) models were generated in a similar fashion with the stochastic regressors and errors generated according to the model $\eta_t = e_t + \theta e_{t-1}$ where e_t is drawn from i.i.d. $N(0, (1+\theta^2)^{-1})$ random variables, which results in η_t having unit variance. We report results for $\theta = 0.3, 0.5, 0.7, 0.99$. In all cases we used 2,000 replications.

Following Andrews (1991) and Andrews and Monahan (1992) we computed type I error probabilities (they computed confidence interval coverage probabilities) for tests of the hypothesis H_0: $\beta_2 = 0$. We extend the results of Andrews (1991) and Andrews and

[7] We also computed asymptotic power for 1%, 2.5%, and 10% significance levels. The relative power of the tests is similar to that depicted in Figure 1 and is available upon request.

Monahan (1992) and also report results for tests of the hypotheses: H_0: $\beta_2 = \beta_3 = 0$, H_0: $\beta_2 = \beta_3 = \beta_4 = 0$, H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. We label the hypotheses according to the number of restrictions being tested, i.e. $q = 1, 2, 3, 4$. The results for the AR(1) models with a sample size of $T = 128$ are reported in Table III. Several patterns emerge from the table. First, in nearly every case, null rejection probabilities of F^* are less distorted and closer to 0.05 than the QS or QS−PW tests. The differences become larger as q increases. Although the F^* test has less distortions, there are many cases in which null rejection probabilities are much greater than 0.05. Nonetheless, the asymptotic approximation of the distribution of F^* is substantially better compared to QS and QS−PW. Second, as ρ approaches one, distortions of the null rejection probabilities increase for all the statistics. This is explained by the fact that the stationary asymptotic approximation becomes less accurate the closer the autoregressive root is to one. Third, for all three statistics, as q increases, null rejection probabilities also increase, indicating the asymptotic approximation is less precise when testing joint hypotheses compared to testing simple hypotheses. This result suggests, in particular, that for joint hypotheses, size distortions of HAC estimator tests can be substantial even when there is only modest serial correlation in the errors.

Results for the MA(1) models with $T = 128$ are given in Table IV. Similar patterns are seen as for the AR(1) models except that distortions overall are much less severe. Rejection probabilities of F^* are rarely above 0.10 while those of QS and QS−PW often exceed 0.10 especially for large q.

In Table V we report results for the AR(1)−HOMO model for sample sizes $T = 256, 512$. The table indicates that the asymptotic approximation improves substantially for all the tests as T increases. For the most part, F^* has rejection probabilities close to 0.05 for $\rho \leq 0.5$. For $\rho > 0.5$ rejection probabilities are inflated but by much less compared to when $T = 128$. Rejection probabilities of QS and QS−PW are, for the most part, more distorted than those of F^*, especially for $\rho \geq 0.9$ and $q \geq 3$.

7. FINITE SAMPLE POWER AND EMPIRICAL EXAMPLE

Using the DGPs from the previous section, we simulated size-adjusted power of the statistics and found that power rankings of the statistics followed patterns qualitatively similar to the local asymptotic power curve depicted in Figure 2. Therefore, we do not report those simulations here and instead report results on finite sample power from simulations based on the following empirical example. Let $\Delta lrev_t$ denote the first difference of the natural logarithm of real aggregate restaurant revenues for the United States, and let $\Delta lgdp_t$ denote the first difference of the natural logarithm of (seasonally adjusted) real gross domestic product (GDP) for the United States. We obtained quarterly observations from 1971:1 to 1996:4 for the nominal versions of these series and constructed the real series by dividing by the implicit GDP deflator. We seasonally adjusted the nominal restaurant revenue series before constructing the real series. The restaurant revenue series was obtained from the *Current Business Reports* published by the Bureau of the Census, and the nominal GDP and deflator series were obtained from the *Survey of Current Business* published by the Bureau of Economic Analysis, U.S. Department of Commerce. The levels of the real revenue and real GDP series are clearly trending over time and may have unit root errors. Therefore, the first differences of the series are likely to be stationary and satisfy Assumptions A1 and A2, so we consider a regression model in first differences of the data. For simplicity, we are ignoring the

TABLE III

FINITE SAMPLE NULL REJECTION PROBABILITIES AR(1) MODELS, $T = 128$
2,000 REPLICATIONS, NOMINAL LEVEL 0.05; ASYMPTOTIC CRITICAL VALUES USED

Model	ρ	F^*	QS	$QS{-}PW$	Model	ρ	F^*	QS	$QS{-}PW$
	−0.5	0.067	0.094	0.079		−0.5	0.082	0.118	0.097
	−0.3	0.058	0.073	0.067		−0.3	0.062	0.090	0.080
AR(1)-	0.0	0.059	0.064	0.068	AR(1)-	0.0	0.054	0.070	0.069
HOMO	0.3	0.073	0.078	0.075	HOMO	0.3	0.065	0.090	0.085
$q = 1$	0.5	0.083	0.103	0.089	$q = 2$	0.5	0.090	0.134	0.109
	0.7	0.099	0.143	0.107		0.7	0.128	0.207	0.147
	0.9	0.197	0.302	0.211		0.9	0.273	0.440	0.322
	0.95	0.307	0.439	0.314		0.95	0.409	0.611	0.448
	−0.5	0.096	0.147	0.127		−0.5	0.104	0.184	0.153
	−0.3	0.074	0.109	0.107		−0.3	0.077	0.124	0.121
AR(1)-	0.0	0.059	0.083	0.093	AR(1)-	0.0	0.074	0.089	0.109
HOMO	0.3	0.071	0.115	0.108	HOMO	0.3	0.089	0.127	0.128
$q = 3$	0.5	0.097	0.169	0.131	$q = 4$	0.5	0.114	0.199	0.162
	0.7	0.141	0.262	0.195		0.7	0.169	0.313	0.237
	0.9	0.344	0.567	0.429		0.9	0.388	0.651	0.515
	0.95	0.491	0.748	0.570		0.95	0.543	0.832	0.658
	−0.5	0.075	0.108	0.093		−0.5	0.085	0.127	0.117
	−0.3	0.070	0.084	0.080		−0.3	0.071	0.091	0.093
AR(1)-	0.0	0.063	0.069	0.068	AR(1)-	0.0	0.062	0.087	0.086
HET1	0.3	0.075	0.093	0.089	HET1	0.3	0.073	0.099	0.097
$q = 1$	0.5	0.088	0.119	0.099	$q = 2$	0.5	0.095	0.136	0.118
	0.7	0.114	0.166	0.128		0.7	0.125	0.211	0.167
	0.9	0.217	0.338	0.267		0.9	0.283	0.450	0.352
	0.95	0.326	0.439	0.348		0.95	0.395	0.579	0.464
	−0.5	0.087	0.153	0.134		−0.5	0.098	0.174	0.157
	−0.3	0.065	0.100	0.103		−0.3	0.075	0.112	0.120
AR(1)-	0.0	0.068	0.088	0.101	AR(1)-	0.0	0.067	0.096	0.117
HET1	0.3	0.086	0.120	0.120	HET1	0.3	0.091	0.128	0.136
$q = 3$	0.5	0.103	0.165	0.142	$q = 4$	0.5	0.117	0.187	0.168
	0.7	0.149	0.262	0.208		0.7	0.163	0.317	0.242
	0.9	0.328	0.535	0.423		0.9	0.361	0.616	0.490
	0.95	0.449	0.687	0.546		0.95	0.506	0.769	0.616
	−0.5	0.073	0.090	0.078		−0.5	0.086	0.121	0.099
	−0.3	0.064	0.077	0.069		−0.3	0.070	0.085	0.087
AR(1)-	0.0	0.056	0.068	0.073	AR(1)-	0.0	0.069	0.077	0.079
HET2	0.3	0.068	0.086	0.082	HET2	0.3	0.077	0.097	0.094
$q = 1$	0.5	0.085	0.108	0.096	$q = 2$	0.5	0.089	0.130	0.116
	0.7	0.100	0.151	0.122		0.7	0.119	0.203	0.163
	0.9	0.203	0.305	0.234		0.9	0.257	0.421	0.318
	0.95	0.303	0.416	0.318		0.95	0.372	0.557	0.443

TABLE III–*Continued*

Model	ρ	F^*	QS	$QS\text{-}PW$	Model	ρ	F^*	QS	$QS\text{-}PW$
	−0.5	0.083	0.142	0.124		−0.5	0.087	0.160	0.136
	−0.3	0.075	0.105	0.104		−0.3	0.079	0.115	0.117
AR(1)-	0.0	0.069	0.086	0.097	AR(1)-	0.0	0.077	0.090	0.110
HET2	0.3	0.079	0.114	0.114	HET2	0.3	0.083	0.123	0.131
$q = 3$	0.5	0.097	0.157	0.142	$q = 4$	0.5	0.101	0.196	0.170
	0.7	0.139	0.258	0.202		0.7	0.169	0.310	0.246
	0.9	0.311	0.529	0.406		0.9	0.351	0.610	0.489
	0.95	0.443	0.671	0.548		0.95	0.505	0.753	0.624

TABLE IV

FINITE SAMPLE NULL REJECTION PROBABILITIES MA(1) MODELS, $T = 128$
2,000 REPLICATIONS, NOMINAL LEVEL 0.05; ASYMPTOTIC CRITICAL VALUES USED

Model	θ	F^*	QS	$QS\text{-}PW$	Model	θ	F^*	QS	$QS\text{-}PW$
MA(1)-	0.3	0.072	0.074	0.071	MA(1)-	0.3	0.068	0.087	0.080
HOMO	0.5	0.073	0.084	0.074	HOMO	0.5	0.074	0.100	0.081
$q = 1$	0.7	0.073	0.089	0.073	$q = 2$	0.7	0.078	0.109	0.082
	0.99	0.073	0.090	0.073		0.99	0.078	0.112	0.083
MA(1)-	0.3	0.066	0.106	0.104	MA(1)-	0.3	0.083	0.117	0.117
HOMO	0.5	0.074	0.124	0.105	HOMO	0.5	0.090	0.135	0.124
$q = 3$	0.7	0.080	0.134	0.102	$q = 4$	0.7	0.093	0.160	0.124
	0.99	0.082	0.139	0.101		0.99	0.094	0.164	0.118
MA(1)-	0.3	0.068	0.087	0.082	MA(1)-	0.3	0.072	0.093	0.090
HET1	0.5	0.082	0.098	0.086	HET1	0.5	0.074	0.103	0.089
$q = 1$	0.7	0.083	0.102	0.084	$q = 2$	0.7	0.082	0.115	0.095
	0.99	0.080	0.104	0.084		0.99	0.080	0.122	0.097
MA(1)-	0.3	0.080	0.118	0.117	MA(1)-	0.3	0.086	0.112	0.122
HET1	0.5	0.085	0.132	0.119	HET1	0.5	0.093	0.139	0.130
$q = 3$	0.7	0.095	0.140	0.116	$q = 4$	0.7	0.095	0.151	0.130
	0.99	0.095	0.146	0.116		0.99	0.093	0.158	0.130
MA(1)	0.3	0.068	0.084	0.081	MA(1)-	0.3	0.073	0.088	0.089
HET2	0.5	0.077	0.096	0.084	HET2	0.5	0.072	0.103	0.091
$q = 1$	0.7	0.082	0.098	0.083	$q = 2$	0.7	0.077	0.110	0.088
	0.99	0.081	0.098	0.076		0.99	0.085	0.104	0.089
MA(1)-	0.3	0.078	0.102	0.105	MA(1)-	0.3	0.077	0.119	0.126
HET2	0.5	0.077	0.122	0.115	HET2	0.5	0.082	0.142	0.138
$q = 3$	0.7	0.082	0.133	0.109	$q = 4$	0.7	0.087	0.155	0.136
	0.99	0.086	0.132	0.106		0.99	0.097	0.156	0.125

TABLE V

FINITE SAMPLE NULL REJECTION PROBABILITIES AR(1)-HOMO MODEL, $T = 256, 512$
2,000 REPLICATIONS, NOMINAL LEVEL 0.05; ASYMPTOTIC CRITICAL VALUES USED

Model	ρ	F^*	QS	$QS\text{-}PW$	Model	ρ	F^*	QS	$QS\text{-}PW$
	−0.5	0.050	0.069	0.059		−0.5	0.062	0.089	0.071
	−0.3	0.051	0.068	0.061		−0.3	0.049	0.064	0.062
AR(1)-	0.0	0.044	0.057	0.058	AR(1)-	0.0	0.053	0.053	0.056
HOMO	0.3	0.052	0.066	0.061	HOMO	0.3	0.057	0.075	0.067
$q = 1$	0.5	0.054	0.081	0.064	$q = 2$	0.5	0.070	0.095	0.077
$T = 256$	0.7	0.067	0.101	0.078	$T = 256$	0.7	0.095	0.137	0.106
	0.9	0.123	0.191	0.141		0.9	0.170	0.289	0.197
	0.95	0.184	0.297	0.207		0.95	0.263	0.442	0.311
	−0.5	0.063	0.104	0.080		−0.5	0.072	0.120	0.095
	−0.3	0.048	0.075	0.071		−0.3	0.056	0.086	0.078
AR(1)-	0.0	0.052	0.060	0.064	AR(1)-	0.0	0.057	0.064	0.068
HOMO	0.3	0.066	0.082	0.079	HOMO	0.3	0.067	0.090	0.086
$q = 3$	0.5	0.073	0.101	0.096	$q = 4$	0.5	0.084	0.132	0.106
$T = 256$	0.7	0.100	0.173	0.125	$T = 256$	0.7	0.122	0.202	0.146
	0.9	0.213	0.386	0.263		0.9	0.250	0.477	0.342
	0.95	0.330	0.565	0.406		0.95	0.394	0.682	0.502
	−0.5	0.062	0.070	0.057		−0.5	0.059	0.080	0.070
	−0.3	0.058	0.060	0.053		−0.3	0.047	0.061	0.056
AR(1)-	0.0	0.055	0.054	0.056	AR(1)-	0.0	0.047	0.053	0.056
HOMO	0.3	0.057	0.063	0.057	HOMO	0.3	0.058	0.062	0.058
$q = 1$	0.5	0.047	0.067	0.062	$q = 2$	0.45	0.058	0.073	0.060
$T = 512$	0.7	0.064	0.081	0.064	$T = 512$	0.7	0.065	0.097	0.072
	0.9	0.092	0.125	0.084		0.9	0.105	0.170	0.118
	0.95	0.124	0.193	0.132		0.95	0.165	0.278	0.195
	−0.5	0.057	0.087	0.073		−0.5	0.057	0.091	0.073
	−0.3	0.049	0.070	0.067		−0.3	0.045	0.076	0.068
AR(1)-	0.0	0.045	0.061	0.060	AR(1)-	0.0	0.060	0.063	0.066
HOMO	0.3	0.053	0.066	0.060	HOMO	0.3	0.057	0.082	0.071
$q = 3$	0.5	0.050	0.077	0.065	$q = 4$	0.5	0.054	0.089	0.071
$T = 512$	0.7	0.068	0.104	0.077	$T = 512$	0.7	0.073	0.117	0.091
	0.9	0.120	0.225	0.156		0.9	0.142	0.268	0.191
	0.95	0.194	0.365	0.240		0.95	0.232	0.452	0.309

possibility that the levels of the series are cointegrated.

Consider the regression

$$\Delta lrev_t = \beta_1 + \beta_2 \, \Delta lgdp_t + u_t. \tag{13}$$

In the notation of Section 2, $\beta = (\beta_1, \beta_2)'$ and $X_t = (1, \Delta lgdp_t)'$. The parameter β_2 measures the change in the growth of real restaurant revenues with respect to a unit increase in the real growth rate of GDP. Thus, β_2 measures the sensitivity of real restaurant revenue growth to changes in real GDP growth. Since shocks to the restaurant sector are likely to have little or no effect on GDP, it is reasonable to think of $\Delta lgdp_t$ as an exogenous regressor. Therefore, OLS provides a consistent estimate of β_2.

We estimated (13) by OLS and obtained $\hat{\beta}_2 = 0.681$. Thus, an increase in the real growth rate of GDP by 1% results in a 0.681% increase in the real growth rate of restaurant revenues. To measure the sampling variability of $\hat{\beta}_2$, we constructed the following 95% confidence intervals: *QS*: (0.059, 1.302), *QS–PW*: (−0.070, 1.431), and F^*: (0.305, 1.056). Confidence intervals based on *QS* and *QS–PW* were computed as $\hat{\beta}_2 \pm 1.96(\hat{V}_{22}/T)^{1/2}$ where $\hat{V}_{22}$ is the second diagonal element of

$$\hat{V} = (T^{-1}\Sigma_{t=1}^{T} X_t X_t')^{-1} \hat{\Omega} (T^{-1}\Sigma_{t=1}^{T} X_t X_t')^{-1},$$

$\hat{\Omega}$ is the *QS* or *QS–PW* HAC estimator respectively of Ω, 1.96 is 97.5% critical value of a standard normal distribution, and $T = 103$. The confidence interval based on F^* was computed as $\hat{\beta}_2 \pm 6.811(\hat{B}_{22}/T)^{1/2}$ where $\hat{B}_{22}$ is the second diagonal element of

$$\hat{B} = (T^{-1}\Sigma_{t=1}^{T} X_t X_t')^{-1} \left(T^{-2}\Sigma_{t=1}^{T} \hat{S}_t \hat{S}_t'\right) (T^{-1}\Sigma_{t=1}^{T} X_t X_t')^{-1}$$

and 6.811 is the 97.5% asymptotic critical value taken from Table I. Interestingly, the tightest confidence interval is obtained using F^*, and there are nontrivial differences in the HAC based confidence intervals whether or not prewhitening is used. This empirical example suggests a situation where power of the F^* statistic may be greater than power of HAC estimator tests and illustrates the sensitivity of inference to the way HAC estimators are constructed (see Den Haan and Levin (1997) for additional evidence on the latter).

To investigate the possibility that F^* is more powerful in the empirical example, we conducted the following simulation experiment. We fit a variety of ARMA models to the OLS residuals from (13) and found that an AR(4) model provided a good fit. We also fit a variety of ARMA models to $\Delta lgdp_t$ and found that an AR(1) and an ARMA(4, 1) model provided good fits. We performed three power simulations using 2,000 replications and $T = 103$. We generated data according to the model $y_t = \beta_2 x_t + u_t$ with $u_t = -0.3429u_{t-1} - 0.3301u_{t-2} - 0.2686u_{t-3} + 0.5947u_{t-4} + \epsilon_t$, $\epsilon_t \sim$ iid $N(0, 0.0197)$. We generated x_t using three DGPs: DGP(1): x_t equal to the actual first differenced quarterly real GDP data ($x_t = \Delta lgdp_t$), DGP(2): $x_t = 0.3249x_{t-1} + \xi_t$, $\xi_t \sim$ iid $N(0, 0.0089)$, and DGP(3): $x_t = 0.9952x_{t-1} - 0.1446x_{t-2} + 0.0411x_{t-3} - 0.1465x_{t-4} + \xi_t - 0.7410\xi_{t-1}$, $\xi \sim$ iid $N(0, 0.0089)$. The null hypotheses was H_0: $\beta_2 = 0$.

We computed finite sample null rejection probabilities of the statistics using 5% asymptotic critical values, which we report in Table VI. Regardless of the DGP used for x_t, rejection probabilities of all the statistics are below 0.05 with those of *QS* and *QS–PW* below that of F^*. Rejection probabilities of *QS* and *QS–PW* are quite low when the actual GDP data are used (DGP(1)) which explains the wide confidence intervals using *QS* and *QS–PW* in the empirical example. Because the tests are all conservative, it

TABLE VI

FINITE SAMPLE SIZE BASED ON THE EMPIRICAL EXAMPLE.
$T = 103$, 2,000 REPLICATIONS, NOMINAL LEVEL 0.05; ASYMPTOTIC CRITICAL VALUES USED

	β_2	F^*	*QS*	*QS-PW*
DGP(1)	0.00	0.026	0.005	0.002
DGP(2)	0.00	0.044	0.038	0.030
DGP(3)	0.00	0.044	0.030	0.023

Note: The regression is $y_t = \beta_1 + \beta_2 x_t + u_t$ with H_0: $\beta_2 = 0$. The DGP is $y_t = \beta_2 x_t + u_t$, $u_t = -0.3429u_{t-1} - 0.3301u_{t-2} - 0.2686u_{t-3} + 0.5947u_{t-4} + \epsilon_t$, $\epsilon_t \sim$ iid $N(0, 0.0197)$. For DGP(1), x_t is first differenced quarterly real GDP. For DGP(2), $x_t = 0.3249x_{t-1} + \xi_t$, $\xi_t \sim$ iid $N(0, 0.0089)$. For DGP(3), $x_t = 0.9952x_{t-1} - 0.1446x_{t-2} + 0.0411x_{t-3} - 0.1465x_{t-4} + \xi_t - 0.7410\xi_{t-1}$, $\xi_t \sim$ iid $N(0, 0.0089)$.

makes sense to compare power functions computed using the asymptotic critical values (this mimics the way the tests are used in practice). We simulated power for $\beta_2 = 0.1, 0.2, \ldots, 1.9, 2.0$. The resulting finite sample power curves are plotted in Figures 3, 4 and 5 corresponding to the three DGPs for x_t. In Figure 3 we see that F^* dominates the HAC estimator tests in terms of power when actual GDP data are used for x_t. This is not an atypical example as real GDP is commonly used in empirical work. In the other cases where x_t is modeled as an ARMA process, the power ranking depends on how far β_2 is from zero.

We conclude this section with an example that illustrates the sensitivity of HAC estimator tests to projections of subsets of regressors in OLS regressions. Using the same data set as above, we regressed the level of nominal aggregate restaurant revenue on a constant, a time trend, and the level of nominal GDP and obtained the following 95% confidence intervals for the estimate of the coefficient on nominal GDP: *QS*: (1.077, 1.321), *QS–PW*: (1.040, 1.358), F^*: (1.031, 1.366). (Because the nominal series almost certainly have unit root errors, this example does not satisfy Assumptions A1 and A2. It is illustrative nonetheless.) We also detrended the data and regressed the detrended revenue series on the detrended constant and detrended GDP (we projected out the time trend). The OLS estimate of the GDP coefficient and F^* are invariant to the method of estimation by the FWL Theorem. Therefore, confidence intervals based on F^* are the same in the two cases. Confidence intervals based on *QS* and *QS–PW* are not invariant, which is illustrated by confidence intervals based on the detrended regression: *QS*: (1.069, 1.329), *QS–PW*: (0.795, 1.603). This lack of invariance arises from using an automatic bandwidth and/or pre-whitening and illustrates a pitfall when using HAC estimator tests.

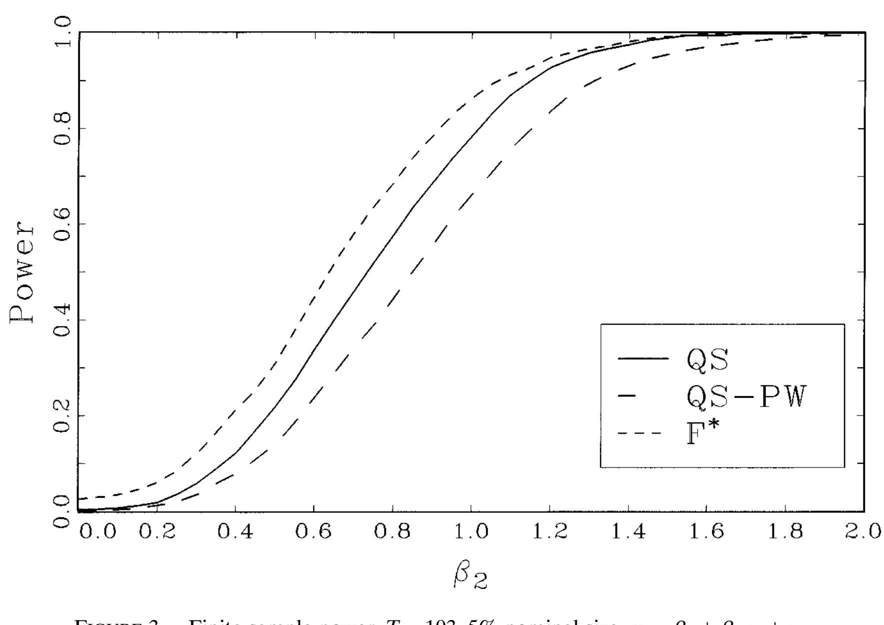

FIGURE 3.—Finite sample power, $T = 103$, 5% nominal size. $y_t = \beta_1 + \beta_2 x_t + u_t$, H_0: $\beta_2 = 0$, $u_t \sim$ AR(4).

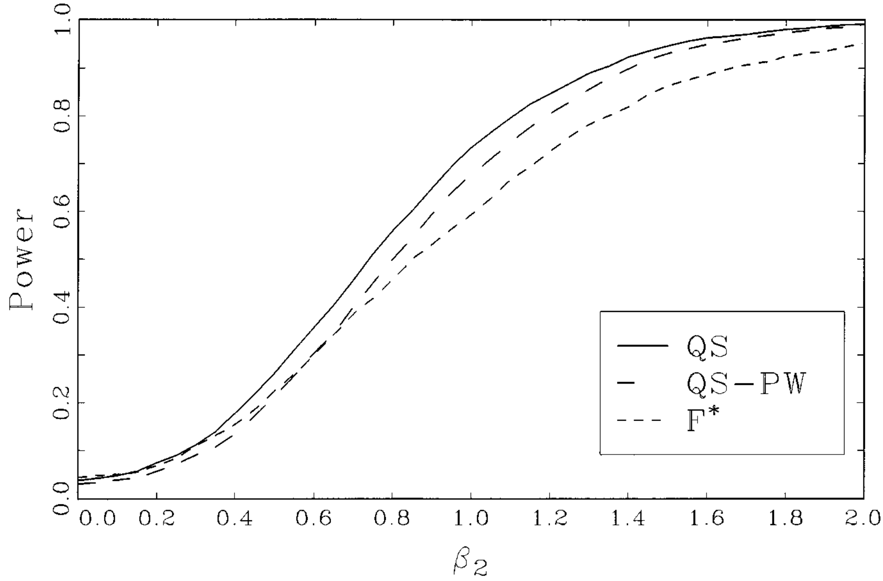

FIGURE 4.—Finite sample power, $T = 103$, 5% nominal size. $y_t = \beta_1 + \beta_2 x_t + u_t$, H_0: $\beta_2 = 0$, $u_t \sim$ AR(4), $x_t \sim$ AR(1).

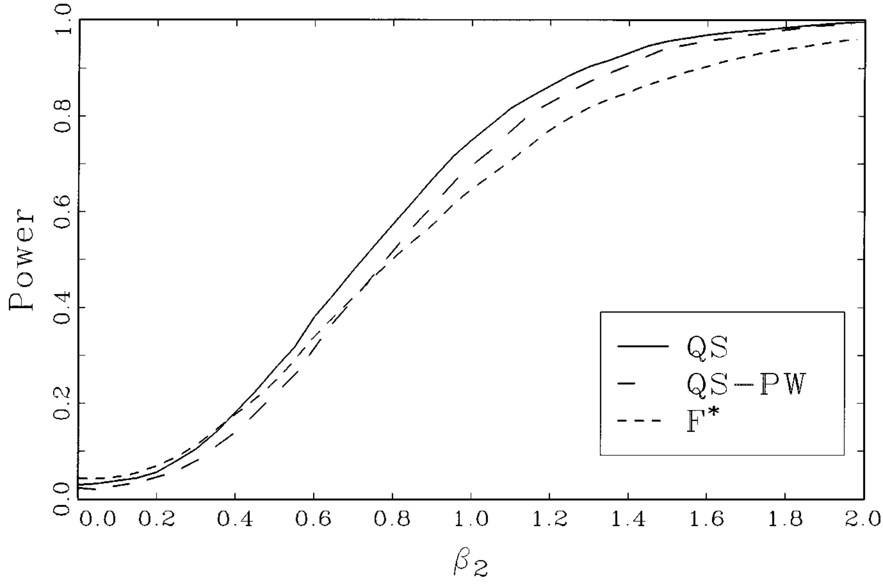

FIGURE 5.—Finite sample power, $T = 103$, 5% nominal size. $y_t = \beta_1 + \beta_2 x_t + u_t$, H_0: $\beta_2 = 0$, $u_t \sim$ AR(4), $x_t \sim$ ARMA(4, 1).

8. CONCLUSIONS

In this paper we propose new test statistics for testing hypotheses in regression models with serial correlation/heteroskedasticity of unknown form. The novel aspect of the new tests is that they are simple to compute and do not require spectral density (HAC) estimators. Our approach is to eliminate nuisance parameters asymptotically with a simple stochastic transformation of the parameter estimates. Since there are many conceivable transformations that will yield asymptotic pivotal statistics, our approach creates a new class of test statistics that are pivotal and robust to heteroskedasticity and serial correlation in the errors. An open research problem is to develop a theory of optimality for this new class of tests. We derive the limiting null distributions of the new tests and show that while they have nonstandard distributions, the distributions only depend on the number of restrictions being tested and critical values are easily simulated. Our results easily extend to GLS and IV estimation, and we conjecture that our approach can be extended to the GMM framework. A simulation experiment showed that the asymptotic approximation of the new test is better (nearly uniformly) than that of more standard HAC estimator tests. But, like HAC estimator tests, the new tests suffer from serious size distortions (although less so) if the data have highly persistent serial correlation and are close to being nonstationary. This is a common problem in time series models when the true form of serial correlation is unknown. Finally, the new tests retain respectable power, and we provide a relevant empirical example where finite sample power of our test dominates finite sample power of HAC estimator tests. Given that new tests compare favorably to HAC methods in finite samples and are simpler to compute, they should become serious competitors to HAC estimator tests in practice.

Dept. of Economics, Uris Hall, Cornell University, Ithaca, NY 14853, U.S.A.,
Dept. of Economics, Uris Hall, Cornell University, Ithaca, NY 14853, U.S.A.; tjv2@cornell.edu
and
Dept. of Economics, 260 Heady Hall, Iowa State University, Ames, IA 50011, U.S.A.

Manuscript received February, 1998; final revision received June, 1999.

APPENDIX

PROOF OF THEOREM 1: Under the null hypothesis and Assumptions A1 and A2 it follows from (2) and (5) that

$$F^* = T(R(\hat\beta-\beta))'[R\hat{B}R']^{-1}(R(\hat\beta-\beta))/q$$
$$= (RT^{1/2}(\hat\beta-\beta))'[R\hat{B}R']^{-1}(RT^{1/2}(\hat\beta-\beta))/q$$
$$\Rightarrow (RQ^{-1}\Lambda W_k(1))'[RQ^{-1}\Lambda P_k \Lambda' Q^{-1}R']^{-1} RQ^{-1}\Lambda W_k(1)/q.$$

Because the matrix $RQ^{-1}\Lambda$ has rank q and $W_k(r)$ is a vector of independent Wiener processes and is Gaussian, $RQ^{-1}\Lambda W_k(r)$ can be written as $\Lambda^* W_q^*(r)$ where $W_q^*(r)$ is a $(q\times 1)$ vector of independent Wiener processes, and Λ^* is the $(q\times q)$ matrix square root of $RQ^{-1}\Lambda\Lambda' Q^{-1}R'$. Λ^* exists and is invertible because $RQ^{-1}\Lambda\Lambda' Q^{-1}R'$ is symmetric and full rank. Therefore, $RQ^{-1}\Lambda W_k(1)$ is equivalent in distribution to $\Lambda^* W_q^*(r)$. In addition $RQ^{-1}\Lambda P_k \Lambda' Q^{-1}R'$ is equivalent in distribution to $\Lambda^* P_q^* \Lambda^{*\prime}$

where

$$P_q^* = \int_0^1 (W_q^*(r) - rW_q^*(1))(W_q^*(r) - rW_q^*(1))'\,dr.$$

This follows because

$$\begin{aligned} &RQ^{-1}\Lambda P_k \Lambda' Q^{-1}R' \\ &\quad = RQ^{-1}\Lambda\left\{\int_0^1 [W_k(r) - rW_k(1)][W_k(r) - rW_k(1)]'\,dr\right\}\Lambda' Q^{-1}R' \\ &\quad = \int_0^1 (RQ^{-1}\Lambda W_k(r) - rRQ^{-1}\Lambda W_k(1))(RQ^{-1}\Lambda W_k(r) - rRQ^{-1}\Lambda W_k(1))'\,dr, \end{aligned}$$

which is equivalent in distribution to

$$\int_0^1 (\Lambda^* W_q^*(r) - r\Lambda^* W_q^*(1))(\Lambda^* W_q^*(r) - r\Lambda^* W_q^*(1))'\,dr = \Lambda^* P_q^* \Lambda^{*\prime}.$$

Thus, $(RQ^{-1}\Lambda W_k(1))'[RQ^{-1}\Lambda P_k \Lambda' Q^{-1}R']^{-1}RQ^{-1}\Lambda W_k(1)/q$ is equivalent in distribution to

$$(\Lambda^* W_q^*(1))'[\Lambda^* P_q^* \Lambda^{*\prime}]^{-1}\Lambda^* W_q^*(1)/q = W_q^*(1)' P_q^{*-1} W_q^*(1)/q. \qquad Q.E.D.$$

PROOF THAT F^* SATISFIES THE FRISCH–WAUGH–LOVELL THEOREM: This proof is simplified by writing the model and F^* in matrix notation. Write regression (1) as $Y = X\beta + u$. Let G denote a $(T \times T)$ lower triangular matrix with elements along the diagonal and below all equal to one. Let $\hat{U}$ denote a $(T \times T)$ diagonal matrix with diagonal elements $(\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_T)$. Define $\hat{S} = G\hat{U}X$. Simple algebra gives $\sum_{t=1}^T \hat{S}_t \hat{S}_t' = \hat{S}'\hat{S} = X'\hat{U}G'G\hat{U}X = X'HX$ where $H = \hat{U}G'G\hat{U}$ so that $\hat{C} = T^{-2}X'HX$. Without loss of generality, partition X into X_1 and X_2 where X_1 contains the first k_1 columns of X, and X_2 contains the last $k - k_1$ columns of X. Partition β into β_1 and β_2 where β_1 is a $(k_1 \times 1)$ vector containing the first k_1 elements of β, and β_2 is a $((k - k_1) \times 1)$ vector containing the last $k - k_1$ elements of β. To show that F^* satisfies the FWL Theorem we must show that F^* for testing $\beta_1 = 0$ in regression (1) is computationally the same as F^* for testing $\beta_1 = 0$ in the regression

$$\text{(A1)} \qquad \tilde{Y} = \tilde{X}_1 \beta_1 + \tilde{u},$$

where $\tilde{Y} = M_2 Y$, $\tilde{X}_1 = M_2 X_1$, $\tilde{u} = M_2 u$, and $M_2 = I_T - X_2(X_2'X_2)^{-1}X_2'$. Let $\tilde{\beta}_1$ denote the OLS estimate of β_1 from regression (A1), and let $R = [I_{k_1}\ 0_{k-k_1}]$. By the FWL Theorem the OLS residuals from regression (1) and (A1) are the same. Therefore the H matrix in regression (A1) is the same as in regression (1). Thus, the F^* statistics from regressions (1) and (A1) can be written respectively as

$$\text{(A2)} \qquad T(R\hat{\beta})'[R(X'X)^{-1}X'HX(X'X)^{-1}R']^{-1}(R\hat{\beta})/q,$$

$$\text{(A3)} \qquad T\tilde{\beta}_1'\left[(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'H\tilde{X}_1(\tilde{X}_1'\tilde{X}_1)^{-1}\right]^{-1}\tilde{\beta}_1/q.$$

By the FWL Theorem $R\hat{\beta} = \tilde{\beta}_1$; therefore, (A2) and (A3) are computationally equivalent if $R(X'X)^{-1}X'HX(X'X)^{-1}R' = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'HX_1(\tilde{X}_1'\tilde{X}_1)^{-1}$. It is sufficient to show that $R(X'X)^{-1}X' = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'$. From the partitioned matrix formula it follows that

$$\begin{aligned} R(X'X)^{-1}X' &= \left[(\tilde{X}_1'\tilde{X}_1)^{-1}, -(\tilde{X}_1'\tilde{X}_1)^{-1}X_1'X_2(X_2'X_2)^{-1}\right][X_1, X_2]' \\ &= (\tilde{X}_1'\tilde{X}_1)^{-1}X_1' - (\tilde{X}_1'\tilde{X}_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2' \\ &= (\tilde{X}_1'\tilde{X}_1)^{-1}X_1'(I_T - X_2(X_2'X_2)^{-1}X_2') \\ &= (\tilde{X}_1'\tilde{X}_1)^{-1}X_1'M_2 = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1' \end{aligned}$$

which completes the proof. *Q.E.D.*

PROOF OF (11) AND (12): The denominators of t_{HAC} and t^* are invariant to c since they are functions of $\hat{u}_t$, which is invariant to c by construction. It directly follows that

$$\text{plim}(\hat{\sigma}_x^{-2}\hat{\sigma}^2\hat{\sigma}_x^{-2})^{1/2} = (\sigma_x^{-2}\sigma^2\sigma_x^{-2})^{1/2} = \sigma/\sigma_x^2,$$

and using (5) with simplifying algebra we have

$$(\hat{\sigma}_x^{-2}\hat{C}\hat{\sigma}_x^{-2})^{1/2} \Rightarrow (\sigma/\sigma_x^2)\left[\int_0^1 (W_1(r) - rW_1(1))^2\, dr\right]^{1/2}.$$

Given these results, all that is needed to prove (11) and (12) is the limit of $T^{1/2}(\hat{\beta} - \beta_0)$ under the local alternative. Using Assumptions A1 and A2 we have

$$T^{1/2}(\hat{\beta} - \beta_0) = c + \left(T^{-1}\sum_{t=1}^{T} x_t^2\right)^{-1} T^{-1/2}\sum_{t=1}^{T} x_t u_t \Rightarrow c + \sigma W_1(1)/\sigma_x^2.$$

Simple algebra completes the proof. *Q.E.D.*

REFERENCES

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.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. New York: Oxford University Press.

DEN HAAN, W. J., AND A. LEVIN (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," Chapter 12 in *Handbook of Statistics: Robust Inference*, Volume 15, ed. by G. S. Maddala and C. R. Rao. New York: Elsevier.

GALLANT, A. R. (1987): *Nonlinear Statistical Models*. New York: Wiley.

HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press.

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

MACKINNON, J. G. (1999): "Computing Numerical Distribution Functions in Econometrics," Mimeo, Department of Economics, Queen's University, forthcoming in *High Performance Computing Systems and Applications*, ed. by A. Pollard, D. Mewhort, and D. Weaver. Amsterdam: Kluwer.

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

PHILLIPS, P. C. B., AND S. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–496.

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

——— (1998): "Inference–Without Smoothing in the Presence of Nonparametric Autocorrelation," *Econometrica*, 66, 1163–1182.

VOGELSANG, T. J. (1998): "Trend Function Hypothesis Testing in the Presence of Serial Correlation," *Econometrica*, 66, 123–148.

WHITE, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.