A NEW APPROACH TO ROBUST INFERENCE IN COINTEGRATION

By

Sainan Jin, Peter C.B. Phillips and Yixiao Sun

October 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1538

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

A New Approach to Robust Inference in Cointegration (Full Version)

Sainan Jin*
Guanghua School of Management,
Peking University

Peter C. B. Phillips†
Cowles Foundation, Yale University,
University of Auckland & University of York

Yixiao Sun
Department of Economics
University of California, San Diego

October 2005

*Corresponding author. Guanghua School of Management, Peking University, Beijing, 100871, PR. China. Tel: 86-10-62756274, fax: 86-10-62753820, email: sainan.jin@gsm.pku.edu.cn

†Partial support is acknowledged from a Kelly Fellowship at the University of Auckland School of Business and the NSF under Grant No. SES 04-142254

ABSTRACT

A new approach to robust testing in cointegrated systems is proposed using nonparametric HAC estimators without truncation. While such HAC estimates are inconsistent, they still produce asymptotically pivotal tests and, as in conventional regression settings, can improve testing and inference. The present contribution makes use of steep origin kernels which are obtained by exponentiating traditional quadratic kernels. Simulations indicate that tests based on these methods have improved size properties relative to conventional tests and better power properties than other tests that use Bartlett or other traditional kernels with no truncation.

JEL Classification: C12; C14; C22

Keywords: Cointegration, HAC estimation, long-run covariance matrix, robust inference, steep origin kernel, fully modified estimation.

1 Introduction

Recent research by Kiefer and Vogelsang (2002a, 2002b, 2005, hereafter KV) and the present authors (2005a, 2005b, hereafter PSJ_a, PSJ_b) has shown that there are certain advantages to constructing robust tests in regression models with inconsistent rather than consistent estimates of the relevant variance matrices. Such 'inconsistent HAC' estimates can be based on standard kernels with the bandwidth parameter set to the sample size (or possibly some fixed proportion of the sample size) or on power kernels which exponentiate traditional kernels. Such variance estimates are inconsistent and tend to random matrices instead of the true variance matrix, but this randomness in the limit succeeds in bringing the finite sample distribution of test statistics based on them closer to the limit distribution, thereby improving the finite sample size properties of the resulting tests. Some formal results of this type have been proved recently by Jansson (2004) and the present authors (2005, hereafter PSJ_c). Using this approach, asymptotically similar tests can be constructed for which the limit distribution is nonstandard but easily computed and easily approximated by series expansions (PSJ_c).

The contribution of the present paper is to apply these methods in the context of cointegrated regression equations. Making use of steep origin kernels, as in PSJ_b, we provide a limit theory for the corresponding cointegrating regression test statistics based on these inconsistent HAC estimates. Simulations show that these alternative robust tests have greater accuracy in size, but also experience some loss of power, including local asymptotic power, in relation to conventional tests. In this sense, the results mirror the earlier findings for regression models in KV and PSJ_b. Steep origin kernels seem to have the best properties in the class of tests considered here.

The plan of the rest of paper is as follows. The test statistics, limit distributions and local asymptotic powers are given in Section 2. Finite sample simulations are presented in Section 3. Some discussion and concluding remarks are given in Section 4. Proofs are given in Section 5.

2 Tests and Limit Theory

We consider the cointegrated regression model

$$y_t = \alpha + x_t'\beta + u_{0t}, \tag{1}$$

where x_t is an m-dimensional vector of full rank integrated regressors generated by

$$\Delta x_t = u_{xt}, \tag{2}$$

where $t = 1, ..., T$. The error vector $u_t = (u_{0t}, u_{xt}')'$ is jointly stationary with zero mean and long run covariance matrix $\Omega > 0$.

The following high level condition for which sufficient conditions are well known (e.g., Phillips and Solo, 1992) facilitates the asymptotic development:

$$T^{-1/2}\sum_{t=1}^{[Tr]} u_t \to_d B(r) = \left[\begin{array}{c} B_0(r) \\ B_x(r) \end{array}\right] \equiv BM(\Omega),\ \ r \in [0,1]\,, \tag{3}$$

with

$$\Omega = \begin{bmatrix} \omega_0^2 & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}, \tag{4}$$

where $B_0(\cdot)$ and $B_x(\cdot)$ are Brownian motions corresponding to u_{0t} and u_{xt} respectively, and where the partition of Ω is conformable with that of u_t. Defining $u_{0.xt} = u_{0t} - \Omega_{0x}\Omega_{xx}^{-1}u_{xt}$, it follows that $T^{-1/2}\sum_{t=1}^{[Tr]} u_{0.xt} \to B_{0.x}(r)$, and

$$T^{-1/2}\begin{pmatrix} \sum_{t=1}^{[Tr]} u_{0.xt} \\ \sum_{t=1}^{[Tr]} u_{xt} \end{pmatrix} \to \begin{bmatrix} B_{0.x}(r) \\ B_x(r) \end{bmatrix} \equiv BM\begin{bmatrix} \omega_{0.x}^2 & 0 \\ 0 & \Omega_{xx} \end{bmatrix}, \tag{5}$$

where $B_{0.x}(r) = B_0(r) - \Omega_{0x}\Omega_{xx}^{-1}B_x(r) = BM(\omega_{0.x}^2) = \omega_{0.x}W_0$, and is independent of $B_x = \Omega_{xx}^{1/2}W_x$, with $\omega_{0.x}^2 = \omega_0^2 - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$. W_0 and W_x are independent standard Brownian motions of dimensions 1 and m, respectively.

Let $\widehat{\beta}^+$ be the FM-OLS estimator of β in (1), $M_{xx} = \sum_{t=1}^{T}(x_t - \overline{x})(x_t - \overline{x})')^{-1}$, where $\overline{x} = T^{-1}\sum_{t=1}^{T} x_t$, and $\widehat{\omega}_{0.x}^{*2}$ be constructed by introducing kernel based estimators of $\omega_{0.x}$ where the bandwidth parameter is set equal to the sample size:

$$\widehat{\omega}_{0.x}^{*2} = \sum_{h=-T+1}^{T-1} k\left(\frac{h}{T}\right)\widehat{\Gamma}(h), \tag{6}$$

with

$$\widehat{\Gamma}(h) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} \widehat{u}_{0.xt+h}\widehat{u}_{0.xt}' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \widehat{u}_{0.xt+h}\widehat{u}_{0.xt}' & \text{for } j < 0 \end{cases} \tag{7}$$

and $\widehat{u}_{0.xt}$ is the FM-OLS regression residual. The class of kernels we consider contains most conventional kernels $k(x)$, as well as steep origin kernels, $k_\rho(x) = (k_{quad}(x))^\rho$, which were introduced in PSJ$_b$ and involve arbitrary powers $\rho > 1$ of a conventional quadratic kernel $k_{quad}(x)$ with Parzen characteristic exponent 2. For the steep origin kernels, we focus on the steep Parzen kernel (i.e. $k_{quad}(x)$ is the traditional Parzen kernel) throughout the paper. The results for other steep origin kernels are similar and will not be reported here.

The cointegrating regression F-statistic for the q-dimensional hypothesis $R\beta = r$ takes the usual form

$$F^* = (R\widehat{\beta}^+ - r)'\left[RM_{xx}R'\right]^{-1}(R\widehat{\beta}^+ - r)/\widehat{\omega}_{0.x}^{*2}. \tag{8}$$

or when $q = 1$, the t-ratio

$$t^* = \left[RM_{xx}R'\right]^{-1/2}(R\widehat{\beta}^+ - r)/\widehat{\omega}_{0.x}^{*}. \tag{9}$$

Similar arguments to those given in Theorems 1 and 5 of PSJ$_a$ lead to the following results. First, the conditional long run variance estimator $\widehat{\omega}_{0.x}^{*2}$ has the random limit

$$\widehat{\omega}_{0.x}^{*2} \Rightarrow \omega_{0.x}^2\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s), \tag{10}$$

where

$$\eta(r) = W_0(r) - rW_0(1) - \int_0^r \underline{W}_x'(\int_0^1 \underline{W}_x\underline{W}_x')^{-1}\int_0^1 \underline{W}_x dW_0, \tag{11}$$

with $\underline{W}_x = W_x - \int_0^1 W_x$. Second, the limit of the regression F-statistic and t-statistic under the null is

$$F^* \Rightarrow \frac{\left(\int_0^1 \underline{W}_{1.2}dW_0\right)'\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1}\left(\int_0^1 \underline{W}_{1.2}dW_0\right)}{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)}, \tag{12}$$

$$t^* \Rightarrow \frac{\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1/2}\left(\int_0^1 \underline{W}_{1.2}dW_0\right)}{\sqrt{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)}}, \tag{13}$$

where $\underline{W}_{1.2} = \underline{W}_1 - \int_0^1 \underline{W}_1\underline{W}_2'(\int_0^1 \underline{W}_2\underline{W}_2')^{-1}\underline{W}_2$ is the L_2 regression residual of $\underline{W}_1$ on $\underline{W}_2$, $\underline{W}_1$ is the vector of the first q coordinates of $\underline{W}_x$, and $\underline{W}_2$ is the last $(m-q)$ coordinates. Third, under the local alternative $R\beta = r + c/T$, the F and t statistics have the following limit distributions

$$F^* \Rightarrow \frac{\left[\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1}\int_0^1 \underline{W}_{1.2}dW_0 + \delta\right]'\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)\left[\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1}\int_0^1 \underline{W}_{1.2}dW_0 + \delta\right]}{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)},$$

$$t^* \Rightarrow \frac{\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1/2}\int_0^1 \underline{W}_{1.2}dW_0 + \delta\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{1/2}}{\sqrt{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)}},$$

where $\delta = \left(R\Omega_{xx}^{-1}R'\right)^{-1/2}c/\omega_{0.x}$.

Remarks

(a) While $\widehat{\omega}_{0.x}^{*2}$ is inconsistent for $\omega_{0.x}^2$, its limit distribution is scale dependent on $\omega_{0.x}^2$, which explains why it is possible to construct asymptotically pivotal tests using $\widehat{\omega}_{0.x}^{*2}$.

(b) The limit distributions of the F-statistic and t-statistic are nonstandard in view of the random limit of the inconsistent HAC estimates.

(c) Under the null, the numerator in the limit distribution follows χ_q^2 or standard normal distribution. As shown in the proof, the numerator is independent of the denominator in the limit distribution. These are the same as in the stationary case. However, the $\eta(r)$ process is not a Brownian bridge process and depends on the number of regressors in the model, which is different from the stationary case.

(d) When the steep origin kernel of the form $k(x) = (k_{quad}(x))^\rho$ is used in the tests, consistency of $\widehat{\omega}^{*2}_{0.x}$ requires the following rate condition

$$T^6/\rho^5 + \rho/T^2 \to 0 \text{ as } T \to \infty, \text{ and } \rho \to \infty \tag{14}$$

(see PSJ$_b$ for details), in which case the limit distributions for F^* and t^* (denoted as F^*_ρ and t^*_ρ to emphasize their dependence on ρ) under the null are $F^*_\rho \Rightarrow \chi^2_q$, and $t^*_\rho \Rightarrow N(0,1)$. The corresponding limit distributions under the local alternative are

$$\begin{aligned} F^*_\rho \;\Rightarrow\; & \left[\left(\int_0^1 \underline{W}_{1.2}\underline{W}'_{1.2}\right)^{-1}\left(\int_0^1 \underline{W}_{1.2}dW_0\right) + \delta\right]'\left(\int_0^1 \underline{W}_{1.2}\underline{W}'_{1.2}\right) \\ & \times\left[\left(\int_0^1 \underline{W}_{1.2}\underline{W}'_{1.2}\right)^{-1}\left(\int_0^1 \underline{W}_{1.2}dW_0\right) + \delta\right], \end{aligned} \tag{15}$$

$$t^*_\rho \Rightarrow \left(\int_0^1 \underline{W}_{1.2}\underline{W}'_{1.2}\right)^{-1/2}\int_0^1 \underline{W}_{1.2}dW_0 + \delta\left(\int_0^1 \underline{W}_{1.2}\underline{W}'_{1.2}\right)^{1/2}. \tag{16}$$

Asymptotic critical values are given in Table 1 for the t^* test for a selection of well known kernels, such as Bartlett, Parzen, Tukey-Hanning, QS, Normal as well as Steep origin kernels (SO) with $\rho = 8, 32$, and the asymptotic case which is represented by "$\rho = \infty$" with $m = 1$. The Brownian motion and detrended Brownian motion paths are calculated using normalized partial sums of $T = 1000$ $iidN(0,1)$ random variables. 50,000 replications are employed.

Table 1: Asymptotic Critical Value of t* Test (m=1)

Kernel	90.0%	95.0%	97.5%	99.0%
Bartlett	3.2299	4.3825	5.4780	6.9642
Parzen	3.5666	5.2742	7.1273	10.1487
Tukey-Hanning	5.2969	8.4123	12.8590	20.7183
QS	7.0547	11.2629	16.9171	27.4247
Normal	3.8622	5.8006	7.9304	11.5824
SO ($\rho = 8$)	1.8871	2.5549	3.2348	4.1203
SO ($\rho = 32$)	1.5537	2.0514	2.5121	3.1046
SO ($\rho = \infty$)	1.2820	1.6450	1.9600	2.3260

Following PSJ$_a$ and PSJ$_b$, we perform local asymptotic power simulations using the t^*_ρ-statistic as the benchmark. We compute asymptotic power for t^* tests with a 5% significance level using the kernels from Table 1 for $\delta \in [0, 9.3]$ with $m = 1$. Results for $m > 1$ are similar and are not presented here. As is apparent in Fig. 1, and similar to KV(2002a), the Bartlett kernel delivers higher power than other well-known kernels, including Parzen, Normal, Tukey-Hanning, and QS. However, steep Parzen kernels produce even better power properties than the Bartlett kernel, and the power curve moves up uniformly as ρ increases. Notice that when $\rho = 32$, the power curve is very close to the power envelope. This

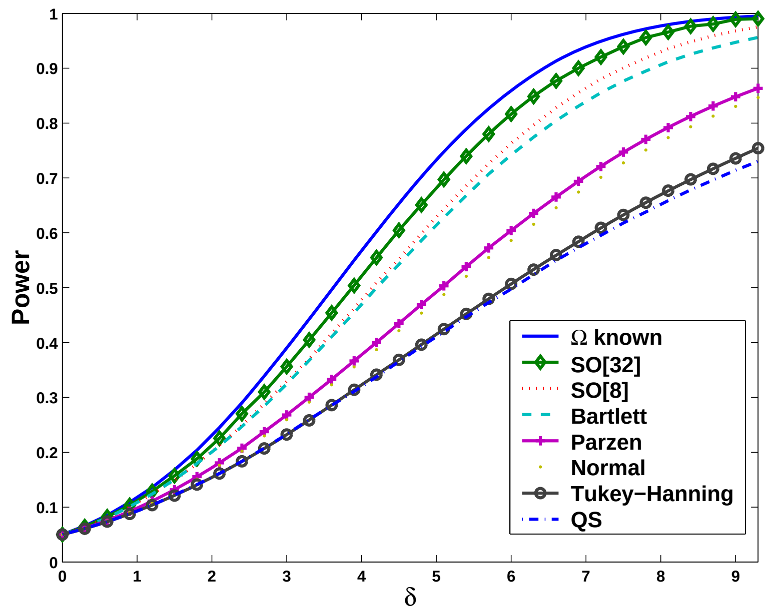

Figure 1: Asymptotic Local Power Function of the t^* Tests with Various Kernels

is expected. When ρ is large, it may be regarded as being roughly compatible with the rate condition in (14) for these sample sizes. In that case, the test statistic is effectively constructed using a consistent estimate of $\omega_{0.x}^2$, and Remark (d) above applies. In sum, tests based on steep Parzen kernels appear to dominate those based on other commonly used kernels.

3 Simulations

This section compares the finite sample performance of the t^* tests and the t-test that uses a HAC estimator, which we denote by t_{HAC}. We use the following simple data generating process

$$y_t = \alpha + x_t\beta + u_{0t}, \tag{17}$$

where $\alpha = \beta = 0$, $u_{0t} = \rho_1 u_{0t-1} + \rho_2 u_{0t-2} + e_t$, $x_t = x_{t-1} + u_{xt}$, $u_{xt} = 0.8u_{xt-1} + \eta_t, e_t$ and η_t are $iid(0,1)$ with $cov(e_t, \eta_t) = 0$, and $x_0 = u_0 = u_{-1} = 0$. Sample sizes $T = 50, 100$ and 200 are used and 2000 replications apply in all cases. We want to test one sided hypothesis $H_0 : \beta \leq 0$ $v.s.$ $H_1 : \beta > 0$. The cointegrating regression (17) is estimated using OLS (which is asymptotically equivalent to FM-OLS in this model) and t^* tests are constructed as in (9). For t_{HAC} , the bandwidth is chosen using the data-driven procedure in Andrews (1991).

We here compare the null rejection probabilities of t_{HAC} computed with a Bartlett kernel with those of t^* tests for the Bartlett, Parzen, and steep Parzen kernels with different values of ρ. Rejections were determined using the asymptotic 5% critical values from Table 1. Two patterns emerge in Table 2. First, finite sample null rejection probabilities are closer to 0.05 for t^* than for t_{HAC}, and the differences become more significant as ρ_1 approaches 1 when the errors follow an AR(1) process. Second, the power improvement of the steep Parzen kernels incurs some cost – the size distortions of the t^* test increases as ρ increases. However, both these patterns diminish with increases in sample size.

Another striking feature of steep Parzen kernels lies in the finite sample power improvement. Fig. 2 depicts the finite sample (size adjusted) power for the DGP with $\rho_1 = 0.85, \rho_2 = 0.0$ with sample size $T = 50$, and Fig. 3 shows the power curves for the DGP with $\rho_1 = 1.9, \rho_2 = -0.95$ with sample size $T = 50$. The power of the t^*-tests increases as ρ increases. While the power of the t_{HAC} test is higher than that of the t^*-test when the Bartlett and QS kernels are used, mirroring results for regression models reported in KV, the power of the two tests is almost the same when steep Parzen kernels with large ρ are used. In fact, for some cases as when $\rho = 32$, the power of the t^*-test dominates that of the t_{HAC} test.

Simulation results are similar when the regressors are allowed to be endogenous in the cointegration system and $cov(e_t, \eta_t) \neq 0$ for various correlation coefficients (see Table 3). The approach therefore turns out to provide a good alternative to tests based on conventional HAC estimation.

4 Conclusion and Extension

This paper shows that asymptotically valid cointegrating regression coefficient tests may be constructed using inconsistent HAC estimates. Simulations and asymptotic local power comparisons indicate that steep origin kernel methods typically work well in such situations, improving size properties and having power that is close to tests based on conventional HAC estimates. As in regression model testing, there remain issues of trade off between power and size in testing. Recent work by the authors (PSJ$_c$) provides one way in which this trade off may be confronted by using power parameter (or bandwidth) selection to minimise loss arising from power loss and size distortion.

Table 2: Finite Sample Null Hypothesis Rejection Probabilities: 50,000 Replications
DGP: $y_t = \alpha + x_t\beta + u_{0t}$; $x_t = x_{t-1} + u_{xt}$; $u_{xt} = 0.8u_{xt-1} + \eta_t$;$u_{0t} = \rho_1 u_{0t-1} + \rho_2 u_{0t-2} + e_t$;
e_t and η_t $iid(0,1)$ with $cov(e_t, \eta_t) = 0$, $x_0 = u_0 = u_{-1} = 0; \alpha = \beta = 0$;
$H_0 : \beta \leq 0$; $H_1 : \beta > 0$. Nominal level = 0.05.

ρ_1	ρ_2	t_{HAC}	t^*-Bartlett	t^*-Parzen	$t^*_{\rho=2}$	$t^*_{\rho=4}$	$t^*_{\rho=8}$	$t^*_{\rho=16}$	$t^*_{\rho=32}$
				$T = 50$					
−0.500	0.000	0.059	0.043	0.060	0.058	0.055	0.054	0.053	0.050
0.000	0.000	0.063	0.057	0.062	0.059	0.056	0.054	0.053	0.052
0.300	0.000	0.090	0.065	0.063	0.061	0.057	0.056	0.055	0.057
0.500	0.000	0.110	0.073	0.066	0.062	0.059	0.059	0.060	0.066
0.700	0.000	0.137	0.086	0.070	0.067	0.065	0.067	0.074	0.088
0.900	0.000	0.191	0.115	0.081	0.081	0.086	0.100	0.119	0.148
0.950	0.000	0.214	0.130	0.086	0.092	0.100	0.118	0.144	0.175
0.990	0.000	0.247	0.152	0.103	0.111	0.126	0.150	0.181	0.214
1.500	−0.750	0.123	0.048	0.057	0.052	0.045	0.035	0.027	0.026
1.900	−0.950	0.263	0.069	0.052	0.045	0.042	0.047	0.062	0.087
0.800	0.100	0.194	0.118	0.081	0.083	0.089	0.102	0.124	0.152
				$T = 100$					
−0.500	0.000	0.054	0.049	0.057	0.058	0.057	0.056	0.055	0.053
0.000	0.000	0.058	0.057	0.058	0.059	0.057	0.056	0.054	0.052
0.300	0.000	0.081	0.061	0.059	0.059	0.058	0.056	0.054	0.054
0.500	0.000	0.094	0.064	0.060	0.060	0.058	0.057	0.056	0.057
0.700	0.000	0.114	0.072	0.062	0.060	0.059	0.059	0.061	0.067
0.900	0.000	0.159	0.094	0.068	0.068	0.069	0.076	0.090	0.110
0.950	0.000	0.182	0.107	0.071	0.074	0.080	0.093	0.113	0.141
0.990	0.000	0.223	0.127	0.082	0.090	0.103	0.125	0.153	0.186
1.500	−0.750	0.090	0.048	0.057	0.057	0.055	0.051	0.043	0.032
1.900	0.950	0.197	0.038	0.047	0.046	0.038	0.025	0.018	0.019
0.800	0.100	0.162	0.096	0.069	0.069	0.071	0.079	0.094	0.116
				$T = 200$					
−0.500	0.000	0.048	0.049	0.055	0.055	0.054	0.054	0.053	0.051
0.000	0.000	0.054	0.053	0.055	0.055	0.054	0.054	0.052	0.051
0.300	0.000	0.073	0.055	0.055	0.055	0.054	0.053	0.052	0.051
0.500	0.000	0.081	0.057	0.055	0.056	0.055	0.054	0.052	0.052
0.700	0.000	0.093	0.061	0.066	0.057	0.055	0.054	0.053	0.055
0.900	0.000	0.130	0.075	0.059	0.058	0.058	0.062	0.067	0.079
0.950	0.000	0.155	0.089	0.062	0.063	0.065	0.074	0.087	0.108
0.990	0.000	0.202	0.112	0.071	0.078	0.089	0.107	0.132	0.164
1.500	−0.750	0.074	0.047	0.054	0.055	0.054	0.053	0.051	0.047
1.900	−0.950	0.133	0.034	0.051	0.053	0.053	0.049	0.036	0.018
0.800	0.100	0.135	0.077	0.059	0.059	0.059	0.063	0.070	0.082

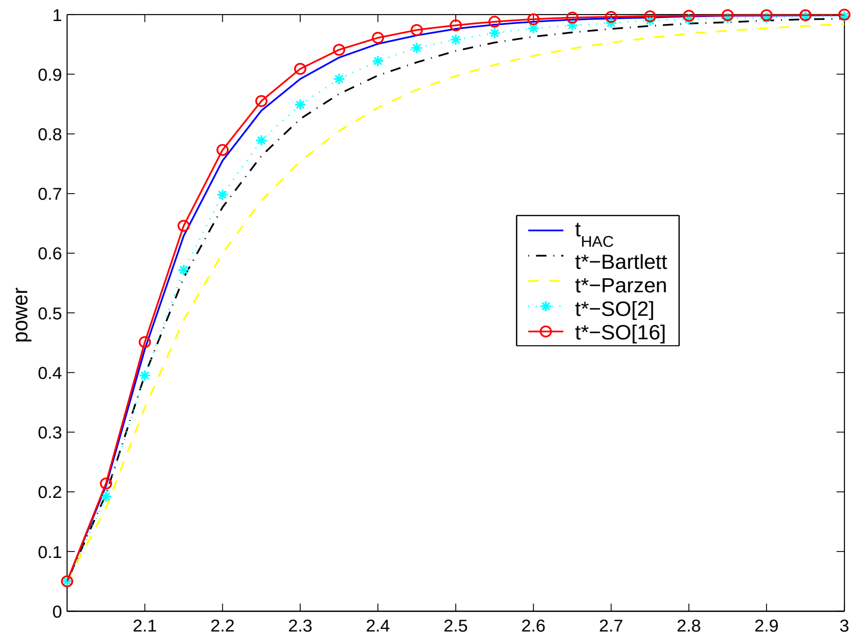

Figure 2: Finite Sample Power at 5% Nominal Level with $T = 50$,
$y_t = \alpha + x_t'\beta + u_{0t}; u_{0t} = 0.85u_{0t-1} + e_t; x_t = x_{t-1} + u_{xt}$,
$u_{xt} = 0.8u_{xt-1} + \eta_t$,
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, With No Prewhitening

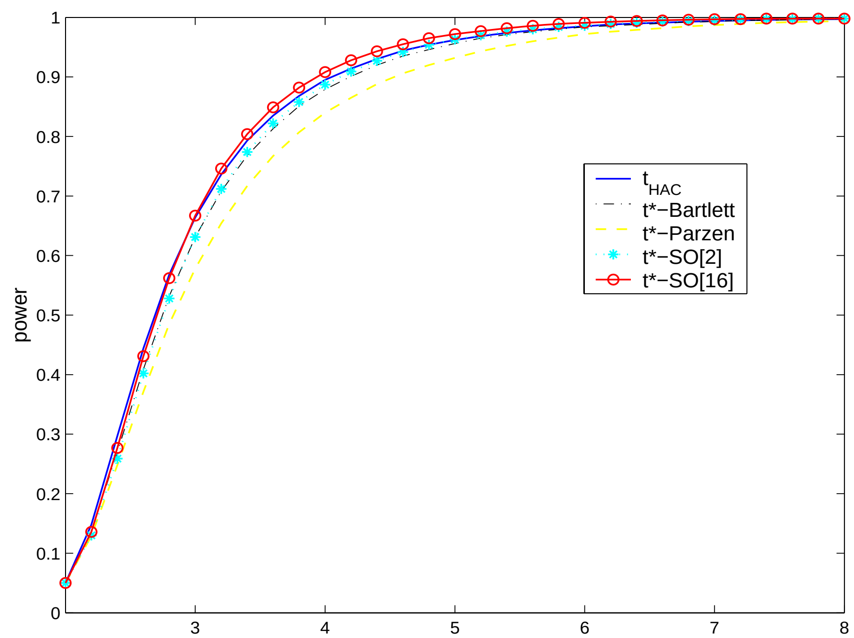

Figure 3: Finite Sample Power at 5% Nominal Level with $T = 50$,
$y_t = \alpha + x_t'\beta + u_{0t}; u_{0t} = 1.9u_{0t-1} - 0.95u_{0t-2} + e_t; x_t = x_{t-1} + u_{xt}$,
$u_{xt} = 0.8u_{xt-1} + \eta_t$,
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, With No Prewhitening

Table 3: Finite Sample Null Hypothesis Rejection Probabilities:
DGP: $y_t = \alpha + x_t\beta + u_{0t}$; $\alpha = \beta = 0$;$x_t = x_{t-1} + u_{xt}$; $u_{xt} = 0.8u_{xt-1} + \eta_t$;
$u_{0t} = \rho_1 u_{0t-1} + \rho_2 u_{0t-2} + e_t$, e_t and η_t $iid(0,1)$, $cov(e_t, \eta_t) = \sigma$, $x_0 = u_0 = u_{-1} = 0$;
$H_0 : \beta \leq 0$; $H_1 : \beta > 0$. Nominal level = 0.05,$T = 50$, with 50,000 Replications

	ρ_1	ρ_2	t_{HAC}	t^*-Bartlett	t^*-Parzen	$t^*_{\rho=2}$	$t^*_{\rho=4}$	$t^*_{\rho=8}$	$t^*_{\rho=16}$
$\sigma = -0.400$	0.900	0.000	0.170	0.075	0.050	0.054	0.059	0.069	0.085
	0.950	0.000	0.147	0.061	0.042	0.043	0.049	0.059	0.074
	0.990	0.000	0.136	0.054	0.038	0.041	0.047	0.058	0.073
$\sigma = 0.400$	0.900	0.000	0.389	0.083	0.051	0.088	0.115	0.144	0.176
	0.950	0.000	0.447	0.155	0.100	0.148	0.182	0.222	0.248
	0.990	0.000	0.510	0.269	0.190	0.231	0.266	0.311	0.339
$\sigma = 0.800$	0.900	0.000	0.484	0.061	0.037	0.075	0.127	0.188	0.253
	0.950	0.000	0.665	0.183	0.115	0.209	0.299	0.389	0.467
	0.990	0.000	0.814	0.461	0.347	0.467	0.550	0.623	0.684

5 Proofs

We derive the limit distribution of the regression F-statistic and t-statistic in the case where Ω is block diagonal. The general case could then be easily obtained.

When Ω is block diagonal, OLS is asymptotically equivalent to FM-OLS. Thus,

$$\begin{aligned} T(\widehat{\beta} - \beta) &\Rightarrow \left[\int_0^1 \underline{B}_x \underline{B}_x'\right]^{-1} \left[\int_0^1 \underline{B}_x dB_0\right] \\ &= \omega_0 \Omega_{xx}^{-1/2} \left(\int_0^1 \underline{W}_x \underline{W}_x'\right)^{-1} \int_0^1 \underline{W}_x dW_0, \end{aligned} \tag{18}$$

where $\widehat{\beta}$ is the OLS estimator for β. Using standard arguments we obtain from (3)

$$T^{-2} M_{xx} \Rightarrow \int_0^1 \underline{B}_x \underline{B}_x' = \Omega_{xx}^{1/2} \left(\int_0^1 \underline{W}_x \underline{W}_x'\right) \Omega_{xx}^{1/2}, \tag{19}$$

and combining these results, the limit distribution of the Wald statistic is of the form

$$F^* = T(R\widehat{\beta} - r)'[R(T^{-2}M_{xx})^{-1}R']^{-1} T(R\widehat{\beta} - r)/\widehat{\omega}_0^{*2} \tag{20}$$

$$\Rightarrow \frac{Q'\left[\widetilde{R}\left(\int_0^1 \underline{W}_x \underline{W}_x'\right)^{-1} \widetilde{R}'\right]^{-1} Q}{\int_0^1 \int_0^1 k(r-s) d\eta(r) d\eta(s)}. \tag{21}$$

where

$$\widetilde{R} = R\Omega_{xx}^{-1/2}, \text{ and } Q = \widetilde{R}\left(\int_0^1 \underline{W}_x \underline{W}_x'\right)^{-1} \left(\int_0^1 \underline{W}_x' dW_0\right) \tag{22}$$

Note that conditional on $\underline{W}_x$, $\int_0^1 \underline{W}_x' dW_0$ and $\eta(r)$ are normal random variables with covariance

$$\begin{aligned}
& E\left\{W_0(r) - rW_0(1) - \int_0^r \underline{W}_x' \left(\int_0^1 \underline{W}_x \underline{W}_x'\right)^{-1} \int_0^1 \underline{W}_x dW_0\right\} \int_0^1 \underline{W}_x' dW_0 \\
= & \int_0^r \underline{W}_x'(s)ds - r\int_0^1 \underline{W}_x'(s)ds - \int_0^r \underline{W}_x'(s)ds \left(\int_0^1 \underline{W}_x \underline{W}_x'\right)^{-1} \int_0^1 \underline{W}_x \underline{W}_x' \\
= & \int_0^r \underline{W}_x'(s)ds - r\int_0^1 \underline{W}_x'(s)ds - \int_0^r \underline{W}_x'(s)ds = 0,
\end{aligned} \tag{23}$$

where we have used $\int_0^1 \underline{W}_x'(s)ds = 0$ by definition. Therefore, $\int_0^1 \underline{W}_x' dW_0$ and $\eta(r)$ are conditionally independent. This implies that $\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)$ and $Q'[\widetilde{R}(\int_0^1 \underline{W}_x \underline{W}_x')^{-1}\widetilde{R}']^{-1}Q$ are conditionally independent. But conditional on $\underline{W}_x$, $Q'[\widetilde{R}(\int_0^1 \underline{W}_x \underline{W}_x')^{-1}\widetilde{R}']^{-1}Q$ follows the χ_q^2 distribution, which does not depend on $\underline{W}_x$. As a result, $\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)$ and $Q'[\widetilde{R}(\int_0^1 \underline{W}_x \underline{W}_x')^{-1}\widetilde{R}']^{-1}Q$ are independent.

Let

$$H = \begin{bmatrix} \widetilde{R} \\ M \end{bmatrix}_{m\times m}, \text{ and } H^{-1} = \begin{bmatrix} \widetilde{R} \\ M \end{bmatrix}^{-1} := \begin{bmatrix} A & B \end{bmatrix}_{m\times m} \tag{24}$$

where M is some $(m-q)\times m$ matrix such that H is of full rank m and $\widetilde{R}'M = 0$, A is a $m\times q$ matrix of rank q, and B is a $m\times(m-q)$ matrix. It is clear that $\widetilde{R}A = I_{q\times q}$ and $\widetilde{R}B = 0$ since $HH^{-1} = I_{m\times m}$. Denote $\widetilde{W}_x = H'^{-1}W_x$ and $\underline{\widetilde{W}}_x = H'^{-1}\underline{W}_x$, then the variance of $\widetilde{W}_x(1)$ is

$$\widetilde{\Omega}_{xx} = \left(HH'\right)^{-1} = \begin{bmatrix} \widetilde{R}\widetilde{R}' & 0 \\ 0 & MM' \end{bmatrix}^{-1} = \begin{bmatrix} (\widetilde{R}\widetilde{R}')^{-1} & 0 \\ 0 & (MM')^{-1} \end{bmatrix}. \tag{25}$$

Partition $\underline{\widetilde{W}}_x$ into submatrices $\left[\underline{\widetilde{W}}_1', \underline{\widetilde{W}}_2'\right]'$ and $\widetilde{W}_x$ into submatrices $\left[\widetilde{W}_1', \widetilde{W}_2'\right]'$, then $\widetilde{W}_1$ and $\widetilde{W}_2$ are independent Brownian motion with $\text{Var}(\widetilde{W}_2(1)) = (MM')^{-1}$ and

$$\text{Var}(\widetilde{W}_1(1)) = (\widetilde{R}\widetilde{R}')^{-1} = (R\Omega_{xx}^{-1}R')^{-1} := \widetilde{\Lambda}_1\widetilde{\Lambda}_1' \tag{26}$$

Using $\underline{W}_x = H'\underline{\widetilde{W}}_x$, we can rewrite Q as

$$\begin{aligned}
Q &= \widetilde{R}\left(\int_0^1 H'\underline{\widetilde{W}}_x\underline{\widetilde{W}}_x' H\right)^{-1} H'\int_0^1 \underline{\widetilde{W}}_x dW_0 \\
&= \widetilde{R}H^{-1}\left(\int_0^1 \underline{\widetilde{W}}_x\underline{\widetilde{W}}_x'\right)^{-1}\int_0^1 \underline{\widetilde{W}}_x dW_0 \\
&= \widetilde{R}\begin{bmatrix} A & B \end{bmatrix}\begin{pmatrix} \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_1' & \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_2' \\ \int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_1' & \int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_2' \end{pmatrix}^{-1}\begin{bmatrix} \int_0^1 \underline{\widetilde{W}}_1 dW_0 \\ \int_0^1 \underline{\widetilde{W}}_2 dW_0 \end{bmatrix}
\end{aligned} \tag{27}$$

which, by the partitioned inverse formula, is equal to

$$
\begin{aligned}
&\left(\int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_1' - \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_2'(\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_2')^{-1}\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_1'\right)^{-1}\int_0^1 \underline{\widetilde{W}}_1 dW_0 \\
&-\left(\int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_1' - \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_2'(\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_2')^{-1}\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_1'\right)^{-1} \\
&\times \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_2'\left(\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_2'\right)^{-1}\int_0^1 \underline{\widetilde{W}}_2 dW_0 \\
= \;&\left(\int_0^1 \underline{\widetilde{W}}_{1.2}\underline{\widetilde{W}}_{1.2}'\right)^{-1}\int_0^1 \underline{\widetilde{W}}_{1.2} d\underline{W}_0 \qquad (28)
\end{aligned}
$$

by the fact that $\widetilde{R}A = I_{q\times q}$ and $\widetilde{R}B = 0$ where

$$
\underline{\widetilde{W}}_{1.2} = \underline{\widetilde{W}}_1 - \int_0^1 \underline{\widetilde{W}}_1\underline{\widetilde{W}}_2'(\int_0^1 \underline{\widetilde{W}}_2\underline{\widetilde{W}}_2')^{-1}\underline{\widetilde{W}}_2, \qquad (29)
$$

the L_2 regression residual from projecting $\underline{\widetilde{W}}_1$ on $\underline{\widetilde{W}}_2$.

Similarly, we can get

$$
R\left(\int_0^1 \underline{W}_x\underline{W}_x'\right)^{-1}R' = \left(\int_0^1 \underline{\widetilde{W}}_{1.2}\underline{\widetilde{W}}_{1.2}'\right)^{-1}. \qquad (30)
$$

Thus,

$$
\begin{aligned}
F^* \;&\Rightarrow\; \frac{(\int_0^1 \underline{\widetilde{W}}_{1.2}d\underline{W}_0)'\left(\int_0^1 \underline{\widetilde{W}}_{1.2}\underline{\widetilde{W}}_{1.2}'\right)^{-1}\int_0^1 \underline{\widetilde{W}}_{1.2}d\underline{W}_0}{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)} \\
&=_d \frac{(\int_0^1 \underline{W}_{1.2}d\underline{W}_0)'\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1}\int_0^1 \underline{W}_{1.2}d\underline{W}_0}{\int_0^1\int_0^1 k(r-s)d\eta(r)d\eta(s)}, \qquad (31)
\end{aligned}
$$

where '$=_d$' denotes distributional equivalence, $\underline{W}_{1.2} = \underline{W}_1 - \int_0^1 \underline{W}_1\underline{W}_2'(\int_0^1 \underline{W}_2\underline{W}_2')^{-1}\underline{W}_2$, the L_2 regression residual of $\underline{W}_1$ on $\underline{W}_2$, and $\underline{W}_1$ is the first q coordinates of $\underline{W}_x$, and $\underline{W}_2$ is the last $(m-q)$ coordinates. The distributional equivalence holds because $(\int_0^1 \underline{W}_{1.2}d\underline{W}_0)'\left(\int_0^1 \underline{W}_{1.2}\underline{W}_{1.2}'\right)^{-1}\int_0^1 \underline{W}_{1.2}d\underline{W}_0$ is independent of $\underline{W}_{1.2}$.

Under the local alternative hypothesis $H_1 : R\beta = r + c/T$ for some $c \neq 0$, we have

$$
T(R\widehat{\beta} - r) = RT(\widehat{\beta} - \beta) + c \Rightarrow_d \omega_0 Q + c. \qquad (32)
$$

So

$$
\begin{aligned}
&T(R\widehat{\beta} - r)'[R(T^{-2}M_{xx})^{-1}R']^{-1}T(R\widehat{\beta} - r) \\
\Rightarrow\; &(\omega_0 Q + c)'\left[\widetilde{R}\left(\int_0^1 \underline{W}_x\underline{W}_x'\right)^{-1}\widetilde{R}'\right]^{-1}\omega_0 Q + c
\end{aligned}
$$

$$= \omega_0^2 \left[\left(\int_0^1 \widetilde{\Lambda}_1^{-1} \underline{\widetilde{W}}_{1.2} \underline{\widetilde{W}}_{1.2}' \widetilde{\Lambda}_1'^{-1} \right)^{-1} \int_0^1 \widetilde{\Lambda}_1^{-1} \underline{\widetilde{W}}_{1.2} d\underline{W}_0 + \widetilde{\Lambda}_1' c/\omega_0 \right]'$$
$$\times \left(\int_0^1 \widetilde{\Lambda}_1^{-1} \underline{\widetilde{W}}_{1.2} \underline{\widetilde{W}}_{1.2}' \widetilde{\Lambda}_1'^{-1} \right)$$
$$\times \left[\left(\int_0^1 \widetilde{\Lambda}_1^{-1} \underline{\widetilde{W}}_{1.2} \underline{\widetilde{W}}_{1.2}' \widetilde{\Lambda}_1'^{-1} \right)^{-1} \int_0^1 \widetilde{\Lambda}_1^{-1} \underline{\widetilde{W}}_{1.2} d\underline{W}_0 + \widetilde{\Lambda}_1' c/\omega_0 \right] \quad (33)$$

and

$$\widehat{\omega}_0^{*2} \Rightarrow \omega_0^2 \int_0^1 \int_0^1 k(r-s) d\eta(r) d\eta(s) \quad (34)$$

where $\eta(r)$ can be represented as

$$W_0(r) - rW_0(1) - \int_0^r \underline{W}_x' \left(\int_0^1 \underline{W}_x \underline{W}_x' \right)^{-1} \int_0^1 \underline{W}_x dW_0$$
$$= W_0(r) - rW_0(1) - \int_0^r \underline{\widetilde{W}}_x' H \left(\int_0^1 H' \underline{\widetilde{W}}_x \underline{\widetilde{W}}_x' H \right)^{-1} \int_0^1 H' \underline{\widetilde{W}}_x dW_0 \quad (35)$$
$$= W_0(r) - rW_0(1) - \int_0^r \left(\widetilde{\Omega}_{xx}^{-1/2} \underline{\widetilde{W}}_x \right)' \left\{ \int_0^1 \widetilde{\Omega}_{xx}^{-1/2} \underline{\widetilde{W}}_x \left(\widetilde{\Omega}_{xx}^{-1/2} \underline{\widetilde{W}}_x \right)' \right\}^{-1} \int_0^1 \widetilde{\Omega}_{xx}^{-1/2} \underline{\widetilde{W}}_x dW_0.$$

Here we have written the weak limits of both $T(R\widehat{\beta} - r)'[R(T^{-2}M_{xx})^{-1}R']^{-1}T(R\widehat{\beta} - r)$ and $\widehat{\omega}_0^{*2}$ in terms of the transformed Brownian motion $\underline{\widetilde{W}}_x$ so that the weak convergence in (33) and (34) holds jointly.

Hence,

$$F^* \Rightarrow \frac{F_{1,\infty}}{\int_0^1 \int_0^1 k(r-s) d\eta(r) d\eta(s)} \quad (36)$$

where

$$F_{1,\infty} = \left[\left(\int_0^1 \underline{W}_{1.2} \underline{W}_{1.2}' \right)^{-1} \int_0^1 \underline{W}_{1.2} d\underline{W}_0 + \delta \right]' \left(\int_0^1 \underline{W}_{1.2} \underline{W}_{1.2}' \right)$$
$$\times \left[\left(\int_0^1 \underline{W}_{1.2} \underline{W}_{1.2}' \right)^{-1} \int_0^1 \underline{W}_{1.2} d\underline{W}_0 + \delta \right], \quad (37)$$

$\delta = \widetilde{\Lambda}_1' c/\omega_0 = \left(R\Omega_{xx}^{-1} R' \right)^{-1/2} c/\omega_{0.x}$, and

$$\eta(r) = W_0(r) - rW_0(1) - \int_0^r \underline{W}_x' \left(\int_0^1 \underline{W}_x \underline{W}_x' \right)^{-1} \int_0^1 \underline{W}_x dW_0. \quad (38)$$

The weak convergence in (36) follows because we can replace the normalized and demeaned Brownian motion process $\widetilde{\Omega}_{xx}^{-1/2} \underline{\widetilde{W}}_x$ in (33) and (34) by any demeaned standard Brownian motion, say, $\underline{W}_x$ without affecting the distribution of their ratio.

The proof of limit distribution of t^* is straightforward and is omitted.

References

Andrews, D. W. K., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817–854.

Jansson, M., 2004, On the Error of Rejection Probability in Simple Autocorrelation Robust Tests, *Econometrica,* 72, 937-946.

Kiefer, N. M. and T. J. Vogelsang, 2002a, Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size, *Econometric Theory* 18, 1350–1366.

———, 2002b, Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation, *Econometrica* 70, 2093–2095.

———, 2005, A New Asymptotic Theory for Heteroskedasiticy-Autocorrelation Robust Tests, forthcoming, *Econometric Theory.*

Phillips, P. C. B. and V. Solo, 1992, Asymptotics for Linear Processes, *Annals of Statistics*, 20, 971–1001.

Phillips, P. C. B., Y. Sun and S. Jin, 2005a, Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation, forthcoming, *Journal of Statistical Planning and Inference.*

———, 2005b, Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation, forthcoming, *International Economic Review.*

———, 2005c, Improved HAR Inference Using Power Kernels without Truncation, Yale University, mimeographed.