CONSISTENT HAC ESTIMATION AND ROBUST REGRESSION TESTING USING SHARP ORIGIN KERNELS WITH NO TRUNCATION

By

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

March 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1407

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation*

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

Yixiao Sun
Department of Economics
University of California, San Diego

and

Sainan Jin
Department of Economics
Yale University

February 2003

*Phillips thanks the NSF for research support under Grant No. SES 00-92509. Jin thanks the Cowles Foundation for Fellowship support.

ABSTRACT

A new family of kernels is suggested for use in heteroskedasticity and autocorrelation consistent (HAC) and long run variance (LRV) estimation and robust regression testing. The kernels are constructed by taking powers of the Bartlett kernel and are intended to be used with no truncation (or bandwidth) parameter. As the power parameter (ρ) increases, the kernels become very sharp at the origin and increasingly downweight values away from the origin, thereby achieving effects similar to a bandwidth parameter. Sharp origin kernels can be used in regression testing in much the same way as conventional kernels with no truncation, as suggested in the work of Kiefer and Vogelsang (2002a, 2002b). A unified representation of HAC limit theory for untruncated kernels is provided using a new proof based on Mercer's theorem that allows for kernels which may or may not be differentiable at the origin. This new representation helps to explain earlier findings like the dominance of the Bartlett kernel over quadratic kernels in test power and yields new findings about the asymptotic properties of tests with sharp origin kernels. Analysis and simulations indicate that sharp origin kernels lead to tests with improved size properties relative to conventional tests and better power properties than other tests using Bartlett and other conventional kernels without truncation.

If ρ is passed to infinity with the sample size (T), the new kernels provide consistent HAC and LRV estimates as well as continued robust regression testing. Optimal choice of ρ based on minimizing the asymptotic mean squared error of estimation is considered, leading to a rate of convergence of the kernel estimate of $T^{1/3}$, analogous to that of a conventional truncated Bartlett kernel estimate with an optimal choice of bandwidth. A data-based version of the consistent sharp origin kernel is obtained which is easily implementable in practical work.

Within this new framework, untruncated kernel estimation can be regarded as a form of conventional kernel estimation in which the usual bandwidth parameter is replaced by a power parameter that serves to control the degree of downweighting. Simulations show that in regression testing with the sharp origin kernel, the power properties are better than those with simple untruncated kernels (where $\rho = 1$) and at least as good as those with truncated kernels. Size is generally more accurate with sharp origin kernels than truncated kernels. In practice a simple fixed choice of the exponent parameter around $\rho = 16$ for the sharp origin kernel produces favorable results for both size and power in regression testing with sample sizes that are typical in econometric applications.

JEL Classification: C13; C14; C22; C51

Keywords: Consistent HAC estimation, data determined kernel estimation, long run variance, Mercer's theorem, power parameter, sharp origin kernel.

1 Introduction

While much practical econometric testing makes use of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimates, it is not necessary that such estimates be employed in order to produce asymptotically similar tests. In this spirit, Kiefer, Vogelsang and Bunzel (2000; hereafter, KVB) and Kiefer and Vogelsang (2002a, 2002b; hereafter, KV) have proposed the use in robust regression testing of kernel based covariance matrix estimates in which the bandwidth parameter (M) is set to the sample size (T). While these estimates are inconsistent for the asymptotic covariance matrix, they nevertheless lead to asymptotically valid regression tests. Simulations reveal that the null asymptotic approximation of these tests is often more accurate in finite samples than that of tests based on consistent HAC estimates, although prewhitening is known to improve size accuracy in the latter (den Haan and Levin, 1997) particularly when model selection is used in the selection of the prewhitening filter (Lee and Phillips, 1994)[1]. Using higher order asymptotics, Jansson (2002) explained this improved accuracy of the null approximation, showing that the error rejection probability (ERP) in a Gaussian location model of these tests is $O(T^{-1})$, where T is the sample size, in contrast to the usual rate of $O(T^{-1/2})$ that is attained by tests using conventional HAC estimates.

While these alternative robust tests based on inconsistent HAC estimates have greater accuracy in size, they also experience a loss of power, including local asymptotic power, in relation to conventional tests. To address this deficiency, Jansson (2002) proposed a weighting scheme, analogous to that used in Anderson and Darling (1952), in the construction of these alternative tests that delivers power improvements while retaining their better size properties in finite samples. Moreover, a local power analysis in Kiefer and Voglesang (2002b) reveals that it is the Bartlett kernel among the common choices of kernel that produces the highest power function when bandwidth is set to the sample size. The latter outcome may appear unexpected in view of the usual preferred choice of quadratic (at the origin) kernels in terms of their better asymptotic mean squared error characteristics in consistent spectral density and HAC estimation (Hannan, 1970; Andrews, 1991). Unlike quadratic kernels, the Bartlett kernel has a tent shape, is not differentiable at the origin, and the reason for its better power performance characteristics is unexplained.

The present paper takes a new look at HAC estimation and robust regression testing using kernel estimation without truncation (or when the bandwidth equals the sample size). One contribution of the paper is to provide a unified representation of the HAC limit theory in such cases and a new proof using Mercer's theorem that allows for kernels which may or may not be differentiable at the origin. This new representation gives a coherent asymptotic theory and thereby helps to explain earlier findings like the better power performance of the untruncated Bartlett kernel mentioned above.

[1] If the autocorrelation is parametric and model selection based prefiltering (within the correct parametric class) is employed in conjunction with conventional kernel HAC estimation using a data determined bandwidth, Lee and Phillips (1994) show that the bandwidth is effectively proportional to the sample size T (up to a slowly varying factor) and a convergence rate of $\sqrt{T}$ (up to a slowly varying factor) for the HAC estimator is attainable.

Our main contribution is to provide a new approach to HAC estimation that embeds the Bartlett kernel in a new class of sharp origin kernels. The new kernels are equal to the Bartlett kernel raised to some positive power (ρ). For $\rho > 1$, the kernels have a sharper peak at the origin and they downweight non zero values more rapidly than the Bartlett kernel. The asymptotic theory for HAC estimation and regression testing with sharp origin kernels turns out to differ in some important ways and yet to be similar in others to that of the conventional Bartlett kernel.

We consider two cases, one where the power parameter ρ is fixed and the other where ρ passes to infinity with T. When $\rho > 1$ is fixed as $T \to \infty$, HAC estimation based on this sharp origin kernel is inconsistent just as it is when $\rho = 1$. However, compared with the Bartlett kernel, sharp origin kernels put less weight on autocovariances with larger lags and correspondingly deliver HAC estimates with smaller asymptotic variance. The reduction in asymptotic variance has important implications in regression testing. Compared with conventional tests that use consistent estimates of the asymptotic variance matrix, tests based on kernel estimates without truncation inevitably introduce additional variability by virtue of the fact that the HAC estimates are inconsistent, much as an F ratio has more variability because of its random denominator than the asymptotic chi-squared limit. This additional variability assists in better approximating finite sample behavior under the null while compromising power. Intuition suggests that test power may be improved if the variability can be reduced while at the same time maintaining more accurate size characteristics in finite samples. Sharp origin kernels can achieve variance reductions in this way, while continuing to deliver better size.

Our findings indicate that sharp origin kernels without truncation deliver asymptotically valid tests with greater accuracy in size and power close to or better than that of conventional tests. The simulations we report below show that tests based on sharp origin kernels with $\rho > 1$ uniformly dominate those based on the Bartlett kernel ($\rho = 1$). As ρ increases there is a tendency for greater size distortion, although even for samples as small as $T = 50$ the size distortion is smaller than that of the conventional tests using data driven bandwidth choices. Overall, our results indicate that a simple fixed choice of the exponent parameter around $\rho = 16$ for the sharp origin kernel produces favorable outcomes for both size and power in regression testing with sample sizes that are typical in econometric applications.

When $\rho \to \infty$ with T, sharp origin kernels provide a new mechanism for consistent HAC (and, more generally, spectral density) estimation. While there is no need to make bandwidth choices in this approach and test validity is retained whatever the choice of ρ, there is an opportunity for optimal choice of ρ. When ρ increases, the variance declines and bias increases, just as in conventional kernel estimation bias decreases and variance increases as the bandwidth shrinks. Accordingly, we develop an asymptotic distribution theory for consistent HAC estimation using sharp origin kernels with no truncation. Optimal choice of the power parameter ρ is then obtained by minimizing the asymptotic mean squared error of the HAC estimate, leading to an explicit rate $\rho = O(T^{2/3})$ which gives a convergence rate for the HAC estimate of $T^{1/3}$. This is precisely the same rate that applies when a truncated Bartlett kernel HAC estimate is implemented with an optimal bandwidth choice (c.f., Hannan, 1970; Andrews, 1991). These new asymptotics for sharp kernels, like those for truncated kernels, offer the opportunity of data-driven methods for selecting ρ in practical work; and an automated version of the new sharp

kernel HAC estimator is provided based on the plug-in approach.

The paper is organized as follows. Section 2 briefly overviews testing problems in the presence of nonparametric autocorrelation. Section 3 introduces sharp origin kernels and establishes the asymptotic properties of HAC estimators using these kernels without truncation when the power parameter ρ is fixed. A new proof using Mercer's theorem is given for problems of this type which allows for kernels which may or may not be differentiable at the origin. Section 4 develops the asymptotic theory for the case when $\rho \to \infty$ with T and extracts optimal values of ρ based on a MSE criterion. Section 5 provides a limit theory for regression tests using sharp origin kernels under both null and local alternatives. Section 6 reports simulation results on the finite sample performance of the proposed tests and makes some suggestions for implementation in practical econometric work. Section 7 concludes. Notation is given in a table at the end of the paper and proofs and additional technical results are in the Appendix.

2 Robust Testing of Regression Hypotheses

As in earlier work by KVB, we use the following linear regression model for exposition

$$y_t = x_t'\beta + u_t, \quad t = 1, 2, ..., T, \tag{1}$$

where u_t is autocorrelated, possibly conditionally heteroskedastic and x_t is such that assumption **A1** below holds. Least squares estimation leads to $\widehat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t$ and the scaled estimation error is written in the form

$$\sqrt{T}(\widehat{\beta} - \beta) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{\sqrt{T}} S_T\right), \tag{2}$$

where

$$S_t = \sum_{\tau=1}^{t} v_\tau, \text{ and } v_\tau = x_\tau u_\tau. \tag{3}$$

Let $\widehat{v}_\tau = x_\tau \widehat{u}_\tau$ be estimates of v_τ constructed from the regression residuals $\widehat{u}_\tau = y_\tau - x_\tau'\widehat{\beta}$, and define the corresponding partial sum process $\widehat{S}_t = \sum_{\tau=1}^{t} \widehat{v}_\tau$.

The following high level condition for which sufficient conditions are well known (e.g. Phillips and Solo, 1992) facilitates the asymptotic development and is in common use (e.g., KVB, Jansson, 2002).

A1:

(a) *$S_{[Tr]}$ satisfies the functional law*

$$T^{-1/2} S_{[Tr]} \Rightarrow \Lambda W_m(r), \quad r \in [0, 1] \tag{4}$$

where $\Lambda\Lambda' = \Omega > 0$ is the long run variance of v_t and $W_m(r)$ is m-dimensional standard Brownian motion.

(b) $\text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{[Tr]} x_t x_t' = rQ$ *uniformly in* $r \in [0, 1]$ *with positive definite* Q.

Under **A1** we have

$$T^{-1/2}\widehat{S}_{[Tr]} \Rightarrow \Lambda V_m(r), \quad r \in [0,1]\,, \tag{5}$$

where V_m is a standard m-dimensional Brownian bridge process, as well as the usual regression limit theory

$$\sqrt{T}(\widehat{\beta}-\beta) \Rightarrow Q^{-1}\Lambda W_m(1) = N(0, Q^{-1}\Omega Q^{-1}), \tag{6}$$

which provides a basis for robust regression testing on β. The conventional approach relies on consistent estimation of the sandwich variance matrix $Q^{-1}\Omega Q^{-1}$ in (6), which in turn involves the estimation of Ω since Q^{-1} is consistently estimated by $\widehat{Q}^{-1}$ where $\widehat{Q} = T^{-1}\sum_{t=1}^{T} x_t x_t'$. Many consistent estimators of Ω have been proposed in the literature (see, for example, White (1980), Newey and West (1987), Andrews (1991), Hansen (1992) and de Jong and Davidson (2000)). Among them, kernel-based nonparametric estimators that involve smoothing and truncation are in common use. When v_t is stationary with spectral density matrix $f_{vv}(\lambda)$, the long run variance (LRV) of v_t is

$$\Omega = \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma(j)+\Gamma(j)'\right) = 2\pi f_{vv}(0)\,, \tag{7}$$

where $\Gamma(j) = E(v_t v_{t-j}')$. Consistent kernel based estimation of Ω typically involves use of formulae motivated by (7) of the general form

$$\widehat{\Omega}(M) = \sum_{j=-T+1}^{T-1} k(\frac{j}{M})\widehat{\Gamma}(j), \tag{8}$$

$$\widehat{\Gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j}\widehat{v}_{t+j}\widehat{v}_t' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T}\widehat{v}_{t+j}\widehat{v}_t' & \text{for } j < 0 \end{cases} \tag{9}$$

involving the sample covariances $\widehat{\Gamma}(j)$ that are based on estimates $\widehat{v}_t = x_t\widehat{u}_t = x_t(y_t - x_t'\widehat{\beta})$ of v_t constructed from regression residuals. In (8), $k(\cdot)$ is a kernel function and M a bandwidth parameter. Consistency of $\widehat{\Omega}(M)$ requires $M \to \infty$ and $M/T \to 0$ as $T \to \infty$.

Various kernel functions $k(\cdot)$ are available for use in (8) and their properties have been extensively explored in the time series literature (e.g., Parzen, 1957; Hannan, 1970; and Priestley, 1981) from which the econometric literature on HAC estimation draws. Some of these properties, such as asymptotic bias and mean squared error, hinge on the behavior of the kernel function around the origin which is often characterized in terms of the Parzen exponent q, the first positive integer for which $k_q = \lim_{x\to 0}\{1-k(x))/|x|^q\} \neq 0$. Most standard kernels (except the Bartlett) have $q = 2$ and hence quadratic behavior around the origin. These kernels have been found to produce estimates $\widehat{\Omega}(M)$ with preferable asymptotic MSE properties and better rates of convergence for optimal choices of the bandwidth than other kernels. When $q = 2$, this rate of convergence is $T^{2/5}$. The Bartlett kernel, which is also commonly used in econometric work (Newey and West, 1987, 1994), has $q = 1$. When an optimal bandwidth rate is used with this kernel, the rate of convergence of $\widehat{\Omega}(M)$ is $T^{1/3}$. While none of these considerations matter

asymptotically when all that is needed is a consistent estimate of Ω, they do play an important role in finite sample behavior. Indeed, higher order expansions, as in Linton (1995) and Xiao and Phillips (1998, 2002) show that improved regression estimation and testing can be accomplished using appropriate bandwidth selection that takes into account higher order behavior. While we will not pursue that line of analysis in the present paper, we do note here the important differences between standard kernels for which $q = 2$ and the Bartlett kernel where $q = 1$.

To test a null such as $H_0 : R\beta = r$, where R is a known $p \times m$ matrix of rank p and r is a specified p-vector, the standard approach relies on the F-ratio statistic of the form

$$F_{\widehat{\Omega}(M)} = T(R\widehat{\beta} - r)' \left(R\widehat{Q}^{-1}\widehat{\Omega}\,(M)\,\widehat{Q}^{-1}R' \right)^{-1} (R\widehat{\beta} - r)/p, \tag{10}$$

which is asymptotically χ^2_p/p. Use of $F_{\widehat{\Omega}(M)}$ is very convenient in empirical work and robustifies the test to a wide range of possible behaviour in the regression error u_t in (1). However, it is well known that the size of tests based on (10) can be poorly approximated by the asymptotic distribution, which neglects the finite sample randomness induced by the consistent HAC estimate $\widehat{\Omega}\,(M)$, although prewhitening in the estimation of Ω does help to ameliorate finite sample performance — see den Haan and Levin (1997) and Jansson (2002) for further details and discussion.

KV proposed a class of kernel based estimators of Ω in which standard kernels are used but where the bandwidth parameter is set equal to the sample size. These estimates are inconsistent and tend to random matrices instead of Ω. Nonetheless, valid asymptotically similar tests can be constructed with these covariance estimators in the same manner as (10) but with a different limit distribution for the test that depends on the form of the kernel. KV showed that the Bartlett kernel delivers tests with the highest power among the standard kernels, including those for which $q = 2$, although this finding is unexplained. Jansson (2002) showed that power improvements are possible by use of a suitably chosen weight function in the construction of the estimate of Ω.

Following KV, the next section proposes a new class of power kernels where the bandwidth is set to the sample size and which dominate the Bartlett kernel in a sense that will be made clear later on. We also provide a new way of deriving the limit theory of these inconsistent HAC estimators.

3 Sharp Origin Kernels and HAC Estimation

We define a class of sharp origin (SO) kernels by taking an arbitrary power $\rho \geq 1$ of the usual Bartlett kernel, giving

$$k_\rho(x) = \begin{cases} (1 - |x|)^\rho, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases} \quad \text{for } \rho \geq 1. \tag{11}$$

When $\rho = 1$, $k_\rho(x)$ is the usual Bartlett kernel. As ρ increases, $k_\rho(x)$ becomes successively more concentrated at the origin and its peak more pronounced and sharp. Fig. 1 graphs $k_\rho(x)$ for several values of ρ illustrating these effects.

SO kernels have the following properties, which may be readily verified.

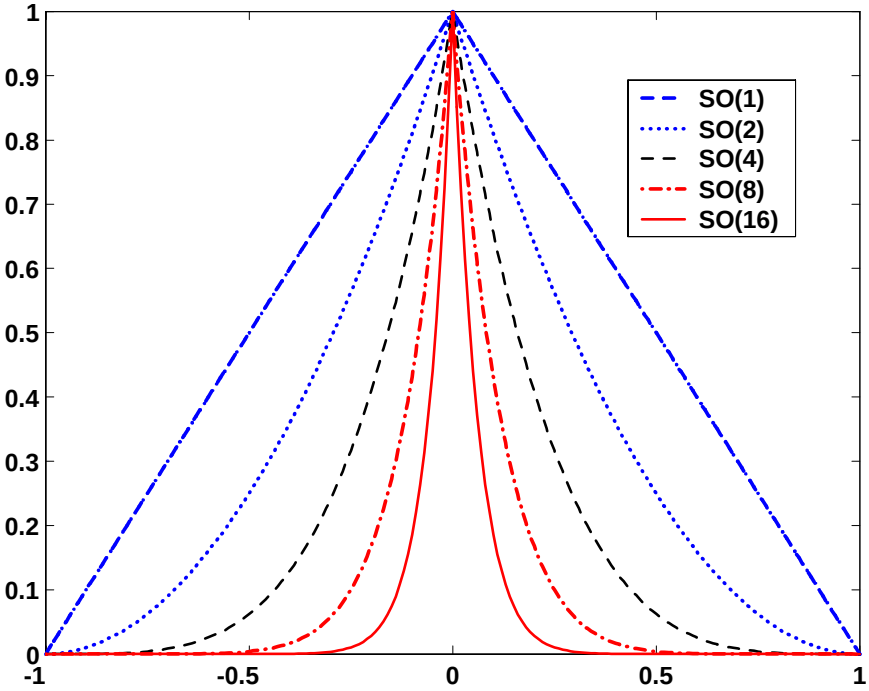

Figure 1: Sharp Origin (SO) Kernels $k_\rho(x)$ for $\rho \in [1, 16]$

(i) $k_\rho(x)$: $(-\infty, \infty) \to [0, 1]$ satisfies $k_\rho(x) = k_\rho(-x)$, $k_\rho(0) = 1$, and $k_\rho(1) = 0$.

(ii) The Parzen exponent (Parzen, 1957) of $k_\rho(x)$ equals 1, i.e. $q = 1$ is the largest integer such that $\lim_{x\to 0} [1 - (1 - |x|)^\rho] |x|^{-q}$ is finite and

$$k_1 = \lim_{x\to 0} \frac{1 - (1 - |x|)^\rho}{|x|} = \rho < \infty. \tag{12}$$

(iii) $k_\rho(x)$ is positive semi-definite in the sense that $\int_{-1}^{1} k(x) \exp(-\lambda x) dx \geq 0$ for any $\lambda \in \mathbb{R}$.

As is well known (e.g. Newey and West, 1987; Andrews, 1991), positive semi-definiteness of $k_\rho(x)$ guarantees the positive semi-definiteness of kernel HAC estimators defined as in (14) below. It also enables us to use Mercer's theorem (e.g., see Shorack and Wellner, 1986), as it implies that for any square integrable function $f(x)$, $\int_0^1 \int_0^1 k(r - s) f(r) f(s) dr ds \geq 0$. The version of Mercer's theorem given below provides a convenient mechanism for extracting the limit distribution of inconsistent HAC estimators and does so in a unified way for kernels with various Parzen exponents, thereby removing the need for separate arguments such as those in Kiefer and Vogelsang (2002a, 2002b).

Mercer's Theorem *If $k(x)$ is positive semi-definite, then*

$$k(r-s) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} f_n(r) f_n(s), \tag{13}$$

where $\lambda_n > 0$ are the eigenvalues of the kernel and $f_n(x)$ are the corresponding eigenfunctions, i.e. $f_n(s) = \lambda_n \int_0^1 k(r-s) f_n(r) dr$. The right side of (13) converges uniformly over $(r,s) \in [0,1] \times [0,1]$.

Using the kernel function k_ρ in expression (8) and letting $M = T$ gives a class of untruncated HAC estimators of the form

$$\widehat{\Omega}_{k_\rho} = \sum_{j=-T+1}^{T-1} k_\rho \left(\frac{j}{T} \right) \widehat{\Gamma}(j). \tag{14}$$

In what follows in this section we will assume that the ρ value in (14) is fixed as $T \to \infty$.

The following theorem establishes the asymptotic properties of a HAC estimator $\widehat{\Omega}_k$ of the general form given in (8) with bandwidth $M = T$ and any positive semi-definite kernel k. The result includes cases such as $\widehat{\Omega}_{k_\rho}$ in (14) where a sharp origin kernel has been used.

Theorem 1 *Let **A1** hold and define $\widehat{\Omega}_k = \widehat{\Omega}(T)$ as in (8) with $M = T$. If the kernel $k(x)$ used in $\widehat{\Omega}_k$ is positive semi-definite, then the following results hold:*

(a) $\widehat{\Omega}_k \Rightarrow \Lambda \Xi \Lambda'$, where $\Xi = \int_0^1 \int_0^1 k(r-s) dV_m(r) dV_m'(s)$,

and $V_m(r)$ is an m-vector of standard Brownian bridges.

(b) $E(\Lambda \Xi \Lambda') = \mu \Omega$, where $\mu = 1 - \int_0^1 \int_0^1 k(r-s) dr ds$.

(c) $var(vec(\Lambda \Xi \Lambda')) = \nu (I_{m^2} + K_{mm}) \Omega \otimes \Omega$ where

$$\nu = \int k(r-s) k(p-q) - 2 \int k(r-s) k(r-q) + \int k(r-s)^2,$$

and the integrals are taken with respect to all the underlying argument variables, for example

$$\int k(r-s) k(p-q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 k(r-s) k(p-q) dr ds dp dq.$$

Remarks

(a) The theorem shows that $\widehat{\Omega}_k$, properly scaled, is centered on the true long run covariance matrix. In particular, $\widehat{\Omega}_k / \mu$ is asymptotically unbiased for Ω, but inconsistent in view of part (a).

(b) Part (c) shows that the asymptotic covariance between the (a,b) and (c,d) elements of $\widehat{\Omega}_k$ is $\nu \left(\Omega_{ac} \Omega_{bd} + \Omega_{ad} \Omega_{bc} \right)$, where Ω_{ab} denotes the (a,b) element of Ω.

(c) KV established asymptotic results similar to Theorem 1(a) under different assumptions. They assumed the kernels are continuously differentiable to the second order. As a consequence, they had to treat the Bartlett kernel separately, obtaining different representations of the limit distributions for these two cases. The unified representation given in Theorem 1(a) brings results of this type together, shortens the technical arguments and allows for a coherent method of proof.

(d) As shown in the proof of the Theorem, an alternative representation of Ξ in part (a) is

$$\Xi = \int_0^1 \int_0^1 k^*(r,s) dW_m(r) dW_m'(s),$$

where

$$k^*(r,s) = k(r-s) - \int_0^1 k(r-t)dt - \int_0^1 k(\tau - s)d\tau + \int_0^1 k(t-\tau)dtd\tau,$$

and then

$$\mu = \int_0^1 \int_0^1 k^*(r,r)dr, \quad \nu = \int_0^1 \int_0^1 \left[k^*(r,s)\right]^2 drds.$$

Explicit expressions for the constants μ and ν appearing in Theorem 1 can be obtained for the sharp origin kernel k_ρ and these are given in the Corollary below.

Corollary 1 *When $k = k_\rho$ in Theorem 1, the constants μ and ν have the explicit forms*

$$\begin{aligned} \mu_\rho &= \frac{\rho}{\rho+2}, \\ \nu_\rho &= \left(\frac{2}{\rho+2}\right)^2 + \frac{1}{\rho+1} - \frac{2}{(\rho+1)^2}\left(4 + \frac{2}{2\rho+3} - \frac{8}{\rho+2} - 2\frac{\Gamma^2(\rho+2)}{\Gamma(2\rho+4)}\right). \end{aligned}$$

Theorem 1 tells us that $\widehat{\Omega}_k/\mu \to_d \xi_\Omega := \Lambda\Xi\Lambda'/\mu$ and

$$\begin{aligned} E\xi_\Omega &= \Omega, \\ \text{var}\left(\text{vec}(\xi_\Omega)\right) &= \nu\mu^{-2}(I_{m^2} + K_{mm})\Omega \otimes \Omega. \end{aligned} \tag{15}$$

Hence, $\widehat{\Omega}_k/\mu$ is asymptotically unbiased with asymptotic variance matrix $\nu\mu^{-2}(I_{m^2} + K_{mm})\Omega \otimes \Omega$.

In seeking a preferred kernel, it might first appear reasonable to search for a kernel that minimizes the scale factor in (15), viz.,

$$\nu\mu^{-2} = \left(\int k(r-s)k(p-q) - 2\int k(r-s)k(r-q) + \int k(r-s)^2\right)\left(1 - \int_0^1 k(r-s)\right)^{-2}, \tag{16}$$

subject to the constraints that (a) $k(0) = 1$, and (b) $k(x)$ is positive semi-definite. This minimization problem is not well defined. Consider, for example, the class of SO kernels. Then, using μ_ρ and ν_ρ in (16) we get $\nu_\rho\mu_\rho^{-2} = f(\rho)$ with

$$f(\rho) = \left(\left(\frac{2}{\rho+2}\right)^2 + \frac{1}{\rho+1} - \frac{2}{(\rho+1)^2}\left(4 + \frac{2}{2\rho+3} - \frac{8}{\rho+2} - 2\frac{\Gamma^2(\rho+2)}{\Gamma(2\rho+4)}\right)\right)\left(\frac{\rho}{\rho+2}\right)^{-2}.$$

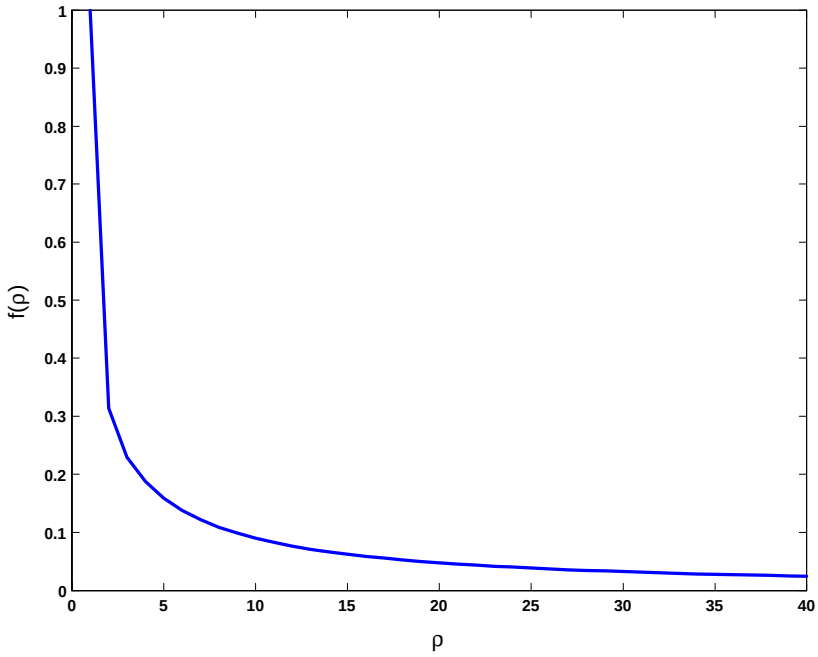

Figure 2: Scale Function $f(\rho)$ in the Asymptotic Variance of $\widehat{\Omega}/\mu$ with Kernel k_ρ

As shown in Fig.2, $f(\rho)$ is a decreasing function of ρ and it is easily seen that $\lim_{\rho\to\infty} f(\rho) = 0$. So, the asymptotic variance of the HAC estimate can be made arbitrarily small by taking an arbitrarily large ρ in the kernel k_ρ.

Below we will show that it is possible to choose an optimal value of ρ by taking into account higher order terms in the asymptotic bias of $\widehat{\Omega}_k/\mu$ in conjunction with the asymptotic variance.

4 Consistent HAC Estimation with SO Kernels

4.1 Some New Asymptotics with $\rho \to \infty$

This section develops an asymptotic theory for the HAC estimator $\widehat{\Omega}_{k_\rho}$ when $\rho \to \infty$ as $T \to \infty$. Under certain rate conditions on ρ, we show that $\widehat{\Omega}_{k_\rho}$ is consistent for Ω and has a limiting normal distribution. Thus, consistent HAC estimation is possible even when the bandwidth is set equal to the sample size. Of course, as is apparent from the graphs in Fig. 1, the action of ρ passing to infinity plays a role similar to that of a bandwidth parameter in that very high order autocorrelations are progressively downweighted as $T \to \infty$.

It is convenient to start the analysis with the HAC estimator that uses the true regression errors u_t rather than the regression residuals $\widehat{u}_t$. Accordingly, let $\widetilde{\Omega}_{k_\rho}$ be this pseudo-estimator, which is identical to $\widehat{\Omega}_{k_\rho}$ except that it is based on the unobserved

sequence $v_t = v_t(\beta)$ rather than $v_t = v_t(\widehat{\beta})$, i.e.

$$\widetilde{\Omega}_{k_\rho} = \sum_{j=-T+1}^{T-1} k_\rho(\frac{j}{T})\,\widetilde{\Gamma}(j),$$

where

$$\widetilde{\Gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j} v_{t+j}(\beta)v_t'(\beta) & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} v_{t+j}(\beta)v_t'(\beta) & \text{for } j < 0 \end{cases}.$$

The spectral matrix of v_t is $f_{vv}(\lambda)$, and $\Omega = 2\pi f_{vv}(0)$. Define

$$f^{(1)} = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} |h|\Gamma(h), \quad \Omega^{(1)} = 2\pi f^{(1)},$$

and

$$MSE(\rho, \widetilde{\Omega}_{k_\rho}, W) = \rho E\left\{vec(\widetilde{\Omega}_{k_\rho} - \Omega)'W\, vec(\widetilde{\Omega}_{k_\rho} - \Omega)\right\},$$

for some $m^2 \times m^2$ weight matrix W.

The following conditions help in the development of the asymptotic theory. The linear process and summability conditions given in **A2** ensure that the matrix $f^{(1)}$ is well defined and enable the use of standard formulae for the covariance properties of periodogram ordinates. **A3** controls the allowable expansion rate of ρ as $T \to \infty$ so that $\rho = o(T/\log T)$. It will often be convenient to set $\rho = aT^b$ for some $a > 0$ and $0 < b < 1$. The optimal expansion rate for ρ is found later to be of this form with $b = 2/3$.

A2: *$v_t = v_t(\beta)$ is a mean zero, fourth order stationary linear process*

$$v_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j^{1+\Delta} \parallel C_j \parallel < \infty, \quad \text{for some } \Delta > 0, \tag{17}$$

where ε_{t-j} is $iid(0, \Sigma_\varepsilon)$ with $E \parallel \varepsilon_t \parallel^4 < \infty$.

A3: $\frac{1}{\rho} + \frac{\rho \log T}{T} \to 0$, *as* $T \to \infty$.

Define the spectral window

$$K_\rho(\lambda) = \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})e^{i\lambda h} \tag{18}$$

corresponding to the SO kernel k_ρ. Analogous to $\widetilde{\Omega}_{k_\rho}$, we define the spectral estimate $\widetilde{f}_{vv}(0) = \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widetilde{\Gamma}(h)$ and let $\{\lambda_s = \frac{2\pi s}{T};\ s = 0, 1, ..., T-1\}$ be the Fourier frequencies and $I_{vv}(\lambda_s)$ be the periodogram of v_t. Using the inversion formula $\widetilde{\Gamma}(h) = \frac{2\pi}{T}\sum_{s=0}^{T-1} I_{vv}(\lambda_s)e^{i\lambda_s h}$, we deduce the smoothed periodogram form of this estimate, viz.,

$$\widetilde{f}_{vv}(0) = \frac{1}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)I_{vv}(\lambda_s), \tag{19}$$

with a corresponding formula for $\widetilde{\Omega}_{k_\rho} = 2\pi\widetilde{f}_{vv}(0)$. It is apparent that the limit behavior of these two quantities depends on the spectral window $K_\rho(\lambda_s)$, whose asymptotic form as $T \to \infty$ is given in the next result.

Lemma 1 *Let $\rho = aT^b \to \infty$ for some $a > 0$ and $0 < b < 1$. Then, for all $\lambda_s = \frac{2\pi s}{T}$, $s = 0, 1, ..., [T/2]$, we have as $T \to \infty$*

$$K_\rho(\lambda_s) = \frac{2\rho T}{2T^2(1-\cos\lambda_s)+\rho^2}\left[1+o(1)\right] \tag{20}$$

$$= \begin{cases} \frac{2\rho T}{(2\pi s)^2+\rho^2}\left[1+o(1)\right], & s = o(T); \\ \frac{\rho}{T(1-\cos(\kappa\pi))}\left[1+o(1)\right], & s = \left[\frac{T\kappa}{2}\right],\ \kappa \in (0,1]. \end{cases} \tag{21}$$

Since $K_\rho(\lambda_s) = K_\rho(-\lambda_s) = K_\rho(-\lambda_s + 2\pi)$, it follows from (20) and (21) that

$$K_\rho(\lambda_s) = \begin{cases} O\left(\frac{T}{\rho}\right) & s \le \rho \text{ and } s \ge T-\rho; \\ O\left(\frac{\rho T}{s^2}\right) & \rho < s < T-\rho. \end{cases} \tag{22}$$

So, for frequencies λ_s in the vicinity of the origin such that $\lambda_s = \frac{2\pi s}{T} = O\left(\frac{\rho}{T}\right)$ with ρ satisfying **A3**, the spectral window $K_\rho(\lambda_s) = O\left(\frac{T}{\rho}\right)$ diverges, while for all frequencies $\lambda_s \to \lambda \in (0, 2\pi)$, $K_\rho(\lambda_s) = O\left(\frac{\rho}{T}\right) = o(1)$. Thus, Lemma 1 shows that when $\rho \to \infty$ the sharp origin spectral estimate (19) effectively smooths periodogram ordinates in the neighborhood of the origin by downweighting frequencies that are removed from the origin (and 2π).

In comparison to (20), the spectral window of the Bartlett kernel ($\rho = 1$) is well known (e.g Priestley, 1981, p. 400) to be given by the exact formula

$$K_1(\lambda) = \sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)\cos(h\lambda) = \frac{1}{T}\frac{\sin^2\left(\frac{T\lambda}{2}\right)}{\sin^2\left(\frac{\lambda}{2}\right)} = 2\pi F_T(\lambda), \tag{23}$$

where $F_T(\lambda)$ is Fejer's kernel. Fig. 3 compares the spectral windows (23) and (18) when $T = 10$ for various ρ.

Note that the side lobes of the Fejer kernel are smoothed out in the sharp origin spectral window even for $\rho = 2$, as we expect from the asymptotic approximation (21). The peaks in the spectral windows at the origin reduce and the window becomes flatter as ρ increases (for fixed T) because $K_\rho(0) = O\left(\frac{T}{\rho}\right)$, as is clear from (22).

The following theorem describes the limit behavior of $\widetilde{\Omega}_{k_\rho}$ and gives the asymptotic form of the mean squared error $MSE(\rho, \widetilde{\Omega}_{k_\rho}, W)$.

Theorem 2 *Suppose **A1 - A3** hold and $\rho = aT^b \to \infty$ for some $a > 0$ and $0 < b < 1$. Then:*

(a) $\lim_{T\to\infty} \rho \mathrm{Var}\left(vec(\widetilde{\Omega}_{k_\rho})\right) = (I + K_{mm})\Omega \otimes \Omega$.

(b) If $b < \frac{2}{3}$ then

$$\sqrt{\rho}\left(vec(\widetilde{\Omega}_{k_\rho}) - vec(\Omega)\right) \to_d N\left(0, (I+K_{mm})\Omega\otimes\Omega\right).$$

(c) $\lim_{T\to\infty}\left(\frac{T}{\rho}\right)\left(E\widetilde{\Omega}_{k_\rho} - \Omega\right) = -\Omega^{(1)}$.

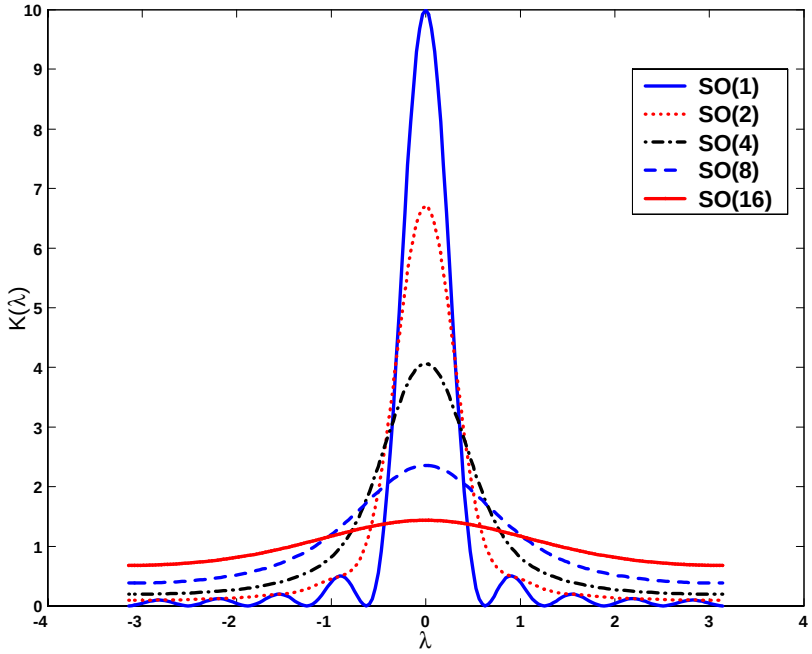

Figure 3: Spectral Window $K_\rho(\lambda)$ for Bartlett ($\rho = 1$) and Sharp Origin Kernels k_ρ with $\rho = 2, 4, 8, 16$ and $T = 10$.

(d) If $\rho^3/T^2 \to \vartheta \in (0, \infty)$, *then*

$$\begin{aligned} &\lim_{T\to\infty} \text{MSE}(\rho, \widetilde{\Omega}_{k_\rho}, W) \\ = \;& \vartheta \; vec\left(\Omega^{(1)}\right)' W vec\left(\Omega^{(1)}\right) + tr\; \{W(I + K_{mm})\Omega \otimes \Omega\}. \end{aligned}$$

Remarks

(a) It is not surprising that the results in Theorem 2 are similar to those for conventional HAC estimates as given, for example, in Andrews (1991). Fig. 4 shows the spectral window $K_\rho(\lambda)$ corresponding to the sharp origin kernel k_ρ with $\rho(T) = O(T^{2/3})$ for various values of T over the domain $(-\pi, \pi)$. Apparently, $K_\rho(\lambda)$ becomes successively more concentrated at the origin as ρ and T increase, so that the overall effect in this approach is analogous to that of conventional HAC estimation where increases in the bandwidth parameter M ensure that the band of frequencies narrows as $T \to \infty$.

(b) Part (b) of Theorem 2 gives a CLT for the new HAC estimator $\widetilde{\Omega}_{k_\rho}$. $\widetilde{\Omega}_{k_\rho}$ is computed using a full set of frequencies as is apparent from (19), but since $\rho \to \infty$ as $T \to \infty$, the spectral window becomes more concentrated at the origin and 2π, as we have seen. The proof of part (b) effectively shows that intermediate frequencies may

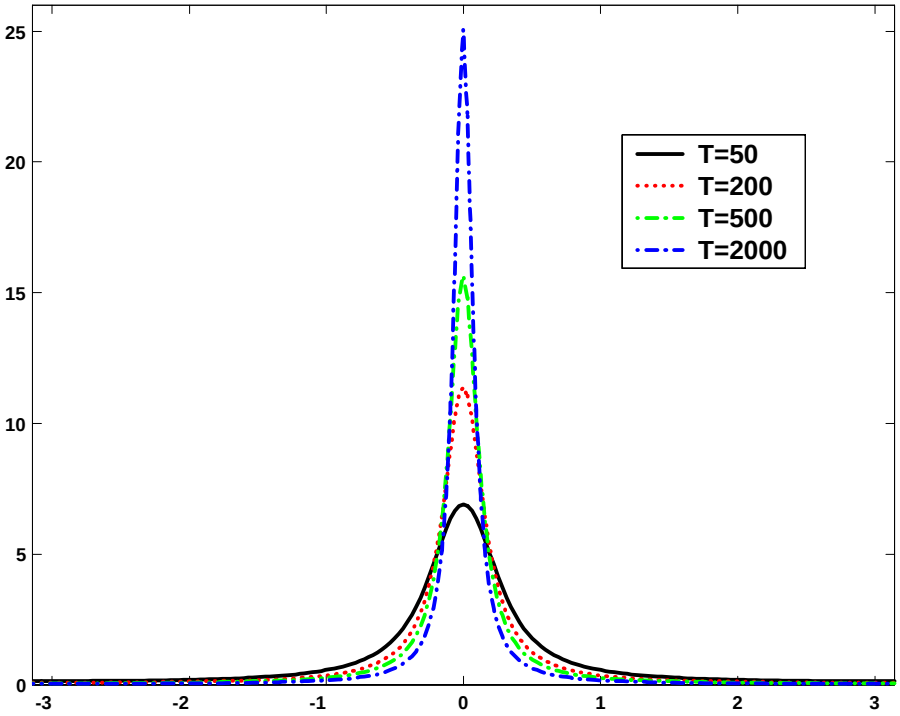

Figure 4: Spectral Window $K_\rho(\lambda)$ of the Sharp Origin Kernel with $\rho = O(T^{2/3})$

be neglected as $T \to \infty$ and that a CLT follows in a manner analogous to what happens when only a narrow band of frequencies is included (c.f., Robinson, 1995).

Next, we give a corresponding result for the feasible HAC estimator $\widehat{\Omega}_{k_\rho}$, showing that essential asymptotic properties are unaffected by the presence of the parametric estimation error arising from the use of the regression residuals in $\widehat{v}_t = v_t(\widehat{\beta})$. In this development, it is convenient to work with a truncated MSE as in Andrews (1991), viz.

$$MSE_h(\rho, \widehat{\Omega}_{k_\rho}, W_T) = E \ \min \left\{ \rho vec(\widehat{\Omega}_{k_\rho} - \Omega)' W_T vec(\widehat{\Omega}_{k_\rho} - \Omega), h \right\},$$

where W_T is a (possibly random) $m^2 \times m^2$ weight matrix that is positive semi-definite (almost surely). The asymptotic form of MSE_h when $T \to \infty$ and $h \to \infty$ is given in the following theorem. Use of MSE_h helps to avoid the effects of heavy tails in coefficient estimation on the criterion. Some additional regularity conditions are needed in this case and are based on those used in Andrews (1991). These are detailed in Assumption **B** prior to the proof of the following theorem in the Appendix.

Theorem 3 *Suppose* ***A1 - A3*** *and* ***B*** *hold. Suppose* $\rho^3/T^2 \to \vartheta \in (0, \infty)$ *as* $T \to \infty$. *Then:*

(a) $\sqrt{\rho}(\widehat{\Omega}_{k_\rho} - \Omega) = O_p(1)$, $\sqrt{\rho}(\widehat{\Omega}_{k_\rho} - \widetilde{\Omega}_{k_\rho}) \to_p 0$; *and*

(b)

$$\begin{aligned} & \lim_{h\to\infty}\lim_{T\to\infty} MSE_h(\rho,\widehat{\Omega}_{k_\rho},W) \\ = & \lim_{h\to\infty}\lim_{T\to\infty} MSE_h(\rho,\widetilde{\Omega}_{k_\rho},W) \\ = & \lim_{T\to\infty} MSE(\rho,\widetilde{\Omega}_{k_\rho},W) \\ = & \vartheta\ vec\left(\Omega^{(1)}\right)' W\, vec\left(\Omega^{(1)}\right) + tr\{W(I+K_{mm})\Omega\otimes\Omega\}\,. \end{aligned} \tag{24}$$

4.2 Optimal Power Parameters

As in optimal bandwidth selection in spectral density and HAC estimation, the criterion MSE_h can be used to determine a value of the power parameter ρ that is optimal in the sense that it minimizes the asymptotic truncated MSE for some given sequence of weight matrices W_T that converge in probability to a positive semi-definite limit matrix W. Let

$$\delta = \delta\left(\Omega,\Omega^{(1)}\right) := \frac{tr\left[W(I+K_{mm})\Omega\otimes\Omega\right]}{2vec\left(\Omega^{(1)}\right)' W\, vec\left(\Omega^{(1)}\right)}. \tag{25}$$

Then, using (24), the optimal ρ is

$$\begin{aligned} \rho_T^* &= \arg\min_\rho \left\{ \frac{\rho^2}{T^2} vec\left(\Omega^{(1)}\right)' W\, vec\left(\Omega^{(1)}\right) + \frac{1}{\rho} tr\left[W(I+K_{mm})\Omega\otimes\Omega\right] \right\} \\ &= \delta^{1/3} T^{2/3}. \end{aligned} \tag{26}$$

When $\rho = \rho_T^*$, the truncated MSE of $\widehat{\Omega}_{k_\rho}$,

$$E\min\left\{ vec(\widehat{\Omega}_{k_\rho}-\Omega)' W_T vec(\widehat{\Omega}_{k_\rho}-\Omega), h \right\},$$

converges to zero at the rate $O_p\left(T^{-2/3}\right)$. This rate is the same as that of the MSE of the conventional truncated Bartlett kernel estimate of Ω where the bandwidth (rather than the power parameter) is chosen to minimize MSE (c.f., Hannan, 1970; Andrews, 1991). Thus, $\widehat{\Omega}_{k_{\rho_T^*}}$ may be expected to have asymptotic performance characteristics similar to those of conventional consistent HAC estimates with optimal bandwidth choices.

The selection ρ_T^* will lead to HAC estimates $\widehat{\Omega}_{k_\rho}$ that are preferred in this class, at least in terms of asymptotic MSE performance. Of course, since ρ_T^* is an infeasible choice because it depends on unknown parameters, practical considerations suggest the use of a plug-in procedure that utilizes the form of (26) in conjunction with preliminary estimates of Ω and $\Omega^{(1)}$ in δ. The plug-in method used here is parametric and is based on the use of simple parametric models for Ω, as suggested in Andrews (1991), Andrews and Monahan (1992) and Lee and Phillips (1994). Model selection methods such as BIC and PIC can be used to assist in finding an appropriate parametric model whose estimates are then used to compute $\widehat{\Omega}$ and $\widehat{\Omega}^{(1)}$, which are then plugged into (25) and (26) to produce the data-determined value $\widehat{\rho}_T^* = \widehat{\delta}^{1/3} T^{2/3}$, where $\widehat{\delta} = \delta\left(\widehat{\Omega},\widehat{\Omega}^{(1)}\right)$. Of course, prefiltering is also an option in practical work.

In applications, the AR(1) is a commonly used simple parametric model for the plug-in method in bandwidth choice for conventional HAC estimation. In this case, if the assumed models are m univariate AR(1) processes and W_T gives weight (w_i) only to the diagonal elements of $\widehat{\Omega}_{k_\rho}$, we have

$$\widehat{\delta} = \sum_{i=1}^{m} w_i \frac{\widehat{\sigma}_i^4}{(1-\widehat{\alpha}_i)^4} \Big/ \sum_{i=1}^{m} w_i \frac{4\widehat{\alpha}_i^2 \widehat{\sigma}_i^4}{(1-\widehat{\alpha}_i)^6 (1+\widehat{\alpha}_i)^2}, \tag{27}$$

where

$$\widehat{\alpha}_i = \frac{\sum_{t=2}^{T} \widehat{v}_{t,i} \widehat{v}_{t-1,i}}{\sum_{t=2}^{T} \widehat{v}_{t-1,i}^2}, \text{ and } \widehat{\sigma}_i^2 = \frac{\sum_{t=2}^{T} \left(\widehat{v}_{t,i} - \widehat{\alpha}_i \widehat{v}_{t-1,i}\right)^2}{T-1}, \tag{28}$$

$\widehat{v}_t = x_t(y_t - x_t'\widehat{\beta})$, $\widehat{v}_{t,i}$ is i-th element of $\widehat{v}_t$, and $\widehat{\beta}$ is defined in (2). As in conventional bandwidth estimation, we find that use of the AR(1) plug-in method produces very reasonable results even when the true model is not an AR(1). The reason for some robustness is that $\widehat{\rho}_T^*$ still produces the optimal rate of expansion $T^{2/3}$ of the power parameter even when the approximating model is wrong. Further, investigation reveals that the MSE functions are generally U-shaped functions of the sharp power parameter ρ and this function is often fairly flat in the neighborhood of ρ_T^*, so good performance in HAC estimation is often achieved when the power parameter ρ is in the general vicinity of ρ_T^*. Similar behavior was found in Andrews' (1991) exploration of the properties of the plug-in method for bandwidth choice.

To illustrate this data-determined rule, we assume that the true model for v_t is

$$v_t = a v_{t-1} + \epsilon_t + b\epsilon_{t-1},\ \epsilon_t \sim iid(0, \sigma^2) \tag{29}$$

for which we have

$$\Omega = \frac{(1+b)^2}{(1-a)^2}\sigma^2, \quad \Omega^{(1)} = \frac{2(1+ab)(a+b)}{(1-a)^3(1+a)}\sigma^2,$$

and

$$\delta = \left(\frac{\Omega}{\Omega^{(1)}}\right)^2 = \left(\frac{(1-a^2)(1+b)^2}{2(1+ab)(a+b)}\right)^2.$$

Note that when $a = -b$, (29) reduces to $v_t = \epsilon_t$. Then, $\Omega = \sigma^2$, $\Omega^{(1)} = 0$, and $\delta = \infty$. As a consequence, ρ^* approaches infinity and $k_{\rho^*}(x) = 0$ for all $x \neq 0$. In this case, $\widehat{\Omega}_{k_{\rho^*}} = \widehat{\Gamma}(0)$ and all sample covariances are given zero weight in the sharp origin HAC estimator.

Using an AR(1) plug-in method as in (28) above, we get

$$\text{plim}_{T\to\infty}\hat{\alpha} = \frac{E(v_t v_{t-1})}{E(v_t^2)} = \frac{(a+b)(1+ab)}{1+b^2+2ab} := \bar{\alpha}, \text{ say,} \tag{30}$$

and so the corresponding data-determined power parameter is $\hat{\rho}_T^* = \bar{\delta}^{1/3} T^{2/3}[1 + o(1)]$ where

$$\bar{\delta} = \frac{(1-\bar{\alpha}^2)^2}{4\bar{\alpha}^2},$$

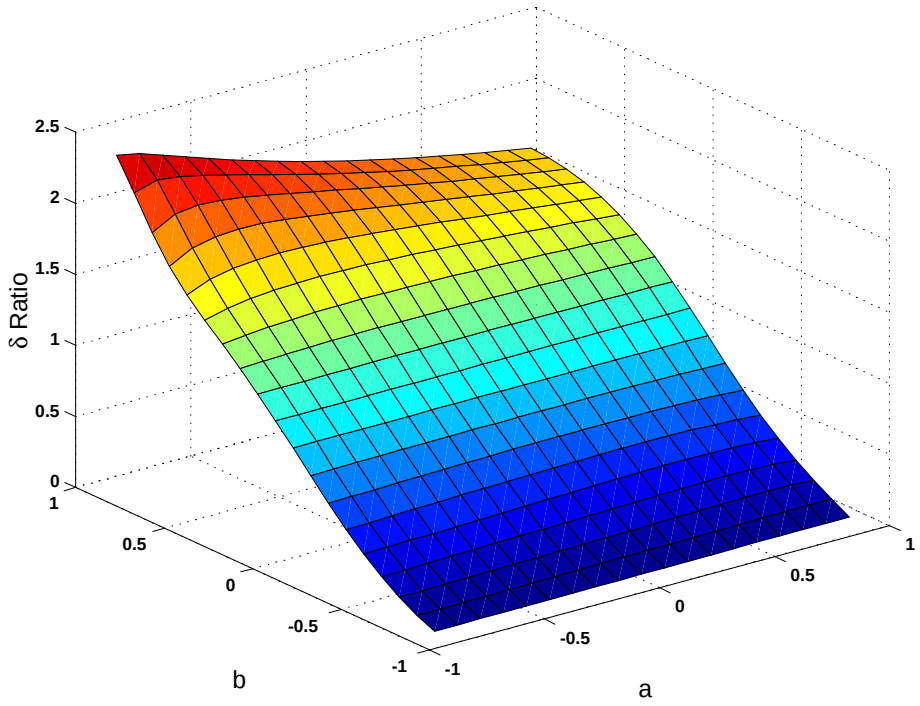

Figure 5: Graph of $\left(\delta/\bar{\delta}\right)^{1/3}$

from (27) and (30). The difference between the optimal choice of ρ, viz. $\rho_T^* = \delta^{1/3}T^{2/3}$, and the data-determined value $\hat{\rho}_T^*$ therefore depends essentially on the difference between $\delta = \delta\left(\Omega, \Omega^{(1)}\right)$ and $\bar{\delta}$. In the present case, we get

$$\frac{\delta}{\bar{\delta}} = \left(\frac{(1-a^2)\,(1+b)^2\,\bar{\alpha}}{(1+ab)(a+b)(1-\bar{\alpha}^2)}\right)^2 = \left(\frac{(1-a^2)\,(1+b)^2\left(1+b^2+2ab\right)}{(1+b^2+2ab)^2-(a+b)^2(1+ab)^2}\right)^2.$$

Fig. 5 presents the surface plot of the ratio $\left(\delta/\bar{\delta}\right)^{1/3}$ against a and b. When $b = 0$, we have $\delta/\bar{\delta} = 1$ for all a, as expected. The ratio $\delta/\bar{\delta} < 1$ for $b < 0$ and is increasing in b, taking very similar values for various $a \geq 0$. When $b > 0$, the ratio $\delta/\bar{\delta} > 1$ and is increasing with b but increases more slowly for larger values of a. Accordingly, we can expect the plug-in value to be less (greater) than the optimal value of ρ when there are positive (respectively, negative) moving average effects. When $b \to -1$, the LRV $\Omega \to 0$ and, correspondingly, $\delta/\bar{\delta} \to 0$, so that the optimal rate for ρ_T^* is no longer $O\left(T^{2/3}\right)$ when $\Omega = 0$, although the plug-in value remains $O(T^{2/3})$. Hence, the plug-in rule becomes progressively less satisfactory as $b \to -1$. For the degenerate case $\Omega = 0$ (or singular Ω in the matrix case), it is known that different asymptotics, bandwidth choices and optimal rates apply in conventional HAC estimation (c.f., Lemma 8.1 and the following discussion in Phillips, 1995). Similar considerations can be expected to be relevant for sharp origin kernels without truncation when $\Omega = 0$, although such a development is beyond the scope of the present paper.

In finite samples, the optimal value, ρ_T^f, of ρ may be calculated directly by simulation.

Table 1 shows the values of ρ_T^f calculated from 50,000 replications against those of the plug-in estimate $\widehat{\rho}_T^*$ and ρ_T^*, both of which are based on asymptotic theory, for various parameter configurations and sample sizes[2]. Apparently, when $b > 0$ both $\widehat{\rho}_T^*$ and ρ_T^* underestimate ρ_T^f, whereas $\widehat{\rho}_T^*$ tends to overestimate ρ_T^f and ρ_T^* when $b < 0$.

Table 1: Asymptotic ρ_T^*, Plug-in $\widehat{\rho}_T^*$, and Finite Sample ρ_T^f
$v_t = av_{t-1} + \epsilon_t + b\epsilon_{t-1}$, $\epsilon_t \sim iid(0,1)$

		$T=50$			$T=100$			$T=200$		
a	b	ρ_T^*	$\widehat{\rho}_T^*$	ρ_T^f	ρ_T^*	$\widehat{\rho}_T^*$	ρ_T^f	ρ_T^*	$\widehat{\rho}_T^*$	ρ_T^f
0.1	0.1	28	25	42	44	39	64	71	62	92
0.1	0.3	22	16	30	35	25	44	55	39	65
0.1	0.5	20	12	26	32	19	40	50	30	60
0.1	0.9	19	10	24	30	16	37	48	25	57
0.1	−0.3	16	25	18	25	40	28	40	64	44
0.1	−0.5	6	17	6	10	26	10	16	42	16
0.5	0.1	11	10	16	17	15	23	28	24	35
0.5	0.3	11	8	15	17	12	22	27	19	33
0.5	0.5	10	7	15	17	11	21	26	17	34
0.5	0.9	9	6	15	16	10	21	26	15	33
0.5	−0.3	14	23	24	23	37	34	36	58	50
0.9	0.1	3	3	4	5	4	7	8	7	11
0.9	0.3	3	2	4	5	4	7	8	6	11
0.9	0.5	3	2	4	5	3	6	8	5	11
0.9	0.9	3	2	4	5	3	6	8	5	11
0.9	−0.3	3	5	4	5	8	7	8	13	11
0.9	−0.5	3	8	5	5	13	7	8	21	11

Table 2 compares the MSE performance of the HAC estimates based on the plug-in value, $\widehat{\rho}_T^*$, the exact asymptotic value, ρ_T^*, and the finite sample optimal value, ρ_T^f, of ρ. The outcomes are given for $T = 50$ and show that the MSE's are generally close, especially when the moving average effect is small or when the autoregressive coefficient is large. The final column of Table 2 gives the ratio $\text{MSE}_{\widehat{\rho}_T^*}/\text{MSE}_{\rho=1}$. The ratio is small over a wide range of parameter values, indicating that on a MSE criterion, use of a sharp origin kernel with a data-determined power parameter will lead to considerable improvements in HAC estimation over use of the Bartlett kernel. Only when b is large and negative ($b \leq -0.5$) is the ratio greater than unity. In such cases, as remarked above, $\widehat{\rho}_T^*$ tends to overestimate ρ_T^f, and it seems likely that use of model selection procedures in finding an appropriate ARMA model for use in the plug-in rule (rather than the mechanical use of an AR(1) model) would lead to improvements, although this has not been investigated.

[2]When $a = -b$, both δ and $\bar{\delta}$ are undefined and $\rho = \infty$. In this case the estimate $\widehat{\Omega}_{k_{\rho^*}} = \widehat{\Gamma}(0)$, so the sharp origin HAC estimate is well defined and is simply the sample variance. However, in this case both $\text{MSE}_{\widehat{\rho}_T^*}$ and $\text{MSE}_{\rho_T^*}$, which are based on asymptotic formulae, are undefined because $\widehat{\rho}_T^* = \rho_T^* = \infty$. We therefore do not include entries for parameter configurations in which $a = -b$ in Tables 1 and 2.

Similar results (not shown) were obtained for larger values of T. As is apparent from Fig. 6, the MSE curve is quite flat around the optimal value, especially when $T = 200$. Fig. 7 shows that the MSE curves based on the asymptotic formula have the shapes similar to those computed by simulation. The case shown in these figures for $a = 0.5$, $b = 0.3$ is typical of other parameter configurations.

Table 2: Ratios of the MSE's of HAC Estimators for Different Choices of ρ with $T = 50$.

a	b	$\text{MSE}_{\widehat{\rho}_T^*}/\text{MSE}_{\rho_T^f}$	$\text{MSE}_{\widehat{\rho}_T^*}/\text{MSE}_{\rho_T^*}$	$\text{MSE}_{\widehat{\rho}_T^*}/\text{MSE}_{\rho=1}$
0.1	0.1	1.16	1.02	0.05
0.1	0.3	1.25	1.11	0.08
0.1	0.5	1.43	1.22	0.09
0.1	0.9	1.56	1.34	0.11
0.1	−0.3	1.08	1.28	0.12
0.1	−0.5	2.02	2.47	0.58
0.5	0.1	1.13	1.02	0.14
0.5	0.3	1.21	1.10	0.16
0.5	0.5	1.34	1.18	0.17
0.5	0.9	1.46	1.26	0.18
0.5	−0.3	1.00	1.28	0.13
0.9	0.1	1.03	1.01	0.49
0.9	0.3	1.13	1.08	0.53
0.9	0.5	1.14	1.15	0.56
0.9	0.9	1.13	1.20	0.59
0.9	−0.3	1.00	1.33	0.64
0.9	−0.5	1.079	2.69	1.29

5 Hypothesis Testing Using HAC Estimator with SO Kernels

As in KV, we use a simple illustrative framework and consider regression tests of the null hypothesis $H_0 : R\beta = r$ against the alternative $H_1 : R\beta \neq r$. Using the estimate $\widehat{\Omega}_{k_\rho}$, we can construct the F-ratio in the usual way

$$F^*\left(\widehat{\Omega}_{k_\rho}\right) = T(R\widehat{\beta} - r)'\left(R\widehat{Q}^{-1}\widehat{\Omega}_{k_\rho}\widehat{Q}^{-1}R'\right)^{-1}(R\widehat{\beta} - r)/p, \tag{31}$$

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

$$t^*\left(\widehat{\Omega}_{k_\rho}\right) = T^{1/2}(R\widehat{\beta} - r)\left(R\widehat{Q}^{-1}\widehat{\Omega}_{k_\rho}\widehat{Q}^{-1}R'\right)^{-1/2}. \tag{32}$$

The limit distributions of F^* and t^* under the null hypothesis and local alternatives when ρ is fixed are given in the following result, which is formulated in a general way to allow for an arbitrary positive semi-definite kernel, thereby including cases where the

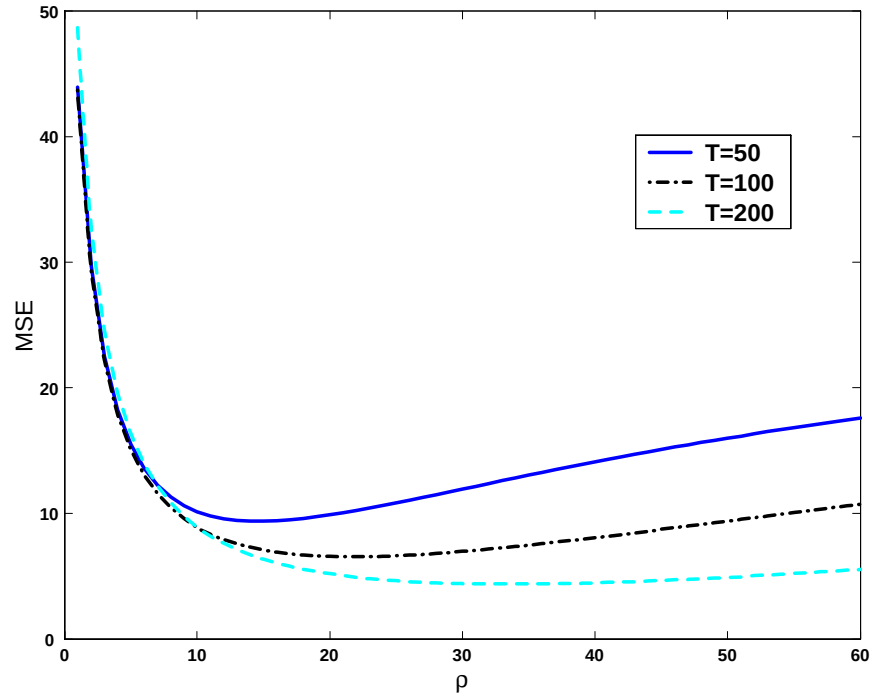

Figure 6: Finite Sample MSE when $v \sim ARMA(1,1)$ with $a = 0.5$ and $b = 0.3$

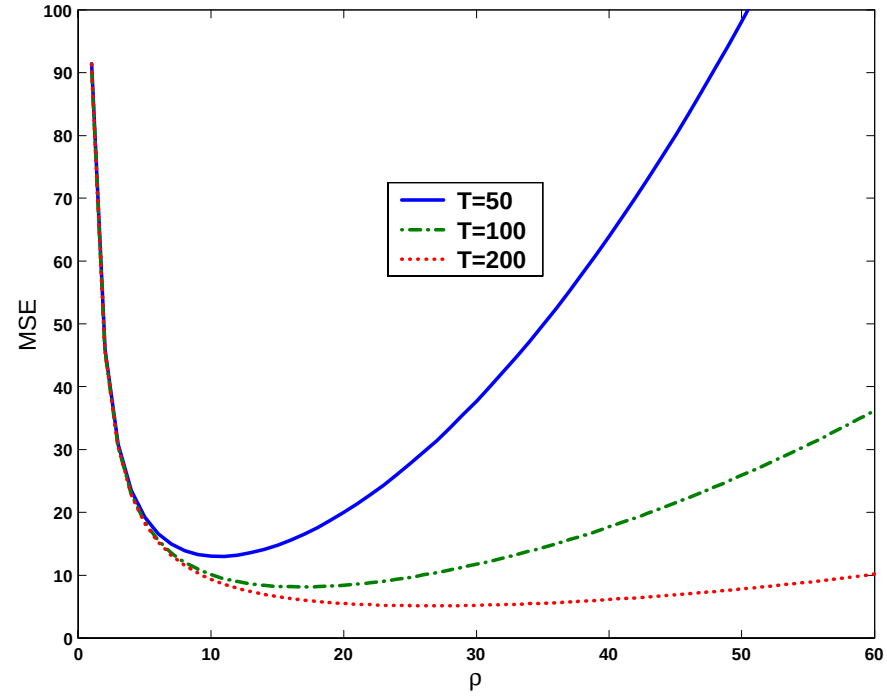

Figure 7: Asymptotic MSE when $v \sim ARMA(1,1)$ with $a = 0.5$ and $b = 0.3$

kernel is continuously differentiable at the origin and those, like the Bartlett and sharp origin kernels, where it has only one sided derivatives at the origin. Both statistics F^* and t^* lead to asymptotically similar tests as shown in the following result.

Theorem 4 *Let **A1** and **A2** hold. If $k(x)$ is positive semi-definite and ρ is fixed, then:*

(a) Under H_0

$$F^*\left(\widehat{\Omega}_k\right) \Rightarrow W_p'(1)\left(\int_0^1\int_0^1 k(r-s)dV_p(r)dV_p'(s)\right)^{-1} W_p(1)/p, \tag{33}$$

$$t^*\left(\widehat{\Omega}_k\right) \Rightarrow W_1(1)\left(\int_0^1\int_0^1 k(r-s)dV_1(r)dV_1'(s)\right)^{-1}. \tag{34}$$

(b) Under the local alternative $H_1 : R\beta = r + cT^{-1/2}$

$$F^*\left(\widehat{\Omega}_k\right) \Rightarrow \left(\Lambda^{*-1}c + W_p(1)\right)'\left(\int_0^1\int_0^1 k(r-s)dV_p(r)dV_p'(s)\right)^{-1}\left(\Lambda^{*-1}c + W_p(1)\right)/p, \tag{35}$$

$$t^*\left(\widehat{\Omega}_k\right) \Rightarrow (\gamma + W_1(1))\left(\int_0^1\int_0^1 k(r-s)dV_1(r)dV_1(s)\right)^{-1}, \tag{36}$$

where $\Lambda^*\Lambda^{*\prime} = RQ^{-1}\Omega Q^{-1}R'$ and $\gamma = c(RQ^{-1}\Omega Q^{-1}R')^{-1/2}$.

When ρ is sample size dependent and satisfies **A3**, we know from Theorem 2 that $\widehat{\Omega}_{k_\rho}$ is consistent. In this case, F^* and t^* have conventional chi-square and normal limits.

Theorem 5 *Let **A1**-**A3** hold. Then:*

(a) under the null hypothesis

$$pF^*\left(\widehat{\Omega}_{k_\rho}\right) \Rightarrow W_p'(1)W_p(1) =_d \chi_p^2,\ t^*\left(\widehat{\Omega}_{k_\rho}\right) \Rightarrow W_1(1) =_d N(0,1). \tag{37}$$

(b) under the local alternative hypothesis $H_1 : R\beta = r + cT^{-1/2}$

$$pF^*\left(\widehat{\Omega}_{k_\rho}\right) \Rightarrow \left(\Lambda^{*-1}c + W_p(1)\right)'\left(\Lambda^{*-1}c + W_p(1)\right),\ t^*\left(\widehat{\Omega}_{k_\rho}\right) \Rightarrow (\gamma + W_1(1)). \tag{38}$$

Remarks From the form of (33)-(36), the statistics F^* and t^* clearly have nonstandard limit distributions arising from the random limit of the HAC estimate when ρ is fixed as $T \to \infty$, just like the KV and Jansson (2002) tests. However, it is also apparent that as ρ increases, the effect of this randomness diminishes. In particular, since

$$k_\rho(r-s) = (1-|r-s|)^\rho \to \begin{cases} 1 & r = s \\ 0 & r \neq s \end{cases}, \quad \text{as } \rho \to \infty,$$

we have

$$\int_0^1\int_0^1 k_\rho(r-s)dV_p(r)dV_p'(s) \to_p \int_0^1 drI_p = I_p, \quad \text{as } \rho \to \infty, \tag{39}$$

in view of the fact that

$$dV_p(r)dV_p'(r) = d[V]_r = drI_p,$$

where $[V_p]_r$ is the quadratic variation (matrix) process of V_p. It follows from (39) that as $\rho \to \infty$ the limit distributions under the null and the alternative approach those of regression tests in which conventional consistent HAC estimates are employed. In consequence, we can expect the tests based on $\widehat{\Omega}_{k_\rho}$ with ρ large to have power similar to that of conventional tests. When $\rho \to \infty$, these tests will have power functions equivalent at the power envelope. Simulations, discussed below, show that these good properties seem to be achieved for quite moderate values of ρ in the range $\rho \in [10, 20]$.

Fig. 8 presents the asymptotic power curves computed by simulation. The Brownian motion and Brownian bridge processes were approximated using normalized partial sums of $T = 1000$ *iid* $N(0,1)$ random variables and the simulation involved 50,000 replications. Asymptotic power was computed for the t^* test at the 5% significance level using sharp origin kernels with $\rho = 1, 2, 4, 8, 16$ and for $\gamma \in [0, 5]$. Critical values of the test for different ρ values (including the asymptotic case which is represented as $\rho = O(T^{2/3})$) are reported in Table 3.

Table 3: Asymptotic Critical Values of the t*-test

Power Parameter	90.0%	95.0%	97.5%	99.0%
$\rho = 1$	2.735	3.767	4.796	6.195
$\rho = 2$	2.132	2.881	3.630	4.600
$\rho = 4$	1.761	2.339	2.902	3.624
$\rho = 8$	1.539	2.018	2.469	3.040
$\rho = 16$	1.418	1.840	2.232	2.694
$\rho = O(T^{2/3})$	1.282	1.645	1.960	2.326

As is apparent from Fig. 8, the power curve moves up uniformly as ρ increases, consonant with the asymptotic theory implied by (36) and (39). When $\rho = 16$, the power curve is very close to the power envelope (the asymptotic power curve when the true Ω or a consistent estimate is used). This is to be expected. When ρ is large, it may be regarded as being roughly compatible with the rate condition in **A3** (e.g., $\rho = 16$ and $T = 1000$ corresponds to $16 \simeq 1000^{0.4}$). In that case, the test statistic is effectively constructed using a consistent estimate of Ω and Theorem 5 applies.

Comparing the asymptotic powers for different ρ in Fig. 8, it is apparent that tests based on sharp origin kernels with $\rho > 1$ outperform those with the Bartlett kernel ($\rho = 1$). KV (2002b) show that the Bartlett kernel delivers the most powerful test within a group of popular kernels (including the Parzen, Tukey-Hanning, and quadratic spectral kernels) when the bandwidth is set to the sample size. Correspondingly, tests based on sharp origin kernels will dominate these other commonly used kernels also.

The reason for the domination by the Bartlett kernel found by KV is related to the argument given above for the sharp origin kernel domination. As is apparent from the general form of the power functions given in (35) and (36), the effect of the choice of kernel k on the power function is manifest in the quadratic functional $\int_0^1 \int_0^1 k(r - s)dV_p(r)dV_p'(s)$. Since k is generally decreasing away from the origin (for many kernels

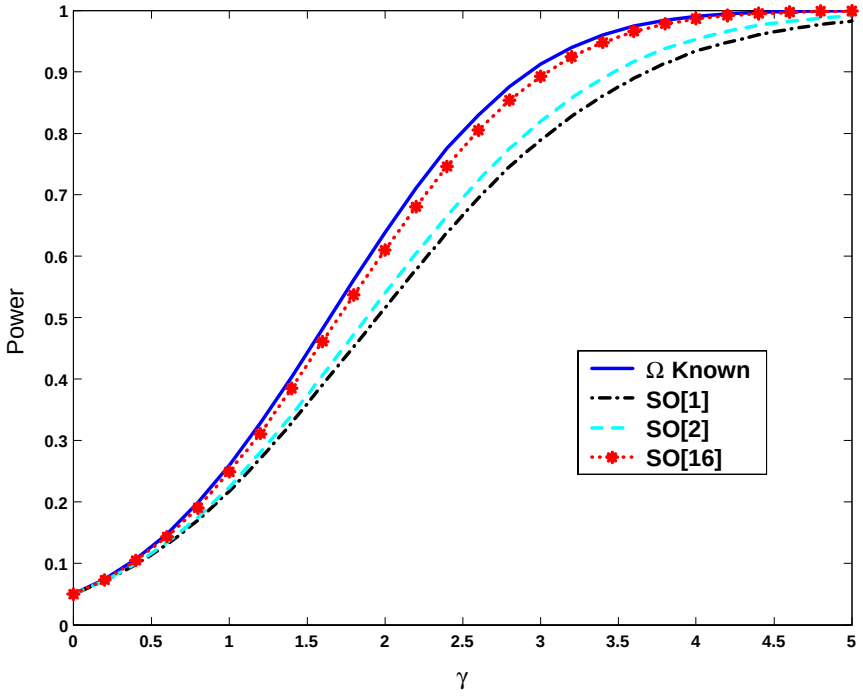

Figure 8: Asymptotic Local Power Function of the t^*-test

it is monotonically decreasing), the major contribution to the value of this functional comes from the neighborhood of the origin. Quadratic kernels (i.e. those kernels with Parzen exponent $q = 2$) have a quadratic shape at the origin with zero first derivative and decay more slowly than the Bartlett kernel (or sharp origin kernels), thereby generally increasing the value of the functional and reducing power (for any given realization of the process V_p).

6 Finite Sample Properties of the t^* Test

This section compares the finite sample performance of the t^* test with a sharp origin kernel for various values of the power parameter. The same data generating process (DGP) as that in KV (2002b) is used here, viz.,

$$y_t = \mu + x_t\beta + u_t,$$

where $\mu = 0$, $u_t = a_1 u_{t-1} + a_2 u_{t-2} + e_t$, $x_t = bx_{t-1} + \eta_t$, $b = 0.5$, e_t and η_t are $iid(0,1)$ with $\text{cov}(e_t, \eta_t) = 0$, and $x_0 = u_0 = u_{-1} = 0$. The simulation results are based on $50,000$ replications. We consider the one-sided null hypothesis $H_0 : \beta \leq 0$ against the alternative $H_1 : \beta > 0$. The regression parameter β is estimated by OLS and the t^* statistic is constructed as in (32). As a benchmark, we also construct the conventional (i.e., bandwidth truncated) t-statistic using the Bartlett kernel which we label t_{HAC}-$SO(1)$. In computing t_{HAC}-$SO(1)$, the bandwidth is chosen by the data driven procedure

proposed in Andrews (1991). We also report t_{HAC} and t^* using AR(1) prewhitening, as suggested by Andrews and Monahan (1992).

We first consider the case that ρ is fixed ($\rho = 1, 2, 4, 8, 16$) and $T = 50, 100$ and 200. Tables 4a and 4b present the finite sample null rejection probabilities with no prewhitening and with prewhitening, respectively. Rejections were determined using asymptotic 95% critical values from Table 3. We draw attention to three aspects of Tables 4a and 4b. First, in all cases the size distortions of the t^* tests are less than those of the t-test. This is true even for large ρ. Prewhitening reduces the difference in the size distortion between the two tests but it does not remove it. Second, the size distortion increases with ρ. (As stated in the previous paragraph, we only report the finite sample size of HAC tests constructed using the Bartlett kernel, which we found in our simulations have the least size distortion among HAC t-tests.) However, as T increases, the null rejection probabilities approach the nominal size for all cases. For $T = 200$, the increasing pattern of the size distortion as a function of ρ is hardly noticeable. Third, when the errors follow an AR(1) process, the size distortion of t_{HAC} becomes larger as a_1 approaches unity. Prewhitening greatly reduces the size distortion for both tests. In short, the asymptotic null approximation of the t^*-test is more accurate than that of the conventional robust t-test, and prewhitening generally improves the quality of the null approximation in both cases.

Figs. 9–10 show the finite sample (size adjusted) power of these tests in two cases where comparisons with the results of KV (2002b) are possible. The typical pattern that is evident in the figures is that the power of the t^*-test increases as ρ increases, just as asymptotic theory predicts. When $\rho = 16$, the power of the t^*-test is equivalent to or better than that of the conventional robust t-test using the Bartlett kernel.

Figures 9a and 9b depict the power for the DGP with $a_1 = 0.85, a_2 = 0.0$. As in KV, we found that the power of the t-test is not sensitive to the kernel used. So we present the power of the t-test only for the Bartlett kernel. Evidently, the power of the t-test is uniformly greater than that of the t^*-$SO(1)$ test, again as found in KV. However, when the sharp origin kernel with $\rho = 16$ is used, the power of the t^*-test (shown as the curve t^*-$SO(16)$ in the figures) slightly exceeds that of the t-test, particularly in the case where prewhitening is employed (Fig. 9b). This dominance is accentuated as ρ continues to increase (but in that event size distortion also increases, although it is still less than that of the t-test). Compared with Fig. 8, it seems that the finite sample power comparisons mimics the asymptotic results well, with larger ρ leading to increases in power. Figs. 10a and 10b show the power curves for the DGP with $a_1 = 1.9, a_2 = -0.95$. The observations made above continue to apply in this case, although the powers are closer, especially when prewhitening is used (Fig. 10b).

Next, we consider the performance of the t^* test when ρ is data-determined. Table 5 reports the asymptotic optimal ρ_T^*, the finite sample optimal $\widehat{\rho}_T^f$, and the plug-in optimal $\widehat{\rho}_T^*$ for both models. Here, Ω and $\Omega^{(1)}$ were obtained by direct computation, using their series representations.

Table 5: Asymptotic ρ_T^*, Finite Sample ρ_T^f, and Plug-in $\widehat{\rho}_T^*$ (50,000 repetitions) when $T = 50$

	$(a_1, a_2) = (0.85, 0)$	$(a_1, a_2) = (1.9, -0.95)$
ρ_T^*	13	13
ρ_T^f	44	27
$\widehat{\rho}_T^*$: mean	31	21
$\widehat{\rho}_T^*$: s.e.	193.6	68.1

Table 5 shows that for both models the plug-in estimate $\widehat{\rho}_T^*$ lies on average between the asymptotic optimal ρ_T^* and finite sample optimal ρ_T^f. However, the sampling distribution of $\widehat{\rho}_T^*$ shows considerable dispersion when T is small, as is clear from the standard errors given in Table 5 and kernel density estimates (not reproduced here). This variability decreases rapidly as T increases — when $T = 200$ the standard error is 14.9 for the first model and 25.5 for the second.

Fig. 11 shows the effects on (size adjusted) power of using the plug-in data-determined exponent $\widehat{\rho}_T^*$ compared with that based on use of the optimal exponents ρ_T^* and ρ_T^f. We report only one case here for brevity, the other parameter configurations and models giving very similar results. As is clear from the figure, the data-determined choice $\widehat{\rho}_T^*$ produces tests with essentially the same power functions. Thus, the variability in $\widehat{\rho}_T^*$ observed in Table 5 does not seem to adversely affect power. However, the variability does affect size adversely. Simulations (not reported here) show that size distortion of the t^*-test using the plug-in exponent $\widehat{\rho}_T^*$ is comparable to that of the conventional t-test (using data-determined consistent HAC estimates) for the various parameter configurations in Table 4a and 4b. The explanation seems to be that both large and small choices of $\widehat{\rho}_T^*$ can arise due to sampling variability and when $\widehat{\rho}_T^*$ is small the nominal asymptotic critical value is too small due to the greater variability in the denominator of the t^* statistic (c.f., Table 3).

In consequence, our findings indicate that use of the plug-in procedure with exponent $\widehat{\rho}_T^*$ does not provide unambiguous improvements in size and power in hypothesis testing over conventional robust procedures. Instead, fixed choices of ρ in the range $[10, 20]$ for sample sizes $T \in [50, 500]$ seem well suited for most practical work in econometrics. Simulations for the various parameter configurations of a_1, a_2 and b suggest that the finite sample power of the t^*-test using sharp parameter values $\rho \in [10, 20]$ is close to that of the conventional robust t-test, while providing clear improvements in size with or without prewhitening. Figs. 12 and Table 6 illustrate these effects for $\rho = 10, 16$ and 20 with no prewhitening. Choices of the exponent in this range stabilize size and do not give up power. Furthermore, in most cases $\rho \in [10, 20]$ lies in the flat area of the MSE curve, which suggest that it will generally be a good substitute for the optimal sharp parameter in practical work with time series sample sizes that are typical in economics. As T increases, it seems that larger ρ's in the range are preferred, while smaller ρ is sufficient when T is small. Unless T is very small (less than 50) we have found that $\rho = 16$ is a suitable choice.

7 Conclusion

The new class of sharp origin kernels introduced in this paper permit consistent HAC and LRV estimation without truncation and use an approach (based on a power parameter) that is different from conventional bandwidth controls to downweight autocorrelations at long lags. Within this class, the Bartlett kernel without truncation is the special case in which the power parameter is fixed at unity. When asymptotically similar regression tests are constructed with such kernels, the size distortion that commonly arises with conventional HAC estimation is reduced. Our findings indicate that as the power parameter increases, test power is enhanced and is arbitrarily close to and sometimes exceeds that of conventional tests, while retaining improvements in size. Data-determined choices of the power parameter are given which are easily implemented in practical work and which lead to HAC estimates with a convergence rate of $T^{1/3}$, analogous to that of a conventional truncated Bartlett kernel estimate with an optimal choice of bandwidth. Simulations show that in practice a simple fixed choice of the exponent parameter around $\rho = 16$ for the sharp origin kernel produces favorable results for both size and power in regression testing with sample sizes that are typical in econometric applications.

The general approach given here of using sharp origin kernels that are formed by taking power functions of conventional kernels without truncation or direct bandwidth control is obviously applicable when the mother kernel is a function other than the Bartlett kernel. It turns out, however, that some modifications of the approach (and the proofs of the limit theory) are required in a more general setting. As one might expect from conventional limit theory for spectral estimation, the optimal rates of divergence for the power parameter and rate of convergence of the corresponding data-driven HAC estimates depend on the choice of the mother kernel. Of course, extensions of the results are also possible to estimation of a spectral density at frequencies other than zero. Details of these extensions will be provided in subsequent work.

Table 4a: Finite Sample Null Hypothesis Rejection Probabilities
DGP's are the same as Table 3 in KV(2002)*, No Prewhitening

	a_1	a_2	$t_{\rho=1}$	$t^*_{\rho=1}$	$t^*_{\rho=2}$	$t^*_{\rho=4}$	$t^*_{\rho=8}$	$t^*_{\rho=16}$
$T=50$	−0.500	0.000	0.056	0.050	0.050	0.049	0.048	0.044
	0.000	0.000	0.071	0.061	0.061	0.062	0.062	0.063
	0.300	0.000	0.087	0.068	0.068	0.070	0.071	0.074
	0.500	0.000	0.099	0.072	0.073	0.075	0.078	0.083
	0.700	0.000	0.110	0.078	0.079	0.082	0.086	0.093
	0.900	0.000	0.122	0.082	0.084	0.088	0.094	0.105
	0.950	0.000	0.124	0.082	0.084	0.089	0.096	0.107
	0.990	0.000	0.130	0.081	0.084	0.090	0.099	0.112
	1.500	−0.750	0.109	0.073	0.074	0.078	0.084	0.094
	1.900	−0.950	0.139	0.080	0.083	0.090	0.101	0.115
	0.800	0.100	0.122	0.081	00.084	0.088	0.094	0.105
$T=100$								
	−0.500	0.000	0.050	0.050	0.050	0.050	0.049	0.047
	0.000	0.000	0.061	0.055	0.055	0.056	0.056	0.056
	0.300	0.000	0.076	0.058	0.060	0.061	0.063	0.064
	0.500	0.000	0.085	0.063	0.065	0.065	0.067	0.070
	0.700	0.000	0.093	0.067	0.068	0.070	0.071	0.076
	0.900	0.000	0.100	0.070	0.072	0.073	0.077	0.084
	0.950	0.000	0.104	0.073	0.074	0.075	0.079	0.086
	0.990	0.000	0.105	0.069	0.070	0.073	0.079	0.086
	1.500	−0.750	0.091	0.065	0.066	0.067	0.070	0.075
	1.900	−0.950	0.101	0.065	0.066	0.067	0.074	0.084
	0.800	0.100	0.100	0.071	0.072	0.073	0.077	0.083
$T=200$								
	−0.500	0.000	0.046	0.049	0.049	0.049	0.048	0.047
	0.000	0.000	0.055	0.054	0.054	0.054	0.053	0.054
	0.300	0.000	0.067	0.055	0.056	0.056	0.056	0.057
	0.500	0.000	0.072	0.056	0.057	0.057	0.058	0.060
	0.700	0.000	0.077	0.058	0.058	0.059	0.060	0.063
	0.900	0.000	0.083	0.061	0.061	0.062	0.063	0.067
	0.950	0.000	0.084	0.061	0.061	0.062	0.063	0.067
	0.990	0.000	0.085	0.062	0.062	0.063	0.065	0.069
	1.500	−0.750	0.072	0.055	0.056	0.057	0.057	0.060
	1.900	−0.950	0.082	0.056	0.056	0.056	0.059	0.066
	0.800	0.100	0.083	0.062	0.061	0.062	0.064	0.067

*50,000 Replications, DGP: $y_t = x_t'\beta + u_t; \beta = 0; x_t = 0.5x_{t-1} + \eta_t, x_0 = 0;$
$u_t = a_1 u_{t-1} + a_2 u_{t-2} + e_t$, $u_0 = u_{-1} = 0$; $\eta_t, e_t \sim iid(0,1)$, $cov(\eta_t, e_t) = 0$.

Table 4b: Finite Sample Null Hypothesis Rejection Probabilities
DGP's are the same as Table 3 in KV(2002)*, with Prewhitening

	a_1	a_2	$t_{\rho=1}$	$t^*_{\rho=1}$	$t^*_{\rho=2}$	$t^*_{\rho=4}$	$t^*_{\rho=8}$	$t^*_{\rho=16}$
$T = 50$	−0.500	0.000	0.067	0.054	0.056	0.058	0.059	0.060
	0.000	0.000	0.076	0.060	0.063	0.064	0.066	0.068
	0.300	0.000	0.083	0.065	0.067	0.069	0.071	0.073
	0.500	0.000	0.086	0.067	0.070	0.071	0.074	0.076
	0.700	0.000	0.091	0.070	0.073	0.075	0.077	0.079
	0.900	0.000	0.097	0.071	0.075	0.078	0.080	0.084
	0.950	0.000	0.099	0.070	0.075	0.078	0.082	0.085
	0.990	0.000	0.104	0.067	0.074	0.079	0.085	0.088
	1.500	−0.750	0.080	0.063	0.065	0.067	0.069	0.071
	1.900	−0.950	0.097	0.069	0.071	0.075	0.080	0.084
	0.800	0.100	0.098	0.072	0.075	0.079	0.081	0.085
$T = 100$								
	−0.500	0.000	0.059	0.052	0.054	0.055	0.055	0.056
	0.000	0.000	0.063	0.054	0.055	0.056	0.058	0.058
	0.300	0.000	0.068	0.056	0.059	0.060	0.061	0.061
	0.500	0.000	0.071	0.060	0.062	0.062	0.064	0.064
	0.700	0.000	0.073	0.062	0.063	0.064	0.065	0.065
	0.900	0.000	0.076	0.064	0.065	0.065	0.066	0.068
	0.950	0.000	0.078	0.065	0.065	0.066	0.068	0.069
	0.990	0.000	0.081	0.059	0.063	0.065	0.068	0.070
	1.500	−0.750	0.062	0.059	0.059	0.059	0.060	0.060
	1.900	−0.950	0.072	0.057	0.058	0.059	0.062	0.064
	0.800	0.100	0.078	0.064	0.065	0.065	0.067	0.068
$T = 200$								
	−0.500	0.000	0.053	0.050	0.050	0.051	0.051	0.051
	0.000	0.000	0.056	0.053	0.054	0.054	0.054	0.054
	0.300	0.000	0.058	0.054	0.055	0.055	0.054	0.056
	0.500	0.000	0.060	0.054	0.055	0.055	0.055	0.056
	0.700	0.000	0.061	0.056	0.056	0.056	0.056	0.057
	0.900	0.000	0.063	0.057	0.057	0.057	0.057	0.058
	0.950	0.000	0.063	0.057	0.057	0.057	0.057	0.056
	0.990	0.000	0.065	0.056	0.057	0.058	0.058	0.058
	1.500	−0.750	0.047	0.052	0.053	0.052	0.052	0.051
	1.900	−0.950	0.058	0.052	0.052	0.051	0.052	0.054
	0.800	0.100	0.064	0.058	0.057	0.057	0.058	0.058

*50,000 Replications, DGP: $y_t = x_t'\beta + u_t; \beta = 0; x_t = 0.5x_{t-1} + \eta_t, x_0 = 0;$
$u_t = a_1 u_{t-1} + a_2 u_{t-2} + e_t$, $u_0 = u_{-1} = 0$; $\eta_t, e_t \sim iid(0,1)$, $cov(\eta_t, e_t) = 0$.

Table 6: Finite Sample Null Rejection Probabilities
Based on 50,000 Replications with No Prewhitening

	a_1	a_2	$t_{\rho=1}$	$t^*_{\rho=10}$	$t^*_{\rho=16}$	$t^*_{\rho=20}$
$T = 50$	0.85	0.00	0.125	0.095	0.102	0.106
	1.90	−0.95	0.139	0.105	0.115	0.121
$T = 100$	0.85	0.00	0.101	0.077	0.081	0.084
	1.90	−0.95	0.101	0.077	0.084	0.088
$T = 200$	0.85	0.00	0.084	0.064	0.067	0.069
	1.90	−0.95	0.082	0.061	0.066	0.069
$T = 500$	0.85	0.00	0.070	0.055	0.057	0.058
	1.9	−0.95	0.071	0.054	0.055	0.057

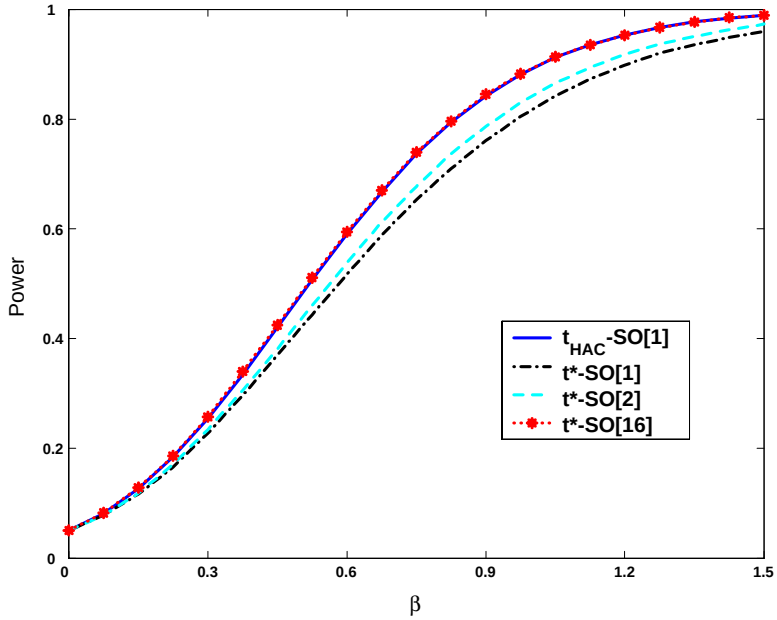

Figure 9a: Finite Sample Power (Size-adjusted, 5% Level), $T = 50$,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 0.85u_{t-1} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, No Prewhitening

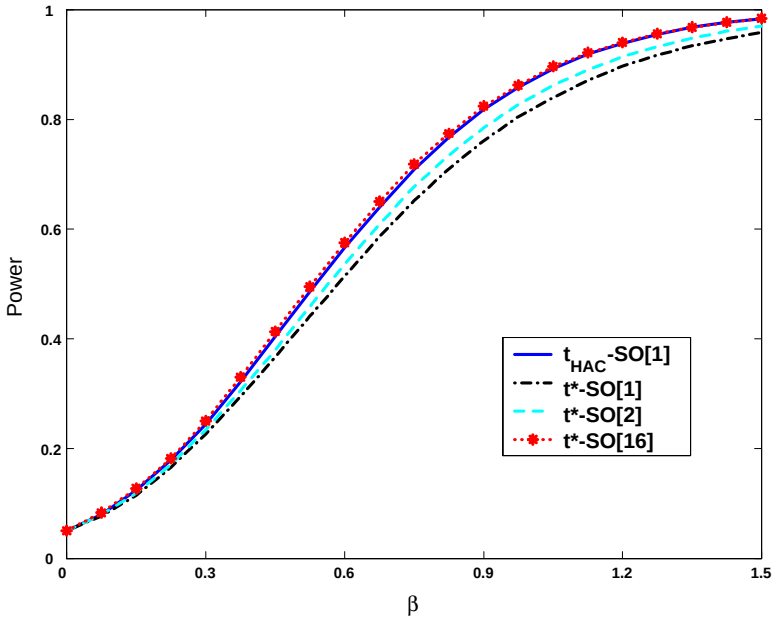

Figure 9b: Finite Sample Power (Size-adjusted, 5% Level), $T = 50$,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 0.85u_{t-1} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, With Prewhitening

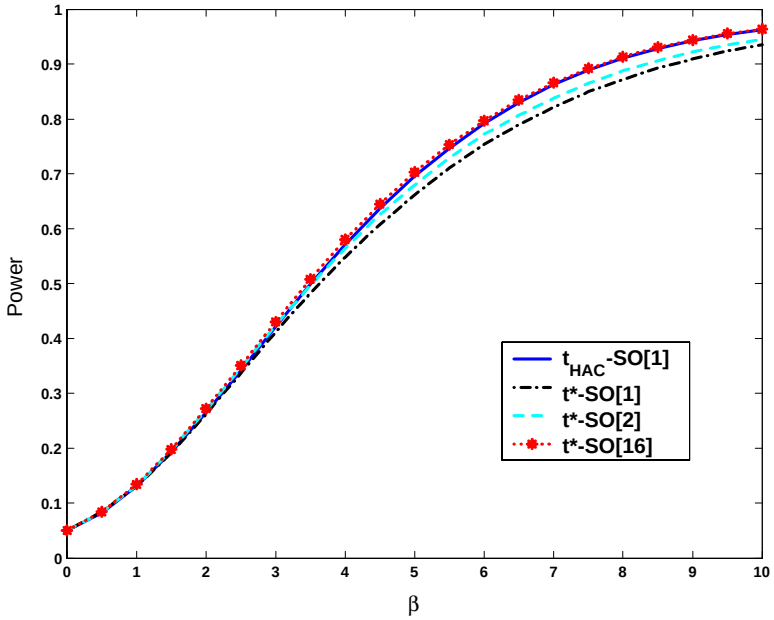

Figure 10a: Finite Sample Power (Size-adjusted, 5% Level), $T = 50$,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 1.9u_{t-1} - 0.95u_{t-2} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0,$ No Prewhitening

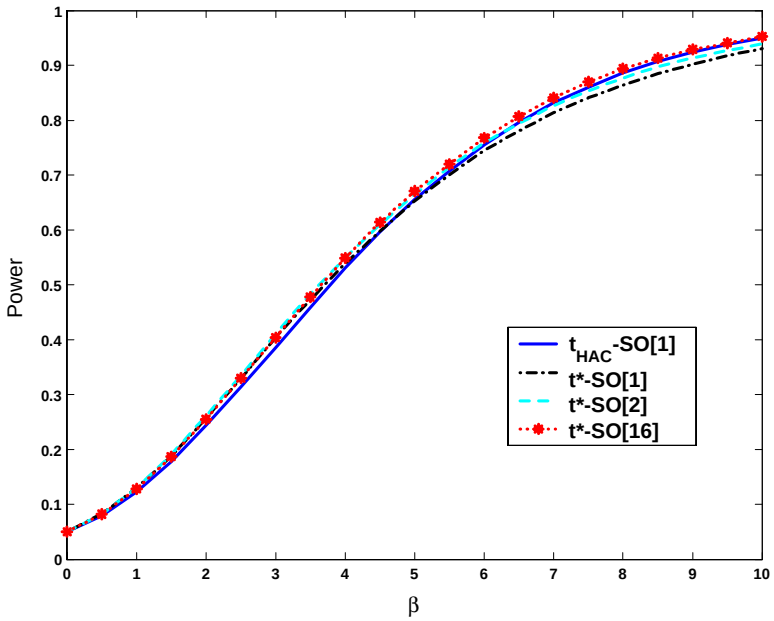

Figure 10b: Finite Sample Power (Size-adjusted, 5% Level), $T = 50$,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 1.9u_{t-1} - 0.95u_{t-2} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0,$ With Prewhitening

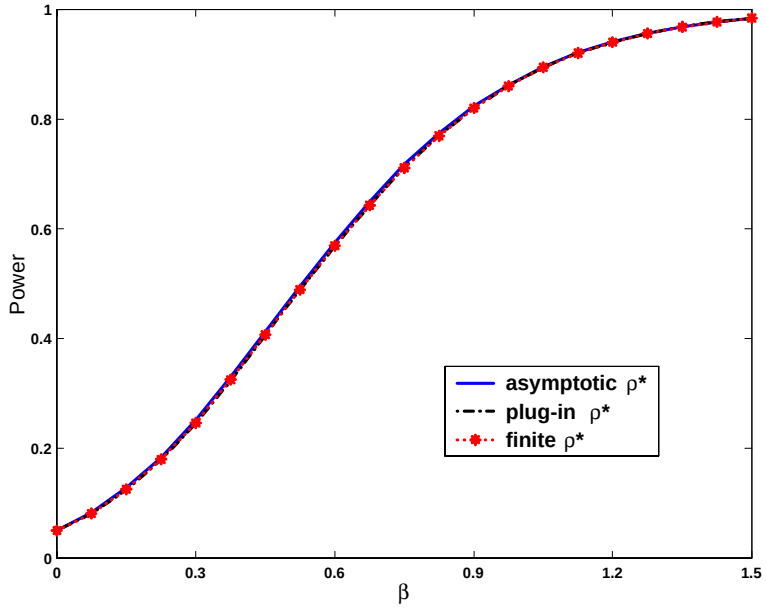

Figure 11: Finite Sample Power (Size adjusted) Comparison ,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 0.85u_{t-1} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$T = 50,\ H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, With Prewhitening

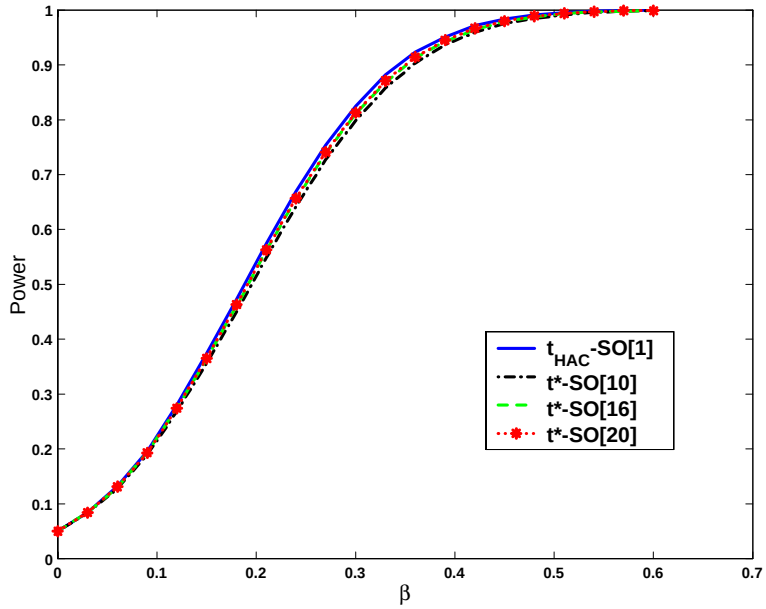

Figure 12: Finite Sample Power Comparison,
$y_t = \mu + x_t'\beta_0 + u_t; u_t = 0.85u_{t-1} + e_t; x_t = 0.5x_{t-1} + \eta_t$
$T = 500,\ H_0 : \beta_0 \leq 0,\ H_1 : \beta_0 > 0$, No Prewhitening

8 Appendix

Proof of Theorem 1 Using the definition of $\widehat{\Omega}_k$, Mercer's Theorem applied to $k(\cdot)$, (5), and standard weak convergence arguments, we have

$$\begin{aligned}
\widehat{\Omega}_k &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\widehat{v}_t k(\frac{t-\tau}{T})\widehat{v}'_\tau \\
&= \sum_{n=1}^{\infty}\frac{1}{\lambda_n}\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\widehat{v}_t f_n(t/T)f_n(\tau/T)\widehat{v}'_\tau \\
&= \sum_{n=1}^{\infty}\frac{1}{\lambda_n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_n(t/T)\widehat{v}_t\right)\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}f_n(\tau/T)\widehat{v}'_\tau\right) \\
&\Rightarrow \sum_{n=1}^{\infty}\frac{1}{\lambda_n}\Lambda\int_0^1 f_n(r)dV_m(r)\int_0^1 f_n(s)dV'_m(s)\Lambda' \\
&= \Lambda\int_0^1\int_0^1\sum_{n=1}^{\infty}\frac{1}{\lambda_n}f_n(r)f_n(s)dV_m(r)dV'_m(s)\Lambda' \\
&= \Lambda\int_0^1\int_0^1 k(r-s)dV_m(r)dV'_m(s)\Lambda'
\end{aligned}$$

So $\widehat{\Omega}_k \Rightarrow \Lambda\Xi\Lambda'$, as required. Next, it is easy to see that

$$\begin{aligned}
\Xi &= \int_0^1\int_0^1 k(r-s)\left(dW_m(r)-drW_m(1)\right)\left(dW_m(s)-dsW_m(1)\right)' \\
&= \int_0^1\int_0^1 k^*(r,s)dW_m(r)dW'_m(s),
\end{aligned}$$

where

$$k^*(r,s) = k(r-s) - \int_0^1 k(r-t)dt - \int_0^1 k(\tau-s)d\tau + \int_0^1 k(t-\tau)dtd\tau. \tag{1}$$

It follows that

$$\begin{aligned}
E\Xi &= \int_0^1\int_0^1 k^*(r,r)drI_m \\
&= \left(1-\int_0^1\int_0^1 k(r-s)drds\right)I_m = \mu I_m.
\end{aligned}$$

Therefore, $E\Lambda\Xi\Lambda' = \mu\Omega$, giving part (b).

For part (c), we write $E\left(\text{vec}(\Xi)\text{vec}(\Xi)'\right)$ as

$$E\left(\int_0^1\int_0^1\int_0^1\int_0^1 k^*(r,s)k^*(p,q)\text{vec}\left(dW_m(r)dW'_m(s)\right)\text{vec}\left(dW_m(p)dW'_m(q)\right)'\right)$$

Some calculations show that $E\left(\text{vec}\left(dW_m(r)dW'_m(s)\right)\text{vec}\left(dW_m(p)dW'_m(q)\right)\right)$ is

$$\begin{cases}
\text{vec}(I_m)\text{vec}(I_m)'drdp, & \text{if } r=s\neq p=q, \\
I_{m^2}drds, & \text{if } r=p\neq s=q, \\
K_{mm}drds, & \text{if } r=q\neq s=p, \\
0, & \text{otherwise.}
\end{cases} \tag{2}$$

Using the above result, we have

$$\begin{aligned} & E\left(\text{vec}(\Xi)\text{vec}(\Xi)'\right) \\ = & \int_0^1\int_0^1 k^*(r,r)k^*(p,p)drdp\text{vec}(I_m)\text{vec}(I_m)' \\ & +\int_0^1\int_0^1 k^*(r,s)k^*(r,s)drdsI_{m^2}+\int_0^1\int_0^1 k^*(r,s)k^*(r,s)drdsK_{mm} \\ = & \left(\int_0^1 k^*(r,r)dr\right)^2\text{vec}(I_m)\text{vec}(I_m)'+\int_0^1\int_0^1\left[k^*(r,s)\right]^2 drds\left(I_{m^2}+K_{mm}\right). \quad (3) \end{aligned}$$

Therefore

$$E\left(\text{vec}(\Xi_i)\,\text{vec}(\Xi_i)'\right)=\mu^2\text{vec}(I_m)\text{vec}(I_m)'+\nu\left(I_{m^2}+K_{mm}\right).$$

Some simple manipulations show that

$$\nu=\int k(r-s)k(p-q)-2\int k(r-s)k(r-q)+\int k(r-s)^2.$$

Hence

$$\begin{aligned} \text{var}(\text{vec}(\Lambda\Xi\Lambda')) &= E\text{vec}(\Lambda\Xi\Lambda')\text{vec}(\Lambda\Xi\Lambda')'-\text{vec}(\Lambda E\Xi\Lambda')\text{vec}(\Lambda E\Xi\Lambda')' \\ &= E\left(\Lambda\otimes\Lambda\right)\text{vec}(\Xi)\text{vec}(\Xi)'(\Lambda'\otimes\Lambda')-\mu^2\text{vec}(\Lambda\Lambda')\text{vec}(\Lambda\Lambda') \\ &= \mu^2\left(\Lambda\otimes\Lambda\right)\text{vec}(I_m)\text{vec}(I_m)'(\Lambda'\otimes\Lambda') \\ &\quad +\nu\left(\Lambda\otimes\Lambda\right)\left(I_{m^2}+K_{mm}\right)(\Lambda'\otimes\Lambda')-\mu^2\text{vec}(\Lambda\Lambda')\text{vec}(\Lambda\Lambda') \\ &= \nu\left(\Lambda\otimes\Lambda\right)\left(I_{m^2}+K_{mm}\right)(\Lambda'\otimes\Lambda') \\ &= \nu\left(\Lambda\Lambda'\right)\otimes\left(\Lambda\Lambda'\right)+\nu K_{mm}\left(\Lambda\Lambda'\right)\otimes\left(\Lambda\Lambda'\right) \\ &= \nu(I_{m^2}+K_{mm})\Omega\otimes\Omega, \end{aligned}$$

giving the stated result.

Proof of Corollary 1 Some algebraic manipulations show that

$$\begin{aligned} \mu_\rho &= 1-\int_0^1\int_0^1(1-|r-s|)^\rho drds=1-2\int_0^1\int_0^s(1+r-s)^\rho drds \\ &= 1-2\int_0^1\frac{1-(1-s)^{\rho+1}}{\rho+1}ds=\frac{\rho}{\rho+2}, \end{aligned}$$

and

$$\begin{aligned} \nu_\rho &= \left(\int k(r-s)\right)^2-2\int k(r-s)k(r-q)+\int k(r-s)^2 \\ &= \left(\frac{2}{\rho+2}\right)^2+\frac{2}{2\rho+2}-2\int_0^1\left(\int_0^r(1-r+s)^\rho ds+\int_r^1(1+r-s)^\rho ds\right)^2 dr \\ &= \left(\frac{2}{\rho+2}\right)^2+\frac{2}{2\rho+2}-2\int_0^1\left(\frac{2-(1-r)^{\rho+1}-r^{\rho+1}}{\rho+1}\right)^2 dr \\ &= \left(\frac{2}{\rho+2}\right)^2+\frac{1}{\rho+1}-\frac{2}{(\rho+1)^2}\left(4+\frac{2}{2\rho+3}-\frac{8}{\rho+2}-2\frac{\Gamma^2(\rho+2)}{\Gamma(2\rho+4)}\right). \end{aligned}$$

Lemma K *For* $\lambda_s = \frac{2\pi s}{T}$, $s = 0, 1, ..., [T/2]$, *and* $\rho = aT^b$ *with* $a > 0$ *and* $0 < b < 1$, *we have*

$$\sum_{h=0}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} e^{i\lambda_s h} = \frac{1}{1 - e^{i\lambda_s - \rho/T}} \left[1 + o\left(1\right)\right],$$

as $T \to \infty$.

Proof of Lemma K We introduce L such that $\frac{L}{T^{1-\frac{b}{2}}} + \frac{T^{1-b}}{L} \to 0$. For example, set $L = T^{1-\frac{3}{4}b}$. Then, we split the sum into two parts as follows:

$$\begin{aligned} \sum_{h=0}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} e^{i\lambda_s h} &= \sum_{h=0}^{L} \left(1 - \frac{h}{T}\right)^{\rho} e^{i\lambda_s h} + \sum_{h=L+1}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} e^{i\lambda_s h} \\ &= A_1 + A_2, \text{ say.} \end{aligned} \tag{4}$$

Consider each of these in turn, starting with

$$A_1 = \sum_{h=0}^{L} e^{\rho \log\left[1 - \frac{h}{T}\right]} e^{i\lambda_s h} = \sum_{h=0}^{L} e^{-h\rho/T + O\left(T^b \frac{h^2}{T^2}\right)} e^{i\lambda_s h} \tag{5}$$

$$= \sum_{h=0}^{L} e^{-h\rho/T} e^{i\lambda_s h} \left[1 + o\left(1\right)\right], \tag{6}$$

as $h^2/T^{2-b} = O\left(L^2/T^{2-b}\right) = o\left(1\right)$ uniformly for $h \leq L$. Next consider

$$\begin{aligned} |A_2| &= \left| \sum_{h=L+1}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} e^{i\lambda_s h} \right| \\ &\leq \sum_{h=L+1}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} = O\left(\int_{L+1}^{T-1} (1-h)^{\rho} dh\right) = O\left(T \int_{L/T}^{1-1/T} (1-y)^{\rho} dy\right) \\ &= O\left(T \left[-\frac{(1-y)^{\rho+1}}{\rho+1}\right]_{\frac{L}{T}}^{1-\frac{1}{T}}\right) = O\left(T \left[\frac{\left(1-\frac{L}{T}\right)^{\rho+1}}{\rho+1} - \frac{1}{T^{\rho+1}\left(\rho+1\right)}\right]\right) \\ &= O\left(\frac{e^{-\frac{L}{T^{1-b}}}}{T^{b-1}}\right) = O\left(\frac{e^{-T^{\frac{1}{4}b}}}{T^{b-1}}\right), \end{aligned} \tag{7}$$

which is exponentially smaller than $O\left(T^{-b}\right)$.

Now go back to consider A_1. First define

$$A_{12} = \sum_{h=L+1}^{T-1} e^{-h\rho/T} e^{i\lambda_s h}, \tag{8}$$

noting that

$$\begin{aligned} |A_{12}| &\leq \sum_{h=L+1}^{T-1} e^{-h\rho/T} = O\left(\int_{L}^{T-1} e^{-x\rho/T} dx\right) = O\left(\left[-\frac{e^{-xaT^{b-1}}}{aT^{b-1}}\right]_{L}^{T-1}\right) \\ &= O\left(\frac{1}{T^{b-1}}\left(e^{-aT^{\frac{1}{4}b}} - e^{-aT^b}\right)\right) \end{aligned} \tag{9}$$

which is exponentially smaller than $O\left(T^{-b}\right)$. Then, using (5)-(9) and for any $d \in (0,a)$, we can write

$$\begin{aligned} A_1 &= \sum_{h=0}^{L} e^{-h\rho/T} e^{i\lambda_s h} \left[1+o\left(1\right)\right] \\ &= \sum_{h=0}^{T-1} e^{-h\rho/T} e^{i\lambda_s h} \left[1+o\left(1\right)\right] + O\left(\frac{e^{-aT^{\frac{1}{4}b}}}{T^{b-1}}\right) \\ &= \sum_{h=0}^{T-1} e^{h(i\lambda_s-\rho/T)} \left[1+o\left(1\right)\right] + O\left(e^{-dT^{\frac{1}{4}b}}\right) \\ &= \frac{e^{T(i\lambda_s-\rho/T)}-1}{e^{i\lambda_s-\rho/T}-1} \left[1+o\left(1\right)\right] + O\left(e^{-dT^{\frac{1}{4}b}}\right) \\ &= \frac{e^{-\rho}-1}{e^{i\lambda_s-\rho/T}-1} \left[1+o\left(1\right)\right] + O\left(e^{-dT^{\frac{1}{4}b}}\right). \end{aligned} \tag{10}$$

Combining (4), (5) and (10), we have

$$\begin{aligned} \sum_{h=0}^{T-1} \left(1-\frac{h}{T}\right)^{\rho} e^{i\lambda_s h} &= \frac{e^{-\rho}-1}{e^{i\lambda_s-\rho/T}-1} \left[1+o\left(1\right)\right] + O\left(e^{-dT^{\frac{1}{4}b}}\right) \\ &= \frac{1}{1-e^{i\lambda_s-\rho/T}} \left[1+o\left(1\right)\right], \end{aligned}$$

as stated.

Proof of Lemma 1 Let $\rho = aT^b$ for some $a>0$ and $0<b<1$. We start by writing

$$\begin{aligned} K_{\rho}\left(\lambda_s\right) &= \sum_{h=-T+1}^{T-1} \left(1-\frac{|h|}{T}\right)^{\rho} \cos\left(\lambda_s h\right) \\ &= 2\sum_{h=0}^{T-1} \left(1-\frac{h}{T}\right)^{\rho} \cos\left(\lambda_s h\right) - 1 \\ &= 2\,\mathrm{Re}\left\{\sum_{h=0}^{T-1} \left(1-\frac{h}{T}\right)^{\rho} e^{i\lambda_s h}\right\} - 1. \end{aligned} \tag{11}$$

From Lemma K, we have

$$\begin{aligned} \sum_{h=0}^{T-1} \left(1-\frac{h}{T}\right)^{\rho} e^{i\lambda_s h} &= \frac{e^{-\rho}-1}{e^{i\lambda_s-\rho/T}-1} \left[1+o\left(1\right)\right] + O\left(\frac{e^{-aT^{\frac{1}{4}b}}}{T^{b-1}}\right) \\ &= \frac{1}{1-e^{i\lambda_s-\rho/T}} \left[1+o\left(1\right)\right]. \end{aligned} \tag{12}$$

Direct evaluation gives

$$\mathrm{Re}\left(\frac{1}{1-e^{ix-\frac{\rho}{T}}}\right) = \frac{1-e^{-\frac{\rho}{T}}\cos x}{1+e^{-\frac{2\rho}{T}}-2\left(\cos x\right)e^{-\frac{\rho}{T}}},$$

and so

$$\begin{aligned}
&\operatorname{Re}\left(\frac{1}{1-e^{ix-\frac{\rho}{T}}}\right)\\
=\;&\frac{1-\cos x\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)\right]}{2-2\cos x\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)\right]-\frac{2\rho}{T}+2\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}\\
=\;&\frac{1-\cos x\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)\right]}{2-2\cos x-\frac{2\rho}{T}(1-\cos x)+\left(\frac{\rho}{T}\right)^2(1-\cos x)+\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}\\
=\;&\frac{1-\cos x\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)\right]}{(1-\cos x)\left(2-\frac{2\rho}{T}+\left(\frac{\rho}{T}\right)^2\right)+\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&2\operatorname{Re}\left(\frac{1}{1-e^{-\frac{\rho}{T}+ix}}\right)-1\\
=\;&\frac{2-2\cos x\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)\right]}{2(1-\cos x)\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2\right]+\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}-1\\
=\;&\frac{\frac{2\rho}{T}-2\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}{2(1-\cos x)\left[1-\frac{\rho}{T}+\frac{1}{2}\left(\frac{\rho}{T}\right)^2\right]+\left(\frac{\rho}{T}\right)^2+o\left(\left(\frac{\rho}{T}\right)^2\right)}\\
=\;&\frac{\frac{2\rho}{T}[1+o(1)]}{2(1-\cos x)[1+o(1)]+\left(\frac{\rho}{T}\right)^2[1+o(1)]}\\
=\;&\frac{2\rho T}{2T^2(1-\cos x)+\rho^2}[1+o(1)],
\end{aligned}$$

which, when combined with (11) and (12), gives

$$K_\rho(\lambda_s)=\frac{2\rho T}{2T^2(1-\cos\lambda_s)+\rho^2}[1+o(1)],$$

as stated.

Proof of Theorem 2: We prove the results for the scalar v_t case, the vector case follows without further complication.

Part (a) From (19)

$$\widetilde{f}_{vv}(0)=\frac{1}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)I_{vv}(\lambda_s). \tag{13}$$

To find the asymptotic variance of $\widetilde{f}_{vv}(0)$, we can work from the following standard formula (e.g., Priestley, 1981, eqn. 6.2.110 on p. 455) for the variance of a weighted

periodogram estimate such as (13)[3], viz.,

$$\mathrm{Var}\left\{\widetilde{f}_{vv}(0)\right\} = 2f_{vv}(0)^2 \frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})^2\left[1+o(1)\right], \tag{14}$$

which follows directly from the covariance properties of the periodogram of a linear process (e.g., Priestley, 1981, p. 426). To evaluate (14), we develop an asymptotic approximation of

$$\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho^2\left(\frac{h}{T}\right) = \frac{1}{T}\sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)^{2\rho} = \frac{2}{T}\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{2\rho} - \frac{1}{T}.$$

This can be accomplished by Euler summation, viz.,

$$\sum_{k=0}^{n} g(k) = \int_0^n g(x)\,dx + \frac{1}{2}\left\{g(0)+g(n)\right\} + \int_0^n \left(x-[x]-\frac{1}{2}\right) g'(x)\,dx$$

applied to $g(x) = (1-x/T)^{2\rho}$ giving

$$\begin{aligned}\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{2\rho} &= \int_0^{T-1}\left(1-\frac{x}{T}\right)^{2\rho}dx + \frac{1}{2}\left\{1+\left(1-\frac{T-1}{T}\right)^{2\rho}\right\} \\ &\quad + \left(-\frac{2\rho}{T}\right)\int_0^{T-1}\left(x-[x]-\frac{1}{2}\right)\left(1-\frac{x}{T}\right)^{2\rho-1}dx. \end{aligned}\tag{15}$$

Note that

$$1+\left(1-\frac{T-1}{T}\right)^{2\rho} = O(1), \tag{16}$$

and

$$\begin{aligned}&\left|\left(-\frac{2\rho}{T}\right)\int_0^{T-1}\left(x-[x]-\frac{1}{2}\right)\left(1-\frac{x}{T}\right)^{2\rho-1}dx\right| \\ \leq\ & 2\frac{\rho}{T}\int_0^{T-1}\left(1-\frac{x}{T}\right)^{2\rho-1}dx = 2\rho\int_0^{1-1/T}(1-y)^{2\rho-1}\,dy \\ =\ & \left[(1-y)^{2\rho}\right]_0^{1-1/T} = O(1), \end{aligned}\tag{17}$$

[3] Note that inversion of $I_{vv}(\lambda) = \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\widetilde{\Gamma}(h)e^{-i\lambda h}$ gives $\widetilde{\Gamma}(j) = \int_{-\pi}^{\pi} I_{vv}(\lambda)e^{i\lambda j}d\lambda$ so that

$$\begin{aligned}\widetilde{f}_{vv}(0) &= \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widetilde{\Gamma}(h) = \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{vv}(\lambda)\left\{\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\,e^{i\lambda h}\right\}d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{vv}(\lambda)K_\rho(\lambda)\,d\lambda, \end{aligned}$$

is an alternate form of (13).

whereas

$$\begin{aligned}
&\int_0^{T-1}\left(1-\frac{x}{T}\right)^{2\rho}dx \\
= \;& T\int_0^{1-1/T}(1-y)^{2\rho}\,dy=\frac{T}{2\rho+1}\left[-(1-y)^{2\rho+1}\right]_0^{1-1/T} \\
= \;& \frac{T}{2\rho+1}\left[1-\left(1-\frac{T-1}{T}\right)^{2\rho+1}\right]=\frac{T}{2\rho+1}+O\left(\frac{1}{\rho T^{2\rho}}\right). \qquad (18)
\end{aligned}$$

It follows from (15) - (18) that

$$\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{2\rho}=\frac{T}{2\rho+1}+O\left(1\right),$$

so that

$$\begin{aligned}
&\frac{1}{T}\sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)^{2\rho}=\frac{2}{T}\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{2\rho}-\frac{1}{T} \\
= \;& \frac{2}{T}\int_0^{T-1}\left(1-\frac{s}{T}\right)^{2\rho}ds+O(\frac{1}{T}) \\
= \;& 2\int_0^{1-\frac{1}{T}}(1-y)^{2\rho}dy+O(\frac{1}{T}) \\
= \;& \frac{2}{2\rho+1}+O(\frac{1}{T})=\frac{1}{\rho}\left[1+o\left(1\right)\right], \qquad (19)
\end{aligned}$$

giving

$$\frac{1}{T}\sum_{h=-T+1}^{T-1}k_\rho^2\left(\frac{h}{T}\right)=\frac{1}{\rho}\left[1+o\left(1\right)\right]. \qquad (20)$$

Using (19) in (14) we have

$$\mathrm{Var}\left\{\widetilde{f}_{vv}(0\right\}=\frac{1}{\rho}2f_{vv}\left(0\right)^2\left[1+o\left(1\right)\right],$$

and so

$$\rho\mathrm{Var}\left\{\widetilde{f}_{vv}(0\right\}=2f_{vv}\left(0\right)^2\left[1+o\left(1\right)\right]\to 2f_{vv}\left(0\right)^2,$$

which gives

$$\lim_{T\to\infty}\rho\mathrm{Var}\left\{\widetilde{\Omega}_{k_\rho}\right\}=2\left(2\pi\right)^2f_{vv}\left(0\right)^2=2\Omega^2,$$

as required. The stated result for the vector case follows in a straightforward way.

Part (b): Since $\widetilde{f}_{vv}(0)=\frac{1}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)I_{vv}(\lambda_s)$ and

$$\sum_{s=0}^{T-1}K_\rho(\lambda_s)=\sum_{h=-T+1}^{T-1}k_\rho(\frac{h}{T})\sum_{s=0}^{T-1}e^{i\lambda_s h}=Tk(0)=T,$$

we can write the scaled estimation error as

$$\begin{aligned}
&\sqrt{\rho}\left\{\widetilde{f}_{vv}(0)-f_{vv}(0)\right\}\\
=\;&\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\left[I_{vv}(\lambda_s)-f_{vv}(0)\right]\\
=\;&\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\left[I_{vv}(\lambda_s)-f_{vv}(\lambda_s)\right]+\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\left[f_{vv}(\lambda_s)-f_{vv}(0)\right]. \quad (21)
\end{aligned}$$

Using Lemma 1, we have

$$K_\rho(\lambda_s)=O\left(\frac{2\rho T}{(2\pi s)^2+\rho^2}\left[1+o(1)\right]\right),\quad s=0,1,...,[T/2]. \quad (22)$$

By **A2**, $|f'_{vv}(\lambda_s)|\leq\frac{1}{2\pi}\sum_{-\infty}^{\infty}|h|\,|\Gamma(h)|$, so that

$$|f_{vv}(\lambda_s)-f_{vv}(0)|\leq\left(\frac{1}{2\pi}\sum_{-\infty}^{\infty}|h|\,|\Gamma(h)|\right)\lambda_s.$$

Hence, the second term of (21) can be bounded as follows:

$$\begin{aligned}
&\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\left[f_{vv}(\lambda_s)-f_{vv}(0)\right]\\
=\;&\frac{2\sqrt{\rho}}{T}\sum_{s=0}^{[T/2]}K_\rho(\lambda_s)\left[f_{vv}(\lambda_s)-f_{vv}(0)\right]=O\left(\frac{\sqrt{\rho}}{T}\sum_{s=0}^{[T/2]}K_\rho(\lambda_s)\lambda_s\right)\\
=\;&O\left(\frac{\rho^{3/2}}{T}\sum_{s=0}^{[T/2]}\frac{2T\lambda_s}{\rho^2+(2\pi s)^2}\right)=O\left(\frac{\rho^{3/2}}{T}\int_0^{T/2}\frac{x}{\rho^2+(2\pi x)^2}dx\right)\\
=\;&O\left(\frac{\rho^{3/2}}{T}\left[\frac{\log\left\{\rho^2+(2\pi x)^2\right\}}{2(2\pi)^2}\right]_0^{T/2}\right)\\
=\;&O\left(\frac{\rho^{3/2}\log T}{T}\right)=o(1), \quad (23)
\end{aligned}$$

since $\rho=aT^b$ with $b<\frac{2}{3}$. Then, by (21) and (23), we have

$$\sqrt{\rho}\left\{\widetilde{f}_{vv}(0)-f_{vv}(0)\right\}=\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\left(I_{vv}(\lambda_s)-f_{vv}(\lambda_s)\right)+o_p(1).$$

In view of **A2**, we have $v_t=C(L)\varepsilon_t=\sum_{j=0}^{\infty}C_j\varepsilon_{t-j}$, where the ε_t are $iid(0,\sigma^2)$ and have finite fourth moment μ_4. The operator $C(L)$ has a valid spectral BN decomposition (Phillips and Solo, 1992):

$$C(L)=C(e^{i\lambda})+\widetilde{C}_\lambda(e^{-i\lambda}L)(e^{-i\lambda}L-1),$$

where $\widetilde{C}_\lambda(e^{-i\lambda}L) = \sum_{j=0}^{\infty} \widetilde{C}_{\lambda j} e^{-ij\lambda} L^j$ and $\widetilde{C}_{\lambda j} = \sum_{s=j+1}^{\infty} C_s e^{is\lambda}$, leading to the representation

$$v_t = C(L)\varepsilon_t = C(e^{i\lambda})\varepsilon_t + e^{-i\lambda}\widetilde{\varepsilon}_{\lambda t-1} - \widetilde{\varepsilon}_{\lambda t} \tag{24}$$

where

$$\widetilde{\varepsilon}_{\lambda t} = \widetilde{C}_\lambda(e^{-i\lambda}L)\varepsilon_t = \sum_{j=0}^{\infty} \widetilde{C}_{\lambda j} e^{-ij\lambda}\varepsilon_{t-j}$$

is stationary. The discrete Fourier transform of v_t has the corresponding representation

$$\begin{aligned} w(\lambda_s) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} v_t e^{it\lambda_s} \\ &= C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + \frac{1}{\sqrt{2\pi T}} (\widetilde{\varepsilon}_{\lambda_s 0} - e^{in\lambda_s} \widetilde{\varepsilon}_{\lambda_s n}) \\ &= C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + O_p(T^{-1/2}). \end{aligned}$$

Thus, using the fact that $\sum_{s=0}^{T-1} |K_\rho(\lambda_s)| = \sum_{s=0}^{T-1} K_\rho(\lambda_s) = T$, we have

$$\begin{aligned} & \sqrt{\rho}\left\{\widetilde{f}_{vv}(0) - f_{vv}(0)\right\} \\ = \; & \frac{\sqrt{\rho}}{T} \sum_{s=0}^{T-1} K_\rho(\lambda_s)(I_{vv}(\lambda_s) - f_{vv}(\lambda_s)) + o_p(1) \qquad (25) \\ = \; & \frac{\sqrt{\rho}}{T} \sum_{s=0}^{T-1} K_\rho(\lambda_s)(w(\lambda_s)w(\lambda_s)^* - f_{vv}(\lambda_s)) + o_p(1) \\ = \; & \frac{\sqrt{\rho}}{T} \sum_{s=0}^{T-1} K_\rho(\lambda_s)\{[C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + O_p(T^{-1/2})] \\ & \times [C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + O_p(T^{-1/2})]^* - f_{vv}(\lambda_s)\} + o_p(1) \\ = \; & \frac{\sqrt{\rho}}{T} \sum_{s=0}^{T-1} K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] + O_p\left(\frac{\sqrt{\rho}}{T} T \frac{1}{T^{1/2}}\right) + o_p(1) \\ = \; & \frac{\sqrt{\rho}}{T} \sum_{s=0}^{T-1} K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] + o_p(1), \qquad (26) \end{aligned}$$

where we have used $\rho/T \to 0$.

Let $m_1 = 0$ and for $t \geq 2$,

$$m_t = \varepsilon_t \sum_{j=1}^{t-1} \varepsilon_j c_{t-j}$$

where

$$c_j = \frac{C^2(1)}{2\pi} \frac{\sqrt{\rho}}{T^2} \sum_{s=0}^{T-1} (K(\lambda_s)\cos(j\lambda_s)).$$

Then we can write

$$\begin{aligned}
&\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s)-\frac{1}{2\pi}\sigma^2)] \\
= \;& 2\sum_{t=1}^{T}m_t+\sqrt{\rho}\frac{C^2(1)}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\frac{1}{2\pi}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2-\sigma^2\right) \\
= \;& 2\sum_{t=1}^{T}m_t+2\sqrt{\rho}\frac{C^2(1)}{T}\left(\sum_{s=0}^{T-1}K_\rho(\lambda_s)\right)O_p(\frac{1}{\sqrt{T}}) \\
= \;& 2\sum_{t=1}^{T}m_t+O_p\left(\frac{\sqrt{\rho}}{T}T\frac{1}{\sqrt{T}}\right) \\
= \;& 2\sum_{t=1}^{T}m_t+o_p(1).
\end{aligned}$$

By the Fourier inversion formula, we have

$$c_j=\frac{C^2(1)}{2\pi}\frac{\sqrt{\rho}}{T}k_\rho(\frac{j}{T}). \tag{27}$$

Hence

$$\sum_{j=1}^{T}c_j^2=O\left(\frac{\rho}{T^2}\sum_{j=1}^{T}k_\rho^2(\frac{j}{T})\right)=O\left(\frac{\rho}{T}\frac{1}{2\rho+1}\right)=O\left(\frac{1}{T}\right). \tag{28}$$

The sequence m_t depends on T via the coefficients c_j and forms a zero mean martingale difference array. Then

$$2\sum_{t=1}^{T}m_t\to_d N\left(0,\frac{\sigma^4C^4(1)}{2\pi^2}\right)=N\left(0,2f_{vv}^2(0)\right),$$

by a standard martingale CLT, provided the following two sufficient conditions hold:

$$\sum_{t=1}^{T}E(m_t^2|\mathcal{F}_{t-1})-\frac{\sigma^4C^4(1)}{8\pi^2}\to_p 0, \tag{29}$$

where $\mathcal{F}_{t-1}=\sigma(\varepsilon_{t-1},\varepsilon_{t-2},...)$ is the filtration generated by the innovations ε_j, and

$$\sum_{t=1}^{T}E(m_t^4)\to_p 0. \tag{30}$$

We now proceed to establish (29) and (30). The left hand side of (29) is

$$\left(\sigma^2\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varepsilon_j^2c_{t-j}^2-\frac{\sigma^4C^4(1)}{8\pi^2}\right)+\sigma^2\sum_{t=2}^{T}\sum_{r\neq j}\varepsilon_r\varepsilon_jc_{t-r}c_{t-j}:=I_1+I_2. \tag{31}$$

The first term, I_1, is

$$\sigma^2\left(\sum_{j=1}^{T-1}(\varepsilon_j^2-\sigma^2)\sum_{s=1}^{T-j}c_s^2\right)+\left(\sigma^4\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}c_j^2-\frac{\sigma^4C^4(1)}{8\pi^2}\right):=I_{11}+I_{12}. \tag{32}$$

The mean of I_{11} is zero and its variance is of order

$$O\left[\sum_{j=1}^{T-1}\left(\sum_{s=1}^{T-j}c_s^2\right)^2\right]=O\left[T\left(\sum_{s=1}^{T}c_s^2\right)^2\right]=O\left(\frac{1}{T}\right),$$

using (28).

Next, consider the second term of (32). We have

$$\begin{aligned}
&\sum_{j=1}^{T-1}\sum_{s=1}^{T-j}c_s^2=\frac{C^4(1)}{4\pi^2}\frac{\rho}{T^2}\sum_{j=1}^{T-1}\sum_{s=1}^{T-j}k_\rho^2(\frac{s}{T})\\
=\;&(1+o(1))\frac{C^4(1)}{4\pi^2}\frac{\rho}{T}\sum_{j=1}^{T-1}\int_{j/T}^{1-1/T}y^{2\rho}dy\\
=\;&(1+o(1))\frac{C^4(1)}{4\pi^2}\frac{\rho}{T}\sum_{j=1}^{T-1}\left(\frac{1}{2\rho+1}\right)\left[\left(1-\frac{1}{T}\right)^{2\rho+1}-\left(\frac{j}{T}\right)^{2\rho+1}\right]\\
=\;&\frac{C^4(1)}{4\pi^2}\frac{\rho}{T}\left(\frac{1}{2\rho+1}\right)T\left(1-\frac{1}{T}\right)^{2\rho+1}(1+o(1))\\
&-\frac{C^4(1)}{4\pi^2}\frac{\rho}{T}\left(\frac{1}{2\rho+1}\right)\frac{T}{2\rho+2}(1+o(1))\\
=\;&\frac{C^4(1)}{8\pi^2}+o(1).
\end{aligned}$$

We have therefore shown that

$$I_1=\sigma^2\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varepsilon_j^2c_{t-j}^2-\frac{\sigma^4C^4(1)}{8\pi^2}\to_p 0.$$

So the first term of (31) is $o_p(1)$.

Now consider the second term, I_2, of (31). I_2 has mean zero and variance

$$\begin{aligned}
&O\left(2\sum_{p,q=2}^{T}\sum_{r\neq j}^{\min(p-1,q-1)}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right)\\
=\;&O\left(2\sum_{p=2}^{T}\sum_{r\neq j}^{p-1}c_{p-r}^2c_{p-j}^2+4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right).
\end{aligned} \tag{33}$$

In view of (28), we have

$$\sum_{p=2}^{T}\sum_{r\neq j}^{p-1}c_{p-r}^2c_{p-j}^2=O\left(T\left(\sum_{j=1}^{T}c_j^2\right)^2\right)=O\left(\frac{1}{T}\right).$$

For the second component in (33), we have, using (28) and the Cauchy inequality,

$$
\begin{aligned}
& 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{p-1,q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j}) \leq 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1}c_{q-r}^2\sum_{r=1}^{q-1}c_{p-r}^2 \\
\leq\ & 4\sum_{i=1}^{T}c_i^2\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1}c_{p-r}^2 \leq 4(\sum_{i=1}^{T}c_i^2)(\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1}c_r^2) \\
=\ & O\left(\frac{1}{T}\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1}c_r^2\right) = O\left(\frac{1}{T}\sum_{r=1}^{T-2}r(T-r-1)c_r^2\right) \\
=\ & O\left(\frac{\rho}{T^3}\sum_{r=1}^{T-2}r(T-r-1)(1-\frac{r}{T})^{\rho}\right) = O\left(\frac{\rho}{T^3}\sum_{r=1}^{T-2}r(T-r)(1-\frac{r}{T})^{\rho}\right) \\
=\ & O\left(\rho\frac{1}{T}\sum_{r=1}^{T-2}\frac{r}{T}(1-\frac{r}{T})^{\rho+1}\right) = O(\rho B(2,\rho+2)) \\
=\ & O\left(\frac{\rho\Gamma(2)\Gamma(\rho+2)}{\Gamma(\rho+4)}\right) = O\left(\frac{\rho}{(\rho+3)(\rho+2)}\right) \\
=\ & O\left(\frac{1}{\rho}\right) = o(1).
\end{aligned}
$$

Hence, $I_2 \to_p 0$ and we have therefore established condition (29).

It remains to verify (30). Let A be some positive constant, then the left hand side of (30) is

$$
\begin{aligned}
& \mu_4\sum_{t=2}^{T}E(\sum_{s=1}^{t-1}\varepsilon_s c_{t-s})^4 \\
\leq\ & A\sum_{t=2}^{T}E\left(\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\sum_{p=1}^{t-1}\sum_{q=1}^{t-1}\varepsilon_s\varepsilon_r\varepsilon_p\varepsilon_q c_{t-s}c_{t-r}c_{t-p}c_{t-q}\right) \\
\leq\ & A\sum_{t=2}^{T}(\sum_{s=1}^{T}c_{t-s}^4) + A\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}c_{t-s}^2c_{t-r}^2 \\
\leq\ & AT(\sum_{t=1}^{T}c_t^2)^2 = O\left(T\frac{1}{T^2}\right) = O\left(\frac{1}{T}\right)
\end{aligned}
$$

using (28), which verifies (30) and the CLT.

With this construction we therefore have

$$
\begin{aligned}
& \frac{\sqrt{\rho}}{T}C^2(1)\sum_{s=0}^{T-1}K_\rho(\lambda_s)[(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] \\
=\ & 2\sum_{t=1}^{T}m_t + o_p(1) \to_d 2N\left(0, \frac{\sigma^4C^4(1)}{8\pi^2}\right) \\
=\ & N\left(0, \frac{\sigma^4C^4(1)}{2\pi^2}\right) = N(0, 2f_{vv}^2(0)).
\end{aligned}
$$

This gives the required limit theory for the spectral estimate at the origin, viz.,

$$\sqrt{\rho}\left\{\widetilde{f}_{vv}(0) - f_{vv}(0)\right\} = \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K(\lambda_s)(I_{vv}(\lambda_s) - f_{vv}(\lambda_s)) + o_p(1)$$

$$\to_d N\left(0, 2f_{vv}^2(0)\right),$$

from which we deduce that

$$\sqrt{\rho}(\widetilde{\Omega}_{k_\rho} - \Omega) \to_d N\left(0, 2\Omega^2\right).$$

The stated result for the vector case follows directly by standard extensions.

Part (c): By definition,

$$\begin{aligned}
& E(\widehat{f}_{vv}(0) - f_{vv}(0)) \\
= & \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})EC_h - \frac{1}{2\pi}\sum_{-\infty}^{\infty}\Gamma(h) \\
= & \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)^{\rho+1}\Gamma(h) - \frac{1}{2\pi}\sum_{-\infty}^{\infty}\Gamma(h) \\
= & \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h) - \frac{1}{2\pi}\sum_{|h|\geq T}\Gamma(h) \\
= & \frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h) \\
& + \frac{1}{2\pi}\sum_{T-1\geq|h|>T/(\rho\log T)}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h) + \frac{1}{2\pi}\sum_{|h|\geq T}\Gamma(h) \qquad (34)
\end{aligned}$$

where the second equality follows from the fact $EC_h = \left(1 - \frac{|h|}{T}\right)\Gamma(h)$. Now

$$\left|\frac{T}{\rho+1}\sum_{|h|\geq T}\Gamma(h)\right| \leq \frac{1}{\rho}\sum_{|h|\geq T}|h|\,|\Gamma(h)| = o\left(\frac{1}{\rho}\right) = o(1),$$

by virtue of **A2,** and

$$\begin{aligned}
& \left|\frac{T}{\rho+1}\sum_{T-1\geq|h|>T/(\rho\log T)}^{T-1}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h)\right| \\
\leq & \frac{T}{\rho+1}\sum_{T-1\geq|h|>T/(\rho\log T)}^{T-1}\left|1-\left(1-\frac{|h|}{T}\right)^{\rho+1}\right||\Gamma(h)| \\
\leq & \frac{T}{\rho+1}\sum_{T-1\geq|h|>T/(\rho\log T)}^{T-1}|\Gamma(h)| \leq \frac{T}{\rho}\left(\frac{\rho\log T}{T}\right)^{1+\Delta}\sum_{T-1\geq|h|>T/(\rho\log T)}^{T-1} h^{1+\Delta}|\Gamma(h)| \\
= & o(1),
\end{aligned}$$

for some small $\Delta > 0$, in view of **A2.**

The first term of (34) can be written as

$$\begin{aligned}
&\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1}-1\right]\Gamma(h)\\
=\;&\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[1-\frac{(\rho+1)|h|}{T}+O\left(\frac{(\rho+1)^2|h|^2}{T^2}\right)-1\right]\Gamma(h)\\
=\;&-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+O\left(\frac{(\rho+1)^2}{T^2}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}h^2\Gamma(h)\right)\\
=\;&-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+o\left(\frac{(\rho+1)^2}{T^2}\frac{T}{\rho\log T}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}|h|\,|\Gamma(h)|\right)\\
=\;&-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+o\left(\frac{\rho}{T\log T}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\lim_{T\to\infty}\frac{T}{\rho}\left(E\widetilde{\Omega}_{k_\rho}-\Omega\right) &= \lim_{T\to\infty}\left(-\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}|h|\,\Gamma(h)\right)\\
&= -2\pi f^{(1)}=-\Omega^{(1)},
\end{aligned}$$

as required.

Part (d) Since $\rho^3/T^2\to\vartheta\in(0,\infty)$, we have $\rho\sim\vartheta^{\frac{1}{3}}T^{\frac{2}{3}}$ and then

$$\frac{T}{\rho}\sim\frac{T}{\vartheta^{\frac{1}{3}}T^{\frac{2}{3}}}=\vartheta^{-\frac{1}{3}}T^{\frac{1}{3}}\sim\vartheta^{-\frac{1}{3}}\sqrt{\frac{\rho}{\vartheta^{\frac{1}{3}}}}=\frac{\sqrt{\rho}}{\sqrt{\vartheta}}. \tag{35}$$

It follows from (35) that

$$\begin{aligned}
&MSE(\rho,\widetilde{\Omega}_{k_\rho},W)\\
=\;&\rho E\left\{vec\left(\widetilde{\Omega}_{k_\rho}-\Omega\right)'W\,vec\left(\widetilde{\Omega}_{k_\rho}-\Omega\right)\right\}\\
=\;&\rho E\left\{vec\left(\widetilde{\Omega}_{k_\rho}-E\widetilde{\Omega}_{k_\rho}+E\widetilde{\Omega}_{k_\rho}-\Omega\right)'W\,vec\left(\widetilde{\Omega}_{k_\rho}-E\widetilde{\Omega}_{k_\rho}+E\widetilde{\Omega}_{k_\rho}-\Omega\right)\right\}\\
=\;&\vartheta\left(\frac{T}{\rho}\right)^2E\left\{vec\left(E\widetilde{\Omega}_{k_\rho}-\Omega\right)'W\,vec\left(E\widetilde{\Omega}_{k_\rho}-\Omega\right)\right\}[1+o(1)]\\
&+\rho\mathrm{tr}\left\{WE\left[vec\left(\widetilde{\Omega}_{k_\rho}-E\widetilde{\Omega}_{k_\rho}\right)\right]vec\left(\widetilde{\Omega}_{k_\rho}-E\widetilde{\Omega}_{k_\rho}\right)'\right\}.
\end{aligned}$$

Using parts (b) and (c), we obtain

$$\begin{aligned}
&\lim_{T\to\infty}\mathrm{MSE}(\rho,\widetilde{\Omega}_{k_\rho},W)\\
=\;&\vartheta\;vec\left(\Omega^{(1)}\right)'Wvec\left(\Omega^{(1)}\right)+\mathrm{tr}\;\left\{W(I+K_{mm})\Omega\otimes\Omega\right\}.
\end{aligned}$$

The corresponding result for the spectral density estimate is

$$\begin{aligned} & \lim_{T\to\infty} MSE(\rho, \widetilde{f}_{vv}(0), W) \\ = \;& \vartheta \; vec\left(f^{(1)}\right)' W \, vec\left(f^{(1)}\right) + \text{tr}\; \{W(I+K_{mm})f\otimes f)\}. \end{aligned}$$

Proof of Theorem 3: Assumption B below is based on corresponding conditions in Andrews (1991). It allows for the effect of using $\widehat{\beta}$ in the HAC estimate and is sufficient for the consistency of $\widehat{\Omega}_{k_\rho}$ and for $\widehat{\Omega}_{k_\rho}$ to have the same asymptotic distribution as $\widetilde{\Omega}_{k_\rho}$. Let $\varkappa$ denote some convex neighborhood of β_0, the true value of β. Let v_{at} denote the a'th element of v_t. Let $\kappa_{a_1...a_8}(0, j_1, j_{2,}..., j_7)$ denote the cumulant of $(v_{a_1 0}, ...v_{a_8\ j_7})$, where $a_1, ..., a_8$ are positive integers less than $p+1$ and $j_1, ..., j_7$ are integers.

Assumption B: (1) $\sqrt{T}(\widehat{\beta}-\beta) = O_p(1)$.

(2) $\sup_{t\geq 1} E \parallel v_t \parallel^2 < \infty$.

(3) $\sup_{t\geq 1} E \sup_{\beta\in\varkappa} \parallel (\partial/\partial\beta')v_t(\beta) \parallel^2 < \infty$.

(4) Assumption **A2** holds with v_t replaced by $(v_t', vec(\frac{\partial}{\partial\beta'}v_t(\beta) - E\frac{\partial}{\partial\beta'}v_t(\beta))')'$.

(5) $\sup_{t\geq 1} E \sup_{\beta\in\varkappa} \parallel (\partial/\partial\beta')v_{at}(\beta) \parallel^2 < \infty$, $\forall a = 1, ..., m$.

(6) $\{v_t\}$ *is eighth order stationary with summable cumulant function* $\kappa_{a_1...a_8}(0, j_1, j_{2,}..., j_7)$, i.e., $\sum_{j_1=-\infty}^{\infty}\cdots\sum_{j_7=-\infty}^{\infty} |\kappa_{a_1...a_8}(0, j_1, j_{2,}..., j_7)| < \infty$.

(7) $W_T \to_p W$.

Proof of Part (a) A Taylor expansion gives

$$\begin{aligned} \sqrt{\rho}(\widehat{\Omega}_{k_\rho} - \widetilde{\Omega}_{k_\rho}) \;=\; & [\sqrt{\rho/T}\frac{\partial}{\partial\beta'}\widetilde{\Omega}_{k_\rho}(\beta)]\sqrt{T}(\widehat{\beta}-\beta) \\ & + \frac{1}{2}\sqrt{T}(\widehat{\beta}-\beta)'[\sqrt{\frac{\rho}{T^2}}\frac{\partial^2}{\partial\beta\partial\beta'}\widetilde{\Omega}_{k_\rho}(\widetilde{\beta})]\sqrt{T}(\widehat{\beta}-\beta), \end{aligned}$$

for some $\widetilde{\beta}$ lies between $\widehat{\beta}$ and β. Manipulations similar to those in the proof of theorem 1 of Andrews (1991) lead to

$$\begin{aligned} & \sqrt{\frac{\rho}{T^2}}\frac{\partial^2}{\partial\beta\partial\beta'}\widetilde{\Omega}_{k_\rho}(\widetilde{\beta}) \\ \leq \;& \sqrt{\frac{\rho}{T^2}}\sum_{-T+1}^{T-1}\left|\left(1-\frac{|h|}{T}\right)^\rho\right|\frac{1}{T}\sum_{t=|h|+1}^{T}\sup_{\beta\in\varkappa}\parallel\frac{\partial^2}{\partial\beta\partial\beta'}v_t(\beta)v_{t-|h|}(\beta)\parallel \\ = \;& \sqrt{\rho}(\frac{1}{T}\sum_{-T+1}^{T-1}\left|\left(1-\frac{|h|}{T}\right)^\rho\right|)O_p(1) \\ = \;& o_p(1), \end{aligned}$$

where the last equality follows from the fact, shown earlier, that $\frac{1}{T}\sum_{-T+1}^{T-1}|(1-\frac{|h|}{T})^\rho| = O(\frac{1}{\rho})$.

The proof of the rest of the theorem involves calculations similar to those given above and in Andrews (1991) and is therefore omitted.

Proof of Theorems 4 and 5 These results follow directly from standard weak convergence arguments.

9 Notation

$o_{a.s.}(1)$	tends to zero almost surely	K_{mm}	$m^2 \times m^2$ commutation matrix
$O_{a.s}(1)$	bounded almost surely	$\otimes$	Kronecker product
$\rightarrow_d, \Longrightarrow$	weak convergence	$vec(A)$	vectorization by columns
$\rightarrow_p \rightarrow_{a.s.}$	convergence in probability, almost surely	$[\cdot]$	integer part
$\int_0^1 f$	$\int_0^1 f(r)dr$	$\mathrm{tr}\{A\}$	trace of A
$W_p(r)$	$p-$dimensional standard Brownian motion	$\sum$	$\sum_{t=1}^{T}$
$V_p(r)$	$p-$dimensional standard Brownian bridge	$\mathbb{R}$	$(-\infty, \infty)$
KVB	Kiefer, Vogelsang and Bunzel (2000)	OLS	Ordinary least squares
KV	Kiefer and Vogelsang (2002a, 2002b)	LRV	Long run variance

References

[1] Anderson, T. W. and D. A. Darling (1952): "Asymptotic Theory of Certain 'goodness of fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212.

[2] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817–854.

[3] Andrews, D. W. K. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica,* 60, 953–966.

[4] de Jong, R. M. and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices," *Econometrica* 68, 407–424.

[5] den Haan, W. J. and A. Levin (1997): "A Practitioners Guide to Robust Covariance Matrix Estimation," in G. Maddala and C. Rao (eds), *Handbook of Statistics: Robust Inference,* Volume 15, Elsevier, New York.

[6] Hannan, E. J. (1970): *Multiple Time Series*, New York, Wiley.

[7] Hansen, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes," *Econometrica* 60, 967–972.

[8] Jansson, M. (2002): "Autocorrelation robust tests with good size and power," Department of Economics, University of California, Berkeley.

[9] Kiefer, N. M., T. J., Vogelsang and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica* 68, 695–714.

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

[11] ——— (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

[12] Lee, C. and P. C. B. Phillips (1994): "An ARMA Prewhitened Long-run Variance Estimator," working paper, Yale University.

[13] Linton, O. B. (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63, 1079–1112.

[14] Magnus, J. R. and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7, 381–394.

[15] ——— (1988): *Matrix Differential Calculus with Application in Statistics and Econometrics*, New York, Wiley.

[16] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55, 703–708.

[17] ——— (1994): "Automatic Lag Selection in Covariance Estimation," *Review of Economic Studies*, 61, 631–654.

[18] Parzen, E. (1957): "On the Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

[19] Phillips, P. C. B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

[20] Phillips, P. C. B. (1995): "Fully Modified Least Squares and Vector Autoregression," *Econometrica* 63, 1023–1078.

[21] Priestley, M. B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press.

[22] Robinson, P. (1995): "Gaussian Semiparametric Estimation of Long Range Dependence," *Annals of Statistics*, 1630–1661.

[23] Shorack, G. R. and A. J. Wellner (1986): *Empirical Processes with Applications to Statistics*, John Wiley & Sons.

[24] White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

[25] Xiao, Z. and P. C. B. Phillips (1998): "Higher Order Approximations for Frequency Domain Time Series Regression," *Journal of Econometrics*, Vol. 86, pp. 297–336.

[26] Xiao, Z. and P. C. B. Phillips (2002): "Higher Order Approximations for Wald Statistics in Time Series Regressions with Integrated Processes," *Journal of Econometrics*, 108, pp. 157–198.