View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Economics School of Economics

3-2007

Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Yixiao SUN
University of California - San Diego

Sainan JIN
Singapore Management University, snjin@smu.edu.sg

DOI: https://doi.org/10.1016/j.jspi.2006.06.033

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research

Part of the Econometrics Commons

Citation

PHILLIPS, Peter C. B.; SUN, Yixiao; and JIN, Sainan. Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation. (2007). *Journal of Statistical Planning and Inference*. 137, (3), 985-1023. Research Collection School Of Economics.
Available at: https://ink.library.smu.edu.sg/soe_research/318

This Journal Article is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in Journal of Statistical Planning and Inference
Volume 137, Issue 3, 1 March 2007, Pages 985-1023
https://doi.org/10.1016/j.jspi.2006.06.033

Long run variance estimation and robust regression testing using sharp origin kernels with no truncation

Peter C.B. Phillips[a, b, c, *], Yixiao Sun[d], Sainan Jin[e]

[a] *Cowles Foundation, Yale University, USA*
[b] *University of Auckland, New Zealand*
[c] *University of York, USA*
[d] *Department of Economics, University of California, San Diego, USA*
[e] *Guanghua School of Management, Peking University, China*

Abstract

A new family of kernels is suggested for use in 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, Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory 18, 1350–1366, 2002b, Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation, Econometrica 70, 2093–2095] 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 LRV estimates. 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. A data-driven method for selecting the power parameter is recommended for hypothesis testing. Simulations show that this method gives arise to a test with more accurate size than the conventional HAC t-test at the cost of a very small power loss.

Keywords: Heteroscedasticity and autocorrelation consistent standard error; Data-determined kernel estimation; Long run variance; Power parameter; Sharp origin kernel

1. Introduction and overview

One of the many areas where Madan Puri has made major contributions to statistical theory is nonparametrics (see Hall et al., 2003). Nonparametric methods are now very popular in econometrics both in time series and cross section data applications. Indeed, much of modern econometrics is concerned with attempts to achieve generality wherever

* Corresponding author. Cowles Foundation, Yale University, USA. Tel.: +1 203 4323695; fax: +1 203 4326167.
E-mail address: peter.phillips@yale.edu (P.C.B. Phillips).

possible by using nonparametric techniques, while retaining specificity wherever the model connects most closely with underlying economic ideas to be tested. In practical empirical econometric work, nonparametric methods have been used most extensively to obtain covariance matrix estimates and statistical tests that are robust to heteroskedasticity and autocorrelation in the data. This topic is the subject of the present contribution. We first overview a recent direction in econometric method on this topic and motivate our own work.

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 regression. In this spirit, Kiefer et al. (2000; hereafter, KVB) and Kiefer and Vogelsang (2002a,b; hereafter, KV) have recently 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 (2004) explained this improved accuracy of the null approximation, showing that the error rejection probability (ERP) in a Gaussian location model of these tests is $O(\log T/T)$, where T is the sample size, in contrast to the rate of $O(T^{-1/2})$ that is attained by tests using conventional HAC estimates.[2]

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. A local power analysis in Kiefer and Vogelsang (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). Our main contribution is to provide a new approach to HAC estimation that embeds the Bartlett kernel in a new class of sharp origin (SO) 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 SO 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 SO kernel is inconsistent just as it is when $\rho = 1$. However, compared with the Bartlett kernel ($\rho = 1$), SO 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. SO kernels can achieve variance reductions in this way, while continuing to deliver tests with better size than conventional Bartlett-kernel-based tests.

Our findings indicate that SO kernels without truncation deliver asymptotically valid tests with greater accuracy in size and power close to or better than that of conventional tests. The simulation results show that tests based on SO kernels ($\rho > 1$) without truncation is uniformly more powerful than those based on the Bartlett kernel ($\rho = 1$) without

[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.

[2] The $O(T^{-1/2})$ rate is obtained by letting $d = 1$ and $M = T^{1/2}$ in Eq. (11) of Velasco and Robinson (2001).

truncation. 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.

When $\rho \to \infty$ with T, SO kernels provide a new mechanism for consistent HAC (and, more generally, spectral density) estimation. While test validity is retained whatever the choice of ρ, there is an opportunity for optimal choice of ρ. To this end, we develop an asymptotic distribution theory for consistent LRV estimation using SO 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 $\rho = \mathrm{O}(T^{2/3})$ 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. Two data-driven methods are proposed and their finite sample performances are investigated using simulations.

The present contribution is related to recent work by Kiefer and Vogelsang (2005) and Hashimzade and Vogelsang (2006). These authors consider LRV and spectral density estimation using traditional kernels when the bandwidth (M) is set proportional to the sample size (T), i.e. $M = bT$ for some $b \in (0, 1)$. Their approach is equivalent to contracting traditional kernels $k(\cdot)$ to get $k_b(x) = k(x/b)$ and using the contracted kernels $k_b(\cdot)$ in the LRV and spectral density estimation without truncation. Our approach is to exponentiate the traditional kernels $k(\cdot)$ to get $k_\rho(x) = k^\rho(x)$ and use it in estimation and testing without truncation. Both contracted kernel and exponentiated kernels are designed to improve the power of existing robust regression tests with no truncation. The associated estimators and tests share many properties. For example, the size distortion and power of the new robust regression tests increase as ρ increases or b decreases. Nevertheless, it is difficult to characterize the exact relationship between these two types of strategies. In the special cases when exponential type kernels such as $k(x) = \exp(-|x|)$ is used, these two strategies lead to identical estimators and statistical tests when ρ and b are appropriately chosen (i.e. $\rho = 1/b$). Exponential kernels of this type have not been used before in LRV estimation and appear in spectral density estimation only in the Abel estimate (c.f. Hannan, 1970, p. 279).

The rest of the paper is organized as follows. Section 2 briefly overviews testing problems in the presence of nonparametric autocorrelation. Section 3 introduces SO kernels and establishes the asymptotic properties of HAC estimators using these kernels without truncation when the power parameter ρ is fixed. Section 4 develops the asymptotic theory for the case when $\rho \to \infty$ with T and extracts optimal values of ρ based on an MSE criterion. Section 5 provides a limit theory for regression tests using SO kernels under both null and local alternatives. Section 6 reports simulation results on the finite sample performances 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 Appendix.

2. Robust testing of regression hypotheses

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

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

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

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

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, 2000; Jansson, 2004).

Assumption 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 (LRV) 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

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

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; 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 LRV of v_t is

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

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

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

$$\widehat{\Gamma}(j) = \begin{cases} \dfrac{1}{T}\sum\limits_{t=1}^{T-j} \widehat{v}_{t+j}\widehat{v}_t' & \text{for } j \geqslant 0, \\ \dfrac{1}{T}\sum\limits_{t=-j+1}^{T} \widehat{v}_{t+j}\widehat{v}_t' & \text{for } j < 0 \end{cases} \tag{9}$$

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 is 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; 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

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

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 behavior 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 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 (2004) 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.

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.

3. SO kernels and HAC estimation

We define a class of SO kernels by taking an arbitrary power $\rho \geqslant 1$ of the usual Bartlett kernel, giving

$$k_\rho(x) = \begin{cases} (1 - |x|)^\rho, & |x| \leqslant 1 \\ 0, & |x| > 1 \end{cases} \quad \text{for } \rho \in \mathbb{Z}^+. \tag{11}$$

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.

(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

$$\kappa_1 = \lim_{x\to 0} \frac{1 - (1 - |x|)^\rho}{|x|} = \rho < \infty. \tag{12}$$

(iii) $k_\rho(x)$ is positive semi-definite for any $\rho \in \mathbb{Z}^+$.

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

$$\widehat{\Omega}_\rho = \sum_{j=-T+1}^{T-1} k_\rho\left(\frac{j}{T}\right)\widehat{\Gamma}(j). \tag{13}$$

In what follows in this section we will assume that the ρ value in (13) is fixed as $T \to \infty$.

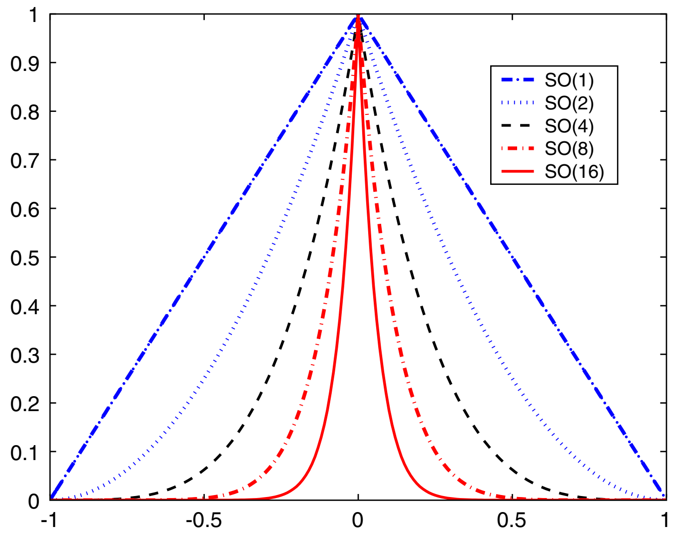

Fig. 1. Sharp origin (SO) kernels $k_\rho(x)$ for $\rho \in [1, 16]$.

The following theorem establishes the asymptotic properties of the HAC estimator $\widehat{\Omega}_\rho$ when a SO kernel is used.

Theorem 1. *Let* A1 *hold, then the following results hold*:

(a) $\widehat{\Omega}_\rho \Rightarrow \Lambda \Xi_\rho \Lambda'$, *where* $\Xi_\rho = \int_0^1 \int_0^1 k_\rho(r-s)\,\mathrm{d}V_m(r)\,\mathrm{d}V_m'(s)$ *and* $V_m(r)$ *is an m-vector of standard Brownian bridges*.
(b) $E(\Lambda \Xi_\rho \Lambda') = \mu_\rho \Omega$, *where* $\mu_\rho = 1 - \int_0^1 \int_0^1 k_\rho(r-s)\,\mathrm{d}r\,\mathrm{d}s$.
(c) $var(\mathrm{vec}(\Lambda \Xi_\rho \Lambda')) = v_\rho (I_{m^2} + K_{mm}) \Omega \otimes \Omega$ *where*

$$v_\rho = \int k_\rho(r-s)k_\rho(p-q) - 2\int k_\rho(r-s)k_\rho(r-q) + \int k_\rho(r-s)^2$$

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

$$\int k_\rho(r-s)k_\rho(p-q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 k_\rho(r-s)k_\rho(p-q)\,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}p\,\mathrm{d}q.$$

Remarks.

(a) Part (a) of Theorem 1 holds for more general kernels. For example, if $k(\cdot)$ is twice continuously differentiable, then it follows from the proof that

$$\sum_{j=-T+1}^{T-1} k(j/T)\widehat{\Gamma}(j) \;\Rightarrow\; \Lambda \Xi \Lambda', \tag{14}$$

where $\Xi = \int_0^1 \int_0^1 k(r-s)\,\mathrm{d}V_m(r)\,\mathrm{d}V_m'(s)$. This result then implies parts (b) and (c) with $k_\rho(\cdot)$ replaced by $k(\cdot)$.
(b) As shown in the proof of the theorem, an alternative representation of Ξ_ρ is

$$\Xi_\rho = \int_0^1 \int_0^1 k_\rho^*(r, s)\,\mathrm{d}W_m(r)\,\mathrm{d}W_m'(s),$$

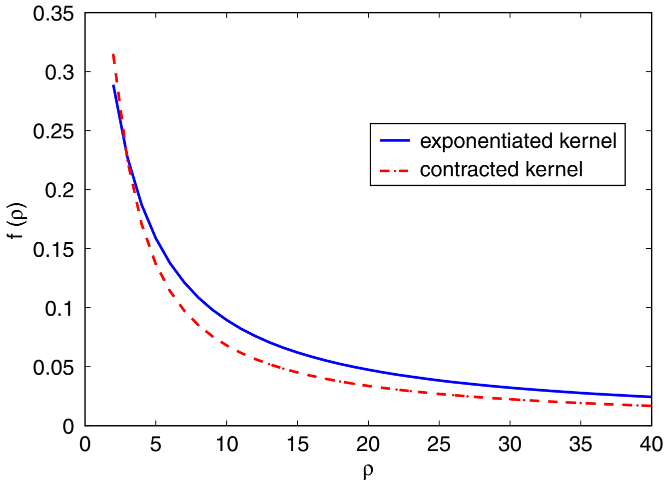

Fig. 2. Scale functions for exponentiated kernels and contracted kernels.

where

$$k^*_\rho(r,s) = k_\rho(r-s) - \int_0^1 k_\rho(r-t)\,\mathrm{d}t - \int_0^1 k_\rho(\tau - s)\,\mathrm{d}\tau + \int_0^1 k_\rho(t-\tau)\,\mathrm{d}t\,\mathrm{d}\tau,$$

and then

$$\mu_\rho = \int_0^1 \int_0^1 k^*_\rho(r,r)\,\mathrm{d}r, \quad v_\rho = \int_0^1 \int_0^1 [k^*_\rho(r,s)]^2\,\mathrm{d}r\,\mathrm{d}s.$$

(c) Theorem 1 tells us that $\widehat{\Omega}_\rho/\mu_\rho \rightarrow_d \xi_\Omega := \Lambda \Xi_\rho \Lambda'/\mu_\rho$ and

$$E\xi_\Omega = \Omega,$$

$$\mathrm{var}(\mathrm{vec}(\xi_\Omega)) = v_\rho \mu_\rho^{-2}(I_{m^2} + K_{mm})\Omega \otimes \Omega. \quad (15)$$

Hence, $\widehat{\Omega}_\rho/\mu_\rho$ is asymptotically unbiased with asymptotic variance matrix $v_\rho\mu_\rho^{-2}(I_{m^2} + K_{mm})\Omega \otimes \Omega$. In seeking a preferred kernel, it might first appear reasonable to choose a ρ that minimizes the scale factor $f(\rho) = v_\rho\mu_\rho^{-2}$. It is easy to show that

$$f(\rho) = \left\{ \left(\frac{2}{\rho+2}\right)^2 + \frac{1}{\rho+1} - \frac{4}{(\rho+1)^2}\left(\frac{4\rho^2+7\rho+2}{(2\rho+3)(\rho+2)} + \frac{\Gamma^2(\rho+2)}{\Gamma(2\rho+4)}\right)\right\}\left(\frac{\rho}{\rho+2}\right)^{-2}.$$

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. However, the vanishing of the asymptotic variance is obtained under the sequential limit in which $T \to \infty$ for a fixed ρ and then passing $\rho \to \infty$. Such a limit theory may not capture the behavior of $\widehat{\Omega}_\rho/\mu_\rho$ in finite samples very well. In the next section, we allow $T \to \infty$ and $\rho \to \infty$ at the same time. We show that the absolute bias of $\widehat{\Omega}_\rho/\mu_\rho$ increases as ρ increases and that there is an opportunity to choose ρ to balance the asymptotic bias and the asymptotic variance.

(d) Since the contracted kernel $\bar{k}_\rho(x) = (1-\rho|x|)1\{\rho|x| \leqslant 1\}$ and the exponentiated kernel $k_\rho(x)$ have the same first order expansion near the origin, it is of interest to find the scale factor analogous to $f(\rho)$ for the estimator

$$\bar{\Omega}_\rho = \sum_{j=-T+1}^{T-1} \left(1 - \rho\left|\frac{j}{T}\right|\right) 1\left\{\rho\left|\frac{j}{T}\right| \leqslant 1\right\} \widehat{\Gamma}(j). \quad (16)$$

Following a similar proof, we can prove that Theorem 1 holds for $\bar{k}_\rho(x)$ with $k_\rho(x)$ replaced by $\bar{k}_\rho(x)$. Direct calculations show that the scale factor $\bar{f}(\rho)$ for $\bar{k}_\rho(x)$ when $\rho \geqslant 2$ is

$$\bar{f}(\rho) = \left(\frac{1}{90}\,\frac{10+42\rho-105\rho^2+60\rho^3}{\rho^4}\right)\left(\frac{1}{3}\,\frac{1-3\rho+3\rho^2}{\rho^2}\right)^{-2}.$$

Fig. 2 shows that $f(\rho)$ is larger than $\bar{f}(\rho)$ when $\rho \geqslant 3$, implying that for a given ρ, $\widehat{\Omega}_\rho/\mu_\rho$ has a larger variability than $\bar{\Omega}_\rho/\bar{\mu}_\rho$.

4. HAC estimation with SO kernels

4.1. Some new asymptotics with $\rho \to \infty$

This section develops an asymptotic theory for the HAC estimator $\widehat{\Omega}_\rho$ when $\rho \to \infty$ as $T \to \infty$. Under certain rate conditions on ρ, we show that $\widehat{\Omega}_\rho$ 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 $\widetilde{\Omega}_\rho$ be this pseudo-estimator, which is identical to $\widehat{\Omega}_\rho$ except that it is based on the unobserved sequence $v_t = v_t(\beta) = x_t u_t$ rather than $v_t = v_t(\widehat{\beta}) = x_t \widehat{u}_t$, i.e.

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

where

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

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

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

and

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

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 $\rho = \mathrm{o}(T/\log T)$. 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 $b = \frac{2}{3}$.

Assumption A2. $v_t = x_t u_t$ is a mean zero, fourth order stationary linear process

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

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

Assumption A3. $(1/\rho) + (\rho \log T)/T \to 0$, as $T \to \infty$.

Define the spectral window

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

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

$$\widetilde{\Gamma}(h) = (2\pi/T)\sum_{s=0}^{T-1} I_{vv}(\lambda_s) e^{i\lambda_s h},$$

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 $\widetilde{\Omega}_\rho = 2\pi \widetilde{f}_{vv}(0)$. 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 2. *Let $\rho = aT^b \to \infty$ for some $a > 0$ and $0 < b < 1$. Then, for all $\lambda_s = 2\pi s/T$, $s = 0, 1, \ldots, [T/2]$, we have as $T \to \infty$*

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

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

where the o(1) *terms hold uniformly over s.*

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

$$K_\rho(\lambda_s) = \begin{cases} O\left(\dfrac{T}{\rho}\right) & s \leqslant \rho \text{ and } s \geqslant T - \rho, \\ O\left(\dfrac{\rho T}{s^2}\right) & \rho < s < T - \rho. \end{cases} \tag{22}$$

So, for frequencies λ_s in the vicinity of the origin such that $\lambda_s = 2\pi s/T = O(\rho/T)$ with ρ satisfying A3, the spectral window $K_\rho(\lambda_s) = O(T/\rho)$ diverges, while for all frequencies $\lambda_s \to \lambda \in (0, 2\pi)$, $K_\rho(\lambda_s) = O(\rho/T) = o(1)$. Thus, Lemma 2 shows that when $\rho \to \infty$ the SO 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

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

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

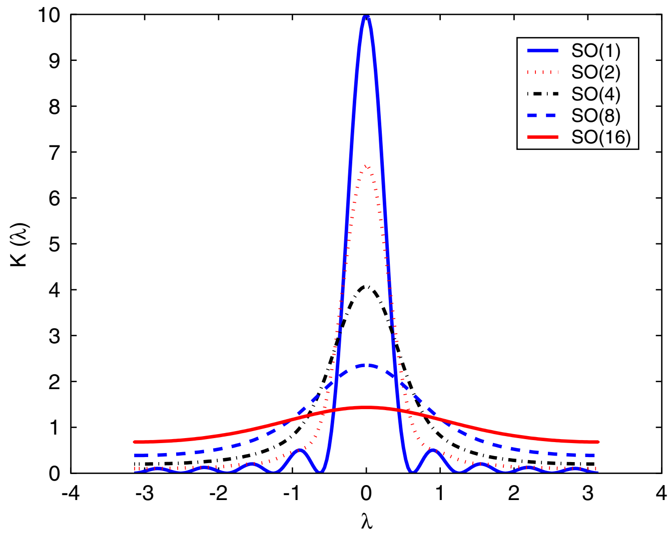

Fig. 3. Spectral window $K_\rho(\lambda)$ for Bartlett ($\rho = 1$) and sharp origin kernels k_ρ with $\rho = 2, 4, 8, 16$ and $T = 10$.

Note that the side lobes of the Fejer kernel are smoothed out in the SO 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 $K_\rho(0) = \mathrm{O}(T/\rho)$, as is clear from (22).

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

Theorem 3. *Suppose* A1–A3 *hold and* $\rho = aT^b \to \infty$ *for some* $a > 0$ *and* $0 < b < 1$. *Then*:

(a) $\lim_{T\to\infty} \rho \mathrm{Var}(\mathrm{vec}(\widetilde{\Omega}_\rho)) = (I_{m^2} + K_{mm})(\Omega \otimes \Omega)$.
(b) *If* $b < \frac{2}{3}$ *then*

$$\sqrt{\rho}(\mathrm{vec}(\widetilde{\Omega}_\rho) - \mathrm{vec}(\Omega)) \to_d \mathrm{N}(0, (I_{m^2} + K_{mm})(\Omega \otimes \Omega)).$$

(c) $\lim_{T\to\infty}(T/\rho)(E\widetilde{\Omega}_\rho - \Omega) = -\Omega^{(1)}$.
(d) *If* $\rho^3/T^2 \to \vartheta \in (0, \infty)$, *then*

$$\lim_{T\to\infty} \mathrm{MSE}(\rho, \widetilde{\Omega}_\rho, W) = \vartheta\, \mathrm{vec}(\Omega^{(1)})' W\, \mathrm{vec}(\Omega^{(1)}) + \mathrm{tr}\{W(I_{m^2} + K_{mm})(\Omega \otimes \Omega)\}.$$

Remarks.

(a) It is not surprising that the results in Theorem 3 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 SO kernel k_ρ with $\rho(T) = \mathrm{O}(T^{2/3})$ 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 3 gives a CLT for the new HAC estimator $\widetilde{\Omega}_\rho$. $\widetilde{\Omega}_\rho$ 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 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).

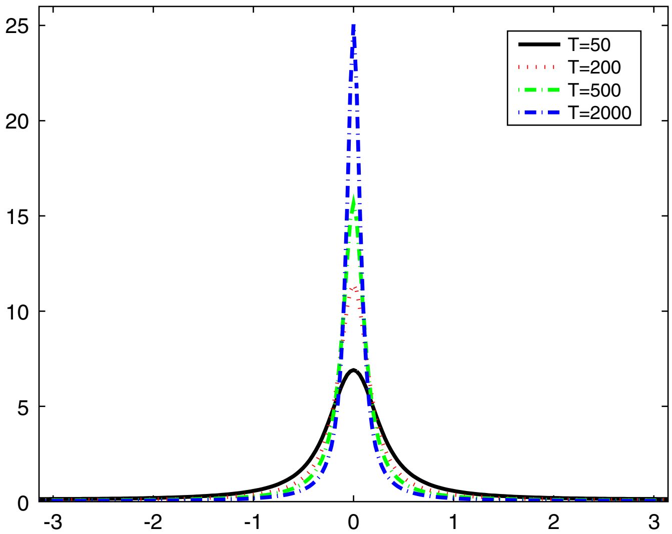

Fig. 4. Spectral window $K_\rho(\lambda)$ of the sharp origin kernel with $\rho = \mathrm{O}(T^{2/3})$.

Next, we give a corresponding result for the feasible HAC estimator $\widehat{\Omega}_\rho$, 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.

$$\mathrm{MSE}_h(\rho, \widehat{\Omega}_\rho, W_T) = E \min\{\rho \operatorname{vec}(\widehat{\Omega}_\rho - \Omega)' W_T \operatorname{vec}(\widehat{\Omega}_\rho - \Omega), h\},$$

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 Appendix.

Theorem 4. *Let* A1–A3 *and* B *hold. Suppose* $\rho^3/T^2 \to \vartheta \in (0, \infty)$ *as* $T \to \infty$. *Then*

(a) $\sqrt{\rho}(\widehat{\Omega}_\rho - \Omega) = \mathrm{O}_\mathrm{p}(1)$, $\sqrt{\rho}(\widehat{\Omega}_\rho - \widetilde{\Omega}_\rho) \to_p 0$; *and*
(b)

$$\begin{aligned}
&\lim_{h\to\infty} \lim_{T\to\infty} \mathrm{MSE}_h(\rho, \widehat{\Omega}_\rho, W) \\
&\quad = \lim_{h\to\infty} \lim_{T\to\infty} \mathrm{MSE}_h(\rho, \widetilde{\Omega}_\rho, W) \\
&\quad = \lim_{T\to\infty} \mathrm{MSE}(\rho, \widetilde{\Omega}_\rho, W) \\
&\quad = \vartheta \operatorname{vec}(\Omega^{(1)})' W \operatorname{vec}(\Omega^{(1)}) + \operatorname{tr}\{W(I_{m^2} + K_{mm})\Omega \otimes \Omega\}.
\end{aligned} \tag{24}$$

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

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

Then, using (24), the optimal ρ is

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

The selection ρ_T^* will lead to HAC estimates $\widehat{\Omega}_\rho$ 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 $\widehat{\delta} = \delta(\widehat{\Omega}, \widehat{\Omega}^{(1)})$. 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 $\widehat{\Omega}_\rho$, we have

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

where

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

$\widehat{v}_t = x_t(y_t - x_t'\widehat{\beta})$, $\widehat{v}_{t,i}$ is ith element of $\widehat{v}_t$, and $\widehat{\beta}$ is defined in (2). In the special case when $m = 1$, the data-determined power parameter is

$$\hat{\rho}_T^* = \hat{\delta}^{1/3} T^{2/3} \quad \text{with } \hat{\delta} = \frac{(1-\hat{\alpha}^2)^2}{4\hat{\alpha}^2}. \tag{29}$$

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

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

converges to zero at the rate $O(T^{-2/3})$. 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, $\widehat{\Omega}_\rho$ may be expected to have asymptotic performance characteristics similar to those of conventional consistent HAC estimates with optimal bandwidth choices.

When $\rho = \rho_T^*$, the asymptotic MSE (AMSE) as given in (24) is

$$\text{AMSE} = \frac{3}{2} \operatorname{tr}[W(I_{m^2} + K_{mm})(\Omega \otimes \Omega)].$$

Using Proposition 1 in Andrews (1991), we can show that for the conventional Bartlett-kernel-based estimator with MSE-optimal bandwidth, the AMSE is

$$\text{AMSE} = \operatorname{tr}[W(I_{m^2} + K_{mm})(\Omega \otimes \Omega)].$$

Therefore, the sharp Bartlett kernel estimator is asymptotically 50% less efficient than the conventional Bartlett kernel estimator.

However, as is well known, a long run variance estimator with optimal asymptotic MSE properties does not necessarily deliver the best estimate in finite samples or, more specifically, a test with good size properties in finite samples. In fact, as argued later in the paper and in keeping with the results in Jansson (2004), some variability in the LRV variance estimator assists in better approximating the finite sample behavior of the t^* or F^* statistics under the null hypothesis.

This is one of the main reasons why inconsistent HAC estimates have been advocated in regression testing. They are particularly useful when size distortion in conventional procedures is a major concern.

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 $\mathrm{H}_0 : R\beta = r$ against the alternative $\mathrm{H}_1 : R\beta \neq r$. Using the estimate $\widehat{\Omega}_\rho$, we can construct the F-ratio in the usual way

$$F^*(\widehat{\Omega}_\rho) = T(R\widehat{\beta} - r)'(R\widehat{Q}^{-1}\widehat{\Omega}_\rho\widehat{Q}^{-1}R')^{-1}(R\widehat{\beta} - r)/p, \tag{30}$$

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

$$t^*(\widehat{\Omega}_\rho) = T^{1/2}(R\widehat{\beta} - r)(R\widehat{Q}^{-1}\widehat{\Omega}_\rho\widehat{Q}^{-1}R')^{-1/2}. \tag{31}$$

The limit distributions of F^* and t^* under the null hypothesis and local alternatives when ρ is fixed are given in the following theorem:

Theorem 5. *Let* A1 *and* A2 *hold. If ρ is fixed, then*

(a) *Under* H_0

$$F^*(\widehat{\Omega}_\rho) \;\Rightarrow\; W_p'(1)\left(\int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}V_p(r)\,\mathrm{d}V_p'(s)\right)^{-1} W_p(1)/p, \tag{32}$$

$$t^*(\widehat{\Omega}_\rho) \;\Rightarrow\; W_1(1)\left(\int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}V_1(r)\,\mathrm{d}V_1'(s)\right)^{-1/2}. \tag{33}$$

(b) *Under the local alternative* $\mathrm{H}_1 : R\beta = r + cT^{-1/2}$

$$F^*(\widehat{\Omega}_\rho) \;\Rightarrow\; (\Lambda^{*-1}c + W_p(1))'\left(\int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}V_p(r)\,\mathrm{d}V_p'(s)\right)^{-1}(\Lambda^{*-1}c + W_p(1))/p, \tag{34}$$

$$t^*(\widehat{\Omega}_\rho) \;\Rightarrow\; (\gamma + W_1(1))\left(\int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}V_1(r)\,\mathrm{d}V_1(s)\right)^{-1/2}, \tag{35}$$

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

When ρ is sample size dependent and satisfies A3, we know from Theorem 3 that $\widehat{\Omega}_\rho$ is consistent. In this case, F^* and t^* have conventional chi-square and normal limits.

Theorem 6. *Let* A1–A3 *hold. Then*

(a) *under the null hypothesis*

$$pF^*(\widehat{\Omega}_\rho) \;\Rightarrow\; W_p'(1)W_p(1) =_{\mathrm{d}} \chi_p^2, \quad t^*(\widehat{\Omega}_\rho) \;\Rightarrow\; W_1(1) =_{\mathrm{d}} \mathrm{N}(0,1); \tag{36}$$

(b) *under the local alternative hypothesis* $\mathrm{H}_1 : R\beta = r + cT^{-1/2}$

$$pF^*(\widehat{\Omega}_\rho) \;\Rightarrow\; (\Lambda^{*-1}c + W_p(1))'(\Lambda^{*-1}c + W_p(1)), \quad t^*(\widehat{\Omega}_\rho) \;\Rightarrow\; (\gamma + W_1(1)). \tag{37}$$

Table 1
Asymptotic critical values and fitted hyperbola for one-sided t-test

	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
$\rho = \infty$	1.282	1.645	1.960	2.326
a	−0.410	−0.457	−0.469	−0.513
b	2.103	3.127	4.329	5.637
c	1.282	1.645	1.960	2.326
s.e.	0.004	0.006	0.008	0.024
R^2	1.000	1.000	1.000	0.999

From the forms of (32)–(35), the F^* and t^* statistics 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 test. 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

$$\int_0^1 \int_0^1 k_\rho(r - s)\, \mathrm{d}V_p(r)\, \mathrm{d}V_p'(s) \to_p \int_0^1 \mathrm{d}r\, I_p = I_p \quad \text{as } \rho \to \infty, \tag{38}$$

in view of the fact that

$$\mathrm{d}V_p(r)\, \mathrm{d}V_p'(r) = \mathrm{d}[V_p]_r = \mathrm{d}r\, I_p,$$

where $[V_p]_r$ is the quadratic variation (matrix) process of V_p. It follows from (38) 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 $\widehat{\Omega}_\rho$ with large ρ to have power similar to that of conventional tests. When $\rho \to \infty$, these tests will have power functions equivalent to the power envelope.

Given the above asymptotic distributions, the critical values can be obtained by simulations. The first part of Table 1 contains the critical values of the t-test for selected ρ values (including the asymptotic case which is represented as $\rho = \infty$). The Brownian motion and Brownian bridge processes are approximated using normalized partial sums of $T = 1000$ iid N(0, 1) random variables and the simulation involves 50,000 replications. For other values of ρ, the critical values can be represented approximately by a hyperbola of the form

$$cv = \frac{b}{\rho - a} + c, \tag{39}$$

where c is the critical value from the standard normal. The second part of Table 1 gives nonlinear least squares estimates of a and b. The standard errors are seen to be very small. The R^2, defined as the ratio of the sum of squared errors to that of the critical values, is almost indistinguishable from one. Both the standard errors and the R^2s indicate that the hyperbola explains the critical values very well. In view of the (a, b) values, it is easy to see that as ρ increases, the fitted hyperbola approaches its asymptote very quickly. For example, when the significance level is 95%, the critical values are very close to those from the standard normal when $\rho > 32$.

Fig. 5 presents the asymptotic power curves computed by simulation. The experiment design is the same as before. Asymptotic power is computed for the t^*-test at the 95% significance level using SO kernels with $\rho = 1, 2, 16$ and for $\gamma \in [0, 5]$. As is apparent from Fig. 5, the power curve moves up uniformly as ρ increases, consonant with the asymptotic theory implied by (35) and (38). 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

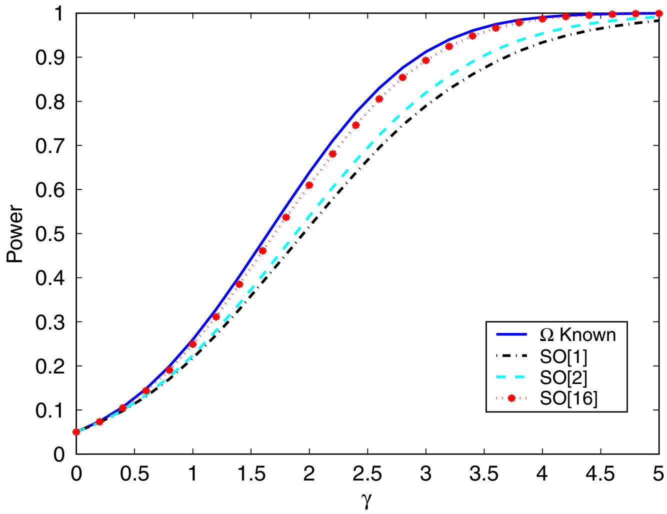

Fig. 5. Asymptotic local power function of the t^*-test.

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 6 applies.

Comparing the asymptotic powers for different ρ in Fig. 5, it is apparent that tests based on SO 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, SO kernels will also dominate these commonly used kernels as far as the power of the test is concerned.

The reason for the domination by the Bartlett kernel found by KV is related to the argument given above for the SO kernel domination. As is apparent from the form of the power functions given in (34) and (35), the effect of the choice of kernel k on the power function is manifest in the quadratic functional $\int_0^1 \int_0^1 k(r-s)\, \mathrm{d}V_p(r)\, \mathrm{d}V_p'(s)$. Since k is generally decreasing away from the origin (for many kernels 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 SO 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 SO 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 $\mu = 0$, $u_t = a_1u_{t-1} + a_2u_{t-2} + e_t$, $x_t = bx_{t-1} + \eta_t$, $b = 0.5$, e_t and η_t are iid N(0, 1) with $\mathrm{cov}(e_t, \eta_t) = 0$, and $x_0 = u_0 = u_{-1} = 0$. The simulation results are based on 50,000 replications. We consider the one-sided null hypothesis $\mathrm{H}_0 : \beta \leqslant 0$ against the alternative $\mathrm{H}_1 : \beta > 0$. The regression parameter β is estimated by OLS and the t^*-statistic is constructed as in (31). As a benchmark, we also construct the conventional (i.e., bandwidth truncated) t-statistic using the Bartlett kernel which we label t_{HAC} or t_{HAC}-SO(1) if we want to emphasize that the Bartlett kernel is used. In computing t_{HAC}, the bandwidth is chosen by the data-driven procedure proposed in Andrews (1991). We also report t 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 2a and b present the finite sample null rejection probabilities with no prewhitening and with prewhitening, respectively. Rejections were determined using asymptotic 95% critical values from Table 1. We draw attention to three aspects of Tables 2a and b.

Table 2
Finite sample null hypothesis rejection probabilities at 5% nominal level

T	a_1	a_2	t_{HAC}	t^*-SO(1)	t^*-SO(2)	t^*-SO(4)	t^*-SO(8)	t^*-SO(16)
(a) *With no prewhitening*								
50								
	−0.500	0.000	0.057	0.050	0.050	0.049	0.048	0.044
	0.000	0.000	0.073	0.061	0.061	0.062	0.062	0.063
	0.300	0.000	0.088	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	0.113	0.078	0.079	0.082	0.086	0.093
	0.900	0.000	0.125	0.082	0.084	0.088	0.094	0.105
	0.950	0.000	0.129	0.082	0.084	0.089	0.096	0.107
	0.990	0.000	0.131	0.081	0.084	0.090	0.099	0.112
	1.500	−0.750	0.113	0.073	0.074	0.078	0.084	0.094
	1.900	−0.950	0.135	0.080	0.083	0.090	0.101	0.115
	0.800	0.100	0.124	0.081	0.084	0.088	0.094	0.105
100								
	−0.500	0.000	0.051	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	0.074	0.058	0.060	0.061	0.063	0.064
	0.500	0.000	0.082	0.063	0.065	0.065	0.067	0.070
	0.700	0.000	0.092	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	0.103	0.073	0.074	0.075	0.079	0.086
	0.990	0.000	0.104	0.069	0.070	0.073	0.079	0.086
	1.500	−0.750	0.088	0.065	0.066	0.067	0.070	0.075
	1.900	−0.950	0.106	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
200								
	−0.500	0.000	0.048	0.049	0.049	0.049	0.048	0.047
	0.000	0.000	0.056	0.054	0.054	0.054	0.053	0.054
	0.300	0.000	0.066	0.055	0.056	0.056	0.056	0.057
	0.500	0.000	0.071	0.056	0.057	0.057	0.058	0.060
	0.700	0.000	0.076	0.058	0.058	0.059	0.060	0.063
	0.900	0.000	0.082	0.061	0.061	0.062	0.063	0.067
	0.950	0.000	0.083	0.061	0.061	0.062	0.063	0.067
	0.990	0.000	0.084	0.062	0.062	0.063	0.065	0.069
	1.500	−0.750	0.073	0.055	0.056	0.057	0.057	0.060
	1.900	−0.950	0.086	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
(b) *With prewhitening*								
50								
	−0.500	0.000	0.065	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	0.084	0.065	0.067	0.069	0.071	0.073
	0.500	0.000	0.089	0.067	0.070	0.071	0.074	0.076
	0.700	0.000	0.094	0.070	0.073	0.075	0.077	0.079
	0.900	0.000	0.102	0.071	0.075	0.078	0.080	0.084
	0.950	0.000	0.103	0.070	0.075	0.078	0.082	0.085
	0.990	0.000	0.105	0.067	0.074	0.079	0.085	0.088
	1.500	−0.750	0.082	0.063	0.065	0.067	0.069	0.071
	1.900	−0.950	0.107	0.069	0.071	0.075	0.080	0.084
	0.800	0.100	0.103	0.072	0.075	0.079	0.081	0.085
100								
	−0.500	0.000	0.058	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	0.067	0.056	0.059	0.060	0.061	0.061

Table 2 *continued*.

T	a_1	a_2	t_{HAC}	t^*-SO(1)	t^*-SO(2)	t^*-SO(4)	t^*-SO(8)	t^*-SO(16)
	0.500	0.000	0.071	0.060	0.062	0.062	0.064	0.064
	0.700	0.000	0.074	0.062	0.063	0.064	0.065	0.065
	0.900	0.000	0.077	0.064	0.065	0.065	0.066	0.068
	0.950	0.000	0.079	0.065	0.065	0.066	0.068	0.069
	0.990	0.000	0.080	0.059	0.063	0.065	0.068	0.070
	1.500	−0.750	0.059	0.059	0.059	0.059	0.060	0.060
	1.900	−0.950	0.081	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
200								
	−0.500	0.000	0.054	0.050	0.050	0.051	0.051	0.051
	0.000	0.000	0.057	0.053	0.054	0.054	0.054	0.054
	0.300	0.000	0.059	0.054	0.055	0.055	0.054	0.056
	0.500	0.000	0.059	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	0.064	0.057	0.057	0.057	0.057	0.058
	0.950	0.000	0.064	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	0.066	0.052	0.052	0.051	0.052	0.054
	0.800	0.100	0.065	0.058	0.057	0.057	0.058	0.058

* 50,000 replications, DGP: $y_t = x_t'\beta + u_t$; $\beta = 0$; $x_t = 0.5x_{t-1} + \eta_t$, $x_0 = 0$; $u_t = a_1u_{t-1} + a_2u_{t-2} + e_t$, $u_0 = u_{-1} = 0$; $\eta_t, e_t \sim$ iid N(0, 1), $\text{cov}(\eta_t, e_t) = 0$.

First, in all cases the size distortions of the t^* tests are less than those of the t-test. 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. 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 both t and t^*-tests increases 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. 6 and 7 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.

Figs. 6a and b 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 SO 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. 6b). This dominance is accentuated as ρ continues to increase (but in that event size distortion also increases). Compared with Fig. 5, it seems that the finite sample power comparisons mimic the asymptotic results well, with larger ρ leading to increases in power. Figs. 7a and b 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. 7b).

Next, we consider the performance of the t^*-test when ρ is data-determined. Simulations (not reported here) show that the size distortion of the t^*-test is very close to that of the conventional t-test (using data-determined consistent HAC estimates) for various parameter configurations as shown in Tables 2a and b. This is not surprising as the t^*-test with a data-driven power parameter utilizes asymptotic normality under the large ρ asymptotics. The normality approximation works well only when $\widehat{\rho}_T^*$ is large. But in finite samples, both large and small choices of $\widehat{\rho}_T^*$ can arise due to sampling variability and when $\widehat{\rho}_T^*$ is small the normal asymptotic critical value is too small due to the greater variability in the denominator of the t^*-statistic (c.f., Table 1). Simulations show that the power of the t^*-test with $\widehat{\rho}_T^*$ is also close to

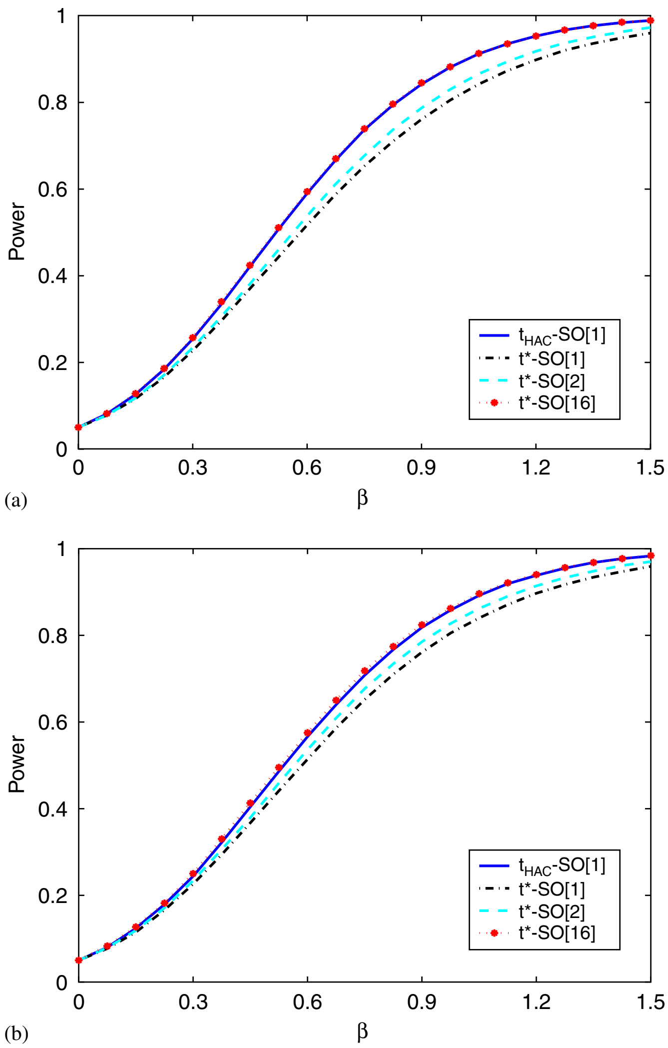

Fig. 6. Finite sample power (size-adjusted, 5% level), $T=50$, $y_t=\mu+x_t'\beta_0+u_t$; $u_t=0.85u_{t-1}+e_t$; $x_t=0.5x_{t-1}+\eta_t$ $\mathrm{H}_0:\beta_0\leqslant 0$, $\mathrm{H}_1:\beta_0>0$, (a) no prewhitening; (b) with prewhitening.

that of the conventional t-test. Therefore, the finite sample performances of the data-driven t^*-test are close to those of the t^*-test with a large fixed power parameter.

To sum up, our findings indicate that ρ is a parameter that tunes the size and power of the t^*-test. As ρ increases, the t^*-test becomes more powerful but also more size distorted. If we are very concerned with the size and are willing to give up some power to reduce size distortion, then we should choose a smaller ρ value. On the other hand, if we care more about power and are less worried about size distortion, we may choose a larger ρ value. There is always a size-power trade-off.

The magnitude of the size-power trade-off depends on the serial correlation structure of the data. In practice, we may use a hybrid procedure: (i) obtain a data-driven power parameter $\widehat{\rho}_T$ and construct the t^*-SO$(\widehat{\rho}_T)$ statistic; (ii) find the critical values using the hyperbola

$$cv(\widehat{\rho}_T)=\frac{b}{\widehat{\rho}_T-a}+c \tag{40}$$

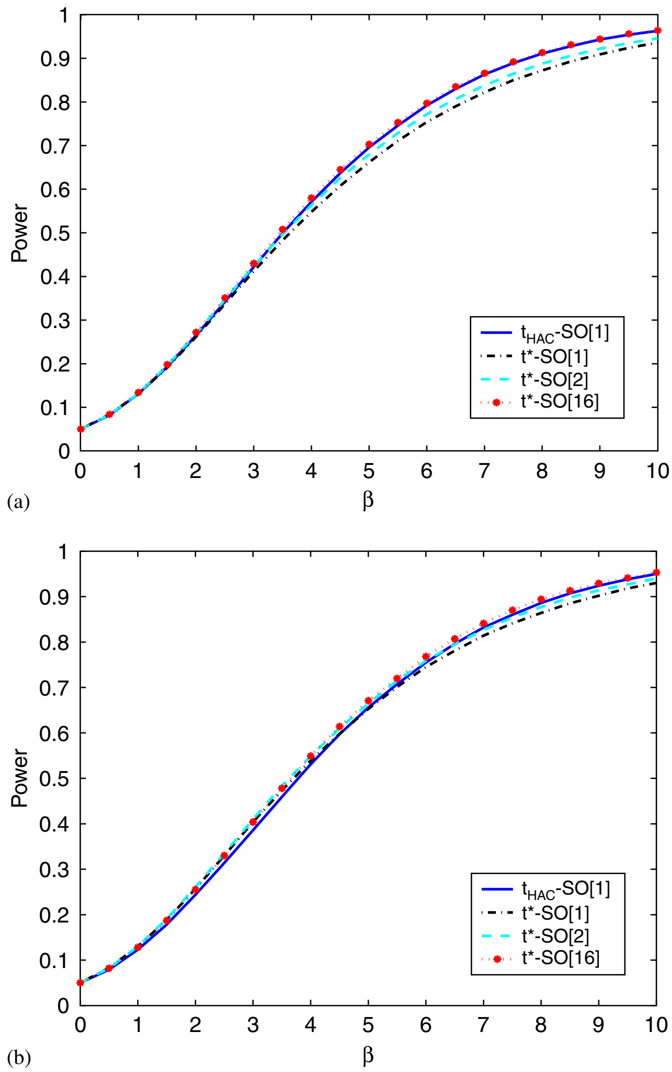

Fig. 7. Finite sample power (size-adjusted, 5% level), $T = 50$, $y_t = \mu + x_t'\beta_0 + u_t$; $u_t = 1.9u_{t-1} - 0.95u_{t-2} + e_t$; $x_t = 0.5x_{t-1} + \eta_t$ $\mathrm{H}_0 : \beta_0 \leqslant 0$, $\mathrm{H}_1 : \beta_0 > 0$, (a) no prewhitening; (b) with prewhitening.

and carry out the test. We refer to the resulting test as the t_h^*-SO$(\widehat{\rho}_T)$ test. We consider two different data-dependent power parameters. The first one is the AR(1) plug-in power parameter given in Section 4.2

$$\widehat{\rho}_1 := \widehat{\rho}_{T1} = \widehat{\rho}_T^* = [((1 - \widehat{\alpha}_c^2)^2/(4\widehat{\alpha}_c^2))^{1/3} T^{2/3}]. \tag{41}$$

The second one is a new data-dependent power parameter defined as

$$\widehat{\rho}_2 := \widehat{\rho}_{T2} = [(1 - \widehat{\alpha}_c) T^{2/3}]. \tag{42}$$

Here $\widehat{\alpha}_c = \min\{|\widehat{\alpha}|, 0.99\}$ and $\widehat{\alpha}$ is given in (28). Both $\widehat{\rho}_1$ and $\widehat{\rho}_2$ increase as $\widehat{\alpha}_c$ decreases. The difference is that $\widehat{\rho}_1$ is larger than $\widehat{\rho}_2$ for small values of $\widehat{\alpha}_c$. For example, when $\widehat{\alpha}_c = 0.3$ and $T = 50$, $\widehat{\rho}_1 = 18$ and $\widehat{\rho}_2 = 10$. Nevertheless, both selection rules deliver asymptotically valid test as the power parameters are compatible with the rate conditions under the large ρ asymptotics and the critical values approach standard normal ones as $\widehat{\rho}_1$and $\widehat{\rho}_2$ go to infinity.

Table 3
Finite sample null rejection probabilities at 5% nominal level

T	a_1	a_2	t_{HAC}	t_h^*-SO$(\widehat{\rho}_1)$	Aver$(\widehat{\rho}_1)$	t_h^*-SO$(\widehat{\rho}_2)$	Aver$(\widehat{\rho}_2)$	$t_{\hat{b}}^*$	Aver$(\hat{b})$
With no prewhitening									
50	0.00	0.00	0.073	0.064	66.44	0.063	12.43	0.063	0.042
	0.50	0.00	0.099	0.090	48.83	0.081	11.67	0.088	0.052
	0.85	0.00	0.123	0.108	29.09	0.098	10.15	0.105	0.073
	1.90	−0.95	0.135	0.116	20.14	0.104	8.80	0.114	0.094
100	0.00	0.00	0.061	0.057	125.27	0.057	20.11	0.057	0.021
	0.50	0.00	0.082	0.077	60.80	0.070	17.72	0.076	0.034
	0.85	0.00	0.098	0.089	30.06	0.082	14.51	0.087	0.052
	1.90	−0.95	0.106	0.093	22.64	0.083	12.57	0.092	0.063
200	0.00	0.00	0.056	0.054	242.59	0.054	32.44	0.054	0.011
	0.50	0.00	0.071	0.067	68.24	0.063	27.10	0.066	0.023
	0.85	0.00	0.082	0.075	38.69	0.070	21.51	0.073	0.034
	1.90	−0.95	0.086	0.078	31.55	0.070	18.67	0.077	0.041
400	0.00	0.00	0.054	0.053	466.41	0.053	52.20	0.053	0.005
	0.50	0.00	0.066	0.064	91.27	0.059	41.98	0.063	0.015
	0.85	0.00	0.071	0.067	57.01	0.062	32.82	0.066	0.022
	1.90	−0.95	0.076	0.070	47.55	0.065	28.56	0.070	0.026
With prewhitening									
50	0.00	0.00	0.076	0.073	256.85	0.067	13.87	0.071	0.021
	0.50	0.00	0.089	0.086	217.41	0.077	13.78	0.084	0.022
	0.85	0.00	0.099	0.095	156.25	0.085	13.56	0.093	0.026
	1.90	−0.95	0.107	0.101	105.74	0.089	13.18	0.100	0.031
100	0.00	0.00	0.063	0.062	569.12	0.059	21.91	0.061	0.010
	0.50	0.00	0.071	0.069	384.59	0.065	21.72	0.068	0.011
	0.85	0.00	0.076	0.074	247.04	0.069	21.34	0.074	0.014
	1.90	−0.95	0.081	0.078	185.25	0.070	20.95	0.078	0.016
200	0.00	0.00	0.057	0.056	1211.19	0.055	34.71	0.056	0.005
	0.50	0.00	0.059	0.059	656.73	0.057	34.23	0.058	0.006
	0.85	0.00	0.064	0.063	419.52	0.059	33.74	0.063	0.007
	1.90	−0.95	0.066	0.065	342.20	0.060	33.37	0.065	0.008
400	0.00	0.00	0.054	0.054	2432.07	0.053	54.87	0.053	0.003
	0.50	0.00	0.055	0.055	1169.71	0.054	54.26	0.055	0.003
	0.85	0.00	0.057	0.057	777.25	0.055	53.66	0.056	0.004
	1.90	−0.95	0.059	0.058	648.05	0.055	53.26	0.058	0.004

* 50,000 replications, DGP: $y_t = x_t'\beta + u_t$; $\beta = 0$; $x_t = 0.5x_{t-1} + \eta_t$, $x_0 = 0$; $u_t = a_1u_{t-1} + a_2u_{t-2} + e_t$, $u_0 = u_{-1} = 0$; η_t, $e_t \sim$ iid N(0, 1), $\mathrm{cov}(\eta_t, e_t) = 0$.

The hybrid procedure has the dual advantage of a choice of power parameter that is data-determined and at the same time the good finite sample size properties of the t^*-SO(ρ) test for a fixed ρ. When the underlying time series are highly persistent, the conventional HAC t-test rejects too often. In this case, $\widehat{\rho}_1$ and $\widehat{\rho}_2$ are expected to be small and the adjustments to the critical values will be large, leading to a test with better size. In contrast, when there is not much autocorrelation in the data, the conventional HAC t-test does not incur much size distortion. In this case, $\widehat{\rho}_1$ and $\widehat{\rho}_2$ are expected to be large and the adjustments to the critical values will be small. The hybrid test is thus close to the conventional test and has good size properties.

Table 3 reports the finite sample null rejection probabilities of t_h^*-tests and the average values of $\widehat{\rho}_1$ and $\widehat{\rho}_2$ over 50 000 replications. Only the economically relevant cases of positive serial correlation are given. The table shows that both the t_h^*-SO$(\widehat{\rho}_1)$ test and the t_h^*-SO$(\widehat{\rho}_2)$ test are less size distorted than the conventional HAC t-test, although the

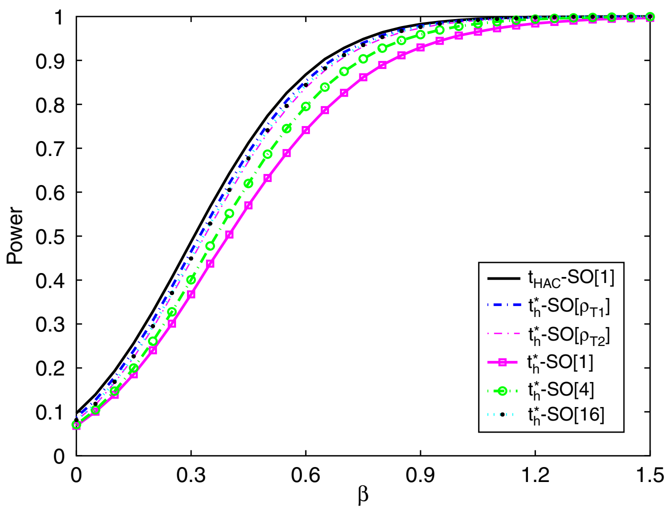

Fig. 8. Finite sample power at 5% nominal level with $T=50$, $y_t=\mu+x_t'\beta_0+u_t$; $u_t=0.85u_{t-1}+e_t$; $x_t=0.5x_{t-1}+\eta_t$ $\mathrm{H}_0:\beta_0\leqslant 0$, $\mathrm{H}_1:\beta_0>0$, with no prewhitening.

size improvement becomes less obvious when prewhitening is used or the sample size is large. Since prewhitening removes most of the autocorrelation, the truncated estimates $\widehat{\alpha}_c$ based on the prewhitened time series become closer to zero, leading to larger exponents. This is true for both the t_h^*-SO$(\widehat{\rho}_1)$ and t_h^*-SO$(\widehat{\rho}_2)$ tests. Comparing the prewhitened case with the non-prewhitened case, we find that the increases in $\hat{\rho}_{T1}$ are more dramatic than that in $\hat{\rho}_{T2}$. This is not surprising as $\hat{\rho}_{T1}$ is more sensitive to the value of $\widehat{\alpha}_c$ when $\widehat{\alpha}_c$ is small. Due to the use of very large exponents, the prewhitened t_h^*-SO$(\widehat{\rho}_1)$ test does not have much size advantage over the conventional HAC t-test. Therefore, if we are more concerned about the size of the test, then the t_h^*-SO$(\widehat{\rho}_2)$ test is preferred.

To reduce the size distortion of the conventional HAC t-test, we may use the nonstandard critical value suggested in Kiefer and Vogelsang (2005). More specifically, we plug the ratio of the data-driven bandwidth to the sample size into a polynomial to obtain the approximate nonstandard critical value and compare it with the conventional t-statistic. The resulting test is a variant of the $t_{\hat{b}}^*$-test in the earlier version of Kiefer and Vogelsang (2005) where the exact nonstandard critical value is used. The last two columns of Table 3 report the rejection probability and average value of $\hat{b}$ for the $t_{\hat{b}}^*$-test implemented using the polynomial approximation. It is clear that the empirical size of the $t_{\hat{b}}^*$-test is almost identical to that of the t_h^*-SO$(\widehat{\rho}_1)$ test. Additional simulations (not reported) show that the empirical sizes of the conventional HAC t-test and the new t^*-test are very close to each other if normal critical values are used for both tests. We may conclude that the size properties of the robust test are very similar for the Bartlett kernel and sharp kernel as long as the same type of critical value is used.

Fig. 8 depicts the finite sample power of various tests with no prewhitening when $a_1=0.85$, $a_2=0.0$ and $T=50$. The results for other cases are similar. The figure shows that the t_h^*-SO$(\widehat{\rho}_2)$ test has very competitive finite sample power but much reduced size distortion. Simulation results not reported show that, as a_1 moves away from unity, the power of the t_h^*-SO$(\widehat{\rho}_2)$ test becomes closer to that of the t_{HAC} test. Fig. 8 also shows the size distortions of different tests, which are in the descending order: t_{HAC}, t_h^*-SO$(\widehat{\rho}_1)$, t^*-SO(16), t_h^*-SO$(\widehat{\rho}_2)$, t^*-SO(4), t^*-SO(1). This pattern is found to be typical in cases where the AR coefficient is large but less than unity. Overall, the t_h^*-SO$(\widehat{\rho}_2)$ test produces favorable results for both size and power in regression testing and is recommended for practical use.

7. Conclusion

The new class of sharp origin (SO) kernels introduced in this paper permits 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 data-driven exponent selection produces favorable results for both size and power in regression testing with sample sizes that are typical in econometric applications.

The results of this paper are obtained for regression models. Following Kiefer and Vogelsang (2005) and Vogelsang (2003), we can easily extend the results to the GMM framework. Kiefer and Vogelsang (2005) set the bandwidth $M = bT$ for some fixed constant b in the context of regression testing. Their fixed-b asymptotics are similar to our fixed-ρ asymptotics. In future work, it would be helpful to compare the large sample and finite sample performances of the two different sets of tests.

The general approach given here of using SO kernels 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. Finally, by means of higher order expansions of the distributions of the HAC estimates used in regression tests, it is possible to investigate trade-offs between size distortion and power increases through the construction of these tests and, in particular, the selection of the power exponent ρ. Details are provided in some ongoing investigations by Phillips et al. (2006a,b).

Notation

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

Acknowledgement

The authors acknowledge helpful comments from two referees. Phillips thanks the NSF for research support under Grants SES 00-92509 & SES 04-142254. Jin thanks the Cowles Foundation for Fellowship support.

Appendix A.

Proof of Theorem 1. The proof follows KV (2002a) and Sun (2004) closely. Using summation by parts twice, we have

$$\begin{aligned}
\widehat{\Omega}_\rho &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\widehat{v}_t k_\rho\left(\frac{t-\tau}{T}\right)\widehat{v}_\tau' \\
&= \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\widehat{S}_t T^2 D_T\left(\frac{t-\tau}{T}\right)\widehat{S}_\tau' + \widehat{S}_T\sum_{\tau=1}^{T-1}\left(k_\rho\left(\frac{T-\tau}{T}\right)-k_\rho\left(\frac{T-\tau-1}{T}\right)\right)\widehat{S}_\tau' \\
&\quad + \sum_{t=1}^{T-1}\widehat{S}_t\left(k_\rho\left(\frac{t-T}{T}\right)-k_\rho\left(\frac{t-T+1}{T}\right)\right)\widehat{S}_T' + \widehat{S}_T\widehat{S}_T' \\
&= \frac{1}{T^2}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\widehat{S}_t T^2 D_T\left(\frac{t-\tau}{T}\right)\widehat{S}_\tau',
\end{aligned} \tag{A.1}$$

where

$$D_T\left(\frac{t-\tau}{T}\right) = 2k_\rho\left(\frac{t-\tau}{T}\right) - k_\rho\left(\frac{t-\tau-1}{T}\right) - k_\rho\left(\frac{t-\tau+1}{T}\right)$$

and we have used the identity $\widehat{S}_T = 0$. Rewrite the double summation in an integral form, we have

$$\begin{aligned}
\widehat{\Omega}_\rho &= \int_0^1\int_0^1 \widehat{S}_{[Tr]}T^2 D_T\left(\frac{[rT]-[sT]}{T}\right)\widehat{S}_{[Ts]}'\,\mathrm{d}r\,\mathrm{d}s \\
&= \int_0^1\int_0^1 1\{r\neq s\}\widehat{S}_{[Tr]}T^2 D_T\left(\frac{[rT]-[sT]}{T}\right)\widehat{S}_{[Ts]}'\,\mathrm{d}r\,\mathrm{d}s \qquad \text{(A.2)} \\
&\quad + \int_0^1 \widehat{S}_{[Tr]}' T D_T(0)\widehat{S}_{[Tr]}'\,\mathrm{d}r. \qquad \text{(A.3)}
\end{aligned}$$

For the term in (A.2), we note that when $r \neq s$,

$$\lim_{T\to\infty} T^2 D_T\left(\frac{[rT]-[sT]}{T}\right) = -k_\rho''(r-s) = -\rho(\rho-1)(1-|r-s|)^{\rho-2}. \tag{A.4}$$

Since $k(\cdot)$ is twice continuously differentiable on $[-1,0)\cup(0,1]$, the above convergence is uniform in $(r,s) \in \{(r,s)|0\leqslant r\leqslant 1, 0\leqslant s\leqslant 1, r\neq s\}$. In other words, for any given $\varepsilon > 0$, there exists a positive Δ which is independent of r and s such that

$$\left|T^2 D_T\left(\frac{[rT]-[sT]}{T}\right) + k_\rho''(r-s)\right| < \varepsilon$$

for all (r,s) in $\{(r,s)|0\leqslant r\leqslant 1, 0\leqslant s\leqslant 1, r\neq s\}$. For a proof of the uniformity, see Weinstock (1957). It now follows from $T^{-1/2}\widehat{S}_{[Tr]} \Rightarrow \Lambda V_m(r)$ and the continuous mapping theorem that

$$\begin{aligned}
&\int_0^1\int_0^1 1\{r\neq s\}\widehat{S}_{[Tr]}T^2 D_T\left(\frac{[rT]-[sT]}{T}\right)\widehat{S}_{[Ts]}'\,\mathrm{d}r\,\mathrm{d}s \\
&\quad\Rightarrow \Lambda\iint_{r\neq s} -\rho(\rho-1)(1-|r-s|)^{\rho-2}V_m(r)V_m'(s)\,\mathrm{d}r\,\mathrm{d}s\,\Lambda'.
\end{aligned} \tag{A.5}$$

For the term in (A.3), we note that

$$\lim_{T\to\infty} T D_T(0) = \lim_{T\to\infty} 2T\left(1-\left(1-\frac{1}{T}\right)^{\rho}\right) = 2\rho$$

which, combined with the continuous mapping theorem, gives

$$\int_0^1 \widehat{S}'_{[Tr]} T D_T(0) \widehat{S}'_{[Tr]}\,\mathrm{d}r \;\Rightarrow\; 2\rho\Lambda \int_0^1 V_m(r)V'_m(r)\,\mathrm{d}r\,\Lambda'. \tag{A.6}$$

Combining (A.2), (A.3), (A.5) and (A.6) yields

$$\widehat{\Omega}_\rho \Rightarrow \Lambda \iint_{r\neq s} -\rho(\rho-1)(1-|r-s|)^{\rho-2} V_m(r)V'_m(s)\,\mathrm{d}r\,\mathrm{d}s\,\Lambda' + 2\rho\Lambda\int_0^1 V_m(r)V'_m(r)\,\mathrm{d}r\,\Lambda' \tag{A.7}$$

$$= \Lambda \int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}V_m(r)\,\mathrm{d}V_m(s)\Lambda' \tag{A.8}$$

which the last equality follows from integration by parts.

Next, it is easy to see that

$$\begin{aligned}\Xi_\rho &= \int_0^1\int_0^1 k_\rho(r-s)(\mathrm{d}W_m(r)-\mathrm{d}r\,W_m(1))(\mathrm{d}W_m(s)-\mathrm{d}s\,W_m(1))' \\ &= \int_0^1\int_0^1 k^*_\rho(r,s)\,\mathrm{d}W_m(r)\,\mathrm{d}W'_m(s),\end{aligned}$$

where

$$k^*_\rho(r,s) = k_\rho(r-s) - \int_0^1 k_\rho(r-t)\,\mathrm{d}t - \int_0^1 k_\rho(\tau-s)\,\mathrm{d}\tau + \int_0^1 k_\rho(t-\tau)\,\mathrm{d}t\,\mathrm{d}\tau. \tag{A.9}$$

It follows that

$$\begin{aligned}E\Xi_\rho &= \int_0^1\int_0^1 k^*_\rho(r,r)\,\mathrm{d}r\,I_m \\ &= \left(1-\int_0^1\int_0^1 k_\rho(r-s)\,\mathrm{d}r\,\mathrm{d}s\right) I_m = \mu_\rho I_m.\end{aligned}$$

Therefore, $E\Lambda\Xi_\rho\Lambda' = \mu_\rho\Omega$, giving part (b).

For part (c), we write $E(\mathrm{vec}(\Xi_\rho)\,\mathrm{vec}(\Xi_\rho)')$ as

$$E\left(\int_0^1\int_0^1\int_0^1\int_0^1 k^*_\rho(r,s)k^*_\rho(p,q)\,\mathrm{vec}(\mathrm{d}W_m(r)\,\mathrm{d}W'_m(s))\,\mathrm{vec}(\mathrm{d}W_m(p)\,\mathrm{d}W'_m(q))'\right).$$

Some calculations show that $E(\mathrm{vec}(\mathrm{d}W_m(r)\,\mathrm{d}W'_m(s))\,\mathrm{vec}(\mathrm{d}W_m(p)\,\mathrm{d}W'_m(q)))$ is

$$\begin{cases} \mathrm{vec}(I_m)\,\mathrm{vec}(I_m)'\,\mathrm{d}r\,\mathrm{d}p & \text{if } r=s\neq p=q, \\ I_{m^2}\,\mathrm{d}r\,\mathrm{d}s & \text{if } r=p\neq s=q, \\ K_{mm}\,\mathrm{d}r\,\mathrm{d}s & \text{if } r=q\neq s=p, \\ 0 & \text{otherwise.} \end{cases} \tag{A.10}$$

Using the above result, we have

$$
\begin{aligned}
E(\operatorname{vec}(\Xi_\rho)\operatorname{vec}(\Xi_\rho)') &= \int_0^1\int_0^1 k_\rho^*(r,r)k_\rho^*(p,p)\,\mathrm{d}r\,\mathrm{d}p\operatorname{vec}(I_m)\operatorname{vec}(I_m)' \\
&\quad + \int_0^1\int_0^1 k_\rho^*(r,s)k_\rho^*(r,s)\,\mathrm{d}r\,\mathrm{d}s\,I_{m^2} + \int_0^1\int_0^1 k_\rho^*(r,s)k_\rho^*(r,s)\,\mathrm{d}r\,\mathrm{d}s\,K_{mm} \\
&= \left(\int_0^1 k_\rho^*(r,r)\,\mathrm{d}r\right)^2 \operatorname{vec}(I_m)\operatorname{vec}(I_m)' + \int_0^1\int_0^1 [k_\rho^*(r,s)]^2\,\mathrm{d}r\,\mathrm{d}s(I_{m^2}+K_{mm}).
\end{aligned}
\tag{A.11}
$$

Therefore

$$
E(\operatorname{vec}(\Xi_\rho)\operatorname{vec}(\Xi_\rho)') = \mu_\rho^2\operatorname{vec}(I_m)\operatorname{vec}(I_m)' + v_\rho(I_{m^2}+K_{mm}).
$$

Some simple manipulations show that

$$
v_\rho = \int k_\rho(r-s)k_\rho(p-q) - 2\int k_\rho(r-s)k_\rho(r-q) + \int k_\rho(r-s)^2.
$$

Hence

$$
\begin{aligned}
\operatorname{var}(\operatorname{vec}(\Lambda\Xi_\rho\Lambda')) &= E\operatorname{vec}(\Lambda\Xi_\rho\Lambda')\operatorname{vec}(\Lambda\Xi\Lambda')' - \operatorname{vec}(\Lambda E\Xi_\rho\Lambda')\operatorname{vec}(\Lambda E\Xi_\rho\Lambda')' \\
&= E(\Lambda\otimes\Lambda)\operatorname{vec}(\Xi_\rho)\operatorname{vec}(\Xi_\rho)'(\Lambda'\otimes\Lambda') - \mu_\rho^2\operatorname{vec}(\Lambda\Lambda')\operatorname{vec}(\Lambda\Lambda') \\
&= \mu_\rho^2(\Lambda\otimes\Lambda)\operatorname{vec}(I_m)\operatorname{vec}(I_m)'(\Lambda'\otimes\Lambda') \\
&\quad + v_\rho(\Lambda\otimes\Lambda)(I_{m^2}+K_{mm})(\Lambda'\otimes\Lambda') - \mu_\rho^2\operatorname{vec}(\Lambda\Lambda')\operatorname{vec}(\Lambda\Lambda') \\
&= v_\rho(\Lambda\otimes\Lambda)(I_{m^2}+K_{mm})(\Lambda'\otimes\Lambda') \\
&= v_\rho(\Lambda\Lambda')\otimes(\Lambda\Lambda') + v_\rho K_{mm}(\Lambda\Lambda')\otimes(\Lambda\Lambda') \\
&= v_\rho(I_{m^2}+K_{mm})(\Omega\otimes\Omega),
\end{aligned}
$$

giving the stated result.

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

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

uniformly over s as $T\to\infty$.

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

$$
\begin{aligned}
\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^\rho \mathrm{e}^{\mathrm{i}\lambda_s h} &= \sum_{h=0}^{L}\left(1-\frac{h}{T}\right)^\rho \mathrm{e}^{\mathrm{i}\lambda_s h} + \sum_{h=L+1}^{T-1}\left(1-\frac{h}{T}\right)^\rho \mathrm{e}^{\mathrm{i}\lambda_s h} \\
&= A_1 + A_2,\ \text{say.}
\end{aligned}
\tag{A.12}
$$

Consider each of these in turn, starting with

$$
A_1 = \sum_{h=0}^{L} \mathrm{e}^{\rho\log[1-(h/T)]}\mathrm{e}^{\mathrm{i}\lambda_s h} = \sum_{h=0}^{L}\mathrm{e}^{-h\rho/T+\mathrm{O}(T^b(h^2/T^2))}\mathrm{e}^{\mathrm{i}\lambda_s h} \tag{A.13}
$$

$$
= \sum_{h=0}^{L}\mathrm{e}^{-h\rho/T}\mathrm{e}^{\mathrm{i}\lambda_s h}[1+\mathrm{o}(1)], \tag{A.14}
$$

as $h^2/T^{2-b} = \mathrm{O}(L^2/T^{2-b}) = \mathrm{o}(1)$ uniformly for $h \leqslant L$. Obviously, the above o(1) term does not depend on s. Next consider

$$
\begin{aligned}
|A_2| &= \left| \sum_{h=L+1}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} \mathrm{e}^{\mathrm{i}\lambda_s h} \right| \\
&\leqslant \sum_{h=L+1}^{T-1} \left(1 - \frac{h}{T}\right)^{\rho} = \mathrm{O}\left(\int_{L+1}^{T-1} (1-h)^{\rho}\, \mathrm{d}h\right) = \mathrm{O}\left(T \int_{L/T}^{1-1/T} (1-y)^{\rho}\, \mathrm{d}y\right) \\
&= \mathrm{O}\left(T\left[-\frac{(1-y)^{\rho+1}}{\rho+1}\right]_{L/T}^{1-(1/T)}\right) = \mathrm{O}\left(T\left[\frac{(1-(L/T))^{\rho+1}}{\rho+1} - \frac{1}{T^{\rho+1}(\rho+1)}\right]\right) \\
&= \mathrm{O}\left(\frac{\mathrm{e}^{-(L/T^{1-b})}}{T^{b-1}}\right) = \mathrm{O}\left(\frac{\mathrm{e}^{-T^{(1/4)b}}}{T^{b-1}}\right),
\end{aligned} \tag{A.15}
$$

where the $\mathrm{O}(\cdot)$ term holds uniformly over s.

Now go back to consider A_1. First define

$$
A_{12} = \sum_{h=L+1}^{T-1} \mathrm{e}^{-h\rho/T} \mathrm{e}^{\mathrm{i}\lambda_s h}, \tag{A.16}
$$

noting that

$$
\begin{aligned}
|A_{12}| &\leqslant \sum_{h=L+1}^{T-1} \mathrm{e}^{-h\rho/T} = \mathrm{O}\left(\int_{L}^{T-1} \mathrm{e}^{-x\rho/T}\, \mathrm{d}x\right) = \mathrm{O}\left(\left[-\frac{\mathrm{e}^{-xaT^{b-1}}}{aT^{b-1}}\right]_{L}^{T-1}\right) \\
&= \mathrm{O}\left(\frac{1}{T^{b-1}}(\mathrm{e}^{-aT^{(1/4)b}} - \mathrm{e}^{-aT^{b}})\right) = \mathrm{O}\left(\frac{\mathrm{e}^{-aT^{(1/4)b}}}{T^{b-1}}\right)
\end{aligned} \tag{A.17}
$$

uniformly over s. Then, using (A.13)–(A.17) and for any $d \in (0, a)$, we can write

$$
\begin{aligned}
A_1 &= \sum_{h=0}^{L} \mathrm{e}^{-h\rho/T} \mathrm{e}^{\mathrm{i}\lambda_s h}[1 + \mathrm{o}(1)] \\
&= \sum_{h=0}^{T-1} \mathrm{e}^{-h\rho/T} \mathrm{e}^{\mathrm{i}\lambda_s h}[1 + \mathrm{o}(1)] + \mathrm{O}\left(\frac{\mathrm{e}^{-aT^{(1/4)b}}}{T^{b-1}}\right) \\
&= \sum_{h=0}^{T-1} \mathrm{e}^{h(\mathrm{i}\lambda_s - \rho/T)}[1 + \mathrm{o}(1)] + \mathrm{O}(\mathrm{e}^{-\mathrm{d}T^{(1/4)b}}) \\
&= \frac{\mathrm{e}^{T(\mathrm{i}\lambda_s - \rho/T)} - 1}{\mathrm{e}^{\mathrm{i}\lambda_s - \rho/T} - 1}[1 + \mathrm{o}(1)] + \mathrm{O}(\mathrm{e}^{-\mathrm{d}T^{(1/4)b}}) \\
&= \frac{\mathrm{e}^{-\rho} - 1}{\mathrm{e}^{\mathrm{i}\lambda_s - \rho/T} - 1}[1 + \mathrm{o}(1)] + \mathrm{O}(\mathrm{e}^{-\mathrm{d}T^{(1/4)b}})
\end{aligned} \tag{A.18}
$$

with the $\mathrm{o}(\cdot)$ and $\mathrm{O}(\cdot)$ terms holding uniformly over s.

Combining (A.12), (A.13) and (A.18), we have

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

uniformly over s, as stated.

Proof of Lemma 2. Let $\rho=aT^b$ for some $a>0$ and $0<b<1$. We start by writing

$$\begin{aligned}K_\rho(\lambda_s)&=\sum_{h=-T+1}^{T-1}\left(1-\frac{|h|}{T}\right)^{\rho}\cos(\lambda_s h)\\&=2\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{\rho}\cos(\lambda_s h)-1\\&=2\,\mathrm{Re}\left\{\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{\rho}\mathrm{e}^{\mathrm{i}\lambda_s h}\right\}-1.\end{aligned}\tag{A.19}$$

From Lemma K, we have

$$\sum_{h=0}^{T-1}\left(1-\frac{h}{T}\right)^{\rho}\mathrm{e}^{\mathrm{i}\lambda_s h}=\frac{1}{1-\mathrm{e}^{\mathrm{i}\lambda_s-\rho/T}}[1+\mathrm{o}(1)],\tag{A.20}$$

uniformly over s. Direct evaluation gives

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

and so

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

uniformly over x.

It follows that

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

uniformly over x. Combining this result with (A.19) and (A.20) gives

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

uniformly over s, as stated.

Proof of Theorem 3. We prove the results for the scalar v_t case, the vector case follows without further complication.

Part (a): From (19)

$$\widetilde{f}_{vv}(0)=\frac{1}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)I_{vv}(\lambda_s). \tag{A.21}$$

To find the asymptotic variance of $\widetilde{f}_{vv}(0)$, we can work from the following standard formula (e.g., Priestley, 1981, Eq. 6.2.110 on p. 455) for the variance of a weighted periodogram estimate such as (A.21),[3] viz.,

$$\mathrm{Var}\{\widetilde{f}_{vv}(0)\}=2f_{vv}(0)^2\frac{1}{T}\sum_{h=-T+1}^{T-1}k_\rho\left(\frac{h}{T}\right)^2[1+\mathrm{o}(1)], \tag{A.22}$$

which follows directly from the covariance properties of the periodogram of a linear process (e.g., Priestley, 1981, p. 426). To evaluate (A.22), we develop an asymptotic approximation of

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

---

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

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

is an alternate form of (A.21).

This can be accomplished by Euler summation, viz.,

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

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

$$\sum_{h=0}^{T-1} \left(1 - \frac{h}{T}\right)^{2\rho} = \int_0^{T-1} \left(1 - \frac{x}{T}\right)^{2\rho} \mathrm{d}x + \frac{1}{2}\left\{1 + \left(1 - \frac{T-1}{T}\right)^{2\rho}\right\} + \left(-\frac{2\rho}{T}\right) \int_0^{T-1} \left(x - [x] - \frac{1}{2}\right)\left(1 - \frac{x}{T}\right)^{2\rho-1} \mathrm{d}x. \quad \text{(A.23)}$$

Note that

$$1 + \left(1 - \frac{T-1}{T}\right)^{2\rho} = \mathrm{O}(1), \quad \text{(A.24)}$$

and

$$\begin{aligned} &\left|\left(-\frac{2\rho}{T}\right) \int_0^{T-1} \left(x - [x] - \frac{1}{2}\right)\left(1 - \frac{x}{T}\right)^{2\rho-1} \mathrm{d}x\right| \\ &\quad \leqslant 2\,\frac{\rho}{T} \int_0^{T-1} \left(1 - \frac{x}{T}\right)^{2\rho-1} \mathrm{d}x = 2\rho \int_0^{1-1/T} (1-y)^{2\rho-1}\,\mathrm{d}y \\ &\quad = [(1-y)^{2\rho}]_0^{1-1/T} = \mathrm{O}(1), \end{aligned} \quad \text{(A.25)}$$

whereas

$$\begin{aligned} &\int_0^{T-1} \left(1 - \frac{x}{T}\right)^{2\rho} \mathrm{d}x \\ &\quad = T \int_0^{1-1/T} (1-y)^{2\rho}\,\mathrm{d}y = \frac{T}{2\rho+1}[-(1-y)^{2\rho+1}]_0^{1-1/T} \\ &\quad = \frac{T}{2\rho+1}\left[1 - \left(1 - \frac{T-1}{T}\right)^{2\rho+1}\right] = \frac{T}{2\rho+1} + \mathrm{O}\left(\frac{1}{\rho T^{2\rho}}\right). \end{aligned} \quad \text{(A.26)}$$

It follows from (A.23)–(A.26) that

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

so that

$$\begin{aligned} &\frac{1}{T} \sum_{h=-T+1}^{T-1} \left(1 - \frac{|h|}{T}\right)^{2\rho} = \frac{2}{T} \sum_{h=0}^{T-1} \left(1 - \frac{h}{T}\right)^{2\rho} - \frac{1}{T} \\ &\quad = \frac{2}{T} \int_0^{T-1} \left(1 - \frac{s}{T}\right)^{2\rho} \mathrm{d}s + \mathrm{O}\left(\frac{1}{T}\right) \\ &\quad = 2 \int_0^{1-1/T} (1-y)^{2\rho}\,\mathrm{d}y + \mathrm{O}\left(\frac{1}{T}\right) \\ &\quad = \frac{2}{2\rho+1} + \mathrm{O}\left(\frac{1}{T}\right) = \frac{1}{\rho}[1 + \mathrm{o}(1)], \end{aligned} \quad \text{(A.27)}$$

giving

$$\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho^2\left(\frac{h}{T}\right) = \frac{1}{\rho}[1+\mathrm{o}(1)]. \tag{A.28}$$

Using (A.27) in (A.22) we have

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

and so

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

which gives

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

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

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

we can write the scaled estimation error as

$$\begin{aligned}
&\sqrt{\rho}\{\widetilde{f}_{vv}(0) - f_{vv}(0)\}\\
&\quad = \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[I_{vv}(\lambda_s) - f_{vv}(0)]\\
&\quad = \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[I_{vv}(\lambda_s) - f_{vv}(\lambda_s)] + \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[f_{vv}(\lambda_s) - f_{vv}(0)].
\end{aligned} \tag{A.29}$$

Using Lemma 2, we have

$$K_\rho(\lambda_s) = \mathrm{O}\left(\frac{2\rho T}{(2\pi s)^2+\rho^2}[1+\mathrm{o}(1)]\right) \quad \text{uniformly over } s = 0, 1, \ldots, [T/2]. \tag{A.30}$$

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

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

Hence, the second term of (A.29) can be bounded as follows:

$$
\begin{aligned}
&\frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[f_{vv}(\lambda_s)-f_{vv}(0)]\\
&=\frac{2\sqrt{\rho}}{T}\sum_{s=0}^{[T/2]} K_\rho(\lambda_s)[f_{vv}(\lambda_s)-f_{vv}(0)] = \mathrm{O}\left(\frac{\sqrt{\rho}}{T}\sum_{s=0}^{[T/2]} K_\rho(\lambda_s)\lambda_s\right)\\
&=\mathrm{O}\left(\frac{\rho^{3/2}}{T}\sum_{s=0}^{[T/2]}\frac{2T\lambda_s}{\rho^2+(2\pi s)^2}\right)=\mathrm{O}\left(\frac{\rho^{3/2}}{T}\int_0^{T/2}\frac{x}{\rho^2+(2\pi x)^2}\,\mathrm{d}x\right)\\
&=\mathrm{O}\left(\frac{\rho^{3/2}}{T}\left[\frac{\log\{\rho^2+(2\pi x)^2\}}{2(2\pi)^2}\right]_0^{T/2}\right)\\
&=\mathrm{O}\left(\frac{\rho^{3/2}\log T}{T}\right)=\mathrm{o}(1),
\end{aligned}
\tag{A.31}
$$

since $\rho = aT^b$ with $b < \frac{2}{3}$. Then, by (A.29) and (A.31), we have

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

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)

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

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

$$
v_t = C(L)\varepsilon_t = C(\mathrm{e}^{\mathrm{i}\lambda})\varepsilon_t + \mathrm{e}^{-\mathrm{i}\lambda}\widetilde{\varepsilon}_{\lambda t-1} - \widetilde{\varepsilon}_{\lambda t},
\tag{A.32}
$$

where

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

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

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

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

$$\begin{aligned}
&\sqrt{\rho}\{\widetilde{f}_{vv}(0) - f_{vv}(0)\} \\
&= \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)(I_{vv}(\lambda_s) - f_{vv}(\lambda_s)) + \mathrm{o_p}(1) \qquad \text{(A.33)} \\
&= \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)(w(\lambda_s)w(\lambda_s)^* - f_{vv}(\lambda_s)) + \mathrm{o_p}(1) \\
&= \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\{[C(\mathrm{e}^{\mathrm{i}\lambda_s})w_\varepsilon(\lambda_s) + \mathrm{O_p}(T^{-1/2})] \\
&\quad \times [C(\mathrm{e}^{\mathrm{i}\lambda_s})w_\varepsilon(\lambda_s) + \mathrm{O_p}(T^{-1/2})]^* - f_{vv}(\lambda_s)\} + \mathrm{o_p}(1) \\
&= \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[C^2(1)\left(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2\right)\right] + \mathrm{O_p}\left(\frac{\sqrt{\rho}}{T}T\frac{1}{T^{1/2}}\right) + \mathrm{o_p}(1) \\
&= \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[C^2(1)\left(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2\right)\right] + \mathrm{o_p}(1), \qquad \text{(A.34)}
\end{aligned}$$

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

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

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

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

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

By the Fourier inversion formula, we have

$$c_j = \frac{C^2(1)}{2\pi} \frac{\sqrt{\rho}}{T} k_\rho \left(\frac{j}{T} \right). \tag{A.35}$$

Hence

$$\sum_{j=1}^{T} c_j^2 = \mathrm{O}\left(\frac{\rho}{T^2} \sum_{j=1}^{T} k_\rho^2 \left(\frac{j}{T} \right) \right) = \mathrm{O}\left(\frac{\rho}{T} \frac{1}{2\rho+1} \right) = \mathrm{O}\left(\frac{1}{T} \right). \tag{A.36}$$

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

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

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

$$\sum_{t=1}^{T} E(m_t^2 | \mathscr{F}_{t-1}) - \frac{\sigma^4 C^4(1)}{8\pi^2} \rightarrow_p 0, \tag{A.37}$$

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

$$\sum_{t=1}^{T} E(m_t^4) \rightarrow_p 0. \tag{A.38}$$

We now proceed to establish (A.37) and (A.38). The left-hand side of (A.37) is

$$\left(\sigma^2 \sum_{t=2}^{T} \sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2 - \frac{\sigma^4 C^4(1)}{8\pi^2} \right) + \sigma^2 \sum_{t=2}^{T} \sum_{r \neq j} \varepsilon_r \varepsilon_j c_{t-r} c_{t-j} := I_1 + I_2. \tag{A.39}$$

The first term, I_1, is

$$\sigma^2 \left(\sum_{j=1}^{T-1} (\varepsilon_j^2 - \sigma^2) \sum_{s=1}^{T-j} c_s^2 \right) + \left(\sigma^4 \sum_{t=1}^{T-1} \sum_{j=1}^{T-t} c_j^2 - \frac{\sigma^4 C^4(1)}{8\pi^2} \right) := I_{11} + I_{12}. \tag{A.40}$$

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

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

using (A.36).

Next, consider the second term of (A.40). We have

$$\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} c_s^2 = \frac{C^4(1)}{4\pi^2}\,\frac{\rho}{T^2}\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} k_\rho^2\left(\frac{s}{T}\right)$$

$$= (1+\mathrm{o}(1))\frac{C^4(1)}{4\pi^2}\,\frac{\rho}{T}\sum_{j=1}^{T-1}\int_{j/T}^{1-1/T} y^{2\rho}\,\mathrm{d}y$$

$$= (1+\mathrm{o}(1))\frac{C^4(1)}{4\pi^2}\,\frac{\rho}{T}\sum_{j=1}^{T-1}\left(\frac{1}{2\rho+1}\right)\left[\left(1-\frac{1}{T}\right)^{2\rho+1} - \left(\frac{j}{T}\right)^{2\rho+1}\right]$$

$$= \frac{C^4(1)}{4\pi^2}\,\frac{\rho}{T}\left(\frac{1}{2\rho+1}\right)T\left(1-\frac{1}{T}\right)^{2\rho+1}(1+\mathrm{o}(1))$$

$$- \frac{C^4(1)}{4\pi^2}\,\frac{\rho}{T}\left(\frac{1}{2\rho+1}\right)\frac{T}{2\rho+2}(1+\mathrm{o}(1))$$

$$= \frac{C^4(1)}{8\pi^2} + \mathrm{o}(1).$$

We have therefore shown that

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

So the first term of (A.39) is $\mathrm{o_p}(1)$.

Now consider the second term, I_2, of (A.39). I_2 has mean zero and variance

$$\mathrm{O}\left(2\sum_{p,q=2}^{T}\sum_{r\neq j}^{\min(p-1,q-1)}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right)$$

$$= \mathrm{O}\left(2\sum_{p=2}^{T}\sum_{r\neq j}^{p-1}c_{p-r}^2 c_{p-j}^2 + 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right). \tag{A.41}$$

In view of (A.36), we have

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

For the second component in (A.41), we have, using (A.36) and the Cauchy inequality,

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

Hence, $I_2 \to_p 0$ and we have therefore established condition (A.37).

It remains to verify (A.38). Let A be some positive constant, then the left-hand side of (A.38) is

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

using (A.36), which verifies (A.38) and the CLT.

With this construction we therefore have

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

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

$$\sqrt{\rho}\{\widetilde{f}_{vv}(0) - f_{vv}(0)\} = \frac{\sqrt{\rho}}{T}\sum_{s=0}^{T-1} K(\lambda_s)(I_{vv}(\lambda_s) - f_{vv}(\lambda_s)) + \mathrm{o_p}(1) \to_d \mathrm{N}(0, 2f_{vv}^2(0)),$$

from which we deduce that

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

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

Part (c): By definition,

$$\begin{aligned}
&E(\widehat{f}_{vv}(0) - f_{vv}(0))\\
&= \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho\left(\frac{h}{T}\right) EC_h - \frac{1}{2\pi}\sum_{-\infty}^{\infty}\Gamma(h)\\
&= \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\left(1 - \frac{|h|}{T}\right)^{\rho+1}\Gamma(h) - \frac{1}{2\pi}\sum_{-\infty}^{\infty}\Gamma(h)\\
&= \frac{1}{2\pi}\sum_{h=-T+1}^{T-1}\left[\left(1 - \frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h) - \frac{1}{2\pi}\sum_{|h|\geqslant T}\Gamma(h)\\
&= \frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[\left(1 - \frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h)\\
&\quad + \frac{1}{2\pi}\sum_{T-1\geqslant|h|>T/(\rho\log T)}\left[\left(1 - \frac{|h|}{T}\right)^{\rho+1} - 1\right]\Gamma(h) + \frac{1}{2\pi}\sum_{|h|\geqslant T}\Gamma(h),
\end{aligned} \tag{A.42}$$

where the second equality follows from the fact $EC_h = (1 - (|h|/T))\Gamma(h)$. Now

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

by virtue of A2, and

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

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

The first term of (A.42) can be written as

$$\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[\left(1-\frac{|h|}{T}\right)^{\rho+1}-1\right]\Gamma(h)$$

$$=\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\left[1-\frac{(\rho+1)|h|}{T}+\mathrm{O}\left(\frac{(\rho+1)^2|h|^2}{T^2}\right)-1\right]\Gamma(h)$$

$$=-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+\mathrm{O}\left(\frac{(\rho+1)^2}{T^2}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}h^2\Gamma(h)\right)$$

$$=-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+\mathrm{o}\left(\frac{(\rho+1)^2}{T^2}\frac{T}{\rho\log T}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}|h||\Gamma(h)|\right)$$

$$=-\frac{1}{2\pi}\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}\frac{(\rho+1)|h|}{T}\Gamma(h)+\mathrm{o}\left(\frac{\rho}{T\log T}\right).$$

Therefore

$$\lim_{T\to\infty}\frac{T}{\rho}(E\widetilde{\Omega}_\rho-\Omega)=\lim_{T\to\infty}\left(-\sum_{h=-T/(\rho\log T)}^{T/(\rho\log T)}|h|\Gamma(h)\right)$$

$$=-2\pi f^{(1)}=-\Omega^{(1)},$$

as required.

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

$$\frac{T}{\rho}\sim\frac{T}{\vartheta^{1/3}T^{2/3}}=\vartheta^{-(1/3)}T^{(1/3)}\sim\vartheta^{-(1/3)}\sqrt{\frac{\rho}{\vartheta^{\frac{1}{3}}}}=\frac{\sqrt{\rho}}{\sqrt{\vartheta}}. \tag{A.43}$$

It follows from (A.43) that

$$\mathrm{MSE}(\rho,\widetilde{\Omega}_\rho,W)$$

$$=\rho E\{\mathrm{vec}(\widetilde{\Omega}_\rho-\Omega)'W\,\mathrm{vec}(\widetilde{\Omega}_\rho-\Omega)\}$$

$$=\rho E\{\mathrm{vec}(\widetilde{\Omega}_\rho-E\widetilde{\Omega}_\rho+E\widetilde{\Omega}_\rho-\Omega)'W\,\mathrm{vec}(\widetilde{\Omega}_\rho-E\widetilde{\Omega}_\rho+E\widetilde{\Omega}_\rho-\Omega)\}$$

$$=\vartheta\left(\frac{T}{\rho}\right)^2E\{\mathrm{vec}(E\widetilde{\Omega}_\rho-\Omega)'W\,\mathrm{vec}(E\widetilde{\Omega}_\rho-\Omega)\}[1+\mathrm{o}(1)]$$

$$+\rho\,\mathrm{tr}\{WE[\mathrm{vec}(\widetilde{\Omega}_\rho-E\widetilde{\Omega}_\rho)]\mathrm{vec}(\widetilde{\Omega}_\rho-E\widetilde{\Omega}_\rho)'\}.$$

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

$$\lim_{T\to\infty}\mathrm{MSE}(\rho,\widetilde{\Omega}_\rho,W)$$

$$=\vartheta\,\mathrm{vec}(\Omega^{(1)})'W\,\mathrm{vec}(\Omega^{(1)})+\mathrm{tr}\{W(I+K_{mm})(\Omega\otimes\Omega)\}.$$

The corresponding result for the spectral density estimate is

$$\lim_{T\to\infty} \mathrm{MSE}(\rho, \widetilde{f}_{vv}(0), W)$$
$$= \vartheta \operatorname{vec}(f^{(1)})' W \operatorname{vec}(f^{(1)}) + \operatorname{tr}\{W(I + K_{mm}) f \otimes f)\}. \qquad \square$$

Proof of Theorem 4. 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 $\widehat{\Omega}_\rho$ and for $\widehat{\Omega}_\rho$ to have the same asymptotic distribution as $\widetilde{\Omega}_\rho$. 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 $\kappa_{a_1\ldots a_8}(0, j_1, j_2, \ldots, j_7)$ denote the cumulant of $(v_{a_1 0}, \ldots, v_{a_8 j_7})$, where $a_1, \ldots, a_8$ are positive integers less than $p+1$ and $j_1, \ldots, j_7$ are integers.

Assumption B. (1) Assumption A2 holds with v_t replaced by $(v_t', \operatorname{vec}((\partial/\partial\beta')v_t(\beta) - E(\partial/\partial\beta')v_t(\beta))')'$.

(2) $\{v_t\}$ is eighth order stationary with summable cumulant function $\kappa_{a_1\cdots a_8}(0, j_1, j_2, \ldots, j_7)$, i.e., $\sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_7=-\infty}^{\infty} |\kappa_{a_1\cdots a_8}(0, j_1, j_2, \ldots, j_7)| < \infty$.

(3) $W_T \to_p W$.

Proof of Part (a). A Taylor expansion gives

$$\sqrt{\rho}(\widehat{\Omega}_\rho - \widetilde{\Omega}_\rho) = \left[\sqrt{\rho/T}\frac{\partial}{\partial\beta'}\widetilde{\Omega}_\rho(\beta)\right]\sqrt{T}(\widehat{\beta}-\beta)$$
$$+ \frac{1}{2}\sqrt{T}(\widehat{\beta}-\beta)'\left[\sqrt{\frac{\rho}{T^2}}\frac{\partial^2}{\partial\beta\,\partial\beta'}\widetilde{\Omega}_\rho(\widetilde{\beta})\right]\sqrt{T}(\widehat{\beta}-\beta),$$

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

$$\sqrt{\frac{\rho}{T^2}}\frac{\partial^2}{\partial\beta\partial\beta'}\widetilde{\Omega}_\rho(\widetilde{\beta})$$
$$\leqslant \sqrt{\frac{\rho}{T^2}}\sum_{-T+1}^{T-1}\left|\left(1-\frac{|h|}{T}\right)^\rho\right|\frac{1}{T}\sum_{t=|h|+1}^{T}\sup_{\beta\in\varkappa}\left\|\frac{\partial^2}{\partial\beta\,\partial\beta'}v_t(\beta)v_{t-|h|}(\beta)\right\|$$
$$= \sqrt{\rho}\left(\frac{1}{T}\sum_{-T+1}^{T-1}\left|\left(1-\frac{|h|}{T}\right)^\rho\right|\right)\mathrm{O_p}(1)$$
$$= \mathrm{o_p}(1),$$

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

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

Proof of Theorems 5 and 6. These results follow directly from standard weak convergence arguments. $\square$

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–854.

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

de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. Econometrica 68, 407–424.

den Haan, W.J., Levin, A., 1997. A practitioners guide to robust covariance matrix estimation. In: Maddala, G., Rao, C. (Eds.), Handbook of Statistics: Robust Inference, vol. 15. Elsevier, New York.

Hall, P.G., Hallin, M., Roussas, G.G., 2003. Madan Lal Puri, Selected Collected Works, Volume 1: Nonparametric Methods in Statistics and Related Topics. VSP International Science Publishers.

Hannan, E.J., 1970. Multiple Time Series. Wiley, New York.

Hansen, B.E., 1992. Consistent covariance matrix estimation for dependent heterogenous processes. Econometrica 60, 967–972.

Hashimzade, N., Vogelsang, T.J., 2006. Fixed-b asymptotic approximation for the sampling behavior of nonparametric spectral density estimators. Working paper, Department of Economics, Cornell University.

Jansson, M., 2004. On the error of rejection probability in simple autocorrelation robust rests. Econometrica 72, 937–946.

Kiefer, N.M., Vogelsang, T.J., 2002a. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory 18, 1350–1366.

Kiefer, N.M., Vogelsang, T.J., 2002b. Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica 70, 2093–2095.

Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasiticy-autocorrelation robust tests. Econometric Theory 21, 1130–1164.

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

Lee, C., Phillips, P.C.B., 1994. An ARMA prewhitened long-run variance estimator. Working Paper, Yale University.

Linton, O.B., 1995. Second order approximation in the partially linear regression model. Econometrica 63, 1079–1112.

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

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance estimation. Rev. Econom. Stud. 61, 631–654.

Parzen, E., 1957. On the consistent estimates of the spectrum of a stationary time series. Annals Math. Statist. 28, 329–348.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Ann. Statist. 20, 971–1001.

Phillips, P.C.B., Sun, Y., Jin, S., 2006a. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. International Economic Review 47(3), forthcoming.

Phillips, P.C.B., Sun, Y., Jin, S., 2006b. Improved HAR inference using power kernels without truncation. Mimeographed, Yale University.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press, New York.

Robinson, P., 1995. Gaussian semiparametric estimation of long range dependence. Ann. Statist., 1630–1661.

Sun, Y., 2004. A convergent t-statistic in spurious regressions. Econometric Theory 20, 943–962.

Velasco, C., Robinson, P., 2001. Edgeworth expansions for spectral density estimates and studentized sample mean. Econometric Theory 17, 497–539.

Vogelsang, T.J., 2003. Testing in GMM models without truncation. In: Fomby, T.B., Hill, R.C. (Eds.), Advances in Econometrics, vol. 17, Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. Elsevier Science, Amsterdam, pp. 199–233 (Chapter 10).

Weinstock, R., 1957. On continuous differentiability. Amer. Math. Month. 64, 492.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

Xiao, Z., Phillips, P.C.B., 1998. Higher order approximations for frequency domain time series regression. J. Econometrics 86, 297–336.

Xiao, Z., Phillips, P.C.B., 2002. Higher order approximations for Wald statistics in time series regressions with integrated processes. J. Econometrics 108, 157–198.