LONG RUN VARIANCE ESTIMATION USING STEEP ORIGIN KERNELS WITHOUT TRUNCATION

By

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

September 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1437

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

Long Run Variance Estimation using Steep Origin Kernels without Truncation*

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

Yixiao Sun
Department of Economics
University of California, San Diego

and

Sainan Jin
Department of Economics
Yale University

June 8, 2003

*The authors thank Graham Elliot, Jim Hamilton, and Hal White for helpful comments. Phillips thanks the NSF for research support under Grant No. SES 00-92509.

ABSTRACT

A new class of kernel estimates is proposed for long run variance (LRV) and heteroskedastic autocorrelation consistent (HAC) estimation. The kernels are called steep origin kernels and are related to a class of sharp origin kernels explored by the authors (2003) in other work. They are constructed by exponentiating a mother kernel (a conventional lag kernel that is smooth at the origin) and they can be used without truncation or bandwidth parameters. When the exponent is passed to infinity with the sample size, these kernels produce consistent LRV/HAC estimates. The new estimates are shown to have limit normal distributions, and formulae for the asymptotic bias and variance are derived. With steep origin kernel estimation, bandwidth selection is replaced by exponent selection and data-based selection is possible. Rules for exponent selection based on minimum mean squared error (MSE) criteria are developed. Optimal rates for steep origin kernels that are based on exponentiating quadratic kernels are shown to be faster than those based on exponentiating the Bartlett kernel, which produces the sharp origin kernel. It is further shown that, unlike conventional kernel estimation where an optimal choice of kernel is possible in terms of MSE criteria (Priestley, 1962; Andrews, 1991), steep origin kernels are asymptotically MSE equivalent, so that choice of mother kernel does not matter asymptotically. The approach is extended to spectral estimation at frequencies $\omega \neq 0$. Some simulation evidence is reported detailing the finite sample performance of steep kernel methods in LRV/HAC estimation and robust regression testing in comparison with sharp kernel and conventional (truncated) kernel methods.

Key words and Phrases: Exponentiated kernel, lag kernel, long run variance, optimal exponent, spectral window, spectrum.

JEL Classification: C22

1 Introduction and Motivation

Following the vast time series literature on spectral estimation, kernel estimates were proposed and analyzed in the econometric literature for long run variance (LRV) and heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation. These procedures have been found to be particularly useful in the construction of robust regression tests, unit root tests, and cointegration estimators. There is now a wide literature discussing these procedures, their various refinements and data-based empirical implementations in econometrics (see den Haan and Levin, 1997, for a recent review).

It is known that in cases like robust hypothesis testing, consistent HAC estimates are not needed in order to produce asymptotically valid tests. In recent work on this issue, Kiefer and Vogelsang (2002a, 2002b) have proposed the use of inconsistent HAC estimates based on conventional kernels but with the bandwidth parameter (M) set equal to the sample size (T). Kiefer and Vogelsang show that such estimates lead to asymptotically valid tests that can have better finite sample size properties than tests based on consistent HAC estimates. Their power analysis and simulations reveal that the Bartlett kernel among the common choices of kernel produces the highest power function in regression testing when $M = T$, although power is noticeably less than that which can be attained using conventional procedures involving consistent HAC estimators.

In other work, the authors (2003) recently showed that sharp origin kernels, constructed by exponentiating the Bartlett kernel, can deliver consistent LRV/HAC estimates while eliminating truncation and retaining some of the size advantages noticed by Kiefer and Vogelsang. The present paper pursues this approach by considering the use of mother kernels other than the Bartlett kernel in the construction of LRV estimates. In particular, we consider as mother kernels a class of quadratic kernels that includes many of the popular kernels that are used in practical work, such as the Parzen and Tukey Hanning kernels. Exponentiating these kernels produces a class of kernels that have steep but smooth behavior at the origin, in contrast to the Bartlett kernel which produces a sharp, non differentiable kernel at the origin. Earlier work on quadratic kernels with the use of bandwidths $M < T$ showed that there are certain advantages, including improved rates of convergence, arising from the smooth behavior of such kernels at the origin. The present paper is motivated to explore whether similar advantages may arise in the use of exponentiated kernels of this type when $M = T$ and the exponent is passed to infinity with the sample size.

Accordingly, the paper develops an asymptotic theory for this new class of steep origin kernel estimates, providing a central limit theory, and giving formulae for asymptotic bias, variance and mean squared error (MSE) of the estimates. It is shown that data-determined selection of the exponent parameter is possible and rules are provided for optimal choice of the exponent based on a minimum MSE error criteria. Optimal rates of convergence for steep origin kernel estimates constructed from quadratic mother kernels are shown to be faster than those based on exponentiating the Bartlett kernel. This steep origin approach to LRV estimation applies more generally to cases of spectral density and probability density estimation and the paper illustrates such extensions by considering spectral estimation at frequencies $\omega \neq 0$. Simulations reveal that steep kernel methods generally outperform sharp kernel and conventional quadratic kernel estimators in both LRV estimation and robust regression testing, although sharp kernel estimators tend to do better in extreme cases like white noise or nearly nonstationary spectra.

The paper is organized as follows. Section 2 describes a class of steep origin kernels, characterizes their asymptotic form, develops a central limit theory, provides bias, variance and MSE formulae and discusses data-determined optimal exponent selection. Section 3 provides a similar analysis for the corresponding spectral density estimates at non-zero frequencies. Section 4 reports some simulation evidence on the finite sample performance of these estimates and associated tests. Conclusions are given in Section 5. Proofs and other technical material are included in the Appendix (Section 6) and a glossary of notation is given in Section 7.

2 LRV Estimation with Steep Origin Kernels

We construct a class of steep origin kernels for use in LRV estimation based on quadratic mother kernels, study the asymptotic form of the associated windows, and develop an asymptotic theory for the estimates.

2.1 Exponentiated Quadratic Kernels

Consider an m-vector stationary process $\{X_t\}_{t=1}^T$ with non-singular spectral density matrix $f_{XX}(\lambda)$. The long run variance matrix of X_t is defined as

$$\Omega = \gamma_0 + \sum_{h=1}^{\infty} \left(\gamma_h + \gamma_h'\right) = 2\pi f_{XX}(0) \tag{1}$$

where $\gamma_h = E(X_t X_{t-h}')$. To estimate Ω, we consider the following lag kernel estimator of $f_{XX}(0)$

$$\widehat{f}_{XX}(0) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widehat{\gamma}_h, \tag{2}$$

$$\widehat{\gamma}_h = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h}X_t' & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-h+1}^{T} X_{t+h}X_t' & \text{for } j < 0 \end{cases}, \tag{3}$$

where $k_\rho(x)$ is equal to $k(x)$ raised to some positive integer[1] power ρ, i.e.

$$k_\rho(x) = k^\rho(x). \tag{4}$$

When $k(x)$ is the Bartlett kernel, $\widehat{f}_{XX}(0)$ is the sharp origin estimator considered by Phillips, Sun and Jin (PSJ hereafter, 2003).

[1]It is often convenient to treat the exponent ρ as an integer. But when $k(x)$ is nonnegative any positive real value of ρ may be considered.

Exponentiating the kernel $k(x)$ induces a class of kernels $\{k_\rho(x)\}_{\rho\in\mathbb{Z}^+}$. The kernel $k(x)$ itself belongs to this class and is called the mother kernel of the class. This paper will consider mother kernels that have quadratic behavior at the origin and satisfy the following assumption.

Assumption 1: *(a)* $k(x) : [-1,1] \to [0,1]$ *is even, nonnegative and differentiable with* $k(0) = 1$ *and* $k(1) = 0$.

(b) For any $\eta > 0$, *there exists* $\xi < 1$ *such that* $k(x) \leq \xi$ *for* $|x| \geq \eta$.

(c) $k(x)$ *has a valid quadratic expansion in a neighborhood of zero:*

$$k(x) = 1 - gx^2 + o\left(x^2\right), \text{ as } x \to 0 \text{ for some } g > 0. \tag{5}$$

Under Assumption 1(c), the kernel $k(x)$ has Parzen (1957) exponent $q = 2$ such that

$$\lim_{x\to 0} \frac{1-k(x)}{|x|^q} = g.$$

The Parzen exponent characterizes the smoothness of $k(x)$ at the origin. Assumptions 1(a) and 1(c) imply that $k'(0) = 0$ and $k''(0) = -2g$. Thus, the kernels satisfying Assumption 1 have quadratic behavior around the origin.

Examples of commonly used kernels satisfying Assumption 1 include the Parzen and Tukey-Hanning kernels:

Parzen $$k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1-|x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Tukey-Hanning $$k_{TH}(x) = \begin{cases} (1 + \cos \pi x)/2 & \text{for } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

For the Parzen kernel, $g = 6$. For the Tukey-Hanning kernel, $g = \pi^2/4$. The Parzen kernel has been used in the literature concerning long run variance estimation and the Tukey-Hanning kernel is popular in the spectral density estimation literature. The quadratic spectral (QS) kernel is also quadratic at the origin and has some optimality properties in conventional LRV/HAC estimation. However, the QS kernel is not used in the present paper because it is unsuitable as a mother kernel, being nonzero for $|x| > 1$ and not always nonnegative.

Since $k_\rho(x) = k^\rho(x)$, the kernel $k_\rho(x)$ satisfies Assumption 1 if $k(x)$ does. Obviously, $k_\rho(x)$ has series expansion

$$k_\rho(x) = 1 - \rho g x^2 + o\left(x^2\right), \text{ as } x \to 0$$

and

$$\lim_{x\to 0} \frac{1-k_\rho(x)}{x^2} = \rho g. \tag{6}$$

Thus, the curvature of $k_\rho(x)$ at the origin increases as ρ increases. In other words, as ρ increases, $k_\rho(x)$ becomes successively more concentrated at the origin and its shape steeper. $k_\rho(x)$ is therefore called a steep origin kernel. Figs. 1 and 2 graph $k_\rho(x)$ for $\rho = 1, 5, 10, 20$ illustrating these effects.

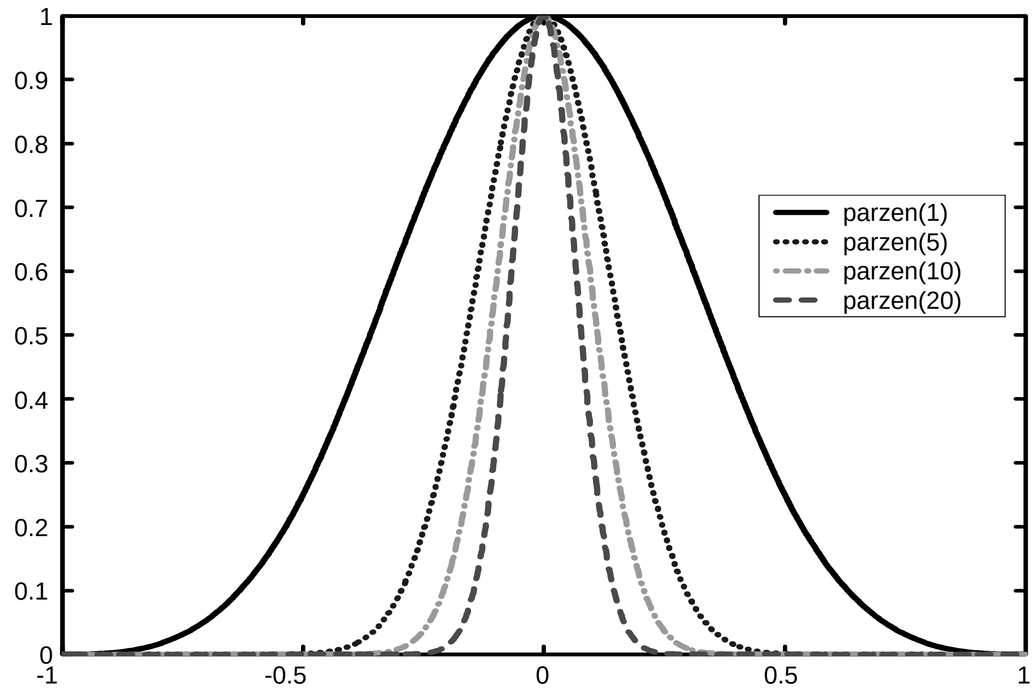

Figure 1: Steep Origin Kernels with Parzen Kernel as the Mother Kernel

2.2 Asymptotic Bias, Variance and MSE Properties of the LRV/HAC estimator

This section develops an asymptotic theory for the spectral estimator $\widehat{f}_{XX}(0)$ when $\rho \to \infty$ as $T \to \infty$. Under certain rate conditions on ρ, we show that $\widehat{f}_{XX}(0)$ is consistent for $f_{XX}(0)$ and has a limiting normal distribution. Of course, as is apparent from Figs. 1 and 2, 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 $T \to \infty$.

To establish the asymptotic bias and variance of $\widehat{f}_{XX}(0)$, we use the conditions below.

Assumption 2: *X_t is a mean zero stationary linear process*

$$X_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j \parallel C_j \parallel < \infty, \tag{7}$$

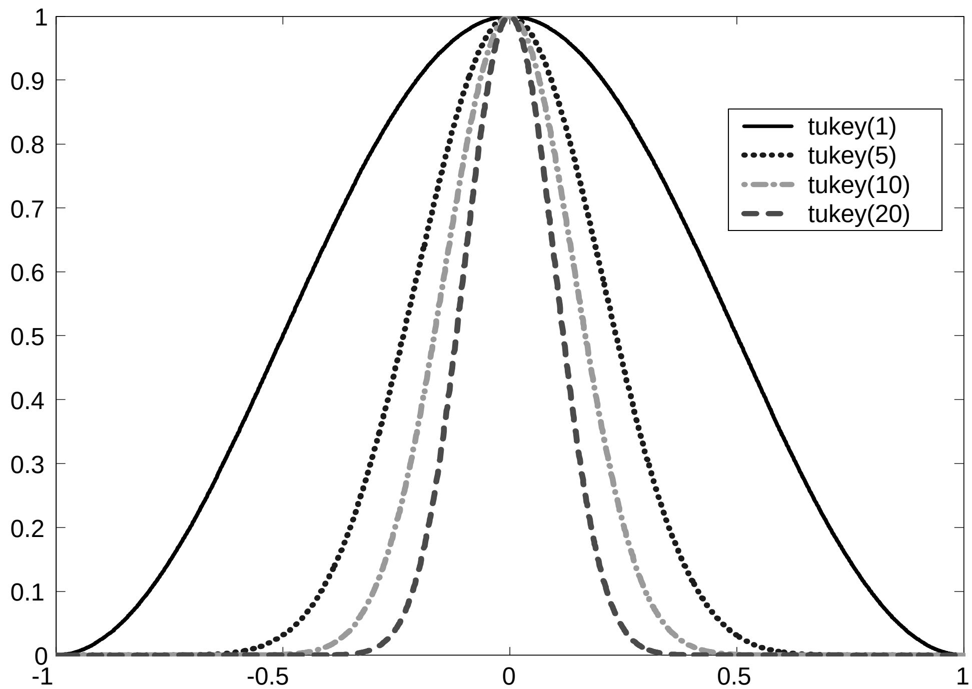

Figure 2: Steep Origin Kernels with a Tukey Kernel as the Mother Kernel

where ε_t *is* $iid(0, \Sigma_\varepsilon)$ *with* $E \parallel \varepsilon_t \parallel^4 < \infty$.

Assumption 3: $T/\rho + \rho/T^2 \to 0$ *as* $T \to \infty$ *and* $\rho \to \infty$.

Assumption 2 is convenient and includes many series of interest in applications, although condition (7) is stronger than necessary in establishing results for the asymptotic bias and variance. Let $f_{XX}^{(2)}(0) = \sum_{-\infty}^{\infty} h^2 \gamma_h$, then Assumption 2 implies that

$$\left| f_{XX}^{(2)}(0) \right| \leq \sum_{-\infty}^{\infty} h^2 \left| \gamma_h \right| < \infty. \tag{8}$$

The boundedness of $f_{XX}^{(2)}(0)$ is often assumed in the LRV estimation and spectral density estimation literature, ensuring that the spectral density has some degree of smoothness. In particular, (8) ensures that $f_{XX}(\lambda)$ is twice continuously differentiable and that results for the asymptotic bias, variance and MSE of kernel estimates can be derived. However, the linear process assumption facilitates asymptotic calculations and is particularly useful in establishing a central limit theory for our estimates.

Assumption 3 imposes both upper and lower bounds on the rate that ρ approaches infinity. Given the lower bound $T/\rho \to 0$, we can use either the biased covariance estimate $\widehat{\gamma}_h$ as in (3) or the unbiased covariance estimate $\widetilde{\gamma}_h$ in the construction of $\widehat{f}_{XX}(0)$. The unbiased covariance estimate $\widetilde{\gamma}_h$ is equal to $\widehat{\gamma}_h$ divided by the factor $(1 - |h|/T)$. Both approaches lead to the same asymptotic results. Some simulations show that the form involving the usual biased covariance estimator works better in practice. The upper bound $\rho/T^2 \to 0$ ensures that the asymptotic bias diminishes as T goes to ∞. It also helps prove the concentration of the spectral window, which is defined in (9) and whose asymptotic form is given in Lemma 2 below.

Assumption 3 holds when $\rho = aT^b$ for some $a > 0$ and $1 < b < 2$. Note that the expansion rate for ρ implied by Assumption 3 is very different from the rate condition $\frac{1}{\rho} + \frac{\rho \log T}{T} \to 0$ that was used in developing an asymptotic theory for the sharp origin kernel in PSJ (2003). While $\rho/T \to 0$ in that case, we require $T/\rho \to 0$ in the present case, so that ρ tends to infinity much faster. The reason for this difference is the Bartlett mother kernel rapidly decays from unity at the origin and less exponentiation is required with this kernel in order to achieve a similar degree of weighting to the autocorrelogram. On the other hand, the sharp behavior of the Bartlett kernel at the origin prevents a second order development that enables a higher rate of convergence in the kernel estimator. So, with this accommodating rate condition on ρ, we have the opportunity to achieve both objectives in exponentiating a quadratic kernel.

Let

$$K_\rho(\lambda) = \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})e^{i\lambda h}, \tag{9}$$

be the spectral window and

$$I_{XX}(\lambda) = \left| \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{i\lambda t} \right|^2 = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} \widehat{\gamma}_h e^{-i\lambda h}, \tag{10}$$

be the periodogram. Then

$$\widehat{f}_{XX}(0) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widehat{\gamma}_h = \frac{1}{T}\sum_{s=0}^{T-1} K(\lambda_s) I_{XX}(\lambda_s). \tag{11}$$

Note that $\widehat{\gamma}_h = \int_{-\pi}^{\pi} I_{XX}(\lambda)e^{i\lambda h}d\lambda$, so $\widehat{f}_{XX}(0)$ can also be written as

$$\widehat{f}_{XX}(0) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widehat{\gamma}_h = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_\rho(\lambda) I_{XX}(\lambda)d\lambda. \tag{12}$$

The two representations in (11) and (12) will be used in establishing the asymptotic variance of $\widehat{f}_{XX}(0)$ in the theorem below.

Before stating the theorem, we introduce some notation. Let K_{mm} be the $m^2 \times m^2$ commutation matrix (e.g. Magnus and Neudecker, 1979). Define the Mean Squared Error (MSE) as

$$\text{MSE}(\widehat{f}_{XX}(0), W) = E\left\{\text{vec}(\widehat{f}_{XX}(0) - f_{XX}(0))' W \,\text{vec}(\widehat{f}_{XX}(0) - f_{XX}(0))\right\},$$

for some $m^2 \times m^2$ positive semi-definite weight matrix W.

Theorem 1 *Let Assumptions 1-3 hold. Then, we have:*

(a) $\lim_{T\to\infty}\left(T^2/\rho\right)\left(E\widehat{f}_{XX}(0)-f_{XX}(0)\right)=-gf_{XX}^{(2)}(0)$.

(b) $\lim_{T\to\infty}\left(\frac{\pi}{2\rho g}\right)^{-1/2}Var\left(\text{vec}(\widehat{f}_{XX}(0))\right)=(I+K_{mm})f_{XX}(0)\otimes f_{XX}(0)$.

(c) If $\rho^5/T^8\to\vartheta\in(0,\infty)$, *then*

$$\begin{aligned}&\lim_{T\to\infty}T^{4/5}\text{MSE}(\widehat{f}_{XX}(0),W)\\ =\ &\vartheta^{2/5}g^2\text{vec}\left(f_{XX}^{(2)}(0)\right)'W\text{vec}\left(f_{XX}^{(2)}(0)\right)\\ &+\left(\frac{\pi}{2g}\right)^{1/2}\vartheta^{-1/10}tr\left\{W(I+K_{mm})f_{XX}(0)\otimes f_{XX}(0)\right\}.\end{aligned}$$

Results (a) and (b) of Theorem 1 are similar to those for the LRV estimate based on a sharp origin kernel in PSJ (2003). They also bear similarities to those for conventional LRV estimates as given, for example, in Andrews (1991). Note that the asymptotic variance of $\widehat{f}_{XX}(0)$ depends explicitly on $f_{XX}(0)$ and the Parzen exponent parameter ρg. In fact, as the proof of part (b) makes clear the asymptotic variance of $\widehat{f}_{XX}(0)$ can be written in the more conventional form

$$\int_{-1}^{1}k_\rho^2(x)\,dx(I+K_{mm})\left(f_{XX}(0)\otimes f_{XX}(0)\right),$$

involving the second moment of the kernel $k_\rho(x)$. However, as $\rho\to\infty$, $k_\rho(x)$ concentrates at the origin and a Laplace approximation gives

$$\int_{-1}^{1}k_\rho^2(x)=\left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)), \tag{13}$$

as shown in (58) in the Appendix. Thus, the critical parameter affecting the asymptotic variance is g, the Parzen exponent of the mother kernel $k(x)$. This point turns out to be important in constructing comparable exponent sequences for comparing kernels as discussed below.

Since $k_\rho(x)$ becomes successively more concentrated at the origin as ρ and T increase, 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 $T\to\infty$. The increase of the asymptotic bias and the decrease of the asymptotic variance with ρg reflect the usual bias/variance trade-off. As in the conventional case, the absolute asymptotic bias increases with the curvature of the true spectral density at the origin.

Observe that when $\rho^5/T^8\to\vartheta\in(0,\infty)$, $\text{MSE}(\widehat{f}_{XX}(0),W)=O\left(T^{-4/5}\right)$. So $\widehat{f}_{XX}(0)$ converges to $f_{XX}(0)$ at the rate of $O\left(T^{-2/5}\right)$, which is a faster rate than in the case of the sharp origin Bartlett kernel. In the latter case, the optimal rate for the exponent was found to be $\rho=O\left(T^{2/3}\right)$ and the rate of convergence of the estimate to be $O\left(T^{-1/3}\right)$. The $T^{-2/5}$ rate of convergence for the steep origin kernel estimate represents an improvement on the sharp origin Bartlett kernel. Note that the

$T^{-2/5}$ rate for the steep origin kernel estimate is the same as that of a conventional (truncated) quadratic kernel estimate with an optimal choice of bandwidth (e.g., Hannan, 1970; Andrews, 1991).

With the given expressions for the asymptotic MSE, we may proceed to compare different mother kernels. However, the mother kernels satisfying Assumption 1 are not subject to any normalization. In other words, both $k(x)$ and $k^{\alpha}(x)$ for any $\alpha \in \mathbb{R}^{+}$ can be used as mother kernels to construct steep origin kernels. It is therefore meaningless to compare two kernels using the same sequence of exponents. To make the comparison meaningful, we use comparable exponents defined in the following sense. Suppose $k_1(x)$ is the reference kernel and $\rho_{T,1}$ is a sequence of exponents to be used with $k_1(x)$. Then the comparable sequence of exponents for kernel $k_2(x)$ is $\rho_{T,2}$ such that

$$\lim_{T\to\infty}\left(\frac{\pi}{2\rho_{T,1}}\right)^{-1/2}\operatorname{Var}\left(\operatorname{vec}(\widehat{f}_{XX}^{(1)}(0))\right)=\lim_{T\to\infty}\left(\frac{\pi}{2\rho_{T,2}}\right)^{-1/2}\operatorname{Var}\left(\operatorname{vec}(\widehat{f}_{XX}^{(2)}(0))\right), \tag{14}$$

where $\widehat{f}_{XX}^{(1)}(0)$ and $\widehat{f}_{XX}^{(2)}(0)$ are spectral density estimates based on $k_1(x)$ and $k_2(x)$, respectively. In view of Theorem 1(b), this definition yields

$$\rho_{T,2} = g_1 \rho_{T,1}/g_2, \tag{15}$$

where g_1 and g_2 are the Parzen parameters for the two kernels (i.e. $g_1 = -1/2k_1^{''}(0)$, $g_2 = -1/2k_2^{''}(0)$). The requirement (15) for $\rho_{T,1}$ and $\rho_{T,2}$ to be comparable exponent sequences adjusts for the scale differences in the kernels that is reflected in the asymptotic approximation (13) of the second moment of the mother kernel.

When comparable exponents are employed, it is easy to see that the pairs $(k_1(x), \rho_{T,1})$ and $(k_2(x), \rho_{T,2})$ produce estimates with the same asymptotic bias, variance and MSE. This is expected, as the second order derivative $k^{''}(0)$ is the only parameter that appears in the expressions for asymptotic bias, variance and MSE. Alternatively, we can normalize the mother kernels first and then compare the mean squared errors of the resulting LRV estimates, using the same exponent. As an example, let the Parzen kernel be the reference kernel. The normalized Tukey kernel is,

$$k_{TH}^{o}(x) = \left(\frac{(1+\cos\pi x)}{2}\right)^{24/\pi^2} 1\left\{|x| \le 1\right\}. \tag{16}$$

Then, for any ρ satisfying Assumption 3, $(k_{PR}(x))^{\rho}$ and $(k_{TH}^{o}(x))^{\rho}$ will deliver LRV estimates with the same asymptotic MSE.

Thus, our asymptotic theory shows that all quadratic kernels are equivalent asymptotically. In effect, since the exponentiated kernels concentrate as $\rho, T \to \infty$, what matters asymptotically is the local shape of the mother kernel at the origin. When comparable exponent sequences as in (15) are employed, it follows that the asymptotic MSE's of the kernel LRV estimates are the same for all mother kernels with the same Parzen exponent (here $q = 2$).

The equivalence of quadratic kernels in our context is in contrast to earlier results in the LRV/spectral density estimation literature. In the conventional spectral density estimation, Priestley (1962; 1981, pp.567-571) showed by a variational argument that the quadratic spectral kernel is preferred in terms of an asymptotic MSE criterion to other quadratic kernels when comparable bandwidths are used. Later, Andrews (1991) utilized this result in the context of LRV/HAC estimation. Priestley's variational argument involves optimizing a quadratic functional with respect to the spectral window. In the case of steep origin kernels, Lemma 2 below shows that the spectral window $K_\rho(\lambda)$ has the same asymptotic normal behavior (up to scaling by the fixed parameter g) for all quadratic kernels windows. This explains the asymptotic MSE equivalence of steep origin quadratic kernels.

Of course, the equivalence of quadratic kernels in our case holds only asymptotically when T is large. In finite samples, different quadratic kernels lead to estimates with different performance characteristics and they are well known to have different properties. For example, the Parzen kernel is positive definite and the resulting LRV estimate is guaranteed to be nonnegative. This property is certainly desirable and is not shared by the Tukey-Hanning kernel which is not positive definite.

2.3 Optimal Exponent Selection

Theorem 1(c) reveals that there is an opportunity for optimal selection of ϑ. The first order condition for minimizing the scaled asymptotic MSE is

$$\begin{aligned} & \frac{2}{5}\vartheta^{-3/5} g^2 \text{vec}\left(f_{XX}^{(2)}(0)\right)' W \text{vec}\left(f_{XX}^{(2)}(0)\right) \\ = & \frac{1}{10}\left(\frac{\pi}{2g}\right)^{1/2} \vartheta^{-11/10} \text{tr}\ \left\{W(I+K_{mm}) f_{XX}(0) \otimes f_{XX}(0)\right\}, \end{aligned} \tag{17}$$

leading to

$$\vartheta = \left(\frac{\left(\frac{\pi}{2g}\right)^{1/2} \text{tr}\ \left\{W(I+K_{mm}) f_{XX}(0) \otimes f_{XX}(0)\right\}}{4g^2 \text{vec}\left(f_{XX}^{(2)}(0)\right)' W \text{vec}\left(f_{XX}^{(2)}(0)\right)}\right)^2.$$

So the optimal ρ is

$$\rho^* = T^{8/5} g^{-1} \left[\frac{\sqrt{\pi}\text{tr}\ \left\{W(I+K_{mm}) f_{XX}(0) \otimes f_{XX}(0)\right\}}{4\sqrt{2}\text{vec}\left(f_{XX}^{(2)}(0)\right)' W \text{vec}\left(f_{XX}^{(2)}(0)\right)}\right]^{2/5}. \tag{18}$$

For illustrative purposes, suppose X_t is a scalar AR(1) process such that $X_t = \alpha X_{t-1} + \epsilon_t$, $\epsilon_t \sim iid(0, \sigma^2)$. Then

$$f_{XX}(0) = \frac{\sigma^2}{2\pi}\frac{1}{(1-\alpha)^2},\ f_{XX}^{(1)} = \frac{\sigma^2}{2\pi}\frac{2\alpha}{(1-\alpha)^3(1+\alpha)},\ \text{and}\ f_{XX}^{(2)} = \frac{\sigma^2}{2\pi}\frac{2\alpha}{(1-\alpha)^4}.$$

Hence,

$$\rho^*_{steep} = T^{8/5} g^{-1} \left(\frac{\sqrt{2\pi}}{16}\frac{(1-\alpha)^4}{\alpha^2}\right)^{2/5}. \tag{19}$$

For this choice of ρ, the RMSE is

$$\mathrm{RMSE}^*_{steep} = 2.1306T^{-2/5}\alpha^{1/5}(1-\alpha)^{-12/5}. \tag{20}$$

In contrast, when sharp origin kernels are used in the construction of $\widehat{f}_{XX}(0)$, PSJ (2003) showed that the optimal exponent satisfies

$$\rho^*_{sharp} = T^{2/3}\left(\frac{(1-\alpha^2)^2}{4\alpha^2}\right)^{1/3} \tag{21}$$

and the resulting RMSE is

$$\mathrm{RMSE}^*_{sharp} = \sqrt{3}T^{-1/3}(1-\alpha)^{-2}. \tag{22}$$

The ratio of the respective RMSE's of the sharp and steep kernel estimates is

$$\frac{\mathrm{RMSE}^*_{sharp}}{\mathrm{RMSE}^*_{steep}} = 0.81294T^{1/15}(1-\alpha)^{2/5}\alpha^{-1/5}. \tag{23}$$

Table 1 tabulates ρ^*_T for the sharp origin kernel and the steep origin kernel for different values of T. For steep kernels, we choose the Parzen kernel as the mother kernel as it is representative of other quadratic kernels. As implied by the asymptotics, the values of ρ^*_T are much larger for the steep origin kernel than the sharp origin kernel. Since the ratio $\mathrm{RMSE}^*_{sharp}/\mathrm{RMSE}^*_{steep}$ is of order $T^{1/15}$, the sharp kernel estimate is 100% less efficient asymptotically than the steep kernel estimate. Finite sample performance may not necessarily follow this ordering, however, and will depend on the magnitudes of $T, f, f^{(1)}$ and $f^{(2)}$. For example, in the AR(1) case, when the autoregression parameter is very close to 1, the sharp kernel estimate may have a smaller RMSE than the steep kernel estimate for moderate T.

Table 1: Asymptotically Optimal ρ^* for the Sharp Kernel and Steep Kernel for AR (1) Processes

	Sharp Kernel				Steep Kernel			
α	T=50	100	200	1000	T=50	100	200	1000
$\alpha = .04$	73	115	184	538	510	1548	4693	61634
$\alpha = .09$	42	67	106	311	245	742	2251	29574
$\alpha = .25$	20	32	52	152	79	240	729	9584
$\alpha = .49$	11	18	28	84	25	75	229	3018
$\alpha = .81$	3	4	7	22	3	10	31	415
$\alpha = .90$	2	3	5	17	1	3	10	136

2.4 Central Limit Theory

We proceed to investigate the limiting distribution of $\widehat{f}_{XX}(0)$. In view of (11) and (12), it is apparent that the asymptotic distribution of $\widehat{f}_{XX}(0)$ is that of a smoothed version of the periodogram and depends on the spectral window $K_\rho(\lambda_s)$, whose asymptotic form as $T \to \infty$ is given in the next result.

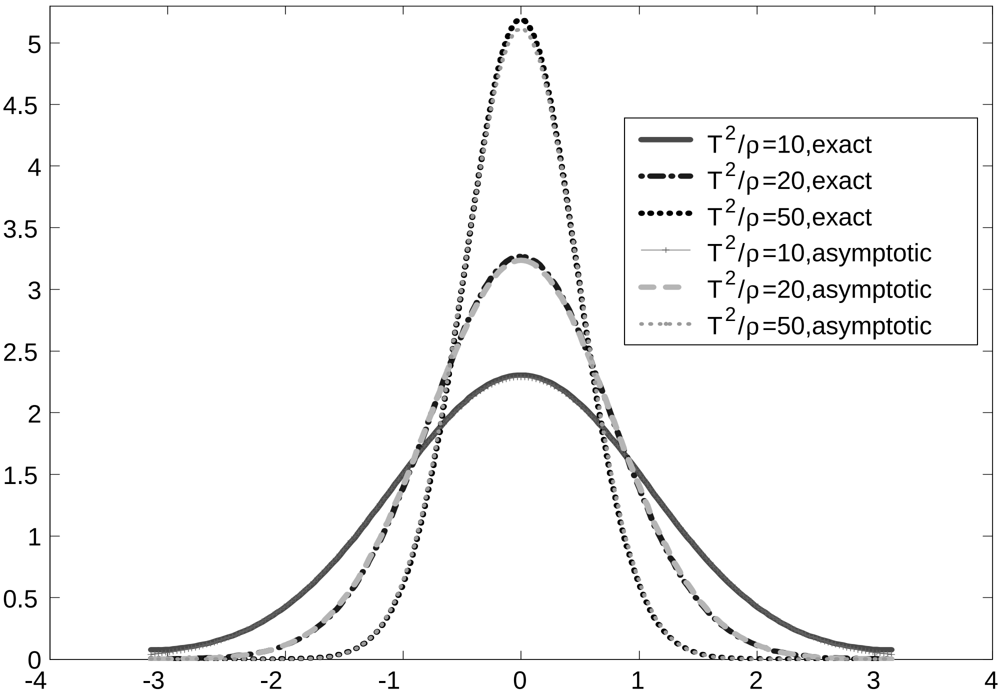

Figure 3: Spectral Windows of Steep Parzen Kernels

Lemma 2 *Let Assumptions 1 and 3 hold. Then, as* $T \to \infty$

$$
\begin{aligned}
K_\rho(\lambda_s) &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \exp\left(-\frac{\pi^2 s^2}{\rho g}\right)(1+o(1)) \\
&= \begin{cases} O\left(\frac{T}{\sqrt{\rho}}\right) & \textit{for } s \le O(\sqrt{\rho}), \\ O\left(\frac{Te^{-\frac{\pi^2 s^2}{\rho g}}}{\sqrt{\rho}}\right) & \textit{for } s > O(\sqrt{\rho}). \end{cases}
\end{aligned}
$$

It follows from Lemma 2 that $K_\rho(\lambda_s)$ is asymptotically equivalent to

$$\frac{\sqrt{\pi}T}{\sqrt{\rho g}} \exp\left(-\frac{T^2\lambda_s^2}{4\rho g}\right), \tag{24}$$

which is proportional to a normal density with mean zero and variance of order $O\left(\rho/T^2\right)$. The graph of $K_\rho(\lambda)$ with Parzen kernel as the mother kernel is shown in Fig. 3 for $T^2/\rho = 10, 20, 50$ and $T = 200$. The graph shows that the exact expression as defined in (9) is almost indistinguishable from the asymptotic expression as defined in (24). The peak in the spectral window at the origin increases and the window becomes steeper as T^2/ρ increases because $K_\rho(0) = O\left(T/\sqrt{\rho}\right)$, as is clear from Lemma 2.

Theorem 3 *Let Assumptions 1-3 hold, then*

$$\rho^{1/4}\left\{\widehat{f}_{XX}(0)-f_{XX}(0)\right\}\rightarrow_d \ N\left(0,\left(\frac{\pi}{2g}\right)^{1/2}(I+K_{mm})f_{XX}(0)\otimes f_{XX}(0)\right).$$

As in the proof of Lemma 2, the derivation of this result makes use of the Laplace method to approximate integrals (see, e.g. De Bruijn 1982). The asymptotic normality result permits us to make inference on $f_{XX}(0)$, which we discuss further in the following section.

3 Spectral Density Estimation with Steep Origin Kernels

We consider estimating the spectral density at an arbitrary point $\omega \in (0,\pi)$ and extend the asymptotic theory of the previous section to this general case. The results for $\omega = 0$ (and also $\omega = \pi$) given above continue to apply with minor modifications. The steep kernel estimator of $f_{XX}(\omega)$ is

$$\widehat{f}_{XX}(\omega)=\frac{1}{2\pi}\sum_{h=-T+1}^{T-1}k_\rho(\frac{h}{T})\widehat{\gamma}_h e^{-ih\omega} \tag{25}$$

where $\widehat{\gamma}_h$ is defined as before. When $\omega = 0$, the estimator reduces to the estimator in (2).

Following arguments similar to those in Section 2.2, we can prove the theorem below.

Theorem 4 *Let Assumptions 1-3 hold. Then for $\omega \neq 0,\pi$,*

(a) $\lim_{T\to\infty}\left(T^2/\rho\right)\left(E\widehat{f}_{XX}(\omega)-f_{XX}(\omega)\right)=-gf_{XX}^{(2)}(\omega)$ *where*

$$f_{XX}^{(2)}(\omega)=\sum_{-\infty}^{\infty}h^2\gamma_h e^{-ih\omega}. \tag{26}$$

(b) $\lim_{T\to\infty}\left(\frac{\pi}{2\rho g}\right)^{-1/2}Var\left(\text{vec}(\widehat{f}_{XX}(\omega))\right)=f_{XX}(\omega)\otimes f_{XX}'(\omega).$

(c) If $\rho^5/T^8\to\vartheta\in(0,\infty)$*, then*

$$\begin{aligned}&\lim_{T\to\infty}T^{4/5}\text{MSE}(\widehat{f}_{XX}(\omega),W)\\ =\ &\vartheta^{2/5}g^2\text{vec}\left(f_{XX}^{(2)}(\omega)\right)'W\text{vec}\left(f_{XX}^{(2)}(\omega)\right)\\ &+\left(\frac{\pi}{2g}\right)^{1/2}\vartheta^{-1/10}tr\ \left\{W[f_{XX}(\omega)\otimes f_{XX}'(\omega)]\right\}.\end{aligned}$$

Theorem 4 shows that the earlier asymptotic results for bias, variance and MSE continue to apply for $\omega \neq 0$. The only difference between the case $\omega = 0$ (or $\omega = \pi$) and $\omega \in (0, \pi)$ lies in the asymptotic variance. This is typical of the literature on spectrum estimation. The rates of convergence are the same for all $\omega \in [0, \pi]$ and the optimal power parameter that minimizes the asymptotic MSE is still of order $T^{8/5}$. The optimal power parameter now depends on $f_{XX}^{(2)}(\omega)$ and $f_{XX}(\omega)$.

To establish the limiting distribution of $\widehat{f}_{XX}(\omega)$, we proceed as in Section 2.2 by developing an asymptotic approximation for the spectral window $K(\lambda_s - \omega)$.

Lemma 5 *Let Assumptions 1 and 3 hold. Then, as* $T \to \infty$

$$\begin{aligned} K(\lambda_s - \omega) &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \exp\left(-\frac{(\omega T - 2\pi s)^2}{4\rho g} \right)(1 + o(1)) \\ &= \begin{cases} O\left(\frac{T}{\sqrt{\rho}}\right) & \text{for } |\omega T - 2\pi s| \leq O(\sqrt{\rho}), \\ O\left(\frac{T}{\sqrt{\rho}} \exp\left(-\frac{(\omega T - 2\pi s)^2}{4\rho g}\right)\right) & \text{for } |\omega T - 2\pi s| > O(\sqrt{\rho}). \end{cases} \end{aligned}$$

The asymptotic approximation is the same as that in Lemma 2 except that $K_\rho(\lambda_s)$ now concentrates around ω. This is apparent, as Lemma 5 shows that $K(\lambda_s - \omega)$ is exponentially small when $|\omega T - 2\pi s| \to \infty$. Note that $K_\rho(\lambda_s)$ can be written as

$$\frac{\sqrt{\pi}T}{\sqrt{\rho g}} \exp\left(-\frac{(\omega - \lambda_s)^2 T^2}{4\rho g} \right)(1 + o(1)). \tag{27}$$

Therefore, the asymptotic approximation to the spectral window is proportional to a normal density with mean ω and variance of order $O(\rho/T^2)$.

Using this asymptotic representation of $K(\lambda_s - \omega)$, we establish the following central limit theorem for $\widehat{f}_{XX}(\omega)$.

Theorem 6 *Let Assumptions 1-3 hold. Then*

$$\rho^{1/4}\left\{\widehat{f}_{XX}(\omega) - f_{XX}(\omega)\right\} \to_d N\left(0, \left(\frac{\pi}{2g}\right)^{1/2} f_{XX}(\omega) \otimes f'_{XX}(\omega)\right),$$

for $\omega \neq 0, \pi$.

Again, the asymptotic distribution continues to hold with obvious modifications. The asymptotic normality results in Theorems 3 and 6 are related, of course, to much earlier results in the time series literature (see, e.g., Anderson, 1971) on the asymptotic normality of conventional spectral density estimates under regularity conditions on the bandwidth expansion rate.

Using the asymptotic normality of $\widehat{f}_{XX}(\omega)$, we may construct pointwise confidence regions for $f_{XX}(\omega)$ in the usual manner. When X_t is a scalar process,

$$\rho^{1/4}\left\{\widehat{f}_{XX}(\omega) - f_{XX}(\omega)\right\} \to_d N\left(0, V^2\right),$$

where

$$V^2 = (1+\delta_{0,\omega})\left(\frac{\pi}{2g}\right)^{1/2} f_{XX}^2(\omega) \text{ and } \delta_{0,\omega} = 1\{\omega = 0,\pi\}.$$

Thus, an approximate $100(1-\alpha)\%$ confidence interval is

$$\left[\widehat{f}_{XX}(\omega) - cv(\alpha/2)\rho^{-1/4}\widehat{V}, \widehat{f}_{XX}(\omega) + cv(\alpha/2)\rho^{-1/4}\widehat{V}\right], \tag{28}$$

where

$$\widehat{V}^2 = (1+\delta_{0,\omega})\left(\frac{\pi}{2g}\right)^{1/2} \widehat{f}_{XX}^2(\omega), \tag{29}$$

and $cv(\alpha/2)$ is the critical value of a standard normal for area $\alpha/2$ in the right tail.

In finite samples, the quantity $\rho^{-1/2}\widehat{V}^2$ in (28) may be estimated by

$$\widetilde{V}^2 = (1+\delta_{0,\omega})\left\{\frac{1}{T}\sum_{s=1}^{T-1} k_\rho^2(\frac{s}{T})\cos^2\omega s\right\}\widehat{f}_{XX}^2(\omega). \tag{30}$$

The latter expression comes from the proofs of Theorems 3 and 6. With this variance estimate, the $100(1-\alpha)\%$ confidence interval becomes

$$\left[\widehat{f}_{XX}(\omega) - cv(\alpha/2)\widetilde{V}, \widehat{f}_{XX}(\omega) - cv(\alpha/2)\widetilde{V}\right]. \tag{31}$$

The above confidence interval may have a coverage probability closer to the nominal significance level than that based on the direct asymptotic expression.

The asymptotic covariance between $\widehat{f}_{XX}(\omega_i)$ and $\widehat{f}_{XX}(\omega_j)$ for $\omega_i \neq \omega_j$ is given in the next result.

Theorem 7 *Let Assumption 1-3 holds, then for* $\omega_i \neq \omega_j$

$$\lim_{T\to\infty} \rho^{1/2}\mathrm{cov}\left(\mathrm{vec}(\widehat{f}_{XX}(\omega_i)), \mathrm{vec}(\widehat{f}_{XX}(\omega_j)\right) = 0. \tag{32}$$

According to this theorem, $\widehat{f}_{XX}(\omega_i)$ is asymptotically uncorrelated with $\widehat{f}_{XX}(\omega_j)$ for any fixed $\omega_i \neq \omega_j$, a result that is analogous to that for conventional spectral density estimators. Intuitively, $\widehat{f}_{XX}(\omega_i)$ is a weighted average of the periodogram with a weight function that becomes more and more concentrated at ω_i. The asymptotic uncorrelatedness of $\widehat{f}_{XX}(\omega_i)$ across points on the spectrum is therefore inherited from that of the periodogram. In fact, the proof of theorem shows that $\widehat{f}_{XX}(\omega_i)$ will be asymptotically uncorrelated with $\widehat{f}_{XX}(\omega_j)$ unless ω_i and ω_j are sufficiently close together in the sense that $|\omega_i - \omega_j| = o(\sqrt{\rho}/T)$. Therefore, $\sqrt{\rho}/T$ may be regarded as the effective width of the spectral window $K_\rho(\lambda)$.

4 Finite Sample Performance

In this section, we examine the finite sample performance of steep Parzen kernel methods in LRV/HAC estimation and robust regression testing in comparison with sharp Bartlett kernel and conventional Parzen kernel methods.

4.1 Spectral Density Estimation

We explore the finite sample properties of the new spectral density estimator $\widehat{f}(\omega)$ at different frequencies. The frequencies considered are $\omega = 0$, $\pi/6$, and $\pi/4$, which include the low frequency and business cycle frequencies. In order to compare the performances in a more demanding setting, we allow for spectral peaks at these frequencies.

To illustrate, suppose X_t is a scalar $AR(2)$ process: $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$ with $\varepsilon_t \sim iid\ N(0,1)$. This process has a spectral peak at ω if

$$b = \frac{a}{a - 4\cos\omega}, \tag{33}$$

provided that

$$b < 0, \text{ and } \left|\frac{a(1-b)}{4b}\right| < 1. \tag{34}$$

See Priestley (1981, pp. 241). Fig. 4 depicts combinations of (a, b) satisfying (33) for $\omega = 0$, $\pi/6$, and $\pi/4$, together with the stationary triangular region of the parameter space for the $AR(2)$ process X_t. Thus, $a \in [0, 2)$ for a stationary $AR(2)$ process with spectral peak at $\omega = 0$, $a \in [0, \sqrt{3})$ for a peak at $\omega = \pi/6$, and $a \in [0, \sqrt{2})$ for a peak at $\omega = \pi/4$. Accordingly, for our simulations, we select $a = 0, 0.4, 0.8, 1.2, 1.6$ in the second case, and $a = 0, 0.4, 0.8, 1.2$ in the third case, together with $b = a/(a - 4\cos\omega)$ for different values of ω. Figs. 5, 6, and 7 display the corresponding spectral densities of the X_t process with peaks at $\omega = 0, \pi/6$, and $\pi/4$, respectively, for $a = 0, 0.4, 0.8$ and 1.2. When $a = 0$, the process is white noise and its spectral density is flat in each case. As a increases, we move closer to the boundary of the stationary region, and the spectral densities become progressively more peaked at the corresponding values of ω. The second order derivative of the spectral density at the origin is zero for an $AR(2)$ process that has a peak at zero, c.f. Fig. 5. Thus, the bias is expected to be of smaller order and our optimal exponent formula does not apply for that case. Instead, we use an $AR(1)$ process which has a spectral peak at zero, and select $a = 0, 0.2, 0.4, 0.6, 0.8$.

From Theorem 4 in the last section, we can show that for steep origin kernels the optimal exponent at $\omega \neq 0, \pi$ is

$$\rho^*_{steep} = T^{8/5} g^{-1} \left[\frac{\sqrt{\pi} f^2_{XX}(\omega)}{4\sqrt{2}\left(f^{(2)}_{XX}(\omega)\right)^2} \right]^{2/5}. \tag{35}$$

An analogous analysis shows that for sharp origin kernels the optimal exponent at $\omega \neq 0, \pi$ is

$$\rho^*_{sharp} = T^{2/3} \left[\frac{f^2_{XX}(\omega)}{2\left(f^{(1)}_{XX}(\omega)\right)^2} \right]^{1/3}, \tag{36}$$

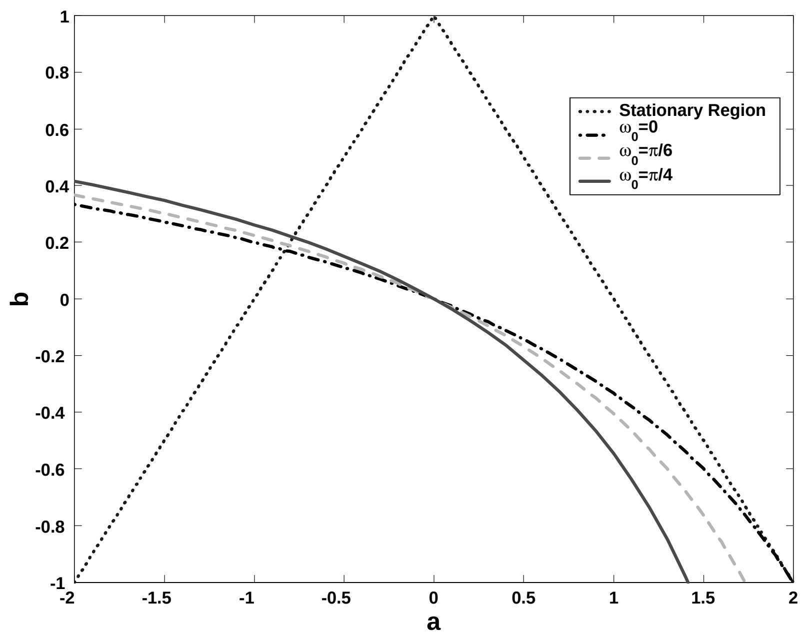

Figure 4: Stationary region and (a, b) combinations satisfying $b = a/(a - 4\cos\omega)$ for $\omega = 0$, $\pi/6$, $\pi/4$

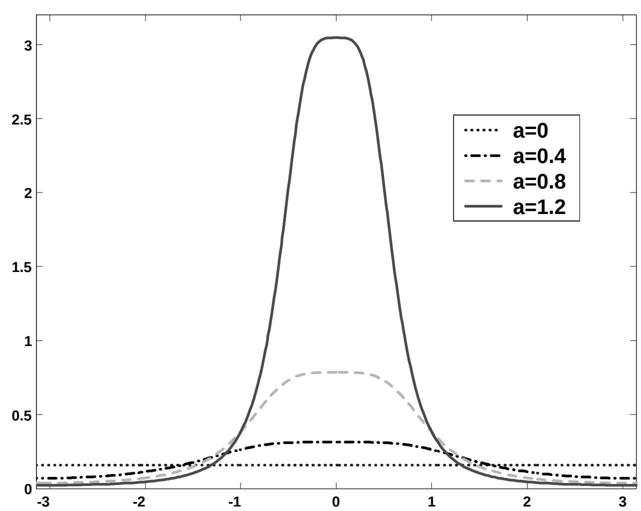

Figure 5: Spectral density of $AR(2)$ process with peak at $\omega = 0$

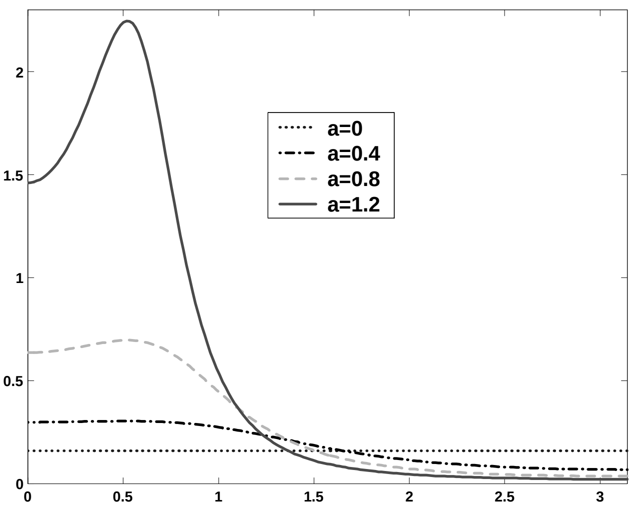

Figure 6: Spectral density of $AR(2)$ process with peak at $\omega = \pi/6$

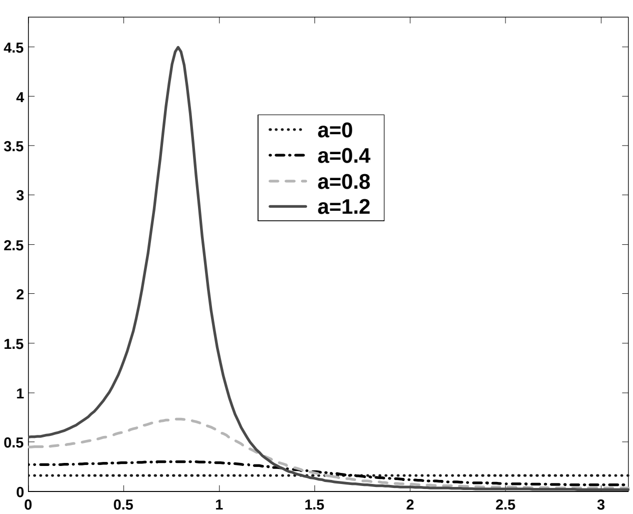

Figure 7: Spectral density of $AR(2)$ process with peak at $\omega = \pi/4$

and for the conventional estimator with Parzen kernel the optimal bandwidth at $\omega \neq 0, \pi$ is

$$S_T = 2.6614 T^{1/5} \left[\frac{2 \left(f_{XX}^{(2)}(\omega) \right)^2}{f_{XX}^2(\omega)} \right]^{1/5}. \tag{37}$$

Table 2: Ratio of $RMSE$ of steep and sharp estimators to $RMSE$ of Parzen estimator using $AR(1)$ plug-in exponents or bandwidths for $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$ with $b = \mathbf{1}_{\{\omega \neq 0\}} a/(a - 4\cos\omega)$ and $\varepsilon_t \sim iid(0,1)$.

		$T = 50$		$T = 100$		$T = 200$	
	a	Sharp	Steep	Sharp	Steep	Sharp	Steep
$\omega = 0$	0.0	0.929	0.987	0.946	0.984	0.954	0.983
	0.2	0.945	0.993	1.007	0.994	1.063	0.995
	0.4	0.955	0.994	1.027	0.995	1.091	0.994
	0.6	0.922	0.988	0.996	0.995	1.067	0.995
	0.8	0.850	0.945	0.917	0.986	0.995	0.995
$\omega = \pi/6$	0.0	0.956	0.988	0.948	0.987	0.949	0.987
	0.4	0.901	1.000	0.949	1.008	0.982	1.015
	0.8	1.138	0.995	1.329	0.994	1.557	0.991
	1.2	1.062	0.993	1.098	0.997	1.137	0.997
	1.6	0.994	1.007	0.960	1.004	0.915	1.008
$\omega = \pi/4$	0.0	0.980	0.991	0.985	0.989	0.972	0.988
	0.4	1.014	1.003	1.084	1.006	1.179	1.010
	0.8	1.157	0.999	1.204	1.003	1.244	1.006
	1.2	1.035	1.005	1.014	1.008	0.984	1.013

For data-based implementation of these formulae, we use both $AR(1)$ and $AR(2)$ plug-in approaches (as in Andrews, 1991). The $RMSE's$ of the different methods are compared for $T = 50, 100, 200$ in Tables 2 and 3. It is clear that for $\omega = 0$, both the steep and sharp estimators outperform the conventional Parzen estimator when the sample size is small. As T becomes larger, the steep estimator continues to outperform the Parzen estimator, while the sharp estimator outperforms the latter only when the $AR(1)$ process is close to extreme cases like white noise or nonstationarity. For $\omega = \pi/6$ and $\pi/4$, the steep kernel estimator seems to outperform the sharp kernel estimator in most cases when we use the $AR(2)$ plug-in approach, while the sharp estimator is better when the $AR(2)$ process is close to extreme cases like white noise or nonstationarity.

Table 3: Ratio of $RMSE$ of steep and sharp estimators to $RMSE$ of Parzen estimator using $AR(2)$ plug-in exponents or bandwidths for $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$ with $b = \mathbf{1}_{\{\omega \neq 0\}} a/(a - 4\cos\omega)$ and $\varepsilon_t \sim iid(0,1)$.

		$T = 50$		$T = 100$		$T = 200$	
	a	Sharp	Steep	Sharp	Steep	Sharp	Steep
$\omega = 0$	0.0	0.877	0.986	0.907	0.983	0.922	0.982
	0.2	0.899	0.990	0.968	0.992	1.028	0.995
	0.4	0.922	0.991	0.990	0.997	1.044	0.998
	0.6	0.896	0.983	0.970	0.996	1.048	0.996
	0.8	0.835	0.932	0.903	0.982	0.985	0.995
$\omega = \pi/6$							
	0.0	0.914	0.990	0.943	0.988	0.960	0.987
	0.4	0.996	0.999	1.048	1.003	1.074	1.008
	0.8	1.052	1.000	1.079	1.003	1.097	1.002
	1.2	0.996	0.988	1.032	0.997	1.074	0.999
	1.6	0.989	0.935	0.960	0.960	0.985	0.986
$\omega = \pi/4$							
	0.0	0.923	0.992	0.936	0.991	0.942	0.991
	0.4	1.022	0.999	1.060	1.000	1.086	1.000
	0.8	0.997	0.993	1.040	0.998	1.093	0.998
	1.2	0.942	0.954	0.978	0.984	1.033	0.997

Table 4: Ratio of $RMSE$ of steep and sharp estimators to $RMSE$ of Parzen estimator using infeasible exponents or bandwidths for $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$, with $b = \mathbf{1}_{\{\omega \neq 0\}} a/(a - 4\cos\omega)$ and $\varepsilon_t \sim iid(0,1)$

		$T = 50$		$T = 100$		$T = 200$	
	a	Sharp	Steep	Sharp	Steep	Sharp	Steep
$\omega = 0$	0.2	0.989	0.992	1.037	0.991	1.091	0.992
	0.4	1.008	0.993	1.061	0.993	1.122	0.992
	0.6	0.999	0.994	1.042	0.996	1.101	0.993
	0.8	0.940	0.996	1.000	0.994	1.045	0.996
$\omega = \pi/6$							
	0.4	1.023	1.000	1.061	1.000	1.122	1.000
	0.8	1.064	1.000	1.113	1.000	1.163	1.000
	1.2	1.006	0.988	1.045	0.994	1.096	0.996
	1.6	0.996	0.974	0.988	0.978	1.021	0.997
$\omega = \pi/4$							
	0.4	1.041	1.000	1.085	0.996	1.129	0.993
	0.8	1.011	0.992	1.053	0.994	1.107	0.995
	1.2	0.981	0.978	1.010	1.000	1.064	0.998

If we assume the data generating processes are known, a comparison of the $RMSE$ of the different methods is presented in Table 4. The steep kernel estimator is better than the other two procedures for most cases when $\omega = 0$, and when $\omega = \pi/6$ and $\pi/4$, especially when the $AR(2)$ parameters are close to the boundary of the stationary region. This suggests that the steep kernel estimator has the potential to do better in worse situations (i.e., when there is a large peak in the spectrum) when $\omega \neq 0$.

4.2 Robust Hypothesis Testing

Using the steep kernel LRV estimator, we propose a new approach to robust hypothesis testing. Consider the linear regression model:

$$y_t = z_t'\beta + u_t, \quad t = 1, 2, ..., T, \tag{38}$$

where u_t is autocorrelated and possibly conditionally heteroskedastic and z_t is an $m \times 1$ vector of regressors. Suppose we want to test the null $H_0 : R\beta = r$ against the alternative $H_1 : R\beta \neq r$ where R is a $p \times m$ matrix. Let $\widehat{\beta}$ be the OLS estimator and $\widehat{Q}$ be $1/T \sum_{t=1}^{T} z_t z_t'$. Then the usual F-statistic is

$$F_\rho^* = T(R\widehat{\beta} - r)' \left(R\widehat{Q}^{-1}\widehat{\Omega}_\rho \widehat{Q}^{-1} R' \right)^{-1} (R\widehat{\beta} - r)/p, \tag{39}$$

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

$$t_\rho^* = T^{1/2}(R\widehat{\beta} - r) \left(R\widehat{Q}^{-1}\widehat{\Omega}_\rho \widehat{Q}^{-1} R' \right)^{-1/2}, \tag{40}$$

where $\widehat{\Omega}_\rho = 2\pi \widehat{f}_{XX}(0)$, $\widehat{f}_{XX}(0)$ is defined in (2) with X_t replaced by $z_t(y_t - z_t'\widehat{\beta})$.

Let $\widehat{\rho}$ be the data-driven exponent as defined in (19) with α replaced by the first order autocorrelation of X_t. Using the results in the previous sections and following the arguments similar to the proof of Theorem 3 in PSJ (2003), we can show that under Assumptions 1-3,

$$pF_{\widehat{\rho}}^* \Rightarrow W_p'(1)W_p(1) =_d \chi_p^2, \; t_{\widehat{\rho}}^* \Rightarrow W_1(1) =_d N(0, 1), \tag{41}$$

under the null hypothesis, and

$$pF_{\widehat{\rho}}^* \Rightarrow \left(\Lambda^{*-1}c + W_p(1) \right)' \left(\Lambda^{*-1}c + W_p(1) \right), \; t_{\widehat{\rho}}^* \Rightarrow (\gamma + W_1(1)), \tag{42}$$

under the local alternative hypothesis $H_1 : R\beta = r + cT^{-1/2}$. Here $\Lambda^*\Lambda^{*\prime} = RQ^{-1}\Omega Q^{-1}R'$, $\gamma = c(RQ^{-1}\Omega Q^{-1}R')^{-1/2}$, and $W_p(r)$ is p-dimensional standard Brownian motion.

The above limiting distributions hold under large ρ asymptotics in which the exponent $\widehat{\rho}$ approaches infinity at a suitable rate so that we have consistent HAC estimates. It is known that consistent HAC estimates are not needed in order to produce asymptotically valid tests. PSJ (2003) showed that under fixed ρ asymptotics, i.e. when $T \to \infty$ for a fixed ρ, the LRV estimator $\widehat{\Omega}_\rho$ is inconsistent. Nevertheless, they showed that the F_ρ^* and t_ρ^* statistics have the following limiting distributions. First, under the null $H_0 : R\beta = r$,

$$pF_\rho^* \Rightarrow W_p'(1) \left(\int_0^1 \int_0^1 k_\rho(r - s) dV_p(r) dV_p'(s) \right)^{-1} W_p(1), \tag{43}$$

and

$$t_\rho^* \Rightarrow W_1(1)\left(\int_0^1\int_0^1 k_\rho(r-s)dV_1(r)dV_1'(s)\right)^{-1/2}. \tag{44}$$

Second, under the local alternative $H_1 : R\beta = r + cT^{-1/2}$,

$$pF_\rho^* \Rightarrow \left(\Lambda^{*-1}c + W_p(1)\right)'\left(\int_0^1\int_0^1 k_\rho(r-s)dV_p(r)dV_p'(s)\right)^{-1}\left(\Lambda^{*-1}c + W_p(1)\right), \tag{45}$$

and

$$t_\rho^* \Rightarrow (\gamma + W_1(1))\left(\int_0^1\int_0^1 k_\rho(r-s)dV_1(r)dV_1(s)\right)^{-1/2}. \tag{46}$$

In these formulae, $k_\rho(\cdot)$ is any positive semi-definite kernel (so the steep Parzen kernel may be used), and $V_p(r)$ is p-dimensional standard Brownian bridge.

Given the above fixed ρ asymptotics, the critical values for different ρ values can be simulated and tabulated. For details, the reader is referred to PSJ (2003). It turns out that the critical values at a given significance level can be represented approximately by a hyperbola of the form:

$$cv = \frac{b}{\rho - a} + c, \tag{47}$$

where c is the critical value from the standard normal. Table 5 presents nonlinear least squares estimates of a and b and the standard errors of the nonlinear regressions. The standard errors are seen to be very small. Fig. 8 depicts the fitted hyperbolae for different significance levels. The figure shows the curves are nearly flat for large ρ and for $\rho > 60$ the critical values are very close to those from the standard normal. This is not surprising as the t-statistic is asymptotically normal under the large ρ asymptotics.

Table 5: Asymptotic critical value functions for the one-sided t_ρ^*-test with steep Parzen kernels

	90.0%	95.0%	97.5%	99.0%
a	−2.152	−1.884	−2.036	−2.370
b	4.260	6.604	10.012	16.015
c	1.282	1.645	1.960	2.326
$s.e.$	0.001	0.002	0.005	0.007

We now proceed to investigate the asymptotic power of the t^* test under both the fixed ρ asymptotics and the large ρ asymptotics. For convenience, we refer to these two tests as the t_ρ^* test and the $t_{\hat\rho}^*$ test, respectively. For the $t_{\hat\rho}^*$ test, the power curve is the same as the power envelope that is obtained when the true Ω or any consistent estimate is used. For the t_ρ^* test, we consider three values of $\rho : \rho = 1, 16$ and 32. For each ρ, we approximate the Brownian motion and Brownian bridge processes by the partial sums of 1000 normal variates. Fig. 9 presents the asymptotic power curves based on 50,000 simulation replications. It is apparent that the power curve moves

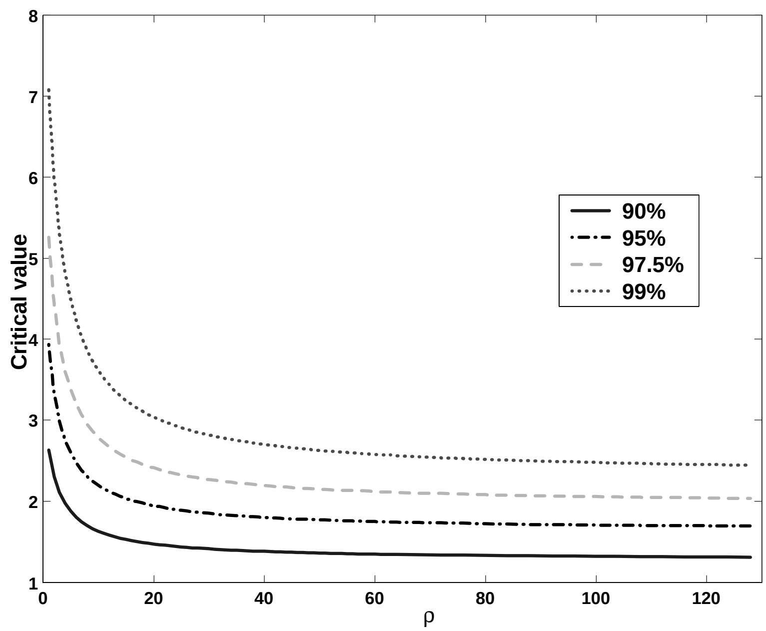

Figure 8: Asymptotic critical value curves for various significance levels

up uniformly as ρ increases, just as it does with sharp origin kernels (PSJ, 2003). The difference is that with sharp origin kernels, when $\rho \geq 16$, the power curve is very close to the power envelope, whereas much bigger values of ρ are needed here, consonant with the power parameter expansion rates established for consistent HAC estimation earlier in the paper.

Compared with the t_{ρ}^{*} test, the $t_{\widehat{\rho}}^{*}$ test has obvious power advantage. However, as with other tests that use consistent LRV estimates, the $t_{\widehat{\rho}}^{*}$ test has larger size distortion than the t_{ρ}^{*} test in finite samples. Before studying the finite sample performances of these two tests, we introduce a new test that seeks to combine the good elements of both procedures. The new test uses the same $t_{\widehat{\rho}}^{*}$ statistic defined in (40) with a data-driven $\widehat{\rho}$. The point of departure is that, instead of using the critical values from the standard normal, we propose to use the critical values from the hyperbola defined in (47). The new testing procedure is thus a mixture of the $t_{\widehat{\rho}}^{*}$ test and the t_{ρ}^{*} test. As a result, the new test has the dual advantage of an optimal choice of power parameter that is data-determined and at the same time the good finite sample size properties of the t_{ρ}^{*} test. The latter point will become clear below. Since the critical value from the hyperbola approaches that from the standard normal as $\rho \to \infty$, the new test is equivalent to the $t_{\widehat{\rho}}^{*}$ test in large samples. We will refer to the new test as the t_{new}^{*} test hereafter.

To compare the finite sample performances of the t^{*} tests (including the t_{ρ}^{*} test, $t_{\widehat{\rho}}^{*}$ test and t_{new}^{*} test) with steep Parzen kernels and the conventional (i.e., bandwidth

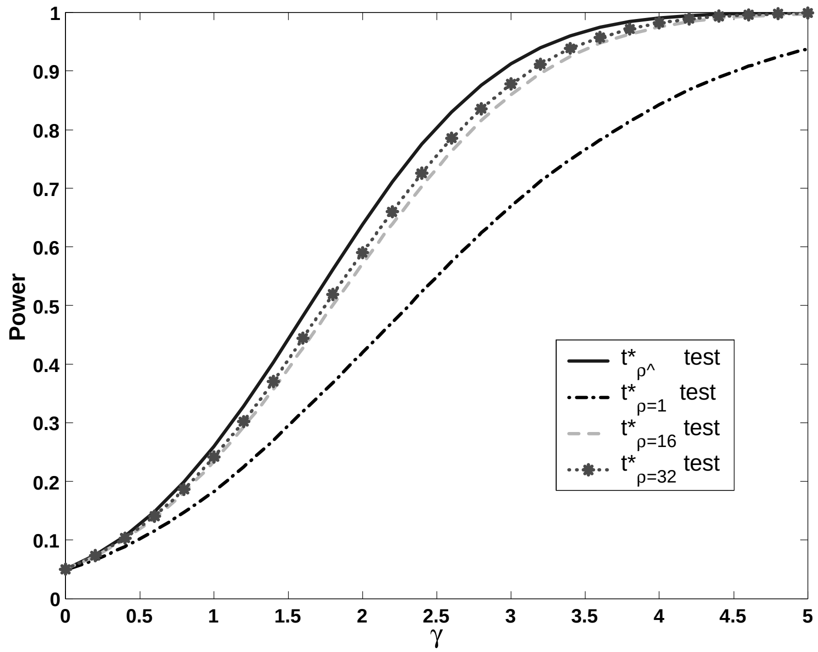

Figure 9: Asymptotic Local Power Function of the t^* Tests

truncated) t test, we use a simple location model

$$y_t = \mu + u_t,$$

where $u_t = a_1 u_{t-1} + a_2 u_{t-2} + e_t$, e_t are $iid(0,1)$. We consider the null hypothesis $H_0 : \mu = 0$ against the one-sided alternative $H_1 : \mu > 0$. We use the Parzen kernel to construct the conventional t-statistic, which is labeled t_{HAC}. In computing t_{HAC}, the bandwidth is chosen by the AR(1) plug-in approach as in Andrews (1991).

Table 6 presents the finite sample null rejection probabilities via simulation for $T = 50$ and 200. The simulation results are based on 50,000 replications. For the t^*_ρ and t^*_{new} tests, rejections were determined using the asymptotic 95% critical value based on the hyperbola formula (47). For the $t^*_{\hat{\rho}}$ and t_{HAC} tests, rejections were determined using the 95% critical value from the standard normal. The results for the t^*_ρ test with a fixed ρ are very similar to those of the test with sharp Bartlett kernels. First, in all cases, the size distortions of the t^*_ρ tests are less than those of the t_{HAC}-test. This is true even for large ρ. Second, the size distortion increases with ρ. But as T increases, the null rejection probabilities approach the nominal size for all cases. Simulation results (not reported here) show that with the increase of ρ, the size distortions of the t^*_ρ test constructed using steep Parzen kernels increase less dramatically than those using sharp Bartlett kernels. Third, when the errors follow an AR(1) process, the size distortion of all tests becomes larger as a_1 approaches unity. However, compared to sharp Bartlett kernels, the incremental size distortion is less (not reported here). The size distortion of the $t^*_{\hat{\rho}}$ test is close to or slightly less than that of the t_{HAC} test, which is expected, since we use the same asymptotic critical value (1.645) for both tests. Compared with the t^*_ρ test for $\rho = 1$, 16, and 32, the $t^*_{\widehat{\rho}}$ test has larger size distortion, especially when the error process is persistent. Using the adjusted critical values, the t^*_{new} test has significantly smaller size distortions than the $t^*_{\widehat{\rho}}$ test, especially in cases where the $t^*_{\widehat{\rho}}$ and t_{HAC} tests perform worse. In fact, the t^*_{new} test achieves the best size properties among all the tests considered except the t^*_1 test.

Table 6: Finite Sample Null Hypothesis Rejection Probabilities for a Location Model $y_t = \mu + u_t$ with $u_t = a_1 u_{t-1} + a_2 u_{t-2} + e_t$, $u_0 = u_{-1} = 0$ and $e_t \sim iid(0, 1)$

	a_1	a_2	t_{HAC}	$t^*_{\widehat{\rho}}$	t^*_{new}	$t^*_{\rho=1}$	$t^*_{\rho=16}$	$t^*_{\rho=32}$
$T = 50$	−0.500	0.000	0.054	0.047	0.042	0.050	0.050	0.054
	0.000	0.000	0.060	0.059	0.056	0.055	0.052	0.060
	0.300	0.000	0.078	0.077	0.069	0.058	0.056	0.066
	0.500	0.000	0.097	0.096	0.076	0.062	0.066	0.078
	0.700	0.000	0.127	0.127	0.082	0.069	0.086	0.110
	0.900	0.000	0.236	0.227	0.117	0.102	0.184	0.219
	0.950	0.000	0.310	0.291	0.155	0.136	0.257	0.288
	0.990	0.000	0.384	0.350	0.191	0.175	0.331	0.364
	1.500	−0.750	0.144	0.129	0.033	0.050	0.022	0.026
	1.900	−0.950	0.361	0.147	0.029	0.029	0.030	0.047
	0.800	0.100	0.238	0.234	0.131	0.106	0.196	0.230
$T = 200$								
	−0.500	0.000	0.046	0.048	0.048	0.059	0.054	0.057
	0.000	0.000	0.057	0.056	0.056	0.059	0.054	0.057
	0.300	0.000	0.068	0.069	0.067	0.061	0.055	0.057
	0.500	0.000	0.074	0.074	0.071	0.061	0.056	0.058
	0.700	0.000	0.086	0.086	0.078	0.062	0.059	0.063
	0.900	0.000	0.129	0.129	0.088	0.069	0.089	0.101
	0.950	0.000	0.175	0.174	0.099	0.084	0.131	0.154
	0.990	0.000	0.326	0.308	0.165	0.148	0.273	0.312
	1.500	−0.750	0.085	0.083	0.051	0.058	0.051	0.050
	1.900	−0.950	0.199	0.173	0.046	0.051	0.027	0.014
	0.800	0.100	0.134	0.133	0.097	0.071	0.094	0.107

There are always trade-offs between finite sample size and power. Fig. 10 shows

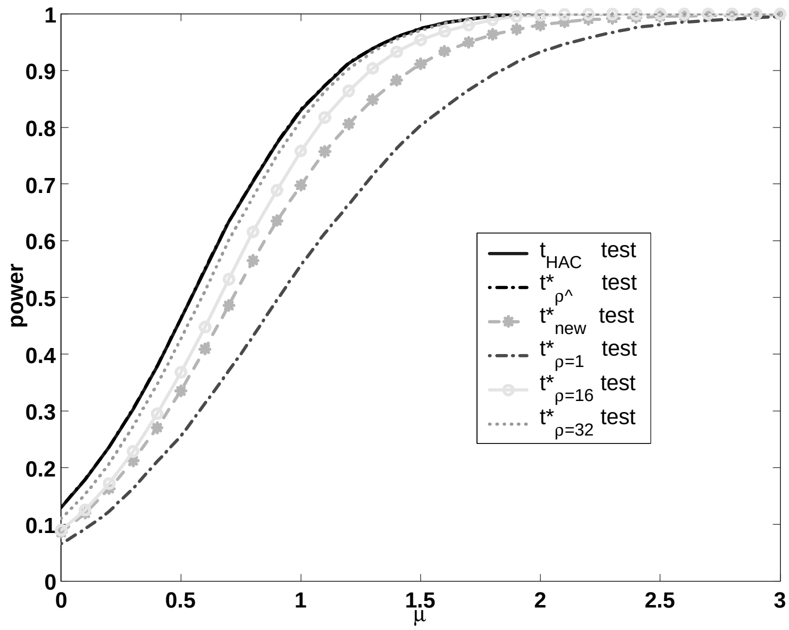

Figure 10: Finite Sample Power for Location Model: $y_t = \mu + u_t, u_t = 0.7u_{t-1} + e_t$ with $T = 50$

the finite sample power of these tests when $a_1 = 0.7$ without size correction. The typical pattern in the figure is that the power curves of t_{HAC} and $t^*_{\widehat{\rho}}$ are indistinguishable, and the power of the t^*_ρ test increases as ρ increases, just as asymptotic theory predicts. When $\rho \geq 32$, the power of the t^*_ρ test is very close to that of the conventional robust t-test using the Parzen kernel. The t^*_{new} test also 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^*_{new} test becomes closer to that of the t_{HAC} and $t^*_{\widehat{\rho}}$ tests. Fig. 10 also shows the size distortions of the different tests, which are in the descending order: $t_{HAC}, t^*_{\widehat{\rho}}, t^*_{\rho=32}, t^*_{\rho=16}, t^*_{new}$, and $t^*_{\rho=1}$. This pattern is found to be typical in cases where the AR coefficient is large but less than unity. Overall, the t^*_{new} test produces favorable results for both size and power in regression testing and is recommended for practical use.

All the tests considered can be combined with prewhitening procedures such as those in Andrews and Monahan (1992) and Lee and Phillips (1994). To save space, we do not report the simulation results for the prewhitening version for the tests. We remark that all the qualitative observations continue to apply but the size distortions are smaller in all cases.

5 Extensions and Conclusion

Exponentiating a mother kernel enables consistent kernel estimation without the use of lag truncation. When the exponent parameter is not too large, the absence of lag truncation influences the variability of the estimate because of the presence of autocovariances at long lags. As has been noted by Kiefer and Vogelsang (2002a & b) and Jansson (2002) and as confirmed in the simulations reported here, such effects can have the advantage of better reflecting finite sample behavior in test statistics that employ LRV/HAC estimates leading to some improvement in test size. When the exponent is passed to infinity with the sample size, the kernels produce consistent LRV/HAC and spectral density estimates, thereby ensuring that there is no loss in test power asymptotically. Similar ideas can, of course, be used in probability density estimation and in nonparametric regression.

One feature of interest in the asymptotic theory is that, unlike conventional kernel estimation where an optimal choice of quadratic kernel is possible in terms of MSE criteria, steep origin kernels are asymptotically MSE equivalent, so that choice of mother kernel does not matter asymptotically, although it may of course do so in finite samples. Another feature of the asymptotic theory of steep origin kernel estimation is that optimal convergence rates (that minimize an asymptotic MSE criterion) are faster for quadratic mother kernels than they are for the Bartlett kernel. The corresponding expansion rate for the exponent is $\rho = O\left(T^{8/5}\right)$ (leading to a convergence rate of $T^{2/5}$ for the kernel estimate $\hat{f}_{XX}$) so that ρ tends to infinity much faster than the sample size T. The reason for this fast expansion rate is that quadratic kernels have a flat shape at the origin and, since no bandwidth or lag truncation is being employed to control the effect of sample autocovariances at long lags, the fast rate of exponentiation ensures that the long lag sample autocovariances are sufficiently downweighted for a central limit theory to apply. The use of flat top kernels with bandwidth parameters and steep decay at long lags has recently attracted interest in the nonparametric literature (e.g., Politis and Romano, 1997) and it may be worthwhile pursuing these new ideas in conjunction with those of the present paper.

6 Appendix

Proof of Theorem 1. Part (a). Using $E\widehat{\gamma}_h = \gamma_h(1-\frac{|h|}{T})$, we have

$$\frac{T^2}{\rho}E(\widehat{f}_{XX}(0) - f_{XX}(0)) = \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})(1-\frac{|h|}{T})\gamma_h - \frac{T^2}{\rho}\frac{1}{2\pi}\sum_{-\infty}^{\infty}\gamma_h$$

$$= \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h - \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\frac{|h|}{T}\gamma_h - \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{|h|\geq T}\gamma_h$$

$$= \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h - \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\frac{|h|}{T}\gamma_h + O\left(\frac{1}{\rho}\right), \quad (48)$$

where the last line follows because $\left|\sum_{|h|\geq T}\gamma_h\right| \leq T^{-2}\sum_{|h|\geq T}\left|h^2\gamma_h\right|$. We now consider the first two terms in (48). The second term is bounded by

$$\frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\frac{|h|}{T}|\gamma_h| = \frac{T}{\rho}\frac{1}{2\pi}\sum_{h=-T+1}^{T-1}|h||\gamma_h| = O(\frac{T}{\rho}) = o(1) \tag{49}$$

using Assumptions 2 and 3. The first term in (48) can be written as

$$\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h = \sum_{h=-T/\log T}^{T/\log T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h + \sum_{T/\log T\leq|h|<T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h. \tag{50}$$

Noting that

$$\begin{aligned}\left|\frac{T^2}{\rho}\sum_{T/\log T\leq|h|<T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h\right| &\leq \frac{T^2}{\rho}\sum_{T/\log T\leq|h|<T}|\gamma_h| \\ &\leq \frac{\log^2 T}{\rho}\sum_{T/\log T\leq|h|<T}h^2|\gamma_h| = o\left(\frac{\log^2 T}{\rho}\right),\end{aligned}$$

and using Assumption 1(c), we obtain

$$\begin{aligned}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h &= \frac{T^2}{\rho}\sum_{h=-T/\log T}^{T/\log T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h + o\left(\frac{\log^2 T}{\rho}\right) \\ &= \sum_{h=-T/\log T}^{T/\log T}\left[\frac{k(\frac{h}{T})-1}{\rho h^2/T^2}\right]h^2\gamma_h + o(1) \\ &= -g\sum_{-\infty}^{\infty}h^2\gamma_h\left(1+o(1)\right).\end{aligned} \tag{51}$$

Combining the above results gives

$$\lim_{T\to\infty}\frac{T^2}{\rho}E(\widehat{f}_{XX}(0)-f_{XX}(0)) = -g\sum_{-\infty}^{\infty}h^2\gamma_h. \tag{52}$$

Part (b). We prove only the scalar case. The vector case follows from standard extensions. Note that

$$\widehat{f}_{XX}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}I_{XX}(\lambda)K_\rho(\lambda)\,d\lambda. \tag{53}$$

To find the asymptotic variance of $\widehat{f}_{XX}(0)$, we can work from the following standard formula (e.g., Priestley, 1981, eqn. 6.2.110 on p. 455) for the variance of a weighted periodogram estimate such as (53), viz.

$$\mathrm{Var}\left\{\widehat{f}_{XX}(0)\right\} = 2f_{XX}^2(0)\frac{1}{T}\sum_{h=-T+1}^{T-1}k_\rho(\frac{h}{T})^2\left[1+o(1)\right], \tag{54}$$

which follows directly from the covariance properties of the periodogram of a linear process (e.g., Priestley, 1981, p. 426).

To evaluate (54), we develop an asymptotic approximation of $T^{-1}\sum_{h=-T+1}^{T-1} k_\rho^2\left(\frac{h}{T}\right)$. Since $k_\rho(x)$ is differentiable by Assumption 1, it follows by Euler summation that the sum can be approximated by an integral as

$$\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho^2\left(\frac{h}{T}\right) = \int_{-1}^{1} k_\rho^2(x)\,dx\,(1+o(1)). \tag{55}$$

We use Laplace's method to approximate the above integral. It follows from Assumption 1(b) that for any $\delta > 0$, there exists $\zeta > 0$ such that $\log k(x) \leq -\zeta(\delta)$ for $|x| \geq \delta$. Therefore, the contribution of the intervals $\delta \leq |x| \leq 1$ satisfies

$$\int_{\delta\leq|x|\leq 1} k_\rho^2(x)\,dx = \int_{\delta\leq|x|\leq 1} \exp\{2\rho\log k(x)\}\,dx \leq \exp[-2(\rho-1)\zeta(\delta)]\int_{-1}^{1} k^2(x). \tag{56}$$

We now deal with the integral from $-\delta$ to δ. From Assumption 2(c),

$$k(x) = 1 - gx^2 + o\left(x^2\right), \text{ as } x \to 0 \text{ for some } g > 0,$$

we have $\log k(x) = -gx^2 + o(x^2)$. So, for any given $\varepsilon > 0$, we can determine $\delta > 0$ such that

$$\left|\log k(x) + gx^2\right| \leq \varepsilon x^2, \ \ |x| \leq \delta.$$

In consequence,

$$\int_{-\delta}^{\delta} \exp\left[-2\rho(g+\varepsilon)x^2\right]dx \leq \int_{-\delta}^{\delta} \exp 2\rho\log k(x)\,dx \leq \int_{-\delta}^{\delta} \exp\left[-2\rho(g-\varepsilon)x^2\right]dx.$$

But

$$\begin{aligned}\int_{-\delta}^{\delta} \exp\left[-2\rho(g+\varepsilon)x^2\right]dx &= \int_{-\infty}^{\infty} \exp\left[-2\rho(g+\varepsilon)x^2\right]dx + O\left(e^{-\rho\alpha}\right)\\ &= \frac{\sqrt{\pi}}{\sqrt{2\rho(g+\varepsilon)}} + O\left(e^{-\rho\alpha}\right),\end{aligned}$$

for some positive α that depends on δ but not ρ.

Similarly,

$$\int_{-\delta}^{\delta} \exp\left[-2\rho(g-\varepsilon)x^2\right]dx = \frac{\sqrt{\pi}}{\sqrt{2\rho(g-\varepsilon)}} + O\left(e^{-\rho\alpha}\right).$$

Therefore

$$\int_{-\delta}^{\delta} \exp 2\rho\log k(x)\,dx = \left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)). \tag{57}$$

Combining (56) and (57) yields

$$\int_{-1}^{1} k_\rho^2(x)\,dx = \left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)), \tag{58}$$

which completes the proof of part (b).

Part (c). Part (c) follows directly from parts (a) and (b). ∎

Proof of Lemma 2. Approximating the sum by an integral, we have

$$\begin{aligned} K_\rho(\lambda_s) &= T\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})e^{i\lambda_s h} = T\int_{-1}^{1} k_\rho(x)e^{i2\pi sx}dx\,(1+o(1)) \\ &= T\int_{-1}^{1} \exp\rho\left(\log k(x)+\log\cos(2\pi sx)\right)(1+o(1)). \end{aligned} \tag{59}$$

Using the Laplace approximation, we find that as $\rho\to\infty$, the contribution to the integral in (59), as in the proof of Theorem 1(b), comes mainly from a small region around $x=0$, say $(-\delta,\delta)$ for some arbitrarily small $\delta>0$. So there exists $\zeta(\delta)>0$ such that

$$\begin{aligned} K_\rho(\lambda_s) &= T\int_{-\delta}^{\delta}\exp\{\rho\left[\log k(x)\cos(2\pi sx)\right]\}(1+o(1)) + T\exp\left[-\rho\zeta(\delta)\right](1+o(1)) \\ &= T\int_{-\delta}^{\delta} e^{\rho\log[k(x)]+2\pi sxi}dx\,(1+o(1)) + T\exp\left[-\rho\zeta(\delta)\right](1+o(1)) \\ &= T\int_{-\delta}^{\delta} e^{-\rho gx^2+2\pi sxi}dx\,(1+o(1)) + T\exp\left[-\rho\zeta(\delta)\right](1+o(1)) \\ &= T\int_{-\infty}^{\infty} e^{-\rho gx^2+2\pi sxi}dx\,(1+o(1)) + T\exp\left[-\rho\zeta(\delta)\right](1+o(1)) \\ &= T\int_{-\infty}^{\infty} e^{-\rho g(x^2+2\pi sxi/\rho g-(\pi s)^2/(\rho g)^2)-(\pi s)^2/(\rho g)} + T\exp\left[-\rho\zeta(\delta)\right](1+o(1)) \\ &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\pi^2 s^2}{\rho g}\right)(1+o(1)). \end{aligned}$$

Hence,

$$K_\rho(\lambda_s) = \begin{cases} O\left(\frac{T}{\sqrt{\rho}}\right) & \text{for } s\le O(\sqrt{\rho}), \\ O\left(\frac{Te^{-\frac{\pi^2 s^2}{\rho g}}}{\sqrt{\rho}}\right) & \text{for } s> O(\sqrt{\rho}), \end{cases}$$

as desired. ∎

Proof of Theorem 3. We prove the results for the scalar case, the vector case follows without further complication. Since $\widehat{f}_{XX}(0)=\frac{1}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)I_{XX}(\lambda_s)$ and

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

we can write the scaled estimation error as

$$\begin{aligned}
& \rho^{1/4}\left\{\widehat{f}_{XX}(0)-f_{XX}\left(0\right)\right\} \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_{\rho}(\lambda_s)\left[I_{XX}(\lambda_s)-f_{XX}\left(0\right)\right] \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_{\rho}(\lambda_s)\left[I_{XX}(\lambda_s)-f_{XX}\left(\lambda_s\right)\right] \\
& +\frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_{\rho}(\lambda_s)\left[f_{XX}(\lambda_s)-f_{XX}\left(0\right)\right]. \qquad (60)
\end{aligned}$$

Using Lemma 2, we have

$$K_{\rho}(\lambda_s)=\frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\pi^2 s^2}{\rho g}\right)(1+o(1)), \quad s=0,1,...,[T/2]. \qquad (61)$$

By Assumption 2, $|f''_{XX}(\lambda_s)|\leq\frac{1}{2\pi}\sum_{-\infty}^{\infty}h^2|\gamma_h|$, so that

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

Hence, the second term of (60) can be bounded as follows:

$$\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_{\rho}(\lambda_s)\left[f_{XX}(\lambda_s)-f_{XX}\left(0\right)\right] \\
= & \frac{\rho^{1/4}}{T}2\sum_{s=0}^{[T/2]}K_{\rho}(\lambda_s)\left[f_{XX}(\lambda_s)-f_{XX}\left(0\right)\right]=O\left(\frac{\rho^{1/4}}{T}\sum_{s=0}^{[T/2]}|K_{\rho}(\lambda_s)|\lambda_s^2\right) \\
= & O\left(\frac{\rho^{1/4}}{T}\sum_{s=0}^{[T/2]}\frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\pi^2 s^2}{\rho g}\right)\lambda_s^2\right) \\
= & O\left(\rho^{-1/4}T^{-2}\int_0^{\infty}\exp\left(-\frac{\pi^2 s^2}{\rho g}\right)s^2 ds\right) \\
= & O(\rho^{-1/4}T^{-2})=o(1). \qquad (62)
\end{aligned}$$

Then, by (60) and (62), we have

$$\rho^{1/4}\left\{\widehat{f}_{XX}(0)-f_{XX}\left(0\right)\right\}=\frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_{\rho}(\lambda_s)\left(I_{XX}(\lambda_s)-f_{XX}(\lambda_s)\right)+o_p\left(1\right).$$

In view of Assumption 2, we have $X_t=C(L)\varepsilon_t=\sum_{j=0}^{\infty}C_j\varepsilon_{t-j}$. The operator $C(L)$ has a valid spectral BN decomposition (Phillips and Solo, 1992)

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

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

$$X_t = C(L)\varepsilon_t = C(e^{i\lambda})\varepsilon_t + e^{-i\lambda}\widetilde{\varepsilon}_{\lambda t-1} - \widetilde{\varepsilon}_{\lambda t}, \tag{63}$$

where

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

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

$$\begin{aligned} w(\lambda_s) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{it\lambda_s} \\ &= C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + \frac{1}{\sqrt{2\pi T}} (\widetilde{\varepsilon}_{\lambda_s 0} - e^{in\lambda_s} \widetilde{\varepsilon}_{\lambda_s n}) \\ &= C(e^{i\lambda_s}) w_\varepsilon(\lambda_s) + O_p(T^{-1/2}). \end{aligned} \tag{64}$$

Thus, using the fact that

$$\begin{aligned} \sum_{s=0}^{T-1} |K_\rho(\lambda_s)| &= \frac{2\sqrt{\pi}T}{\sqrt{\rho g}} \sum_{s=0}^{[T/2]} \exp\left(-\frac{\pi^2 s^2}{\rho g}\right)(1+o(1)) \\ &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \int_{-\infty}^{\infty} \exp\left(-\frac{\pi^2 s^2}{\rho g}\right) ds\,(1+o(1)) \\ &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \int_{-\infty}^{\infty} \exp\left(-\frac{s^2}{2\rho g/(2\pi^2)}\right) ds\,(1+o(1)) \\ &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \left(\frac{1}{\sqrt{2\pi\rho g/(2\pi^2)}}\right)^{-1} (1+o(1)) \\ &= O(T), \end{aligned} \tag{65}$$

we get

$$
\begin{aligned}
& \rho^{1/4}\left\{\widehat{f}_{XX}(0)-f_{XX}\left(0\right)\right\} \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)(I_{XX}(\lambda_s)-f_{XX}(\lambda_s))+o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)(w(\lambda_s)w(\lambda_s)^*-f_{XX}(\lambda_s))+o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)\{[C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2})] \\
& \times[C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2})]^*-f_{XX}(\lambda_s)\}+o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s)-\frac{1}{2\pi}\sigma^2)]+O_p\left(\frac{\rho^{1/4}}{T}T\frac{1}{T^{1/2}}\right)+o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s)-\frac{1}{2\pi}\sigma^2)]+o_p(1), \qquad (66)
\end{aligned}
$$

where we have used $\rho/T^2 \to 0$. The fourth equality follows because $K_\rho(\lambda_s)$ becomes progressively concentrated at the origin. It can be proved rigorously using the same steps as those after (**??**).

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

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

where

$$c_j = \frac{C^2(1)}{2\pi}\frac{\rho^{1/4}}{T^2}\sum_{s=0}^{T-1}(K_\rho(\lambda_s)\cos(j\lambda_s)).$$

Then we can write

$$
\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s)-\frac{1}{2\pi}\sigma^2)] \\
= & 2\sum_{t=1}^{T}m_t+\frac{\rho^{1/4}}{T}C^2(1)\sum_{s=0}^{T-1}K_\rho(\lambda_s)\frac{1}{2\pi}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2-\sigma^2\right) \\
= & 2\sum_{t=1}^{T}m_t+\frac{\rho^{1/4}}{T}C^2(1)\left|\sum_{s=0}^{T-1}K_\rho(\lambda_s)\right|O_p(\frac{1}{\sqrt{T}}) \\
= & 2\sum_{t=1}^{T}m_t+O_p\left(\frac{\rho^{1/4}}{T}T\frac{1}{\sqrt{T}}\right) \\
= & 2\sum_{t=1}^{T}m_t+o_p(1). \qquad (67)
\end{aligned}
$$

By the Fourier inversion formula, we have

$$c_j = \frac{C^2(1)}{2\pi} \frac{\rho^{1/4}}{T} k_\rho(\frac{j}{T}). \tag{68}$$

Hence

$$\sum_{j=1}^{T} c_j^2 = O\left(\frac{\rho^{1/2}}{T^2} \sum_{j=1}^{T} k_\rho^2(\frac{j}{T})\right) = O\left(\frac{\rho^{1/2}}{T^2} \left(\frac{\pi}{2\rho g}\right)^{1/2} T\right) = O\left(\frac{1}{T}\right). \tag{69}$$

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

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

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

$$\sum_{t=1}^{T} E(m_t^2|\mathcal{F}_{t-1}) - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2} \to_p 0, \tag{70}$$

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

$$\sum_{t=1}^{T} E(m_t^4) \to_p 0. \tag{71}$$

We now proceed to establish (70) and (71). The left hand side of (70) is

$$\left(\sigma^2 \sum_{t=2}^{T}\sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2 - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) + \sigma^2 \sum_{t=2}^{T}\sum_{r\neq j} \varepsilon_r \varepsilon_j c_{t-r} c_{t-j} := I_1 + I_2. \tag{72}$$

The first term, I_1, is

$$\sigma^2\left(\sum_{j=1}^{T-1}(\varepsilon_j^2 - \sigma^2)\sum_{s=1}^{T-j} c_s^2\right) + \left(\sigma^4 \sum_{t=1}^{T-1}\sum_{j=1}^{T-t} c_j^2 - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) := I_{11} + I_{12}. \tag{73}$$

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

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

using (69). Next, consider the second term of (73). We have

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

Here we have used the following result, obtained by means of the Laplace approximation:

$$
\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\frac{s}{T}) &= \int_0^\infty (1-x)k_\rho^2(x)dx(1+o(1)) \\
&= \int_0^\infty \exp\left\{-x-\frac{1}{2}x^2-\rho g x^2\right\}dx(1+o(1)) \\
&= \int_0^\infty \exp\left\{-x-(\rho g+\frac{1}{2})x^2\right\}dx(1+o(1)) \\
&= \frac{1}{2}\exp\left\{2(\rho g+\frac{1}{2})^{-1}\right\}\frac{\sqrt{2\pi}}{\sqrt{2\rho g+1}} \\
&= \frac{1}{2}\left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)).
\end{aligned}
\tag{74}
$$

We have therefore shown that

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

So the first term of (72) is $o_p(1)$.

Now consider the second term, I_2, of (72). I_2 has mean zero and variance

$$
\begin{aligned}
& O\left(2\sum_{p,q=2}^{T}\sum_{r\neq j}^{\min(p-1,q-1)}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right) \\
= \; & O\left(2\sum_{p=2}^{T}\sum_{r\neq j}^{p-1}c_{p-r}^2 c_{p-j}^2 + 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right).
\end{aligned}
\tag{75}
$$

In view of (69), we have

$$\sum_{p=2}^{T}\sum_{r\neq j}^{p-1} c_{p-r}^2 c_{p-j}^2 = O\left(T\left(\sum_{j=1}^{T} c_j^2\right)^2\right) = O\left(\frac{1}{T}\right). \tag{76}$$

For the second component in (75), we have, using (69) and the Cauchy inequality,

$$\begin{aligned}
&4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j}) \le 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1} c_{q-r}^2 \sum_{r=1}^{q-1} c_{p-r}^2 \\
\le\ & 4\sum_{i=1}^{T} c_i^2 \sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1} c_{p-r}^2 \le 4\left(\sum_{i=1}^{T} c_i^2\right)\left(\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1} c_r^2\right) \\
=\ & O\left(\frac{1}{T}\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1} c_r^2\right) = O\left(\frac{1}{T}\sum_{r=1}^{T-2} r(T-r-1)c_r^2\right) \\
=\ & O\left(\frac{\rho^{1/2}}{T^3}\sum_{r=1}^{T-2} r(T-r-1)k_\rho(\frac{r}{T})\right) = O\left(\rho^{1/2}\int_0^1 x(1-x)k_\rho(x)dx\right).
\end{aligned} \tag{77}$$

We now show that

$$\rho^{1/2}\int_0^1 x(1-x)k_\rho(x)dx = o(1). \tag{78}$$

To this end, we need the following result: if the function $p(x) = x(1-x)k_\rho(x)$ achieves its maximum at $x^*(\rho) \in (0,1)$, then $x^*(\rho) \to 0$ as $\rho \to \infty$. The result can be proved by contradiction. Suppose for any ρ, there exists an $\varepsilon > 0$ and $\rho_0 \ge \rho$ such that $x^*(\rho_0) \ge \varepsilon$. Since $x^*(\rho_0) \ge \varepsilon$, it follows from Assumption 1 that there exist a positive number $\zeta(\varepsilon)$ such that $k(x^*(\rho_0)) \le 1 - \zeta(\varepsilon)$. Therefore

$$p(x^*(\rho_0)) \le x^*(\rho_0)(1 - x^*(\rho_0))\left[1-\zeta(\varepsilon)\right]^{\rho_0} \le 4\left[1-\zeta(\varepsilon)\right]^{\rho_0}. \tag{79}$$

But for large ρ_0,

$$p(1/\rho_0) = \frac{1}{\rho_0}\left(1-\frac{1}{\rho_0}\right)\left(1-\frac{g}{\rho_0^2}\right)^{\rho_0}(1+o(1)) = \frac{1}{\rho_0}(1+o(1)). \tag{80}$$

Hence

$$p(1/\rho_0) > p(x^*(\rho_0)) \tag{81}$$

for large ρ_0. This contradicts with the fact that $x^*(\rho_0)$ is a maximizing point. So $\lim x^*(\rho)$ must be zero. We note, in passing, that we have effectively shown that $x^*(\rho)$ is of order $O(1/\rho)$. Since the function $p(x)$ is strictly concave in a neighborhood of zero, $x^*(\rho)$ is the unique maximizer for any fixed ρ.

Given that $p(x)$ has a unique maximizer x^*, we can apply Laplace's method to approximate the integral $\int_0^1 p(x)dx$. Let

$$\kappa(x^*) = \frac{k''(x^*)}{k(x^*)} - \frac{(k'(x^*))^2}{k^2(x^*)}, \tag{82}$$

then

$$
\begin{aligned}
& \int_0^1 x(1-x)k_\rho(x)dx \\
= & \int_0^1 \exp[\log(x)+\log(1-x)+\log k_\rho(x)]dx \\
= & x^*(1-x^*)k_\rho(x^*)\int_0^\infty \exp-\left(\frac{1}{2\left(x^*-1\right)^2}+\frac{1}{2(x^*)^2}-\frac{1}{2}\rho\kappa(x^*)\right)y^2dy(1+o(1)) \\
= & o\left(\frac{1}{\sqrt{\rho}}\right), \qquad (83)
\end{aligned}
$$

using $\lim_{\rho\to\infty} x^*=0$, $\lim_{\rho\to\infty}\kappa(x^*)=-2g$ and $k_\rho(x^*)=O(1)$ as $\rho\to\infty$.

Combining (76), (77) and (83) completes the proof of $I_2\to_p 0$. We have therefore established condition (70).

It remains to verify (71). Let A be some positive constant, then the left hand side of (71) is

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

using (69), which verifies (71) and the CLT.

With this construction, we therefore have

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

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

$$
\begin{aligned}
\rho^{1/4}\left\{\widehat{f}_{XX}(0)-f_{XX}(0)\right\} &= \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1}K(\lambda_s)(I_{XX}(\lambda_s)-f_{XX}(\lambda_s))+o_p(1) \\
&\to_d N\left(0,2\left(\frac{\pi}{2g}\right)^{1/2}f_{XX}^2(0)\right).
\end{aligned}
$$

■

Proof of Theorem 4. Part (a) follows from the same arguments as in the proof of Theorem 1(a). It remains to prove part (b) as part (c) can be easily proved using parts (a) and (b). To prove part (b), we write

$$\begin{aligned}\widehat{f}_{XX}(\omega) &= \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\widehat{\gamma}_h e^{-ih\omega} = \frac{1}{2\pi}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\frac{2\pi}{T}\sum_{s=0}^{T-1} I_{XX}(\lambda_s)\, e^{i(\lambda_s-\omega)h} \\ &= \sum_{s=0}^{T-1} I_{XX}(\lambda_s)\, e^{i(\lambda_s-\omega)h}\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T}) \\ &= \frac{1}{T}\sum_{s=0}^{T-1} K(\lambda_s-\omega)\, I_{XX}(\lambda_s).\end{aligned}$$

As before, the variance of $\widehat{f}_{XX}(\omega)$ can be calculated using a standard formula (e.g., Priestley, 1981, eqn. 6.2.110 on p. 455):

$$\begin{aligned}\mathrm{Var}\left\{\widehat{f}_{XX}(\omega)\right\} &= f^2_{XX}(\omega)\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})^2\,[1+o(1)], \\ &= f^2_{XX}(\omega)\left(\frac{\pi}{2\rho g}\right)^{1/2}[1+o(1)], \qquad (84)\end{aligned}$$

where the last line uses (58). This complete the proof of part (b).

The stated result for the vector case follows directly by standard extensions (e.g. Hannan, 1970, page 280). ■

Proof of Lemma 5. Approximating the sum by an integral, we have

$$\begin{aligned}K(\lambda_s-\omega) &= T\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})e^{i(\omega-\lambda_s)h} = T\int_{-1}^{1} k_\rho(x)e^{(\omega xT-2\pi sx)i}dx\,(1+o(1)) \\ &= T\int_{-1}^{1}\exp\rho\left(\log k(x)+\log\cos(\omega xT-2\pi sx)\right)(1+o(1)). \qquad (85)\end{aligned}$$

Proceeding as before, we approximate the integral using Laplace's method. For some small $\delta>0$, we have

$$\begin{aligned}K_\rho(\lambda_s) &= T\int_{-\delta}^{\delta}\exp\{\rho\log[k(x)\cos(\omega xT-2\pi sx)]\}(1+o(1)) \\ &= T\int_{-\delta}^{\delta}\exp\{\rho\log[k(x)]+(\omega T-2\pi s)\,xi\}\,dx\,(1+o(1)) \\ &= T\int_{-\delta}^{\delta}\exp\left\{-\rho g x^2+(\omega T-2\pi s)\,xi\right\}dx\,(1+o(1)) \\ &= T\int_{-\infty}^{\infty}\exp\left[-\rho g x^2+(\omega T-2\pi s)\,xi\right]dx\,(1+o(1)).\end{aligned}$$

Simple calculations give

$$\begin{aligned}
K_\rho(\lambda_s) &= T\exp\left[-(\omega T-2\pi s)^2/(4\rho g)\right] \\
&\quad\times\int_{-\infty}^{\infty}\exp -\rho g\left[x^2+(\omega T-2\pi s)\,xi/\rho g-((\omega T-2\pi s)^2/(2\rho g)^2\right] \\
&= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{(\omega T-2\pi s)^2}{4\rho g}\right)(1+o(1)) \\
&= \begin{cases} O\left(\frac{T}{\sqrt{\rho}}\right) & \text{for } |\omega T-2\pi s|\le O(\sqrt{\rho}), \\ O\left(\frac{T}{\sqrt{\rho}}\exp\left(-\frac{(\omega T-2\pi s)^2}{4\rho g}\right)\right) & \text{for } |\omega T-2\pi s|> O(\sqrt{\rho}), \end{cases}
\end{aligned}$$

and this completes the proof. ■

Proof of Theorem 6. As before, we consider the scalar case as the vector case can be proved by standard extensions. Note that

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

and $K(\lambda_s)$ is a real even periodic function of λ_s with periodicity 2π.

Without loss of generality, we assume that T is even. Let λ_J be the Fourier frequency that is closest to ω and

$$B_\omega=\{s: s=J-T/2+1, J-[T]/2,...,J,J+1,...,J+T/2\}. \tag{86}$$

Then the scaled estimation error can be written as

$$\begin{aligned}
&\rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}(\omega)\right\} \\
&= \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[I_{XX}(\lambda_s)-f_{XX}(\omega)\right] \\
&= \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[I_{XX}(\lambda_s)-f_{XX}(\lambda_s)\right] \\
&\quad+\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[f_{XX}(\lambda_s)-f_{XX}(\omega)\right].
\end{aligned} \tag{87}$$

By Assumption 2, $|f'_{XX}(\lambda_s)|\le\frac{1}{2\pi}\sum_{-\infty}^{\infty}|h|\,|C(h)|$, so that

$$|f_{XX}(\lambda_s)-f_{XX}(\omega)|\le\left(\frac{1}{2\pi}\sum_{-\infty}^{\infty}|h|\,|\Gamma(h)|\right)|\lambda_s-\omega|.$$

Hence, the second term of (87) is

$$\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho\left(\lambda_s-\omega\right)\left[f_{XX}(\lambda_s)-f_{XX}\left(\omega\right)\right] \\
= & \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho\left(\lambda_s-\omega\right)\left|\lambda_s-\omega\right| \\
= & O\left(\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}\frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\left(\omega-\lambda_s\right)^2T^2}{4\rho g}\right)\left|\lambda_s-\omega\right|\right) \\
= & O\left(\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}\frac{\sqrt{\pi}}{\sqrt{\rho g}}\exp\left(-\frac{\left(\omega T-2\pi s\right)^2}{4\rho g}\right)\left|2\pi s-\omega T\right|\right) \\
= & O\left(\frac{\rho^{1/4}}{T}\int_0^\infty \exp\left(-\frac{v^2}{4g}\frac{T^2}{\rho}\right)vdv\right) \\
= & O(\frac{\rho^{1/4}}{T})=o(1), \qquad (88)
\end{aligned}$$

where we have used Lemma 5.

Combining (87) and (88) leads to

$$\rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}\left(\omega\right)\right\}=\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho\left(\lambda_s-\omega\right)\left[I_{XX}(\lambda_s)-f_{XX}\left(\lambda_s\right)\right]+o\left(1\right) \qquad (89)$$

In view of (64), the frequency domain BN decomposition, we have

$$w(\lambda_s)=C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2}). \qquad (90)$$

Following the same steps in (65), we can show that

$$\sum_{s=0}^{T-1}\left|K_\rho(\lambda_s-\omega)\right|=O(T). \qquad (91)$$

Plugging (90) into (89) and using (65), we have

$$\begin{aligned}
& \rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}\left(\omega\right)\right\} \\
= & \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho\left(\lambda_s-\omega\right)\left\{\left[C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2})\right]\times\right. \\
& \left.\left[C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2})\right]^*-f_{XX}(\lambda_s)\right\}+o\left(1\right) \\
= & \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho\left(\lambda_s-\omega\right)\left[\left|C(e^{i\lambda_s})\right|^2 I_{\varepsilon\varepsilon}\left(\lambda_s\right)-f_{XX}\left(\lambda_s\right)\right]+O_p\left(\frac{\rho^{1/4}}{T}T\frac{1}{T^{1/2}}\right)+o\left(1\right)
\end{aligned}$$

$$= \frac{\rho^{1/4}}{T} \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) \left| C(e^{i\lambda_s}) \right|^2 \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi} \right] + o_p(1)$$

$$= \frac{\rho^{1/4}}{T} \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) \left| C(e^{iw}) \right|^2 \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi} \right]$$

$$+ \frac{\rho^{1/4}}{T} \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) \left(\left| C(e^{i\lambda_s}) \right|^2 - \left| C(e^{iw}) \right|^2 \right) \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi} \right] + o_p(1), \tag{92}$$

where we have used $\rho/T^2 \to 0$. But the second term in (92) is bounded by

$$\frac{\rho^{1/4}}{T} \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) |\lambda_s - \omega| = o_p(1) \tag{93}$$

by the smoothness of $C(e^{iw})$ and (88). Hence

$$\rho^{1/4} \left\{ \widehat{f}_{XX}(\omega) - f_{XX}(\omega) \right\} = \frac{\rho^{1/4}}{T} \left| C(e^{iw}) \right|^2 \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi} \right]. \tag{94}$$

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

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

where

$$c_j(\omega) = \frac{\left| C(e^{iw}) \right|^2}{2\pi} \frac{\rho^{1/4}}{T^2} \sum_{s \in B_\omega} (K_\rho(\lambda_s - \omega) \cos(j\lambda_s)).$$

Following the same steps as in (67), we can write

$$\frac{\rho^{1/4}}{T} \left| C(e^{iw}) \right|^2 \sum_{s \in B_\omega} K_\rho(\lambda_s - \omega) \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi} \right] = 2 \sum_{t=1}^{T} m_t + o_p(1). \tag{95}$$

Simple calculations show that

$$c_j(\omega) = \frac{\left| C(e^{iw}) \right|^2}{4\pi} \frac{\rho^{1/4}}{T} k_\rho(\frac{j}{T}) \cos \omega j. \tag{96}$$

Hence

$$\sum_{j=1}^{T} c_j^2(\omega) = O\left(\frac{\rho^{1/2}}{T^2} \sum_{j=1}^{T} k_\rho^2(\frac{j}{T}) \right) = O\left(\frac{1}{T} \right). \tag{97}$$

We proceed to show that

$$2 \sum_{t=1}^{T} m_t \to_d N\left(0, \frac{\sigma^4 \left| C(e^{iw}) \right|^4}{4\pi^2} \left(\frac{\pi}{2g} \right)^{1/2} \right) = N\left(0, f_{XX}^2(\omega) \left(\frac{\pi}{2g} \right)^{1/2} \right), \tag{98}$$

by verifying the following two sufficient conditions for a martingale CLT:

$$\sum_{t=1}^{T} E(m_t^2|\mathcal{F}_{t-1}) - \frac{\sigma^4 \left|C(e^{iw})\right|^4}{16\pi^2} \left(\frac{\pi}{2g}\right)^{1/2} \to_p 0, \tag{99}$$

and

$$\sum_{t=1}^{T} E(m_t^4) \to_p 0. \tag{100}$$

The left hand side of (99) is

$$\left(\sigma^2 \sum_{t=2}^{T} \sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2(\omega) - \frac{\sigma^4 C^4(1)}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) + \sigma^2 \sum_{t=2}^{T} \sum_{r\neq j} \varepsilon_r \varepsilon_j c_{t-r}(\omega) c_{t-j}(\omega) := \mathcal{I}_1 + \mathcal{I}_2. \tag{101}$$

The first term, $\mathcal{I}_1$, is

$$\sigma^2 \left(\sum_{j=1}^{T-1} (\varepsilon_j^2 - \sigma^2) \sum_{s=1}^{T-j} c_s^2(\omega)\right) + \left(\sigma^4 \sum_{t=1}^{T-1} \sum_{j=1}^{T-t} c_j^2(\omega) - \frac{\sigma^4 C^4(1)}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) := \mathcal{I}_{11} + \mathcal{I}_{12}. \tag{102}$$

The mean of $\mathcal{I}_{11}$ is zero and its variance is of order

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

using (97). Next, consider the second term of (102). We have

$$\begin{aligned}
\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} c_s^2(\omega) &= \frac{\left|C(e^{i\omega})\right|^4}{4\pi^2}\frac{\rho^{1/2}}{T^2}\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} k_\rho^2(\frac{s}{T})\cos^2\omega s \\
&= \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\frac{\rho^{1/2}}{T^2}\sum_{s=1}^{T-1}\sum_{j=1}^{T-s} k_\rho^2(\frac{s}{T})\cos^2\omega s \\
&= \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\frac{\rho^{1/2}}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\frac{s}{T})\cos^2\omega s \\
&= \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\rho^{1/2}\left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)) \qquad (103) \\
&= \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2} + o(1),
\end{aligned}$$

where (103) follows from the approximation

$$\frac{1}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\frac{s}{T})\cos^2\omega s = \left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)), \tag{104}$$

which can be proved using Laplace's method. To save space, the details of the proof are omitted.

The above derivations therefore demonstrate that

$$\mathcal{I}_1 = \sigma^2 \sum_{t=2}^{T} \sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2 (\omega) - \frac{\sigma^4 \left| C(e^{i\omega}) \right|^4}{16\pi^2} \left(\frac{\pi}{2g} \right)^{1/2} \to_p 0.$$

So the first term of (101) is $o_p(1)$. Following arguments similar to the proof of Theorem 3, we can show that $\mathcal{I}_2 \to_p 0$. In fact, since $c_j(\omega) \leq 1/2c_j$, all steps go through with no modifications. Similarly, condition (100) can be verified in the same way.

Combining (94), (95) and (98) yields

$$\frac{\rho^{1/4}}{T} \left| C(e^{iw}) \right|^2 \sum_{s \in B_\omega} K_\rho (\lambda_s - \omega) \left[I_{\varepsilon\varepsilon} (\lambda_s) - \frac{\sigma^2}{2\pi} \right]$$

$$\to_d N \left(0, \frac{\sigma^4 \left| C(e^{i\omega}) \right|^4}{4\pi^2} \left(\frac{\pi}{2g} \right)^{1/2} \right) = N \left(0, \left(\frac{\pi}{2g} \right)^{1/2} f_{XX}^2(\omega) \right).$$

From this, we obtain the limit theory for the spectral estimate at $\omega \neq 0, \pi$:

$$\rho^{1/4} \left\{ \widehat{f}_{XX}(\omega) - f_{XX}(\omega) \right\} \to_d N \left(0, \left(\frac{\pi}{2g} \right)^{1/2} f_{XX}^2(\omega) \right),$$

as desired. ■

Proof of Theorem 7. Note that

$$\widehat{f}_{XX}(\omega_i) = \frac{1}{T} \sum_{s=0}^{T-1} K (\lambda_s - \omega_i) I_{XX} (\lambda_s),$$

so

$$cov \left(\widehat{f}_{XX}(\omega_i), \widehat{f}_{XX}(\omega_j) \right) = \frac{1}{T^2} \sum_{\tau=0}^{T-1} \sum_{s=0}^{T-1} K (\lambda_s - \omega_i) K (\lambda_\tau - \omega_j) cov (I_{XX} (\lambda_s), I_{XX} (\lambda_\tau)). \tag{105}$$

Under Assumption 2, we have

$$\begin{aligned} (i) \mathrm{Var} (I_{XX}(\lambda_s) - f_{XX}(\lambda_s)) &= 4\pi^2 \delta_{0,\lambda_s} f_{XX}^2(\lambda_s)(1 + O(T^{-1/2})), \quad (106) \\ (ii) Cov (I_{XX}(\lambda_s), I_{XX}(\lambda_\tau)) &= O (f_{XX}(\lambda_s) f_{XX}(\lambda_\tau)/T), s \neq \tau, \end{aligned}$$

where $\delta_{0,\lambda_s} = 1 + 1_{\{\lambda_s = 0, \pi\}}$, and $O(\cdot)$ holds uniformly in λ_s and λ_τ (see 6.2.37 of Priestley (1981)). Therefore

$$\begin{aligned} cov \left(\widehat{f}_{XX}(\omega_i), \widehat{f}_{XX}(\omega_j) \right) &= \left(T^{-2} \sum_{s=0}^{T-1} \delta_s K (\lambda_s - \omega_i) K (\lambda_s - \omega_j) \right) \left((1 + O(T^{-1/2})) \right) \\ &\quad + O \left(T^{-3} \sum_{s \neq \tau} \left| K (\lambda_s - \omega_i) K (\lambda_\tau - \omega_j) \right| \right). \quad (107) \end{aligned}$$

The second term in (107) is

$$O\left\{T^{-3}\left(\sum_{s=0}^{T-1}\left|K\left(\lambda_s-\omega_i\right)\right|\right)\left(\sum_{\tau=0}^{T-1}\left|K\left(\lambda_\tau-\omega_j\right)\right|\right)\right\}=O(1/T),$$

using $\sum_{\tau=0}^{T-1}\left|K\left(\lambda_\tau-\omega_j\right)\right|=O(T)$. The first term in (107) is bounded by

$$\begin{aligned}
& O\left(\frac{1}{\rho}\sum_{s=0}^{T-1}\exp\left(-\frac{\left(\omega_i-\lambda_s\right)^2T^2}{4\rho g}-\frac{\left(\omega_j-\lambda_s\right)^2T^2}{4\rho g}\right)\right)\\
= & O\left(\frac{T}{\rho}\int_0^{2\pi}\exp\left(-\frac{T^2}{4\rho g}\left(\left(\omega_i-x\right)^2+\left(\omega_j-x\right)^2\right)\right)dx\right)\\
= & O\left\{\frac{T}{\rho}\int_0^{2\pi}\exp\left[-\frac{T^2}{4\rho g}\left(2\left(x-\frac{\omega_i+\omega_j}{2}\right)^2+\frac{\left(\omega_i-\omega_j\right)^2}{2}\right)\right]dx\right\}\\
= & O\left\{\frac{T}{\rho}\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\frac{\sqrt{\rho}}{T}\right\}\\
= & O\left\{\frac{1}{\sqrt{\rho}}\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\right\}. \qquad (108)
\end{aligned}$$

Therefore

$$\begin{aligned}
\rho^{1/2}cov\left(\widehat{f}_{XX}(\omega_i),\widehat{f}_{XX}(\omega_j)\right) &= O\left(\rho^{1/2}/T\right)+O\left(\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\right)\\
&= o(1) \qquad (109)
\end{aligned}$$

using $\rho/T^2\to 0$. ■

7 Notation

LRV	Long Run Variance	K_{mm}	$m^2\times m^2$ commutation matrix
MSE	Mean Squared Error	$\otimes$	Kronecker product
HAC	Heteroskedastic and autocorrelation consistent	$vec(A)$	vectorization by columns
$\to_d$	weak convergence	$[\cdot]$	integer part
$o_p(1)$	tends to zero in probability	$\text{tr}\{A\}$	trace of A
$\mathbb{Z}^+$	set of positive integers	$\mathbb{R}$	$(-\infty,\infty)$
$\mathbb{R}^+$	$(0,\infty)$	$\|\|A\|\|$	Euclidian norm of A

References

[1] Anderson, T. (1971):*The Statistical Analysis of Time Series,* Wiley, New York.

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

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

[4] De Bruijn, N. G. (1982): *Asymptotic Methods in Analysis,* Dover Publications.

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

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

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

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

[9] Kiefer, N. M. and T. J. Vogelsang (2002b): "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica,* 70, 2093–2095.

[10] Lee, C. C. and Peter C. B. Phillips (1994). "An ARMA-prewhitened Long Run Variance Estimator" Yale University, mimeographed (korora.econ.yale.edu/phillips/papers/prewhite.pdf).

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

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

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

[14] Phillips, P. C. B., Y. Sun and S. Jin (2003): "Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," Cowles Foundation Discussion Paper No. 1407 (http://cowles.econ.yale.edu/P/au/d_phillips.htm).

[15] Politis, D. N. and J. P. Romano (1998): "Multivariate density estimation with general flat-top kernels of infinite order". *Journal of Multivariate Analysis*, 68, 1-25.

[16] Politis, D. N. and J. P. Romano (1995): "Bias corrected nonparametric spectral density estimator" *Journal of Time Series Analysis,* 16, 67-103

[17] Priestley, M. B. (1962): "Basic considerations in the estimation of spectra," *Technometrics*, 4, 551-564.

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