Spectral Density Estimation and Robust Hypothesis Testing 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
Guanghua School of Management
Peking University

August 2004

*The authors thank Graham Elliot, Jim Hamilton, Petra Todd, Hal White and anonymous referees for helpful comments. Phillips thanks the NSF for research support under Grant No. SES 00-92509 and SES 04-142254.

ABSTRACT

A new class of kernel estimates is proposed for long run variance and spectral density 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. Depending on whether the exponent is allowed to grow with the sample size, we establish different asymptotic approximations to the sampling distribution of the proposed estimator. When the exponent is passed to infinity with the sample size, the new estimator is consistent and shown to be asymptotically normal. It is 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. When the exponent is fixed, the new estimator is inconsistent and has a nonstandard limiting distribution. It is shown via Monte Carlo experiments that, when the chosen exponent is small in practical applications, the nonstandard limit theory provides better approximations to the finite sampling distributions of the spectral density estimator and the associated test statistic in regression settings.

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 improve the power of linear hypothesis tests 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 quadratic spectral (QS) 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 or assumed to be fixed as the sample size increases.

Accordingly, the paper first develops an asymptotic theory for this new class of steep origin kernel estimates, assuming that the exponent (ρ) goes to infinity at an appropriate rate with the sample size. For convenience, we call this type of asymptotics "large-ρ asymptotics". The paper establishes the consistency and asymptotic normality, and gives formulae for asymptotic bias, variance and mean squared error (MSE) of the new kernel estimator. 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 density estimation at frequencies $\omega \neq 0$.

The paper next considers the "fixed-ρ" asymptotics in which ρ is assumed to be fixed as the sample size increases. Both the LRV estimation and spectral density estimation at nonzero frequencies are considered. Under the fixed-ρ asymptotics, the LRV and spectral density estimators are inconsistent and converge to nonstandard distributions. Statistical inference can be made in a similar way to that under the large-ρ asymptotics. Since no rate condition is imposed on the exponent and the limiting distribution reflects the kernel used, the fixed-ρ asymptotics may be better able to approximate the finite sample distribution than the large-ρ asymptotics when ρ is not large in practice.

Finally, the paper conducts three Monte Carlo experiments to examine the finite sample properties of the proposed spectral density estimators and associated tests. In the first experiment, we compare the root mean squared errors (RMSEs) of different kernel estimators using data-driven exponents or bandwidths. Simulation results show that the steep kernel estimators have very competitive RMSE-performance relative to the conventional QS estimator in an overall sense for the sample sizes and frequencies considered. In particular, for spectral density estimation at nonzero frequencies, the steep kernel estimators outperform the conventional QS estimator when there is a large peak at the target frequency. In the second experiment, we compare the finite sample coverage and length of different 95% confidence intervals. We find that confidence intervals based on the fixed-ρ asymptotics have the best finite sample performance when both the length and coverage error are taken into account. In the last experiment, we compare the size and power of robust regression tests using steep kernels. We propose a new $t-$ test which produces favorable results for both size and power in regression testing and this test is recommended for practical use.

The present contribution is related to recent work by Kiefer and Vogelsang (2003) and Hashimzade and Vogelsang (2004). These authors consider LRV and spectral density estimation using traditional kernels when the bandwidth (M) is set proportional to the sample size (T), i.e. $M = bT$ for some $b \in (0, 1)$. Their approach is equivalent to contracting traditional kernels $k(\cdot)$ to get $k_b(x) = k(x/b)$ and using the contracted kernels $k_b(\cdot)$ in the LRV and spectral density estimation without truncation. Both contracted and exponentiated kernels are designed to improve the power of existing robust regression tests with truncation. The associated estimators and tests share many properties. For example, the size distortion and power of the new robust regression tests increase as ρ increases or b decreases. Nevertheless, it is difficult to characterize the exact relationship between these two types of strategies. In the special cases when exponential type kernels such as $k(x) = \exp(-|x|)$ and $k(x) = \exp(-|x|^2)$ are used, these two strategies lead to identical estimators and statistical tests when ρ and b are appropriately chosen. Exponential kernels of this type have not been used before in LRV estimation and appear in spectral density

estimation only in the Abel estimate (c.f. Hannan, 1970, p. 279).

The rest of the paper is organized as follows. Section 2 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. This section assumes that the exponent goes to infinity as the sample size increases. Section 3 provides a similar analysis for the corresponding spectral density estimates at non-zero frequencies. Section 4 assumes that the exponent is fixed and develops alternative asymptotic approximations to the finite sample distributions of the LRV estimator and spectral density estimator. Section 5 reports some simulation evidence on the finite sample performance of these estimates and associated tests. Conclusions are given in Section 6. Proofs and other technical material are included in the Appendix (Section 7) and a glossary of notation is given in Section 8.

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

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

where

$$\widehat{\gamma}_h = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-h}\left(X_{t+h} - \bar{X}\right)\left(X_t - \bar{X}\right)' & \text{for } h \geq 0 \\ \frac{1}{T}\sum_{t=-h+1}^{T}\left(X_{t+h} - \bar{X}\right)\left(X_t - \bar{X}\right)' & \text{for } h < 0 \end{cases} \tag{3}$$

$\bar{X} = 1/T\sum_{t=1}^T X_t$, $k_\rho(x)$ is equal to $k(x)$ raised to some positive integer 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).

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 quadratic spectral (QS) kernels:

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

Quadratic Spectral
$$k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

For the Parzen kernel, $g = 6$. For the quadratic spectral kernel, $g = 18\pi^2/125$. The Parzen kernel has been used in the literature concerning long run variance estimation. The quadratic spectral (QS) kernel has some optimality properties in conventional LRV/HAC estimation. Since the bandwidth is set equal to the sample size, we effectively restrict the domain of the QS kernel to be $[-1,1]$, over which the kernel is positive and may be exponentiated as in (4).

The exponentiated 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.

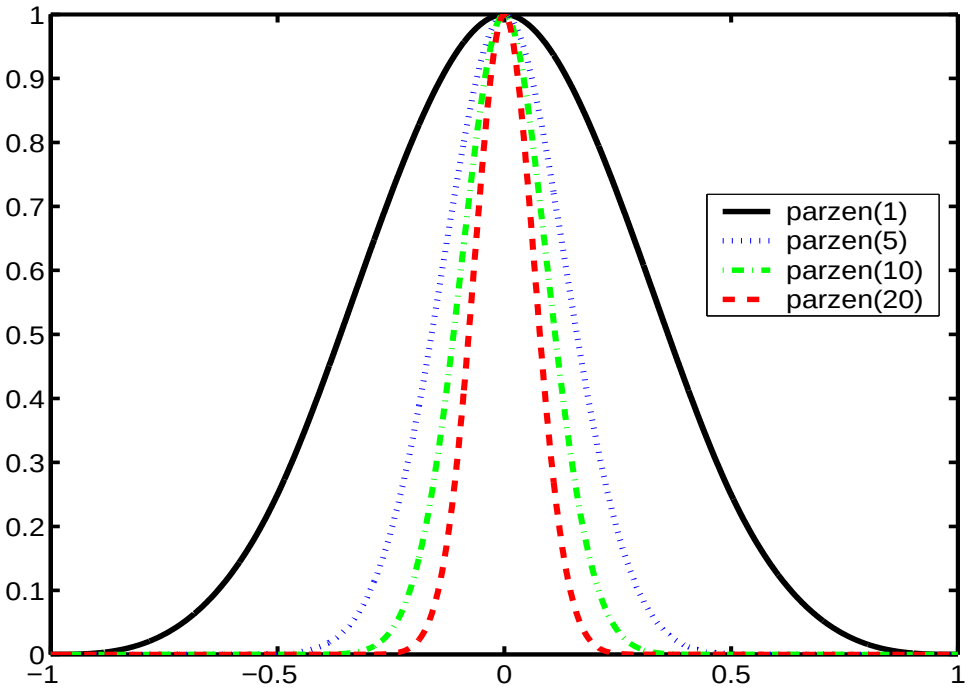

Figure 1: Steep origin kernels with the Parzen kernel as the mother kernel

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 m-vector stationary linear process with mean μ*

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

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

Assumption 3: *$T^6/\rho^5 + \rho/T^2 \to 0$ as $T \to \infty$ and $\rho \to \infty$.*

Assumption 2 is convenient and includes many time series of interest in applications, although condition (7) is stronger than necessary in establishing results for the

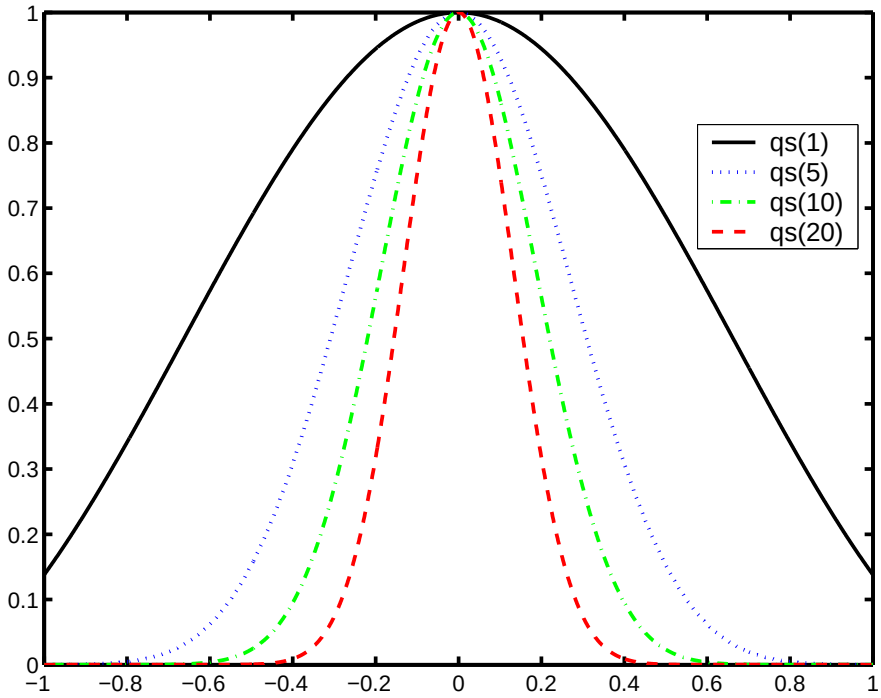

Figure 2: Steep origin kernels with the QS kernel as the mother kernel

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 $T^6/\rho^5 \to 0$, 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 by the authors (not reported here) show that the form involving the usual biased covariance estimator works better in practice. The bounds in Assumption 3 ensure that the asymptotic bias diminishes as T goes to ∞. They are also used in the proof of Lemma 2 in Section 2.4.

Assumption 3 holds when $\rho = aT^b$ for some $a > 0$ and $6/5 < b < 2$. Note

that the expansion rate for ρ implied by Assumption 3 is very different from the rate condition $\frac{T}{\rho^2} + \frac{\rho \log T}{T} \to 0$ 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 $T^6/\rho^5 \to 0$ 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

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

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) and I_{m^2} be the $m^2 \times m^2$ identity matrix. 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) $\lim_{T\to\infty} \left(T^2/\rho\right) \left(E\widehat{f}_{XX}(0) - f_{XX}(0) \right) = -g f^{(2)}_{XX}(0)$.

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

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

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

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 (65) 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 $T\to\infty$. When ρ is large, 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, for large ρ the absolute asymptotic bias increases with the curvature of the true spectral density at the origin. It should be mentioned that when ρ is not very large and the data are demeaned, the bias/variance trade-off becomes more complicated. Theorem 8 in Section 4 shows that when ρ is small, the asymptotic bias decreases and the asymptotic variance may increase for certain kernels as ρ increases.

Observe that when $\rho^5/T^8 \to \vartheta \in (0,\infty)$, $\mathrm{MSE}(\widehat{f}_{XX}(0),W) = O\left(T^{-4/5}\right)$. 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 ρ_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 ρ_T,2 such that

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

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 ρ_T,1 and ρ_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 QS kernel is,

$$k_{QS}^{o}(x)=\left[\frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5}-\cos(6\pi/5x)\right)\right]^{125/(3\pi^2)}1\{|x|\leq 1\}. \tag{16}$$

Then, for any ρ satisfying Assumption 3, $(k_{PR}(x))^{\rho}$ and $(k_{QS}^{o}(x))^{\rho}$ 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. Some comparisons with the asymptotic MSE of conventional (bandwidth driven) quadratic kernels will be given in the following section.

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 and QS kernels are positive definite and the resulting LRV estimate is guaranteed to be nonnegative. This property is certainly desirable and is not shared by kernels which are 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

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

leading to

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

So the optimal ρ is

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

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

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

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

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

and the resulting RMSE is

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

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

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

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 $\text{RMSE}^*_{sharp}/\text{RMSE}^*_{steep}$ 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 and and steep kernels for AR (1) processes

	Sharp Bartlett Kernel				Steep Parzen 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

When $\rho = \rho^*$, direct calculation shows that the MSE satisfies:

$$\lim_{T\to\infty} T^{4/5}\text{MSE}(\widehat{f}_{XX}(0), W) = \frac{5}{4}\pi^{\frac{2}{5}}\kappa$$

where

$$\kappa = \{\text{tr } \{W(I + K_{mm})f_{XX}(0) \otimes f_{XX}(0)\}\}^{4/5} \left\{\text{vec}\left(f_{XX}^{(2)}(0)\right)' W \text{vec}\left(f_{XX}^{(2)}(0)\right)\right\}^{1/5}.$$

As expected, the asymptotic MSE does not depend on the kernel used. Using Proposition 1 in Andrews (1991), we can show that for the conventional kernel estimators $\tilde{f}_{XX}(0)$ with MSE-optimal bandwidth, the MSE satisfies:

$$\lim_{T\to\infty} T^{4/5}\text{MSE}(\tilde{f}_{XX}(0), W) = \frac{5}{4}(2g)^{\frac{2}{5}} \left\{\int_{-\infty}^{\infty} k^2(x)dx\right\}^{4/5} \kappa.$$

Therefore, the asymptotic relative efficiency (ARE) of these two types of estimators is

$$\text{ARE} = \lim_{T\to\infty} \text{MSE}(\widehat{f}_{XX}(0), W)\left\{\text{MSE}(\tilde{f}_{XX}(0), W)\right\}^{-1} = \left(\frac{\pi}{2g}\right)^{\frac{2}{5}} \left\{\int_{-\infty}^{\infty} k^2(x)dx\right\}^{-\frac{4}{5}}.$$

For the QS kernel, $\int_{-\infty}^{\infty} k^2(x)dx = 1$, $g = 18\pi^2/125$. So ARE = 1.040 8. For the Parzen kernel, $\int_{-\infty}^{\infty} k^2(x)dx = 0.539285$, $g = 6$. So ARE = 0.958 81. Therefore, the asymptotic efficiency of the new kernel estimator lies between that of the traditional Parzen-based and QS-based estimators. Clearly, the asymptotic efficiency of the steep kernel estimator is very close to that of the traditional QS estimator and has the advantage that its asymptotic MSE is independent of the kernel employed in the estimator's construction.

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.

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

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

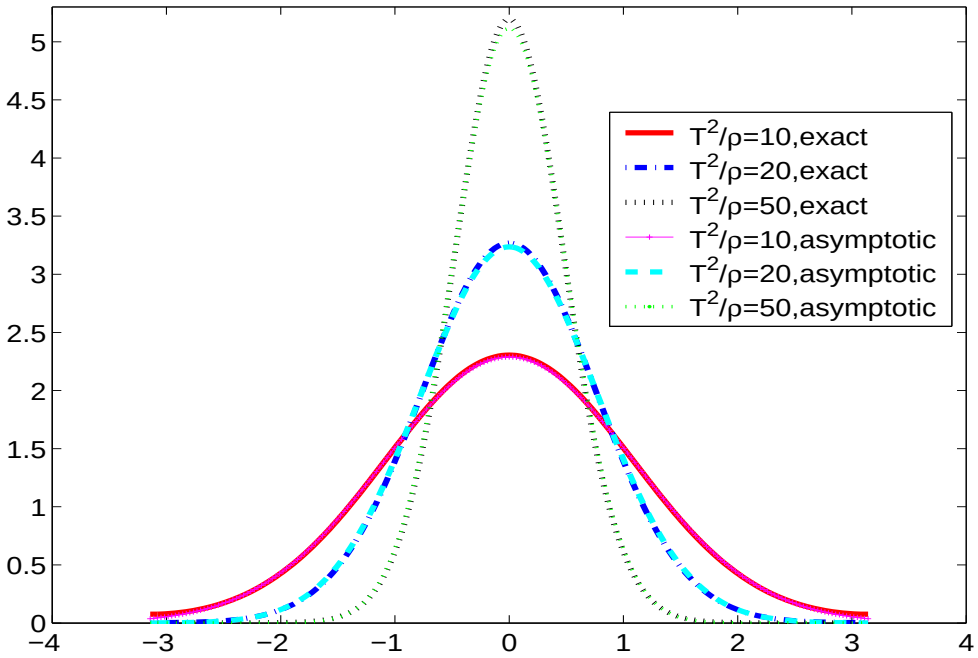

Figure 3: Spectral windows of steep Parzen kernels

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*

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

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

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

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

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

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

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 $f^{(2)}_{XX}(\omega)$ 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 $O\left(\rho/T^2\right)$.

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,

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

where

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

Thus, an approximate $100(1-\alpha)\%$ confidence interval (CI) for $f_{XX}(\omega)$ is

$$\begin{cases} \widehat{f}_{XX}(\omega)\left[\frac{1}{1+(4\pi)^{1/4}(2\rho g)^{-1/4} c v(\alpha/2)}, \frac{1}{1-(4\pi)^{1/4}(2\rho g)^{-1/4} c v(\alpha/2)}\right] & \text{for } \omega = 0, \pi \\ \widehat{f}_{XX}(\omega)\left[\frac{1}{1+(\pi)^{1/4}(2\rho g)^{-1/4} c v(\alpha/2)}, \frac{1}{1-(\pi)^{1/4}(2\rho g)^{-1/4} c v(\alpha/2)}\right] & \text{for } \omega \neq 0, \pi \end{cases} \tag{28}$$

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

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* $\omega_i \neq \omega_j$

$$\lim_{T\to\infty} \rho^{1/2} \text{cov}\left(\text{vec}(\widehat{f}_{XX}(\omega_i), \text{vec}(\widehat{f}_{XX}(\omega_j)\right) = 0. \tag{29}$$

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 the 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 An Alternative Asymptotic Approximation

In previous sections, we assumed that ρ approaches infinity as the sample size increases. This specification embeds the spectral density estimates in a triangular array. Under this large-ρ specification, the spectral density estimator is consistent and asymptotically normal. However, the required rate condition $T^6/\rho^5 + \rho/T^2 \to 0$ may not be compatible with (ρ, T)-combinations that work well in practice, especially when ρ is small. Also, the limit distribution theory, being independent of the kernel employed, does not reflect certain aspects of the finite sample behavior of the spectral density estimator. To alleviate these problems, we consider an alternative limit theory called fixed-ρ asymptotics in which ρ is fixed as T goes to infinity. Simulations reveal that fixed-ρ asymptotics generally provide better approximations to the finite sample distribution and these will be reported in the following section.

Define the partial sum discrete Fourier transform process

$$S_\omega(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (X_t - \mu) e^{-i\omega t}.$$

To derive the fixed-ρ asymptotics, we maintain the following assumptions:

Assumption 4. *$k(x) : \mathbb{R} \to [0, 1]$ is twice continuously differentiable.*

Assumption 5. *$S_\omega(r)$ satisfies a functional central limit theorem: $S_\omega(r) \Rightarrow \Lambda_\omega W_\omega(r)$ with*

$$W_\omega(r) = \begin{cases} W_\omega(r) & \text{for } \omega = 0 \text{ or } \pi, \\ W_{\omega R}(r) + iW_{\omega I}(r) & \text{for } \omega \neq 0,\ \pi, \end{cases}$$

and

$$\Lambda_\omega \Lambda_\omega' = \begin{cases} 2\pi f_{XX}(\omega) & \text{for } \omega = 0 \text{ or } \pi, \\ \pi f_{XX}(\omega) & \text{for } \omega \neq 0,\ \pi, \end{cases}$$

where $W_0(r)$, $W_\pi(r)$, $W_R(r)$ *and* $W_I(r)$ *are standard vector Brownian motions and* $W_{\omega R}(r)$ *is independent of* $W_{\omega I}(r)$.

The limit distribution theory of $\widehat{f}_{XX}(\omega)$ under fixed-ρ asymptotics is characterized in the following result.

Theorem 8 *Let Assumptions 4 and 5 hold, then for a fixed* ρ *as* T *goes to infinity:*
(a) $\widehat{f}_{XX}(\omega) \to_d (2\pi)^{-1}\Lambda_\omega \Xi_\omega \Lambda_\omega'$ *where*

$$\Xi_\omega = \begin{cases} \int_0^1 \int_0^1 k_\rho^*(t,\tau) dW_0(t) dW_0'(\tau) & \text{for } \omega = 0 \\ \int_0^1 \int_0^1 k_\rho(t-\tau) dW_\omega(t) dW_\omega'(\tau) & \text{for } \omega \neq 0, \end{cases}$$

and

$$k_\rho^*(t,\tau) = k_\rho(t-\tau) - \int_0^1 k_\rho(t-r)dr - \int_0^1 k_\rho(s-\tau)ds + \int_0^1 k_\rho(r-s)drds.$$

(b) The mean of the limiting distribution is:

$$E(2\pi)^{-1}\Lambda_\omega \Xi_\omega \Lambda_\omega' = \mu_\omega f_{XX}(\omega)$$

where $\mu_\omega = \left(1 - \int_0^1 \int_0^1 k_\rho(r-s)drds\right) 1_{\{\omega=0\}} + 1_{\{\omega \neq 0\}}$.
(c) The variance of the limiting distribution is:

$$var\left(vec\left((2\pi)^{-1}\Lambda_\omega \Xi_\omega \Lambda_\omega'\right)\right) = \begin{cases} \nu_\omega (I_{m^2} + K_{mm}) \left(f_{XX}(\omega) \otimes f_{XX}(\omega)\right), & \text{for } \omega = 0, \pi \\ \nu_\omega \left(f_{XX}(\omega) \otimes f_{XX}(\omega)\right), & \text{for } \omega \neq 0, \pi \end{cases}$$

where $\nu_\omega = \int_0^1 \int_0^1 \left[k_\rho(r-s)\right]^2 drds 1_{\{\omega \neq 0\}} + \int_0^1 \int_0^1 \left[k_\rho^*(r,s)\right]^2 drds 1_{\{\omega=0\}}$.

Theorem 8 shows that under fixed-ρ asymptotics, the spectral density estimator converges weakly to a random variable. More specifically,

$$\rho^{1/4}\Lambda_\omega^{-1}\left[\widehat{f}_{XX}(\omega) - \mu_\omega f_{XX}(\omega)\right]\left(\Lambda_\omega'\right)^{-1} \to_d (2\pi)^{-1}\rho^{1/4}\left(\Xi_\omega - E\Xi_\omega\right). \tag{30}$$

Like the finite sample distribution, the new limit distribution is random and depends on the kernel. In contrast, the large-ρ limit theory gives a consistent estimate and the limiting normal distribution

$$\rho^{1/4}\Lambda_\omega^{-1}\left(\widehat{f}_{XX}(\omega) - f_{XX}(\omega)\right)\left(\Lambda_\omega'\right)^{-1} \to_d N\left(0, \left(\frac{\pi}{2g}\right)^{1/2}\left(I_{m^2} + 1_{\omega \in \{0,\pi\}} K_{mm}\right)\right), \tag{31}$$

which is unaffected by the kernel. Thus, fixed-ρ limit theory retains features of the finite sample distribution that are lost as $\rho \to \infty$ and, in doing so, are suggestive that this asymptotic theory may better capture finite sample behavior than large-ρ asymptotics which rely on central limit arguments.

The statistics given in (30) and (31) are centred and scaled in a similar way and, as mentioned below, the limit distributions may be related by using sequential limit arguments as $T \to \infty$ followed by $\rho \to \infty$. Accordingly, inferences based on fixed-ρ asymptotic theory can be conducted in the same way as those based on the large-ρ asymptotics. For example, when $m = 1$ and ρ is fixed, (30) implies that

$$\left(\frac{\widehat{f}_{XX}(\omega)}{f_{XX}(\omega)} - \mu_\omega \right) \to_d \begin{cases} (2\nu_\omega)^{1/2} \Xi^*_\omega & \text{for } \omega = 0, \pi \\ (v_\omega)^{1/2} \Xi^*_\omega & \text{for } \omega \neq 0, \pi \end{cases}, \tag{32}$$

where $\Xi^*_\omega = \Xi^*_\omega(\rho) = (\Xi_\omega - E\Xi_\omega) / \sqrt{var(\Xi_\omega)}$. Thus, an approximate $100(1-\alpha)\%$ confidence interval for $f_{XX}(\omega)$ is given by

$$\begin{cases} \widehat{f}_{XX}(\omega) \left[\frac{1}{\mu_\omega + (2\nu_\omega)^{1/2} cv_\omega(\alpha/2)}, \frac{1}{\mu_\omega + (2\nu_\omega)^{1/2} cv_\omega(1-\alpha/2)} \right] & \text{for } \omega = 0, \pi \\ \widehat{f}_{XX}(\omega) \left[\frac{1}{\mu_\omega + (\nu_\omega)^{1/2} cv_\omega(\alpha/2)}, \frac{1}{\mu_\omega + (\nu_\omega)^{1/2} cv_\omega(1-\alpha/2)} \right] & \text{for } \omega \neq 0, \pi \end{cases} \tag{33}$$

where $cv_\omega(\alpha)$ is the α quantile of Ξ^*_ω, i.e. $P(\Xi^*_\omega \leq cv_\omega(\alpha)) = \alpha$.

Direct computations show that when ρ goes to infinity, Ξ^*_ω is approximately standard normal, $\mu_\omega = 1 + o(1)$ and $\nu_\omega = (\pi/(2\rho g))^{1/2}(1 + o(1))$. Therefore, the preceding fixed-ρ confidence interval coincides with the large-ρ confidence interval in (28) as $\rho \to \infty$. However, in finite samples with a particular ρ, the fixed-ρ confidence interval may differ significantly from the large-ρ confidence interval because of three factors (i) Ξ^*_ω is not standard normal for small ρ; (ii) $\mu_0 \neq 1$; and (iii) $\nu_\omega \neq (\pi/(2\rho g))^{1/2}$.

To construct the fixed-ρ confidence interval, we first need to compute μ_ω and ν_ω by analytical or numerical integration. We then need to find the quantiles of Ξ^*_ω by simulation. It is easy to see that the distribution of Ξ^*_ω is the same for all nonzero ω's. So it suffices to simulate two distributions, one for $\omega = 0$ and the other one for $\omega \neq 0$. Table 2 reports the 2.5%, 5%, 95% and 97.5% quantile functions of Ξ^*_ω. The Brownian motion process is approximated using normalized partial sums of 1000 normal variates and the simulation involves 10,000 replications. For each $\alpha = 2.5\%, 5\%, 95\%, 97.5\%$ and $\rho = 1, 2, ..., 1500$, we obtain the quantiles and represent them approximately using a hyperbola of the form:

$$cv = \frac{b}{\rho - a} + c, \tag{34}$$

where c is the quantile from the standard normal distribution. This formulation can be justified in terms of a continued fraction approximation to the Cornish Fisher expansion of the limit distribution as $\rho \to \infty$, which is being developed by the authors in other work.

Table 2 gives nonlinear least squares estimates of a and b in (34). The standard errors (s.e.) are small, indicating that the hyperbola explains the quantiles very well.

As ρ increases, the fitted hyperbola approaches its asymptote and critical values from the fixed-ρ and large-ρ asymptotics become arbitrarily close to each other. However, for small ρ, both the upper and lower quantiles are larger than the respective normal quantiles, reflecting the fact that the distribution of Ξ^*_ω is skewed to the right and has a fat right tail. In fact, Ξ^*_ω can be written as an infinite weighted sum of independent χ^2_1 random variables. So, it is not surprising that the distribution of Ξ^*_ω inherits some properties of a χ^2 distribution.

Table 2: Approximate quantile functions of Ξ^*_ω ($P(\Xi^*_\omega < \frac{b}{\rho - a} + c = \alpha)$)

	$\omega = 0$				$\omega \neq 0$			
	2.5%	5%	95%	97.5%	2.5%	5%	95%	97.5%
				Parzen Kernel				
a	−2628.2	−1659.0	−1846.3	−1539.6	−2776	−2391.3	−1650.3	−2267.6
b	865.95	384.53	349.43	519.45	890.52	488.74	324.43	721.1
c	−1.96	−1.645	1.645	1.96	−1.96	−1.645	1.645	1.96
$s.e.$	0.0352	0.0231	0.009	0.0278	0.0325	0.0233	0.009	0.0270
				QS Kernel				
a	−3181.9	−2614.9	−9598.8	−3093.4	−3614.5	−2702.3	−6253.9	−3960
b	1397.1	772.19	1734.5	1263.3	1499.4	748.47	1237.7	1540.2
c	−1.96	−1.645	1.645	1.96	−1.96	−1.645	1.645	1.96
$s.e.$	0.0487	0.0367	0.0149	0.0295	0.0475	0.0357	0.0110	0.0336

5 Finite Sample Performance

This section examines the finite sample performance of steep kernel methods in spectral density estimation and robust regression testing in comparison with sharp Bartlett kernel and conventional kernel methods.

5.1 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 low frequency and business cycle frequencies. In order to compare performance in a more demanding setting, we allow for spectral peaks at these frequencies.

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

$$b = \frac{a}{a - 4\cos\omega}, \tag{35}$$

provided that

$$b < 0 \text{ and } \left|\frac{a(1-b)}{4b}\right| < 1. \tag{36}$$

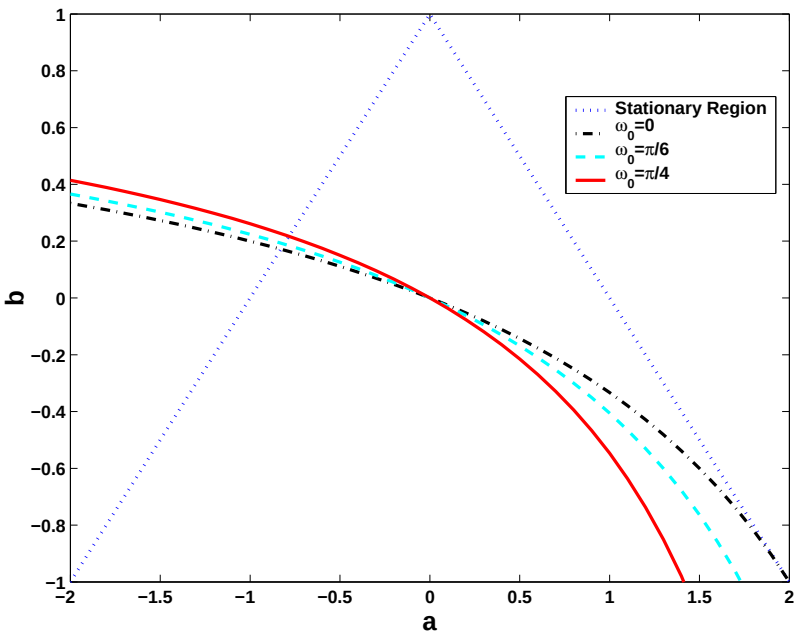

Figure 4: Stationary region and (a, b) combinations satisfying $b = a/(a - 4\cos\omega)$ for $\omega = 0,\ \pi/6,\ \pi/4$

See Priestley (1981, pp. 241). Fig. 4 depicts combinations of (a, b) satisfying (35) 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 $a = 0, 0.2, 0.4, 0.6, 0.8$.

We first compare the RMSE performance of different kernel estimators using the data-driven exponent or bandwidth. From Theorem 4 in Section 3, we can show that

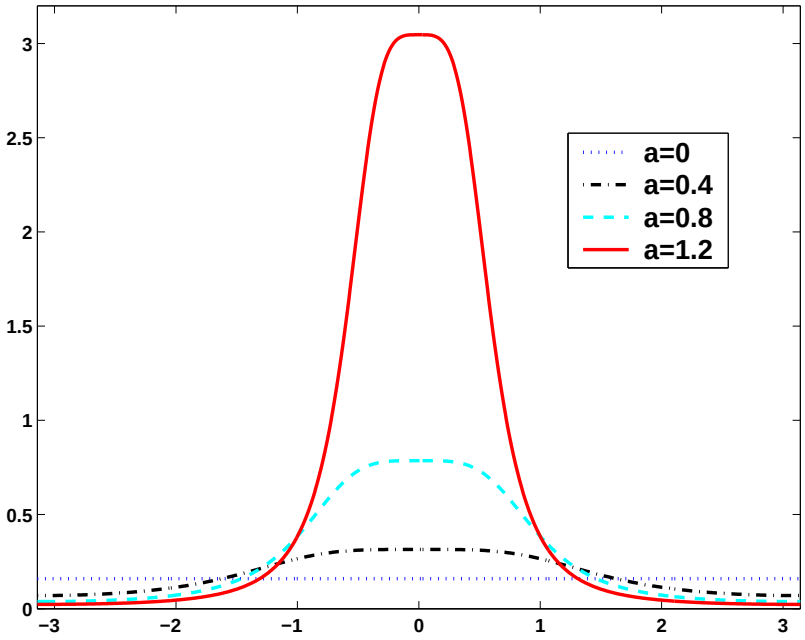

Figure 5: Spectral density of $AR(2)$ process with peak at $\omega = 0$

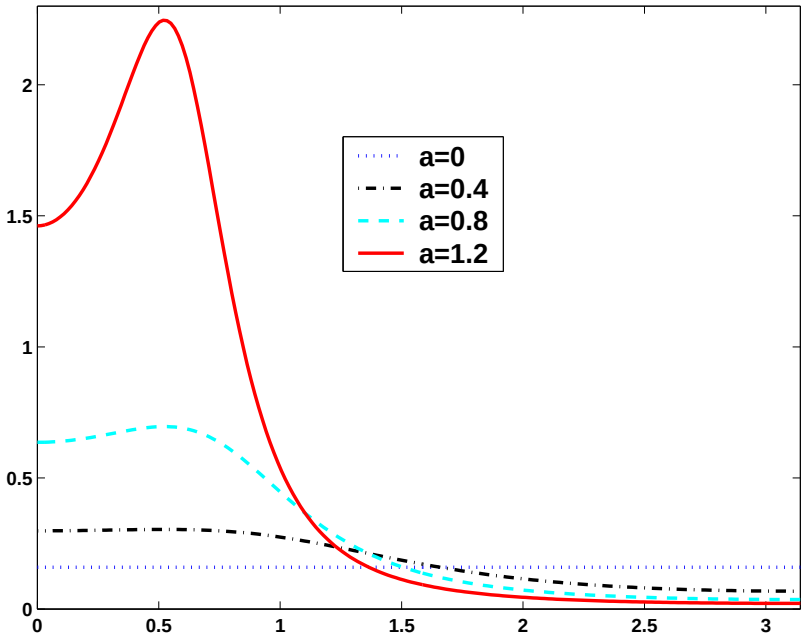

Figure 6: Spectral density of $AR(2)$ process with peak at $\omega = \pi/6$

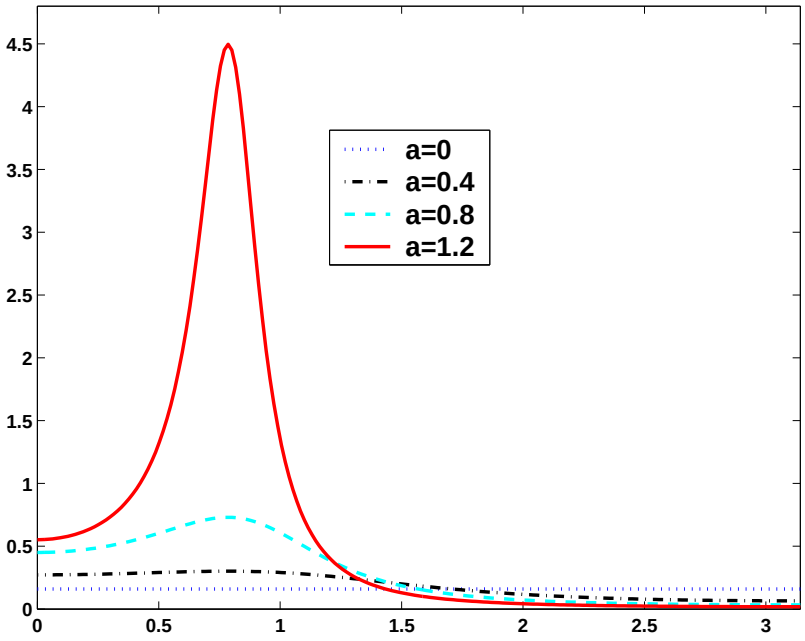

Figure 7: Spectral density of $AR(2)$ process with peak at $\omega = \pi/4$

for steep origin kernels the optimal exponent is

$$\rho^*_{steep} = \begin{cases} T^{8/5} g^{-1} \left[\frac{\sqrt{\pi} f^2_{XX}(\omega)}{4\sqrt{2}\left(f^{(2)}_{XX}(\omega)\right)^2} \right]^{2/5} & \text{for } \omega \neq 0, \pi \\ T^{8/5} g^{-1} \left[\frac{\sqrt{\pi} f^2_{XX}(\omega)}{2\sqrt{2}\left(f^{(2)}_{XX}(\omega)\right)^2} \right]^{2/5} & \text{for } \omega = 0, \pi \end{cases} . \tag{37}$$

An analogous analysis shows that for sharp origin kernels the optimal exponent is

$$\rho^*_{sharp} = \begin{cases} T^{2/3} \left[\frac{f^2_{XX}(\omega)}{2\left(f^{(1)}_{XX}(\omega)\right)^2} \right]^{1/3} & \text{for } \omega \neq 0, \pi \\ T^{2/3} \left[\frac{f^2_{XX}(\omega)}{\left(f^{(1)}_{XX}(\omega)\right)^2} \right]^{1/3} & \text{for } \omega = 0, \pi \end{cases} , \tag{38}$$

and, for the conventional estimator with the QS kernel the optimal bandwidth is

$$S_T = \begin{cases} 1.3221 T^{1/5} \left[\frac{2\left(f^{(2)}_{XX}(\omega)\right)^2}{f^2_{XX}(\omega)} \right]^{1/5} & \text{for } \omega \neq 0, \pi \\ 1.3221 T^{1/5} \left[\frac{\left(f^{(2)}_{XX}(\omega)\right)^2}{f^2_{XX}(\omega)} \right]^{1/5} & \text{for } \omega = 0, \pi \end{cases} \tag{39}$$

To implement these formulae, we use the $AR(1)$ and $AR(2)$ plug-in approaches (as in Andrews, 1991) in the simulation study.

We report results for the AR(1) plug-in approach as those for the AR(2) plug-in approach are qualitatively similar. Table 3 reports the ratio of $RMSE$ of the steep and sharp estimators to that of the conventional QS estimator for sample sizes $T = 50, 100, 200$. We use the conventional QS estimator as the benchmark as it minimizes the asymptotic MSE among the kernel estimators that are guaranteed to be positive semidefinite. We consider only Parzen, QS and Bartlett kernels as they are positive semidefinite and have been used in practice. It is clear that the steep kernel estimators have very competitive performance relative to the conventional QS estimator in an overall sense for the sample sizes and values of ω considered. When $\omega = 0$, the conventional QS estimator outperforms the steep kernel estimators by less than 2%. In contrast, for $\omega \neq 0$ the steep kernel estimators have the potential to outperform the conventional QS estimator, especially when the AR process is close to nonstationary, i.e. when there is a large peak in the spectrum. While the relative performance of the steep kernel estimators is quite robust, the relative performance of the sharp Bartlett kernel is sensitive to the sample size, frequency and DGP considered, with the RMSE ratio ranging from 86.3% to 169%. The simulation results show that the sharp Bartlett estimator has superior performance when the DGP is close to a white noise or a nonstationary process but its performance deteriorates dramatically for other cases.

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

		$T = 50$			$T = 100$			$T = 200$		
ω	a	Parzen	QS	Bartlett	Parzen	QS	Bartlett	Parzen	QS	Bartlett
0	0.0	1.007	1.004	0.961	1.008	1.006	0.971	1.012	1.011	0.991
	0.2	1.013	1.010	0.994	1.012	1.010	1.044	1.021	1.019	1.117
	0.4	1.012	1.008	1.034	1.019	1.016	1.098	1.022	1.019	1.161
	0.6	1.008	1.005	1.042	1.018	1.014	1.095	1.026	1.022	1.157
	0.8	1.014	1.010	1.058	1.012	1.008	1.069	1.024	1.019	1.105
$\pi/6$	0.0	1.010	1.008	0.984	1.006	1.005	0.971	1.007	1.006	0.968
	0.4	1.008	1.006	0.928	0.991	0.991	0.942	0.973	0.974	0.935
	0.8	1.005	1.000	1.153	1.012	1.007	1.367	1.026	1.022	1.690
	1.2	1.026	1.022	1.113	1.030	1.023	1.140	1.025	1.020	1.178
	1.6	0.980	0.983	0.970	0.964	0.971	0.923	0.953	0.960	0.863
$\pi/4$	0.0	1.006	1.004	0.986	1.004	1.003	0.996	1.004	1.003	0.984
	0.4	0.993	0.992	0.985	0.982	0.982	1.026	0.972	0.973	1.127
	0.8	1.004	1.002	1.186	1.004	1.003	1.225	1.002	1.002	1.255
	1.2	0.968	0.975	1.001	0.956	0.964	0.964	0.946	0.953	0.919

We now compare the coverage probabilities of confidence intervals based on the different kernel estimators and asymptotic theories. For steep quadratic kernels, the confidence intervals are given in (28) and (33) for different asymptotic specifications. For the conventional QS kernel estimator, we have, under some regularity conditions (see Anderson (1971)), that

$$\sqrt{\frac{T}{M}}\left(\frac{\widehat{f}_{XX}(\omega)-f_{XX}(\omega)}{f_{XX}(\omega)}\right)\rightarrow_d N(0,\left(1+1_{\omega\in\{0,\pi\}}\right)\int_{-\infty}^{\infty}k^2(x)dx), \tag{40}$$

where, for the QS kernel, $\int_{-\infty}^{\infty}k^2(x)dx=1$. Thus, an approximate $100(1-\alpha)\%$ confidence interval based on the conventional QS kernel estimator is given by

$$\begin{cases} \widehat{f}_{XX}(\omega)\left[\frac{1}{1+\sqrt{2M/T}cv(\alpha/2)},\frac{1}{1-\sqrt{2M/T}cv(\alpha/2)}\right] & \text{for } \omega=0,\pi \\ \widehat{f}_{XX}(\omega)\left[\frac{1}{1+\sqrt{M/T}cv(\alpha/2)},\frac{1}{1-\sqrt{M/T}cv(\alpha/2)}\right] & \text{for } \omega\neq 0,\pi \end{cases}, \tag{41}$$

where, as in (28), $cv(\alpha/2)$ is the critical value of a standard normal for area $\alpha/2$ in the right tail.

To make the comparison meaningful, we use comparable exponents and bandwidths. We consider $\rho_{parzen}=1,16,32$ for the steep Parzen kernels. It follows from equation (15) that the comparable exponents for the steep QS kernel are $\rho_{QS}=4,67$ and 135. For these values of ρ_{QS}, we choose the bandwidth according to $M=T/\sqrt{\rho_{QS}}$. Such a bandwidth choice rule ensures that the local behavior of $k_{QS,\rho}(x/T)$ at the origin matches that of $k_{QS}(x/M)$.

For each data generating process and exponent or bandwidth, we compute $\widehat{f}_{XX}(\omega)$ and construct the confidence intervals using (28), (33), or (41) for $T=50,100,200$. To evaluate the information content of these confidence intervals, we calculate their lengths. Obviously, the shorter a confidence interval is, the more informative it is. We focus on the steep Parzen and QS kernel estimators under the fixed and large ρ asymptotics and the conventional QS kernel estimator under the usual 'small-M' asymptotics given in (40). Tables 4a and 4b report the finite sample coverages and relative lengths of different 95% confidence intervals for $T=100$ and 200 based on 10,000 replications. The length reported is the median length over the 10,000 replications divided by the median length of the conventional QS confidence interval.

We draw attention to four aspects of Tables 4a and 4b. First, for the confidence interval constructed using the asymptotic normality results, the length and finite sample coverages are more or less the same. Second, the confidence interval based on the fixed-ρ asymptotics are shorter than those based on the large-ρ and small-M asymptotics. This is especially true when ρ is small or M is large as given in Table 4a, in which case the upper limits of the large-ρ and small-M confidence intervals are infinity and the relative lengths of the fixed-ρ confidence intervals approach zero. Therefore, the fixed-ρ confidence intervals are more informative than those based on consistent estimation and central limit theorems. Third, when $\omega=0$ and ρ is small, the fixed-ρ asymptotics delivers the shortest confidence interval whose empirical coverage is closest to the nominal coverage. When $\omega=0$ and ρ is relatively large, the

fixed-ρ confidence interval still has the best performance except when the DGP is very persistent. This finding shows that the fixed-ρ asymptotic distribution generally gives a more accurate approximation to the finite sample distribution than the large-ρ asymptotic distribution. Fourth, when $\omega \neq 0$, the empirical coverage of the fixed-ρ confidence interval is less satisfactory when the sample size is small and the DGP is close to nonstationarity but it improves substantially when the sample size increases. On the other hand, when $\omega \neq 0$, the fixed-ρ confidence interval is about 30% or 100% shorter than the large-ρ and small-M confidence intervals.

To sum up, the fixed-ρ steep-kernel-based confidence interval is the best in an overall sense when ρ is not large.

Table 4a: Finite sample coverages and relative lengths of different 95% confidence intervals[a,b] for $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$ with $b = \mathbf{1}_{\{\omega \neq 0\}} a/(a - 4\cos\omega)$ and $\varepsilon_t \sim iid(0,1)$.

		fixed-ρ Parzen		large-ρ Parzen		fixed-ρ QS		large-ρ QS		QS
		Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage
ω	a					$T = 100$				
0	0.0	0.950	0.000	0.998	0.894	0.950	0.000	0.998	0.935	0.995
	0.2	0.948	0.000	0.998	0.892	0.950	0.000	0.998	0.939	0.996
	0.4	0.949	0.000	0.998	0.890	0.951	0.000	0.998	0.934	0.996
	0.6	0.951	0.000	0.998	0.889	0.952	0.000	0.996	0.934	0.996
	0.8	0.947	0.000	0.999	0.880	0.950	0.000	0.998	0.925	0.998
$\pi/4$	0.0	0.863	0.000	0.959	0.986	0.868	0.000	0.962	0.990	0.961
	0.4	0.861	0.000	0.959	0.987	0.864	0.000	0.962	0.990	0.961
	0.8	0.857	0.000	0.961	0.981	0.861	0.000	0.964	0.988	0.963
	1.2	0.829	0.000	0.969	0.99	0.832	0.000	0.972	0.993	0.971
ω	a					$T = 200$				
0	0.0	0.953	0.000	0.998	0.894	0.954	0.000	0.998	0.937	0.996
	0.2	0.953	0.000	0.998	0.891	0.954	0.000	0.998	0.932	0.997
	0.4	0.953	0.000	0.998	0.892	0.954	0.000	0.998	0.933	0.997
	0.6	0.953	0.000	0.998	0.893	0.955	0.000	0.998	0.938	0.997
	0.8	0.953	0.000	0.998	0.889	0.955	0.000	0.998	0.933	0.997
$\pi/4$	0.0	0.864	0.000	0.959	0.986	0.868	0.000	0.962	0.990	0.961
	0.4	0.861	0.000	0.959	0.987	0.864	0.000	0.962	0.990	0.961
	0.8	0.857	0.000	0.961	0.981	0.861	0.000	0.964	0.988	0.963
	1.2	0.829	0.000	0.969	0.990	0.832	0.000	0.972	0.993	0.971

a. The exponents for the steep Parzen and QS kernels are 1 and 4 respectively.

b. The bandwidth for the conventional QS kernel is $T/2$.

Table 4b: Finite sample coverages and relative lengths of different 95% confidence intervals[c,d] for $X_t = aX_{t-1} + bX_{t-2} + \varepsilon_t$ with $b = \mathbf{1}_{\{\omega \neq 0\}} a/(a - 4\cos\omega)$ and $\varepsilon_t \sim iid(0,1)$.

		fixed-ρ Parzen		large-ρ Parzen		fixed-ρ QS		large-ρ QS		QS
		Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage
ω	a					$T = 100$				
0	0.0	0.948	0.539	0.975	1.098	0.946	0.531	0.976	1.108	0.973
	0.2	0.946	0.538	0.977	1.096	0.946	0.528	0.977	1.104	0.974
	0.4	0.943	0.538	0.978	1.095	0.942	0.528	0.979	1.102	0.978
	0.6	0.928	0.541	0.981	1.101	0.928	0.530	0.983	1.108	0.983
	0.8	0.816	0.558	0.970	1.136	0.808	0.543	0.970	1.134	0.963
$\pi/4$	0.0	0.935	0.701	0.955	1.045	0.935	0.694	0.956	1.047	0.954
	0.4	0.928	0.701	0.960	1.045	0.929	0.693	0.962	1.046	0.962
	0.8	0.888	0.701	0.968	1.055	0.888	0.698	0.970	1.053	0.970
	1.2	0.481	0.748	0.765	1.115	0.461	0.730	0.757	1.103	0.697
ω	a					$T = 200$				
0	0.0	0.952	0.538	0.976	1.096	0.952	0.529	0.977	1.105	0.975
	0.2	0.952	0.538	0.976	1.096	0.951	0.529	0.977	1.105	0.975
	0.4	0.951	0.539	0.978	1.097	0.949	0.530	0.978	1.107	0.976
	0.6	0.947	0.538	0.981	1.095	0.947	0.528	0.981	1.103	0.979
	0.8	0.923	0.542	0.985	1.103	0.922	0.531	0.987	1.109	0.984
$\pi/4$	0.0	0.934	0.702	0.952	1.046	0.934	0.695	0.955	1.049	0.955
	0.4	0.931	0.703	0.954	1.048	0.931	0.695	0.956	1.049	0.957
	0.8	0.925	0.703	0.958	1.047	0.924	0.695	0.961	1.049	0.963
	1.2	0.800	0.720	0.944	1.074	0.792	0.707	0.945	1.068	0.938

c. The exponents for the steep Parzen and QS kernels are 32 and 135 respectively.
d. The bandwidth for the conventional QS kernel is $T/\sqrt{135}$.

5.2 Robust Hypothesis Testing

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

$$y_t = z_t'\beta + u_t, \quad t = 1, 2, ..., T, \tag{42}$$

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

$$F_\rho^* = T(R\widehat{\beta} - r)' \left(R\widehat{Q}^{-1}\widehat{\Omega}_\rho\widehat{Q}^{-1}R'\right)^{-1} (R\widehat{\beta} - r)/p, \tag{43}$$

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

$$t_\rho^* = T^{1/2}(R\widehat{\beta} - r) \left(R\widehat{Q}^{-1}\widehat{\Omega}_\rho\widehat{Q}^{-1}R'\right)^{-1/2}, \tag{44}$$

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,

$$pF_{\widehat{\rho}}^* \Rightarrow W_p'(1)W_p(1) =_d \chi_p^2,\ t_{\widehat{\rho}}^* \Rightarrow W_1(1) =_d N(0,1), \tag{45}$$

under the null hypothesis, and

$$pF_{\widehat{\rho}}^* \Rightarrow \left(\Lambda^{*-1}c + W_p(1)\right)' \left(\Lambda^{*-1}c + W_p(1)\right),\ t_{\widehat{\rho}}^* \Rightarrow \left(\gamma + W_1(1)\right), \tag{46}$$

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

The above limiting distributions hold under the 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. Using Theorem 8, we can show that the F_ρ^* and t_ρ^* statistics have the following limiting distributions under the fixed-ρ asymptotics. First, under the null $H_0 : R\beta = r$,

$$\begin{aligned} pF_\rho^* &\Rightarrow W_p'(1)\left(\int_0^1\int_0^1 k_\rho^*(r-s)dW_p(r)dW_p'(s)\right)^{-1} W_p(1) \\ &: = W_p'(1)\left(\int_0^1\int_0^1 k_\rho(r-s)dV_p(r)dV_p'(s)\right)^{-1} W_p(1) \end{aligned} \tag{47}$$

and

$$t_\rho^* \Rightarrow W_1(1)\left(\int_0^1\int_0^1 k_\rho(r-s)dV_1(r)dV_1'(s)\right)^{-1/2}. \tag{48}$$

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

$$pF_\rho^* \Rightarrow \left(\Lambda^{*-1}c + W_p(1)\right)' \left(\int_0^1\int_0^1 k_\rho(r-s)dV_p(r)dV_p'(s)\right)^{-1} \left(\Lambda^{*-1}c + W_p(1)\right), \tag{49}$$

and

$$t_\rho^* \Rightarrow \left(\gamma + W_1(1)\right)\left(\int_0^1\int_0^1 k_\rho(r-s)dV_1(r)dV_1(s)\right)^{-1/2}. \tag{50}$$

In these formulae, $k_\rho(\cdot)$ is any positive semi-definite kernel (so the steep Parzen and QS kernels may be used), and $V_p(r)$ is p-dimensional standard Brownian bridge. For derivations of the preceding formulae, see PSJ (2003).

Table 5: Asymptotic critical value functions for
the one-sided t^*_ρ-test with steep Parzen and QS kernels

	Parzen				QS			
	90.0%	95.0%	97.5%	99.0%	90.0%	95.0%	97.5%	99.0%
a	−2.152	−1.884	−2.036	−2.370	−281.328	−43.119	−40.141	−21.394
b	4.260	6.604	10.012	16.015	120.806	63.656	83.741	93.564
c	1.282	1.645	1.960	2.326	1.282	1.645	1.960	2.326
$s.e.$	0.001	0.002	0.005	0.007	0.035	0.031	0.050	0.054

Given the above fixed-ρ asymptotics, the critical values for different ρ values can be simulated and tabulated. As in Table 2, we approximate the Brownian motion by normalized partial sums of 1000 *iid* $N(0,1)$ random variables and simulate the t^*_ρ statistic 10,000 times. It turns out that the critical values at a given significance level can be represented approximately by a hyperbola of the following form (again motivated by a continued fraction asymptotic approximation):

$$cv = \frac{b}{\rho - a} + c, \tag{51}$$

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 small, especially when the steep Parzen kernel is used. Figs. 8 and 9 depict the fitted hyperbolae at different significance levels. These figures show that the curves are nearly flat for large ρ and the critical values are very close to those from the standard normal as $\rho \to \infty$. This is not surprising as the t-statistic is asymptotically normal under the large-ρ asymptotics.

We now proceed to investigate the asymptotic power of the t^* test under both fixed-ρ asymptotics and large-ρ asymptotics. For convenience, we refer to these two tests as the t^*_ρ test and the $t^*_{\widehat{\rho}}$ test, respectively. Note that for a given exponent the t statistic is constructed in exactly the same way regardless of the asymptotics used. The difference between the two tests is that for the t^*_ρ test the exponent is fixed a priori and critical values from the fixed-ρ asymptotics are used, while for the $t^*_{\widehat{\rho}}$ test the exponent is data driven and critical values from the large-ρ asymptotics are used. For the $t^*_{\widehat{\rho}}$ test, the power curve is the same as the power envelope that is obtained when the true Ω or any consistent estimate is used. This holds because the consistency approximation is being used. For the t^*_ρ test, we consider three values of $\rho : \rho = 1, 16$ and 32 for the steep Parzen kernel and $\rho = 6, 96$ and 192 for the steep QS kernel. For each ρ, we approximate the Brownian motion and Brownian bridge processes by the partial sums of 1000 normal variates.

Fig. 10 presents the asymptotic power curves when the steep Parzen kernel is used. The figure is based on 50,000 simulation replications. It is apparent that the power curve moves 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 larger values of ρ are needed here, consonant with the power parameter expansion rates established

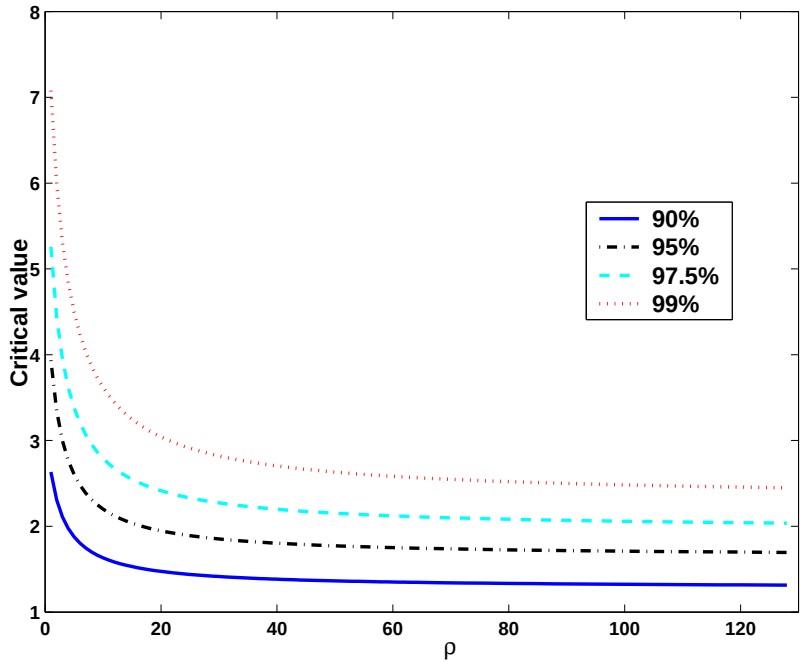

Figure 8: Asymptotic critical value curves at various significance levels for one-sided t test using the steep Parzen kernel

for consistent HAC estimation earlier in the paper. To save space, we do not report the local power curves when the steep QS kernel is used but instead comment on it briefly. The curves are similar to those in Fig. 10 but we need to take a larger ρ to attain the same power. This result is consistent with the curvature difference at the origin between the Parzen and QS kernels.

Compared with the t^*_ρ test, the $t^*_{\widehat{\rho}}$ test has an 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 (44) with a data-driven $\widehat{\rho}$. The point of departure is that, instead of using the critical values from the standard normal, we propose using the critical values from the hyperbola defined in (51). 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 of 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,

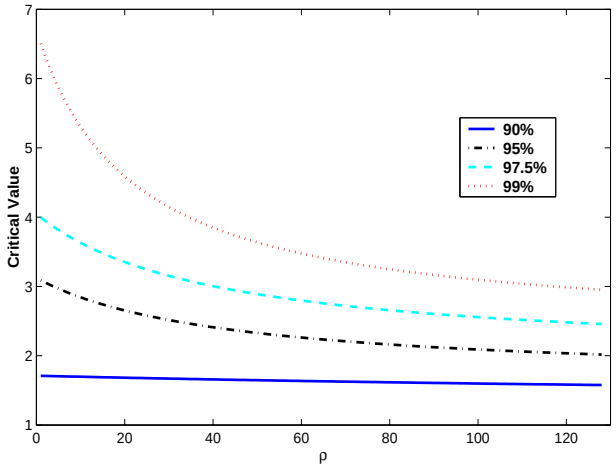

Figure 9: Asymptotic critical value curves at various significance levels for one-sided t test using the steep QS kernel

$t^*_{\widehat{\rho}}$ test and t^*_{new} test) and the conventional (i.e., bandwidth 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 $H_1 : \mu > 0$. In computing the conventional t-statistic, t_{HAC}, the bandwidth is chosen by the AR(1) plug-in approach as in Andrews (1991).

Tables 6a and 6b present 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 (51). For the $t^*_{\widehat{\rho}}$ 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 almost 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 increases in ρ, the size distortions of the t^*_ρ test constructed using steep Parzen or QS kernels increase less dramatically than those using sharp Bartlett kernels. Third, when the

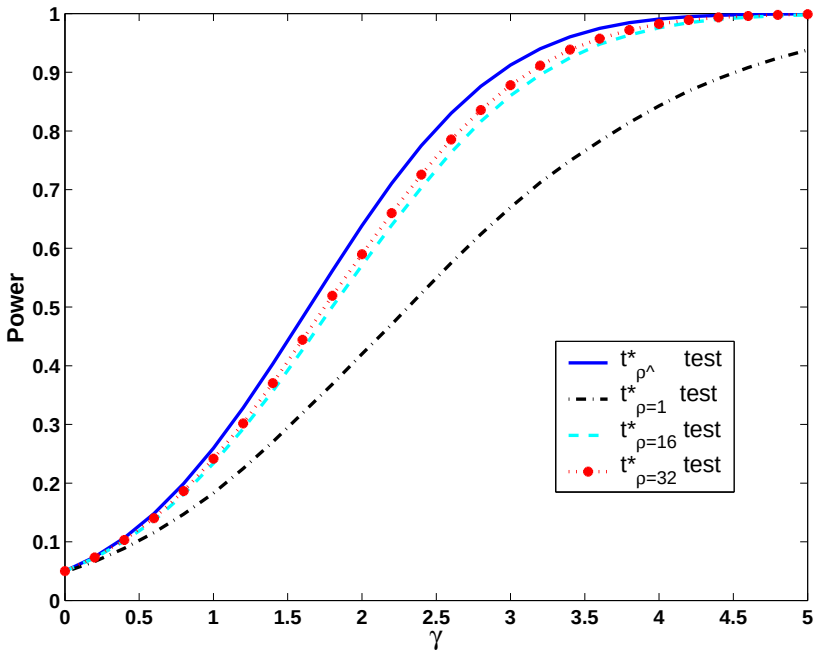

Figure 10: Asymptotic Local Power Function of the t^* Tests with the Steep Parzen Kernel

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 $t^*_{\widehat{\rho}}$ test is close to 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, the $t^*_{\widehat{\rho}}$ test generally 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 with the Parzen kernel and the t^*_6 test with the QS kernel.

Table 6a: Finite Sample Null Hypothesis Rejection Probabilities
for the 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 $e_t \sim iid(0,1)$ with the Parzen kernel

	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

Table 6b: Finite Sample Null Hypothesis Rejection Probabilities
for the 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 $e_t \sim iid(0,1)$ with the QS kernel

	a_1	a_2	t_{HAC}	$t^*_{\widehat{\rho}}$	t^*_{new}	$t^*_{\rho=6}$	$t^*_{\rho=96}$	$t^*_{\rho=192}$
$T = 50$	-0.500	0.000	0.057	0.037	0.032	0.064	0.037	0.042
	0.000	0.000	0.059	0.058	0.054	0.068	0.041	0.047
	0.300	0.000	0.076	0.079	0.067	0.071	0.045	0.056
	0.500	0.000	0.097	0.097	0.070	0.074	0.056	0.070
	0.700	0.000	0.124	0.125	0.075	0.082	0.079	0.104
	0.900	0.000	0.228	0.224	0.123	0.129	0.184	0.228
	0.950	0.000	0.302	0.292	0.179	0.177	0.259	0.298
	0.990	0.000	0.378	0.360	0.249	0.232	0.333	0.366
	1.500	-0.750	0.140	0.112	0.034	0.059	0.015	0.019
	1.900	-0.950	0.360	0.260	0.140	0.030	0.029	0.049
	0.800	0.100	0.231	0.232	0.135	0.134	0.197	0.240
$T = 200$								
	-0.500	0.000	0.050	0.043	0.042	0.071	0.043	0.049
	0.000	0.000	0.057	0.056	0.056	0.072	0.044	0.048
	0.300	0.000	0.068	0.071	0.068	0.072	0.044	0.049
	0.500	0.000	0.074	0.076	0.072	0.073	0.045	0.051
	0.700	0.000	0.084	0.088	0.075	0.075	0.049	0.056
	0.900	0.000	0.127	0.129	0.080	0.084	0.079	0.105
	0.950	0.000	0.168	0.171	0.096	0.103	0.128	0.163
	0.990	0.000	0.315	0.308	0.195	0.191	0.276	0.319
	1.500	-0.750	0.083	0.074	0.039	0.071	0.038	0.039
	1.900	-0.950	0.191	0.166	0.063	0.066	0.012	0.007
	0.800	0.100	0.131	0.135	0.091	0.086	0.086	0.112

There are inevitable trade-offs between finite sample size and power. Fig. 11 shows the finite sample power of the Parzen-kernel-based tests when $a_1 = 0.7$ without size correction. The graph is similar for the QS-kernel-based test and that is omitted to save space. 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. The t^*_{new} test also has very competitive finite sample power but much reduced size distortion. Simulation results (not reported here) 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. 11 also shows the size distortions of the different tests, which are shown 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.

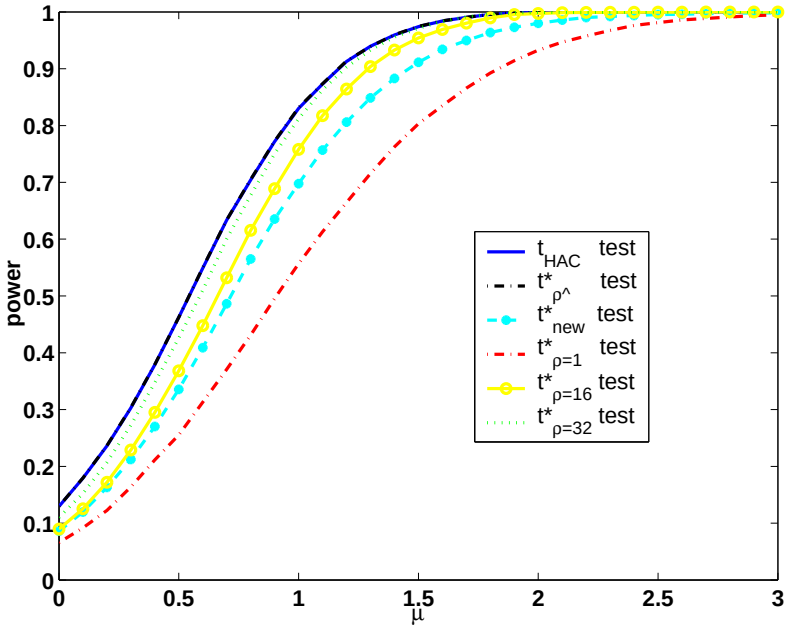

Figure 11: Finite Sample Power of Parzen-Kernel-Based Test for Location Model: $y_t = \mu + u_t, u_t = 0.7u_{t-1} + e_t$ with $T = 50$.

6 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 (2004) 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.

The simulation results of the present paper confirm earlier findings of the existence of a trade-off between power and size in econometric testing. Methods that reduce size distortion, such as fixed ρ asymptotics, also lead to some reduction in power relative to alternate methods based on large-ρ asymptotics. This trade-off between improvement in test size and reduction in test power can be quantified using higher order expansions of the limit theory and criteria for the optimal selection of control parameters such as the power exponent ρ or the bandwidth in conventional kernel estimation can be developed. Some research along these lines for the case of sharp kernels is reported by the authors (2004) in other work.

7 Appendix

Proof of Theorem 1. Part (a). We first show that

$$E\widehat{\gamma}_h = \gamma_h(1 - \frac{|h|}{T}) + O(\frac{|h|}{T}). \tag{52}$$

For $h \geq 0$, some simple manipulations yield

$$E\widehat{\gamma}_h = E\frac{1}{T}\sum_{t=1}^{T-h}\left(X_{t+h} - \bar{X}\right)\left(X_t - \bar{X}\right)' = E\gamma_h^0 + O(\frac{|h|}{T}) \tag{53}$$

where

$$\gamma_h^0 = \frac{1}{T}\sum_{t=1}^{T-h} X_{t+h}X_t' - \left(\frac{1}{T-h}\sum_{s=1}^{T-h} X_{s+h}\right)\left(\frac{1}{T-h}\sum_{s=1}^{T-h} X_s\right)'.$$

Direct computation leads to

$$\begin{aligned}
E\gamma_h^0 &= \gamma_h(1 - \frac{h}{T}) - \frac{1}{T(T-h)}E\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(X_{s+h} - \mu\right)\left(X_t - \mu\right)' \\
&= \gamma_h(1 - \frac{h}{T}) - \frac{1}{T(T-h)}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\gamma_{s+h-t} \\
&= \gamma_h(1 - \frac{|h|}{T}) - \frac{1}{T}\left[\frac{1}{T-h}\gamma_h + \sum_{\tau=1}^{T-h-1}\left(1 - \frac{\tau}{T-h}\right)\gamma_{-\tau-h}\right. \\
&\quad \left. + \sum_{\tau=1}^{T-h-1}\left(1 - \frac{\tau}{T-h}\right)\gamma_{\tau-h}\right] \\
&= \gamma_h(1 - \frac{|h|}{T}) + O(\frac{1}{T}).
\end{aligned} \tag{54}$$

Combining (53) and (54) gives the desired result.

Similarly, we can show that (52) holds for $h < 0$. Using (52), we have

$$\begin{aligned}
&\frac{T^2}{\rho}E(\widehat{f}_{XX}(0) - f_{XX}(0)) \\
= \;& \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})(1 - \frac{|h|}{T})\gamma_h - \frac{T^2}{\rho}\frac{1}{2\pi}\sum_{-\infty}^{\infty}\gamma_h \\
& + O\left(\frac{T^3}{\rho}\frac{1}{T}\sum_{h=-T+1}^{T-1}\frac{|h|}{T}k_\rho(\frac{h}{T})\right) \\
= \;& \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T}) - 1\right]\gamma_h - \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})\frac{|h|}{T}\gamma_h
\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{2\pi}\frac{T^2}{\rho}\sum_{|h|\geq T}\gamma_h + O\left(\frac{T^3}{\rho}\int_0^1 xk_\rho(x)dx\right) \\
= \; & \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h - \frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}k_\rho(\frac{h}{T})\frac{|h|}{T}\gamma_h \\
& + O\left(\frac{1}{\rho}\right) + O\left(\frac{T^3}{\rho^{5/2}}\right) \qquad (55)
\end{aligned}
$$

where the last line follows because $\left|\sum_{|h|\geq T}\gamma_h\right| \leq T^{-2}\sum_{|h|\geq T}\left|h^2\gamma_h\right|$ and $\int_0^1 xk_\rho(x)dx = O\left(\frac{1}{\rho^{3/2}}\right)$. The latter order of magnitude can be proved using the Laplace approximation method. The details are similar to the proof of (85) in what follows and are omitted. We now consider the first two terms in (**??**). The second term is bounded by

$$
\frac{1}{2\pi}\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\frac{|h|}{T}|\gamma_h| = \frac{T}{\rho}\frac{1}{2\pi}\sum_{h=-T+1}^{T-1}|h||\gamma_h| = O(\frac{T}{\rho}) = o(1) \qquad (56)
$$

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

$$
\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h = \sum_{h=-T/\log T}^{T/\log T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h + \sum_{T/\log T\leq|h|<T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h. \qquad (57)
$$

Noting that

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

and using Assumption 1(c), we obtain

$$
\begin{aligned}
\frac{T^2}{\rho}\sum_{h=-T+1}^{T-1}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h &= \frac{T^2}{\rho}\sum_{h=-T/\log T}^{T/\log T}\left[k_\rho(\frac{h}{T})-1\right]\gamma_h + o\left(\frac{\log^2 T}{\rho}\right) \\
&= \sum_{h=-T/\log T}^{T/\log T}\left[\frac{k(\frac{h}{T})-1}{\rho h^2/T^2}\right]h^2\gamma_h + o\left(1\right) \\
&= -g\sum_{-\infty}^{\infty}h^2\gamma_h\left(1+o(1)\right). \qquad (58)
\end{aligned}
$$

Combining the above results gives

$$
\lim_{T\to\infty}\frac{T^2}{\rho}E(\widehat{f}_{XX}(0)-f_{XX}(0)) = -g\sum_{-\infty}^{\infty}h^2\gamma_h. \qquad (59)
$$

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

$$\widehat{f}_{XX}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{XX}(\lambda) K_\rho(\lambda)\, d\lambda. \tag{60}$$

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 (60), viz.

$$\mathrm{Var}\left\{\widehat{f}_{XX}(0)\right\} = 2f_{XX}^2(0)\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})^2\left[1+o(1)\right], \tag{61}$$

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

To evaluate (61), 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

$$\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho^2\left(\frac{h}{T}\right) = \int_{-1}^{1} k_\rho^2(x)\, dx\,(1+o(1)). \tag{62}$$

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 $\log k(x) \le -\zeta(\delta)$ for $|x| \ge \delta$. Therefore, the contribution of the intervals $\delta \le |x| \le 1$ satisfies

$$\int_{\delta\le|x|\le1} k_\rho^2(x)\, dx = \int_{\delta\le|x|\le1} \exp\left\{2\rho\log k(x)\right\} dx \le \exp\left[-2(\rho-1)\zeta(\delta)\right]\int_{-1}^{1} k^2(x). \tag{63}$$

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

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

In consequence,

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

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

$$\int_{-\delta}^{\delta} \exp 2\rho \log k\left(x\right) dx = \left(\frac{\pi}{2\rho g}\right)^{1/2} (1+o(1)). \tag{64}$$

Combining (63) and (64) yields

$$\int_{-1}^{1} k_\rho^2\left(x\right) dx = \left(\frac{\pi}{2\rho g}\right)^{1/2} (1+o(1)), \tag{65}$$

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

$$\begin{aligned} K_\rho\left(\lambda_s\right) &= T \sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T}) e^{i\lambda_s h} = T\int_{-1}^{1} k_\rho(x) e^{i2\pi sx} dx\,(1+o(1)) \\ &= T\int_{-1}^{1} \exp\rho\left(\log k(x) + \log\cos(2\pi sx)\right)(1+o(1)). \end{aligned} \tag{66}$$

Using the Laplace approximation, we find that as $\rho \to \infty$, the contribution to the integral in (66), 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 $\zeta\left(\delta\right) > 0$ such that

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

Hence,

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

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

$$\begin{aligned}
& \rho^{1/4}\left\{\widehat{f}_{XX}(0) - f_{XX}(0)\right\} \\
= \; & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[I_{XX}(\lambda_s) - f_{XX}(0)\right] \\
= \; & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[I_{XX}(\lambda_s) - f_{XX}(\lambda_s)\right] \\
& + \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[f_{XX}(\lambda_s) - f_{XX}(0)\right]. \qquad (67)
\end{aligned}$$

Using Lemma 2, we have

$$K_\rho(\lambda_s) = \frac{\sqrt{\pi}T}{\sqrt{\rho g}} \exp\left(-\frac{\pi^2 s^2}{\rho g}\right)(1 + o(1)), \quad s = 0, 1, ..., [T/2]. \qquad (68)$$

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

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

Hence, the second term of (67) can be bounded as follows:

$$\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\left[f_{XX}(\lambda_s) - f_{XX}(0)\right] \\
= \; & \frac{\rho^{1/4}}{T} 2\sum_{s=0}^{[T/2]} K_\rho(\lambda_s)\left[f_{XX}(\lambda_s) - f_{XX}(0)\right] = O\left(\frac{\rho^{1/4}}{T}\sum_{s=0}^{[T/2]}|K_\rho(\lambda_s)|\lambda_s^2\right) \\
= \; & O\left(\frac{\rho^{1/4}}{T}\sum_{s=0}^{[T/2]}\frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\pi^2 s^2}{\rho g}\right)\lambda_s^2\right) \\
= \; & O\left(\rho^{-1/4}T^{-2}\int_0^\infty \exp\left(-\frac{\pi^2 s^2}{\rho g}\right)s^2 ds\right) \\
= \; & O(\rho^{-1/4}T^{-2}) = o(1). \qquad (69)
\end{aligned}$$

Then, by (67) and (69), we have

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

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 $\widetilde{C}_{\lambda j}=\sum_{s=j+1}^{\infty}C_se^{is\lambda}$, leading to the representation

$$X_t=C(L)\varepsilon_t=C(e^{i\lambda})\varepsilon_t+e^{-i\lambda}\widetilde{\varepsilon}_{\lambda t-1}-\widetilde{\varepsilon}_{\lambda t}, \tag{70}$$

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

$$\begin{aligned}
w(\lambda_s) &= \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}X_te^{it\lambda_s}\\
&= C(e^{i\lambda_s})w_{\varepsilon}(\lambda_s)+\frac{1}{\sqrt{2\pi T}}(\widetilde{\varepsilon}_{\lambda_s 0}-e^{in\lambda_s}\widetilde{\varepsilon}_{\lambda_s n})\\
&= C(e^{i\lambda_s})w_{\varepsilon}(\lambda_s)+O_p(T^{-1/2}).
\end{aligned} \tag{71}$$

Thus, using the fact that

$$\begin{aligned}
\sum_{s=0}^{T-1}|K_{\rho}(\lambda_s)| &= \frac{2\sqrt{\pi}T}{\sqrt{\rho g}}\sum_{s=0}^{[T/2]}\exp\left(-\frac{\pi^2s^2}{\rho g}\right)(1+o(1))\\
&= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\int_{-\infty}^{\infty}\exp\left(-\frac{\pi^2s^2}{\rho g}\right)ds\,(1+o(1))\\
&= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\int_{-\infty}^{\infty}\exp\left(-\frac{s^2}{2\rho g/(2\pi^2)}\right)ds\,(1+o(1))\\
&= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\left(\frac{1}{\sqrt{2\pi\rho g/(2\pi^2)}}\right)^{-1}(1+o(1))\\
&= O(T),
\end{aligned} \tag{72}$$

we get

$$\begin{aligned}
& \rho^{1/4}\left\{\widehat{f}_{XX}(0) - f_{XX}(0)\right\} \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)(I_{XX}(\lambda_s) - f_{XX}(\lambda_s)) + o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)(w(\lambda_s)w(\lambda_s)^* - f_{XX}(\lambda_s)) + o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)\{[C(e^{i\lambda_s})w_\varepsilon(\lambda_s) + O_p(T^{-1/2})] \\
& \times[C(e^{i\lambda_s})w_\varepsilon(\lambda_s) + O_p(T^{-1/2})]^* - f_{XX}(\lambda_s)\} + o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] + O_p\left(\frac{\rho^{1/4}}{T}T\frac{1}{T^{1/2}}\right) + o_p(1) \\
= & \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] + o_p(1), \qquad (73)
\end{aligned}$$

where we have used $\rho/T^2 \to 0$. The fourth equality follows because $K_\rho(\lambda_s)$ becomes progressively concentrated at the origin.

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

$$\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s=0}^{T-1} K_\rho(\lambda_s)[C^2(1)(I_{\varepsilon\varepsilon}(\lambda_s) - \frac{1}{2\pi}\sigma^2)] \\
= & 2\sum_{t=1}^{T} m_t + \frac{\rho^{1/4}}{T}C^2(1)\sum_{s=0}^{T-1} K_\rho(\lambda_s)\frac{1}{2\pi}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 - \sigma^2\right) \\
= & 2\sum_{t=1}^{T} m_t + \frac{\rho^{1/4}}{T}C^2(1)\left|\sum_{s=0}^{T-1} K_\rho(\lambda_s)\right| O_p(\frac{1}{\sqrt{T}}) \\
= & 2\sum_{t=1}^{T} m_t + O_p\left(\frac{\rho^{1/4}}{T}T\frac{1}{\sqrt{T}}\right) \\
= & 2\sum_{t=1}^{T} m_t + o_p(1). \qquad (74)
\end{aligned}$$

By the Fourier inversion formula, we have

$$c_j = \frac{C^2(1)}{2\pi}\frac{\rho^{1/4}}{T}k_\rho(\frac{j}{T}). \tag{75}$$

Hence

$$\sum_{j=1}^{T} c_j^2 = O\left(\frac{\rho^{1/2}}{T^2}\sum_{j=1}^{T} k_\rho^2(\frac{j}{T})\right) = O\left(\frac{\rho^{1/2}}{T^2}\left(\frac{\pi}{2\rho g}\right)^{1/2} T\right) = O\left(\frac{1}{T}\right). \tag{76}$$

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:

$$\sum_{t=1}^{T} E(m_t^2|\mathcal{F}_{t-1}) - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2} \to_p 0, \tag{77}$$

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

$$\sum_{t=1}^{T} E(m_t^4) \to_p 0. \tag{78}$$

We now proceed to establish (77) and (78). The left hand side of (77) is

$$\left(\sigma^2\sum_{t=2}^{T}\sum_{j=1}^{t-1}\varepsilon_j^2 c_{t-j}^2 - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) + \sigma^2\sum_{t=2}^{T}\sum_{r\neq j}\varepsilon_r\varepsilon_j c_{t-r}c_{t-j} := I_1 + I_2. \tag{79}$$

The first term, I_1, is

$$\sigma^2\left(\sum_{j=1}^{T-1}(\varepsilon_j^2 - \sigma^2)\sum_{s=1}^{T-j} c_s^2\right) + \left(\sigma^4\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} c_j^2 - \frac{\sigma^4 C^4(1)}{8\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) := I_{11} + I_{12}. \tag{80}$$

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 (76). Next, consider the second term of (80). 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:

$$\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\frac{s}{T}) &= \int_0^\infty (1-x)k_\rho^2(x)dx(1+o(1)) \\
&= \int_0^\infty \exp\left\{-x-\frac{1}{2}x^2-\rho g x^2\right\}dx(1+o(1)) \\
&= \int_0^\infty \exp\left\{-x-(\rho g+\frac{1}{2})x^2\right\}dx(1+o(1)) \\
&= \frac{1}{2}\exp\left\{2(\rho g+\frac{1}{2})^{-1}\right\}\frac{\sqrt{2\pi}}{\sqrt{2\rho g+1}} \\
&= \frac{1}{2}\left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)). \qquad (81)
\end{aligned}$$

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 (79) is $o_p(1)$.

Now consider the second term, I_2, of (79). I_2 has mean zero and variance

$$\begin{aligned}
& O\left(2\sum_{p,q=2}^{T}\sum_{r\neq j}^{\min(p-1,q-1)}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right) \\
= \; & O\left(2\sum_{p=2}^{T}\sum_{r\neq j}^{p-1}c_{p-r}^2 c_{p-j}^2 + 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j})\right). \qquad (82)
\end{aligned}$$

In view of (76), we have

$$\sum_{p=2}^{T}\sum_{r\neq j}^{p-1} c_{p-r}^2 c_{p-j}^2 = O\left(T\left(\sum_{j=1}^{T} c_j^2\right)^2\right) = O\left(\frac{1}{T}\right). \tag{83}$$

For the second component in (82), we have, using (76) and the Cauchy inequality,

$$\begin{aligned}
& 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r\neq j}^{q-1}(c_{q-r}c_{q-j}c_{p-r}c_{p-j}) \leq 4\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1} c_{q-r}^2 \sum_{r=1}^{q-1} c_{p-r}^2 \\
\leq\ & 4\sum_{i=1}^{T} c_i^2 \sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=1}^{q-1} c_{p-r}^2 \leq 4\left(\sum_{i=1}^{T} c_i^2\right)\left(\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1} c_r^2\right) \\
=\ & O\left(\frac{1}{T}\sum_{p=3}^{T}\sum_{q=2}^{p-1}\sum_{r=p-q+1}^{p-1} c_r^2\right) = O\left(\frac{1}{T}\sum_{r=1}^{T-2} r(T-r-1)c_r^2\right) \\
=\ & O\left(\frac{\rho^{1/2}}{T^3}\sum_{r=1}^{T-2} r(T-r-1)k_\rho(\frac{r}{T})\right) = O\left(\rho^{1/2}\int_0^1 x(1-x)k_\rho(x)dx\right). \qquad (84)
\end{aligned}$$

We now show that

$$\rho^{1/2}\int_0^1 x(1-x)k_\rho(x)dx = o(1). \tag{85}$$

To this end, we need the following result: if the function $p_\rho(x) = x(1-x)k_\rho(x)$ 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 $\rho_0 \geq \rho$ such that $x^*(\rho_0) \geq \varepsilon$. Since $x^*(\rho_0) \geq \varepsilon$, it follows from Assumption 1 that there exist a positive number $\zeta(\varepsilon)$ such that $k(x^*(\rho_0)) \leq 1 - \zeta(\varepsilon)$. Therefore

$$p_{\rho_0}(x^*(\rho_0)) \leq x^*(\rho_0)(1-x^*(\rho_0))\left[1-\zeta(\varepsilon)\right]^{\rho_0} \leq 1/4\left[1-\zeta(\varepsilon)\right]^{\rho_0}. \tag{86}$$

But for large ρ_0,

$$p_{\rho_0}(1/\rho_0) = \frac{1}{\rho_0}\left(1-\frac{1}{\rho_0}\right)\left(1-\frac{g}{\rho_0^2}\right)^{\rho_0}(1+o(1)) = \frac{1}{\rho_0}(1+o(1)). \tag{87}$$

Hence

$$p_{\rho_0}(1/\rho_0) > p_{\rho_0}(x^*(\rho_0)) \tag{88}$$

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

$$\kappa(x^*) = \frac{k''(x^*)}{k(x^*)} - \frac{(k'(x^*))^2}{k^2(x^*)}, \tag{89}$$

then

$$\begin{aligned}
&\int_0^1 x(1-x)k_\rho(x)dx \\
= &\int_0^1 \exp[\log(x)+\log(1-x)+\log k_\rho(x)]dx \\
= &\; x^*(1-x^*)k_\rho(x^*)\int_0^\infty \exp\left[-\left(\frac{1}{2\left(x^*-1\right)^2}+\frac{1}{2(x^*)^2}-\frac{1}{2}\rho\kappa(x^*)\right)y^2\right]dy(1+o(1)) \\
= &\; o\left(\frac{1}{\sqrt{\rho}}\right), \qquad (90)
\end{aligned}$$

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 (83), (84) and (90) completes the proof of $I_2 \to_p 0$. We have therefore established condition (77).

It remains to verify (78). Let A be some positive constant, then the left hand side of (78) is

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

using (76), which verifies (78) 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):

$$\begin{aligned}
\mathrm{Var}\left\{\widehat{f}_{XX}(\omega)\right\} &= f_{XX}^2(\omega)\frac{1}{T}\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})^2\left[1+o(1)\right], \\
&= f_{XX}^2(\omega)\left(\frac{\pi}{2\rho g}\right)^{1/2}\left[1+o(1)\right], \qquad (91)
\end{aligned}$$

where the last line uses (65). 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

$$\begin{aligned}
K(\lambda_s-\omega) &= T\sum_{h=-T+1}^{T-1} k_\rho(\frac{h}{T})e^{i(\omega-\lambda_s)h} = T\int_{-1}^{1} k_\rho(x)e^{(\omega xT-2\pi sx)i}dx\,(1+o(1)) \\
&= T\int_{-1}^{1}\exp\rho\left(\log k(x)+\log\cos(\omega xT-2\pi sx)\right)(1+o(1)). \qquad (92)
\end{aligned}$$

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

$$\begin{aligned}
K_\rho(\lambda_s) &= T\int_{-\delta}^{\delta}\exp\left\{\rho\log\left[k(x)\cos(\omega xT-2\pi sx)\right]\right\}(1+o(1)) \\
&= T\int_{-\delta}^{\delta}\exp\left\{\rho\log\left[k(x)\right]+(\omega T-2\pi s)\,xi\right\}dx\,(1+o(1)) \\
&= T\int_{-\delta}^{\delta}\exp\left\{-\rho g x^2+(\omega T-2\pi s)\,xi\right\}dx\,(1+o(1)) \\
&= T\int_{-\infty}^{\infty}\exp\left[-\rho g x^2+(\omega T-2\pi s)\,xi\right]dx\,(1+o(1)). \qquad (93)
\end{aligned}$$

Simple calculations give

$$\begin{aligned} K_\rho(\lambda_s) &= T\exp\left[-(\omega T-2\pi s)^2/(4\rho g)\right] \\ &\quad \times\int_{-\infty}^{\infty}\exp -\rho g\left[x^2+(\omega T-2\pi s)\,xi/\rho g-((\omega T-2\pi s)^2/(2\rho g)^2\right] \\ &= \frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{(\omega T-2\pi s)^2}{4\rho g}\right)(1+o(1)) \\ &= \begin{cases} O\left(\frac{T}{\sqrt{\rho}}\right) & \text{for } |\omega T-2\pi s|\le O(\sqrt{\rho}), \\ O\left(\frac{T}{\sqrt{\rho}}\exp\left(-\frac{(\omega T-2\pi s)^2}{4\rho g}\right)\right) & \text{for } |\omega T-2\pi s|>O(\sqrt{\rho}), \end{cases} \end{aligned} \tag{94}$$

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

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

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

$$B_\omega=\{s: s=J-T/2+1, J-[T]/2, ..., J, J+1, ..., J+T/2\}. \tag{95}$$

Then the scaled estimation error can be written as

$$\begin{aligned} &\rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}(\omega)\right\} \\ =\ &\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[I_{XX}(\lambda_s)-f_{XX}(\omega)\right] \\ =\ &\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[I_{XX}(\lambda_s)-f_{XX}(\lambda_s)\right] \\ &+\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho(\lambda_s-\omega)\left[f_{XX}(\lambda_s)-f_{XX}(\omega)\right]. \end{aligned} \tag{96}$$

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 (96) is

$$\begin{aligned}
& \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho\left(\lambda_s-\omega\right)\left[f_{XX}(\lambda_s)-f_{XX}\left(\omega\right)\right] \\
= & \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho\left(\lambda_s-\omega\right)\left|\lambda_s-\omega\right| \\
= & O\left(\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}\frac{\sqrt{\pi}T}{\sqrt{\rho g}}\exp\left(-\frac{\left(\omega-\lambda_s\right)^2T^2}{4\rho g}\right)\left|\lambda_s-\omega\right|\right) \\
= & O\left(\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}\frac{\sqrt{\pi}}{\sqrt{\rho g}}\exp\left(-\frac{\left(\omega T-2\pi s\right)^2}{4\rho g}\right)\left|2\pi s-\omega T\right|\right) \\
= & O\left(\frac{\rho^{1/4}}{T}\int_0^\infty \exp\left(-\frac{v^2}{4g}\frac{T^2}{\rho}\right)vdv\right) \\
= & O(\frac{\rho^{1/4}}{T})=o(1),
\end{aligned} \tag{97}$$

where we have used Lemma 5.

Combining (96) and (97) leads to

$$\rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}\left(\omega\right)\right\}=\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega}K_\rho\left(\lambda_s-\omega\right)\left[I_{XX}(\lambda_s)-f_{XX}\left(\lambda_s\right)\right]+o\left(1\right) \tag{98}$$

In view of (71), the frequency domain BN decomposition, we have

$$w(\lambda_s)=C(e^{i\lambda_s})w_\varepsilon(\lambda_s)+O_p(T^{-1/2}). \tag{99}$$

Following the same steps in (72), we can show that

$$\sum_{s=0}^{T-1}\left|K_\rho(\lambda_s-\omega)\right|=O(T). \tag{100}$$

Plugging (99) into (98) and using (72), we have

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

$$
\begin{aligned}
&= \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left|C(e^{i\lambda_s})\right|^2\left[I_{\varepsilon\varepsilon}(\lambda_s)-\frac{\sigma^2}{2\pi}\right]+o_p(1)\\
&= \frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left|C(e^{iw})\right|^2\left[I_{\varepsilon\varepsilon}(\lambda_s)-\frac{\sigma^2}{2\pi}\right]\\
&\quad+\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left(\left|C(e^{i\lambda_s})\right|^2-\left|C(e^{iw})\right|^2\right)\left[I_{\varepsilon\varepsilon}(\lambda_s)-\frac{\sigma^2}{2\pi}\right]+o_p(1),
\end{aligned}
\tag{101}
$$

where we have used $\rho/T^2\to 0$. But the second term in (101) is bounded by

$$
\frac{\rho^{1/4}}{T}\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left|\lambda_s-\omega\right|=o_p(1) \tag{102}
$$

by the smoothness of $C(e^{iw})$ and (97). Hence

$$
\rho^{1/4}\left\{\widehat{f}_{XX}(\omega)-f_{XX}(\omega)\right\}=\frac{\rho^{1/4}}{T}\left|C(e^{iw})\right|^2\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left[I_{\varepsilon\varepsilon}(\lambda_s)-\frac{\sigma^2}{2\pi}\right]. \tag{103}
$$

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 (74), we can write

$$
\frac{\rho^{1/4}}{T}\left|C(e^{iw})\right|^2\sum_{s\in B_\omega} K_\rho(\lambda_s-\omega)\left[I_{\varepsilon\varepsilon}(\lambda_s)-\frac{\sigma^2}{2\pi}\right]=2\sum_{t=1}^{T} m_t+o_p(1). \tag{104}
$$

Simple calculations show that

$$
c_j(\omega)=\frac{\left|C(e^{iw})\right|^2}{4\pi}\frac{\rho^{1/4}}{T}k_\rho(\frac{j}{T})\cos\omega j. \tag{105}
$$

Hence

$$
\sum_{j=1}^{T} c_j^2(\omega)=O\left(\frac{\rho^{1/2}}{T^2}\sum_{j=1}^{T}k_\rho^2(\frac{j}{T})\right)=O\left(\frac{1}{T}\right). \tag{106}
$$

We proceed to show that

$$
2\sum_{t=1}^{T} m_t\to_d N\left(0,\frac{\sigma^4\left|C(e^{iw})\right|^4}{4\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right)=N\left(0,f_{XX}^2(\omega)\left(\frac{\pi}{2g}\right)^{1/2}\right), \tag{107}
$$

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

$$\sum_{t=1}^{T} E(m_t^2|\mathcal{F}_{t-1}) - \frac{\sigma^4 \left|C(e^{iw})\right|^4}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2} \to_p 0, \tag{108}$$

and

$$\sum_{t=1}^{T} E(m_t^4) \to_p 0. \tag{109}$$

The left hand side of (108) is

$$\left(\sigma^2 \sum_{t=2}^{T}\sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2(\omega) - \frac{\sigma^4 C^4(1)}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) + \sigma^2 \sum_{t=2}^{T}\sum_{r\neq j} \varepsilon_r \varepsilon_j c_{t-r}(\omega)\, c_{t-j}(\omega) := \mathcal{I}_1 + \mathcal{I}_2. \tag{110}$$

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

$$\sigma^2\left(\sum_{j=1}^{T-1}(\varepsilon_j^2 - \sigma^2)\sum_{s=1}^{T-j} c_s^2(\omega)\right) + \left(\sigma^4 \sum_{t=1}^{T-1}\sum_{j=1}^{T-t} c_j^2(\omega) - \frac{\sigma^4 C^4(1)}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2}\right) := \mathcal{I}_{11} + \mathcal{I}_{12}. \tag{111}$$

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 (106). Next, consider the second term of (111). We have

$$\begin{aligned}
&\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} c_s^2(\omega) = \frac{\left|C(e^{i\omega})\right|^4}{4\pi^2}\frac{\rho^{1/2}}{T^2}\sum_{j=1}^{T-1}\sum_{s=1}^{T-j} k_\rho^2(\tfrac{s}{T})\cos^2\omega s \\
= \;& \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\frac{\rho^{1/2}}{T^2}\sum_{s=1}^{T-1}\sum_{j=1}^{T-s} k_\rho^2(\tfrac{s}{T})\cos^2\omega s \\
= \;& \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\frac{\rho^{1/2}}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\tfrac{s}{T})\cos^2\omega s \\
= \;& \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\rho^{1/2}\left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)) \qquad (112)\\
= \;& \frac{\left|C(e^{i\omega})\right|^4}{16\pi^2}\left(\frac{\pi}{2g}\right)^{1/2} + o(1),
\end{aligned}$$

where (112) follows from the approximation

$$\frac{1}{T}\sum_{s=1}^{T-1}(1-\frac{s}{T})k_\rho^2(\tfrac{s}{T})\cos^2\omega s = \left(\frac{\pi}{2\rho g}\right)^{1/2}(1+o(1)), \tag{113}$$

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

The above derivations therefore demonstrate that

$$\mathcal{I}_1 = \sigma^2 \sum_{t=2}^{T} \sum_{j=1}^{t-1} \varepsilon_j^2 c_{t-j}^2 (\omega) - \frac{\sigma^4 \left|C(e^{i\omega})\right|^4}{16\pi^2} \left(\frac{\pi}{2g}\right)^{1/2} \to_p 0. \tag{114}$$

So the first term of (110) 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 (109) can be verified in the same way.

Combining (103), (104) and (107) yields

$$\begin{aligned} &\frac{\rho^{1/4}}{T} \left|C(e^{iw})\right|^2 \sum_{s \in B_\omega} K_\rho (\lambda_s - \omega) \left[I_{\varepsilon\varepsilon}(\lambda_s) - \frac{\sigma^2}{2\pi}\right] \\ \to_d \; & N\left(0, \frac{\sigma^4 \left|C(e^{i\omega})\right|^4}{4\pi^2} \left(\frac{\pi}{2g}\right)^{1/2}\right) = N\left(0, \left(\frac{\pi}{2g}\right)^{1/2} f_{XX}^2(\omega)\right). \end{aligned} \tag{115}$$

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

$$\rho^{1/4} \left\{\widehat{f}_{XX}(\omega) - f_{XX}(\omega)\right\} \to_d N\left(0, \left(\frac{\pi}{2g}\right)^{1/2} f_{XX}^2(\omega)\right), \tag{116}$$

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

$$cov\left(\widehat{f}_{XX}(\omega_i), \widehat{f}_{XX}(\omega_j)\right) = \frac{1}{T^2} \sum_{\tau=0}^{T-1} \sum_{s=0}^{T-1} K(\lambda_s - \omega_i) K(\lambda_\tau - \omega_j) cov\left(I_{XX}(\lambda_s), I_{XX}(\lambda_\tau)\right). \tag{117}$$

Under Assumption 2, we have

$$\begin{aligned} (i) \mathrm{Var}\left(I_{XX}(\lambda_s) - f_{XX}(\lambda_s)\right) &= 4\pi^2 \delta_{0,\lambda_s} f_{XX}^2(\lambda_s)(1 + O(T^{-1/2})), \\ (ii) Cov\left(I_{XX}(\lambda_s), I_{XX}(\lambda_\tau)\right) &= O\left(f_{XX}(\lambda_s) f_{XX}(\lambda_\tau)/T\right), s \neq \tau, \end{aligned} \tag{118}$$

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

$$\begin{aligned} cov\left(\widehat{f}_{XX}(\omega_i), \widehat{f}_{XX}(\omega_j)\right) &= \left(T^{-2} \sum_{s=0}^{T-1} \delta_s K(\lambda_s - \omega_i) K(\lambda_s - \omega_j)\right) \left((1 + O(T^{-1/2}))\right) \\ &\quad + O\left(T^{-3} \sum_{s \neq \tau} \left|K(\lambda_s - \omega_i) K(\lambda_\tau - \omega_j)\right|\right). \end{aligned} \tag{119}$$

The second term in (119) 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 (119) is bounded by

$$\begin{aligned}
& O\left(\frac{1}{\rho}\sum_{s=0}^{T-1}\exp\left(-\frac{\left(\omega_i-\lambda_s\right)^2T^2}{4\rho g}-\frac{\left(\omega_j-\lambda_s\right)^2T^2}{4\rho g}\right)\right) \\
= & O\left(\frac{T}{\rho}\int_0^{2\pi}\exp\left(-\frac{T^2}{4\rho g}\left(\left(\omega_i-x\right)^2+\left(\omega_j-x\right)^2\right)\right)dx\right) \\
= & O\left\{\frac{T}{\rho}\int_0^{2\pi}\exp\left[-\frac{T^2}{4\rho g}\left(2\left(x-\frac{\omega_i+\omega_j}{2}\right)^2+\frac{\left(\omega_i-\omega_j\right)^2}{2}\right)\right]dx\right\} \\
= & O\left\{\frac{T}{\rho}\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\frac{\sqrt{\rho}}{T}\right\} \\
= & O\left\{\frac{1}{\sqrt{\rho}}\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\right\}. \qquad (120)
\end{aligned}$$

Therefore

$$\begin{aligned}
\rho^{1/2}cov\left(\widehat{f}_{XX}(\omega_i),\widehat{f}_{XX}(\omega_j)\right) &= O\left(\rho^{1/2}/T\right)+O\left(\exp\left[-\frac{T^2}{8\rho g}\left(\omega_i-\omega_j\right)^2\right]\right) \\
&= o(1) \qquad (121)
\end{aligned}$$

using $\rho/T^2\to 0$. ■

Proof of Theorem 8. (a) Let $\tilde{X}_t=X_t-1/T\sum_{s=1}^{T}X_s$. Simple calculations show that

$$\widehat{f}_{XX}(\omega)=\frac{1}{2\pi}\sum_{t=1}^{T}\sum_{\tau=1}^{T}k_\rho(\frac{t-\tau}{T})\left(\tilde{X}_te^{-it\omega}\right)\left(\tilde{X}_\tau e^{-i\tau\omega}\right)' \qquad (122)$$

where for a complex matrix z, z' denotes the conjugate and transpose of z. Let $\tilde{S}_\omega(r)=\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\tilde{X}_te^{-i\omega t}$. Using summation by parts twice, we have

$$\begin{aligned}
\widehat{f}_{XX}(\omega) &= \frac{1}{2\pi}\sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1}\tilde{S}_\omega(t/T)D_\rho(\frac{t-\tau}{T})\tilde{S}'_\omega(\tau/T) \\
&\quad+\frac{1}{2\pi}\tilde{S}_\omega(1)\sum_{\tau=1}^{T-1}\left(k_\rho(\frac{T-\tau}{T})-k_\rho(\frac{T-\tau-1}{T})\right)\tilde{S}'_\omega(\tau/T) \qquad (123) \\
&\quad+\frac{1}{2\pi}\sum_{t=1}^{T-1}\tilde{S}_\omega(t/T)\left(k_\rho(\frac{t-T}{T})-k_\rho(\frac{t-T+1}{T})\right)\tilde{S}'_\omega(1)+\tilde{S}_\omega(1)\tilde{S}'_\omega(1)
\end{aligned}$$

where

$$D_\rho(\frac{t-\tau}{T})=2k_\rho(\frac{t-\tau}{T})-k_\rho(\frac{t-\tau-1}{T})-k_\rho(\frac{t-\tau+1}{T}). \qquad (124)$$

We now consider three cases (i) $\omega = 0$, (ii) $\omega = \pi$ and (iii) $\omega \neq 0, \pi$ separately. First, when $\omega = 0$, we have

$$\tilde{S}_0(1) = 0,\ \tilde{S}_\omega(r) \Rightarrow \Lambda_0 \left(W_0(r) - W_0(1) r \right) := \Lambda_0 V_0(r) \tag{125}$$

where $V_0(r)$ is a standard Brownian Bridge. Note that when $T \to \infty$ such that $(t/T, \tau/T) \to (r, s)$ in the Euclidean metric $(||\cdot||)$ for some r and s, we have

$$\lim_{T \to \infty} T^2 D_T(\frac{t-\tau}{T}) = -k_\rho''(r-s). \tag{126}$$

Since $k(\cdot)$ is twice continuously differentiable, the above convergence is uniform in r and s. In other words, for any given $\varepsilon > 0$, there exists a positive Δ which is independent of r and s such that

$$\left| T^2 D_\rho(\frac{t-\tau}{T}) + k_\rho''(r-s) \right| < \varepsilon$$

whenever $\|(t/T, \tau/T) - (r,s)\| < \Delta$ for all (r,s) in $[0,1] \times [0,1]$. For a proof of the uniformity, see Weinstock (1957).

Combining (123), (125), and (126), and invoking the continuous mapping theorem, we get

$$\begin{aligned} \widehat{f}_{XX}(0) &\Rightarrow -\frac{1}{2\pi}\Lambda_0 \int_0^1 \int_0^1 k_\rho''(t-\tau) V_0(t) V_0'(\tau) dt d\tau \Lambda_0' \\ &= \frac{1}{2\pi}\Lambda_0 \int_0^1 \int_0^1 k_\rho(t-\tau) dV_0(t) dV_0'(\tau) \Lambda_0' \end{aligned} \tag{127}$$

where the last line follows from integration by parts. Some simple algebraic manipulations show that the last expression is the same as $(2\pi)^{-1} \Lambda_0 \int\int k_\rho^*(t-\tau) dW_0(t) dW_0'(\tau) \Lambda_0'$ as required.

Second, when $\omega = \pi$, we have

$$\tilde{S}_\pi(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (X_t - \mu) e^{-i\pi t} - \left(\bar{X} - \mu \right) \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} e^{-i\pi t} \Rightarrow \Lambda_\pi W_\pi(r). \tag{128}$$

Combining the above result with the the continuous mapping theorem leads to

$$\begin{aligned} &\widehat{f}_{XX}(\pi) \Rightarrow \frac{1}{2\pi}\Lambda_\pi \int_0^1 \int_0^1 W_\pi(t) k_\rho''(t-\tau) W_\pi'(\tau) \Lambda_\pi' + \frac{1}{2\pi}\Lambda_\pi W_\pi(1) W_\pi(1)' \Lambda_\pi' \\ &+ \frac{1}{2\pi}\Lambda_\pi \left(\int_0^1 k_\rho(1-\tau) W_\pi(\tau) d\tau \right) W_\pi'(1) \Lambda_\pi' + \frac{1}{2\pi}\Lambda_\pi W_\pi(1) \left(\int_0^1 k_\rho(1-\tau) W_\pi'(\tau) d\tau \right) \Lambda_\pi' \\ :\ &= \frac{1}{2\pi}\Lambda_\pi \int_0^1 \int_0^1 k_\rho(t-\tau) dW_\pi(t) dW_\pi'(\tau) \Lambda_\pi'. \end{aligned} \tag{129}$$

Finally, we consider $\omega \neq 0, \pi$. Note that

$$\tilde{S}_\omega(r) \Rightarrow \Lambda_\omega \left(W_{\omega R}(r) + i W_{\omega I}(r) \right). \tag{130}$$

Again using (123), (126), and the continuous mapping theorem, we get

$$\widehat{f}_{XX}(\omega) \Rightarrow \frac{1}{2\pi}\Lambda_\omega \int\int k_\rho(t-\tau)dW_\omega(t)dW'_\omega(\tau)\Lambda'_\omega. \tag{131}$$

Details are omitted.

b) For $\omega = 0$, we have

$$\begin{aligned} & E(2\pi)^{-1}\Lambda_0\Xi_0\Lambda'_0 = (2\pi)^{-1}\Lambda_0 E\int_0^1\int_0^1 k^*_\rho(t,\tau)dW_0(t)dW'_0(\tau)\Lambda'_0 \\ = \;& f_{XX}(0)\int_0^1 k^*_\rho(t,t)dt = f_{XX}(0)\left(1-\int_0^1\int_0^1 k_\rho(t,\tau)dtd\tau\right). \end{aligned} \tag{132}$$

For $\omega \neq 0$, we have

$$\begin{aligned} & E(2\pi)^{-1}\Lambda_\omega\Xi_\omega\Lambda'_\omega = (2\pi)^{-1}\Lambda_\omega E\int_0^1\int_0^1 k_\rho(t-\tau)dW_\omega(t)dW'_\omega(\tau)\Lambda'_\omega \\ = \;& (2\pi)^{-1}\Lambda_\omega\Lambda'_\omega + (2\pi)^{-1}\Lambda_\omega\Lambda'_\omega 1\{\omega \neq \pi\} = f_{XX}(\omega). \end{aligned} \tag{133}$$

c) We prove the case when $\omega \neq 0, \pi$, as the proofs for the other cases are similar and simpler. Write $E\left(\text{vec}(\Xi_\rho)\text{vec}(\Xi_\rho)'\right)$ as

$$\begin{aligned} & E\left(\int_0^1\int_0^1\int_0^1\int_0^1 k_\rho(r,s)k_\rho(p,q)\text{vec}\left(dW_\omega(r)dW'_\omega(s)\right)\text{vec}\left(dW_\omega(p)dW'_\omega(q)\right)'\right) \\ = \;& E\left(\int_0^1\int_0^1\int_0^1\int_0^1 k_\rho(r,s)k_\rho(p,q)\sum_{k_1,k_2,k_3,k_4\in\{R,I\}} i^{\#\{k_1,k_4\}}(-i)^{\#\{k_2,k_3\}}\right. \\ & \left.\times\,\text{vec}\left(dW_{\omega k_1}(r)dW'_{\omega k_2}(s)\right)\text{vec}\left(dW_{\omega k_3}(p)dW'_{\omega k_4}(q)\right)'\right) \end{aligned}$$

where $\#A$ denotes the number of elements in A which are equal to 'I'.

Some calculations show that $E\left(\text{vec}\left(dW_{\omega k_1}(r)dW_{\omega k_2}(s)\right)\text{vec}\left(dW_{\omega k_3}(p)dW'_{\omega k_4}(q)\right)'\right)$ is

$$\begin{cases} \text{vec}(I_m)\text{vec}(I_m)'drdp1_{\{k_1=k_2\}}1_{\{k_3=k_4\}}, & \text{if } r=s\neq p=q, \\ I_{m^2}drds1_{\{k_1=k_3\}}1_{\{k_2=k_4\}} & \text{if } r=p\neq s=q, \\ K_{mm}drds1_{\{k_1=k_4\}}1_{\{k_2=k_3\}} & \text{if } r=q\neq s=p, \\ 0, & \text{otherwise.} \end{cases} \tag{134}$$

Using the above result, we have

$$\begin{aligned}
& E\left(\text{vec}(\Xi_\omega)\text{vec}(\Xi_\omega)'\right) \\
= & \sum_{k_1,k_2,k_3,k_4\in\{R,I\}} i^{\#\{k_1,k_4\}}\,(-i)^{\#\{k_2,k_3\}}\,\text{vec}(I_m)\text{vec}(I_m)'1_{\{k_1=k_2\}}1_{\{k_3=k_4\}} \\
& + \sum_{k_1,k_2,k_3,k_4\in\{R,I\}} i^{\#\{k_1,k_4\}}\,(-i)^{\#\{k_2,k_3\}}\int_0^1\int_0^1 k_\rho^2(r-s)drdsI_{m^2}1_{\{k_1=k_3\}}1_{\{k_2=k_4\}} \\
& + \sum_{k_1,k_2,k_3,k_4\in\{R,I\}} i^{\#\{k_1,k_4\}}\,(-i)^{\#\{k_2,k_3\}}\int_0^1\int_0^1 k_\rho^2(r-s)drdsK_{mm}1_{\{k_1=k_4\}}1_{\{k_2=k_3\}} \\
= & \ 4\text{vec}(I_m)\text{vec}(I_m)' + 4\int_0^1\int_0^1 k_\rho^2(r-s)drdsI_{m^2} \\
= & \ 4\text{vec}(I_m)\text{vec}(I_m)' + 4\int_0^1\int_0^1 k_\rho^2(r-s)drdsI_{m^2}. \qquad (135)
\end{aligned}$$

Hence

$$\begin{aligned}
& \text{var}(\text{vec}(\Lambda_\omega\Xi_\omega\Lambda_\omega')) \\
= & \ E\text{vec}(\Lambda_\omega\Xi_\omega\Lambda_\omega')\text{vec}(\Lambda_\omega\Xi_\omega\Lambda_\omega')' - \text{vec}(\Lambda_\omega E\Xi_\omega\Lambda_\omega')\text{vec}(\Lambda_\omega E\Xi_\omega\Lambda_\omega')' \\
= & \ E\left(\Lambda_\omega\otimes\Lambda_\omega\right)\text{vec}(\Xi_\omega)\text{vec}(\Xi_\omega)'(\Lambda_\omega'\otimes\Lambda_\omega') - 4\text{vec}(\Lambda_\omega\Lambda_\omega')\text{vec}(\Lambda_\omega\Lambda_\omega') \\
= & \ 4\left(\Lambda_\omega\otimes\Lambda_\omega\right)\text{vec}(I_m)\text{vec}(I_m)'(\Lambda_\omega'\otimes\Lambda_\omega') \\
& + 4\int_0^1\int_0^1 k_\rho^2(r-s)drds\left(\Lambda_\omega\otimes\Lambda_\omega\right)\left(\Lambda_\omega'\otimes\Lambda_\omega'\right) - 4\text{vec}(\Lambda_\omega\Lambda_\omega')\text{vec}(\Lambda_\omega\Lambda_\omega') \\
= & \ 4\int_0^1\int_0^1 k_\rho^2(r-s)drds\left(\Lambda_\omega\otimes\Lambda_\omega'\right)\left(\Lambda_\omega\otimes\Lambda_\omega'\right) \\
= & \ 4\pi^2\int_0^1\int_0^1 k_\rho^2(r-s)drds\left(f_{XX}(\omega)\otimes f_{XX}(\omega)\right), \qquad (136)
\end{aligned}$$

giving the stated result. ■

8 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
$\rightarrow_d$	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] Hashimzade, N. and T. J. Vogelsang (2004): "Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators," Working paper, Department of Economics, Cornell University.

[8] Jansson, M. (2004): "On the Error of Rejection Probability in Simple Autocorrelation Robust Tests" *Econometrica* 72, 937-946.

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

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

[11] Kiefer, N. M. and T. J. Vogelsang (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Working paper, Department of Economics, Cornell University.

[12] 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).

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

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

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

[16] 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).

[17] Phillips, P. C. B., Y. Sun and S. Jin (2004): "Improved HAR Inference using Power Kernels without Truncation". Yale University, mimeographed.

[18] 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.

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

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

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

[22] Weinstock R. (1957) On continuous differentiability. *American Mathematical Monthly* 64, 492.