

# **The Size-Power Tradeoff in HAR Inference**

January 17, 2021

Eben Lazarus

MIT Sloan School of Management

Daniel J. Lewis

Federal Reserve Bank of New York

and

James H. Stock\*

Department of Economics, Harvard University

and the National Bureau of Economic Research

\*We are grateful to Yixao Sun for helpful comments on multiple drafts of this paper. We also thank Ian Dew-Becker, Frank Diebold, Michael Jansson, Lutz Kilian, Tom Kolokotronis, Mikkel Plagborg-Møller, Ulrich Müller, Benedikt Pötscher, Tim Vogelsang, Mark Watson, Ken West, numerous seminar participants, the editor, and three anonymous referees for discussions and/or comments, and Tim de Silva for excellent research assistance. Replication files for simulations in the Online Supplement are on Lewis's website.

## ABSTRACT

Heteroskedasticity and autocorrelation-robust (HAR) inference in time series regression typically involves kernel estimation of the long-run variance. Conventional wisdom holds that, for a given kernel, the choice of truncation parameter trades off a test's null rejection rate and power, and that this tradeoff differs across kernels. We formalize this intuition: using higher-order expansions, we provide a unified size-power frontier for both kernel and weighted orthonormal series tests using nonstandard "fixed- $b$ " critical values. We also provide a frontier for the subset of these tests for which the fixed- $b$  distribution is  $t$  or  $F$ . These frontiers are respectively achieved by the QS kernel and equal-weighted periodogram. The frontiers have simple closed-form expressions, which show that the price paid for restricting attention to tests with  $t$  and  $F$  critical values is small. The frontiers are derived for the Gaussian multivariate location model, but simulations suggest the qualitative findings extend to stochastic regressors.

JEL codes: C22, C32

Keywords: heteroskedasticity- and autocorrelation-robust estimation, HAR, long-run variance estimator

## 1. INTRODUCTION

Heteroskedasticity- and autocorrelation-robust (HAR) tests and confidence intervals are used in time series regression when the product of the regressors  $x_t$  and the regression errors  $u_t$ ,  $x_t u_t \equiv z_t$ , is potentially serially correlated and  $u_t$  is potentially heteroskedastic. Computing HAR standard errors entails estimating the long-run variance (LRV) of  $z_t$ ,  $\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j$ , where  $\Gamma_j = \text{cov}(z_t, z_{t-j}')$ ,  $j = 0, 1, \dots$ . The challenge of HAR inference is that  $\Omega$  depends on infinitely many autocovariances, but this infinite sum must be estimated using only  $T$  observations.

The foundational papers on HAR inference in the econometrics literature are Newey and West (1987) and Andrews (1991). The Newey-West/Andrews method, which dominates empirical practice, estimates  $\Omega$  using a kernel-weighted average of the first  $S$  sample autocovariances of  $\hat{z}_t = x_t \hat{u}_t$ , where  $\hat{u}_t$  are the OLS residuals. Andrews (1991) and Newey and West (1994) recommend choosing the truncation parameter sequence  $S_T$  to minimize the mean squared error (MSE) of the LRV estimator  $\hat{\Omega}$ . Under that sequence,  $\hat{\Omega}$  is consistent and inference proceeds using standard normal or chi-squared critical values. Drawing on classical results in the spectral estimation literature, Andrews (1991) further suggests using the Epanechnikov (1969) kernel, also called the quadratic spectral (QS) kernel, which minimizes the asymptotic MSE of  $\hat{\Omega}$  among kernel estimators that are positive semidefinite (psd).

Unfortunately, tests using the MSE-optimal truncation parameter can have large size distortions (e.g., den Haan and Levin (1997)). In fact, Edgeworth expansions of rejection probabilities in the Gaussian location model formally show that the testing problem entails a bias-variance tradeoff, in contrast to MSE minimization which entails a tradeoff between *squared* bias and variance, so the testing-optimal sequence  $S_T$  increases faster than the MSE-optimal sequence (Velasco and Robinson (2001), Sun, Phillips and Jin (2008)). The testing-optimal sequence introduces sampling variability in  $\hat{\Omega}$  which leads to  $t$ -like behavior of the  $t$ -statistic. That variability can be handled using Kiefer and

Vogelsang's (2005) "fixed  $b$ " (or "fixed smoothing") critical values, which model the truncation parameter as increasing proportionally to  $T$ , i.e.,  $S_T = bT$  with  $b$  fixed; doing so provides a higher-order refinement to the null rejection rate of HAR tests (Jansson (2004), Sun, Phillips and Jin (2008), and Sun (2014b)). The lesson is thus to combine a testing-optimal bandwidth rate for  $S_T$  with fixed- $b$  critical values. This literature, however, has two loose ends. First, given a kernel, it suggests but does not formalize a tradeoff between size and power that depends on whether  $S$ , while growing at the testing-optimal rate, is large or small (e.g., Kiefer and Vogelsang (2005, Section 5)). Moreover, there are no theoretical results on which kernel, if any, is optimal for testing.

This paper fills this gap by using the asymptotic expansions of Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2011, 2013, 2014b) for the Gaussian location model to study the tradeoff between the size distortion and power loss for HAR tests using fixed- $b$  critical values and the testing-optimal rate for  $S$ . By size distortion, we mean the difference between the null rejection rate and the desired nominal significance level  $\alpha$ . Following Rothenberg (1984) and the literature on higher-order comparisons of tests, by power we mean size-adjusted power, that is, the rejection rate under the alternative when the test is evaluated using (generally infeasible) critical values that have been adjusted so that the rejection rate under the null is  $\alpha$ .<sup>1</sup>

---

<sup>1</sup> The Neyman-Pearson Lemma ranks tests by their probability of rejecting a point alternative among tests with the same null rejection rate. This principle extends to the second-order comparison of tests based on Edgeworth expansions, which entails (i) obtaining expressions for second-order corrections to critical values, (ii) imposing those corrections so that tests have the same second-order size, then (iii) obtaining and comparing their size-adjusted power. Sun, Phillips, and Jin (2008, Corollary 5) use these three steps to derive a higher order approximation to the power of kernel HAR tests. The practice of using size-adjusted critical values is commonplace in Monte Carlo studies comparing competing tests; for example, see Kiefer and Vogelsang (2002), Sun (2013), Long and Ervin (2000), Ng and Perron (2001), and Clark and West (2007).

This paper makes four main contributions. First, we derive theoretical expressions characterizing the tradeoff between the size distortion  $\Delta_S$  and its size-adjusted power loss  $\Delta_P$ , that is, the difference between the local asymptotic power of the candidate HAR test and the infeasible oracle test with  $\Omega$  known.

Second, we derive the envelope of these size-power tradeoffs and show that this size-power frontier is achieved by the QS kernel. Let  $\Delta_P^{\max}$  be the maximum size-adjusted power loss of the test over all alternatives. For a 5% test in the one-dimensional Gaussian location model, the size-power frontier is,

$$\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{0.3368}{T} + o(T^{-1}), \quad (1)$$

where  $\omega^{(2)}$  is the normalized curvature of the spectral density of  $z_t$  at frequency zero (in the scalar case, the negative of the ratio of the second derivative of the spectral density to the spectral density at frequency zero). For the  $m$ -dimensional location model, the only change to (1) is that the constant increases with  $m$ .

The frontier is plotted in Figure 1 for 5% tests for  $m = 1, 2$ , and 3. Choosing the sequence for  $b$  to equate the asymptotic rates at which  $\Delta_S$  and  $\Delta_P^{\max}$  converge to zero in (1) yields  $\Delta_S, \Delta_P^{\max} = O(T^{2/3})$ , and this rate is used to derive (1) and to scale the axes in Figure 1. For the Bartlett kernel used in the Newey-West (1987) test, equating these rates yields  $\Delta_S, \Delta_P = O(T^{-1/2})$ , so the Bartlett kernel HAR test is asymptotically dominated.

Third, we extend these results for kernel HAR tests to the family of weighted orthogonal series (WOS) tests and in the process provide unified expressions covering the two families. WOS estimators of  $\Omega$  are weighted sums of the squared projections of  $\hat{z}_t$  onto low-frequency orthonormal functions, typically the first  $B$  terms of a basis of  $L^2[0,1]$  excluding the constant function. The WOS family includes weighted periodogram tests (for which the orthogonal series are Fourier series) and, in the location model, Ibragimov and Müller's (2010) subsample estimator. If the weights are equal, WOS HAR tests have standard  $t$  and  $F$  fixed- $b$  distributions (Brillinger (1975, exercise 5.13.25), Müller (2007), Phillips (2005), and Sun (2013)). Building on Sun (2013), we characterize the size-power

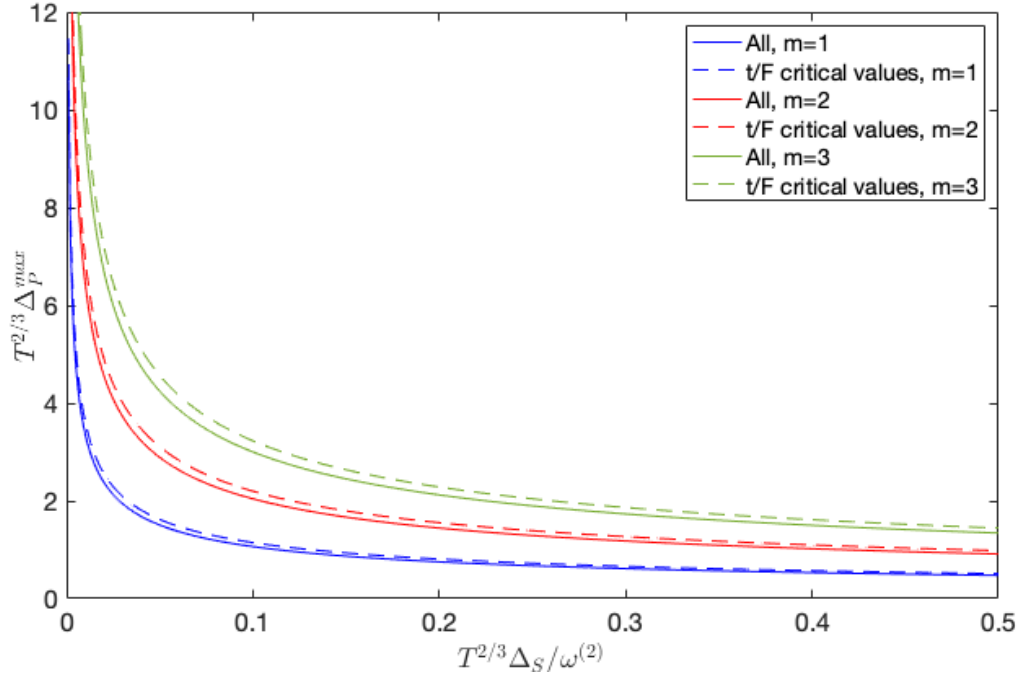


FIGURE 1. Higher-order frontier between the size distortion  $\Delta_S$  and the maximum power loss  $\Delta_P^{\max}$  of HAR tests in the Gaussian location model with dimension  $m$ , for stationary processes with normalized spectral curvature  $\omega^{(2)}$ . Solid line: all kernel and orthonormal series HAR tests; dashed: tests with standard  $t$  and  $F$  critical values.

tradeoff for WOS tests and show that the bound (1) applies to WOS tests as well.

Fourth, we derive the size-power frontier among HAR tests that have standard  $t$  and  $F$  fixed- $b$  distributions. For a 5% test with  $m = 1$ , this frontier is,

$$\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{0.3624}{T} + o(T^{-1}). \quad (2)$$

The frontier (2) is achieved by the equal-weighted periodogram (EWP) test and by the closely related equal-weighted cosine (EWC) test, in which  $\Omega$  is estimated using the Type II cosine basis functions. As can be seen in Figure 1, the cost of this restriction to  $t$  or  $F$  inference is small. For example, the power loss of the EWP test using the first four periodogram ordinates, relative to the same-sized QS test, is at most 0.0074.

The frontiers in Figure 1 are obtained under the sequence  $b = b_0 T^{-2/3}$ , while the

tradeoff for the Bartlett kernel is obtained under  $b = b_0 T^{-1/2}$ . Implementing the HAR tests studied here requires choosing  $b_0$ , which is equivalent to choosing a point on the test's size-power tradeoff curve. We briefly discuss possible criteria for determining this choice after providing our main results. A more complete treatment and recommendations for practical implementation are provided in a companion paper, Lazarus, Lewis, Stock, and Watson (LLSW, 2018). That paper proposes a loss function trading off size and power, which, using the formulas for the tradeoffs derived in this paper, is then minimized to obtain rule-of-thumb guidelines for HAR tests. LLSW also provide extensive Monte Carlo simulations (including for data-based designs), which corroborate the theoretical size-power tradeoffs and frontiers we establish in this paper.

The remainder of the paper is organized as follows. Section 2 defines the kernel and WOS estimators. Section 3 provides unified expressions for their higher-order bias and variance. Section 4 provides the main results, and Section 5 concludes. Proofs are given in the Appendix and in the Online Supplement.

## 2. MODEL, TESTS, AND LRV ESTIMATORS

We consider two-sided HAR tests of  $\beta = \beta_0$  in the Gaussian location model,

$$y_t = \beta + u_t, \quad t = 1, \dots, T, \quad (3)$$

where  $y_t$  is  $m \times 1$ ,  $\beta$  is the vector of means of  $y_t$ , and  $u_t$  is an  $m \times 1$  vector of disturbances following a stationary Gaussian process that is potentially heteroskedastic and/or autocorrelated. We consider rejection rates both under the null,  $H_{0T} : \beta = \beta_0$ , and under the local alternative,

$$H_{1T} : \beta = \beta_0 + T^{-1/2} \Omega^{1/2} \tilde{\delta}, \quad (4)$$

where  $\tilde{\delta}$  is uniformly distributed on the real  $m$ -dimensional sphere centered at the origin and with radius  $\delta$ , as in Sun (2013, 2014b).

The LRV estimator  $\hat{\Omega}$  is computed using estimated values  $\hat{z}_t = y_t - \hat{\beta} = y_t - \bar{y}$ , where  $\bar{y}$  is the sample mean of  $y_t$ . For  $m = 1$ , the  $t$ -statistic testing  $\beta = \beta_0$  is  $t_T = \sqrt{T} \bar{z}_0 / \sqrt{\hat{\Omega}}$ ,

where  $\bar{z}_0 = T^{-1} \sum_{t=1}^T z_t(\beta_0)$ ,  $z_t(\beta_0) = y_t - \beta_0$ , and  $\hat{\Omega}$  is an estimator of  $\Omega$ . For  $m > 1$ , as in Stock and Watson (2008) and Sun (2013), we consider the scaled  $F$  statistic,  $F_T^* = ((B - m + 1)/B) F_T$ , with  $F_T = T\bar{z}_0' \hat{\Omega}^{-1} \bar{z}_0 / m$  and  $B = b^{-1}$  (or its integer part). As discussed below, with this scaling,  $F_T^*$  is asymptotically distributed  $F_{m, B-m+1}$  under fixed- $b$  asymptotics when  $\hat{\Omega}$  is an equal-weighted WOS estimator.

## 2.1. KERNEL ESTIMATORS

The kernel estimator of  $\Omega$  sums the sample autocovariances, weighted by a kernel  $k$ :

$$\hat{\Omega}^{SC} = \sum_{j=-(T-1)}^{T-1} k(j/S) \hat{\Gamma}_j, \quad \text{where } \hat{\Gamma}_j = \frac{1}{T} \sum_{t=\max(1, j+1)}^{\min(T, T+j)} \hat{z}_t \hat{z}_{t-j}', \quad (5)$$

where  $S$  is the truncation parameter and the superscript “ $SC$ ” denotes sum-of-covariances.

The sum-of-covariances estimator can alternatively be computed in the frequency domain as a weighted average of the periodogram:

$$\hat{\Omega}^{WP} = 2\pi \sum_{j=-[T/2]}^{[T/2]} \tilde{w}_j I_{\hat{z}\hat{z}}(2\pi j/T), \quad (6)$$

where  $[T/2]$  denotes the integer part of  $T/2$ ,  $I_{zz}(\omega)$  is the periodogram of  $\hat{z}_t$  at frequency  $\omega$ ,  $I_{\hat{z}\hat{z}}(\omega) = (2\pi)^{-1} d_{\hat{z}}(\omega) \overline{d_{\hat{z}}(\omega)'}$  where  $d_{\hat{z}}(\omega) = T^{-1/2} \sum_{t=1}^T \hat{z}_t e^{-i\omega t}$ , and where the weights  $\{\tilde{w}_j\}$  in (6) satisfy  $\tilde{w}_j = T^{-1} \sum_{u=-(T-1)}^{T-1} k(u/S) e^{i2\pi ju/T}$ .<sup>2</sup> Kernel estimators are positive semidefinite with probability one if  $\tilde{w}_j \geq 0, j \in \mathbb{R}$ . Toward aligning  $\hat{\Omega}^{WP}$  with WOS estimators as defined below, note that (6) may be rewritten as  $\hat{\Omega}^{WP} = 4\pi \sum_{j=1}^{[T/2]} \tilde{w}_j \text{Re}(I_{\hat{z}\hat{z}}(2\pi j/T))$ .

---

<sup>2</sup> For large  $S$ ,  $\tilde{w}_j \sim 2\pi b K(2\pi j b)$ , where  $b = S/T$ , and  $K(\omega) = (2\pi)^{-1} \int_{u=-\infty}^{\infty} k(u) e^{-i\omega u} du$  is the spectral window generator; see Priestley (1981, pp. 447, 580-581) or Andrews (1991).



Three important kernel estimators are the Bartlett (Newey-West), EWP, and QS estimators. The Bartlett kernel is the tent function,  $k(x) = (1 - |x|)\mathbf{1}(|x| \leq 1)$ . The EWP estimator  $\hat{\Omega}^{EWP}$  is computed using  $\tilde{w}_j = 2B^{-1}\mathbf{1}(|j| \leq B/2)$  (the Daniell spectral kernel) in (6). The quadratic spectral estimator is so named because its weights in (6) are quadratic in  $j$ :  $\tilde{w}_j \propto \left[1 - (|j|/(B/2))^2\right]\mathbf{1}(|j| \leq B/2)$ .

## 2.2. WEIGHTED ORTHONORMAL SERIES ESTIMATORS

WOS estimators are computed by projecting  $\hat{z}_t$  onto a set of  $B$  mean-zero low-frequency orthonormal functions, typically the first mean-zero elements of a basis for  $L^2[0,1]$ , and then evaluating a weighted sum of these projections (Hannan (1970), Brillinger (1975), Priestley (1981), and Stoica and Moses (2005)). Following Sun (2013), let  $\{\phi_j(s)\}$ ,  $j = 0, \dots, B$ ,  $0 \leq s \leq 1$ , denote the first  $B+1$  functions in a real orthonormal basis for  $L^2[0,1]$ , where  $\phi_0(s) = 1$  and  $\int_0^1 \phi_j(s) ds = 0$  for  $j \geq 1$ . The WOS estimator is,

$$\hat{\Omega}^{WOS} = \sum_{j=1}^B w_j \hat{\Omega}_j^{OS}, \text{ where } \sum_{j=1}^B w_j = 1, \hat{\Omega}_j^{OS} = \hat{\Lambda}_j \hat{\Lambda}_j', \text{ and } \hat{\Lambda}_j = \sqrt{\frac{1}{T}} \sum_{t=1}^T \phi_j(t/T) \hat{z}_t. \quad (7)$$

Note that  $\hat{\Omega}^{WOS}$  omits the  $j = 0$  (constant) function since  $\hat{\Lambda}_0 = \sqrt{T} \bar{\hat{z}} = 0$ . The condition for  $\hat{\Omega}^{WOS}$  to be psd with probability one is that  $\{w_j\}$  are nonnegative.

The theory in this paper covers basis functions with two continuous and bounded derivatives. The leading case uses Fourier basis functions, for which psd WOS estimators and psd kernel estimators asymptotically coincide (see Priestley (1981, pp. 578-581)). The use of Fourier basis functions with equal weights produces the EWP estimator. Other examples of basis functions include Type II cosine basis functions and Legendre polynomials. The theory developed here extends to the batch means estimator studied by Ibragimov and Müller (2010): we show in the Online Supplement (Proposition S1) that in the location model (3), this estimator can be expressed as a WOS estimator, using what we refer to below as the split-sample (SS) basis functions.

### 3. UNIFIED EXPRESSIONS FOR BIAS, VARIANCE, AND REJECTION RATES

Our unification of expressions for the bias, variance, and higher order rejection rates of HAR kernel and WOS tests relies on what we call the implied mean kernel of WOS tests. The implied mean kernel  $k_{B,T}^{WOS}$  of  $\hat{\Omega}^{WOS}$  depends on the WOS weights and basis functions  $\{\phi_j\}$ . Using the definition of  $\hat{\Omega}_j^{OS}$  in (7) and the device in Grenander and Rosenblatt (1957, p. 125), write the mean of the  $j^{\text{th}}$  contribution to a WOS estimator as,

$$E\hat{\Omega}_j^{OS} = E\left[\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)'\right] = \sum_{u=-(T-1)}^{T-1}\tilde{k}_{j,T}^{OS}(u/T)\Gamma_u + O(1/T), \quad (8)$$

where  $\tilde{k}_{j,T}^{OS}(u/T) = T^{-1}\sum_{t=1}^T\phi_j(t/T)\phi_j((t-u)/T)\mathbf{1}(1 \leq t-u \leq T)$ . Thus,

$$E\hat{\Omega}^{WOS} = \sum_{j=1}^B E(w_j\hat{\Omega}_j) = \sum_{u=-(T-1)}^{T-1}k_{B,T}^{WOS}(u/S)\Gamma_u + O(1/T), \quad (9)$$

with  $k_{B,T}^{WOS}(u/S) = \sum_{j=1}^B w_j\tilde{k}_{j,T}^{OS}\left(B^{-1}\frac{u}{S}\right)$ , where for WOS estimators we define  $S = T/B$  so

that kernels and implied mean kernels have the same domain (cf. Priestley (1981, eq. (6.2.120)) and Brillinger (1975, eq. (5.8.6))); see the Online Supplement for details.

The  $j^{\text{th}}$  contribution to the implied mean kernel has the limit  $\lim_{T \rightarrow \infty}\tilde{k}_{jT}^{OS} = \tilde{k}_j^{OS}$ , and the implied mean kernel has the limit  $\lim_{T \rightarrow \infty}k_{B,T}^{WOS} = k_B^{WOS}$ , where

$$k_B^{WOS}(x) = \sum_{j=1}^B w_j\tilde{k}_j^{OS}(B^{-1}x) \text{ and } \tilde{k}_j^{OS}(v) = \int_{\max(0,v)}^{\min(1,1+v)}\phi_j(s)\phi_j(s-v)ds, \quad (10)$$

where the limit is pointwise holding  $B$  fixed. Note that  $k_B^{WOS}(0) = 1$ .

#### 3.1. PROPERTIES OF KERNEL AND WOS ESTIMATORS

The asymptotic bias of a kernel LRV estimator depends on the behavior of the kernel at the origin. Below we provide an analogous result for WOS estimators. Let  $k$  be a

kernel or WOS implied mean kernel. Its  $q_0^{\text{th}}$  generalized derivative at the origin is,

$$k^{(q_0)}(0) = \lim_{x \rightarrow 0} \frac{1 - k(x)}{|x|^{q_0}}. \quad (11)$$

The Parzen (1957) characteristic exponent of  $k$ , denoted by  $q$ , is the maximum integer  $q_0$  such that  $0 < k^{(q_0)}(0) < \infty$ . The Bartlett kernel has  $q = 1$ , while EWP and QS both have  $q = 2$ . For kernel estimators, a necessary condition for  $\hat{\Omega}$  to be psd with probability 1 is that  $q \leq 2$  (e.g., Priestley (1981)).

Bias also depends on the stochastic process for  $z_t$  through the behavior of its spectral density at frequency zero. Let  $s_z(\lambda)$  be the spectral density of  $z_t$  at frequency  $\lambda$ , and let  $s_z^{(q)}(0)$  be its Parzen generalized  $q^{\text{th}}$  derivative at the origin,  $s_z^{(q)}(0) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j$ . It is convenient to work with the trace of a scaled version of this generalized derivative,

$$\omega^{(q)} = \text{tr} \left( m^{-1} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j \Omega^{-1} \right), \quad (12)$$

which measures long-run persistence or anti-persistence of  $z_t$ . For  $m = 1$  and  $q = 2$ ,  $\omega^{(2)}$  is the relative curvature of the spectral density at frequency zero:  $\omega^{(2)} = -s_z''(0) / s_z(0)$ .

In addition to the “smoothing” bias indexed by  $k^{(q)}(0)$  and  $\omega^{(q)}$ , for kernel estimators bias arises from the need to estimate the mean of  $y_t$  (Hannan (1958)). This “demeaning” bias depends on the kernel’s asymptotic mean. We show below, as in Sun (2011), that no such demeaning bias arises for WOS estimators, as  $\int_0^1 \phi_j(s) ds = 0$ . Accordingly, define

$$\mu = \begin{cases} \int_{-\infty}^{\infty} k(x) dx & \text{for kernel estimators,} \\ \sum_{j=1}^B w_j \int_0^1 \phi_j(s) ds = 0 & \text{for WOS estimators.} \end{cases} \quad (13)$$

If  $z_t$  is Gaussian, then both kernel and WOS LRV estimators are distributed as

weighted averages of independent chi-squared random variables.<sup>3</sup> For kernel estimators and scalar processes, Tukey (1950) proposed approximating this mixture distribution by a chi-squared with “equivalent degrees of freedom”  $\nu$  chosen to match the estimator’s asymptotic variance. Tukey’s approximation, extended to include WOS estimators, is

$$\hat{\Omega} \sim \left(\chi_\nu^2 / \nu\right)\Omega, \text{ where } \nu = (b\psi)^{-1} \text{ and } \psi = \begin{cases} \int_{-\infty}^{\infty} k^2(x)dx & \text{for kernel estimators,} \\ B \sum_{j=1}^B w_j^2 & \text{for WOS estimators,} \end{cases} \quad (14)$$

where we set  $b = B^{-1}$  for WOS estimators.

For equal-weighted WOS estimators with  $m = 1$ , the approximation (14) is asymptotically exact, with  $\nu = B$ , for fixed  $B$  and  $T \rightarrow \infty$  (e.g., Brillinger (1975), Phillips (2005)). This extends straightforwardly to the vector case,  $m > 1$ , as

$\hat{\Omega}^{WOS} \xrightarrow{d} \Omega^{1/2} (\Xi_B / B) \Omega^{1/2'}$ , where  $\Xi_B$  follows a standard  $m$ -dimensional Wishart

distribution with  $B$  degrees of freedom (e.g., Sun (2011)). Given  $\int_0^1 \phi_j(s)ds = 0$  for  $j \geq 1$ ,

the equal-weighted estimator is asymptotically independent of  $\bar{z}_0$ , and thus the equal-

weighted WOS test  $F_T^*$  is asymptotically distributed  $F_{B,m-B+1}$ . Among the class of kernel and WOS tests we consider, this property holds only for equal-weighted WOS tests, and the fixed- $b$  limiting distribution and critical values for  $F_T^*$  are in general nonstandard.

### 3.2. RESULTS ON BIAS, VARIANCE, AND REJECTION RATES

We now present unified expressions for the bias, variance, and rejection rates of kernel and WOS LRV estimators and HAR tests. Henceforth, let  $k$  denote either a kernel or an implied mean kernel. For kernel estimators,  $b = S/T$  and for WOS estimators,  $b = B^{-1}$ .

We make the following assumptions:

---

<sup>3</sup> Without Gaussianity, this holds asymptotically under fixed- $b$  asymptotics (Sun (2013, 2014b)).

Assumption 1 (Gaussian stochastic processes).  $z_t$  is a stationary Gaussian process generated according to the multivariate location model (3), with spectral density matrix  $s_z(\lambda)$  that is positive definite in a neighborhood around  $\lambda = 0$  and autocovariances  $\Gamma_u$  that satisfy  $\sum_{u=-\infty}^{\infty} |u|^r |\Gamma_u| < \infty$  for  $r \in [0, 2 + \zeta]$ , for some  $\zeta > 0$ .

Assumption 2 (kernels). For a kernel LRV estimator, the kernel  $k(x) : \mathbb{R} \rightarrow [-1, 1]$  is continuous, piecewise continuously differentiable, satisfies  $k(x) = k(-x)$ ,  $k(0) = 1$ ,  $\int_{-\infty}^{\infty} |x| k(x) dx < \infty$ , has frequency-domain weights  $\{\tilde{w}_j\}$  in (6) satisfying  $\tilde{w}_j \geq 0$ ,  $j \in \mathbb{R}$ , and has Parzen characteristic exponent  $q = 1$  or  $2$ .

Assumption 3 (WOS). For a WOS LRV estimator, for  $j = 1, \dots, B$ , the orthonormal series  $\phi_j \in L^2[0, 1]$  satisfy  $\int_0^1 \phi_j(s) ds = 0$  for  $j \geq 1$  and have two continuous derivatives, such that the  $n^{\text{th}}$  derivative  $\phi_j^{(n)}(s)$  satisfies  $\sup_{s \in [0, 1]} |\phi_j^{(n)}(s)| \leq C_{n, \phi} j^{2n+1/2}$  for some constant  $C_{n, \phi}$  for all  $j$  and  $n = 0, 1, 2$ . The weights  $w_j \geq 0$  are  $O(B^{-1})$  and satisfy  $\sum_{j=1}^B w_j = 1$ .

Assumption 4 (rates). The sequence  $b$  is assumed to satisfy  $b^q T^{q-1} + (bT)^{-1} \rightarrow 0$ .

These assumptions are the same as or modifications of those of Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2013, 2014b). Assumption 1 states the model and provides conditions under which the bias expressions and fixed- $b$  distributions hold, and it implies that  $\omega^{(q)}$  is finite for  $q \leq 2$ . Assumption 2 states standard conditions on psd kernel estimators. Assumption 3 strengthens slightly the conditions in Sun's (2013) Assumption 3.1 so that the orthonormal series have two derivatives, each of the order  $j^{2n+1/2}$ . Bases that satisfy this condition include Fourier, Type II cosine, and Legendre polynomials, as shown in Online Supplement Proposition S2. Assumption 4 strengthens the corresponding condition in Sun, Phillips, and Jin (2008), who require

$b + (bT)^{-1} \rightarrow 0$ . The more restrictive rate condition in Assumption 4 is used to express the limiting results for WOS tests in terms of the implied mean kernel when  $q = 2$ .

Theorem 1 collects expansions for kernel and equal-weighted WOS estimators in Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2013, 2014b) (among others) and extends them to include general WOS estimators.

**THEOREM 1.** Under Assumptions 1–4,

(i) The asymptotic bias of kernel and WOS LRV estimators is,

$$E\hat{\Omega} - \Omega = -2\pi (bT)^{-q} k^{(q)}(0) s_z^{(q)}(0) - b\mu\Omega + o(b) + o((bT)^{-q}). \quad (15)$$

(ii) The first two generalized derivatives of the WOS implied mean kernel are,

$$k^{(1)}(0) = \lim_{B \rightarrow \infty} \frac{1}{B} \sum_{j=1}^B w_j [\phi_j(0)^2 + \phi_j(1)^2] / 2, \text{ and}$$

$$k^{(2)}(0) = -\lim_{B \rightarrow \infty} \frac{1}{B^2} \sum_{j=1}^B w_j \int_0^1 \phi_j(s) \phi_j''(s) ds / 2. \quad (16)$$

If  $k^{(1)}(0) \neq 0$ , then  $q = 1$ ; otherwise,  $q = 2$ .

(iii) The asymptotic variance of kernel and WOS LRV estimators is,

$$\text{var}(\text{vec } \hat{\Omega}) = v^{-1} (I_{m^2} + K_{mm}) \Omega \otimes \Omega + o(b), \quad (17)$$

where  $K_{mm}$  is the  $m^2 \times m^2$  commutation matrix and  $\otimes$  is the Kronecker product.

(iv) Let  $c_m^\alpha(b)$  denote the fixed- $b$  asymptotic critical value for the level  $\alpha$  test with  $m$  degrees of freedom. The asymptotic expansion of the null rejection rate is,

$$\Pr_0 [F_T^* > c_m^\alpha(b)] = \alpha + G'_m(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} + o(b) + o((bT)^{-q}), \quad (18)$$

where  $G_m$  is the chi-squared cdf with  $m$  degrees of freedom,  $G'_m$  is the first derivative of  $G_m$ , and  $\chi_m^\alpha$  is the  $1-\alpha$  quantile of  $G_m$ .

(v) The rejection rate against the local alternative (4) using the fixed- $b$  critical value has the expansion,

$$\Pr_{\delta} \left[ F_T^* > c_m^{\alpha}(b) \right] = \left[ 1 - G_{m, \delta^2}(\chi_m^{\alpha}) \right] + G'_{m, \delta^2}(\chi_m^{\alpha}) \chi_m^{\alpha} \omega^{(q)} k^{(q)}(0) (bT)^{-q} - \frac{1}{2} \delta^2 G'_{m+2, \delta^2}(\chi_m^{\alpha}) \chi_m^{\alpha} \nu^{-1} + o(b) + o((bT)^{-q}), \quad (19)$$

where  $G_{m, \delta^2}$  is the noncentral chi-squared cdf with  $m$  degrees of freedom and noncentrality parameter  $\delta^2$  and  $G'_{m, \delta^2}$  is its first derivative.

- (vi) The expansions in (18) and (19) also hold for the split-sample (SS) series estimator, for which  $q = 1$ , although it does not satisfy Assumption 3.

The term in  $(bT)^{-q}$  in the null rejection rate expansion (18) arises from the bias of the LRV estimator. Under the local alternative, the rejection rate expansion (19) depends both on bias (the first term) and on its variance through the term in  $\nu^{-1}$ . This latter term is the power loss analogous to that from using a  $t$  distribution in the i.i.d. location model because the variance is estimated, not known.

#### 4. SIZE-POWER TRADEOFFS AND THE SIZE-POWER FRONTIER

This section uses the expansions in Theorem 1 to characterize the size-power tradeoff, the size-power frontier, and optimality results for kernel and WOS HAR tests evaluated using fixed- $b$  critical values. Section 4.1 provides our results, which are discussed in detail in Section 4.2. Proofs are provided in the Appendix.

##### 4.1. MAIN RESULTS

Assume throughout that Assumptions 1–4 hold.

**THEOREM 2.** Let  $c_{m,T}^{\alpha}(b)$  be the size-adjusted fixed- $b$  critical value,

$$c_{m,T}^{\alpha}(b) = \left[ 1 + \omega^{(q)} k^{(q)}(0) (bT)^{-q} \right] c_m^{\alpha}(b). \quad (20)$$

Then  $\Pr_0[F_T^* > c_{m,T}^\alpha(b)] = \alpha + o(b) + o((bT)^{-q})$  and the higher order size-adjusted power of the test is,

$$\Pr_\delta[F_T^* > c_{m,T}^\alpha(b)] = \left[1 - G_{m,\delta^2}(\chi_m^\alpha)\right] - \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha)\chi_m^\alpha \nu^{-1} + o(b) + o((bT)^{-q}). \quad (21)$$

**THEOREM 3.** Consider two HAR test statistics  $F_{1,T}^*$  and  $F_{2,T}^*$  based on different kernels or implied mean kernels with the same value of  $q$ , with equivalent degrees of freedom respectively given by  $\nu_1$  and  $\nu_2$ , and with fixed- $b$  critical values respectively given by  $c_{1,m}^\alpha(b_1)$  and  $c_{2,m}^\alpha(b_2)$ . Choose sequences  $b_1$  and  $b_2$  meeting Assumption 4 such that  $F_{1,T}^*$  and  $F_{2,T}^*$  have the same higher-order size. Then the difference between their higher-order rejection rates under the local alternative indexed by  $\delta$  is,

$$\begin{aligned} \Pr_\delta[F_{1T}^* > c_{1,m}^\alpha(b_1)] - \Pr_\delta[F_{2T}^* > c_{2,m}^\alpha(b_2)] &= \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha)\chi_m^\alpha (\nu_2^{-1} - \nu_1^{-1}) \\ &\quad + o(b_1) + o((b_1T)^{-q}) + o(b_2) + o((b_2T)^{-q}). \end{aligned} \quad (22)$$

Our main results concern the tradeoff between size and size-adjusted power. The size distortion  $\Delta_S$  of the candidate test is,

$$\Delta_S = \Pr_0[F_T^* > c_m^\alpha(b)] - \alpha. \quad (23)$$

The power of the oracle test, in which  $\Omega$  is known, is  $1 - G_{m,\delta^2}(\chi_m^\alpha)$ . Let  $\Delta_P(\delta)$  denote the power loss of the candidate test, compared to the oracle test, under the local alternative indexed by  $\delta$ , and let  $\Delta_P^{\max}$  denote the maximum such power loss, so  $\Delta_P^{\max}$  is the maximum gap between the power curves of the oracle test and the candidate test:

$$\Delta_P(\delta) = \left[1 - G_{m,\delta^2}(\chi_m^\alpha)\right] - \Pr_\delta[F_T^* > c_{m,T}^\alpha(b)], \text{ and} \quad (24)$$

$$\Delta_P^{\max} = \sup_\delta \Delta_P(\delta). \quad (25)$$

Because  $\nu = (b\nu)^{-1}$ , equations (18) and (21) constitute a pair of parametric equations that determine  $\Delta_S$  and  $\Delta_P$  for a given  $b$ . Both expressions are monotonic in  $b$ , so



eliminating  $b$  yields the higher-order tradeoff between the size and power of a given test. Requiring that  $\Delta_S$  and  $\Delta_P$  maintain the same asymptotic order further restricts the rate of the sequence  $b$ ; Corollary 1 provides that restriction, which satisfies Assumption 4.<sup>4</sup> Theorem 4 then provides the higher-order tradeoff between size and power. The envelope of these tradeoffs, provided in Theorem 5, is the size-power frontier.

**COROLLARY 1.**  $\Delta_P(\delta)$  and  $\Delta_S$  are of the same asymptotic order if and only if  $b = O\left(T^{-q/(q+1)}\right)$  and  $T^{-q/(q+1)} = O(b)$ .

**THEOREM 4.** For a given HAR test evaluated using fixed- $b$  critical values, under the sequence for  $b$  in Corollary 1:

- (i) The small- $b$  asymptotic tradeoff between the size distortion and the power loss against the local alternative indexed by  $\delta$  is,

$$T\Delta_P(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta)\ell^{(q)}(k)|\omega^{(q)}|^{1/q} + o(1), \quad (26)$$

where  $a_{m,\alpha,q}(\delta) = \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha)\chi_m^\alpha(G'_m(\chi_m^\alpha)\chi_m^\alpha)^{1/q}$  and  $\ell^{(q)}(k) = (k^{(q)}(0))^{1/q}\psi$ .

- (ii) Let  $\bar{a}_{m,\alpha,q} = \sup_\delta a_{m,\alpha,q}(\delta)$ . The small- $b$  asymptotic size-power tradeoff is,

$$T\Delta_P^{\max}|\Delta_S|^{1/q} = \bar{a}_{m,\alpha,q}\ell^{(q)}(k)|\omega^{(q)}|^{1/q} + o(1). \quad (27)$$

- (iii) The size-power tradeoffs of tests based on LRV estimators with Parzen characteristic exponent  $q = 2$  asymptotically dominate the tradeoffs for tests with  $q = 1$ , both within and across the two families of tests.

**THEOREM 5.**

- (i) For psd kernel and WOS HAR tests evaluated using fixed- $b$  critical values, under the

---

<sup>4</sup> Equating the order of  $\Delta_S$  and  $\Delta_P$  is desirable as long as one places non-vanishing weight on both size and power in assessing their tradeoff; see LLSW (2018) for further discussion.

sequence for  $b$  in Corollary 1,

$$T\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{3\pi\sqrt{10}}{25} \bar{a}_{m,\alpha,2} + o(1), \quad (28)$$

where  $\bar{a}_{m,\alpha,2}$  is defined in Theorem 4. This frontier is achieved by the QS kernel. For tests with  $\alpha = .05$ ,  $\bar{a}_{m,\alpha,2} 3\pi\sqrt{10} / 25 \approx 0.3368$  for  $m = 1$  (yielding (1)),

$\bar{a}_{m,\alpha,2} 3\pi\sqrt{10} / 25 \approx 0.6460$  for  $m = 2$ , and  $\bar{a}_{m,\alpha,2} 3\pi\sqrt{10} / 25 \approx 0.9491$  for  $m = 3$ .

(ii) For psd kernel and WOS HAR tests with exact  $t$  and  $F$  asymptotic fixed- $b$  distributions and critical values, under the sequence for  $b$  in Corollary 1,

$$T\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{\pi}{\sqrt{6}} \bar{a}_{m,\alpha,2} + o(1) \text{ (exact } t \text{ or } F \text{ critical values)}. \quad (29)$$

This frontier is achieved by the EWP test. For  $\alpha = .05$ ,  $\bar{a}_{m,\alpha,2} \pi / \sqrt{6} \approx 0.3624$  for  $m = 1$  (yielding (2)),  $\approx 0.6950$  for  $m = 2$ , and  $\approx 1.0211$  for  $m = 3$ .

## 4.2. REMARKS

1. For a given  $\alpha$  and  $m$ , the testing frontier depends only on the sample size and the average normalized curvature of the spectral density at frequency zero. As a result, the scaled fixed- $b$  frontier plotted in Figure 1 applies universally to all psd kernel and weighted orthonormal series HAR tests evaluated using fixed- $b$  critical values at the optimal rates in Corollary 1.
2. The rate for  $b$  in Corollary 1,  $b = O\left(T^{-q/(q+1)}\right)$ , is the same rate found by Sun, Phillips, and Jin (2008) and Sun (2014b) to minimize a weighted average of type I and type II testing errors in the case that  $\Delta_S > 0$ . Although we derive the frontier only for this sequence, we conjecture that it holds more generally. This conjecture is supported by the generally good ability of the frontier to describe simulation results (LLSW (2018)). This conjecture could be proven by strengthening remainder terms in  $o(b)$  and  $o((bT)^{-q})$  in the underlying Edgeworth expansions to  $O$  of a somewhat higher

order; doing so is left for future work.

3. For kernel tests, the size-power frontier is obtained first by noting that the frontier for  $q = 2$  tests asymptotically dominates the frontier for  $q = 1$  tests, then by minimizing, over  $q = 2$  kernels, the expression  $\ell^{(2)}(k) = \sqrt{k^{(2)}(0)} \int_{-\infty}^{\infty} k^2(x) dx$ . This quantity is minimized by the Epanechnikov/QS kernel (Epanechnikov (1969)). The quantity  $\ell^{(2)}(k)$  has a long history in spectral density estimation. Priestley (1981, Section 7.3.2) dates it to Grenander's (1951) uncertainty principle for spectral estimation: as summarized by Priestley, "bias and variance are antagonistic." In our application, bias produces a size distortion while variance degrades size-adjusted power.

Our results indicate that Grenander's uncertainty principle extends beyond the minimal-MSE spectral density estimation problem. In addition to the size-power tradeoff in Theorem 4, the following objective functions depend on the (implied mean) kernel only through  $\ell^{(2)}(k)$  when evaluated using the optimal  $b$  for  $q = 2$  (where a scalar process,  $m = 1$ , is assumed for simplicity):

- (a) An objective function for the spectral estimation problem (given known  $\beta$ ) that minimizes  $MSE(\hat{s}_z(0)) = \text{bias}^2(\hat{s}_z(0)) + \text{var}(\hat{s}_z(0))$ ;
- (b) The previous objective function modified to  $a|\text{bias}(\hat{s}_z(0))| + (1-a)\text{var}(\hat{s}_z(0))$  with  $0 < a < 1$ ;
- (c) An objective function for the HAR testing problem that minimizes size distortions plus power, specifically  $a|\Delta_s| + (1-a)\Delta_p^{\max}$  or alternatively  $a|\Delta_s| + (1-a)\int \Delta_p(\delta) d\Pi_\delta(\delta)$ , where  $a$  is a weight  $0 < a < 1$  and where  $\Pi_\delta$  is a density function over the noncentrality parameter  $\delta$ ;
- (d) A quadratic version of the previous objective function,  $a(\Delta_s)^2 + (1-a)(\Delta_p^{\max})^2$ ;
- (e) The objective function considered by Sun, Phillips, and Jin (2008) that minimizes the weighted average of the type I and type II error.

Minimizing (a) is the classic problem of optimal spectral estimation; its optimum is

achieved at a rate  $b = O(T^{-2q/(2q+1)})$  converging to zero faster than the testing-optimal rate in Corollary 1 (which is optimal for the remaining objective functions). Objective function (b) is not of primitive interest, but (c) and (e) reduce to (b). The objective function (c), which trades off size distortion and power loss linearly, also depends on the kernel solely through  $\ell^{(2)}(k)$  under the optimal  $b$ . Minimizing (d) does the same with quadratic loss and is the approach used by LLSW (2018). Objective function (e) differs from (c) because the type II error is not size-adjusted, yet its minimal value also depends on the kernel only through  $\ell^{(2)}(k)$  (see Sun and Yang (2020) and LLSW (2018, rejoinder). Each of these objective functions is minimized by the QS kernel, or, among equal-weighted WOS estimators, by the EWP estimator. See Proposition S3 in the Online Supplement for derivations.<sup>5</sup>

4. Obtaining the frontier in Theorem 5(i) requires proving the optimality of the QS kernel relative to all WOS tests, not just among kernel tests. The proof of this new optimality result extends Grenander and Rosenblatt's (1957, Section 4.2) result on the optimality of "spectrograph" (i.e., weighted Fourier series) estimators to show that, for WOS tests, the Fourier basis functions are optimal. This extension also implies that the EWP test achieves the restricted frontier provided in Theorem 5(ii), as exact  $t$  or  $F$  inference obtains only among equal-weighted WOS tests.
5. The price paid for the convenience of exact  $t$  or  $F$  fixed- $b$  critical values can be computed from Theorem 3. Let  $v_{EWP} = B$ . It is shown in the Online Supplement that, for EWP and QS tests with the same higher-order size,

$$\begin{aligned} & \Pr_{\delta}[F_{QS,T}^* > c_{QS,\alpha}(b_{QS})] - \Pr_{\delta}[F_{EWP,T}^* > c_{EWP,\alpha}(b_{EWP})] \\ & \approx \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha (v_{EWP}^{-1} - v_{QS}^{-1}) = \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \left(1 - \frac{6\sqrt{3}}{5\sqrt{5}}\right) B^{-1}. \quad (30) \end{aligned}$$

Table I reports the maximum higher-order power loss from using EWP over all

---

<sup>5</sup> We thank a referee and Yixao Sun for pointing out that our results extend to (c) and (e), respectively.

TABLE I. Maximum power loss of same-sized EWP (with  $B$  series) compared to QS.

$m$	$B = 4$	$B = 8$	$B = 16$
1	0.0147	0.0074	0.0037
2	0.0247	0.0123	0.0062
3	0.0335	0.0168	0.0084
4	0.0419	0.0209	0.0105

*Note:*  $b$  for QS is chosen so that its higher order size is the same as EWP.

alternatives  $\delta$ . The cost of using EWP relative to QS is small: for  $B = 8$  and  $m = 1$ , the maximum equivalent-size power gap is 0.0074 over all alternatives. This explains the numerical finding in Kiefer and Vogelsang (2005) that the local asymptotic power curves for these two tests are very close. Figure S3 in the Online Supplement plots the final expression in (30) as a function of  $\delta$  for various values of  $B$  and  $m = 1$ .

6. While Theorem 5 provides results on optimal kernel choice, our framework also allows us to rank any two HAR tests using their asymptotic size-power tradeoffs from Theorem 4. For example, as shown in Proposition S5 in the Online Supplement, the Bartlett kernel dominates the equal-weighted split-sample WOS estimator (both of which have  $q = 1$ ), as the Bartlett small- $b$  size-power tradeoff curve is strictly below the SS tradeoff curve. We also find that, among  $q = 2$  equal-weighted WOS tests, the tradeoff for the EWP test (i.e., using Fourier basis functions) is asymptotically equivalent to that obtained using Type II cosine basis functions as proposed by Müller (2007); see LLSW (2018) for further discussion.
7. The tradeoffs in Theorem 4 are expressed in terms of absolute size distortions. For processes with  $s_z''(0) < 0$  (loosely, positive serial correlation), the HAR tests are oversized and the tradeoff is between size and power. Positive serial correlation is common in practice, e.g. in multiperiod return regressions and multistep-ahead forecasts. In the negative serial correlation case (specifically,  $s_z''(0) > 0$ ), the HAR test is undersized. If our size-power tradeoffs are used to construct truncation

parameter rules, one might therefore want to treat these two cases separately. For example, Sun, Phillips, and Jin (2008) consider a pretest approach that distinguishes between these two cases based on the sign of a preliminary estimate of  $s_z''(0)$ , and their approach could be extended to our framework, where the size-power tradeoff is used to obtain a truncation parameter rule in the positive serial correlation case. For additional discussion, see LLSW (2018).

8. The multivariate results focus on inference on all  $m$  elements of  $\beta$ . The question arises as to whether they extend to inference on only  $m' < m$  of those parameters or, more generally, to inference on  $m' < m$  linear combinations of those parameters.

Accordingly, consider the null hypothesis  $R\beta = \tilde{\beta}_0$ , where  $R$  is  $m' \times m$  and  $\tilde{\beta}_0$  is  $m' \times 1$ . The  $F$ -statistic testing this hypothesis is  $T(R\bar{z}_0)'(R\hat{\Omega}R')^{-1}(R\bar{z}_0)/m'$ , where  $R\bar{z}_0 = T^{-1} \sum_{t=1}^T (Rz_t - \tilde{\beta}_0)$ . Because all the estimators of  $\Omega$  we consider are quadratic forms in  $\hat{z}$ , this  $F$  statistic testing  $R\beta = \tilde{\beta}_0$  is equivalent to the usual  $F$  statistic testing a full vector hypothesis (i.e.,  $T\bar{z}_0'\hat{\Omega}^{-1}\bar{z}_0/m$ ), but computed using the  $m' \times 1$  vector of transformed data  $Ry_t$ . Thus, the results for full vector inference apply directly to subvector inference.

#### 4.3. TESTS WITH UNIFORM SIZE CONTROL

The foregoing results, like most of the HAR literature, consider the performance of tests pointwise in the nuisance parameter  $\omega^{(q)}$ . An alternative approach is to consider controlling the rejection rate uniformly over a region of  $\omega^{(q)}$ , in particular for all  $\omega^{(q)}$  less than some finite upper bound  $\bar{\omega}^{(q)}$ , and choosing the test that maximizes weighted average power among those that control size uniformly over  $|\omega^{(q)}| \leq \bar{\omega}^{(q)}$ . This uniform-size-control approach follows a small HAR literature developed by Müller (2007, 2014), Preinerstorfer and Pötscher (2016), and Pötscher and Preinerstorfer (2017). The calculations here differ from earlier work by restricting the space of nuisance parameters

to be a closed subset representing moderate (bounded) persistence.

Uniform size control can be achieved for any sequence  $b \propto T^{-q/(1+q)}$  by using the size-adjusted critical value corresponding to the worst-case (least favorable) value of the nuisance parameter. It can be seen from (18) and (23) that the higher-order size distortion is increasing in  $\omega^{(q)}$ , so the least favorable value of the nuisance parameter is the maximum  $\bar{\omega}^{(q)}$ . The size-adjusted critical value (20), evaluated using this least favorable value, therefore results in a test that controls size uniformly to higher order under the condition that the remainder in (18) is of the stated order uniformly over  $|\omega^{(q)}| \leq \bar{\omega}^{(q)}$ .

As an illustration, we derive the maximum weighted average power (WAP) test for the case where  $z_t$  follows an AR(1) with coefficient  $\rho$ . First, given a kernel or WOS test, choose  $b$  to maximize the WAP among tests using the size-adjusted critical value (20) with  $\omega^{(q)} = \bar{\omega}^{(q)}$ :

$$b^{WAP} = \arg \max_b \int \int_{\delta \mid \rho \leq \bar{\rho}} \Delta_p(\omega^{(q)}(\rho), \delta) d\Pi_\rho(\rho) d\Pi_\delta(\delta), \quad (31)$$

where  $\Delta_p(\omega^{(q)}(\rho), \delta) = \frac{1}{2} \delta^2 G'_{m+2, \delta^2}(\chi_m^\alpha) \chi_m^\alpha \nu^{-1} + G'_{m, \delta^2}(\chi_m^\alpha) \chi_m^\alpha [\bar{\omega}^{(q)} - \omega^{(q)}(\rho)] k^{(q)}(0) (bT)^{-q}$ ,

$\omega^{(1)}(\rho) = 2\rho / (1 - \rho^2)$ ,  $\omega^{(2)}(\rho) = 2\rho / (1 - \rho)^2$ ,  $\bar{\rho} = \max \rho$  s.t.  $\omega^{(q)}(\rho) \leq \bar{\omega}^{(q)}$ , and the weight functions  $\Pi_\rho$  and  $\Pi_\delta$  are independent and each integrate to one. The solution to (31), as shown in Proposition S6 in the Online Supplement, is

$$b^{WAP} = q^{\frac{1}{1+q}} \tilde{d}_{m, \alpha, q} \left( \frac{k^{(q)}(0)}{\psi} \right)^{\frac{1}{1+q}} \left( \tilde{\omega}^{(q)} \right)^{\frac{1}{1+q}} T^{\frac{-q}{1+q}}, \quad (32)$$

where expressions for the constants  $\tilde{\omega}^{(q)}$  and  $\tilde{d}_{m, \alpha, q}$  are provided in the Supplement with the derivation of the result. We can see immediately that  $b^{WAP}$  declines with  $T$  at the same rate as given in Corollary 1. Further (again see the Online Supplement), the power loss of the test using the WAP-maximizing sequence (32) depends on  $k$  only through  $\ell^{(q)}(k)$ .

Once again, the term in Grenander's (1951) uncertainty principle appears, and the test asymptotically delivering the highest WAP uses the QS kernel, with a numerically small cost to using EWP. Further,  $q = 1$  kernels are again asymptotically dominated by  $q = 2$

kernels. Thus, the main qualitative findings from the pointwise analysis carry through to uniform-size-control, maximum-WAP tests.

#### 4.4. MONTE CARLO SIMULATIONS

We conducted extensive Monte Carlo simulations to assess the accuracy of the asymptotic tradeoffs and frontiers. Results are reported in Section S1 of the Online Supplement and in LLSW (2018). We draw three overall conclusions. First, the theoretical tradeoff (27) provides a good description of finite-sample test performance in the Gaussian location model. The fit is better for  $q = 2$  kernels than  $q = 1$ . Second, consistent with the theory, the performance of  $q = 2$  kernels is superior to that of  $q = 1$  kernels for sufficiently large sample sizes; however, for persistent processes with small  $T$ , some  $q = 1$  kernels (such as the Bartlett kernel) have size-power tradeoffs that cross the  $q = 2$  frontier, both in theory and in simulations. Third, we also examined the regression case with stochastic regressors. In this case,  $z_t$  is non-Gaussian even if the error term is Gaussian, so Assumption 1(i) does not hold. Still, the Monte Carlo tradeoffs and rankings across tests accord qualitatively (although not quantitatively) with the theoretical results for the Gaussian location model. Further results using designs constructed to match relevant empirical settings, reported in LLSW (2018), accord with these findings as well.

#### 5. DISCUSSION AND CONCLUSIONS

The size-power tradeoff and frontier are obtained under the sequence,  $b = b_0 T^{-q/(q+1)}$ , which equates the order of the size distortion and power loss. In practice, one needs to know the coefficient  $b_0$ . While one might be tempted to select  $b_0$  to maximize size-adjusted power, doing so would lead to a corner solution with the smallest possible  $b_0$  while respecting the optimal sequence. That choice would have large size distortions, necessitating feasible size-adjusted critical values. In simulations, however, we find that feasible size adjustment (implemented according to (20), with an estimator of  $\omega^{(q)}$ )



works poorly in sample sizes typically encountered; this is perhaps unsurprising because feasible size adjustment replaces the difficult problem of estimating the spectral density at frequency zero with the more difficult problem of estimating its curvature. Choosing a point on the frontier thus requires a judgement by the user. The approach of Section 4.3 is to specify a region over which one requires uniform size control; this region, along with the kernel, determines the constants in (32) and thus  $b_0$ . A second approach, explored in depth in LLSW (2018), is to specify a loss function trading off size and size-adjusted power. A third, related approach (discussed in Remark 3) is to specify that loss function in terms of the Type I and Type II error as in Sun, Phillips, and Jin (2008). Our results on optimal kernel choice apply to all three approaches.

Our results suggest directions for additional research. First, the WOS and kernel estimators are both contained in the larger family of quadratic estimators (e.g., Müller (2007), Sun (2014a)), and we conjecture that our frontier applies to that larger class. Second, we do not consider bootstrap tests. Results in Gonçalves and Vogelsang (2011) suggest that tests with critical values from the moving block bootstrap might also satisfy our size/power tradeoff expressions and the frontiers (1) and (2); an open question is whether bootstrap tests using QS or EWP kernels achieve those frontiers. Third, for certain processes and sample sizes, the tradeoffs for  $q = 1$  and  $q = 2$  kernels cross (LLSW (2018)), raising the question of whether one can improve upon the Bartlett kernel among  $q = 1$  kernels, a topic taken up in Kolokotronis and Stock (2019). Fourth, additional theoretical work on the regression model with stochastic regressors is in order.

## REFERENCES

- Andrews, D.W. K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- Brillinger, D.R. (1975). *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.
- Clark, T.E. and K.D. West (2007). "Approximately Normal Tests for Equal Predictive

- Accuracy in Nested Models,” *Journal of Econometrics* 138, 291-311.
- den Haan, W.J. and A. Levin (1997). “A Practitioner’s Guide to Robust Covariance Matrix Estimation,” *Handbook of Statistics* 15, ch. 12, 291-341.
- Epanechnikov, V.A. (1969). “Non-Parametric Estimation of a Multivariate Probability Density,” *Theory of Probability & Its Applications* 14, 153-158.
- Gonçalves, S. and T. Vogelsang (2011). “Block Bootstrap HAC Robust Tests: The Sophistication of the Naïve Bootstrap,” *Econometric Theory* 27, 745-791.
- Grenander, U. (1951). “On Empirical Spectral Analysis of Stochastic Processes,” *Arkiv För Matematik*, 1, 503-530.
- Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: John Wiley and Sons.
- Hannan, E.J. (1958). “The Estimation of the Spectral Density After Trend Removal,” *Journal of the Royal Statistical Society. Series B (Methodological)* 20, 323-333.
- Hannan, E.J. (1970). *Multiple Time Series*. New York: John Wiley and Sons.
- Ibragimov, R. and Müller, U.K. (2010). “ $t$ -statistic based correlation and heterogeneity robust inference,” *Journal of Business and Economic Statistics* 28, 453-468.
- Jansson, M. (2004). “The Error in Rejection Probability of Simple Autocorrelation Robust Tests,” *Econometrica*, 72, 937-946.
- Kiefer, N.M. and T.J. Vogelsang (2002). “Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size,” *Econometric Theory* 18, 1350-1366.
- Kiefer, N.M. and T.J. Vogelsang (2005). “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *Econometric Theory*, 21, 2093-2095.
- Kolokotronis, T. and J.H. Stock (2019), “Is Newey-West Optimal Among  $q = 1$  Kernels?,” manuscript, Harvard University.
- Lazarus, E., D.J. Lewis, J.H. Stock, and M.W. Watson (2018). “HAR Inference: Recommendations for Practice” (with discussion and rejoinder). *Journal of Business and Economic Statistics* 36, 541-575.
- Long, J.S. and L.H. Ervin (2000). “Using Heteroskedasticity Consistent Standard Errors

- in the Linear Regression Model,” *The American Statistician* 54, 217-224.
- Müller, U. (2007). “A Theory of Robust Long-Run Variance Estimation,” *Journal of Econometrics*, 141, 1331-1352.
- Müller, U. (2014). “HAC Corrections for Strongly Autocorrelated Time Series,” *Journal of Business and Economic Statistics* 32, 311-322.
- Newey, W.K. and K.D. West (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* 55, 703-708.
- Newey, W.K. and K.D. West (1994). “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies* 61, 631-653.
- Ng, S. and P. Perron (2001). “Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power,” *Econometrica* 69, 1519-1554.
- Parzen, E. (1957). “On Consistent Estimates of the Spectrum of a Stationary Time Series,” *Annals of Mathematical Statistics*, 28, 329-348.
- Phillips, P. C. B. (2005). “HAC Estimation by Automated Regression,” *Econometric Theory* 21, 116–142.
- Pötscher, B. and D. Preinerstorfer (2017). “Controlling the Size of Autocorrelation Robust Tests,” *Journal of Econometrics* 207, 406-431.
- Preinerstorfer, D. and B. Pötscher (2016). “On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests,” *Econometric Theory* 32, 261-358.
- Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.
- Rothenberg, T.G. (1984). “Approximating the Distributions of Econometric Estimators and Test Statistics,” Ch. 15 in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics, Volume II*. Amsterdam: Elsevier, 881-935.
- Stock, J.H. and M.W. Watson (2008), “Heteroskedasticity-Robust Standard Errors for Fixed-Effect Panel Data Regression,” *Econometrica* 76, 155-174.
- Stoica, P. and R. Moses (2005), *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice Hall.
- Sun, Y. (2011). “Autocorrelation Robust Trend Inference with Series Variance Estimator

- and Testing-Optimal Smoothing Parameter,” *Journal of Econometrics* 164, 345-366.
- Sun, Y. (2013). “Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator,” *The Econometrics Journal*, 16, 1–26.
- Sun, Y. (2014a). “Fixed-Smoothing Asymptotics in a Two-Step Generalized Method of Moments Framework,” *Econometrica* 82, 2327-2370.
- Sun, Y. (2014b). “Let’s Fix It: Fixed- $b$  Asymptotics versus Small- $b$  Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference,” *Journal of Econometrics* 178, 659-677.
- Sun Y., P.C.B. Phillips, and S. Jin (2008), “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76(1): 175-194.
- Sun, Y. and J. Yang (2020). “Testing-optimal Kernel Choice in HAR Inference,” manuscript, University of California San Diego.
- Tukey, J.W. (1950). “The Sampling Theory of Power Spectrum Estimates,” in *Symposium on Applications of Autocorrelation Analysis to Physical Problems, Woods Hole, Massachusetts, 13-14 June, 1949*. Washington: Office of Naval Research, Dept. of the Navy, 47-67.
- Velasco, C. and P.M. Robinson (2001). “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory* 17, 497-539.

## APPENDIX: PROOFS OF MAIN RESULTS

**Proof of Theorem 1:** Theorem 1(i)-(ii) generalize Theorem 1(i) of Phillips (2005), Theorem 2(a) of Sun (2011), and Theorem 4.1 of Sun (2013), all of which apply only to equal-weighted WOS estimators with  $q = 2$ . Theorem 1(iii) for WOS estimators generalizes Theorem 2(b) of Sun (2011). See the Online Supplement for the proof.

**Proof of Theorem 2:** Write  $c_{m,T}^\alpha(b) = c_m^\alpha(b) + d_{m,T}^\alpha(b)$  for some  $d_{m,T}^\alpha(b) = o(1)$ , where  $c_m^\alpha(b)$  is as in (18), and denote  $f(z) = \Pr_0[F_T^* > z]$ . Taylor expanding  $f(z)$  around  $c_m^\alpha(b)$ ,

$$f(c_{m,T}^\alpha(b)) = \alpha + G'_m(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} - d_{m,T}^\alpha(b) G'_m(\chi_m^\alpha) [1 + O(b) + O((bT)^{-q})] \\ + o(b) + o((bT)^{-q}) + o(d_{m,T}^\alpha(b)), \quad (33)$$

where  $f'(c_{m,T}^\alpha(b)) = -G'_m(\chi_m^\alpha) [1 + O(b) + O((bT)^{-q})]$  follows from (S.29) in the proof of Theorem 1(iv) in the Online Supplement. Then using  $f(c_{m,T}^\alpha(b)) = \alpha + o(b) + o((bT)^{-q})$  in (33), we have  $d_{m,T}^\alpha(b) = k^{(q)}(0) (bT)^{-q} \chi_m^\alpha \omega^{(q)}$ , from which (20) follows.

Taking a similar Taylor expansion and using (S.30) in the Online Supplement,

$$\Pr_\delta[F_T^* > c_{m,T}^\alpha] = [1 - G'_{m,\delta^2}(\chi_m^\alpha)] + G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} - \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha \nu^{-1} \\ - d_{m,T}^\alpha(b) G'_{m,\delta^2}(\chi_m^\alpha) [1 + O(b) + O((bT)^{-q})] + o(b) + o((bT)^{-q}) + o(d_{m,T}^\alpha(b)). \quad (34)$$

From (20),  $d_{m,T}^\alpha(b) G'_{m,\delta^2}(\chi_m^\alpha) = k^{(q)}(0) (bT)^{-q} G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)}$ . So the terms in  $G'_{m,\delta^2}(\chi_m^\alpha)$  in (34) cancel to higher order, which with  $d_{m,T}^\alpha(b) = O((bT)^{-q})$  gives (21).  $\square$

**Proof of Theorem 3:** Fix a sequence  $b_1$  for test  $F_{1,T}^*$ . Given equivalent  $q$  for the two tests and using (18), equivalent higher-order size requires that  $b_2 = (k_2^{(q)}(0) / k_1^{(q)}(0))^{1/q} b_1$ .

Thus

$$G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k_2^{(q)}(0) (b_2 T)^{-q} = G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k_1^{(q)}(0) (b_1 T)^{-q}, \quad (35)$$

which along with (19) yields the stated relation.  $\square$

**Proof of Corollary 1:** Using (18) and (21), and the fact that  $\nu = (b\nu)^{-1}$ ,

$$\Delta_S = G'_m(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} + o(b) + o((bT)^{-q}), \quad (36)$$

$$\Delta_P(\delta) = \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha b\nu + o(b) + o((bT)^{-q}). \quad (37)$$

The leading terms in (36) and (37) are of equivalent asymptotic order if and only if  $b$  and  $(bT)^{-q}$  are of equivalent asymptotic order, which leads to the stated sequence.  $\square$

**Proof of Theorem 4:**

(i) Under the assumed sequence, rewrite (36) as

$$|\Delta_S|^{1/q} = \left( G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q} \left| \omega^{(q)} \right|^{1/q} \left( k^{(q)}(0) \right)^{1/q} (bT)^{-1} [1 + o(1)]^{1/q}, \quad (38)$$

which gives that  $T|\Delta_S|^{1/q} = \left( G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q} \left| \omega^{(q)} \right|^{1/q} \left( k^{(q)}(0) \right)^{1/q} b^{-1} + o(1/b)$ . Multiplying this with (37) (for which  $o(b) = o((bT)^{-q})$ ) and defining  $a_{m,\alpha,q}(\delta)$  and  $\ell^{(q)}(k)$  as stated,

$$T\Delta_P(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta) \left[ \left( k^{(q)}(0) \right)^{1/q} \psi \right] \left| \omega^{(q)} \right|^{1/q} + o(1). \quad (39)$$

(ii) Write  $\Delta_P^{\max} = \sup_\delta \left\{ \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha \right\} b\psi + o(b)$ , since  $\delta$  does not enter  $b\psi$ . Using

this with the same steps as in part (i) yields (27).

(iii) From (38),  $\sqrt{|\Delta_S|} = \sqrt{G'_m(\chi_m^\alpha) \chi_m^\alpha \left| \omega^{(q)} \right| \left( k^{(q)}(0) \right)^{2/q} (bT)^{-q/2} (1 + o(1))}$ . Multiplying by  $\Delta_P^{\max}$ ,

$$\Delta_P^{\max} \sqrt{|\Delta_S|} = \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha \sqrt{G'_m(\chi_m^\alpha) \chi_m^\alpha \left| \omega^{(q)} \right| \left( k^{(q)}(0) \right)^{2/q} \psi T^{-1} (bT)^{1-q/2} (1 + o(1))}. \quad (40)$$

Since  $bT \rightarrow \infty$ , comparing any kernel or WOS tests with  $q = 1$  and  $2$ ,  $\exists \underline{T}$  s.t.  $\forall T > \underline{T}$ ,

$$\Delta_P^{\max,q=2} \sqrt{|\Delta_S^{q=2}|} < \Delta_P^{\max,q=1} \sqrt{|\Delta_S^{q=1}|}, \text{ so } \Delta_P^{\max,q=2} < \Delta_P^{\max,q=1} \text{ for } |\Delta_S^{q=2}| = |\Delta_S^{q=1}|. \quad \square$$

**Proof of Theorem 5:**

(i) From Theorem 4(iii), we can confine attention to the  $q = 2$  case. First, for kernel tests, from Theorem 4(ii), the optimal tradeoff is achieved by minimizing  $\sqrt{k^{(2)}(0)} \int_{-\infty}^{\infty} k^2(x) dx$ .

This minimum is achieved by the QS estimator (Priestley (1981, p. 569-571)).

For WOS tests, from (6), the QS estimator can be represented as a WOS estimator (7) with the Fourier basis and weights  $w_j \propto [1 - (j/B)^2]$  (Priestley (1981, pp. 444, 581)), where we have transformed  $B/2 \mapsto B$  for notational simplicity (without loss, as  $B$  can be understood to be any  $B^*/2$  set as the upper limit in (7)). We show in two parts that QS again dominates among WOS estimators: first, for any set of weights  $\{w_j\}$ , the Fourier basis is optimal; second, the QS weights dominate given the choice of Fourier basis. We sketch the proof here, with technical details in the Online Supplement.

For the first step, fixing  $B$  and the set of weights  $\{w_j\}$ , from Theorem 4(ii) the size-power tradeoff depends on the choice of basis only through  $\sqrt{k^{(2)}(0)}$ , since  $\psi$  is fixed from (14). Lemma S1 in the Online Supplement shows that the Fourier basis minimizes  $|\sum_{j=1}^B w_j \int_0^1 \phi_j(s) \phi_j''(s) ds|$  (and thus  $\sqrt{k^{(2)}(0)}$ , from Theorem 1) for any set of weights, up to  $o(1/T)$ . The proof proceeds by considering the complex Fourier expansion of any  $\phi_j$  from a given basis,  $\phi_j(s) = \sum_{l=-\infty}^{\infty} a_{jl} e^{-i2\pi ls}$ . For any orthonormal series,

$$1 = \int_0^1 |\phi_j(s)|^2 ds = \sum_{l,l'=-\infty}^{\infty} a_{jl} \bar{a}_{j l'} \int_0^1 e^{-i2\pi ls} e^{i2\pi l' s} ds = \sum_l |a_{jl}|^2, \quad (41)$$

$$\text{and } 0 = \int_0^1 \phi_j(s) \overline{\phi_{j' \neq j}(s)} ds = \sum_l a_{jl} \bar{a}_{j' \neq j, l}, \quad (42)$$

where  $\bar{a}_{jl}$  is the complex conjugate of  $a_{jl}$ . The minimization problem for real  $\phi_j$  is then

$$\min_{\{a_{jl}\}} \left| \sum_{j=1}^B w_j \int_0^1 \phi_j(s) \overline{\phi_j''(s)} ds \right| \Leftrightarrow \min_{\{a_{jl}\}} \sum_{j=1}^B w_j \sum_l |a_{jl}|^2 l^2, \text{ s.t. (41)-(42)}. \quad (43)$$

The proof provided in the Supplement then considers a finite truncation of the infinite Fourier series for  $\phi_j$ , so that the problem may be re-expressed as a constrained trace minimization problem for a doubly stochastic matrix containing the values  $\{|a_{jl}|^2\}$ .

Birkhoff's Theorem gives that the extreme points of the set of doubly stochastic matrices are the permutation matrices. Since the objective (43) is linear in the values  $\{|a_{jl}|^2\}$ , and since the set of doubly stochastic matrices is compact and convex, a permutation matrix achieves the minimum; we show that the minimizing permutation matrix features

$a_{2j'-1, j'} = a_{2j', -j'} = 1$ ,  $j' = 1, \dots, B/2$ ,  $a_{jl} = 0$  otherwise. Thus the solution is

$\{\phi_{2j'-1}(s), \phi_{2j'}(s)\} = \{e^{-i2\pi j' s}, e^{i2\pi j' s}\} = \{\sqrt{2} \cos(2\pi j' s), \sqrt{2} \sin(2\pi j' s)\}$ ,  $j' = 1, \dots, B/2$ , so

we have in fact selected the Fourier basis as the minimizing basis for  $\sqrt{k^{(2)}(0)}$  for any set of weights. See the Supplement for details.

For the second step, given the use of Fourier basis functions, we wish to minimize

$$\ell^{(2)}(k) = (k^{(2)}(0))^{1/2} \psi \propto \left( \frac{1}{B^2} \sum_{j=1}^B w_j j^2 \right)^{1/2} \left( B \sum_{j=1}^B w_j^2 \right) = \left( \sum_{j=1}^B w_j j^2 \right)^{1/2} \left( \sum_{j=1}^B w_j^2 \right) \quad (44)$$

over the weights  $\{w_j\}$  (subject to Assumption 3) at all points on the sequence for  $B$ , where the fact that  $k^{(2)}(0) \propto B^{-2} \sum_{j=1}^B w_j j^2$  arises from Theorem 1(ii) and the use of Fourier basis functions (as can be seen from the proof of Lemma S1 in the Supplement). We then follow Priestley's (1981, p. 569-571) proof that the QS kernel minimizes  $\ell^{(2)}(k)$  among kernel functions, modified so that the proof is with respect to WOS estimators using the Fourier basis. This result is provided in Lemma S2 in the Online Supplement. Combined with the fact that the Fourier basis achieves the size-power frontier for any set of weights, QS thus dominates the size-power tradeoff for WOS estimators, and therefore globally among the families considered here.

From Priestley (1981, Tables 6.1 and 7.1),  $k^{(2)}(0) = \pi^2 / 10$ ,  $\int_{-\infty}^{\infty} k^2(x) dx = 6 / 5$  for QS. Combining these with (27) yields (28). Numerically computing  $\bar{a}_{m,\alpha,q} = \sup_{\delta} a_{m,\alpha,q}(\delta)$  for  $q = 2$  and  $\alpha = 0.05$  yields  $\bar{a}_{m,\alpha,q} 3\pi\sqrt{10} / 25 \approx 0.3368$  for  $m = 1$ ,  $\bar{a}_{m,\alpha,q} 3\pi\sqrt{10} / 25 \approx 0.6460$  for  $m = 2$ , and  $\bar{a}_{m,\alpha,q} 3\pi\sqrt{10} / 25 \approx 0.9491$  for  $m = 3$ .

(ii) As after (14), only equal-weighted orthonormal series estimators yield fixed- $b$  asymptotic distributions that are exact  $t$  or  $F$ . The proof of part (i) of the theorem implies immediately that with equal weights, the Fourier basis achieves the frontier. Thus the EWP test is optimal among tests with exact  $t$  and  $F$  asymptotic fixed- $b$  distributions.

From Priestley (1981, Table 7.1),  $k^{(2)}(0) = \pi^2 / 6$  for the Daniell kernel (i.e. the EWP estimator). Further,  $\psi = 1$  for this estimator. Combining these with (27) yields (29).

Again computing  $\bar{a}_{m,\alpha,q}$  for  $q = 2$  and  $\alpha = 0.05$ , we have  $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 0.3624$  for  $m = 1$ ,  $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 0.6950$  for  $m = 2$ , and  $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 1.0211$  for  $m = 3$ .  $\square$