# The Size-Power Tradeoff in HAR Inference

August 14, 2020

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 thank Ian Dew-Becker, Frank Diebold, Michael Jansson, Lutz Kilian, Tom Kolokotrones, Mikkel Plagborg-Møller, Ulrich Müller, Benedikt Pötscher, Yixiao Sun, Tim Vogelsang, Mark Watson, Ken West, numerous seminar participants, and three anonymous referees for helpful 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 upon evaluation 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 $z_t = x_t u_t$, the product of the regressors $x_t$ and the regression errors $u_t$, 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 $\Gamma_j = \text{cov}(z_t, z'_{t-j})$, $j = 0, 1, \ldots$. 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 $\hat{z}_t = x_t \hat{u}_t$, 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, the MSE-optimal truncation parameter can yield large size distortions (den Haan and Levin (1994, 1997)). In fact, Edgeworth expansions of rejection probabilities in the Gaussian location model formally show that the testing problem entails a bias-variance tradeoff, whereas MSE minimization entails a tradeoff between *squared* bias and variance, so the testing-optimal sequence $S_T$ increases more rapidly than the MSE-optimal sequence (Velasco and Robinson (2001), Sun, Phillips and Jin (2008)). The testing-optimal sequence introduces sampling variability in $\hat{\Omega}$ and thus $t$-like behavior of the $t$-statistic, but that variability can be addressed by using Kiefer and Vogelsang's (2005) "fixed $b$" (or "fixed smoothing") critical values,[1] which model the truncation parameter as increasing proportionally to $T$, i.e., $S_T = bT$. The lesson is thus to combine a testing-optimal bandwidth rate for $S_T$ (Sun, Phillips and Jin

[1] Jansson (2004), Sun, Phillips and Jin (2008), and Sun (2014b) show that, in the Gaussian location model, using fixed-$b$ critical values provides a higher-order refinement to the null rejection rate of HAR test statistics.

(2008)) with fixed-$b$ critical values (e.g., Sun (2014b)). This literature, however, has two loose ends. First, given a kernel, it points to, 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 small-$b$ 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 $\alpha$. 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 $\alpha$. Using size-adjusted power is the standard method for making higher-order comparisons between tests (e.g., Rothenberg (1984)) and ensures an "apples to apples" comparison of the ability of two different tests to detect violations of the null when the two tests have different unadjusted null rejection rates.[2]

This paper makes four main contributions. First, we derive theoretical expressions characterizing the tradeoff between the size distortion $\Delta_S$, that is, the size of the test minus the nominal level $\alpha$, 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-

[2] The classical theory of optimal testing (e.g., the Neyman-Pearson Lemma) ranks tests by their power among tests that have the same rejection rate under the null. This principle of finite-sample testing theory extends to the second-order comparison of tests based on their 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) comparing expressions for their size-adjusted power. See, e.g., the discussion in Rothenberg (1984), which draws on Pfanzagl and Wefelmeyer (1978), on the second-order efficiency of the one-sided Lagrange multiplier, likelihood ratio, and Wald tests in the scalar normal means model. More closely related, Sun, Phillips, and Jin (2008, Corollary 5) use these three steps to derive higher order approximations to the power of HAR tests based on second-order corrected critical values. The practice of using size-adjusted critical values is commonplace in Monte Carlo studies that compare competing tests; for but a few examples see Kiefer and Vogelsang (2002), Sul, Phillips, and Choi (2005), and Sun (2013) in the HAR literature; Long and Ervin (2000) in the heteroskedasticity-robust testing literature; Ng and Perron (2001) in the unit root literature; and Clark and West (2007) in the forecast comparison literature. We follow this size-adjusted power approach (and specifically steps (i)-(iii)) to study the higher-order efficiency of HAR tests, though see Sections 4.2 and 4.3 for discussions of alternative approaches.

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,

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

where $\omega^{(2)}$ is the normalized curvature of the spectral density of $z_t$ at frequency zero (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 (tent) 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 a unified representation for 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.[3] 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.[4] Building on Sun (2013), we characterize the size-power 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% level test with $m = 1$, this frontier is,

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

This frontier is achieved by Brillinger's (1975) equal-weighted periodogram (EWP) test and by

[3] WOS tests have a long history in spectral estimation, see for example Grenander and Rosenblatt (1957). WOS HAR papers include Müller (2007), Phillips (2005), and Sun (2013).
[4] See Brillinger (1975, exercise 5.13.25), Müller (2007), Phillips (2005), and Sun (2013).

the closely related equal-weighted cosine (EWC) test, in which Ω is estimated using the Type II cosine basis functions (Müller (2007)). 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 remainder of the paper is organized as follows. Section 2 provides notation and describes the family of kernel and series LRV estimators considered. Section 3 unifies these sets of estimators and, building on the literature on fixed-$b$ asymptotics and asymptotic expansions, provides results on the estimators' limiting behavior. Section 4 provides our main results. Section 5 concludes. Selected proofs of our main results are given in the Appendix, with additional proofs contained in the Online Supplement. Extensive Monte Carlo simulations are provided in the Online Supplement and in a companion paper, Lazarus, Lewis, Stock, and Watson (LLSW, 2018).

## 2. Model, Tests, and LRV Estimators

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

$$y_t = \beta + u_t,\ \ t = 1,\ldots, T, \tag{3}$$

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,

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

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 $\hat{z}_t = y_t - \hat{\beta} = y_t - \overline{y}$, where $\overline{y}$ is the sample mean of $y_t$. For $m = 1$, the $t$-statistic testing $\beta = \beta_0$ is $t_T = \sqrt{T}\overline{z}_0 \big/ \sqrt{\hat{\Omega}}$, where $\overline{z}_0 = T^{-1}\sum_{t=1}^{T} z_t(\beta_0)$, $z_t(\beta_0) = y_t - \beta_0$, and $\hat{\Omega}$ is an estimator of Ω. For $m > 1$, as in Stock and Watson (2008) and Sun (2013), we consider the scaled $F$ statistic, $F_T^* = \big((B-m+1)/B\big)F_T$, with $F_T = T\overline{z}_0{}'\hat{\Omega}^{-1}\overline{z}_0 / m$ 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

Kernel estimators of Ω are sums of sample autocovariances weighted by a kernel $k(\cdot)$:

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

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:

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

where $\lfloor T/2 \rfloor$ denotes the integer part of *T*/2, $I_{zz}(\omega)$ is the periodogram of $\hat{z}_t$ at frequency *ω*, $I_{\hat{z}\hat{z}}(\omega) = (2\pi)^{-1} d_{\hat{z}}(\omega)\overline{d_{\hat{z}}(\omega)}'$, $d_{\hat{z}}(\omega) = T^{-1/2}\sum_{t=1}^{T} \hat{z}_t e^{-i\omega t}$, and where the weights $\{\tilde{w}_j\}$ in (6) satisfy $\tilde{w}_j = T^{-1}\sum_{u=-(T-1)}^{T-1} k(u/S)e^{i2\pi ju/T}$.[5] 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 $\hat{\Omega}^{WP} = 4\pi\sum_{j=1}^{\lfloor T/2 \rfloor} \tilde{w}_j \operatorname{Re}\left(I_{\hat{z}\hat{z}}(2\pi j/T)\right)$.

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 $\tilde{w}_j = 2B^{-1}\mathbf{1}\left(|j| \leq B/2\right)$ (the Daniell spectral kernel) in (6). The quadratic spectral estimator is so named because its weights in (6) are quadratic in *j*: $\tilde{w}_j \propto \left[1 - \left(|j|/(B/2)\right)^2\right]\mathbf{1}\left(|j| \leq B/2\right)$.

---

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

### 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.[6] Following Sun (2013), let $\{\phi_j(s)\}$, $j = 0, \ldots, B$, $0 \le s \le 1$, 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 $j \ge 1$. The WOS estimator is,

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

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. While the spectral density estimation literature has considered general WOS estimators (see Hannan (1970), Brillinger (1975), Priestley (1981), and Stoica and Moses (2005)), the HAR testing literature specializes (7) to the case of equal weights $w_j = 1/B$, $j = 1,\ldots, B$ (Phillips (2005), Müller (2007), and Sun (2013)).

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.[7] The use of Fourier basis functions with equal weights produces the EWP estimator. Other examples of basis functions include Type II cosine basis functions (see, e.g., Müller (2007) and Müller and Watson (2008)) and Legendre polynomials (e.g., Kolokotrones and Stock (2019)). The theory developed here also includes the Ibragimov-Müller (2010) estimator, which is the sample variance of subsample estimators of $\beta$.[8] We show in Proposition S1 in the Online Supplement 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.

---

[6] As noted by Sun (2013), these series estimators can be written as "orthogonal multitaper" or "multiple window" estimators; see, for example, Brillinger (1975), Thomson (1982), and Stoica and Moses (2005) for discussions of properties of these estimators in spectral density estimation.

[7] This can be seen using the representation for $\hat{\Omega}^{WP}$ provided after (6). The asymptotic caveat arises given the different summation limits for $\hat{\Omega}$ in (6) and (7) (the effect of which vanishes as $B$, $T \to \infty$), along with the fact that (5) and (6) may not align exactly in finite samples (see Priestley (1981, pp. 578-581)).

[8] See Conway (1963), Fishman (1978), and Song and Schmeiser (1993) for earlier references to such a subsample (or "batch mean") estimator. We thank Yixiao Sun for pointing us to this literature.

## 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.

### 3.1 Implied Mean Kernel of WOS Estimators

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,

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

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

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

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 proof of Theorem 1 in the Online Supplement for details.

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

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

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

### 3.2 Properties of Kernel and WOS Estimators

The performance of a given kernel or WOS test depends on certain properties of the LRV estimator and stochastic process, for which we introduce terminology and notation here.

***Properties related to bias.*** 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,

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

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

Bias depends as well 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 will be convenient to work with the trace of a scaled version of this generalized derivative,

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

which is a measure of the long-run persistence or anti-persistence of $z_t$. With $m = 1$ and $q = 2$, for example, $\omega^{(2)} = -s_z''(0)/s_z(0)$.

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

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

***Properties related to variance.*** If $z_t$ is Gaussian, then both kernel and WOS LRV estimators are distributed as weighted averages of independent chi-squared random variables.[9] For kernel estimators and scalar processes, Tukey (1950) proposed approximating this mixture distribution by a chi-squared with degrees of freedom chosen to match the estimator's

[9] Without the Gaussianity assumption, this holds in the limit under fixed-$b$ asymptotics (e.g., Sun (2013, 2014b)).

asymptotic variance.[10] Tukey's approximation, extended to include WOS estimators, is

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

where $\nu$ is the "equivalent degrees of freedom" of $\hat{\Omega}$. For WOS estimators, we set $b = B^{-1}$.

For equal-weighted WOS estimators with $m = 1$, the approximation (14) is asymptotically exact, with $\nu = B$, for fixed $B$ and $T \to \infty$ (see Footnote 4). This extends straightforwardly to the vector case, $m > 1$, as $\hat{\Omega}^{WOS} \xrightarrow{d} \Omega^{1/2}\left(\Xi_B / B\right)\Omega^{1/2\prime}$, 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 LRV estimator is asymptotically independent of $\overline{z}_0$, 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.

### 3.3 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, we use $k$ to denote both kernels and WOS implied mean kernels; for kernel estimators, $b = S/T$ and for WOS estimators, $b = B^{-1}$.

We make the following assumptions.

<u>Assumption 1 (stochastic processes)</u>.

(a) $z_t$ 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 $\lambda = 0$.

(b) $\sum_{u=-\infty}^{\infty} |u|^r |\Gamma_u| < \infty$ for $r \in [0, 2+\zeta]$, for some $\zeta > 0$.

---

[10] Tukey's approximating distribution is derived under the assumption of an unbiased estimator $\hat{\Omega}$. Under the optimal sequence provided in Section 4, bias and variance are of the same asymptotic order, so that the chi-squared approximation here matches the estimator's asymptotic variance (as proven in Theorem 1) but not its mean.

Assumption 2 (kernels). The kernel $k(x):\mathbb{R}\to[-1,1]$ used for a kernel LRV estimator 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 (orthonormal series). For $j = 1, \ldots, B$, 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 $\sup_{s\in[0,1]}\left|\phi_j^{(n)}(s)\right|\leq C_{n,\phi}\,j^{2n+1/2}$ for some constant $C_{n,\phi}$ for all $j$ and $n = 0, 1, 2$. The WOS 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 $b^qT^{q-1} + (bT)^{-1} \to 0$.

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 the Online Supplement. Assumption 4 strengthens the corresponding condition in Sun, Phillips, and Jin (2008), who require $b + (bT)^{-1} \to 0$. 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.[11] A proof sketch is provided in the Appendix,

[11] In particular, the results for WOS estimators in Theorem 1(i)-(ii) generalize Theorem 1(i) of Phillips (2005), Theorem 2(a) of Sun (2011), and Theorem 4.1 of Sun (2013), as those earlier results apply only to equal-weighted WOS estimators with $q = 2$. Similarly, Theorem 1(iii) for WOS estimators generalizes Theorem 2(b) of Sun (2011).

with technical details relegated to the Online Supplement.

*Theorem* 1. Under Assumptions 1–4,

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

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

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

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

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

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

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

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

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,

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

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,

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

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 when the variance is estimated in the i.i.d. scalar location model, relative to Gaussian inference with a known variance.

## 4. Size-Power Tradeoffs and the Size-Power Frontier

This section uses the unified 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 further detail in Section 4.2. Proofs for Section 4.1 are provided in the Appendix.

### 4.1 Main Results

Assume throughout that Assumptions 1–4 hold.

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

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

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

$$\Pr_\delta\left[F^*_T > c^{\alpha}_{m,T}(b)\right] = \left[1-G_{m,\delta^2}(\chi^{\alpha}_m)\right] - \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi^{\alpha}_m)\chi^{\alpha}_m\nu^{-1} + o(b) + o\left((bT)^{-q}\right). \tag{21}$$

*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 $c^{\alpha}_{1,m}(b_1)$ and $c^{\alpha}_{2,m}(b_2)$. 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,

$$\Pr_\delta\left[F^*_{1T} > c^\alpha_{1,m}(b_1)\right] - \Pr_\delta\left[F^*_{2T} > c^\alpha_{2,m}(b_2)\right] = \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi^\alpha_m)\chi^\alpha_m\left(\nu_2^{-1} - \nu_1^{-1}\right)$$

$$+ o(b_1) + o((b_1T)^{-q}) + o(b_2) + o((b_2T)^{-q}). \qquad (22)$$

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

$$\Delta_S = \Pr_0\left[F^*_T > c^\alpha_m(b)\right] - \alpha\,. \qquad (23)$$

The power of the oracle test, in which $\Omega$ is known, is $1 - G_{m,\delta^2}(\chi^\alpha_m)$. 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 $\Delta^{\max}_P$ denote the maximum such power loss, so that $\Delta^{\max}_P$ is the maximum gap between the power curves of the oracle test and the candidate test:

$$\Delta_P(\delta) = \left[1 - G_{m,\delta^2}(\chi^\alpha_m)\right] - \Pr_\delta\left[F^*_T > c^\alpha_{m,T}(b)\right], \text{ and} \qquad (24)$$

$$\Delta^{\max}_P = \sup_\delta \Delta_P(\delta)\,. \qquad (25)$$

Because $\nu = (b\psi)^{-1}$, 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 $b$ can be eliminated to obtain expressions for 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.[12] 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

[12] 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:

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

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

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 $\ell^{(q)}(k) = \left(k^{(q)}(0)\right)^{1/q}\psi$.

(ii) The small-$b$ asymptotic tradeoff between $\Delta_S$ and the maximum power loss $\Delta_P^{\max}$ is,

$$T\Delta_P^{\max}\left|\Delta_S\right|^{1/q} = \overline{a}_{m,\alpha,q}\ell^{(q)}(k)\left|\omega^{(q)}\right|^{1/q} + o(1), \tag{27}$$

where $\overline{a}_{m,\alpha,q} = \sup_\delta a_{m,\alpha,q}(\delta)$.

(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 sequence for $b$ in Corollary 1,

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

where $\overline{a}_{m,\alpha,2}$ is defined in Theorem 4. This frontier is achieved by the QS kernel. For tests with $\alpha = .05$, $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.3368$ for $m = 1$ (yielding (1)), $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.6460$ for $m = 2$, and $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.9491$ 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,

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

This frontier is achieved by the EWP test. For $\alpha = .05$, $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 0.3624$ for $m = 1$ (yielding (2)), $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 0.6950$ for $m = 2$, and $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 1.0211$ 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 order for *b* in Corollary 1, $b = O\left(T^{-q/(q+1)}\right)$, is the same as 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 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 $MSE\left(\hat{s}_z(0)\right) = \text{bias}^2\left(\hat{s}_z(0)\right) + \text{var}\left(\hat{s}_z(0)\right)$;
   (b) The previous objective function modified to $a\left|\text{bias}\left(\hat{s}_z(0)\right)\right| + (1-a)\,\text{var}\left(\hat{s}_z(0)\right)$;
   (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 $0 \le a \le 1$ 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.

See the Appendix for derivations. Minimizing (a) is, as above, 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). We thank a referee for pointing out that (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), which uses the results in Theorem 4 to obtain a rule of thumb for $b$. 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 $\ell^{(2)}(k)$.[13] Each of these objective functions is minimized by the QS kernel, or, among equal-weighted WOS estimators, by the EWP estimator.[14]

4. The optimality of the QS kernel in minimizing $\ell^{(2)}(k)$ is well known for kernel tests. Obtaining the frontier in Theorem 5(i), however, requires proving the optimality of the QS kernel relative to all WOS tests, and not just among kernel tests. The proof of this new optimality result proceeds in two parts: first, we show for WOS tests that given any set of weights, the Fourier basis functions are optimal; second, among WOS tests using the Fourier basis functions, choosing weights to deliver an estimator equivalent to the QS kernel is optimal. The first part of this argument then implies immediately 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. The fact that the EWP test achieves the restricted frontier follows Grenander and Rosenblatt's (1957, Section 4.2) result on the optimality of "spectrograph"

[13] We thank Yixiao Sun for pointing this out to us; see Sun and Yang (2020) for a complete derivation of this result, a version of which also appears in LLSW (2018, rejoinder).

[14] Objective functions (a) and (b) also depend on the (implied mean) kernel only through $\ell^{(2)}(k)$ in the case of unknown $\beta$ when restricting the class of estimators to WOS estimators.

(i.e., weighted Fourier series) estimators of the spectral density; our proof of the more general result in Theorem 5(i) builds on theirs, with some technical modifications.

5. The price one must pay for the convenience of exact *t* or *F* fixed-*b* critical values can be computed from Theorem 3. For EWP and QS tests with the same higher-order size,

$$\Pr_\delta[F^*_{QS,T} > c_{QS,\alpha}(b_{QS})] - \Pr_\delta[F^*_{EWP,T} > c_{EWP,\alpha}(b_{EWP})] \approx \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi^\alpha_m)\chi^\alpha_m\left(\nu^{-1}_{EWP} - \nu^{-1}_{QS}\right)$$

$$= \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi^\alpha_m)\chi^\alpha_m\left(1 - \frac{6\sqrt{3}}{5\sqrt{5}}\right)B^{-1}, \tag{30}$$

where $\nu_{EWP} = B$ and the final expression is derived in the Online Supplement (see Proposition S3).

Table 1 reports the maximum higher-order power loss from using EWP over all alternatives $\delta$, that is, (30) maximized over $\delta$. 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 S.3 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.[15] For example, as shown in Proposition S4 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.[16] 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

[15] For WOS estimators, assessing the expressions in Theorem 4 requires calculating generalized derivatives using Theorem 1(ii). For kernel estimators, the required values are typically known (e.g., Priestley (1981)). We note as well that Theorem 3 allows for test comparisons for any sequence meeting Assumption 4 (as in Remark 5), and for tests with different values of $q$, but it requires calculating the values $b_1$ and $b_2$ to equate the two tests' size.

[16] Kolokotrones and Stock (2019) consider $q = 1$ HAR tests in more detail.

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 $T\left(R\bar{z}_0\right)'\left(R\hat{\Omega}R'\right)^{-1}\left(R\bar{z}_0\right)/m'$, where $R\bar{z}_0 = T^{-1}\sum_{t=1}^{T}\left(Rz_t - \tilde{\beta}_0\right)$. 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 Optimal Kernels and Rates for 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 $|\omega^{(q)}| \leq \bar{\omega}^{(q)}$.[17]

Uniform size control can be achieved for any sequence $b \propto T^{-q/(1+q)}$ by using the size-

[17] This uniform-size-control approach follows a small 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 and by focusing on the higher-order approximations used throughout the current paper.

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.[18]

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 $\omega^{(q)} = \bar{\omega}^{(q)}$:

$$b^{WAP} = \arg\max_b \int_{\delta} \int_{|\rho|\leq\bar{\rho}} \Delta_P\left(\omega^{(q)}(\rho), \delta\right) d\Pi_\rho(\rho) d\Pi_\delta(\delta), \tag{31}$$

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

$\omega^{(1)}(\rho) = 2\rho/(1-\rho^2)$, $\omega^{(2)}(\rho) = 2\rho/(1-\rho)^2$, $\bar{\rho} = \max \rho$ 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 S5 in the Online Supplement, is

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

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

---

[18] This statement requires that the remainder terms in (18) are $o(b) + o\left((bT)^{-q}\right)$ uniformly in $\left|\omega^{(q)}\right| \leq \bar{\omega}^{(q)}$, which we assume holds. We thank Yixiao Sun for alerting us to this subtlety; see the Online Supplement for details.

asymptotic tradeoffs and frontiers. The results are reported in Section S2 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

Our results connect the theory of optimal HAR testing to classic results on optimal spectral estimation. Although the optimal bandwidth rates differ for the testing and estimation problems, in both cases the effect of the choice of kernel appears via the same functional. Results on kernel choice from the classical estimation literature accordingly extend to HAR testing.

We obtain the size-power tradeoff and frontier under what is shown to be the optimal sequence, $b = b_0 T^{-q/(q+1)}$. In practice, one of course needs to know the coefficient $b_0$. Different choices of $b_0$ lead to different points on the tradeoff curve and different points on the frontier, and the analysis in this paper does not tell the user which of those points to use. One might be tempted to select $b_0$ to maximize size-adjusted power, which would lead to a corner solution with the smallest possible $b_0$ while respecting the optimal sequence. Because that choice would have large size distortions, it would be necessary to use 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 implementing feasible size adjustment replaces the difficult problem of estimating the spectral density at frequency zero with the more difficult problem of estimating its curvature.[19] Choosing a point on the frontier thus requires a judgement by the user. One approach 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 for the user to specify an explicit tradeoff between size and size-adjusted power. A third approach, discussed in Remark 3, is for the user to follow Sun, Phillips, and Jin (2008) and specify an explicit tradeoff between Type I and Type II errors. 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, however 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, the fact that the tradeoffs for $q = 1$ and $q = 2$ kernels cross for certain processes and sample sizes suggests that it is of interest to explore whether one can improve upon the Bartlett kernel among $q = 1$ kernels, a topic taken up in Kolokotrones and Stock (2019). Fourth, additional theoretical work on the regression model with stochastic regressors is also in order.

## References


Abramowitz, M. and I.A. Stegun (1965). *Handbook of Mathematical Functions*. New York: Dover.

Andrews, D.W. K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Bhatia, R. (1997). *Matrix Analysis*. New York: Springer.

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.

Conway, R.W. (1963). "Some Tactical Problems in Digital Simulation," *Management Science* 10, 47-61.

den Haan, W.J. and A. Levin (1994). "Vector Autoregressive Covariance Matrix Estimation," manuscript, Board of Governors of the Federal Reserve.

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.


[19] These simulation results are available upon request.

Fishman, G.S. (1978). "Grouping Observations in Digital Simulation," *Management Science* 24, 510-521.
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.
Helland, I. S. (1982). "Central Limit Theorems for Martingales with Discrete or Continuous Time," *Scandinavian Journal of Statistics* 9, 79-94.
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., T.J. Vogelsang, and H. Bunzel (2000). "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 69, 695-714.
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.
Kolokotrones, T. and J.H. Stock (2019), "Is Newey-West Optimal Among $q = 1$ Kernels?," manuscript, Harvard University.
Lazarus, E., D.J. Lewis, and J.H. Stock (2017). "The Size-Power Tradeoff in HAR Inference," 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.
Magnus, J.R. and H. Neudecker (1979). "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics* 7, 381-394.
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.
Müller, U. K. and M.W. Watson (2008). "Testing Models of Low-Frequency Variability," *Econometrica* 76, 979–1016.
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.
Pfanzagl, J. and W. Wefelmeyer (1978). "An Asymptotically Complete Class of Tests," *Zeitschrift fur Wahrscheinlichkeitstheorie*, 45, 49-72.
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.
Shen, J., T. Tang, and L.-L. Wang (2011). *Spectral Methods: Algorithms, Analysis and Applications*. Berlin: Springer-Verlag.
Song, W.T. and B.W. Schmeiser (1993). "Variance of the Sample Mean: Properties and Graphs of Quadratic-Form Estimators," *Operations Research* 41, 501-517.
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.
Sul, D., P.C.B. Phillips, and C-Y. Choi (2005). "Prewhitening Bias in HAC Estimation," *Oxford Bulletin of Economics and Statistics* 67(4), 517-546.
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.
Thomson, D.J. (1982). "Spectrum Estimation and Harmonic Analysis," *Proceedings of the IEEE* 70, 1055-1096.
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: Selected Proofs

**Proof of Theorem 1:**

(i)-(ii) For kernel estimators, (15) follows from a well-known result (see, e.g., Velasco and Robinson (2001, Lemmas 2 and 6) for the scalar case, and the proof of Theorem 2 in Sun (2014b) for general $m \geq 1$). For WOS estimators, (i) and (ii) follow from using the representation (9) to derive that

$$E\hat{\Omega}^{WOS} - \Omega = \sum_{u=-L_T}^{L_T} \left[ k_{B,T}^{WOS}\left(\frac{u}{S}\right) - 1 \right] \Gamma_u + o\left( \left(\frac{B}{T}\right)^q \right), \tag{33}$$

where $L_T \to \infty$ such that $L_T = o(T / B)$ and $(T / B)^q = o(L_T^{q+\zeta})$ for some $\zeta > 0$. Equation (15) then follows using standard calculations for the bias of kernel spectral density estimators, which we extend to account for the sequential nature of the implied mean kernel functions in (33). The expressions for the generalized derivatives of the implied mean kernel in (16) then follow by direct calculation. See Section S3 of the Online Supplement for details.

(iii) For kernel estimators, (17) follows from Proposition 1(a) of Andrews (1991). For WOS estimators, the result follows from the proof of Theorem 2(b) of Sun (2011), along with a derivation provided in the Online Supplement showing that

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\left[\sum_{j=1}^{B} w_j \phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{\tau}{T}\right)\right]^2 = \sum_{j=1}^{B} w_j^2 . \tag{34}$$

(iv)-(v) These results are extensions and unifications of results in the literature that take advantage of the unified definition of $v$ in (14). Details are again given in the Online Supplement.

(vi) The result obtains by expressing the SS estimator as a projection on a certain basis function, which in turn leads to an expression for the SS implied mean kernel and its generalized first derivative. When evaluated, one finds $k_B^{SS(1)}(0) = (B+2)/(B+1) \to 1$. See the Online Supplement for details. □

Before providing the proofs of our main results, we state and prove the following preliminary result.

*Lemma A1.* Let Assumptions 1, 3, and 4 hold. For any weights $\{w_j\}$, the Fourier basis minimizes $\left|\sum_{j=1}^{B} w_j \int_0^1 \phi_j(s)\phi_j''(s)ds\right|$ across all WOS estimators up to an error of order $o(1/T)$.

**Proof of Lemma A1:** Consider the complex Fourier expansion of any $\phi_j$ from a given basis,

$$\phi_j(s) = \sum_{l=-\infty}^{\infty} a_{jl} e^{-i2\pi ls} , \tag{35}$$

where $\{a_{jl}\}_l$ are the (inverse) Fourier coefficients of $\phi_j(s)$, with $a_{j0} = \int_0^1 \phi_j(s)ds = 0$. For any orthonormal series, denoting the complex conjugate of $a_{jl}$ by $\bar{a}_{jl}$, we have

$$1 = \int_0^1 \left|\phi_j(s)\right|^2 ds = \sum_{l=-\infty}^{\infty}\sum_{l'=-\infty}^{\infty} a_{jl}\bar{a}_{jl'} \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} ds = \sum_l \left|a_{jl}\right|^2 , \tag{36}$$

$$\text{and } \; 0 = \int_0^1 \phi_j(s)\overline{\phi_{j'\neq j}(s)}ds = \sum_l a_{jl}\bar{a}_{j'\neq j,l} . \tag{37}$$

Our minimization problem for real $\phi_j$ can then be written as

$$\min_{\{a_{jl}\}}\left|\sum_{j=1}^{B} w_j \int_0^1 \phi_j(s)\overline{\phi_j''(s)}ds\right| \Leftrightarrow \min_{\{a_{jl}\}}\left|\sum_{j=1}^{B} w_j \sum_l \sum_{l'} a_{jl}\bar{a}_{jl'} 4\pi^2 l^2 \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} ds\right| \Leftrightarrow \min_{\{a_{jl}\}} \sum_{j=1}^{B} w_j \sum_l \left|a_{jl}\right|^2 l^2 , \tag{38}$$

subject to the constraints (36), (37), and $a_{j0} = 0$. For now set the limits of summation in (35) to be $\pm\bar{T}$, where $\bar{T}$ is an integer satisfying $T / \bar{T} = o(1)$. Then (38) can be written equivalently as

$$\min_A \operatorname{tr}\left(\left(AW\right)^* D\left(AW\right)\right) \Leftrightarrow \min_A \operatorname{tr}\left(W^2 A^* D A\right) \tag{39}$$

subject to $A^* A = I_B$, where $A = [A_1 \quad A_{-1} \quad A_2 \quad A_{-2} \quad \ldots \quad A_{\bar{T}} \quad A_{-\bar{T}}]'$, $A_l = [a_{1l} \quad a_{2l} \quad \ldots \quad a_{Bl}]'$, $A^*$ is the conjugate transpose of $A$, $W = \operatorname{diag}\left(\left[\sqrt{w_1} \quad \sqrt{w_2} \quad \ldots \quad \sqrt{w_B}\right]\right)$, and $D = \operatorname{diag}([1 \quad 1 \quad 4 \quad 4 \quad \ldots \quad \bar{T}^2 \quad \bar{T}^2])$. Note from (38) that the objective is linear in the entries of the matrix $\mathbf{A}_2 = A \circ \bar{A}$, where $\circ$ is the Hadamard product and $\bar{A}$ is the elementwise complex conjugate of $A$.

It will be convenient to transform this problem to $\min_{\tilde{A}} \operatorname{tr}(\tilde{W}^2 \tilde{A}^* D \tilde{A})$ subject to $\tilde{A}^* \tilde{A} = I_{2\bar{T}}$, where $\tilde{W}$ is padded with zeros relative to $W$ (i.e., $\tilde{W} = [I_B \quad 0_{B\times(2\bar{T}-B)}]' W [I_B \quad 0_{B\times(2\bar{T}-B)}]$) and $\tilde{A} = [A \quad H]$, where $H$ is some $2\bar{T} \times (2\bar{T} - B)$ matrix, so that $\tilde{A}$, $\tilde{W}$, and $D$ are all $2\bar{T} \times 2\bar{T}$. The objective is again linear in the entries of $\widetilde{\mathbf{A}}_2 = \tilde{A} \circ \overline{(\tilde{A})}$. Note that $\widetilde{\mathbf{A}}_2$ is doubly stochastic, since $\tilde{A}^* \tilde{A} = I_{2\bar{T}}$ implies that $\tilde{A}\tilde{A}^* = I_{2\bar{T}}$. Thus

$$\min_{\tilde{A} \text{ s.t. } \tilde{A}^* \tilde{A} = I_{2\bar{T}}} \operatorname{tr}\left(\tilde{W}^2 \tilde{A}^* D \tilde{A}\right) \geq \min_{\Upsilon} \sum\nolimits_{j,l} w_j l^2 \gamma_{jl}, \tag{40}$$

where $\Upsilon$ is a doubly stochastic matrix constructed conformably with $A$ and containing the values $\{\gamma_{jl}\}$. The right side of this inequality is linear in the entries of $\Upsilon$, and the set of doubly stochastic matrices is compact and convex. Thus the minimum of the right side is obtained at an extreme point of this set. By Birkhoff's Theorem (e.g., Bhatia (1997, p. 37)), the extreme points of the set of doubly stochastic matrices are the permutation matrices. And any permutation matrix $P$ is unitary, so (40) in fact holds with equality, and we can select $\tilde{A} = \arg\min_P \operatorname{tr}(\tilde{W}^2 P' D P)$.

Next, note that $D$ and $\tilde{W}^2$ are psd and diagonal, and $D$ has its diagonal terms (and eigenvalues) in ascending order. Given some set of weights $\{w_j\}$, assume first that they are ordered descendingly, so that $w_1 \geq w_2 \geq \ldots \geq w_B$, and therefore $\tilde{W}^2$ has its diagonal terms (eigenvalues) in descending order. In this case, the minimum of the objective is achieved trivially by $\tilde{A} = P = I_{2\bar{T}}$, so that the minimizing $A$ is given by the first $B$ columns of $I_{2\bar{T}}$; equivalently, $a_{2j'-1,j'} = a_{2j',-j'} = 1$, $j' = 1, \ldots, B/2$, $a_{jl} = 0$ otherwise. Thus from (35), $\{\phi_{2j'-1}(s), \phi_{2j'}(s)\} = \{e^{-i2\pi j's}, e^{i2\pi j's}\} = \{\sqrt{2}\cos(2\pi j's), \sqrt{2}\sin(2\pi j's)\}$, $j' = 1, \ldots, B/2$, so that we have in fact selected the Fourier basis. This applies for any $\bar{T}$ used in the approximation of the expansion (35), so we can set $\bar{T} \gg T$ as large as necessary and then apply Jackson's inequality to obtain that the given statement holds up to an error of order $o(1/T)$.

If the weight values are not in descending order, use that $\tilde{W}^2$ is psd and diagonal to write its singular value decomposition as $\tilde{W}^2 = V \tilde{W}_{desc}^2 V'$, where $\tilde{W}_{desc}^2$ is the diagonal matrix containing the eigenvalues (diagonal terms) of $\tilde{W}^2$ ordered descendingly. Then the problem can be rewritten as $\min_{\tilde{A}_V} \operatorname{tr}(\tilde{W}_{desc}^2 \tilde{A}_V^* D \tilde{A}_V)$ subject to $\tilde{A}_V^* \tilde{A}_V = I_{2\bar{T}}$, where $\tilde{A}_V = \tilde{A} V$, so that $V$ has been absorbed into the argument to be minimized. But this is the same problem as in the case above, with $\tilde{W}^2$ having its values in descending order. Thus the minimum achieved for (40) is equivalent for any set of weights regardless of their ordering. This implies that for any set of weights, it is without loss of generality to set them in descending order, in which case the Fourier basis again achieves the minimum, completing 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 $c_m^{\alpha}(b)$,

$$\begin{aligned} f\left(c_{m,T}^{\alpha}(b)\right) = \alpha + G'_m(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q} - d_{m,T}^{\alpha}(b)G'_m(\chi_m^{\alpha})\left[1 + O(b) + O\left((bT)^{-q}\right)\right] \\ + o(b) + o\left((bT)^{-q}\right) + o\left(d_{m,T}^{\alpha}(b)\right), \end{aligned} \tag{41}$$

where the fact that $f'\left(c_m^\alpha(b)\right) = -G'_m(\chi_m^\alpha)\left[1 + O(b) + O\left((bT)^{-q}\right)\right] + o(b) + o\left((bT)^{-q}\right)$ follows from equation (S.23) in the proof of Theorem 1(iv) in the Online Supplement. Using (41) and $f\left(c_{m,T}^\alpha(b)\right) = \alpha + o(b) + o\left((bT)^{-q}\right)$ by definition, $d_{m,T}^\alpha(b) = k^{(q)}(0)(bT)^{-q}\chi_m^\alpha\omega^{(q)}$, from which (20) follows.

Taking a similar Taylor expansion and using equation (S.24) in the Online Supplement, size-adjusted power is

$$\Pr_\delta[F_T^* > c_{m,T}^\alpha] = [1 - G_{m,\delta^2}(\chi_m^\alpha)] + G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k^{(q)}(0)(bT)^{-q} - \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\nu^{-1}$$
$$-d_{m,T}^\alpha(b)G'_{m,\delta^2}(\chi_m^\alpha)\left[1 + O(b) + O\left((bT)^{-q}\right)\right] + o(b) + o\left((bT)^{-q}\right) + o(d_{m,T}^\alpha(b)). \quad (42)$$

We have $d_{m,T}^\alpha(b)G'_{m,\delta^2}(\chi_m^\alpha) = k^{(q)}(0)(bT)^{-q}G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}$ from (20). Thus the terms in $G'_{m,\delta^2}(\chi_m^\alpha)$ in (42) cancel to higher order, which along with $d_{m,T}^\alpha(b) = O((bT)^{-q})$ yields (21). □

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

$$G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k_2^{(q)}(0)(b_2T)^{-q} = G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k_1^{(q)}(0)(b_1T)^{-q}, \quad (43)$$

which along with (19) yields the stated relation. □

**Proof of Corollary 1:** Using (18) and the definition of $\Delta_S$,

$$\Delta_S = G'_m(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right). \quad (44)$$

Using the definition of $\Delta_P(\delta)$ and the fact that $\nu = (b\psi)^{-1}$,

$$\Delta_P(\delta) = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha b\psi + o(b) + o\left((bT)^{-q}\right). \quad (45)$$

The first terms in (44) and (45) are of equivalent asymptotic order if and only if $b$ and $(bT)^{-q}$ are of equivalent asymptotic order, which leads to the stated sequence. Further, given that $O(b) = O\left((bT)^{-q}\right)$ under this sequence, the leading term in (44) is in fact the first term, and similarly for (45). □

**Proof of Theorem 4:**

(i) Under the assumed sequence, rewrite (44) as

$$|\Delta_S|^{1/q} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q}\left|\omega^{(q)}\right|^{1/q}\left(k^{(q)}(0)\right)^{1/q}(bT)^{-1}[1 + o(1)]^{1/q}$$
$$= \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q}\left|\omega^{(q)}\right|^{1/q}\left(k^{(q)}(0)\right)^{1/q}(bT)^{-1} + o\left((bT)^{-1}\right). \quad (46)$$

This can be rewritten further as $T|\Delta_S|^{1/q} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q}\left|\omega^{(q)}\right|^{1/q}\left(k^{(q)}(0)\right)^{1/q}b^{-1} + o(1/b)$. Similarly, rewrite (45) as $\Delta_P(\delta) = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha b\psi + o(b)$. Define $a_{m,\alpha,q}(\delta)$ as in the statement of the theorem, and multiply the two previous expressions to obtain

$$T\Delta_P(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta)\left[\left(k^{(q)}(0)\right)^{1/q}\psi\right]\left|\omega^{(q)}\right|^{1/q} + o(1), \quad (47)$$

as stated, since $\ell^{(q)}(k) = \left(k^{(q)}(0)\right)^{1/q}\psi$.

(ii) Write

$$\Delta_P^{\max} = \sup_\delta \left\{ \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha \right\} b\psi + o(b), \tag{48}$$

since $\delta$ does not enter into the term $b\psi$. Using this with the same steps as in part (i) above yields (27).

(iii) Express $\sqrt{(|\Delta_S|)}$ as

$$\sqrt{|\Delta_S|} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/2} \left|\omega^{(q)}\right|^{1/2} \left(k^{(q)}(0)\right)^{1/2} (bT)^{-q/2} + o\left((bT)^{-q/2}\right). \tag{49}$$

Multiplying this by (48), under the assumed sequence for *b*,

$$\Delta_P^{\max} \sqrt{|\Delta_S|} = \overline{a}_{m,\alpha,2} \left(\sqrt{k^{(q)}(0)}\psi\right) \left|\omega^{(q)}\right|^{1/2} T^{-1}(bT)^{1-q/2} + o(b^{3/2}), \tag{50}$$

so $\Delta_P^{\max}\sqrt{|\Delta_S|}$ tends to zero at a slower rate for $q$ = 1 than for $q$ = 2 given that $bT \to \infty$. Thus comparing arbitrary kernel or WOS tests with $q$ = 1 and $q$ = 2, for any two sets of values $k^{(q)}(0)$ and $\left|\omega^{(q)}\right|$, it must be that $\exists \overline{b}, \underline{T}$ such that $\forall b < \overline{b}, T > \underline{T}$, the $q$ = 2 test dominates the size/power tradeoff of the $q$ = 1 test (i.e., $\Delta_P^{\max,q=2}\sqrt{|\Delta_S^{q=2}|} < \Delta_P^{\max,q=1}\sqrt{|\Delta_S^{q=1}|}$). □

**Proof of Theorem 5:**

(i) Theorem 4(iii) implies that we can confine attention to $q$ = 2 estimators. We first consider kernel estimators. From Theorem 4(ii), the lower envelope of the size/power tradeoff is achieved by minimizing $\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx$. As in Priestley (1981, pp. 569-570), this is equivalent to minimizing $\left\{\int_{-\infty}^{\infty} \omega^2 K(\omega)d\omega\right\}^{1/2} \left\{\int_{-\infty}^{\infty} K(\omega)d\omega\right\}$, where *K* is the frequency-domain weight function (spectral window generator) corresponding to *k*; see Footnote 5. This minimum is achieved by the QS estimator (Priestley (1981, p. 571)).

For WOS estimators, from (6), the QS estimator can be represented as a WOS estimator (7) with the Fourier basis and weights $w_j \propto \left[1 - (j/B)^2\right]$, where we have transformed $B/2 \mapsto B$ for notational simplicity (and without loss of generality, as *B* can be understood to be equal to any $B^*/2$ specified as the upper limit in the sum in (7)). We proceed to show in two parts that the QS estimator again dominates among WOS estimators: first, given any set of weights $\{w_j\}$, the Fourier basis is optimal; second, the QS weights dominate given the choice of Fourier basis functions.

For the first step, fixing *B* and the set of weights $\{w_j\}$, it can be seen from Theorem 4(ii) that the size-power tradeoff depends on the choice of basis only through $\sqrt{k^{(2)}(0)}$, since $\psi$ is fixed from (14). Lemma A1 and Theorem 1(ii) then give that the Fourier basis functions minimize $\sqrt{k^{(2)}(0)}$ for any given choice of weights up to a term $o(1/T)$.

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

$$\ell^{(2)}(k) = \left(k^{(2)}(0)\right)^{1/2} \psi \propto \left(\frac{1}{B^2}\sum_{j=1}^{B} w_j j^2\right)^{1/2} \left(B\sum_{j=1}^{B} w_j^2\right) = \left(\sum_{j=1}^{B} w_j j^2\right)^{1/2} \left(\sum_{j=1}^{B} w_j^2\right) \tag{51}$$

over the weights $\{w_j\}$ (subject to Assumption 3) at all points on the sequence for *B*, where the fact that $k^{(2)}(0) \propto \left(\sum_{j=1}^{B} w_j j^2\right) / B^2$ arises from Theorem 1(ii) and the use of Fourier basis functions (as can be seen from (38)), and where the constant of proportionality does not depend on *B* or the choice of weights. We 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.

Set $w_j^*$ to be the QS weights, $w_j^* = \overline{w}_{QS}\left[1-\left(j/B\right)^2\right]$, where $\overline{w}_{QS} = 6B/\left[(B-1)(4B+1)\right]$ is set so that $\sum_{j=1}^{B} w_j^* = 1$, and write $B_{QS} = B$. For any alternative set of weights $w_j$, write $w_j = w_j^* + \epsilon_j$. We allow for the value $B_{\text{alt}}$ corresponding to this set of weights to differ from $B_{QS}$ (while still following a sequence according to the rate in Corollary 1): we will equate higher-order size for the QS and alternative estimators and show that QS dominates with respect to power. If $B_{\text{alt}} > B_{QS}$, then set $w_j^* = 0$ for $j > B_{QS}$, so that $w_j = \epsilon_j \geq 0$ for $B_{QS} < j \leq B_{\text{alt}}$. If $B_{\text{alt}} < B_{QS}$, then correspondingly $w_j = 0$ for $B_{\text{alt}} < j \leq B_{QS}$. To handle both possible cases simultaneously, write $\overline{B} = \max\left(B_{\text{alt}}, B_{QS}\right)$.

Since $\sum_{j=1}^{\overline{B}} w_j = 1$, we have $\sum_{j=1}^{\overline{B}} \epsilon_j = 0$. Equating higher-order size, from Theorem 1, requires $\sum_{j=1}^{\overline{B}} w_j j^2 = \sum_{j=1}^{\overline{B}} w_j^* j^2$, so $\sum_{j=1}^{\overline{B}} \epsilon_j j^2 = 0$. Further, $\sum_{j=1}^{\overline{B}} w_j^2 = \sum_{j=1}^{\overline{B}} \left(w_j^*\right)^2 + 2\sum_{j=1}^{\overline{B}} \epsilon_j w_j^* + \sum_{j=1}^{\overline{B}} \epsilon_j^2$, and

$$\sum_{j=1}^{\overline{B}} \epsilon_j w_j^* = \overline{w}_{QS} \sum_{j=1}^{B_{QS}} \left[1-\left(\frac{j}{B_{QS}}\right)^2\right]\epsilon_j = \overline{w}_{QS}\left\{\sum_{j=1}^{\overline{B}} \left[1-\left(\frac{j}{B_{QS}}\right)^2\right]\epsilon_j + \sum_{j=B_{QS}+1}^{B_{\text{alt}}} \left[\left(\frac{j}{B_{QS}}\right)^2 - 1\right]\epsilon_j\right\}. \quad (52)$$

The first term in this expression is zero given the steps above. For the second term, if $B_{\text{alt}} > B_{QS}$, then as above $w_j = \epsilon_j \geq 0$ for $B_{QS} < j \leq B_{\text{alt}}$, and therefore $\sum_{j=B_{QS}+1}^{B_{\text{alt}}} [(j/B_{QS})^2 - 1]\epsilon_j \geq 0$ (with equality if $B_{\text{alt}} < B_{QS}$). Thus $\sum_{j=1}^{\overline{B}} \epsilon_j w_j^* \geq 0$. It is further trivially the case that $\sum_{j=1}^{\overline{B}} \epsilon_j^2 \geq 0$. We conclude that

$$\sum_{j=1}^{\overline{B}} w_j^2 \geq \sum_{j=1}^{\overline{B}} \left(w_j^*\right)^2, \quad (53)$$

with equality if and only if $\epsilon_j = 0$ for all $j$. Thus QS attains greater higher-order power for equivalent higher-order size relative to all alternative WOS estimators using the Fourier basis,[20] and it minimizes the expression in (51). Combined with the fact that the Fourier basis achieves the size-power frontier for any set of weights as established above, QS thus dominates the size-power tradeoff for WOS estimators, and therefore globally among the families considered here.

For the QS kernel, $\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx = 3\pi\sqrt{10}/25$, since Priestley (1981) gives that $\sqrt{k^{(2)}(0)} = \pi^2/10$ (Table 7.1) and $\int_{-\infty}^{\infty} k^2(x)dx = 6/5$ (Table 6.1). Combining this with (27) yields (28) up to higher-order terms. Numerically computing $\overline{a}_{m,\alpha,q} = \sup_{\delta} a_{m,\alpha,q}(\delta)$ for $q = 2$ and $\alpha = 0.05$ yields $\overline{a}_{m,\alpha,q} 3\pi\sqrt{10}/25 \approx 0.3368$ for $m = 1$, $\overline{a}_{m,\alpha,q} 3\pi\sqrt{10}/25 \approx 0.6460$ for $m = 2$, and $\overline{a}_{m,\alpha,q} 3\pi\sqrt{10}/25 \approx 0.9491$ for $m = 3$, as stated.

(ii) As after (14), only equal-weighted orthonormal series estimators yield fixed-*b* asymptotic distributions that are exact *t* (or exact *F* in the multivariate case). The proof of part (i) of the theorem

[20] Note that while the WOS representation of a kernel estimator in (6) may not align exactly with (5) in finite samples (see Footnote 7), an immediate calculation gives that $\left(\nu_{WOS}^{QS}\right)^{-1} = \sum_{j=1}^{B}\left(w_j^*\right)^2 = 6b/5 + o(b) = \left(\nu_{\text{kernel}}^{QS}\right)^{-1} + o(b)$, where the final equality uses (14) and Priestley (1981, Table 6.1). Thus QS minimizes $\ell^{(2)}(k)$ up to (at worst) $o(b/T)$, so (28) holds.

implies immediately that given equal weights, the Fourier basis functions achieve the frontier, so the EWP test achieves the frontier among tests with exact *t* and *F* asymptotic fixed-*b* distributions.

Priestley (1981) Table 7.1 gives that $k^{(2)}(0) = \pi^2 / 6$ for the Daniell kernel, i.e. the EWP estimator. Further, $\psi = 1$ for this estimator. Combining these with (27) yields (29) up to higher-order terms. Again computing $\bar{a}_{m,\alpha,q} = \sup_{\delta} a_{m,\alpha,q}(\delta)$ for $q = 2$ and $\alpha = 0.05$, we obtain $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 0.3624$ 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 $m = 3$, as stated. □

**Proof of Remark 3:**

(a) For kernel estimators, Priestley (1981, eq. (6.2.123)) extends a result of Parzen (1957) to show that given a process with known mean, equation (15) holds without the terms in *b* if $b^q T^{q-1} \to 0$. Thus

$$\text{bias}\left(\hat{s}_z(0)\right) = E\hat{s}_z(0) - s_z(0) = -\left(bT\right)^{-q} k^{(q)}(0)\, s_z^{(q)}(0) + o\left(\left(bT\right)^{-q}\right), \tag{54}$$

and this equation holds as well for WOS estimators (including with unknown mean) by (13). For variance, (17) holds in both cases, so that

$$\text{var}\left(\hat{s}_z(0)\right) = 2\nu^{-1}\left(s_z(0)\right)^2 + o(b). \tag{55}$$

Thus, up to higher-order terms, $MSE\left(\hat{s}_z(0)\right) = \left(bT\right)^{-2q}\left(k^{(q)}(0)\, s_z^{(q)}(0)\right)^2 + 2b\psi\left(s_z(0)\right)^2$, which is shown by Priestley (1981, eq. (7.5.9)) to satisfy $\min_b MSE\left(\hat{s}_z(0)\right) \propto \left(\ell^{(q)}(k)\right)^{2q/(2q+1)}$.

(b) Using (54) and (55), the objective function evaluates to

$$a\left(bT\right)^{-q} k^{(q)}(0)\left|s_z^{(q)}(0)\right| + 2(1-a)b\psi\left(s_z(0)\right)^2 + o\left(\left(bT\right)^{-q}\right) + o(b). \tag{56}$$

The minimizing value of *b* is invariant, up to a multiplicative constant, to transformations of the objective function of the form

$$a_1\left(bT\right)^{-q} k^{(q)}(0)\left|s_z^{(q)}(0)\right| + a_2 b\psi\left(s_z(0)\right)^2 + o\left(\left(bT\right)^{-q}\right) + o(b), \tag{57}$$

for $a_1, a_2 > 0$. Sun and Yang (2020, p. 11) show that (i) objective function (e) can be expressed in this form, and (ii) its minimum is achieved for $b \propto \left(k^{(q)}(0)/\psi\right)^{1/(q+1)} T^{-q/(q+1)}$ (see also LLSW (2018, rejoinder eq. (1)), so that the minimized objective function is, to higher order and up to an additive constant, proportional to $\left(\ell^{(q)}(k)\right)^{q/(q+1)}$.

(c) By (44) and (45), both objectives can be expressed in the form (57), so part (b) applies.

(d) LLSW (2018, eq. (24)-(25)) show that at the optimum with this objective, $\Delta_S, \Delta_P^{\max}$ are both proportional to $\left(\ell^{(q)}(k)\right)^{q/(q+1)}$.

(e) See part (b). □

## Tables and Figures

**Table 1**. Maximum power loss of same-sized EWP (with $B$ series) compared to QS.

| | ***B*** | | |
|---|---|---|---|
| *m* | 4 | 8 | 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 with $B$ terms.

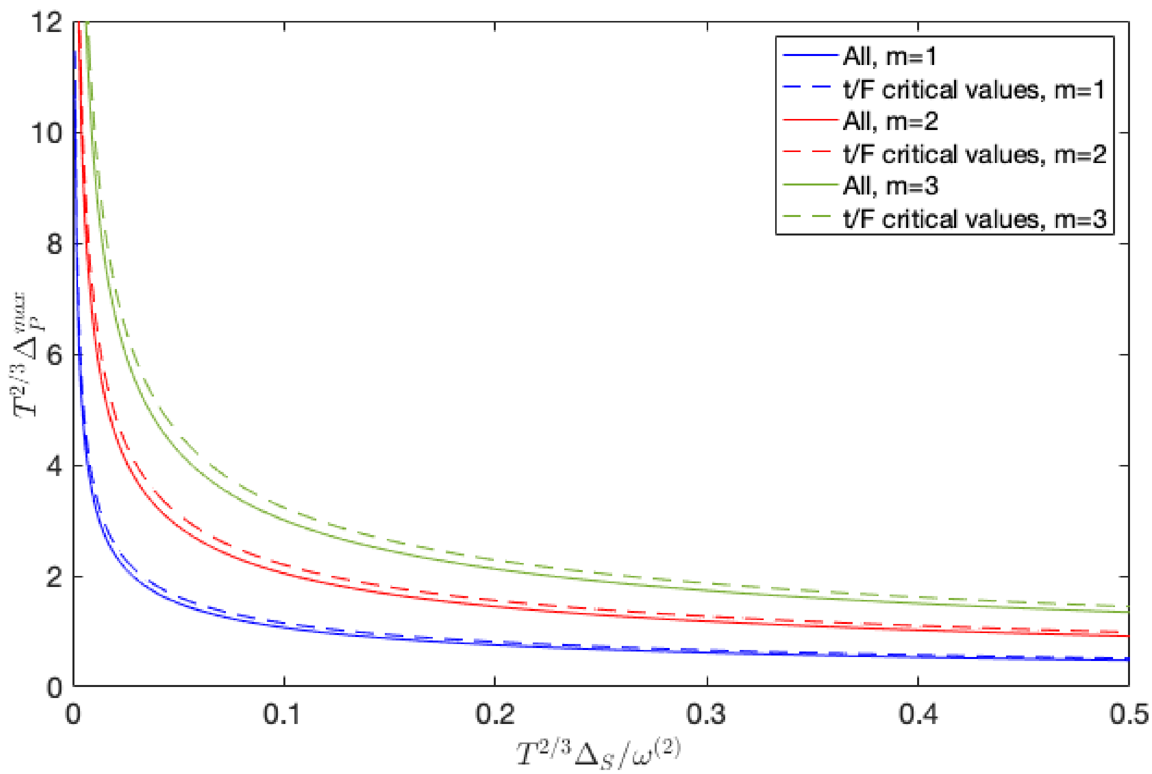


**Figure 1**. Higher-order frontier between the size distortion $\Delta_S$ and the maximum power loss $\Delta_P^{\max}$ of HAR tests in the Gaussian location model with dimension $m$, for stationary processes with normalized spectral curvature $\omega^{(2)}$. Solid line: all kernel and orthonormal series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