

The Size-Power Tradeoff in HAR Inference

August 21, 2019

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 Kolokotronis, 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 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 use higher-order expansions to provide a size-power frontier for 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 multivariate location model that dominates the theoretical literature, but simulations suggest the qualitative findings extend to stochastic regressors.

JEL codes: C12, C13, C18, C22, C32, C51

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

1. Introduction

Heteroskedasticity- and autocorrelation-robust (HAR) standard errors are used to construct test statistics and confidence intervals for the coefficients in time series regression when the regression errors u_t are potentially heteroskedastic and/or serially correlated. Computing HAR standard errors entails estimating the long-run variance (LRV) Ω , which is the sum of the autocovariances of $z_t = x_t u_t$, where x_t is the regressor.

The foundational papers on HAR inference in econometrics are Newey and West (1987) and Andrews (1991). The Newey-West/Andrews method estimates the LRV 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. A truncation parameter sequence S_T is chosen to ensure consistency, and inference uses standard normal or chi-squared critical values.

The Newey-West and Andrews papers stimulated a large theoretical literature on HAR inference, surveyed by Müller (2014). This literature reaches three broad conclusions. First, choosing S to minimize the mean squared error (MSE) of the LRV estimator, as suggested by Andrews (1991) and Newey and West (1994), produces a value of S that is generally too small from a testing perspective in the sense that it can lead to rejection rates under the null that differ substantially from the nominal level.¹ The asymptotic expansions of Velasco and Robinson (2001) and Sun, Phillips and Jin (2008) show that the leading higher order terms of the null rejection rate of the test are a weighted sum of the variance and the bias, not the squared bias, which enters the MSE. The testing problem calls for less bias, and thus a larger truncation parameter, than minimizing the MSE.

Second, using a large truncation parameter introduces another problem: a large value of S can introduce enough sampling variation into the LRV estimator that the standard asymptotic chi-squared approximation to the distribution of the HAR test deteriorates.² Thus addressing the

¹ Den Haan and Levin (1994, 1997) provided early Monte Carlo evidence of the large size distortions of HAR tests computed using the Newey-West/Andrews approach.

² This apparent bias-variance tradeoff holds when increasing S from an initially small value. In general, however, the use of estimated regression coefficients in constructing \hat{z}_t produces a downward bias as S increases further; see Hannan (1958), Ng and Perron (1996), and Velasco and Robinson (2001) for analysis in the location model. The latter effect tends not to perfectly

“bias” size distortion introduces a “variance”-related size distortion when chi-squared critical values are used. Fortunately, this additional size distortion can be addressed by replacing chi-squared critical values by Kiefer and Vogelsang’s (2005) “fixed b ” (or “fixed smoothing”) critical values, which are in general nonstandard. Fixed- b distributions are obtained by letting S_T grow proportionately to the sample size, that is, by fixing $b = S_T/T$ as T increases. Jansson (2004), Sun, Phillips and Jin (2008), and Sun (2014b) show that using fixed- b critical values provides a higher-order refinement to the null rejection rate of HAR test statistics in the location model.

Third, numerical results and some theory in the literature indicate that, for a given kernel, larger values of S reduce power, and that this tradeoff depends on the kernel. However, formal results laying out this size-power tradeoff have remained elusive, as have results on the optimal choice of kernel for testing. Consequently, no clear guidance exists for HAR kernel choice. And the practitioner who chooses a kernel and b still needs to generate (or approximate) nonstandard fixed- b critical values. It is unsurprising that despite the substantial theoretical progress in HAR testing theory, empirical practice remains dominated by the Andrews/Newey-West methodology with small truncation parameters and normal/chi-squared critical values.

This paper characterizes the tradeoff between the size distortion and the power of HAR tests. By size distortion, we mean the difference between the null rejection rate and the desired nominal significance level α . 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 α . 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.³

offset the former, but the relative strengths of the effects depend on the data-generating process in question. Fixed- b critical values, however, account for both of these effects.

³ 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 explicit expressions for second-order corrections to critical values, (ii) imposing those corrections to ensure that tests have the same second-order size, then (iii) comparing expressions for their size-adjusted power. See for example the

The class of LRV estimators we consider is the union of two families: the familiar positive semidefinite (psd) kernel estimators considered by Andrews (1991), and weighted orthonormal series (WOS) estimators (see for example Grenander and Rosenblatt (1957) and, more recently, Müller (2007), Phillips (2005), and Sun (2013)).⁴ All tests are evaluated using fixed- b critical values. Weighted orthonormal series estimators are computed by projecting \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, and then evaluating the weighted sum of these squared projections. The leading example of a WOS estimator is the equal-weighted periodogram (EWP) estimator, which equivalently can be thought of as a series estimator using the first B Fourier series $\{\sin(2\pi jt/T), \cos(2\pi jt/T)\}, j = 1, \dots, B/2$. In the location model, we show that Ibragimov and Müller's (2010) subsample estimator is an equal-weighted WOS estimator. As shown by Brillinger (1975, exercise 5.13.25) for Fourier series in the location model, and more generally by Phillips (2005), Müller (2007), and Sun (2013), the fixed- b asymptotic distribution of equal-weighted series HAR tests using B orthonormal series is t_B or, in the case of tests of m restrictions, is $F_{m, B-m+1}$ (after rescaling).⁵ This feature makes equal-weighted WOS tests attractive in practice.

This paper makes five main contributions. First, using the small- b asymptotic expansions of Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2011, 2013, 2014b) for

discussion in Rothenberg (1984), which draws on Pfanzagl and Wefelmeyer (1978), about how the one-sided Lagrange multiplier, likelihood ratio, and Wald tests are all second-order efficient in the scalar normal means model. More closely related, Sun, Phillips, and Jin (2008, Corollary 5) used 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 Sul, Phillips, and Choi (2005), Kiefer and Vogelsang (2002), 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.

⁴ These two families represent the leading cases considered in the post-Andrews/Newey-West theoretical literature (discussed further at the end of this section), excluding bootstrap estimators and estimators that are not psd with probability one (which we rule out following the bulk of the empirical literature).

⁵ As part of our unification of kernel and WOS tests, we establish in Section 3 that, for WOS tests, it is natural to set $b = B^{-1}$.

the Gaussian location model, we derive theoretical expressions characterizing the tradeoff between the size distortion and the power loss arising from the choice of b for a given HAR test. In so doing, we confirm a conjecture by Kiefer and Vogelsang (2005) that small- b expansions around fixed- b critical values could be used to characterize the size-power tradeoff for HAR tests. These results apply when $b \rightarrow 0$ at the rate for which the size distortion and power loss have the same asymptotic order, which coincides with the optimal rate in Sun, Phillips, and Jin (2008).

Second, we derive the size/power frontier in the Gaussian location model, which is the envelope of the size-power tradeoffs in the class of tests we consider, and we show that this frontier is achieved by the QS kernel. This frontier has a simple form. Let Δ_S be the size distortion of the test, implemented using fixed- b asymptotic critical values, and let Δ_P^{\max} be the maximum size-adjusted power loss of the test over all alternatives, relative to the infeasible test with known LRV. For a 5% test in the one-dimensional location model ($m = 1$), this 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 (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 the frontier (1) is that the constant increases with m (the constants are provided in Section 4). The frontier is plotted in Figure 1 for 5% tests for $m = 1, 2$, and 3. Choosing a sequence of values for b to equate the rates at which Δ_S and Δ_P converge to zero in (1) yields $\Delta_S, \Delta_P = O(T^{-2/3})$, and this is used to derive (1) and to scale the axes in Figure 1. For the Bartlett (Newey-West, tent) kernel, equating these rates yields $\Delta_S, \Delta_P = O(T^{-1/2})$, so the Bartlett kernel HAR test is asymptotically dominated.

Third, we consider the effect of restricting the class of HAR tests further to tests for which the limiting fixed- b distributions are standard t and F , so that the test does not require simulation or special tables for critical values. For a 5% level test with $m = 1$, this frontier is given by

$$\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{0.3623}{T} + o(T^{-1}). \quad (2)$$

This bound is plotted for $m = 1, 2,$ and 3 as the dashed lines in Figure 1. We show that this frontier is achieved by Brillinger’s (1975) EWP test and by the closely related equal-weighted cosine (EWC) test, in which the LRV is estimated using the Type II cosine basis functions (Müller (2007)). Relative to the EWP test, the EWC test has the advantage of admitting all integer values of B , whereas B must be even for the EWP test. As suggested by the numerically close constants in (1) and (2) and by Figure 1, the cost of this restriction to t - or F -based fixed- b inference is quite small. In a separate result in Section 4, we provide an expression for the power difference between two same-sized tests. This expression does not depend on the stochastic process followed by z_t or on T . For the EWP test with $B = 8$ (first four sines and cosines) and $m = 1$, the power loss, relative to the same-sized QS test, is at most 0.0074.

Fourth, we find that, in Monte Carlo simulations of the scalar and multivariate location model, the theoretical size/power tradeoffs provide an accurate description of the size distortions and power losses observed in finite samples, for sample sizes and degrees of persistence typically found in empirical work. The fit of these bounds breaks down at high levels of persistence, as expected from Müller (2007, 2014).⁶

Fifth, we also perform Monte Carlo simulations of the multiple regression model. The theoretical results for the Gaussian location model do not apply here because the process for $z_t = x_t u_t$ is non-Gaussian even if x_t and u_t are Gaussian, and we find that the location model frontier is more favorable than the Monte Carlo frontier. Still, the theoretical findings in the location model seem to hold numerically in the regression case. Extensive additional simulation results are given in a companion paper (Lazarus, Lewis, Stock, and Watson (2018)). That paper uses the

⁶ We have further considered the performance of tests with feasible (data-dependent) adjustments to the fixed- b critical value, as it can be shown that consistent estimation of the normalized curvature parameter $\omega^{(2)}$ allows for a higher-order refinement relative to the bounds (1) and (2). But the Monte Carlo performance of tests with these adjusted critical values is mixed at best, with improvements relative to the frontier only evident in very large sample sizes. (We find this unsurprising in light of the difficulty of estimating the curvature of the spectral density at frequency zero.) We accordingly do not pursue this procedure in much detail.

theoretical results in this paper, along with simulation evidence, to provide recommendations for truncation parameter rules based on trading off size and size-adjusted power.

This paper relates to a large literature. The starting point for our results is the fixed- b asymptotic expansions in Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2013, 2014b). Relative to these papers, the technical distinction is our focus on size-adjusted power rather than unadjusted rejection rates under the alternative. This paper also relates to the literature on orthonormal series estimators, see Phillips (2005), Müller (2007), and Sun (2013) for multiple references. Relative to this literature, we extend the results to weighted orthonormal series estimators, provide a ranking of the small- b performance of the estimators, and unify existing small- b expansions for kernel and weighted orthonormal series expansions using what we refer to as the implied mean kernel of weighted orthonormal series LRV estimators. Although this paper does not consider bootstrap procedures, some papers in the bootstrap literature are germane. In particular, the results in Gonçalves and Vogelsang (2011) suggest that tests with critical values from the moving block bootstrap also satisfy our size/power tradeoff expressions and the frontiers (1) and (2), although we do not pursue this conjecture. The results of Zhang and Shao (2013) suggest that their Gaussian bootstrap improves upon the frontier bounds derived here. There is also a literature on HAR tests using non-psd estimators (e.g. Sun, Phillips, and Jin (2006) and Politis (2011)). Following the empirical literature, we restrict attention to tests that are psd with probability one and do not address non-psd tests.

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 collects results from the literature on fixed- b asymptotics and asymptotic expansions. Section 4 provides our main results. The Monte Carlo study is summarized in Section 5, with additional results provided in the Supplement. Section 6 concludes. Proofs are given in the Appendix.

2. Notation and Class of LRV Estimators

2.1 The HAR testing problem

Let z_t be an $m \times 1$ time series with autocovariances $\Gamma_j = \text{cov}(z_t, z'_{t-j})$, $j = 0, 1, \dots$, and long-run variance

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j. \quad (3)$$

In general z_t depends on a vector β of unknown parameters, so that $z_t = z_t(\beta)$, although after preliminary definitions we suppress this dependence. We consider tests of the null hypothesis $\beta = \beta_0$ against the alternative $\beta \neq \beta_0$.⁷ We consider test statistics of the form,

$$t = \frac{\sqrt{T}\bar{z}_0}{\sqrt{\hat{\Omega}}} \text{ if } m = 1, \text{ and} \quad (4)$$

$$F = T\bar{z}_0' \hat{\Omega}^{-1} \bar{z}_0 / m \text{ if } m > 1, \quad (5)$$

where $\bar{z}_0 = T^{-1} \sum_{t=1}^T z_t(\beta_0)$ and $\hat{\Omega}$ is an estimator of Ω . The test statistics (4) and (5) arise in time series regression, in GMM estimation, and in the multivariate location model.

For our theoretical results, we specialize the HAR problem to testing the mean of a stationary Gaussian process. Doing so allows us to exploit the well-developed literature on Edgeworth expansions for that problem. Specifically, we consider HAR tests of $\beta = \beta_0$ in the model,

$$y_t = \beta + u_t, \quad t = 1, \dots, T, \quad (6)$$

where y_t is $m \times 1$, β is the vector of means of y_t , and u_t is an $m \times 1$ vector of Gaussian disturbances that is potentially heteroskedastic and/or autocorrelated. In this model, the usual t -statistic (if $m = 1$) or F -statistic (if $m > 1$) testing $\beta = \beta_0$ is respectively (4) or (5), where $z_t(\beta_0) = y_t - \beta_0$.⁸

⁷ As discussed in Section 4, our results extend to subvector inference.

⁸ In our Monte Carlo study in Section 5, we additionally consider the more general time series regression model, $y_t = x_t' \beta + u_t$, where the dependent variable y_t is scalar and there are m regressors x_t . In this model, $z_t(\beta_0) = x_t(y_t - x_t' \beta_0)$, and the statistics (4) and (5) test the null hypothesis that $\beta = \beta_0$. In this model, $\hat{z}_t = x_t \hat{u}_t$, where $\hat{u}_t = y_t - x_t' \hat{\beta}$ is the OLS residual and $\hat{\beta}$ is the OLS estimator of β , and the estimator $\hat{\Omega}$ is computed using the estimated $\hat{\beta}$ and $\hat{z}_t = z_t(\hat{\beta})$.

We consider the class of estimators $\hat{\Omega}$ comprised of two overlapping families of estimators, the family of psd kernel estimators and the family of weighted orthonormal series estimators. These estimators are computed using estimated coefficients $\hat{\beta}$ and $\hat{z}_t = z_t(\hat{\beta})$. In the multivariate location model, $\hat{z}_t = y_t - \bar{y}$, where \bar{y} is the sample mean of y_t .

These estimators belong to the larger class of estimators that take a quadratic form in the \hat{z}_t series, i.e.,

$$\hat{\Omega}^{OF} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T A_b \left(\frac{s}{T}, \frac{t}{T} \right) \hat{z}_t \hat{z}_s' = \frac{1}{T} \hat{z}' \mathbf{A}_b \hat{z}, \quad (7)$$

where \mathbf{A}_b is symmetric psd matrix with (s, t) element $A_b(s/T, t/T)$, and $\hat{z} = [\hat{z}_1 \ \dots \ \hat{z}_T]'$.⁹

2.2 Kernel estimators

Kernel estimators of Ω are weighted sums of sample autocovariances using the weight function, or kernel, $k(\cdot)$:

$$\hat{\Omega}^{SC} = \sum_{j=-(T-1)}^{T-1} k(j/S) \hat{\Gamma}_j, \quad \text{where } \hat{\Gamma}_j = \frac{1}{T} \sum_{t=\max(1, j+1)}^{\min(T, T+j)} \hat{z}_t \hat{z}_{t-j}', \quad (8)$$

where S is the truncation parameter and the superscript “SC” denotes sum-of-covariances. The Newey and West (1987) estimator uses the Bartlett kernel, $k(x) = (1 - |x|)\mathbf{1}(|x| \leq 1)$; see Priestley (1981) and Andrews (1991) for other examples.

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

$$\hat{\Omega}^{WP} = 2\pi \sum_{j=1}^{T-1} \tilde{w}_j I_{\hat{z}\hat{z}}(2\pi j/T), \quad (9)$$

⁹ See Grenander and Rosenblatt (1957) for an early consideration of such estimators in the context of spectral density estimation; Müller (2007) and Sun (2014a) provide more recent examples.

where $I_{zz}(\omega)$ is the periodogram of \hat{z}_t at frequency ω ,

$$I_{\hat{z}\hat{z}}(\omega) = (2\pi)^{-1} d_{\hat{z}}(\omega) \overline{d_{\hat{z}}(\omega)'} , \text{ where } d_{\hat{z}}(\omega) = T^{-1/2} \sum_{t=1}^T \hat{z}_t e^{-i\omega t} \quad (10)$$

and where the weights $\{\tilde{w}_j\}$ in (9) satisfy $\tilde{w}_j = T^{-1} \sum_{u=0}^{T-1} k(u/S) e^{i2\pi ju/T}$. For large S ,

$\tilde{w}_j \sim \pi SK(S \cdot 2\pi j / T)$, where

$$K(\omega) = \frac{1}{2\pi} \int_{u=-\infty}^{\infty} k(u) e^{-i\omega u} du \quad (11)$$

is the spectral window generator; see Priestley (1981, pp. 447-448 and 580-581) and Andrews (1991). Kernel estimators are positive semidefinite with probability one if $K(\omega)$ is nonnegative.

An important special case of kernel estimators is the equal-weighted periodogram (EWP) estimator, which is computed using the Daniell kernel which in the frequency domain places equal weight on the first $B/2$ periodogram terms, where B is even:

$$\hat{\Omega}^{EWP} = \frac{2}{B} \sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j / T) = \frac{2}{B} \sum_{j=1}^{B/2} d_{\hat{z}}(2\pi j / T) \overline{d_{\hat{z}}(2\pi j / T)'} . \quad (12)$$

A second important special case is the quadratic-spectral (QS) or Epanechnikov (1960) kernel, for which $\tilde{w}_j \propto [1 - (|j| / (B/2))^2]$ for $|j| \leq B/2$ and $\tilde{w}_j = 0$ for $j > B/2$.

2.3 Weighted orthonormal series estimators

Weighted orthonormal series (WOS) estimators are obtained by projecting \hat{z}_t onto B mean-zero low-frequency functions of a set of orthonormal functions, typically the first mean-

zero elements of a basis for $L^2[0,1]$, and then evaluating the weighted sum of these projections.¹⁰ Following Sun (2013), let $\{\phi_j(s)\}, j = 0, \dots, B, 0 \leq s \leq 1$, denote the first $B+1$ functions in an orthonormal basis for $L^2[0,1]$, where $\phi_0(s) = 1$ and $\int_0^1 \phi_j(s)ds = 0$ for $j \geq 1$. Let Φ denote the $T \times B$ matrix consisting of $\{\phi_j(s)\}, j = 1, \dots, B$, evaluated at t/T :

$$\Phi = [\Phi_1 \dots \Phi_B], \text{ where } \Phi_j = [\phi_j(1/T) \ \phi_j(2/T) \ \dots \ \phi_j(1)]', \Phi' \Phi / T = I_B, \text{ and } \iota_T' \Phi = 0, \quad (13)$$

where ι_T is the T -vector of ones.¹¹

The weighted orthonormal series (WOS) LRV estimator is,

$$\hat{\Omega}^{WOS} = \sum_{j=1}^B w_j \hat{\Omega}_j^{OS}, \text{ where } \sum_{j=1}^B w_j = 1, \hat{\Omega}_j^{OS} = \hat{\Lambda}_j \hat{\Lambda}_j', \text{ and } \hat{\Lambda}_j = \sqrt{\frac{1}{T}} \sum_{t=1}^T \phi_j(t/T) \hat{z}_t, \quad (14)$$

and $w_j > 0, j = 1, \dots, B$. Alternatively, we can write $\hat{\Omega}^{WOS} = \hat{z}' \tilde{\Phi} \tilde{\Phi}' \hat{z} / T$,

$\tilde{\Phi} = [\sqrt{w_1} \Phi_1 \ \dots \ \sqrt{w_B} \Phi_B]$. Note that Φ and $\hat{\Omega}^{OS}$ omit the $j = 0$ function, for which $\Phi_0 = \iota_T$

and $\Phi_0' \Phi = 0$. Evidently $\hat{\Omega}^{WOS}$ is a member of the class of quadratic form estimators (7), where $\mathbf{A}_b = \tilde{\Phi} \tilde{\Phi}'$. By construction, $\hat{\Omega}^{WOS}$ is psd with probability one.

While the spectral density estimation literature has considered general weighted orthonormal series estimators (see Hannan (1970), Brillinger (1975), Priestley (1981), and Stoica and Moses (2005)), recent literature on this class for purposes of HAR estimation specializes (14) to the case of equal weights $w_j = 1/B, j = 1, \dots, B$ (see Phillips (2005), Müller (2007), and

¹⁰ 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.

¹¹ If for a given T , $\Phi_j = [\phi_j(1/T) \ \phi_j(2/T) \ \dots \ \phi_j(1)]'$ does not satisfy $\iota_T' \Phi_j = 0$ and $\Phi' \Phi / T = I_B$, the finite-sample version Φ can be constructed as the orthonormalization of the demeaned $\{[\phi_j(1/T) \ \phi_j(2/T) \ \dots \ \phi_j(1)]'\}$. This is worked out explicitly for the step-function basis below. We assume that sequence which is the difference between these finite-sample adjusted bases and the basis defined in (13) converges to zero; this follows from Lemma A of Phillips (2005), which shows that $\Phi' \Phi / T = I_B + O(1/T)$ for unadjusted basis functions.

Sun (2013)). In keeping with the HAR literature, we refer to equal-weighted WOS estimators simply as orthonormal series (OS) estimators:

$$\hat{\Omega}^{OS} = \frac{1}{B} \sum_{j=1}^B \hat{\Omega}_j^{OS}, \quad (15)$$

where $\hat{\Omega}_j^{OS}$ is defined in (14).

Kernel estimators computed using a frequency domain weight function with bounded support (so that $\tilde{w}_j = 0$ for $j > B$ in (9)) are a special case of WOS estimators, where the basis functions are the Fourier functions and $w_j \propto \tilde{w}_j$. These include the EWP and QS estimators. Because the frequency-domain weights \tilde{w}_j corresponding to the Bartlett (Newey-West) kernel are in general nonzero for all $j = 1, \dots, T/2$, the Newey-West estimator does not fall in this class, however the frequency-domain approximation to the Newey-West estimator computed by truncating the weights \tilde{w}_j does fall in the WOS class. In our asymptotics, we consider increasing sequences of B , so the WOS family encompasses all kernel estimators asymptotically but not in finite samples.

The theory in this paper covers all basis functions that have three bounded derivatives, plus a non-differentiable basis function based on splitting the sample. The basis functions we examine explicitly in our simulation study are Fourier, Type II cosine, and split-sample.

Fourier basis functions are comprised of the B real-valued sine and cosine series,

$$\{\phi_{2j-1}(s), \phi_{2j}(s)\} = \left\{ \sqrt{2} \cos(2\pi js), \sqrt{2} \sin(2\pi js) \right\}, j = 1, \dots, B/2. \quad (16)$$

Cosine basis functions. Müller (2007) and Müller and Watson (2008) suggest using as basis functions the type II discrete cosine transform, which are the eigenvectors of the covariance kernel of a demeaned Brownian motion:

$$\{\phi_{Tj}(s)\} = \left\{ \sqrt{2} \cos \left[\pi j \left(\frac{s-1/2}{T} \right) \right] \right\}, j = 1, \dots, B. \quad (17)$$

Hwang and Sun (2018) refer to this as the shifted cosine function.¹²

Split-sample step function. Ibragimov and Müller (2010) propose estimating the long-run variance by estimating β on $B+1$ equal-sized subsamples and estimating Ω using the sample variance of these subsample estimators.¹³ For convenience, suppose $T/(B+1)$ is an integer. For a single coefficient, their split-sample (SS) test statistic is,

$$t^{SS} = \sqrt{B+1} \left(\bar{\hat{\beta}} - \beta_0 \right) / \sqrt{S_{\hat{\beta}}^2}, \text{ where } S_{\hat{\beta}}^2 = \frac{1}{B} \sum_{i=1}^{B+1} \left(\hat{\beta}^{(i)} - \bar{\hat{\beta}} \right)^2, \quad (18)$$

where $\hat{\beta}^{(i)}$ is the estimator of β computed using the i^{th} subsample and $\bar{\hat{\beta}} = \frac{1}{B+1} \sum_{i=1}^{B+1} \hat{\beta}^{(i)}$.

In the location model, the SS t -statistic (18) can be written in the standard form (4), where the LRV estimator is the equal-weighted orthonormal series estimator,

$$\hat{\Omega}^{SS} = \hat{z}' \Phi^{SS} \Phi^{SS'} \hat{z} / (BT), \text{ where } \Phi^{SS} = \sqrt{(B+1)} \left(I_{B+1} \otimes \iota_{T/(B+1)} \right) M_t^B, \quad (19)$$

where ι_r denotes the r -vector of ones, M_t^B is the $(B+1) \times B$ matrix of eigenvectors corresponding to the B unit eigenvalues of the idempotent matrix $M_t = I_{B+1} - \iota_{B+1} \iota_{B+1}' / (B+1)$, and \otimes is the Kronecker product (the derivation of (19) is given in the Appendix).

We will refer to Φ^{SS} in (19) as the SS orthonormal series and to $\hat{\Omega}^{SS}$ as the SS series LRV estimator. The SS basis functions are discontinuous step functions. For $B = 2^n - 1$, where n is an integer, the SS and Haar basis functions are equivalent, however for other B the Haar functions

¹² A closely related alternative is Phillips's (2005) proposal of using $\{\sqrt{2} \sin[\pi j((s-1/2)/T)]\}$, $j = 1, \dots, B$, which are the eigenvectors of the covariance kernel of Brownian motion. Phillips (2005) shows that, for $B \rightarrow \infty$ and $B/T \rightarrow 0$, the equal-weighted sine series estimator is asymptotically equivalent to $\hat{\Omega}^{EWP}$ to second order; we obtain a similar result below for the cosine basis.

¹³ This subsample variance estimator is also referred to as the "batch mean estimator" in previous literature, for example Song and Schmeiser (1993) compare the batch mean estimator to conventional kernel LRV estimators. We thank Yixiao Sun for pointing us to this literature.

do not span Φ^{SS} . Although the expression for $\hat{\Omega}^{SS}$ in (19) was developed for the location model with $m = 1$, it generalizes directly to $m > 1$ (and further to the time series regression model).¹⁴

Implied mean kernels of weighted orthonormal series estimators. Toward unifying the WOS and kernel estimators, we now show that the mean of every WOS LRV estimator has an approximate kernel representation, which becomes exact as $T \rightarrow \infty$. We refer to the limiting kernel of this mean as the implied mean kernel.

Use the definition of $\hat{\Omega}_j^{OS}$ in (14) and the device in Grenander and Rosenblatt (1957, p. 125) to express the mean of the j^{th} contribution to a WOS estimator as,

$$\begin{aligned} E\hat{\Omega}_j^{OS} &= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)' \\ &= \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\phi_j(t/T)\phi_j(s/T)\Gamma_{s-t} + O(1/T) \\ &= \sum_{u=-(T-1)}^{T-1}\tilde{k}_{j,T}^{OS}(u/T)\Gamma_u + O(1/T), \end{aligned} \tag{20}$$

where¹⁵

$$\tilde{k}_{j,T}^{OS}(u/T) = \sum_{t=1}^T\phi_j(t/T)\phi_j((t-u)/T)\mathbf{1}(1 \leq t-u \leq T). \tag{21}$$

¹⁴Both the Ibragimov-Müller test statistic (18) and the test based on $\hat{\Omega}^{SS}$ in (19) generalize to time series regression with stochastic regressors, however the test statistics are no longer the same outside the location model. Mechanically, this distinction arises because the IM t -statistic is based on the sample variance of the subsample estimators of β in which both the numerator and denominator are computed using the subsample, whereas plugging $\hat{\Omega}^{SS}$ into (4) uses the subsample estimates of β and the full-sample estimator of the x_t second moment matrix. This distinction prevents giving the IM statistic in (18) an orthonormal series interpretation in the general regression model.

¹⁵ The second equality in (20) follows from writing $\hat{z}_t = y_t - \bar{y} = (y_t - \beta) - (\bar{y} - \beta)$, taking the expectation, and noting that $E(y_t - \beta)(\bar{y} - \beta)' = T^{-1}\sum_{s=1}^T\Gamma_{t-s} = O(T^{-1})$, that $E(\bar{y} - \beta)(\bar{y} - \beta)' = T^{-2}\Omega + o(T^{-2})$, and that $\Phi_j'1_T = 0$ (as assumed in (13)).

Thus

$$E\hat{\Omega}^{WOS} = \sum_{j=1}^B E\left(w_j \hat{\Omega}_j\right) = \sum_{u=-(T-1)}^{T-1} k_{B,T}^{WOS}(u/S) \Gamma_u + O(1/T), \quad (22)$$

where $k_{B,T}^{WOS}(u/S) = \sum_{j=1}^B w_j \tilde{k}_{j,T}^{OS}\left(B^{-1} \frac{u}{S}\right)$, 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)). We show below that these definitions imply that, for WOS estimators, it is natural to define $b = B^{-1}$.

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

$$k_B^{WOS}(x) = \sum_{j=1}^B w_j \tilde{k}_j^{WOS}(B^{-1}x) \quad \text{and} \quad \tilde{k}_j^{WOS}(x) = \int_{\max(0,x)}^{\min(1,1+x)} \phi_j(s) \phi_j(s-x) ds, \quad (23)$$

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

Equations (22) and (23) show that the mean, and thus bias, of WOS estimators have the same approximate form as kernel estimators, and that this form becomes exact as $T \rightarrow \infty$. Sections 3 and 4 below develop rejection-rate expansions and associated theoretical results under the large- B (i.e., small- b) sequence under which $B \rightarrow \infty$, and we accordingly define

$k^{WOS}(x) = \lim_{B \rightarrow \infty} k_B^{WOS}(x)$ for WOS estimators.¹⁶ Any expressions below for $k^{WOS}(\cdot)$ without the

¹⁶ More formally, one can define $k^{WOS}(\cdot)$ as the limit of $k_{B,T}^{WOS}(\cdot)$ as $B, T \rightarrow \infty$ s.t. $B/T \rightarrow 0$ in $L^2(\mathbb{R})$, as implied by Assumptions 2 and 3 in Section 4. The validity of the sequential-limit definition given here in the text in the context of these joint-limit assumptions is verified formally in the proof of Theorem 1(i) in the Appendix, and this point is discussed further in Section 4.

subscript B accordingly refer to these limiting implied mean kernels, and $k^{OS}(\cdot)$ refers to the equal-weighted case.

The implied mean kernels $k_{B,T}^{WOS}(u/S)$ for the equal-weighted Fourier, cosine, and SS basis functions all concentrate their mass on low frequencies (Supplemental Figures S.1 and S.2). In the frequency domain the SS implied mean kernel has considerably more leakage than the Fourier and cosine kernels.

The value ψ , defined as

$$\psi = \begin{cases} \int_{-\infty}^{\infty} k^2(x) dx & \text{for kernel estimators} \\ B \sum_{j=1}^B w_j^2 & \text{for WOS estimators,} \end{cases} \quad (24)$$

plays an important role in the size-power tradeoff for our families of estimators.

3. Summary of Fixed- b Asymptotics and Small- b Rejection Rate Expansions

Following Sun, Phillips, and Jin (2008) and Sun (2013, 2014a, 2014b), we consider HAR tests evaluated using fixed- b asymptotic critical values and approximate their rejection rates using “small- b ” asymptotic expansions.

3.1 Fixed- b asymptotics

Kernel and WOS estimators have the fixed- b limiting distribution,

$$\hat{\Omega} \xrightarrow{d} \Omega^{1/2} \left(\sum_{j=1}^{\infty} \lambda_j \Xi_j \right) \Omega^{1/2'}, \text{ where } \Xi_j \sim \text{i.i.d. } W_m(I_m, 1), \quad (25)$$

where $W_m(I_m, 1)$ denotes the standard m -dimensional Wishart distribution with one degree of freedom and where $\{\lambda_j\}$ are constants.

In the scalar case, the limiting distribution (25) is a classic result for truncated kernel estimators, for which $\tilde{w}_j = 0$ for $j > B$ in (9); then $\lambda_j = \tilde{w}_j$ for $j \leq B$ and $\lambda_j = 0$ for $j > B$, and

$\Xi_j \sim \text{i.i.d. } \chi_1^2$ (Brillinger (1981, p.145), Priestley (1981, p. 466)). For general kernel estimators, Sun (2014b) obtained the result (25) in the m -dimensional case, in which case $\{\lambda_j\}$ are the eigenvalues of a centered version of the kernel (see Sun (2014b, p. 662)).

For WOS estimators, the result (25) follows from the orthonormality of the projection matrix and the assumption that $T^{-1/2} \sum_{t=1}^T \phi_j(t/T) \hat{z}_t \xrightarrow{d} N(0, \Omega)$, which implies that

$\hat{\Omega}_j^{OS} \xrightarrow{d} \Omega^{1/2} \Xi_j \Omega^{1/2'}$, where again $\{\Xi_j\}$ are independent standard Wishart random variables with one degree of freedom. (This limit holds under Assumptions 1 and 3 in Section 4; see Phillips (2005), Müller (2007), and Sun (2013) for additional discussion and primitive conditions.) In the WOS case, $\lambda_j = w_j$, $j \leq B$, and $\lambda_j = 0$, $j > B$.

An important special case of (25) is equal-weighted WOS estimators. In the scalar case, $\hat{\Omega} / \Omega$ has a χ_B^2 / B distribution if and only if the weights are equal. Thus, the EWP/Daniell estimator is unique among kernel estimators in having a fixed- b asymptotic distribution that is exact $(\chi_B^2 / B)\Omega$. More generally, among the class of kernel and WOS series considered here, only equal-weighted orthonormal series estimators have a distribution proportional to χ_B^2 / B .

In the case of non-equal weights, Tukey (1949) proposed approximating the distribution of kernel spectral density estimators by a chi-squared with degrees of freedom chosen to match the first two moments of (25). Tukey's approximation is,

$$\hat{\Omega} \sim (\chi_\nu^2 / \nu)\Omega, \text{ where } \nu = (b\psi)^{-1}, \quad (26)$$

where ν is the "equivalent degrees of freedom" of $\hat{\Omega}^{WP}$ and ψ is as defined in (24).

Although Tukey (1949) proposed (26) for kernel estimators, it holds as well for WOS estimators, where $b = B^{-1}$.¹⁷

¹⁷ This result follows from first noting that $\text{var}(\hat{\Omega}^{WOS} / \Omega) = \text{var}\left(\sum_{j=1}^B w_j \hat{\Omega}_j^{OS} / \Omega\right) \approx 2 \sum_{j=1}^B w_j^2$. Equating second moments of the approximating distribution delivers $\nu = \left(\sum_{j=1}^B w_j^2\right)^{-1} = (B^{-1}\psi)^{-1} = (b\psi)^{-1}$, where the second equality uses (24) and the third uses $b = B^{-1}$.

Fixed- b distributions of HAR t -statistic. The fixed- b asymptotic distribution of the HAR test statistic obtains from the fixed- b limiting distributions of $\hat{\Omega}$ and the asymptotically independent normal distribution of $\sqrt{T}(\hat{\beta} - \beta)$. In general, this distribution is nonstandard, but for equal-weighted WOS estimators it is t or F .

Specifically, in the scalar case, if the LRV estimator has a fixed- b χ_B^2 distribution, then the corresponding HAR t -statistic has a fixed- b asymptotic distribution that is t_B . This result appears to date to Brillinger (1975, exercise 5.13.25), who considered the fixed- B EWP HAR test in the location model. An implication of the discussion in Section 3.1 is that the set of estimators with chi-squared fixed- b asymptotic distributions is the set of orthonormal series estimators with $w_j = 1/B$ for all j . Thus, among the family of psd kernel and WOS HAR tests, equal-weighted orthonormal series tests uniquely are t -distributed under fixed- b distributions.

For general $m \geq 1$, $\hat{\Omega}^{OS}$ has the fixed- B asymptotic distribution,

$$\hat{\Omega}^{OS} \xrightarrow{d} \Omega^{1/2}(\Xi_B/B)\Omega^{1/2}, \text{ where } \Xi_B \sim W_m(\mathbf{I}_m, B), \quad (27)$$

where $W_m(\mathbf{I}, B)$ denotes the standard Wishart distribution with dimension m and degrees of freedom B . Thus m times $F = F_T$ in (5) has an asymptotic Hotelling T^2 distribution.

As in Stock and Watson (2008) and Sun (2013), it is convenient to rescale F_T so that it has a fixed- b F distribution. We therefore consider the rescaled F statistic,

$$F_T^* = \frac{B - m + 1}{B} F_T. \quad (28)$$

where F_T is given in (5). When F_T is evaluated using $\hat{\Omega}^{OS}$, $F_T^* \xrightarrow{d} F_{m, B-m+1}$. This is again only the case for the equal-weighted WOS estimators.

3.2 Small- b Rejection Rate Expansions

Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2014b) (among others) provide higher-order expansions of the rejection rate of HAR tests in the Gaussian location model using kernel LRV estimators for small- b sequences satisfying $b \rightarrow 0$, $T \rightarrow \infty$, and

$S_T = bT \rightarrow \infty$. Sun (2013) provides small- b rejection rate expansions for orthonormal series HAR tests. Here, we provide expressions that use the implied mean kernel of WOS tests to unify these results and to extend them to WOS tests. In our unified expressions, k can be either a kernel or an implied mean kernel.

Like classical expansions of the MSE for spectral estimators, the expansions for kernel HAR tests depend on the kernel through the so-called Parzen characteristic exponent. We show below that the expansions for WOS tests also depend on the Parzen characteristic exponent of the implied mean kernel. The Parzen characteristic exponent q is the maximum integer such that

$$k^{(q)}(0) = \lim_{x \rightarrow 0} \frac{1 - k(x)}{|x|^q} < \infty. \quad (29)$$

The term $k^{(q)}(0)$ is called the q^{th} generalized derivative of k , evaluated at the origin. For the Bartlett (Newey-West) kernel, $q = 1$, while for the QS and Daniell kernels, $q = 2$.

The expansions also depend on the Parzen generalized derivative of the spectral density at the origin. Define

$$\omega^{(q)} = \text{tr} \left(m^{-1} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j \Omega^{-1} \right). \quad (30)$$

Without Ω^{-1} , (30) is the trace of 2π times the Parzen (1957) generalized q^{th} derivative of the spectral density at frequency zero; $\omega^{(q)}$ normalizes this by 2π times the spectral density at frequency zero. For the case $m = 1$ and $q = 2$, $\omega^{(2)} = -s_z''(0) / s_z(0)$, where $s_z(\omega)$ is the spectral density of z_t at frequency ω . If z_t follows a stationary AR(1) process with autoregressive coefficient ρ , then $\omega^{(2)} = 2\rho / (1 - \rho)^2$.

Sun's (2013, 2014b) small- b expansions of rejection rates for tests using fixed- b critical values play a central role in our analysis, so we summarize them here in unified notation. Let F_T^* denote the modified F statistic in (28), and let $c_m^\alpha(b)$ denote the fixed- b asymptotic critical value for the level α test with m degrees of freedom. In these expressions, k refers to the kernel or

implied mean kernel, where for WOS estimators $b = B^{-1}$. The asymptotic expansion of the null rejection rate is

$$\Pr_0 \left[F_T^* > c_m^\alpha(b) \right] = \alpha + G'_m(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} + o(b) + o\left((bT)^{-q}\right), \quad (31)$$

where G_m is the chi-squared cdf with m degrees of freedom, G'_m is the first derivative of G_m , and χ_m^α is the $1-\alpha$ quantile of G_m . As discussed in Sun (2013, 2014b) and in the proof of Theorem 1 in the Appendix, the term in $(bT)^{-q}$ in (31) arises from the bias of the LRV estimator.

Following Sun (2013, 2014b), we consider a sequence of local alternatives $H_{1T} : \beta = \beta_0 + T^{-1/2} \Sigma_{XX}^{-1} \Omega^{1/2} \tilde{\delta}$ (against a null of $H_0 : \beta = \beta_0$), where $\tilde{\delta}$ is uniformly distributed on the real m -dimensional sphere with squared radius $\delta^2 \equiv \|\tilde{\delta}\|^2$. We will typically index a given alternative by the length $\delta = \sqrt{\delta^2}$. The rejection rate under such an alternative using the fixed- b critical value then has the expansion,

$$\begin{aligned} \Pr_\delta \left[F_T^* > c_m^\alpha(b) \right] &= [1 - G_{m,\delta^2}(\chi_m^\alpha)] + G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \omega^{(q)} k^{(q)}(0) (bT)^{-q} \\ &\quad - \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \nu^{-1} + o(b) + o\left((bT)^{-q}\right), \end{aligned} \quad (32)$$

where G_{m,δ^2} is the noncentral chi-squared cdf with m degrees of freedom and noncentrality parameter δ^2 , G'_{m,δ^2} is its first derivative, and $\nu = (b\nu)^{-1}$ (see (26)).¹⁸ This expression depends on both the bias of the LRV estimator, as reflected in the second term on the right side of (32), and on its variance, as reflected in the third term (the term in ν^{-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. location model, relative to Gaussian inference with a known variance.

¹⁸ As noted below, these expressions hold in the context of the Gaussian location model. In a slightly more general Gaussian GMM setting, equation (32) would include a term in $O(\log T / \sqrt{T})$, but as in Sun (2014b) this term would not depend on b and can therefore be ignored for our purposes.

4. Main Results

This section provides our theoretical results describing the size and size-adjusted power. The class of tests considered is comprised of tests using psd kernel LRV estimators and tests using weighted orthonormal series LRV estimators. Unless stated otherwise, all HAR tests are evaluated using fixed- b critical values.

4.1 Assumptions

We make the following assumptions throughout, which are similar to the assumptions in Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2013, 2014b).

Throughout, we refer to the $O((bT)^{-q})$ term in (31) as the higher-order size distortion, and to this term plus α as the higher-order size.

Assumption 1 (stochastic processes).

(a) z_t is a stationary Gaussian process generated according to the multivariate location model (6), with spectral density matrix $s_z(\omega)$ that is positive definite in a neighborhood around $\omega = 0$.

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

Assumption 2 (kernels). The kernel $k(x) : \mathbb{R} \rightarrow [-1, 1]$ used for a kernel LRV estimator is

piecewise smooth, satisfies $k(x) = k(-x)$, $k(0) = 1$, $\int_{-\infty}^{\infty} |x| k(x) dx < \infty$, and has Parzen

characteristic exponent $q = 1$ or 2 .

Assumption 3 (orthonormal series). For $j = 1, \dots, B$, with $b = B^{-1}$, the orthonormal series

$\phi_j \in L^2[0, 1]$ satisfy $\int_0^1 \phi_j(s) ds = 0$ and have three continuous derivatives, such that the

n^{th} derivative $\phi_j^{(n)}(x)$ satisfies $\sup_B \sup_{j=1, \dots, B} \sup_{x \in [0, 1]} |\phi_j^{(n)}(x) / j^n| \leq C_d$ for $n = 1, 2, 3$ and

some constant C_d . The WOS weights $w_j \geq 0$ satisfy $\sum_{j=1}^B w_j = 1$ and

$\sup_B \sup_{j=1, \dots, B} w_j B \leq C_w$ for some constant C_w .

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

Assumptions 1(a)-(b) assume the multivariate Gaussian location model and provide conditions under which the bias expressions and fixed- b distributions hold. Assumption 1(b) further implies that $\omega^{(q)}$ in (30) is finite for $q \leq 2$. Assumption 2 states standard conditions on kernel estimators.

Assumption 3 strengthens slightly the conditions in Sun's (2013) Assumption 3.1 so that the basis functions have three derivatives, each of the order j^n . All basis functions discussed above aside from the SS basis functions meet this assumption, so the SS basis functions are treated separately. It further assumes regularity conditions for the WOS weights, which hold immediately for equal-weighted OS estimators.

Assumption 4 is stronger than needed for some of the results. For example, the expansions for kernel HAR tests in Sun, Phillips, and Jin (2008) require only that $b \rightarrow 0$ and $bT \rightarrow \infty$ (i.e., $b + (bT)^{-1} \rightarrow 0$), which are implied by Assumption 4. The more restrictive rate condition in Assumption 4 is used to express the expansions for weighted orthonormal series estimators in terms of the implied mean kernel when $q = 2$.

4.2 Results

Theorems 1-5 and Corollary 1 provide our main theoretical results.

Theorem 1. Under Assumptions 1, 3, and 4,

- (i) For WOS estimators, the scaled asymptotic bias of the LRV estimate is,

$$\left(E\hat{\Omega}^{WOS} - \Omega\right)\Omega^{-1} = -\left(\frac{B}{T}\right)^q k^{WOS(q)}(0) \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j \Omega^{-1} + o\left(\left(\frac{B}{T}\right)^q\right), \quad (33)$$

with $k^{WOS(1)}(0) = \lim_{B \rightarrow \infty} \frac{1}{B} \sum_{j=1}^B w_j [\phi_j(0)^2 + \phi_j(1)^2] / 2$ and $k^{WOS(2)}(0) = -\lim_{B \rightarrow \infty} \frac{1}{B^2} \sum_{j=1}^B w_j \int_0^1 \phi_j(s) \phi_j''(s) ds / 2$. If $k^{WOS(1)}(0) \neq 0$, then $q = 1$; otherwise, $q = 2$.

- (ii) For both psd kernel and weighted orthonormal series HAR tests, the small- b asymptotic expansions (31) and (32) apply. These expansions also hold for the SS series estimator, for which $q = 1$, although it does not satisfy Assumption 3.

Henceforth, let k denote either a kernel or an implied mean kernel.

Theorem 2. Let $c_{m,T}^\alpha(b)$ be the size-adjusted fixed- b critical value, that is, the critical value such that $\Pr_0[F_T^* > c_{m,T}^\alpha(b)] = \alpha + o(b) + o((bT)^{-q})$, and assume that Assumptions 1-4 hold. Then

$$c_{m,T}^\alpha(b) = [1 + \omega^{(q)} k^{(q)}(0)(bT)^{-q}] c_m^\alpha(b), \quad (34)$$

and the higher order size-adjusted power of the test is,

$$\Pr_\delta[F_T^* > c_{m,T}^\alpha(b)] = [1 - G_{m,\delta^2}(\chi_m^\alpha)] - \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha \nu^{-1} + o(b) + o((bT)^{-q}). \quad (35)$$

Theorem 3. Consider two HAR test statistics F_1^* and F_2^* based on different kernels or implied mean kernels with the same value of q , which have equivalent degrees of freedom respectively given by ν_1 and ν_2 , and which have fixed- b critical values respectively given by $c_{1,m}^\alpha(b_1)$ and $c_{2,m}^\alpha(b_2)$. Choose b_1 and b_2 such that F_1^* and F_2^* have the same higher-order size. Then, under Assumptions 1-4, the difference between their higher-order rejection rates under the local alternative indexed by δ is,

$$\begin{aligned} \Pr_\delta[F_{1T}^* > c_{1,m}^\alpha(b_1)] - \Pr_\delta[F_{2T}^* > c_{2,m}^\alpha(b_2)] &= \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha) \chi_m^\alpha (\nu_2^{-1} - \nu_1^{-1}) \\ &+ o(b_1) + o((b_1 T)^{-q}) + o(b_2) + o((b_2 T)^{-q}). \end{aligned} \quad (36)$$

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

$$\Delta_S = \Pr_0 \left[F_T^* > c_m^\alpha(b) \right] - \alpha. \quad (37)$$

The power of the oracle test, in which Ω 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 δ , and let Δ_P^{\max} denote the maximum such power loss, so that Δ_P^{\max} is the maximum gap between the power curves of the oracle test and the candidate test. Then,

$$\Delta_P(\delta) = \left[1 - G_{m,\delta^2}(\chi_m^\alpha) \right] - \Pr_\delta \left[F_T^* > c_{m,T}^\alpha(b) \right], \text{ and} \quad (38)$$

$$\Delta_P^{\max} = \sup_\delta \Delta_P(\delta). \quad (39)$$

Because $\nu = (b\psi)^{-1}$, equations (31) and (35) constitute a pair of parametric equations that determine Δ_S and Δ_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. The way that b enters those two expressions restricts the rate of the sequence b such that Δ_S and Δ_P are of the same order. Corollary 1 provides that restriction, which meets Assumption 4, and which is then used in Theorem 4 to provide the higher-order tradeoff between size and power.

Corollary 1. $\Delta_P(\delta)$ and Δ_S are of the same asymptotic order if and only if $b \sim CT^{-\frac{q}{q+1}}$ for some positive constant C .

Theorem 4. For a given HAR test evaluated using fixed- b critical values, if $b^{\frac{q+1}{q}} T \rightarrow C$ as in Corollary 1, then:

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

$$T\Delta_P(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta)\ell^{(q)}(k)|\omega^{(q)}|^{1/q} + o(1), \quad (40)$$

where $a_{m,\alpha,q}(\delta) = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\mathcal{X}_m^\alpha)\mathcal{X}_m^\alpha (G'_m(\mathcal{X}_m^\alpha)\mathcal{X}_m^\alpha)^{1/q}$, $\ell^{(q)}(k) = (k^{(q)}(0))^{1/q} \psi$, and ψ is defined in (24).

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

$$T\Delta_P^{\max}|\Delta_S|^{1/q} = \bar{a}_{m,\alpha,q}\ell^{(q)}(k)|\omega^{(q)}|^{1/q} + o(1), \quad (41)$$

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

- (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 provides the size/power frontier, which is the envelope of the tradeoffs given in Theorem 4.

Theorem 5.

- (i) For psd kernel and WOS HAR tests, under the sequence for b in Corollary 1,

$$T\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{3\pi\sqrt{10}}{25} \bar{a}_{m,\alpha,2} + o(1) \quad (42)$$

where $\bar{a}_{m,\alpha,2}$ is given 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 $m = 1$, $\bar{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.6460$ 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, under the sequence for b in Corollary 1,

$$T\Delta_P^{\max} \sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{\pi}{\sqrt{6}} \bar{a}_{m,\alpha,2} + o(1) \text{ (exact } t \text{ or } F \text{ critical values)}. \quad (43)$$

This frontier is achieved by the EWP test. For $\alpha = .05$, $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 0.3623$ 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 $m = 3$.

4.3 Remarks

1. For a given α and m , the 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 is universal and applies to all psd kernel and weighted orthonormal series HAR tests evaluated using fixed- b critical values under the asymptotic sequence given in Corollary 1. The frontier furthermore applies to all processes satisfying Assumption 1.
2. The sequence given in Corollary 1, $b \sim CT^{-q/(q+1)}$, is of the same order as the sequence found in 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. Inspection of the proof reveals that strengthening terms in $o(b)$ and $o((bT)^{-q})$ in the underlying Edgeworth expansions, to O of any higher order, would broaden the range of sequences for which this frontier holds. This conjecture is supported by the generally good ability of the frontier to describe simulation results as discussed below.
3. Because the sign of the size distortion and the sign of $\omega^{(2)}$ are the same, the absolute values in (41) are eliminated in Theorem 5 by expressing the tradeoff for $q = 2$ tests in terms of the ratio $\Delta_S/\omega^{(2)}$, which in the case $m = 1$ yields (1).
4. To gain some intuition into the overall frontier (42), note that the frontier for kernel tests is achieved by the $q = 2$ HAR test for which the expression $\ell^{(2)}(k) = \sqrt{k^{(2)}(0)} \int_{-\infty}^{\infty} k^2(x) dx$ in (41) is minimized. 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, which Priestley summarizes as “bias and variance are antagonistic.” In our application, the bias produces a size distortion while the variance degrades size-adjusted power. Our results indicate that this uncertainty principle extends beyond the squared bias-variance tradeoff of spectral density estimation. In addition to appearing in the size-power frontier in Theorem 4, the kernel appears in the following objective functions only through $\ell^{(2)}(k)$ when they are evaluated using the (kernel-specific) optimal value of b for $q = 2$ kernels and implied mean kernels:

- (a) $MSE(\hat{s}_z(0)) = \text{bias}^2(\hat{s}_z(0)) + \text{var}(\hat{s}_z(0))$;
- (b) An objective function 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 $0 \leq a \leq 1$ and where Π_δ is a weight function over the noncentrality parameter δ ;
- (c) A quadratic version of the previous objective function, $a(\Delta_s)^2 + (1-a)(\Delta_p^{\max})^2$;
- (d) An objective function of the form considered by Sun, Phillips, and Jin (2008) that minimizes the weighted average of the Type I and Type II error; and
- (e) An objective function of the form $a|\text{bias}(\hat{s}_z(0))| + (1-a)\text{var}(\hat{s}_z(0))$.

Minimizing (a) is the classical problem of optimal spectral estimation. A referee pointed out that (b), which trades off size distortion and power loss linearly, also depends on the kernel solely through $\ell^{(2)}(k)$ at the optimal truncation parameter. Minimizing (c) does the same with quadratic loss and is the approach used by Lazarus, Lewis, Stock, and Watson (2018a), which uses the results in Theorem 4 to obtain a rule-of-thumb rate for b . Objective function (d) differs from (b) because the Type II error is not size-adjusted, yet its optimized value depends also on the kernel only through $\ell^{(2)}(k)$.¹⁹ Objective function (e) is not of primitive interest but (b) and (d) are fundamentally of the form (e). Each of these objective functions is minimized by the QS kernel. Theorem 5 also implies that the EWP estimator minimizes the objective function among equal-weighted orthonormal series estimators.

¹⁹ We thank Yixiao Sun for pointing this out to us; see Lazarus, Lewis, Stock, and Watson (2018b) for this calculation.

5. The proof for the restricted frontier in (43) entails considering the $q = 2$ equal-weighted orthonormal series, which generate t or F critical values; then expressing a candidate set of equal-weighted orthonormal series in terms of Fourier coefficients and computing their implied value of $\ell^{(2)}(k) = \sqrt{k^{(2)}(0)}$ (since in this equal-weighted case $\psi = B \sum_{j=1}^B w_j^2 = 1$); then concluding that this value must be larger than the equal-weighted orthonormal series that places all weight on the first $B/2$ Fourier terms.²⁰ But that dominating series delivers the Daniell kernel, that is, the EWP estimator.
6. The price one must pay for the convenience of exact t or F fixed- b critical values can be computed from Theorem 3 by letting F_1 be the QS test and F_2 be EWP. Suppose the EWP test is computed using $B/2$ periodogram ordinates (B Fourier basis functions). Then, from (36), the power cost of using EWP relative to the higher-order best test (QS) with the same higher-order size is, neglecting the remainder terms,

$$\begin{aligned} \Pr_{\delta}[F_{QS,T}^* > c_{QS,\alpha}(b_{QS})] - \Pr_{\delta}[F_{EWP,T}^* > c_{EWP,\alpha}(b_{EWP})] &\approx \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha (v_{EWP}^{-1} - v_{QS}^{-1}) \\ &= \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha \left(1 - \frac{6\sqrt{3}}{5\sqrt{5}}\right) B^{-1}, \end{aligned} \quad (44)$$

where $v_{EWP} = B$ and the final expression is derived in the Appendix. (Equation (44) holds under the more general rate condition of Assumption 4, not just under the optimal rate condition of Corollary 1.)

The maximum higher-order power loss from using EWP over all alternatives δ (that is, (44) maximized over δ) is tabulated in Table 1 for various values of B and $m = 1, 2, 3$, and 4. It is apparent that the cost of using EWP relative to QS is small: for $B = 8$ and $m = 1$, the maximum size-equivalent power gap is 0.0074 over all alternatives. This aligns with the numerical finding in Kiefer and Vogelsang (2005) that the local asymptotic power curves for

²⁰ See the working paper version of this paper (Lazarus, Lewis, and Stock (2017)) for the proof in this restricted case. The proof of Theorem 5(i) given here in fact provides a more general result: given *any* set of weights $\{w_j\}$ (and not just equal weights), the choice of Fourier basis functions delivers the size-power frontier. Part (ii) then applies this result to the equal-weighted case delivering t - or F -based inference.

these two tests are very close. And while the maximum power loss increases in m , for $B = 8$ it remains small, approximately 0.02, even when testing $m = 4$ restrictions. Figure S.3 in the Supplement plots the final expression in (44) as a function of δ for various values of B and $m = 1$. Given that there is little apparent loss from considering equal weights for the WOS tests – the best WOS test cannot outperform the QS kernel by Theorem 5(i), and EWP exhibits only minimal loss relative to QS – the remainder of our discussion and our Monte Carlo simulations focus on this equal-weighted case whenever we consider series estimators.

7. The expressions for the generalized derivatives of the implied mean kernel at the origin in Theorem 1(i) are displayed as a sequential calculation (letting $T \rightarrow \infty$ to obtain (23), then differentiating). However, the theorem is proven under the small- b sequence in Assumption 4, then it is shown that this result coincides with the sequential heuristic. We also obtain the generalized derivative of the implied mean kernel (and thus the bias) of the SS series estimator without appealing to the derivative expressions in Theorem 1(i).
8. Theorem 1 allows us to assess the effect of a given choice of orthonormal series (and weights w_j) on the bias of the LRV estimator by calculating the generalized derivatives given in those expressions. We briefly consider some specific choices of equal-weighted orthonormal series.

- (a) **$q = 1$ implied mean kernels.** It is shown in the Appendix that the SS-basis has $q = 1$ and $k^{SS(1)}(0) = 1$. Thus for small b , $k^{SS(1)}(0) = k^{NW(1)}(0)$. Because SS and Bartlett are $q = 1$ kernels or implied mean kernels, their size distortion/size-adjusted power tradeoffs are given by (40) with $q = 1$. Assessing the second term in those tradeoffs, for SS, $\ell^{(1)}(k^{SS}) = k^{SS(1)}(0)\psi = 1$ (since $\psi = 1$ for equal-weighted OS), while for Bartlett, $\ell^{(1)}(k^{NW}) = k^{NW(1)}(0) \int_{-\infty}^{\infty} k^2(x)dx = 2/3$. Thus NW dominates SS to higher order: the NW small- b tradeoff curve is strictly below the SS tradeoff curve.²¹

²¹ Another set of orthonormal basis functions are the Legendre polynomials, which are the B functions of the Gram-Schmidt orthonormalization of $\{s^j\}, j = 1, \dots, B$ on $[0,1]$. It is shown in the Appendix that these basis functions also have $q = 1$, and, surprisingly, the generalized first derivative at the origin is the same for small b (large B) as that of the SS-basis: $k^{Leg(1)}(0) = k^{SS(1)}(0) = 1$, so that the SS and Legendre basis functions have a common size-power tradeoff curve.

(b) $q = 2$ *implied mean kernels*. The Fourier and Type II cosine bases both satisfy Assumption 2, and by calculations in the Appendix, are both $q = 2$ and are

asymptotically equivalent: $k^{EWP(2)}(0) = k^{cos(2)}(0) = \frac{\pi^2}{6}$. The higher-order term in

the bias expansion is smaller for the equal-weighted cosine than for the Fourier basis, as it can be seen in the Appendix that for finite B ,

$k_B^{EWP(2)}(0) = \pi^2(B+1)(B+2)/6B^2$ while $k_B^{cos(2)}(0) = \pi^2(B+1/2)(B+1)/6B^2$. This

suggests that for small B , the cosine basis might slightly outperform the Fourier basis. The Fourier basis further requires that B be even, while the cosine basis is not subject to this restriction. These considerations lead Lazarus, Lewis, Stock, and Watson (2018a) to suggest using equal-weighted cosine basis functions if a researcher would like to use t or F critical values for testing.

9. Theorem 1(i) has several precedents in the literature. In a slightly different context, Brillinger (1975, Theorem 5.8.1) provides a result similar to the first part of our Theorem 1(i), but does not provide expressions for the generalized derivatives of the implied mean kernel. Theorem 1(i) extends the results of Theorem 1(i) in Phillips (2005), Theorem 2(a) in Sun (2011), and Theorem 4.1 in Sun (2013) to the case of a general and/or weighted orthonormal series estimator (some of those earlier results apply only to $q = 2$ or to specific series). Our result unifies the asymptotic bias for kernel and WOS LRV estimators by expressing the asymptotic bias in both cases in terms of the Parzen characteristic exponent and the generalized derivatives of the kernel or implied mean kernel.
10. The tradeoffs in Theorems 4 and 5 are expressed in terms of the absolute size distortions. For processes with $s_z''(0) < 0$ (loosely, series with positive serial correlation), the HAR tests are oversized and the tradeoff is between size and power. This positive serial correlation case is frequently found in practice, e.g. it is typical 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

results where the size-power tradeoff is used to obtain a truncation parameter rule in the positive serial correlation case. For additional discussion, see Lazarus, Lewis, Stock, and Watson (2018a).

11. 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 $|\omega^{(q)}| \leq \bar{\omega}^{(q)}$.²²

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 (31) and (37) 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 (34), evaluated using this least favorable value, therefore results in a test that controls size uniformly to higher order.²³ Then, in maximizing weighted average power (WAP), we make the simplifying assumption that z_t follows an AR(1) with coefficient ρ , and consider the problem of choosing b to solve

$$b^{WAP} = \arg \min_b \int \int_{\delta \mid \rho \leq \bar{\rho}} \Delta_p(\omega^{(q)}(\rho), \delta) d\Pi_\rho(\rho) d\Pi_\delta(\delta), \quad (45)$$

where $\Delta_p(\omega^{(q)}(\rho), \delta) = \frac{1}{2} \delta^2 G'_{m+2, \delta^2}(\chi_m^\alpha) \chi_m^\alpha v^{-1} + G'_{m, \delta^2}(\chi_m^\alpha) \chi_m^\alpha [\bar{\omega}^{(q)} - \omega^{(q)}(\rho)] k^{(q)}(0) (bT)^{-q}$,
 $\omega^{(1)}(\rho) = 2\rho / (1 - \rho^2)$, $\omega^{(2)}(\rho) = 2\rho / (1 - \rho)^2$, $\bar{\rho} = \max \rho$ s.t. $\omega^{(q)}(\rho) \leq \bar{\omega}^{(q)}$,

²² This uniform-size-control approach follows a small literature developed by Müller (2007, 2014), Preinerstorfer and Pötscher (2016), and Pötscher and Preinerstorfer (2017, 2019). 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.

²³ This statement requires the regularity condition that the remainder terms in (31) together are $o(b) + o((bT)^{-q})$ uniformly in $|\omega^{(q)}| \leq \bar{\omega}^{(q)}$, which we assume holds. We thank Yixiao Sun for alerting us to this subtlety; see the Appendix for details.

and the weight functions Π_ρ and Π_δ are independent and each integrate to one. The solution to (45) is,

$$b^{WAP} = q^{\frac{1}{1+q}} \tilde{d}_{m,\alpha,q} \left(\frac{k^{(q)}(0)}{\psi} \right)^{\frac{1}{1+q}} \left(\tilde{\omega}^{(q)} \right)^{\frac{1}{1+q}} T^{\frac{-q}{1+q}}, \quad (46)$$

where expressions for the constants $\tilde{\omega}^{(q)}$ and $\tilde{d}_{m,\alpha,q}$ are provided in the Appendix. We can see immediately that b^{WAP} declines with T at the same rate as given in Corollary 1. Further (again see the Appendix), the power loss of the test using the WAP-maximizing sequence (46) depends on k only through $\ell^{(q)}(k)$. Thus 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. Also, $q = 1$ kernels are again asymptotically dominated by $q = 2$ kernels. Thus, many of the qualitative findings from the baseline analysis carry through to this uniform-size-control, maximum-WAP analysis.

12. The multivariate results focus on inference on all m elements of β . 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 $T(R\bar{z}_0)' (R\hat{\Omega}R')^{-1} (R\bar{z}_0) / m'$ (cf. (5)), where $R\bar{z}_0 = T^{-1} \sum_{t=1}^T (Rz_t - \tilde{\beta}_0)$. Because all the estimators of Ω we consider are quadratic forms in \hat{z} , this F statistic testing $R\beta = \tilde{\beta}_0$ is numerically equivalent to the F statistic (5) testing a full vector hypothesis, computed using the $m' \times 1$ vector of transformed data Ry_t . Thus, the results for full vector inference apply directly to subvector inference.

5. Monte Carlo Analysis

The purpose of this Monte Carlo analysis is twofold. First, we assess the quality of the small- b approximations to the size/power tradeoffs in the Gaussian location model. Second, we investigate the extent to which the theory derived for the Gaussian multivariate location model generalizes to time series regression with stochastic regressors.

5.1 Estimators and Design

For a given kernel or orthonormal series estimator, we use four values of b , chosen so that $\nu = 8, 16, 32$, and 64 . The tests are labeled accordingly, for example NW16 is the Newey-West (Bartlett) test with $\nu = 16$ equivalent degrees of freedom. As a reference, for $T = 200$, NW32 has a truncation parameter of $(3/2)T/\nu$, which rounds up to 10. For the orthonormal series estimators, we consider tests with equal weights $w_j = 1/B$, so that $\nu = B$. Tests use fixed- b critical values unless explicitly stated otherwise.

We specifically examine the following HAR tests:

1. NW: Kernel estimator with Bartlett/Newey-West kernel, $k(x) = (1-|x|)\mathbf{1}(|x|\leq 1)$
2. KVB: The Kiefer-Vogelsang-Bunzel (2000) test, which is NW with $S = T$ (so $\nu = 3/2$).
3. QS implemented in the covariance domain: $k(x) = 3[\sin(\pi u)/(\pi u) - \cos(\pi u)] / (\pi u)^2$ for $u = 6x/5$.
4. EWP: Equal-weighted orthonormal series estimator using the Fourier basis (12).
5. cos: Equal-weighted orthonormal series estimator using the Type-2 cosine basis (17).
6. SS-basis: Equal-weighted split-sample orthonormal series estimator using (19).

In the location model, the data are generated according to (6), where $u_{it}, i = 1, \dots, m$ are independent and follow either a Gaussian AR(1) or an ARMA(2,1), with all m disturbances having the same parameter values. For the regression model, the data are generated according to $y_t = x_t'\beta + u_t$, with $x_{it}, i = 1, \dots, m$ and u_t being independent Gaussian AR(1) processes. Under the null, $\beta = 0$. Under the local alternative, $\beta = T^{-1/2}\Sigma_{XX}^{-1}\Omega^{1/2}\delta$, where δ is the local alternative index value (in the location model, $\Sigma_{XX} = I$).

5.2 Monte Carlo Results

This section presents a small number of representative Monte Carlo results; additional results are contained in the Supplement and in Lazarus, Lewis, Stock and Watson (2018a). All results are displayed in finite-sample counterparts of Figure 1. For these figures, the axes are not scaled, so that the units are the size distortion and the power loss. The theoretical tradeoffs from (41) are shown as lines, and the Monte Carlo results are presented as scatter points. We conduct 100,000 replications for each simulation design considered here.

Location model. Figure 2 presents results for QS, EWP, and NW tests in the location model with Gaussian AR(1) disturbances in the $m = 1$ case with AR parameter $\rho = 0.5$ and $T =$

200. The Monte Carlo results for QS and EWP are close to their theoretical curves. The small- b approximation is less good for Newey-West: the NW Monte Carlo scatter appears to follow a curve that has the same shape as the theoretical curve, but is shifted out. KVB is a limiting case of Newey-West with $S_T = T$ (so $b = 1$ and $\nu = 1.5$), that is, KVB is NW1.5, so KVB lies on the NW Monte Carlo curve.

Figure 3 presents results for $m = 2$ with AR(1) errors, $\rho = 0.5$, and $T = 200$. Compared to the results for $m = 1$, the $m = 2$ frontier fits the simulations slightly better for QS and EWP, but somewhat worse for NW.

The supplement provides additional results for the location model for other AR(1) parameters, other sample sizes, ARMA(2,1) disturbances, and other kernels and orthonormal series. Those results indicate that the fit (distance from the scatter points to their theoretical tradeoff) improves with T , deteriorates as $\omega^{(2)}$ increases, is better for $q = 2$ kernels than $q = 1$, and does not appreciably deteriorate as process parameters are changed holding $\omega^{(2)}$ constant. The first two results are unsurprising. Our interpretation of the third finding is that the order of approximation of the expansions is $o((bT)^{-q})$, so the remainder is of a smaller order for $q = 2$ than for $q = 1$ kernels. The larger values of b used with the NW kernel for a given ν may also play a role. Overall, the simulation results accord with the theory.

Stochastic regressor. Figure 4 shows the QS, EWP, and NW tests on a the coefficient on a single stochastic regressor, where both the regressor and dependent variable have AR(1) disturbances with $\rho = 0.5$ and $T = 200$ (intercept included in the regression but not tested). In this DGP, z_t is AR(1) but non-Gaussian. For reference, the theoretical tradeoff curves are shown for the Gaussian location model. It appears that this departure from Gaussianity results in poor approximations of the Gaussian small- b asymptotic approximation and that there are missing terms in the expansion as suggested by the calculations in Velasco and Robinson (2001). This said, several key qualitative results in the theory continue to apply to the single stochastic regressor. First, for a given estimator, the Monte Carlo results map out a size-power tradeoff that has a shape similar to the Gaussian theoretical shape, just shifted out. Second, the tradeoff for the QS and EWP estimators are very close to each other. Third, the ranking across estimators is the same as suggested by the theory and confirmed in the Monte Carlo analysis of the location model, that is, the $q = 1$ tests are outperformed by the $q = 2$ tests. These findings reflect results

for other designs, kernels, and $m = 2$ in the Supplement. Further, additional simulations show that the approximation improves for higher values of T .

Overall, we can draw three conclusions. First, the theoretical frontiers provide a good description of estimator 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, at least for this design. Third, the qualitative results for stochastic regressors are consistent with the theory for the location model, however the Monte Carlo points no longer lie on the tradeoff derived for the Gaussian location model. We attribute this divergence of the theory and Monte Carlo results to the non-Gaussianity of z_t in the stochastic regressor case.²⁴

6. Conclusions

By combining new theoretical results with previous results from the associated literature, we characterize optimal HAR tests that are implemented using fixed- b critical values. Tests using the QS kernel achieve the size distortion/power loss frontier for all psd kernel tests and weighted orthonormal series tests, but they require nonstandard critical values. Restricting attention to tests admitting exact fixed- b t and F distributions entails a very small sacrifice in the size/power frontier. Among tests using t or F fixed- b critical values, the test using the equal-weighted periodogram estimator achieves the size/power frontier. Our Monte Carlo experiments confirm that the theory works well in the Gaussian location model, confirm the theoretical rankings of the tests, confirm the finding that tests with large truncation parameters and fixed- b critical values provide meaningful size improvements over tests with small truncation parameters while sacrificing little power, and suggest that these qualitative results extend outside the Gaussian location model to tests involving stochastic regressors.

²⁴ In unreported results, we also examined the performance of tests based on plug-in higher-order corrected critical values given by (34), using an estimated value of $\omega^{(q)}$. HAR tests using these plug-in critical values generally worked poorly compared to tests using standard fixed- b critical values. Results are given in the working paper version of this paper (Lazarus, Lewis, and Stock (2017)).

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.
- 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 Practitioners Guide to Robust Covariance Matrix Estimation,” *Handbook of Statistics* 15, ch. 12, 291-341.
- 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.
- Hillmer, C. E. and M. J. Hilmer (2014). *Practical Econometrics: Data Collection, Analysis, and Application, 1st Edition*. New York: McGraw-Hill.
- Hwang, J. and Y. Sun (2018). “Simple, Robust, and Accurate F and t Tests in Cointegrated Systems,” *Econometric Theory* 34, 949-984.
- 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.
- Kandelaki, N.P. and Sozanov, V.V. (1964). “A Central Limit Theorem for Random Elements With Values in Hilbert Space,” *Theory of Probability and its Applications*, 9, 38-46.
- 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.
- Lazarus, E., D.J. Lewis, and J.H. Stock (2017). “The Size-Power Tradeoff in HAR Inference,” working paper, Harvard University.
- Lazarus, E., D.J. Lewis, J.H. Stock, and M.W. Watson (2018a). “HAR Inference: Recommendations for Practice,” *Journal of Business and Economic Statistics* 36, 541-559.
- Lazarus, E., D.J. Lewis, J.H. Stock, and M.W. Watson (2018b). “HAR Inference: Recommendations for Practice: Rejoinder,” *Journal of Business and Economic Statistics* 36, 574-575.
- Long, J.S. and L.H. Ervin (2000). “Using Heteroskedasticity Consistent Standard Errors in the Linear Regression Model,” *The American Statistician* 54(3), 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.
- 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 (1996). "The Exact Error in Estimating the Spectral Density at the Origin," *Journal of Time Series Analysis* 17, 379-408.
- 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 für Wahrscheinlichkeitstheorie*, 45, 49-72.
- Phillips, P. C. B. (2005). "HAC Estimation by Automated Regression," *Econometric Theory* 21, 116-142.
- Phillips, P. C. B. and H. R. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67, 1057-1111.
- Phillips, P.C.B., Y. Sun, and S. Jin (2006). "Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation," *International Economic Review* 47, 837-894.
- Politis, D. (2011). "Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices," *Econometric Theory* 27, 703-744.
- Pötscher, B. and D. Preinerstorfer (2017). "Controlling the Size of Autocorrelation Robust Tests," *Journal of Econometrics* 207, 406-431.
- Pötscher, B. and D. Preinerstorfer (2019). "Further Results on Size and Power of Heteroskedasticity and Autocorrelation Robust Tests, with an Application to Trend Testing," manuscript, Dept. of Statistics, University of Vienna.
- 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*. Elsevier, 881-935.
- 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.
- Stock, J.H. and M.W. Watson (2015), “Factor Models and Structural Vector Autoregressions in Macroeconomics,” in preparation for *Handbook of Macroeconomics*, vol. 2.
- 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., P.C.B. Phillips, and S. Jin (2008), “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” *Econometrica*, 76(1): 175-194.
- 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.
- Thomson, D.J. (1982). “Spectrum Estimation and Harmonic Analysis,” *Proceedings of the IEEE* 70, 1055-1096.
- Velasco, C. and P.M. Robinson (2001). “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory* 17, 497-539.
- Zhang, X. and X. Shao (2013). “Fixed-Smoothing Asymptotics for Time Series,” *Annals of Statistics* 41, 1329-1349.

Appendix

Derivation of Equation (19): First write t^{SS} in (18) in the standard form (4) for HAR tests by (i) noting that $\bar{\beta} - \beta_0 = \bar{z}_0$ and (ii) defining $\hat{\Omega}^{SS} = [T / (B + 1)] S_{\bar{\beta}}^2$. It remains to show that this definition of $\hat{\Omega}^{SS}$ yields the expression for $\hat{\Omega}^{SS}$ given in (19).

Let $\tilde{\beta}$ be the $B+1$ vector with i^{th} element $\tilde{\beta}_i = \hat{\beta}^{(i)}$, so that $S_{\tilde{\beta}}^2 = B^{-1} \tilde{\beta}' \left(I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}' \right) \tilde{\beta}$, where I_{B+1} is the $(B+1) \times (B+1)$ identity matrix and ι_{B+1} is the $(B+1)$ -vector of 1's. Thus,

$$\begin{aligned} \hat{\Omega}^{SS} &= [T / (B + 1)] S_{\tilde{\beta}}^2 = [T / (B + 1)] B^{-1} \tilde{\beta}' \left(I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}' \right) \tilde{\beta} \\ &= [T / (B + 1)] B^{-1} [T / (B + 1)]^{-2} y' \left(I_{B+1} \otimes \iota_{T/(B+1)} \right) \left(I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}' \right) \left(I_{B+1} \otimes \iota_{T/(B+1)}' \right) y \\ &= (BT)^{-1} (B + 1) \hat{z}' \left(I_{B+1} \otimes \iota_{T/(B+1)} \right) \left(I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}' \right) \left(I_{B+1} \otimes \iota_{T/(B+1)}' \right) \hat{z} \\ &= (BT)^{-1} (B + 1) \hat{z}' \left(I_{B+1} \otimes \iota_{T/(B+1)} \right) M_i^B M_i^{B'} \left(I_{B+1} \otimes \iota_{T/(B+1)}' \right) \hat{z} \\ &= \hat{z}' \Phi^{SS} \Phi^{SS'} \hat{z} / BT, \end{aligned}$$

where the first equality uses the definition of $\hat{\Omega}^{SS}$, the second substitutes in the expression for $S_{\tilde{\beta}}^2$, the third uses $\tilde{\beta} = [T / (B + 1)]^{-1} \left[I_{B+1} \otimes \iota_{T/(B+1)}' \right] y$, the fourth exploits the definition that $\hat{z} = y - \bar{\beta}$ and the properties of $I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}'$, the fifth uses the idempotence of $I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}'$ and the definition that M_i^B is the $(B+1) \times B$ matrix of eigenvectors of $I_{B+1} - \iota_{B+1} (\iota_{B+1}' \iota_{B+1})^{-1} \iota_{B+1}'$ associated with its B unit eigenvalues, and the final equality uses the definition of Φ^{SS} following (19). Note that Φ^{SS} is $T \times B$, that $\iota_T' \Phi^{SS} = 0$, and $\Phi^{SS'} \Phi^{SS} / T = I_B$ as required for series estimators. Thus $\hat{\Omega}^{SS}$ is an OS estimator as given following (15) with basis matrix Φ^{SS} .

Proof of Theorem 1: (i) This part of the theorem extends Theorem 1(i) of Phillips (2005) and Theorem 2(a) of Sun (2011), among others, to the case of a general weighted orthonormal series estimator, and additionally derives the expressions for $k^{WOS(1)}(0)$ and $k^{WOS(2)}(0)$ provided in the

theorem. We note that rather than taking sequential limits to prove this result as done in the heuristic derivation of the implied mean kernel in Section 2, we obtain results in which the relevant limits have been taken jointly according to the sequence in Assumption 4.

Using the same steps as for equation (20), we can write,

$$\begin{aligned}
E\hat{\Omega}_j^{OS} &= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^T\phi_j(t/T)\hat{z}_t\right)' \\
&= \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\phi_j(t/T)\phi_j(s/T)\Gamma_{s-t} + O(1/T) \\
&= \sum_{u=-(T-1)}^{T-1}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\Gamma_u + O(1/T), \tag{47}
\end{aligned}$$

where the $O(1/T)$ term in the second line arises due to the approximation of \hat{z}_t with z_t under Assumption 1 (see, for example, the proof of Theorem 2 in Sun (2011), or Footnote 15 above).

Thus,

$$E\hat{\Omega}^{WOS} - \Omega = \sum_{u=-(T-1)}^{T-1}\left\{\left[\frac{1}{B}\sum_{j=1}^B w_j \frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u - \sum_{|u|\geq T}\Gamma_u + O\left(\frac{1}{T}\right). \tag{48}$$

Let q be either 1 or 2; which value of q applies for a given set of basis functions $\{\phi_j\}$ will be shown to be given by the rule stated in the theorem. By assumptions 1(b) and 4,

$$\left|\sum_{|u|\geq T}\Gamma_u\right| \leq \sum_{|u|\geq T}|\Gamma_u| \leq \frac{1}{T^2}\sum_{|u|\geq T}|u|^2|\Gamma_u| = o(T^{-2}) = o((B/T)^q), \tag{49}$$

so we may focus on the first summation in (48).

Following the proof of Theorem 1(i) in Phillips (2005), we may then write

$$\begin{aligned}
E\hat{\Omega}^{WOS} - \Omega &= \sum_{u=-L_T}^{L_T} \left\{ \left[\frac{1}{B} \sum_{j=1}^B \frac{W_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j \left(\frac{t}{T} \right) \phi_j \left(\frac{t-u}{T} \right) \right] - 1 \right\} \Gamma_u \\
&+ \sum_{L_T < |u| < T} \left\{ \left[\frac{1}{B} \sum_{j=1}^B \frac{W_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j \left(\frac{t}{T} \right) \phi_j \left(\frac{t-u}{T} \right) \right] - 1 \right\} \Gamma_u + o \left(\left(\frac{B}{T} \right)^q \right) + O \left(\frac{1}{T} \right), \quad (50)
\end{aligned}$$

where $L_T < T$ is a positive integer sequence chosen such that

$$\frac{T}{L_T^{q+\zeta} B^q} + \frac{L_T B}{T} \rightarrow 0, \quad (51)$$

where ζ is as in Assumption 1(b). Similar to the steps taken in (49), we have that

$$\begin{aligned}
&\left| \sum_{L_T < |u| < T} \left\{ \left[\frac{1}{B} \sum_{j=1}^B \frac{W_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j \left(\frac{t}{T} \right) \phi_j \left(\frac{t-u}{T} \right) \right] - 1 \right\} \Gamma_u \right| \\
&\leq C \sum_{L_T < |u| < T} |\Gamma_u| \leq \frac{C}{L_T^{2+\zeta}} \sum_{L_T < |u| < T} |u|^{2+\zeta} |\Gamma_u| = o(L_T^{-2+\zeta}) = o \left(\left(\frac{B}{T} \right)^q \right), \quad (52)
\end{aligned}$$

for some constant C , by Assumptions 1(b) and 3, and where the fact that $o(L_T^{-2+\zeta}) = o \left(\left(\frac{B}{T} \right)^q \right)$

follows from the sequence (51).

We may accordingly focus attention on the terms in (50) for which $|u| \leq L_T$, and can thus write that equation as

$$E\hat{\Omega}^{WOS} - \Omega = \sum_{u=-L_T}^{L_T} \left\{ \left[\frac{1}{B} \sum_{j=1}^B \frac{W_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j \left(\frac{t}{T} \right) \phi_j \left(\frac{t-u}{T} \right) \right] - 1 \right\} \Gamma_u + o \left(\left(\frac{B}{T} \right)^q \right) + O \left(\frac{1}{T} \right). \quad (53)$$

We now consider the value in square brackets in the first term in this equation (or in (48)). This value is similar to, but defined slightly differently than, the implied mean kernel in (22); accordingly, define

$$\hat{k}_{B,T}^{WOS}\left(\frac{u}{T}\right) = \frac{1}{B} \sum_{j=1}^B \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j\left(\frac{t}{T}\right) \phi_j\left(\frac{t-u}{T}\right). \quad (54)$$

Using a mean-value expansion, we have, for some values $h_{t,u} \in (t, t-u)$,

$$\begin{aligned} \hat{k}_{B,T}^{WOS}\left(\frac{u}{T}\right) - 1 &= \frac{1}{B} \sum_{j=1}^B \left\{ \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j\left(\frac{t}{T}\right) \left[\phi_j\left(\frac{t}{T}\right) - \frac{u}{T} \phi_j'\left(\frac{t}{T}\right) + \frac{1}{2} \left(\frac{u}{T}\right)^2 \phi_j''\left(\frac{t}{T}\right) \right. \right. \\ &\quad \left. \left. - \frac{1}{6} \left(\frac{u}{T}\right)^3 \phi_j'''\left(\frac{t}{T}\right) \left(\frac{h_{t,u}}{T}\right) \right] \right\} - 1 \\ &= \frac{1}{B} \sum_{j=1}^B \left\{ \frac{w_j}{T} \sum_{t=1}^T \phi_j\left(\frac{t}{T}\right)^2 - 1 - \frac{w_j}{T} \sum_{1 \leq t \leq u, T+u \leq t \leq T} \phi_j\left(\frac{t}{T}\right)^2 - \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j\left(\frac{t}{T}\right) \phi_j'\left(\frac{t}{T}\right) \frac{u}{T} \right. \\ &\quad \left. + \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{2} \phi_j\left(\frac{t}{T}\right) \phi_j''\left(\frac{t}{T}\right) \left(\frac{u}{T}\right)^2 - \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{6} \phi_j\left(\frac{t}{T}\right) \phi_j'''\left(\frac{h_{t,u}}{T}\right) \left(\frac{u}{T}\right)^3 \right\} \\ &= -\frac{1}{B} \sum_{j=1}^B \left\{ \frac{w_j}{T} \sum_{1 \leq t \leq u, T+u \leq t \leq T} \phi_j\left(\frac{t}{T}\right)^2 + \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \phi_j\left(\frac{t}{T}\right) \phi_j'\left(\frac{t}{T}\right) \frac{u}{T} \right. \\ &\quad \left. - \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{2} \phi_j\left(\frac{t}{T}\right) \phi_j''\left(\frac{t}{T}\right) \left(\frac{u}{T}\right)^2 + \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{6} \phi_j\left(\frac{t}{T}\right) \phi_j'''\left(\frac{h_{t,u}}{T}\right) \left(\frac{u}{T}\right)^3 \right\} \\ &\quad + O(1/T), \end{aligned} \quad (55)$$

where $\sum_{1 \leq t \leq u, T+u \leq t \leq T}$ refers to the sum over either the indices $1 \leq t \leq u$ (if $u > 0$) or the indices

$T+u \leq t \leq T$ (if $u \leq 0$).

Note that for the last term in (55),

$$\begin{aligned} \left| \sum_{u=-L_T}^{L_T} \left\{ \frac{1}{B} \sum_{j=1}^B \frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{6} \phi_j\left(\frac{t}{T}\right) \phi_j'''\left(\frac{h_{t,u}}{T}\right) \left(\frac{u}{T}\right)^3 \right\} \Gamma_u \right| &\leq CL_T O\left(\left(\frac{B}{T}\right)^3\right) \sum_{u=-L_T}^{L_T} |u|^2 \Gamma_u \\ &= O\left(\frac{L_T B^{3-q}}{T^{3-q}}\right) O\left(\left(\frac{B}{T}\right)^q\right) = o\left(\left(\frac{B}{T}\right)^q\right), \end{aligned} \quad (56)$$

for some constant C , where the first inequality uses that $\phi_j^m\left(\frac{h_{t,u}}{T}\right) = O(B^3)$ by Assumption 3 and the second line uses Assumption 1(b) and (51). We accordingly do not consider this term in (55) when evaluating (53), given that this introduces an error only of order $o((B/T)^q)$.

Now considering the term $\sum_{1 \leq t \leq u, T+u \leq t \leq T} \phi_j\left(\frac{t}{T}\right)^2$ in (55), for any value of u such that

$|u| \leq L_T$, we have,

$$\begin{aligned}
\frac{w_j}{T} \sum_{1 \leq t \leq u, T+u \leq t \leq T} \phi_j\left(\frac{t}{T}\right)^2 &= \frac{w_j}{T} \sum_{t=1}^u \phi_j\left(\frac{t}{T}\right)^2 1\{u > 0\} + \frac{w_j}{T} \sum_{t=T+u}^T \phi_j\left(\frac{t}{T}\right)^2 1\{u \leq 0\} \\
&= \frac{w_j}{T} \sum_{t=1}^u \phi_j(0)^2 \left\{1 + O\left(\frac{u}{T}\right)\right\} 1\{u > 0\} + \frac{w_j}{T} \sum_{t=T+u}^T \phi_j(1)^2 \left\{1 + O\left(\frac{-u}{T}\right)\right\} 1\{u \leq 0\} \\
&= \left[\frac{uw_j}{T} \phi_j(0)^2 + \frac{w_j}{T} O\left(\frac{u^2}{T}\right) \right] 1\{u > 0\} + \left[\frac{-uw_j}{T} \phi_j(1)^2 + \frac{w_j}{T} O\left(\frac{u^2}{T}\right) \right] 1\{u \leq 0\} \\
&= \frac{uw_j}{T} \phi_j(0)^2 1\{u > 0\} + \frac{-uw_j}{T} \phi_j(1)^2 1\{u < 0\} + o\left(\frac{1}{T}\right), \tag{57}
\end{aligned}$$

where the second line conducts a Taylor expansion of $\phi_j(\cdot)$ around 0 and 1, respectively, in the two sums, and the last line uses that $|u| \leq L_T = o(T^{1/2})$ by (51) and Assumption 4.

We now consider the second term in (55). Assume for now that $u > 0$. We first note that,

$$\frac{w_j}{T} \sum_{t=1}^T \phi_j\left(\frac{t}{T}\right) \phi_j'\left(\frac{t}{T}\right) \frac{u}{T} - \frac{w_j}{T} \sum_{t=u}^T \phi_j\left(\frac{t}{T}\right) \phi_j'\left(\frac{t}{T}\right) \frac{u}{T} = \frac{w_j}{T} \sum_{t=1}^{u-1} \phi_j\left(\frac{t}{T}\right) \phi_j'\left(\frac{t}{T}\right) \frac{u}{T}, \tag{58}$$

and further that,

$$\begin{aligned}
\left| \sum_{u=1}^{L_T} \left\{ \left[\frac{1}{B} \sum_{j=1}^B \frac{w_j}{T} \sum_{t=1}^{u-1} \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) \frac{u}{T} \right] - 1 \right\} \Gamma_u \right| &\leq \sum_{u=1}^{L_T} \left| \left[\frac{1}{B} \sum_{j=1}^B \frac{w_j}{T} \sum_{t=1}^{u-1} \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) \frac{u}{T} \right] - 1 \right| |\Gamma_u| \\
&\leq \frac{CO(B/T)}{T} \sum_{u=1}^{L_T} |u|^2 |\Gamma_u| = o(1/T), \tag{59}
\end{aligned}$$

where the second line uses that $\left| \frac{w_j}{T} \sum_{t=1}^{u-1} \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) \right| \leq \left| \frac{w_j}{T} \sum_{t=1}^{u-1} C_a \times O(B) \right| \leq u \times |C \times O(B)|$ for some constants C_a and C by Assumption 3, and further uses Assumptions 1(b) and 4. The same logic applies for $u < 0$.

We thus deal directly with $\frac{w_j}{T} \sum_{t=1}^T \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) \frac{u}{T}$. It can be seen that uniformly in

$|u| \leq L_T$, this value can be approximated by Euler summation (as in Phillips (2005), Lemma A) as,

$$\begin{aligned}
\frac{w_j}{T} \sum_{t=1}^T \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) \frac{u}{T} &= \frac{w_j}{T} \frac{u}{T} \int_1^T \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) dt + \frac{w_j}{2T} \frac{u}{T} \left[\phi_j \left(\frac{1}{T} \right) \phi_j' \left(\frac{1}{T} \right) + \phi_j(1) \phi_j'(1) \right] \\
&\quad + \frac{w_j}{T} \frac{u}{T} \int_1^T \left\{ t - [t] - \frac{1}{2} \right\} \left\{ \phi_j \left(\frac{t}{T} \right) \phi_j'' \left(\frac{t}{T} \right) + \phi_j' \left(\frac{t}{T} \right)^2 \right\} \frac{1}{T} dt \\
&= \frac{w_j}{T} \frac{u}{T} \int_1^T \phi_j \left(\frac{t}{T} \right) \phi_j' \left(\frac{t}{T} \right) dt + \frac{w_j}{T} \frac{u}{T} \int_1^T \left\{ t - [t] - \frac{1}{2} \right\} \left\{ \phi_j \left(\frac{t}{T} \right) \phi_j'' \left(\frac{t}{T} \right) + \phi_j' \left(\frac{t}{T} \right)^2 \right\} \frac{1}{T} dt \\
&\quad + o(1/T) \\
&= \frac{w_j u}{T} \int_0^1 \phi_j(s) \phi_j'(s) ds + O(1/T), \tag{60}
\end{aligned}$$

where $[\cdot]$ is the greatest lesser integer function, and the approximations in the second and third equalities hold by Assumption 3 and (51).

For the third term in (55), by similar steps as in (58)–(59), we can approximate the sum

$$\frac{w_j}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{2} \phi_j \left(\frac{t}{T} \right) \phi_j'' \left(\frac{t}{T} \right) \left(\frac{u}{T} \right)^2 \text{ by } \frac{w_j}{T} \sum_{t=1}^T \frac{1}{2} \phi_j \left(\frac{t}{T} \right) \phi_j'' \left(\frac{t}{T} \right) \left(\frac{u}{T} \right)^2.$$

Further, as in (60), we can

approximate this latter value with the integral $\frac{w_j}{2} \left(\frac{u}{T} \right)^2 \int_0^1 \phi_j(s) \phi_j''(s) ds$.

Combining all the results above, we can thus write (53) as,

$$\begin{aligned} E\hat{\Omega}^{WOS} - \Omega &= \sum_{u=-L_T}^{L_T} -\frac{1}{B} \sum_{j=1}^B w_j \left\{ \left[\phi_j(0)^2 1\{u > 0\} + \phi_j(1)^2 1\{u < 0\} \right] \frac{|u|}{T} \right. \\ &\quad \left. + \frac{u}{T} \int_0^1 \phi_j(s) \phi_j'(s) ds - \frac{1}{2} \left(\frac{u}{T} \right)^2 \int_0^1 \phi_j(s) \phi_j''(s) ds \right\} \Gamma_u + o\left(\left(\frac{B}{T} \right)^q \right) + O\left(\frac{1}{T} \right). \end{aligned} \quad (61)$$

Integrating the second term by parts,

$$\int_0^1 \phi_j(s) \phi_j'(s) ds = \phi_j(1)^2 - \phi_j(0)^2 - \int_0^1 \phi_j(s) \phi_j''(s) ds = \frac{1}{2} [\phi_j(1)^2 - \phi_j(0)^2]. \quad (62)$$

We can use this to write the first two terms in (61) as follows, for each value j :

$$\begin{aligned} &\sum_{u=-L_T}^{L_T} \left\{ \left[\phi_j(0)^2 1\{u > 0\} + \phi_j(1)^2 1\{u < 0\} \right] \frac{|u|}{T} + \frac{u}{T} \int_0^1 \phi_j(s) \phi_j'(s) ds \right\} \Gamma_u \\ &= \sum_{u=-L_T}^{L_T} \left\{ \left[\phi_j(0)^2 1\{u > 0\} + \phi_j(1)^2 1\{u < 0\} \right] \frac{|u|}{T} + \frac{1}{2} \frac{u}{T} [\phi_j(1)^2 - \phi_j(0)^2] \right\} \Gamma_u \\ &= \sum_{u=1}^{L_T} \frac{1}{2} \frac{u}{T} [\phi_j(0)^2 + \phi_j(1)^2] \Gamma_u + \sum_{u=-L_T}^{-1} \frac{1}{2} \frac{-u}{T} [\phi_j(0)^2 - \phi_j(1)^2] \Gamma'_{-u} \\ &= \sum_{u=1}^{L_T} \frac{1}{2} \frac{u}{T} [\phi_j(0)^2 + \phi_j(1)^2] (\Gamma_u + \Gamma'_{-u}) \\ &= \sum_{u=-L_T}^{L_T} \frac{1}{2} \frac{|u|}{T} [\phi_j(0)^2 + \phi_j(1)^2] \Gamma_u. \end{aligned} \quad (63)$$

Thus (61) becomes,

$$E\hat{\Omega}^{WOS} - \Omega = - \sum_{u=-L_T}^{L_T} \frac{1}{B} \sum_{j=1}^B w_j \left\{ \frac{1}{2} \left| \frac{u}{T} \right| [\phi_j(0)^2 + \phi_j(1)^2] - \frac{1}{2} \left(\frac{u}{T} \right)^2 \int_0^1 \phi_j(s) \phi_j''(s) ds \right\} \Gamma_u + o\left(\left(\frac{B}{T}\right)^q\right), \quad (64)$$

which further uses that $O(1/T) = o((B/T)^q)$. Thus,

$$\begin{aligned} (E\hat{\Omega}^{WOS} - \Omega)\Omega^{-1} &= -\frac{1}{B} \sum_{j=1}^B w_j \frac{\phi_j(0)^2 + \phi_j(1)^2}{2} \frac{1}{T} \sum_{u=-L_T}^{L_T} |u| \Gamma_u + \frac{1}{B} \sum_{j=1}^B w_j \frac{\int_0^1 \phi_j(s) \phi_j''(s) ds}{2} \frac{1}{T^2} \sum_{u=-L_T}^{L_T} |u|^2 \Gamma_u \\ &\quad + o\left(\left(\frac{B}{T}\right)^q\right) \\ &= \begin{cases} -\left(\frac{B}{T}\right) \left(\frac{1}{B^2} \sum_{j=1}^B w_j \frac{\phi_j(0)^2 + \phi_j(1)^2}{2} \right) \sum_{u=-\infty}^{\infty} |u| \Gamma_u + o\left(\left(\frac{B}{T}\right)\right), & \text{if } \lim_{B \rightarrow \infty} \frac{1}{B^2} \sum_{j=1}^B w_j \frac{\phi_j(0)^2 + \phi_j(1)^2}{2} \neq 0 \\ -\left(\frac{B}{T}\right)^2 \left(-\frac{1}{B^3} \sum_{j=1}^B w_j \frac{\int_0^1 \phi_j(s) \phi_j''(s) ds}{2} \right) \sum_{u=-\infty}^{\infty} |u|^2 \Gamma_u + o\left(\left(\frac{B}{T}\right)^2\right), & \text{otherwise.} \end{cases} \quad (65) \end{aligned}$$

Equation (33) in the theorem follows from these expressions. These expressions and their relation to equations (53)–(54), along with the definition of the generalized derivative in (29),

make clear that in the limit, we have $\hat{k}^{WOS(1)}(0) = \lim_{B \rightarrow \infty} \frac{1}{B^2} \sum_{j=1}^B w_j \frac{\phi_j(0)^2 + \phi_j(1)^2}{2}$,

$\hat{k}^{WOS(2)}(0) = -\lim_{B \rightarrow \infty} \frac{1}{B^3} \sum_{j=1}^B w_j \frac{\int_0^1 \phi_j(s) \phi_j''(s) ds}{2}$, where \hat{k} is as in (54). We note further that in a

neighborhood around $x = 0$ and with $|x| \leq L_T$, it must be the case in the limit that

$\hat{k}^{WOS}(x) - k^{WOS}(x) \rightarrow 0$ given the definitions in (22) and (54), so that these definitions hold as well for k .

Further, as in Priestley (1981, p. 460), it is apparent from the definition of the generalized derivative that if $0 < k^{WOS(1)}(0) < \infty$, then $\frac{1 - k^{WOS}(x)}{x^2} \rightarrow \infty$ as $x \rightarrow 0$, so that $q = 1$. If not, then

from (65) it must be the case that $q = 2$ since $\hat{\Omega}^{WOS}$ is psd,²⁵ analogous to the Priestley (1981, p. 568) result that the Parzen characteristic exponent of a kernel must be no greater than $q = 2$ for a psd kernel estimator. (Assumption 3 then guarantees the finiteness of $k^{WOS(q)}(0)$.)

Finally, as noted in Remark 7, while the above proof proceeded under the small- b sequence in Assumption 4, we can also show that using the heuristic sequential-limit definition of the implied mean kernel in Section 2 in fact yields the same result as in the theorem, despite not being formally justified under the assumed sequence. To see this, first note that because $k(0) = 1$ for implied mean kernels, if the implied mean kernel is differentiable at zero, then $k^{(1)}(0) = -k'(0)$. Accordingly, using equation (23) and differentiating with respect to x for $x > 0$,

$$\frac{d\tilde{k}_j^{WOS}(B^{-1}x)}{dx} = \tilde{k}_j^{WOS'}(B^{-1}x) = B^{-1} \left[\frac{\tilde{k}_j(B^{-1}x) - \phi_j(B^{-1}x)\phi_j(0) - \int_0^1 \phi_j(s)\phi_j'(s - B^{-1}x)ds}{1 - B^{-1}x} \right], \quad (66)$$

so

$$\tilde{k}_j^{WOS'}(0) = B^{-1} \left[\tilde{k}_j(0) - \phi_j(0)^2 - \int_0^1 \phi_j(s)\phi_j'(s)ds \right] = B^{-1} \left[1 - \frac{1}{2}(\phi_j(0)^2 + \phi_j(1)^2) \right], \quad (67)$$

where the final expression uses $\tilde{k}_j(0) = 1$ and integrates by parts, as in (62). (The same expression for $\tilde{k}_j'(0)$ obtains starting from the expression for $k(x)$, $x \leq 0$.) Plugging this into (23) and taking the limit as $B \rightarrow \infty$ yields the same result as stated in the theorem for $q = 1$ estimators.

Priestley (1981, p. 460) also shows that for q even, $k^{(q)}(0) = -\frac{1}{q!} \left[\frac{d^q(k(x))}{dx^q} \right]_{x=0}$, yielding the relation $k^{(2)}(0) = -\frac{1}{2}k''(0)$. Then for $x > 0$, differentiating the expression in (66) again yields

²⁵ See also the proof of Theorem 5(i) below: all orthonormal series in $L^2[0,1]$ are spanned by the Fourier basis functions, so there is no basis function for which $k^{(2)}(0) = 0$.

$$\tilde{k}_j^{WOS''}(B^{-1}x) = B^{-2} \left[\frac{2\tilde{k}_j^{WOS'}(B^{-1}x) + \phi_j(B^{-1}x)\phi_j'(0) - \phi_j'(B^{-1}x)\phi_j(0) + \int_V^1 \phi_j(s)\phi_j''(s - B^{-1}x)ds}{1 - B^{-1}x} \right], \quad (68)$$

so that

$$\tilde{k}_j^{WOS''}(0) = B^{-2} \int_0^1 \phi_j(s)\phi_j''(s)ds, \quad (69)$$

and again the same expression obtains starting from $x \leq 0$. It follows from (23) that

$$k^{WOS''}(0) = \frac{1}{B} \sum_{j=1}^B w_j \tilde{k}_j^{WOS''}(0), \text{ and substituting (69) into this final expression and taking } B \rightarrow \infty$$

yields the same result as stated in the theorem for $q = 2$ estimators.

To complete the heuristic re-derivation of the result in the theorem under the sequential limit, using the implied mean kernel representation for $E\hat{\Omega}^{WOS}$ in equation (22), we can follow Priestley (1981, p. 459) and write

$$E\hat{\Omega}^{WOS} - \Omega = \sum_{u=-(T-1)}^{T-1} \{k_{B,T}^{WOS}(u/S) - 1\} \Gamma_u - \sum_{|u| \geq T} \Gamma_u + o((B/T)^q). \quad (70)$$

For the second term,

$$\left| \sum_{|u| \geq T} \Gamma_u \right| \leq \sum_{|u| \geq T} |\Gamma_u| \leq \frac{1}{T^q} \sum_{|u| \geq T} |u|^q |\Gamma_u| = o(T^{-q}) = o((B/T)^q), \quad (71)$$

by Assumptions 1(b) and 4. And for the first term in the bias expression, we can write

$$\sum_{u=-(T-1)}^{T-1} \{k_{B,T}^{WOS}(u/S) - 1\} \Gamma_u = -S^{-q} \sum_{u=-(T-1)}^{T-1} \left\{ \frac{1 - k_{B,T}^{WOS}(u/S)}{(|u|/S)^q} \right\} |u|^q \Gamma_u$$

$$= (B/T)^q k^{WOS(q)}(0) \left[-2\pi s_z^{(q)}(0) \right] \{1 + o(1)\}, \quad (72)$$

where the second equality holds from the definition $SB = T$, from Parzen (1957, Theorem 5B) or Priestley (1981, p. 459) under the assumed sequence, and where $s_z^{(q)}(0)$ is the q^{th} generalized derivative of the spectral density at frequency zero. This completes the heuristic re-derivation of the more formal proof above for the result in Theorem 1(i), and note that in this case we did not require differentiability of the basis functions in order to obtain this result (so that the heuristic derivation applies to the SS basis functions, as will be confirmed more formally below in the derivation for Remark 8).

(ii) For WOS estimators, first note that Assumption 1 directly implies that a multivariate martingale functional central limit theorem holds for the partial sums of z_t (see, e.g., Helland

(1982)): for $\lambda \in [0, 1]$, we have that $T^{-1/2} \sum_{t=1}^{\lfloor T\lambda \rfloor} z_t \xrightarrow{d} \Omega^{1/2} W_m(\lambda)$, where $\lfloor \cdot \rfloor$ is the greatest lesser

integer function and W_m is an m -dimensional standard Brownian motion on the unit interval.

(This verifies an assumption by Sun (2013, 2014b), whose results we apply.) We thus have as in

(25) that $\hat{\Omega} \xrightarrow{d} \Omega^{1/2} \left(\sum_{j=1}^B w_j \Xi_j \right) \Omega^{1/2'}$, where $\Xi_j \sim \text{i.i.d. } W_m(I_m, 1)$.

Therefore, as in Sun (2014b), we have in this case that

$$mF_T \xrightarrow{d} \eta' \left(\sum_{j=1}^B w_j \Xi_j \right)^{-1} \eta \equiv mF_{\infty, m, B}, \quad (73)$$

where $\eta \sim N(0, I_m)$ and η is independent of Ξ_j for all j . Write

$$\sum_{j=1}^B w_j \Xi_j = \begin{pmatrix} \varsigma_{11} & \varsigma_{12} \\ \varsigma_{21} & \varsigma_{22} \end{pmatrix}, \quad (74)$$

where $\varsigma_{11} \in \mathbb{R}$, $\varsigma_{22} \in \mathbb{R}^{(m-1) \times (m-1)}$, and so on. Then as in Sun (2014b, equation (10)), we have

equivalently that $mF_{\infty, m, B} \sim \|\eta\|^2 / (\varsigma_{11} - \varsigma_{12} \varsigma_{22}^{-1} \varsigma_{21})$. We then proceed to take a Taylor expansion of

$G_m(z \times (\varsigma_{11} - \varsigma_{12}\varsigma_{22}^{-1}\varsigma_{21}))$ around $G_m(z)$ for arbitrary argument z . Note first that it can be shown quickly (as in Lemma 3 of Sun (2014b)) that

$$\begin{aligned} E(\varsigma_{11}) &= \sum_{j=1}^B w_j = 1, \\ E(\varsigma_{11} - \varsigma_{12}\varsigma_{22}^{-1}\varsigma_{21}) &= 1 - (m-1) \left(\sum_{j=1}^B w_j^2 \right) (1 + o(1)) = 1 - \frac{\Psi}{B}(m-1) + o(b), \\ E\left[(\varsigma_{11} - \varsigma_{12}\varsigma_{22}^{-1}\varsigma_{21})^2 \right] &= 1 + 2(2-m) \frac{\Psi}{B} + o(b), \end{aligned} \quad (75)$$

where again $B = b^{-1}$. Thus a Taylor expansion gives, after some algebra, that

$$\begin{aligned} P(mF_{\infty, m, B} \leq z) &= E\left[G_m\left(z(\varsigma_{11} - \varsigma_{12}\varsigma_{22}^{-1}\varsigma_{21})\right) \right] = G_m(z) - G'_m(z)z(m-1) \frac{\Psi}{B} + \frac{1}{2} G''_m(z)z^2 \left(2 \frac{\Psi}{B} \right) + o(b) \\ &= G_m(z) + \frac{\Psi}{B} \left[G''_m(z)z^2 - G'_m(z)z(m-1) \right] + o(b). \end{aligned} \quad (76)$$

Denote by $\tilde{c}_{m, B}^\alpha$ the $1 - \alpha$ quantile of the distribution $F_{\infty, m, B}$, so that $P(F_{\infty, m, B} > \tilde{c}_{m, B}^\alpha) = \alpha$, and define the critical value $c_m^\alpha(b)$ in the text as $c_m^\alpha(b) = m\tilde{c}_{m, B}^\alpha$. Then from (76),

$$1 - \alpha = G_m(c_m^\alpha(b)) + \frac{\Psi}{B} \left[G''_m(c_m^\alpha(b))(c_m^\alpha(b))^2 - G'_m(c_m^\alpha(b))c_m^\alpha(b)(m-1) \right] + o(b). \quad (77)$$

Now, following Sun (2011, Lemma 3) and Sun (2014b, Lemma 1), we have for the finite-sample test statistic that

$$P(F_T^* < z) = E\left[G_m\left(z\Theta_T^{-1}\right) \right] + O(1/T), \quad (78)$$

where $\Theta_T = e'_T \left[\Omega^{1/2} \hat{\Omega}^{-1} \Omega^{1/2'} \right] e_T$ and $e_T = \left[\text{var}\left(\sqrt{T}\bar{z}_0\right) \right]^{-1/2} \sqrt{T}\bar{z}_0 / \left\| \left[\text{var}\left(\sqrt{T}\bar{z}_0\right) \right]^{-1/2} \sqrt{T}\bar{z}_0 \right\|$. Then

as in Sun (2011, Theorem 4), Sun (2014b, Theorem 2), we can take a Taylor expansion

$\Theta^{-1} = 1 + L + Q + o_p\left((bT)^{-q} + b\right) + O_p(1/T)$, where L is linear in $\hat{\Omega} - \Omega$ and Q is quadratic in that difference (see Sun (2014b, p. 675) for definitions of these values applied directly here). From part (i) of the theorem, $E\hat{\Omega} - \Omega = -(B/T)^q k^{WOS(q)}(0) \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j + o\left((B/T)^q\right)$,²⁶ and therefore, again following the steps in Sun (2011, Theorem 4), Sun (2014b, Theorem 2),

$$E[L] = (B/T)^q k^{WOS(q)}(0) \omega^{(q)} + o\left((bT)^{-q} + b\right). \quad (79)$$

Similarly,

$$E[L^2] = 2 \frac{\Psi}{B} + o\left((bT)^{-q} + b\right) \text{ and } E[Q] = -\frac{\Psi}{B}(m-1) + o\left((bT)^{-q} + b\right). \quad (80)$$

Thus, expanding (78),

$$\begin{aligned} P\left(F_T^* < z\right) &= G_m\left(z \frac{B}{B-m+1}\right) + G'_m(z) z E[L+Q] + \frac{1}{2} E G''_p(z) z^2 E[L^2] + o\left((bT)^{-q} + b\right) + O\left(\frac{1}{T}\right) \\ &= G_m(z) + G'_m(z) z \omega^{(q)} k^{WOS(q)}(0) (bT)^{-q} - G'_m(z) z \frac{\Psi}{B} (m-1) + G''_p(z) z^2 \frac{\Psi}{B} \\ &\quad + o(b) + o\left((bT)^{-q}\right). \end{aligned} \quad (81)$$

Using $z = c_m^\alpha(b)$ and combining this with (77), and additionally noting that $c_m^\alpha(b) = \chi_m^\alpha + O(b)$ as in Sun (2014b, p. 665), immediately gives the null expansion (31) for WOS tests.

The expansion under the alternative follows Sun (2011, Theorem 5(b)) and Sun (2014b, Theorem 5). We omit the calculations here, since they follow the same steps as in those papers and above, but they are available upon request. After these calculations, we obtain

²⁶ While this expression has not been shown formally for the SS series estimator, it is shown below in the derivation of Remark 8, which does not use this theorem in the case of the SS estimator.

$$\begin{aligned} \Pr_{\delta} [F_T^* > c_m^{\alpha}(b)] &= [1 - G_{m,\delta^2}(\chi_m^{\alpha})] + G'_{m,\delta^2}(\chi_m^{\alpha}) \chi_m^{\alpha} \omega^{(q)} k^{(q)}(0) (B/T)^q \\ &\quad - \frac{1}{2} \delta^2 G'_{m+2,\delta^2}(\chi_m^{\alpha}) \chi_m^{\alpha} \frac{\Psi}{B} + o(b) + o((bT)^{-q}), \end{aligned} \quad (82)$$

as stated.

For kernel estimators, given Assumptions 1 and 2, equation (31) follows from Sun (2014b) equation (16), along with $c_m^{\alpha}(b) = \chi_m^{\alpha} + O(b)$. Equation (32) follows directly from the proof of Sun (2014b) Theorem 5 for the case of the Gaussian location model.

We note that in the scalar case with $m = 1$, we have that the higher-order size distortion (in both cases) is proportional to $|s_z''(0)/s_z(0)|$, so that all results below hold uniformly over all stochastic processes satisfying Assumption 1 with $|s_z''(0)/s_z(0)| \leq \kappa$ for finite κ . \square

Proof of Theorem 2: We can define the size-adjusted critical value as

$c_{m,T}^{\alpha}(b) = c_m^{\alpha}(b) + \delta_{m,T}^{\alpha}(b)$, where $c_m^{\alpha}(b)$ is the fixed- b critical value as in equation (31) and $\delta_{m,T}^{\alpha}(b)$ is defined implicitly by $\Pr_0[F_T^* > c_{m,T}^{\alpha}(b)] = \alpha + o(b) + o((bT)^{-q})$. Taking a Taylor expansion of the null rejection rate around $c_m^{\alpha}(b)$,

$$\begin{aligned} \Pr_0[F_T^* > c_{m,T}^{\alpha}(b)] &= \alpha + G'_m(\chi_m^{\alpha}) \chi_m^{\alpha} \omega^{(q)} k^{(q)}(0) (bT)^{-q} - \delta_{m,T}^{\alpha}(b) G'_m(\chi_m^{\alpha}) [1 + O(b) + O((bT)^{-q})] \\ &\quad + o(b) + o((bT)^{-q}) + o(\delta_{m,T}^{\alpha}(b)), \end{aligned} \quad (83)$$

where the fact that $\Pr'_0[F_T^* > c_{m,T}^{\alpha}(b)] = -G'_m(\chi_m^{\alpha}) [1 + O(b) + O((bT)^{-q})] + o(b) + o((bT)^{-q})$

follows from Sun (2014b, Theorem 2 and p. 665). Using this and

$\Pr_0[F_T^* > c_{m,T}^{\alpha}(b)] = \alpha + o(b) + o((bT)^{-q})$ by definition, we can solve for $\delta_{m,T}^{\alpha}(b)$ as

$\delta_{m,T}^{\alpha}(b) = k^{(q)}(0) (bT)^{-q} \chi_m^{\alpha} \omega^{(q)}$, from which (34) follows directly.

Then taking a similar Taylor expansion, size-adjusted power is

$$\begin{aligned} \Pr_{\delta}[F_T^* > c_{m,T}^{\alpha}] &= [1 - G_{m,\delta^2}(\chi_m^{\alpha})] + G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q} - \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}v^{-1} \\ &\quad - \delta_{m,T}^{\alpha}(b)G'_{m,\delta^2}(\chi_m^{\alpha})\left[1 + O(b) + O((bT)^{-q})\right] + o(b) + o((bT)^{-q}) + o(\delta_{m,T}^{\alpha}(b)), \end{aligned} \quad (84)$$

which analogously uses Theorem 5 of Sun (2014b). We have

$\delta_{m,T}^{\alpha}(b)G'_{m,\delta^2}(\chi_m^{\alpha}) = k^{(q)}(0)(bT)^{-q}G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}$ from the solution for $\delta_{m,T}^{\alpha}(b)$ above. Thus the second and fourth terms in (84) cancel, and using this along with $\delta_{m,T}^{\alpha}(b) = O((bT)^{-q})$ yields the size-adjusted power relation given in equation (35). \square

Proof of Theorem 3: This follows directly from equations (31) and (32). Fix a sequence b_1 for test F_1 . Given equivalent values of q for tests F_1 and F_2 , equation (31) gives that we must set $b_2 = \left[k_2^{(q)}(0)/k_1^{(q)}(0)\right]^{1/q} b_1$ in order to obtain equivalent higher-order size. We thus have that

$$G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k_2^{(q)}(0)(b_2T)^{-q} = G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k_1^{(q)}(0)(b_1T)^{-q}, \quad (85)$$

so that the corresponding second terms in the power expression (32) for F_1^* and F_2^* are equivalent. Using this along with equation (32) yields the desired relation. \square

Proof of Corollary 1: Using equation (31) and the definition of the size distortion Δ_S , that size distortion can be written as follows:

$$\Delta_S = G'_m(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q} + o(b) + o((bT)^{-q}). \quad (86)$$

Using the definition of the size-adjusted power loss $\Delta_P(\delta)$ and the fact that $v = (b\psi)^{-1}$, we also have,

$$\Delta_P(\delta) = \frac{1}{2} \delta^2 G'_{(m+2), \delta^2}(\chi_m^\alpha) \chi_m^\alpha b \psi + o(b) + o((bT)^{-q}). \quad (87)$$

We can see that the leading terms in (86) and (87) are of equivalent asymptotic order if and only if b is of equivalent asymptotic order as $(bT)^{-q}$, requiring that $O(b^{-(q+1)}) = O(T^q)$, or equivalently that $b \sim CT^{-\frac{q}{q+1}}$ for some constant $C \in \mathbb{R}_{>0}$, as stated.

Further, given that $O(b) = O((bT)^{-q})$ under this sequence, the $o(b)$ remainder term in (86) is also $o((bT)^{-q})$, confirming that the leading term in (86) is in fact the first term. The same applies to the remainder terms in (87). \square

Proof of Theorem 4: (i) Under the assumed sequence, we can rewrite the size distortion in (86) as

$$\begin{aligned} |\Delta_S|^{1/q} &= \left(G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q} \left| \omega^{(q)} \right|^{1/q} \left(k^{(q)}(0) \right)^{1/q} (bT)^{-1} [1 + o(1)]^{1/q} \\ &= \left(G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q} \left| \omega^{(q)} \right|^{1/q} \left(k^{(q)}(0) \right)^{1/q} (bT)^{-1} + o((bT)^{-1}), \end{aligned} \quad (88)$$

where the first equality uses the argument in the last paragraph of the proof of Corollary 1, and the second equality uses that $|1 + o(1)|^{1/q} \leq |1 + o(1)|$ for $q \leq 2$. This can be rewritten further as

$$T |\Delta_S|^{1/q} = \left(G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q} \left| \omega^{(q)} \right|^{1/q} \left(k^{(q)}(0) \right)^{1/q} b^{-1} + o(1/b). \quad (89)$$

Similarly, rewrite (87) as

$$\Delta_P(\delta) = \frac{1}{2} \delta^2 G'_{(m+2), \delta^2}(\chi_m^\alpha) \chi_m^\alpha b \psi + o(b). \quad (90)$$

Multiplying (89) and (90), and defining $a_{m, \alpha, q}(\delta) = \frac{1}{2} \delta^2 G'_{(m+2), \delta^2}(\chi_m^\alpha) \chi_m^\alpha \left(G'_m(\chi_m^\alpha) \chi_m^\alpha \right)^{1/q}$,

we obtain

$$T\Delta_P(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta)\left[\left(k^{(q)}(0)\right)^{1/q}\psi\right]|\omega^{(q)}|^{1/q} + o(1), \quad (91)$$

as stated, since $\ell^{(q)}(k) = \left(k^{(q)}(0)\right)^{1/q}\psi$. This applies for both kernel and WOS tests.

(ii) We can express the maximum size-adjusted power loss Δ_P^{\max} using its definition in (39) and equation (87) as,

$$\Delta_P^{\max} = \sup_{\delta} \left\{ \frac{1}{2} \delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha \right\} b\psi + o(b), \quad (92)$$

since δ does not enter into the term $b\psi$. Thus following the same steps as in part (i) above, multiplying (89) and (92) yields equation (41).

(iii) Again using equation (31) and the same steps as in parts (i)-(ii), we can express $\sqrt{(|\Delta_S|)}$ for any test ($q = 1$ or $q = 2$) as,

$$\sqrt{|\Delta_S|} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/w} |\omega^{(q)}|^{1/2} \left(k^{(q)}(0)\right)^{1/2} (bT)^{-q/2} + o\left((bT)^{-q/2}\right). \quad (93)$$

Multiplying this by (92), under the assumed sequence for b ,

$$\Delta_P^{\max} \sqrt{|\Delta_S|} = \bar{a}_{m,\alpha,2} \left(\sqrt{k^{(q)}(0)\psi}\right) |\omega^{(q)}|^{1/2} T^{-1} (bT)^{1-q/2} + o(b^{3/2}). \quad (94)$$

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

$\Delta_P^{\max,q-2} \sqrt{|\Delta_S^{q-2}|} < \Delta_P^{\max,q-1} \sqrt{|\Delta_S^{q-1}|}$). This proves the stated result. \square

Proof of Theorem 5: (i) As in the proof of Theorem 1(i) above, we can confine our analysis to kernels (or implied mean kernels) with $q \leq 2$, and given that $q = 2$ kernels dominate $q = 1$ kernels from Theorem 4(iii), we focus on the $q = 2$ case.

We first consider kernel estimators. From Theorem 4, the lower envelope of the size/power frontier is achieved for any data-generating process 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 , as in (11). This minimum is achieved by the QS estimator (Priestley (1981, p. 571)). Thus the QS estimator's size/size-adjusted power tradeoff defines the frontier for kernel tests.

For WOS estimators, note from the representation (9)-(10) that the QS estimator can be represented as a WOS estimator (14) with the Fourier basis and weights $w_j \propto [1 - (|j|/B)^2]$, as after (12), but with $B/2$ modified here to B for notational simplicity; see also Priestley (1981, p. 444).²⁷ We proceed to show in two parts that this estimator again dominates among WOS estimators: first, we show that given any set of weights $\{w_j\}$, the Fourier basis is optimal; second, we show that the QS weights dominates 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 ψ is fixed from (24). We accordingly will show that the Fourier basis functions minimize $\sqrt{k^{(2)}(0)}$ for any choice of weights. From Theorem 1(i), a sufficient condition for this result can be obtained by showing that for all values B along a sequence chosen according to Corollary 1, the

²⁷ For purposes of more fully aligning the notation here with the notation given after (12), one may instead define $\tilde{B} = B/2$ and then proceed using \tilde{B} in place of B for the remainder of the proof; the results in this case would be equivalent to the results derived here. More generally, renormalizations of B are without loss of generality for comparisons across estimators, since such a renormalization for a given estimator is reflected in the equivalent degrees of freedom ν and the approximating distribution (26).

Fourier basis minimizes $\left| \sum_{j=1}^B w_j k_j^{(2)}(0) \right|$, where $k_j^{(2)}(0) = \int_0^1 \phi_j(s) \phi_j''(s) ds$, across orthonormal series.

Given that the Fourier basis functions span $L^2[0,1]$, we can write any basis function as

$$\phi_j(s) = \sum_{l=1}^T a_{jl} e^{-i2\pi ls}, \quad (95)$$

where the a_{jl} values are as-yet undetermined projection coefficients. (This representation is sufficient to match any basis function at all points $s = t/T$ for given T ; by instead setting the upper limit of the summation in (95) to some value \bar{T} and taking $\bar{T} \rightarrow \infty$, it becomes equivalent to $\phi_j(s)$ for all $s \in [0,1]$. We proceed with the representation (95) for notational ease, but the arguments in the proof carry through for all values \bar{T} and therefore for all basis functions with arbitrarily small error.) We know that for any orthonormal series,

$$1 = \int_0^1 |\phi_j(s)|^2 ds = \sum_l \sum_{l'} a_{jl} a_{jl'} \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} ds = \sum_l a_{jl}^2, \quad (96)$$

and similarly

$$0 = \int_0^1 \phi_j(s) \overline{\phi_{j' \neq j}(s)} ds = \sum_l a_{jl} a_{j'l}. \quad (97)$$

Our minimization problem is then

$$\min_{a_{jl}} \left| \sum_{j=1}^B w_j \int_0^1 \phi_j(s) \overline{\phi_j''(s)} ds \right| \Leftrightarrow \min_{a_{jl}} \left| \sum_{j=1}^B w_j \sum_{l=1}^T \sum_{l'=1}^T a_{jl} a_{jl'} 4\pi^2 l^2 \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} ds \right| \Leftrightarrow \min_{a_{jl}} \sum_{j=1}^B w_j \sum_{l=1}^T a_{jl}^2 l^2 \quad (98)$$

subject to the two constraints (96) and (97). This can be written equivalently as

$$\min_A \text{tr}((AW)'D(AW)) \Leftrightarrow \min_A \text{tr}(W^2 A' D A) \quad (99)$$

subject to $A'A = I_B$, where $A = (a_{jl})$, $W = \text{diag}\left(\left[\sqrt{w_1} \quad \sqrt{w_2} \quad \dots \quad \sqrt{w_B}\right]\right)$, and $D = \text{diag}\left(\left[1 \quad 4 \quad \dots \quad T^2\right]\right)$.

Given $B < T$, this is equivalent to $\min_{\tilde{A}} \text{tr}(\tilde{W}^2 \tilde{A}' D \tilde{A})$ subject to $\tilde{A}' \tilde{A} = I_T$, where \tilde{W} is padded with zeros relative to W , so that \tilde{A} and \tilde{W} are both $T \times T$, as is D ; then given a solution \tilde{A} , the solution A selects the first B columns of this solution. From (98), the objective is linear in the entries of the matrix $\tilde{A} \circ \tilde{A}$, where \circ denotes the Hadamard (or element-wise) product, and the matrix $\tilde{A} \circ \tilde{A}$ is orthostochastic. So

$$\min_{\tilde{A} \text{ s.t. } \tilde{A}' \tilde{A} = I_T} \text{tr}(\tilde{W}^2 \tilde{A}' D \tilde{A}) \geq \min_Y \sum_{j,l} w_j l^2 \gamma_{jl}, \quad (100)$$

where $Y = (\gamma_{jl})$ is orthostochastic. The right side is linear in the entries of Y , and the set of all orthostochastic matrices is compact and convex. Thus the minimum of the right side of this inequality is obtained at an extreme point of this set. By Birkhoff's Theorem (e.g., Bhatia (1997, p. 37)), the extreme points of this set are the permutation matrices.

All permutation matrices have orthonormal columns, so (100) holds in fact with equality, and \tilde{A} minimizes $\text{tr}(\tilde{W}^2 P' D P)$ over the set of permutation matrices P . Further, \tilde{W}^2 and D are psd and diagonal. D has its diagonal terms (and eigenvalues) in ascending order. Given some unordered set of weight values, assume first that (as is the case for QS) we order them descendingly, so that $w_1 \geq w_2 \geq \dots \geq w_B$, and therefore \tilde{W}^2 has its diagonal terms (eigenvalues) in descending order. Thus in this case, the minimum of the objective problem (100) is obtained trivially by $\tilde{A} = P = I_T$, so that the minimizing A is given by the first B columns of I_T . Equivalently, $a_{jj} = 1$, $a_{j,j' \neq j} = 0$, so that from (95), we have in fact selected the Fourier basis functions as the minimizing basis.

In the case in which the weight values are not in descending order, note first that from equation (24), regardless of how the weights are ordered, ψ is nonetheless fixed. Second, the

minimal $\sqrt{k^{(2)}(0)}$ obtained from solving (100) is equivalent for a fixed set of weights regardless of their orderings: the solution A is still given by the first B columns of a permutation matrix P , where $a_{jl} = 1$ if w_l is the j^{th} largest value among the set of weights, and 0 otherwise, and therefore the same minimal value is obtained by rearranging the weights in descending order and having $a_{jj} = 1$, $a_{j,j' \neq j} = 0$. This implies that for any given set of weights, it is without loss of generality – in the sense that the size-power tradeoff is no worse – to set the weights in descending order, and to use the Fourier basis.

We have thus obtained that the frontier is obtained for any set of weights by the Fourier basis. Further, given that, as above, all orthonormal series in $L^2[0,1]$ are spanned by these basis functions, there is no set of basis functions for which $k^{(2)}(0) = 0$. This extends the classic result that all psd kernel estimators have $q \leq 2$ to the implied mean kernels of WOS estimators.

It now remains to be shown that the QS weights are optimal given the Fourier basis. Given the use of Fourier basis functions, we therefore wish to minimize

$$\ell^{(2)}(k) = \left(k^{(2)}(0)\right)^{1/2} \psi \propto \left(\frac{1}{B^2} \sum_{j=1}^B w_j j^2\right)^{1/2} \left(B \sum_{j=1}^B w_j^2\right) = \left(\sum_{j=1}^B w_j j^2\right)^{1/2} \left(\sum_{j=1}^B w_j^2\right) \quad (101)$$

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(i) and the use of Fourier basis functions (as can be seen from (98) with $a_{jj} = 1$, $a_{j,j' \neq j} = 0$), 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, but modified so that the proof is with respect to WOS estimators using the Fourier basis.

Set w_j^* to be the QS weights, $w_j^* = \bar{w}_{QS} \left[1 - (|j|/B)^2\right]$, where $\bar{w}_{QS} = 6B / [(B-1)(4B+1)]$

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_{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 $B_{\text{alt}} > B_{\text{QS}}$, then set $w_j^* = 0$ for $j > B_{\text{QS}}$, so that $w_j = \epsilon_j \geq 0$ for $B_{\text{QS}} < j \leq B_{\text{alt}}$. If $B_{\text{alt}} < B_{\text{QS}}$, then correspondingly $w_j = 0$ for $B_{\text{alt}} < j \leq B_{\text{QS}}$. To handle both possible cases simultaneously, write $\bar{B} = \max(B_{\text{alt}}, B_{\text{QS}})$.

Since $\sum_{j=1}^{\bar{B}} w_j = 1$, we must have that

$$\sum_{j=1}^{\bar{B}} \epsilon_j = 0. \quad (102)$$

Equating higher-order size, from Theorem 1, requires

$$\sum_{j=1}^{\bar{B}} w_j j^2 = \sum_{j=1}^{\bar{B}} w_j^* j^2, \quad (103)$$

so we have further that

$$\sum_{j=1}^{\bar{B}} \epsilon_j j^2 = 0. \quad (104)$$

Now note that

$$\sum_{j=1}^{\bar{B}} w_j^2 = \sum_{j=1}^{\bar{B}} (w_j^*)^2 + 2 \sum_{j=1}^{\bar{B}} \epsilon_j w_j^* + \sum_{j=1}^{\bar{B}} \epsilon_j^2, \quad (105)$$

and

$$\sum_{j=1}^{\bar{B}} \epsilon_j w_j^* = \bar{w}_{QS} \sum_{j=1}^{B_{QS}} \left[1 - \left(\frac{j}{B_{QS}} \right)^2 \right] \epsilon_j = \bar{w}_{QS} \left\{ \sum_{j=1}^{\bar{B}} \left[1 - \left(\frac{j}{B_{QS}} \right)^2 \right] \epsilon_j + \sum_{j=B_{QS}+1}^{B_{alt}} \left[\left(\frac{j}{B_{QS}} \right)^2 - 1 \right] \epsilon_j \right\}, \quad (106)$$

where the second term in the final equality is a rearrangement of $\left\{ - \sum_{j=B_{QS}+1}^{B_{alt}} [1 - (j/B_{QS})^2] \epsilon_j \right\}$. The

first term in this equality is zero from (102) and (104). For the second term, if $B_{alt} > B_{QS}$, then as

above $w_j = \epsilon_j \geq 0$ for $B_{QS} < j \leq B_{alt}$, and therefore $\sum_{j=B_{QS}+1}^{B_{alt}} [(j/B_{QS})^2 - 1] \epsilon_j \geq 0$ (and this of course

holds with equality if $B_{alt} < B_{QS}$). Thus $\sum_{j=1}^{\bar{B}} \epsilon_j w_j^* \geq 0$. It is further trivially the case that $\sum_{j=1}^{\bar{B}} \epsilon_j^2 \geq 0$.

Therefore, from (105), we have that

$$\sum_{j=1}^{\bar{B}} w_j^2 \geq \sum_{j=1}^{\bar{B}} (w_j^*)^2, \quad (107)$$

with equality if and only if $\epsilon_j = 0$ for all j . From (32), (101), and (103), this proves that QS attains greater higher-order power for equivalent higher-order size relative to all alternative WOS estimators using the Fourier basis. Combined with the fact that the frontier is obtained for any set of weights by the Fourier basis as above, we have shown that QS 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 (41) yields equation (42) up to higher-order terms. Numerically calculating $\bar{a}_{m,\alpha,q} = \sup_{\delta} a_{m,\alpha,q}(\delta)$ for $q = 2$ and $\alpha = .05$ 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 $m = 3$, as stated.

(ii) As after (25), 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 implies immediately that given equal weights, the Fourier basis functions achieve the size-power frontier for any data-generating process and value B . We conclude that the EWP test achieves the frontier for size/size-adjusted power 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 (see also the derivation for Remark 8(b) below). Further, $\psi = 1$ for this estimator. Combining these with (41) yields (43) up to higher-order terms. Again using numerical calculations for $\bar{a}_{m,\alpha,q} = \sup_{\delta} a_{m,\alpha,q}(\delta)$ for $q = 2$ and $\alpha = .05$, we obtain $\bar{a}_{m,\alpha,2}\pi / \sqrt{6} \approx 0.3623$ 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. \square

Derivation of Equation (44): Fix a sequence $B = 1/b_{\text{EWP}}$. To obtain equivalent higher-order size using the QS test, equation (31) gives that we must set

$$b_{\text{QS}} = \sqrt{\frac{k^{\text{QS}(2)}(0)}{k^{\text{EWP}(2)}(0)}} b_{\text{EWP}} = \sqrt{\frac{\pi^2 / 10}{\pi^2 / 6}} B^{-1} = \sqrt{\frac{3}{5}} B^{-1}, \quad (108)$$

where the $k^{(2)}(0)$ values for the two tests are as in the proof of Theorem 5. That proof also uses

that $\int_{-\infty}^{\infty} k^2(x) dx = \frac{6}{5}$ for QS, so that given equivalent higher-order size, we have

$$v_{\text{EWP}}^{-1} - v_{\text{QS}}^{-1} = B^{-1} - \frac{6}{5} \sqrt{\frac{3}{5}} B^{-1}. \text{ Plugging this into the higher-order power difference in equation (36)}$$

(Theorem 3) yields the desired result.

Derivations for Remark 8: (a) For SS, we can calculate $E\hat{\Omega}^{\text{SS}}$ directly (without appealing to Theorem 1(i)) to observe that the SS implied mean kernel is similar to the Bartlett kernel on a

subsample of $T/(B+1)$ observations. First, note that $\bar{x}_i - \bar{x} \equiv \frac{1}{T} \sum_{t \in T_i} x_t - \frac{1}{T} \sum_{t=1}^T x_t$ (where, abusing

notation, T_i denotes both the number of observations in subsample i and the subsample that t

indexes) can be written as $\bar{x}_i - \bar{x} = \frac{1}{T} \sum_{t=1}^T ((B+1)1\{t \in T_i\} - 1)x_t = \frac{B+1}{T} \sum_{t=1}^T (1\{t \in T_i\} - \frac{1}{B+1})x_t$.

Thus summing over subsamples and squaring, we have,

$$\frac{1}{B} \sum_{i=1}^{B+1} (\bar{x}_i - \bar{x})^2 = \frac{1}{B} \sum_{i=1}^{B+1} \left(\frac{B+1}{T}\right)^2 \sum_{t=1}^T \sum_{s=1}^T (1\{t \in T_i\} - \frac{1}{B+1})(1\{s \in T_i\} - \frac{1}{B+1})x_t x_s. \quad (109)$$

Taking the expectation of this value and performing the same change of variables as in (22),

$$\begin{aligned} E \frac{1}{B} \sum_{i=1}^{B+1} (\bar{x}_i - \bar{x})^2 &= \frac{B+1}{B} \frac{1}{T/B+1} \sum_{u=-(T-1)}^{T-1} \left[\left(1 - \left| \frac{u}{T/B+1} \right| \right) 1\left\{ |u| \leq \frac{T}{B+1} \right\} - \frac{1}{B+1} \left(1 - \left| \frac{u}{T} \right| \right) \right] \Gamma_u \\ &= \frac{B+1}{T} \sum_{u=-(T-1)}^{T-1} \left[\left(\frac{B+1}{B} - \frac{B+1}{B} \left| \frac{u}{T/(B+1)} \right| \right) 1\left\{ |u| \leq \frac{T}{B+1} \right\} - \frac{1}{B} + \frac{1}{B} + \left| \frac{u}{T} \right| \right] \Gamma_u. \end{aligned} \quad (110)$$

Converting $E \frac{1}{B} \sum_{i=1}^{B+1} (\bar{x}_i - \bar{x})^2$ to $E\hat{\Omega}^{SS}$ requires multiplying by $T/(B+1)$ given the form of the statistic given in (18) as compared to the usual t -statistic in (4). Thus in this case defining S such that $T = S(B+1)$ given that there are $B+1$ subsamples and setting $\tilde{v} = u/S$, we can write the SS implied mean kernel (i.e., the expression in brackets in (110)) as

$$k_B^{SS}(\tilde{v}) = \left(\frac{B+1}{B} - \frac{B+1}{B} |\tilde{v}| \right) 1\left\{ |\tilde{v}| \leq 1 \right\} - \frac{1}{B} + \frac{1}{B(B+1)} |\tilde{v}|. \quad (111)$$

Thus using the definition of the generalized first derivative in (29), we have

$$k_B^{SS(1)}(0) = \frac{B+1}{B} - \frac{1}{B(B+1)} = \frac{B+2}{B+1} \rightarrow 1 \text{ as } B \rightarrow \infty. \text{ Because } k^{SS(1)}(0) \neq 0, q = 1 \text{ for the SS}$$

estimator. Further, comparing $E\hat{\Omega}^{SS}$ with Ω using (110) makes apparent that the stated result in the proof of Theorem 1(ii) above holds for the SS estimator as well.

For the Bartlett/Newey-West test, Priestley (1981) Table 7.1 gives that $k^{(1)}(0) = 1$ and $q = 1$, while Table 6.1 gives that $\int_{-\infty}^{\infty} k^2(x) dx = 2/3$, so that $k^{(1)}(0) \int_{-\infty}^{\infty} k^2(x) dx = 2/3$, as stated.

For the Legendre basis (see Footnote 21), let the shifted (to $[0,1]$) but non-normalized j^{th} Legendre polynomial be $p_j(s)$. Then $\int_0^1 p_j(s)p_k(s)dx = \frac{1}{2k+1}1\{j=k\}$ (Abramowitz and Stegun (1965)), so that the j^{th} normalized shifted Legendre polynomial is $\phi_j(s) = p_j(s)\sqrt{2j+1}$. Because $p_j(0) = (-1)^j$ and $p_j(1) = 1$, we have $\phi_j(0) = \sqrt{2j+1}(-1)^j$, $\phi_j(1) = \sqrt{2j+1}$. Thus from Theorem 1(i), abusing notation slightly, we have $k_j^{\text{Leg}(1)}(0) = \frac{1}{B^2} \frac{1}{2}(2j+1) = \frac{2j+1}{B^2}$ for each j , and thus $k_B^{\text{Leg}(1)}(0) = B^{-2} \sum_{j=1} B(2j+1) = (B+2)/B \rightarrow 1$ as $B \rightarrow \infty$, as stated. As in Theorem 1(i), this also implies that $q = 1$ for the Legendre polynomials.

(b) For the Fourier basis functions, we have $\phi_{2j-1}'' = -4\sqrt{2}\pi^2 j^2 \cos(2\pi js)$ and $\phi_{2j}'' = -4\sqrt{2}\pi^2 j^2 \sin(2\pi js)$. Thus $\int_0^1 \phi_{2j-1}(s)\phi_{2j-1}''(s)ds = \int_0^1 \phi_{2j}(s)\phi_{2j}''(s)ds = -4\pi^2 j^2$. Summing over j and using Theorem 1(i), we have

$$k_B^{\text{EWP}(2)}(0) = -\frac{1}{2} \frac{1}{B} \sum_{j=1}^{B/2} \frac{1}{B^2} 2(-4\pi^2 j^2) = \frac{\pi^2}{6} \frac{(B+1)(B+2)}{B^2} \xrightarrow{B \rightarrow \infty} \frac{\pi^2}{6}. \quad (112)$$

Similarly, for cosine basis functions, using their limiting implied mean kernel form, we have $\phi_j''(s) = -\sqrt{2}\pi^2 j^2 \cos(\pi js)$ and $\int_0^1 \phi_{2,j-1}(s)\phi_{2,j-1}''(s)ds = -\pi^2 j^2$. Summing over j yields,

$$k_B^{\text{cos}(2)} = -\frac{1}{2} \frac{1}{B} \sum_{j=1}^B \frac{1}{B^2} (-\pi^2 j^2) = \frac{\pi^2}{6} \frac{(B+1)(B+1/2)}{B^2} \xrightarrow{B \rightarrow \infty} \frac{\pi^2}{6}. \quad (113)$$

Results (112) and (113) and Theorem 1(i) further imply that $q = 2$ for the Fourier and cosine estimators.

Derivations for Remark 11: First, let $\bar{c}_{m,T}^\alpha(b)$ be the size-adjusted critical value (34) based on the boundary value of $\bar{\omega}^{(q)}$:

$$\bar{c}_{m,T}^\alpha(b) = \left[1 + \bar{\omega}^{(q)} k^{(q)}(0)(bT)^{-q}\right] c_m^\alpha(b). \quad (114)$$

It follows from (32) that the null rejection rate of the test using this size-adjusted critical value, evaluated at the true value of $\omega^{(q)}$, is,

$$\Pr_0 \left[F_T^* > \bar{c}_{m,T}^\alpha(b) \right] = \alpha + G'_m(\chi_m^\alpha) \chi_m^\alpha \left(\omega^{(q)} - \bar{\omega}^{(q)} \right) k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right), \quad (115)$$

from which it follows that, for a given sequence b and under the condition in Footnote 23,

$$\sup_{\omega^{(q)} \leq \bar{\omega}^{(q)}} \Pr_0 \left[F_T^* > \bar{c}_{m,T}^\alpha(b) \right] \leq \alpha + o(b) + o\left((bT)^{-q}\right). \quad (116)$$

The expression for size-adjusted power loss $\Delta_p(\omega^{(q)}(\rho), \delta)$ in (45) then follows directly from (84) in the proof of Theorem 2 (omitting higher-order remainder terms). Solving (45) yields (46), where

$$\begin{aligned} \tilde{\omega}^{(q)} &= \int_{|\rho| \leq \bar{\rho}} \left[\bar{\omega}^{(q)}(\rho) - \omega^{(q)}(\rho) \right] d\Pi(\rho), \\ \tilde{d}_{m,\alpha,q} &= \left\{ \int_{\delta} G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha d\Pi_{\delta}(\delta) \right\} / \left\{ \left[\frac{1}{2} \int_{\delta} \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha d\Pi_{\delta}(\delta) \right]^{\frac{1}{1+q}} \right\}. \end{aligned}$$

Finally, substituting b^{WAP} in (46) into the expression for $\Delta_p(\omega^{(q)}(\rho), \delta)$, we obtain that the power loss of the test using the WAP-maximizing sequence b^{WAP} is:

$$\Delta_P^{WAP} = \left(q^{-q/(1+q)} + q^{1/(1+q)} \right) \tilde{d}_{m,\alpha,q} \left[\left(k^{(q)}(0) \right)^{1/q} \psi \right]^{\frac{q}{1+q}} \left(\tilde{\omega}^{(q)} \right)^{\frac{1}{1+q}} T^{\frac{-q}{1+q}}, \quad (117)$$

where $\tilde{d}_{m,\alpha,q} = \left[\int G'_{m,\delta^2}(\chi_m^\alpha) \chi_m^\alpha d\Pi_{\delta}(\delta) \right]^{1/(1+q)} \left[\frac{1}{2} \int \delta^2 G'_{m+2,\delta^2}(\chi_m^\alpha) \chi_m^\alpha d\Pi_{\delta}(\delta) \right]^{q/(1+q)}$. Note that

$\ell^{(q)}(k) = \left(k^{(q)}(0) \right)^{1/q} \psi$. The remaining stated results then follow.

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

m	B		
	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.

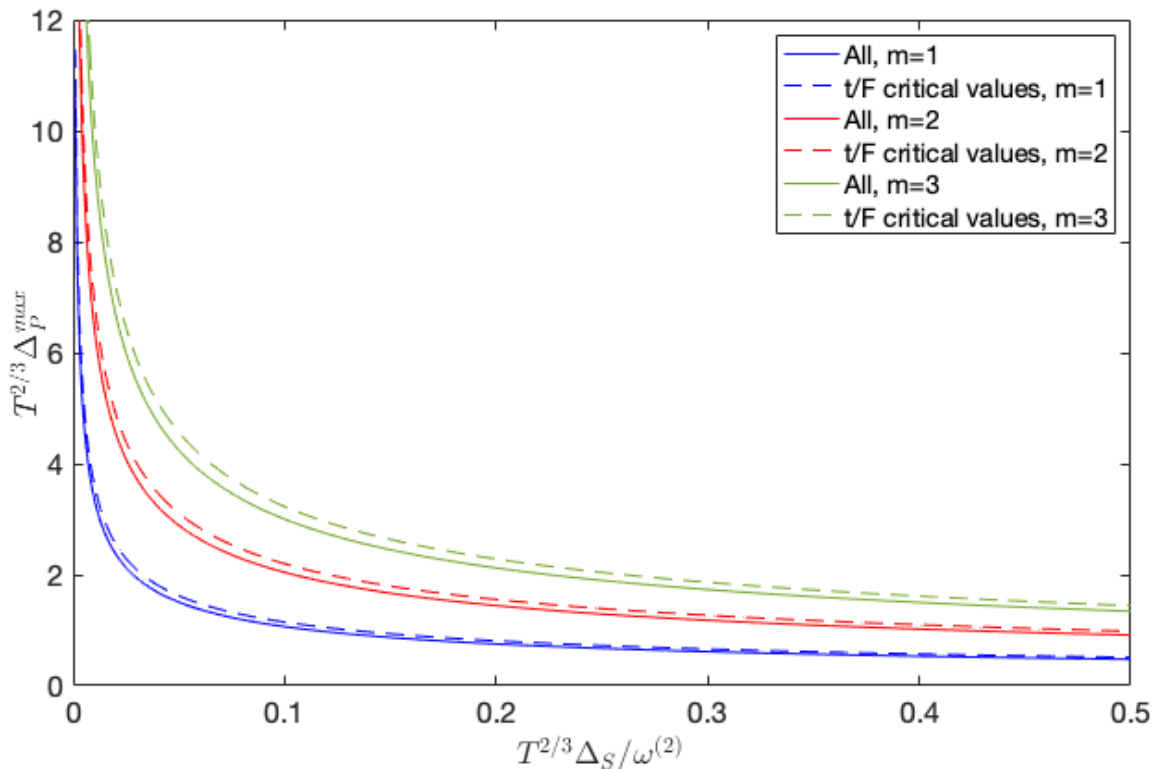


Figure 1. Higher-order frontier between the size distortion Δ_S and the maximum power loss Δ_P^{\max} of HAR tests in the Gaussian location model with dimension m , for stationary processes with average normalized spectral curvature $\omega^{(2)}$. Solid line: all kernel and orthonormal series HAR tests;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

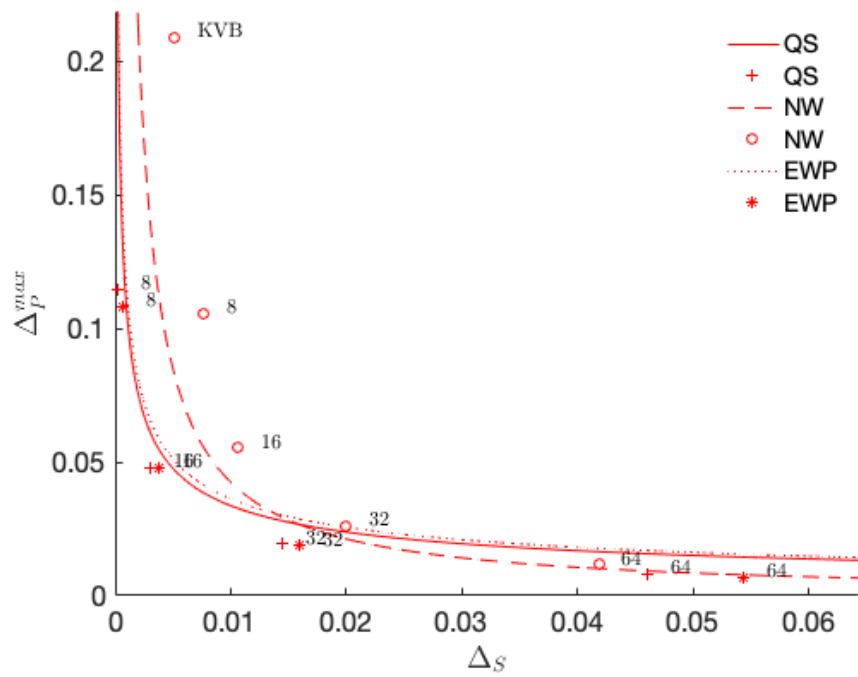


Figure 2. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Location model, $m = 1$, AR(1), $\rho = 0.5$, and $T = 200$.

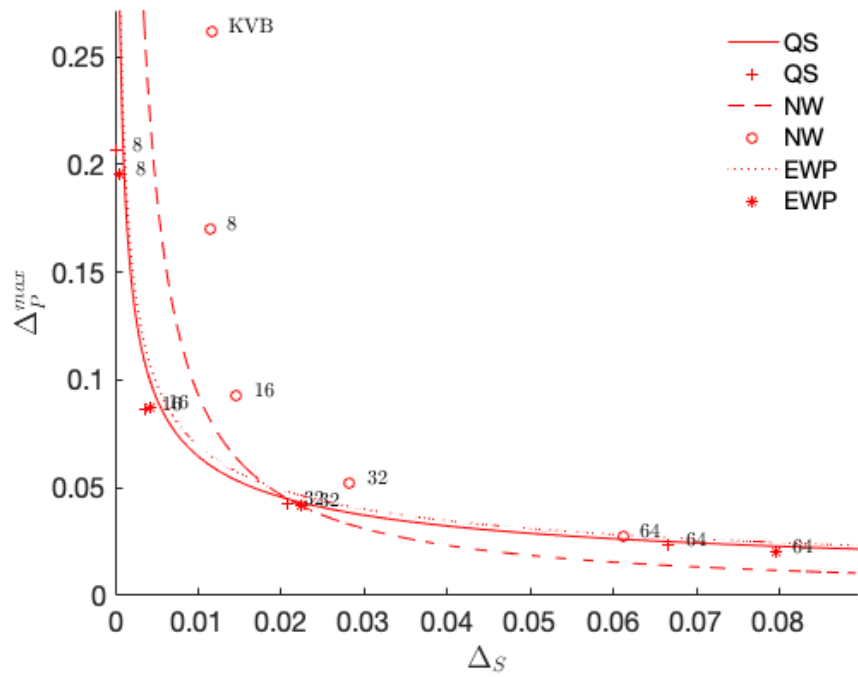


Figure 3. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Location model, $m = 2$, AR(1), $\rho = 0.5$, $T = 200$.

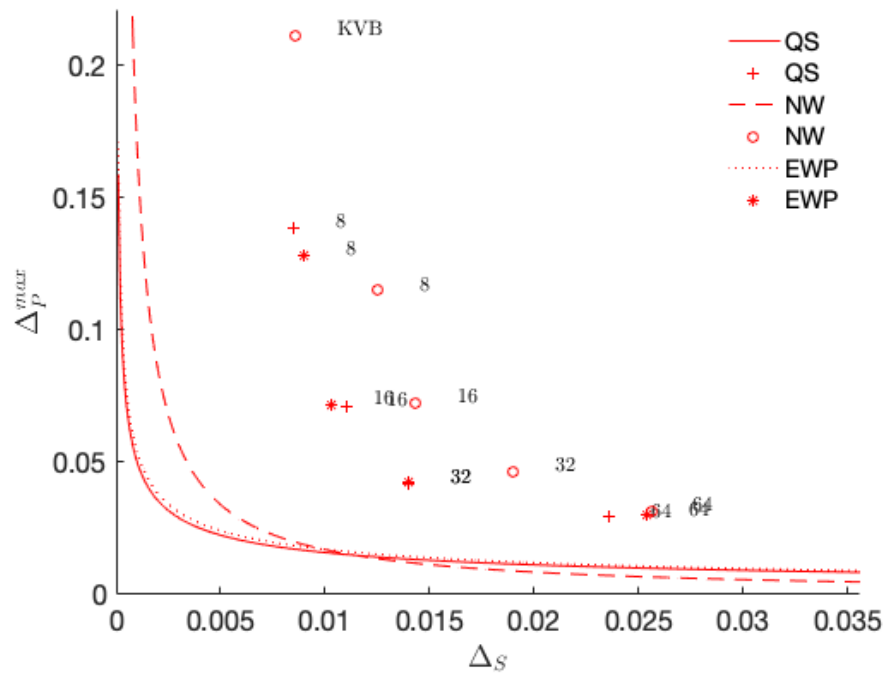


Figure 4. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Stochastic regressor, $m = 1$, $AR(1)$, $\rho = 0.5$, $T = 200$. Theoretical curves are for the Gaussian location model.