The Size-Power Tradeoff in HAR Inference

June 6, 2017

Eben Lazarus

Department of Economics, Harvard University

Daniel J. Lewis

Department of Economics, Harvard University

and

James H. Stock*

Department of Economics, Harvard University
and the National Bureau of Economic Research

*We thank Frank Diebold, Lutz Kilian, Mikkel Plagborg-Møller, Ulrich Müller, Yixiao Sun, and Mark Watson for helpful discussions and/or comments. Replication files are on Lewis's Harvard Web site.

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 orthogonal series tests using nonstandard "fixed-*b*" critical values. We also provide a frontier for the subset of these tests for which the fixed-*b* distribution is *t* or *F*. These frontiers are respectively achieved by the QS kernel and equal-weighted periodogram. The frontiers have simple closed-form expressions, which show that the price paid for restricting attention to tests with *t* and *F* critical values is small. The frontiers are derived for the 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. The truncation parameter sequence S_T is chosen to ensure consistency, and inference proceeds using 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 either by the rule $S_T = 4(T/100)^{2/9}$ suggested by Newey and West (1987), or to minimize the mean squared error (MSE) of the LRV estimator as suggested by Andrews (1991), 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 bias^2 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: increasing S increases the variance of the LRV estimator, altering the null distribution of the test. Thus solving the "bias" size distortion introduces a "variance" size distortion when chi-squared critical values are used. Fortunately, this "variance" size distortion can be addressed by replacing chi-squared critical values by Kiefer and Vogelsang's (2005) "fixed b" 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

[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.

(2008), and Sun (2014) 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 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 bandwidths 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.

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 so-called orthonormal series estimators (see for example Grenander and Rosenblatt (1957), also called orthogonal multitapers as in Brillinger (1975)). All tests are evaluated using fixed-b critical values. 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. Orthonormal series LRV estimators are an attractive family because they are psd and inference relies on standard t and F critical values. The leading example of a series LRV 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,…, B/2. In the location model, this family includes Ibragimov and Müller's (2010) subsample estimator. To align notation across the two families of tests, for orthonormal series tests we set b = 1/B (as discussed further in Section 2). 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 HAR tests using B series is t_B or in the case of tests of m restrictions is $F_{m,B-m+1}$ (after rescaling).

This paper makes six main contributions. First, using the small-b asymptotic expansions of Velasco and Robinson (2001), Sun, Phillips, and Jin (2008), and Sun (2011, 2013, 2014) for 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. These results apply when $b \to 0$ 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 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 $\Delta_P^{\max}$ 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,

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

where $\omega^{(2)}$ is the normalized curvature of the spectral density of z_t at frequency zero (the negative of the ratio of the second derivative of the spectral density to the spectral density, at frequency zero). For the m-dimensional location model, the only change to 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 b to equate the rates at which Δ_S and Δ_P converge to zero in (1) yields Δ_S, $\Delta_P = O(T^{-2/3})$, and this rate is used to derive (1) and to scale the axes in Figure 1. For the Bartlett (Newey-West, tent) kernel, equating these rates yields Δ_S, $\Delta_P = O(T^{-1/2})$, so the Barlett kernel HAR test is asymptotically dominated.

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

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

This bound is plotted for m = 1, 2, and 3 as the dashed line in Figure 1. We also show that this frontier is achieved by Brillinger's (1975) 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 propose a feasible higher-order adjustment to the fixed-b critical value which, when implemented using a consistent estimator of the normalized curvature parameter $\omega^{(2)}$, provides a higher-order improvement to the null rejection rate of the test in the location model. Because the use of the adjusted critical value does not alter the size-adjusted power, the use of the data-dependent adjusted critical value allows for tests that asymptotically improve upon the bounds (1) and (2). Notably, for tests with t or F fixed-b critical values, these adjusted critical values do not require simulation or special tables. For example, for the EWP test, the adjusted critical value is given by $\left[1+\hat{\omega}^{(2)}\left(\pi/\sqrt{6}\right)\left(B/T\right)^2\right]F_{m,B-m+1}^{\alpha}$, where $F_{m,B-m+1}^{\alpha}$ is the α-level critical value of the $F_{m,B-m+1}$ distribution.

Fifth, 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 based on Müller (2007, 2014). The performance of tests with feasible adjusted critical values is mixed at best, with clear 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.

Sixth, 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 orderings in the location model seem to hold numerically in the regression case, specifically the EWP test has a favorable size/power tradeoff, essentially the same as QS and clearly better than Bartlett and related competitors.

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, 2014). 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 provide a ranking of small-*b* performance of orthonormal series estimators and unify existing small-*b* expansions for kernel and orthonormal series expansions using what we refer to as the implied mean kernel of 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 will 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 a growing 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 a $m\times 1$ time series with autocovariances $\Gamma_j = \text{cov}(z_t, z_{t-j})$, $j = 0, 1,\ldots$ and long-run variance

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j \ . \tag{3}$$

In general z_t depends on unknown parameters, so that $z_t = z_t(\beta)$, although after preliminary definitions we suppress this dependence. We consider test statistics of the form,

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

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

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. The time series regression model with dependent variable y_t, m regressors x_t, and potentially heteroskedastic and autocorrelated error u_t is,

$$y_t = x_t'\beta + u_t,\ t = 1,\ldots, T. \tag{6}$$

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 $z_t(\beta_0) = x_t(y_t - x_t'\beta_0)$.

In the multivariate location model, y_t is $m\times 1$, β is the vector of means of y_t, and u_t is $m\times 1$:

$$y_t = \beta + u_t. \tag{7}$$

In this model, $z_t(\beta_0) = y_t - \beta_0$ and the statistics (4) and (5) test the null hypothesis that $\beta = \beta_0$.

The class of estimators $\hat{\Omega}$ considered here is comprised of two families of estimators, the family of psd kernel estimators and the family of 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. In the regression 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 in (6).

2.2 Kernel estimators

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

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

where S is the truncation parameter and the superscript "SC" denotes sum-of-covariances. Examples of kernels $k(v)$ include the Bartlett kernel used by Newey and West (1987), $k(v) = (1 - |v|)\mathbf{1}(|v|\le 1)$, and the Bartlett-Priestley-Epanechnikov quadratic-spectral (QS) kernel, $k(v) = 3[\sin(\pi x)/\pi x - \cos\pi x]/(\pi x)^2$ for $x = 6v/5$; 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:

$$\hat{\Omega}^{WP} = 2\pi\sum_{j=1}^{T-1} K_T(2\pi j/T) I_{\hat{z}\hat{z}}(2\pi j/T), \tag{9}$$

where $K_T(\omega) = T^{-1}\sum_{u=0}^{T-1} k(u/S)e^{-i\omega u}$ and where $I_{zz}(\omega)$ is the periodogram of $\hat{z}_t$ at frequency ω,

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

Kernel estimators are positive semidefinite with probability one if the frequency-domain weight function $K_T(\omega)$ 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:

$$\hat{\Omega}^{EWP} = \frac{2\pi}{B}\sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j/T) = \frac{1}{B}\sum_{j=1}^{B/2} d_{\hat{z}}(2\pi j/T)\overline{d_{\hat{z}}(2\pi j/T)}'. \tag{11}$$

2.3 Orthonormal series estimators

Series 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 $L^2[0,1]$.[2] The EWP estimator (11) is an orthonormal series estimator using the Fourier basis.

Following Sun (2013), let $\{\varphi_j(s)\}$, $j = 0, \ldots, B$, $0 \le s \le 1$, denote the first B+1 functions in an orthonormal basis for [0,1], where $\varphi_0(s) = 1$ and $\int_0^1 \phi_j(s)ds = 0$ for $j \ge 1$. Let Φ denote the $T\times B$ matrix consisting of $\{\varphi_j(s)\}$, $j = 1,\ldots, B$, evaluated at t/T:

$$\Phi = [\Phi_1 \ldots \Phi_B], \text{ where } \Phi_j = [\varphi_j(1/T)\ \ \varphi_j(2/T) \ldots \varphi_j(1)]',\ \Phi'\Phi/T = \mathrm{I}_B, \text{ and } \iota_T'\Phi = 0, \tag{12}$$

where ι_T is the T-vector of 1s. The orthonormal series LRV estimator is,

$$\hat{\Omega}^{OS} = \frac{1}{B}\sum_{j=1}^{B}\hat{\Omega}_j, \text{ where } \hat{\Omega}_j = \hat{\Lambda}_j\hat{\Lambda}_j' \text{ and } \hat{\Lambda}_j = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t. \tag{13}$$

Note that Φ and $\hat{\Omega}^{OS}$ omit the $j = 0$ function, for which $\Phi_0 = \iota = (1\ 1 \ldots 1)'$ and $\Phi_0'\Phi = 0$. By construction, $\hat{\Omega}^{OS}$ is psd with probability one.

[2] These series estimators are equivalent to estimators referred to in previous literature 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.

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 four basis functions we examine explicitly are Fourier, cosine, Legendre, and split-sample.

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

$$\{\varphi_{2j-1}(s), \varphi_{2j}(s)\} = \{\sqrt{2}\cos(2\pi js), \sqrt{2}\sin(2\pi js)\}, j = 1, \ldots, B/2. \tag{14}$$

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

$$\{\varphi_{Tj}(s)\} = \left\{\sqrt{2}\cos\left[\pi j\left(\frac{s-1/2}{T}\right)\right]\right\}, j = 1,\ldots, B. \tag{15}$$

Hwang and Sun (2015) refer to this as the shifted cosine function. A closely related alternative is Phillips' (2005) proposal of using $\{\sqrt{2}\sin[\pi j((s\text{-}1/2)/T)]\}, j = 1,\ldots, B$, which are the eigenvectors of the covariance kernel of Brownian motion. Phillips (2005) shows that, for $B \to \infty$ and $B/T \to 0$, the sine series estimator is asymptotically equivalent to $\hat{\Omega}^{EWP}$ to second order.

Legendre polynomials can be constructed as the B functions of the Gram-Schmidt orthonormalization of $\{s^j\}, j = 1,\ldots, B$ on [0,1]. Abramowitz and Stegun (1965, ch. 8) give recursions for Legendre polynomials on [-1,1], which for the purpose here are then shifted to [0,1] and renormalized.

Split-sample step function. Ibragimov and Müller (2010) proposed 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.[3] For a single coefficient, their split-sample (SS) test statistic is,

[3] 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.

$$t^{SS} = \sqrt{B+1}\left(\bar{\hat{\beta}} - \beta_0\right) \Big/ \sqrt{S_{\hat{\beta}}^2}\ , \text{ where } S_{\hat{\beta}}^2 = \frac{1}{B}\sum_{i=1}^{B+1}\left(\hat{\beta}^{(i)} - \bar{\hat{\beta}}\right)^2 , \tag{16}$$

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 (16) can be written in the standard form (4), where the LRV estimator is the series estimator,[4]

$$\hat{\Omega}^{SS} = \left(T^{-1/2}\Phi^{SS}\hat{z}\right)'\left(T^{-1/2}\Phi^{SS}\hat{z}\right)/B \text{, where } \Phi^{SS} = \sqrt{B+1}\left[I_{B+1} \otimes \iota_{T/(B+1)}\right]M_\iota^B , \tag{17}$$

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

We will refer to Φ^{SS} in (17) 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 do not span Φ^{SS}. Although the expression for $\hat{\Omega}^{SS}$ in (17) was developed for the location model with $m = 1$, it generalizes directly to $m > 1$ and to the time series regression model.[5]

---

[4] Write $\hat{\beta}^{(i)} = [T/(B+1)]^{-1}\sum_{t=[T/(B+1)](i-1)+1}^{[T/(B+1)]i} y_t$ and $\bar{\hat{\beta}} = \bar{y}$ (the full-sample mean of y_t), so $S_{\hat{\beta}}^2 = \tilde{\beta}'M_\iota\tilde{\beta}/B$ where $\tilde{\beta} = \left(\hat{\beta}^{(1)}...\hat{\beta}^{(B+1)}\right)'$. Then $\tilde{\beta}'M_\iota\tilde{\beta} = \tilde{\beta}'M_\iota^B M_\iota^{B\prime}\tilde{\beta} = y'\left(I_{B+1}\otimes\iota_{T/(B+1)}\right)M_\iota^B M_\iota^{B\prime}\left(I_{B+1}\otimes\iota_{T/(B+1)}\right)'y = \hat{z}'\left(I_{B+1}\otimes\iota_{T/(B+1)}\right)M_\iota^B M_\iota^{B\prime}\left(I_{B+1}\otimes\iota_{T/(B+1)}\right)'\hat{z}$, from which (17) follows. Note that Φ^{SS} is $T\times B$ and $\Phi^{SS\prime}\Phi^{SS}/T = I_B$ as required for series estimators.

[5] Both the Ibragimov-Müller test statistic (16) and the test based on $\hat{\Omega}^{SS}$ in (17) 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 the x_t second moment matrix. This distinction prevents giving the IM statistic in (16) an orthonormal series interpretation in the

Implied mean kernels of series estimators. Although the only series estimator with an exact kernel representation is the Fourier series/EWP estimator, the mean of every series LRV estimator has an approximate kernel representation, which becomes exact as $T \to \infty$. We refer to the limiting kernel of this mean as the implied mean kernel.

We now provide an expression for the implied mean kernel in terms of the underlying basis functions. Use the definition of $\hat{\Omega}_j$ in (13) and the device in Grenander and Rosenblatt (1957, p. 125) to express the mean of the j^{th} contribution to an orthonormal series estimator as,

$$
\begin{aligned}
E\hat{\Omega}_j &= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\phi_j(t/T)\phi_j(s/T)\Gamma_{s-t} + O(1/T) \\
&= \sum_{u=-(T-1)}^{T-1} k_{j,T}^{OS}(u/T)(1-|u/T|)\Gamma_u + O(1/T),
\end{aligned}
\tag{18}
$$

where

$$
k_{j,T}^{OS}(u/T) = \frac{1}{T-|u|}\sum_{t=1}^{T}\phi_j(t/T)\phi_j((t-u)/T)\mathbf{1}(1 \le t-u \le T). \tag{19}
$$

The second equality in (18) follows from $\Phi_j' \iota_T = 0$ (as assumed in (12)). Thus

$$
E\hat{\Omega}^{OS} = \frac{1}{B}\sum_{j=1}^{B}E\hat{\Omega}_j = \sum_{u=-(T-1)}^{T-1} k_{B,T}^{OS}(u/S)(1-|u/T|)\Gamma_u + O(1/T), \tag{20}
$$

where $k_{B,T}^{OS}(u/S) = \frac{1}{B}\sum_{j=1}^{B}k_{j,T}^{OS}\left(B^{-1}\frac{u}{S}\right)$, and $SB = T$, so that S corresponds to the usual time-domain truncation parameter. The change of variables from u/T to u/S in (20) aligns the implied

---

general regression model. In this paper we only consider the orthonormal series version of the split-sample test, that is, HAR tests using $\hat{\Omega}^{SS}$.

mean kernel with the usual expression for kernels as a function of u/S, so that (20) matches (for example) Priestley (1981, eq. (6.2.120)); also see Brillinger (1975, eq. (5.8.6)). This definition and the fact that $S = bT$ for kernel estimators motivates setting $b = 1/B$ for orthonormal series estimators.

Asymptotically, $\lim_{T\to\infty} k_{jT}^{OS} = k_j^{OS}$, so the implied mean kernel generating function is,

$$k_B^{OS}(v) = \frac{1}{B}\sum_{j=1}^{B} k_j^{OS}(B^{-1}v), \text{ where } k_j^{OS}(v) = \frac{1}{(1-|v|)}\int_{\max(0,v)}^{\min(1,1+v)} \phi_j(s)\phi_j(s-v)ds . \quad (21)$$

Note that $k_B^{OS}(0) = 1$.

Equations (20) and (21) show that the mean, and thus bias, of orthonormal series estimators have the same approximate form as kernel estimators, and that this form becomes exact as $T\to\infty$. Sections 3 and 4 below develop rejection-rate expansions and associated theoretical results under the large-B (i.e., small-b) sequence under which $B\to\infty$, and we accordingly define $k^{OS}(v) = \lim_{B\to\infty} k_B^{OS}(v)$ for orthonormal series estimators.[6] Any expressions below for $k^{OS}(\cdot)$ without the subscript B accordingly refer to these limiting implied mean kernels.

In the frequency domain, the implied mean kernels $k_{B,T}^{OS}(u/S)$ for the Fourier, cosine, Legendre, and SS basis functions all concentrate their mass on low frequencies (Supplemental Figures S.1 and S.2). Interestingly, in the time domain the Legendre implied mean kernel is indistinguishable from the SS implied mean kernel near the origin, a result shown formally below; however, the Legendre kernel places weight on more distant autocovariances, whereas the SS mean kernel truncates. In the frequency domain both the Legendre and SS implied mean kernels have considerably more leakage than the Fourier and cosine kernels.

[6] More formally, one can define $k^{OS}(\cdot)$ as the limit of $k_{B,T}^{OS}(\cdot)$ as $B,T\to\infty$ s.t. $B/T\to 0$ in the space $L^2(-\infty,\infty)$ endowed with the sup metric, 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.

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

Following Jansson (2004), Sun, Phillips, and Jin (2008), and Sun (2013, 2014), 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 estimators. The earliest fixed-b asymptotic results for kernel estimators were provided in the classical spectral density estimation literature for weighted periodogram (WP) estimators with truncated frequency-domain weights. For clarity, we begin with the scalar case. Consider $\hat{\Omega}^{WP}$ in (9), let $K_{jT} = K_T(2\pi j/T)$, and suppose that $K_{jT} = 0$ for $j > B$ and that $K_{jT} \to K_j$ as $T \to \infty$. These assumptions cover many important kernels when the frequency domain bandwidth B is fixed, including the QS, Parzen, and Daniell kernels (and thus the EWP estimator). For example, for the QS kernel, $K_j = (3\pi B/2)^{-1}(1 - (j/B)^2)\mathbf{1}(j \le B)$ and for the Daniell kernel $K_j = (2\pi \mathrm{B})^{-1}\mathbf{1}(j \le B)$. Under these conditions, the first-order asymptotic distribution of $\hat{\Omega}^{WP}$ is the weighted average of finitely many chi-squareds,

$$\hat{\Omega}^{WP} \xrightarrow{d} \left(2\pi \sum_{j=1}^{B/2} K_j \xi_j\right)\Omega \text{, where } \xi_j \text{ are i.i.d. } \chi_2^2 / 2 \text{, } j = 1,\ldots, B/2 \text{,} \tag{22}$$

see Brillinger (1981, p.145) and Priestley (1981, p. 466).

Kiefer and Vogelsang (2005) obtained the fixed-b asymptotic distribution for kernel estimators by working directly with the time-domain representation. They showed that,

$$\hat{\Omega}^{SC} \xrightarrow{d} \left(\int_0^1\int_0^1 k\left(\frac{r-s}{b}\right) dV(r)dV(s)\right)\Omega \text{,} \tag{23}$$

where V is a Brownian bridge. Sun (2014) applies Mercer's theorem to (23) to represent the limiting distribution of $\hat{\Omega}^{SC}$ as a weighted average of infinitely many independent chi-squared random variables; his representation reduces to (22) for frequency-domain kernels for which K_j =

0, $j > B/2$. This weighted average has a χ_B^2 distribution if and only if $B/2$ of the weights equal 1 and the rest are zero. Thus the EWP/Daniell estimator is unique among WP and SC estimators in having a fixed-b asymptotic distribution that is $(\chi_B^2 / B)\Omega$.

To simplify computing fixed-b critical values for (22), Tukey (1949) proposed approximating the distribution in (22) by a chi-squared, specifically,

$$\hat{\Omega}^{WP} \sim \left(\chi_\nu^2 / \nu\right)\Omega, \text{ where } \nu = \left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}, \tag{24}$$

where ν is the "equivalent degrees of freedom" of $\hat{\Omega}^{WP}$. This approximation is exact only for the EWP/Daniell estimator, for which $\nu = B$, $\int_{-\infty}^{\infty} k^2(x)dx = 1$, and $b = 1/B$.

Orthonormal series estimators. If $|\varphi_j'|$ is bounded, then a standard central limit theorem for stationary processes shows that $T^{-1/2}\sum_{t=1}^{T}\phi_j(t/T)z_t \xrightarrow{d} N(0,\Omega)$, e.g. Sun (2013). For B fixed, it follows that $\hat{\Omega}^{OS}$ in (13) has the fixed B asymptotic distribution,

$$\hat{\Omega}^{OS} \xrightarrow{d} \left(\frac{1}{B}\sum_{j=1}^{B}\zeta_j\right)\Omega \sim \left(\chi_B^2 / B\right)\Omega, \text{ where } \zeta_j \text{ are i.i.d. } \chi_1^2. \tag{25}$$

For orthonormal series estimators, Tukey's approximation (24) to the asymptotic distribution holds exactly with $\nu = B$.

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 $\sqrt{T}\left(\hat{\beta}-\beta\right)$. In general, this distribution is nonstandard and requires tabulated critical values.

For LRV estimators with the chi-squared fixed-b asymptotic distribution in (25), the fixed-b asymptotic distribution is t_B. This result seems 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. Thus, among the family of psd kernel and orthonormal series HAR tests considered in this paper, orthonormal series tests uniquely have exact asymptotic t_v fixed-b distributions.

Multivariate extension. For general $p \geq 1$, $\hat{\Omega}^{OS}$ has the fixed B asymptotic distribution,

$$\hat{\Omega}^{OS} \xrightarrow{d} \Omega^{1/2}\Xi\,\Omega^{1/2\prime}, \text{ where } \Xi \sim W_p(\mathrm{I},B), \tag{26}$$

where $W_p(\mathrm{I},B)$ denotes the standard Wishart distribution with dimension p and degrees of freedom B. Thus m times F_T in (5) will have 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 test,

$$F_T^* = \frac{B-m+1}{B} F_T . \tag{27}$$

where F_T is given in (5). When F_T is evaluated using $\hat{\Omega}^{OS}$, $F_T^* \xrightarrow{d} F_{m,B-m+1}$.

3.2 Small-*b* Rejection Rate Expansions

Velasco and Robinson (2001), Jansson (2004), Sun, Phillips, and Jin (2008), and Sun (2014) (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 $b \to 0$, $T \to \infty$, and $S_T = bT \to \infty$. Sun (2013) provides small-b rejection rate expansions for orthonormal series HAR tests.

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 orthonormal series tests depend on the Parzen characteristic exponent of the implied mean kernel. The Parzen characteristic exponent q is the maximum integer such that

$$k^{(q)}(0) = \lim_{x\to 0} \frac{1-k(x)}{|x|^q} < \infty \ . \tag{28}$$

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

$$\omega^{(q)} = tr\left(m^{-1} \sum_{j=-\infty}^{\infty} | j |^{q} \, \Gamma_j \Omega^{-1} \right). \tag{29}$$

Without Ω^{-1}, (29) 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 $q = 2$, $\omega^{(2)}$ is the negative of the ratio of the second derivative of the spectral density of $z_t = x_{jt}u_t$ at frequency zero to the value of the spectral density of z_t at frequency zero. If z_t follows a stationary AR(1) process with autoregressive coefficient ρ, then $\omega^{(2)} = 2\rho / (1-\rho)^2$.

Sun's (2013, 2014) 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 (27), 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 orthonormal series estimators $b = 1/B$. The asymptotic expansion of the null rejection rate is

$$\Pr_0\left[F_T^* > c_m^{\alpha}(b) \right] = \alpha + G_m'(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right), \tag{30}$$

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-α quantile of G_m. As discussed in Sun (2013, 2014) and in the proof of Theorem 1 in the Appendix, the term in $(bT)^{-q}$ in (30) arises from the bias of the LRV estimator.

Under the local alternative $\delta = T^{1/2}\Omega^{-1/2}\Sigma_{XX}\beta$, the rejection rate using the fixed-b critical value has the expansion,

$$\Pr_{\delta}\left[F_T^* > c_m^{\alpha}(b)\right] = [1 - G_{m,\delta^2}(\chi_m^{\alpha})] + G'_{m,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\omega^{(q)}k^{(q)}(0)(bT)^{-q}$$
$$-\frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\nu^{-1} + o(b) + o\left((bT)^{-q}\right), \tag{31}$$

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 $\nu = \left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}$ for kernel estimators and $\nu = 1/b$ for orthonormal series estimators.[7] This expression depends on both the bias of the LRV estimator, as reflected in the second term on the right hand side of (31), 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.

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 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 restate assumptions in Sun (2013, 2014) and derive from and extend those in Sun, Phillips, and Jin (2008) and Velasco and Robinson (2001).

Throughout, we refer to the $O((bT)^{-q})$ term in (30) as the higher-order size distortion, and to this term plus α as the higher-order size. For kernel estimators, k refers to the kernel. For

[7] As noted below, these expressions hold in the context of the Gaussian location model. In a slightly more general Gaussian GMM setting, equation (31) would include a term in $O(\log T/\sqrt{T})$, but as in Sun (2014) this term would not depend on b and can therefore be ignored for our purposes.

orthonormal series estimators, k refers to the implied mean kernel, for which $b = 1/B$. For both families of tests, q is the Parzen characteristic exponent of the kernel or implied mean kernel.

Assumption 1 (stochastic processes).

(a) The spectral density of z_t, denoted $s_z(\omega)$ at frequency ω, is twice continuously differentiable, and $0 < s_z(\omega) < \infty$ in a neighborhood around $\omega = 0$.

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

(c) A functional central limit theorem holds for z_t: $T^{-1/2}\sum_{t=1}^{[T\lambda]} z_t \xrightarrow{d} \Omega^{1/2} W_p(\lambda)$, where $[\cdot]$ is the greatest lesser integer function and W_p is a p-dimensional standard Brownian motion on the unit interval.

(d) z_t is a stationary Gaussian process, and it is generated according to the multivariate location model (7).

Assumption 2 (kernels). The kernels $k(v): \mathbb{R} \to [-1,1]$ are piecewise smooth, satisfy $k(v) = k(-v)$, $k(0) = 1$, $\int_{-\infty}^{\infty} |v|\, k(v) dv < \infty$, and have Parzen characteristic exponent $q \geq 1$.

Assumption 3 (orthonormal series). The orthonormal series $\{\varphi_j\} \in L^2[0,1]$ have three continuous derivatives, $j = 1, \ldots, B$, such that the n^{th} derivative $\phi_j^{(n)}(x)$ is $O\left(j^n\right)$, and their limiting implied mean kernels satisfy Assumption 2.

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

Assumptions 1(a)-(c) provide conditions under which the bias expressions and fixed-b distributions hold. Assumption 1(a) (or (b)) further implies that $\omega^{(q)}$ in (29) is finite for $q \leq 2$.

Assumption 1(d) assumes the multivariate Gaussian location model.

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. The further assumption that the basis functions' implied mean kernels meet Assumption 2 is not restrictive, as it follows directly from the form of the implied mean kernel in (21).

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 $b \to 0$ and $bT \to \infty$ (i.e., $b + (bT)^{-1} \to 0$), which are implied by Assumption 4. The more restrictive rate condition in Assumption 4 is used to express the expansions for 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 1–4,

(i) For orthonormal series estimators, the scaled asymptotic bias of the LRV estimate is,

$$\left(E\hat{\Omega}^{OS} - \Omega\right)\Omega^{-1} = -\left(\frac{B}{T}\right)^q k^{OS(q)}(0)\sum_{j=-\infty}^{\infty}|j|^q \Gamma_j \Omega^{-1} + o\left(\left(\frac{B}{T}\right)^q\right), \tag{32}$$

with $k^{OS(1)}(0) = \lim_{B\to\infty}\frac{1}{B^2}\sum_{j=1}^{B}\frac{\phi_j(0)^2 + \phi_j(1)^2}{2}$ and $k^{OS(2)}(0) = -\lim_{B\to\infty}\frac{1}{B^3}\sum_{j=1}^{B}\frac{\int_0^1 \phi_j(s)\phi_j''(s)ds}{2}$.

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

(ii) For both psd kernel and orthonormal series HAR tests, the small-b asymptotic expansions (30) and (31) apply. These expansions also hold for the SS series estimator although it does not satisfy Assumption 3.

Theorem 2. Let $c_{m,T}^{\alpha}(b)$ be the size-adjusted fixed-*b* critical value, that is, the critical value such that $\Pr_0\left[F_T^* > c_{m,T}^{\alpha}(b)\right] = \alpha + o(b) + o\left((bT)^{-q}\right)$, and assume that Assumptions 1-4 hold. Then for a test that is either a psd kernel HAR test or an orthonormal series HAR test,

$$c_{m,T}^{\alpha}(b) = \left[1 + \omega^{(q)} k^{(q)}(0)(bT)^{-q}\right] c_m^{\alpha}(b), \tag{33}$$

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

$$\Pr_\delta\left[F_T^* > c_{m,T}^{\alpha}(b)\right] = \left[1 - G_{m,\delta^2}(\chi_m^{\alpha})\right] - \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\nu^{-1}$$
$$+ o(b) + o\left((bT)^{-q}\right). \tag{34}$$

Theorem 3. Consider two tests 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 $c1_m^{\alpha}(b_1)$ and $c2_m^{\alpha}(b_2)$. 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 δ is,

$$\Pr_\delta[F_{1T}^* > c1_m^{\alpha}(b_1)] - \Pr_\delta[F_{2T}^* > c2_m^{\alpha}(b_s)] = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^{\alpha})\chi_m^{\alpha}\left(\nu_2^{-1} - \nu_1^{-1}\right)$$
$$+ o(b_1) + o((b_1T)^{-q}) + o(b_2) + o((b_2T)^{-q}). \tag{35}$$

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

$$\Delta_S = \Pr_0\left[F_T^* > c_m^{\alpha}(b)\right] - \alpha. \tag{36}$$

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 δ, and let $\Delta_P^{\max}$ denote the maximum such power loss, so that $\Delta_P^{\max}$ is the maximum gap between the power curves of the oracle test and the candidate test. Then,

$$\Delta_P(\delta) = \left[1-G_{m,\delta^2}(\chi_m^\alpha)\right] - \Pr_\delta\left[F_T^* > c_{m,T}^\alpha(b)\right], \text{ and} \tag{37}$$

$$\Delta_P^{\max} = \sup_\delta \Delta_p(\delta). \tag{38}$$

Because v can be expressed in terms of b for both families of tests, equations (30) and (34) constitute a pair of parametric equations that determine Δ_S and Δ_P for a given sequence b. Both expressions are monotonic in b so the sequence b can be eliminated to obtain expressions for the higher-order tradeoff between the size and power of a given test. The manner in which b enters into 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 $b^{\frac{q+1}{q}}T \to C$ 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 δ is,

$$T\Delta_p(\delta)\left|\Delta_S\right|^{1/q} = a_{m,\alpha,q}(\delta)\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty}k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q} + o(1) \tag{39}$$

for kernel tests, and

$$T\Delta_p(\delta)\left|\Delta_S\right|^{1/q} = a_{m,\alpha,q}(\delta)\left(k^{(q)}(0)\right)^{1/q}\left|\omega^{(q)}\right|^{1/q} + o(1) \tag{40}$$

for orthonormal series tests, where in both cases $a_{m,\alpha,q}(\delta)$ = $\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q}$, and k is the kernel or implied mean kernel.

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

$$T\Delta_p^{\max}\left|\Delta_S\right|^{1/q} = \bar{a}_{m,\alpha,q}\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty}k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q} + o(1) \tag{41}$$

for kernel tests, and

$$T\Delta_p^{\max}\left|\Delta_S\right|^{1/q} = \bar{a}_{m,\alpha,q}\left(k^{(q)}(0)\right)^{1/q}\left|\omega^{(q)}\right|^{1/q} + o(1) \tag{42}$$

for orthonormal series tests, where in both cases $\bar{a}_{m,\alpha,q}$ = $\sup_\delta a_{m,\alpha,q}(\delta)$.

(iii) The size/power tradeoffs of tests based on LRV estimators with Parzen characteristic exponent $q = 2$ 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 orthonormal series HAR tests, under the sequence for b in Corollary 1,

$$T\Delta_P^{\max}\sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{3\pi\sqrt{10}}{25}\bar{a}_{m,\alpha,2} + o(1) \tag{43}$$

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

(ii) For psd kernel and orthonormal series HAR tests with exact *t*- and *F*- asymptotic fixed-*b* distributions, under the sequence for *b* in Corollary 1,

$$T\Delta_P^{\max}\sqrt{\frac{\Delta_S}{\omega^{(2)}}} \geq \frac{\pi}{\sqrt{6}}\overline{a}_{m,\alpha,2} + o(1) \text{ (exact } t \text{ or } F \text{ critical values).} \tag{44}$$

This frontier is achieved by the EWP test. For $\alpha = .05$, $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 0.3623$ for $m = 1$, $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 0.6950$ for $m = 2$, and $\overline{a}_{m,\alpha,2}\pi/\sqrt{6} \approx 1.0211$ for $m = 3$.

4.3 Remarks

1. For a given α and *m*, the frontier depends only on the sample size and the average 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 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. In practice, the quality of this approximation to the frontier presumably depends on the properties of the stochastic process and on the sample size, and this quality is explored in the Monte Carlo analysis in the next section.

2. The sequence given in Corollary 1, $b \sim CT^{-\frac{q}{q+1}}$, is of the same order as the sequence found in Sun, Phillips, and Jin (2008) and Sun (2014) 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 (43), note that the frontier for kernel tests is achieved by the $q = 2$ HAR test for which the term $\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx$ in (41) is minimized. This minimization problem turns out to be the same problem that gives rise to the QS/Bartlett-Priestley-Epanechnikov kernel (Priestley (1981, pp. 569-70)), for which $\sqrt{k^{(2)}(0)}\int_{-\infty}^{\infty} k^2(x)dx = 3\pi\sqrt{10}/25$. This frontier then can be shown to dominate the frontier for orthonormal series tests (see Remark 5 below), so that it is in fact the overall frontier across both families.
5. The proof for the restricted frontier in part (b) entails expressing a candidate set of $q = 2$ orthogonal series in terms of Fourier coefficients, computing their implied value of $\sqrt{k^{(2)}(0)}$, and concluding that it must be larger than the orthogonal series that places all weight on the first $B/2$ Fourier terms. 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 (35), 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,

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

$$= \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi^\alpha_m)\chi^\alpha_m\left(1 - \frac{6\sqrt{3}}{5\sqrt{5}}\right)B^{-1}, \tag{45}$$

where $v_{EWP} = B$ and the final expression is derived in the Appendix. Note that (45) 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, (45) 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. 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 (45) as a function of δ for various values of B and m = 1.

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 $T \to \infty$ to obtain (21), 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 orthogonal series on the bias of the LRV estimator by calculating the generalized derivatives given in those expressions.
    a. ***q = 1 implied mean kernels.*** It is shown in the Appendix that: the Legendre basis function has q = 1; by a direct calculation, the SS-basis also has q = 1; and, surprisingly, the generalized first derivative at the origin of the two bases is the same for small b (large B): $k^{Leg(1)}(0) = k^{SS(1)}(0) = 1$. Thus for small b, $k^{Leg(1)}(0) = k^{SS(1)}(0) = k^{NW(1)}(0)$. Because Legendre, SS, and Bartlett are all q = 1 kernels or implied mean kernels, their size distortion/size-adjusted power tradeoffs are given by (39)-(40) with q = 1. Assessing the second term in those tradeoffs, for Legendre and SS, $k^{(1)}(0) = 1$ while for Bartlett, $k^{(1)}(0)\int_{-\infty}^{\infty} k^2(x)dx = 2/3$. Thus NW dominates Legendre and SS: the NW small-b tradeoff curve is strictly below the Legendre and SS tradeoff curves. Surprisingly, the Legendre and SS share a common tradeoff curve.

b. ***q* = 2 implied mean kernels.** The Fourier and 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 cosine than for the Fourier basis, as it can be seen in the Appendix that for finite B, $k_B^{EWP(2)}(0) = \frac{\pi^2}{6}\frac{(B+1)(B+2)}{B^2}$ while $k_B^{cos(2)}(0) = \frac{\pi^2}{6}\frac{(B+1)(B+1/2)}{B^2}$. This suggests that for small B, the cosine basis might slightly outperform the Fourier basis.

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 orthonormal series estimator (some of those results apply only to $q = 2$ or to specific kernels). Our result unifies the asymptotic bias for kernel and orthonormal series 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. An implication of Theorem 2 and Assumption 4 is that use of the infeasible adjusted critical value (33) provides a higher-order improvement to the null rejection rate of the test, so that the null rejection rate is $\Pr_0\left[F_T^* > c_{m,T}^{\alpha}(b)\right] - \alpha = o((bT)^{-2}) + o(b) = o(T^{-2/3})$ under the Corollary 1 sequence. This improvement is a refinement over the fixed-b critical value. It is also an improvement over the results provided by Gonçalves and Vogelsang (2011) for the moving block bootstrap applied to a HAR test evaluated using the asymptotic Gaussian/chi-squared critical values, which they showed to provide a critical value equivalent to the fixed-b critical value. The improvement from using $c_{m,T}^{\alpha}(b)$ appears not to be as good as the refinement from using Zhang and Shao's (2013) Gaussian dependent bootstrap, however their Gaussian bootstrap entails generation of bootstrap samples from a consistently estimated autocovariance function whereas the

adjusted critical values $c^{\alpha}_{m,T}(b)$ depend on the process solely through $\omega^{(2)}$. In addition, Zhang and Shao's (2013) expansion is around the fixed-*b* limiting distribution, not around its small-*b* chi-squared limit, so their results are not directly comparable to those here.

11. The preceding remark suggests considering a feasible version of the adjusted critical value, in which $\omega^{(2)}$ is replaced with a consistent estimator:

$$\hat{c}^{\alpha}_{m,T}(b) = \left[1+\hat{\omega}^{(q)}k^{(q)}(0)(bT)^{-q}\right]c^{\alpha}_{m}(b). \tag{46}$$

For the EWP estimator, the feasible higher-order adjusted critical value is

$$\hat{c}^{\alpha}_{m,T}(b) = \left[1+\frac{\pi^2}{6}\left(\frac{B}{T}\right)^2\hat{\omega}^{(2)}\right]F^{\alpha}_{m,B-m+1} \qquad \text{(EWP)}, \tag{47}$$

where $F^{\alpha}_{m,B-m}$ is the α-level critical value for the $F_{m,B\text{-}m}$ distribution. The results in Theorem 2 suggest that these feasible higher-order adjusted critical values provide size improvements that will place the test below the size/power frontier of Theorem 5, because they have better asymptotic size control with the same size-adjusted power.

5. Monte Carlo Analysis

The purpose of this Monte Carlo analysis is threefold. First, we assess the quality of the small-*b* approximations to the size/power tradeoffs in the Gaussian location model. Second, we examine whether the empirically adjusted critical values proposed in (46) and (47) actually provide improvements in relevant sample sizes in the Gaussian location model. Third, 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 v = 8, 16, 32, and 64. The tests are labeled accordingly, for example NW16 is the Newey-West (Bartlett) test with v = 16 equivalent degrees of freedom. As a reference, for T = 200, NW32 has a truncation parameter of $(3/2)T/v$, which rounds up to 10. For the orthonormal series estimators, $v = B$. Tests use fixed-b critical values unless explicitly stated otherwise.

We examine the following HAR tests:

1. NW: Kernel estimator with Bartlett/Newey-West kernel, $k(v) = (1-|v|)\mathbf{1}(|v|\leq 1)$
2. KVB: The Kiefer-Vogelsang-Bunzel (2000) test, which is NW with $S = T$ (so $v = 3/2$).
3. QS: $k(u) = 3[\sin(\pi x)/\pi x - \cos\pi x]/(\pi x)^2$ for $x = 6u/5$.
4. EWP: Orthonormal series estimator using Fourier basis in (11).
5. cos: Orthonormal series estimator using Type-2 cosine basis (15).
6. Legendre: Orthonormal series estimator computed using Legendre polynomials.
7. SS-basis: Split-sample orthonormal series estimator using (17).

We also consider three estimators of $\omega^{(2)}$ for use in computing empirical higher-order adjusted critical values, using the formula in (46).We only investigated these adjusted critical values in the m = 1 case so the estimators are given for scalar ω. The first estimator is the sample analog of (29): $\hat{\omega}^{(2)} = \sum_{j=1}^{S} j^2\hat{\Gamma}_{\hat{z},j}\left(\sum_{j=1}^{S}\hat{\Gamma}_{\hat{z},j}\right)^{-1}$, where $S = 10(T/200)^{1/3}$. The second estimator is obtained from the coefficients of a quadratic estimator of the spectrum fit to the first M periodogram ordinates: $\hat{\hat{\omega}}^{(2)} = -\hat{\hat{S}}_{\hat{z}}(0)^{-1}\hat{\hat{S}}''_{\hat{z}}(0)$, where $\hat{\hat{S}}_{\hat{z}}(0)$ and $\hat{\hat{S}}''_{\hat{z}}(0)$ are the estimated coefficients in the regression, $I_{\hat{z}\hat{z}}(2\pi j/T) = S_{\hat{z}}(0) + (2\pi j/T)^2 S''_{\hat{z}}(0) + error$, $j = 1,\ldots,M$, where M is chosen to be $10(T/200)^{1/3}$. The third estimator is a plug-in parametric estimator based on estimating an AR(1) using $\hat{z}_t$: $\hat{\omega}^{(2),plugin} = 2\hat{\alpha}_{\hat{z}}/(1-\hat{\alpha}_{\hat{z}})^2$, where $\hat{\alpha}_{\hat{z}}$ is the AR(1) coefficient.

In the location model, the data are generated according to (7), where u_{it}, $i = 1,\ldots, m$ 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 (6), with x_{it}, $i = 1,\ldots,m$ and u_t being independent Gaussian AR(1) processes. Under the null, β =

0. Under the local alternative, $\beta = T^{-1/2}\Sigma_{XX}^{-1}\Omega^{1/2}\delta$, where δ is the local alternative (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. All results are displayed on a finite-sample counterpart 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 (39) and (40) are shown as lines, and the Monte Carlo results are presented as scatter points.

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 $T = 200$. The results are much the same as for $m = 1$, except that the theory actually fits better for the NW kernel with $m = 2$ than with $m = 1$.

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 orthogonal 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. 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.

EWP with data-dependent critical values. Figure 5 shows the performance of the data-dependent critical value refinements proposed in (47) in the location model, using the three estimators for $\omega^{(2)}$ discussed in Section 5.1 for $m= 1$, $\rho = 0.5$ and 0.7, and $T = 200$. Analysis is limited to the EWP estimator. The plug-in estimator performs well, offering excellent size control. This is confirmation of the applicability of the theory but is not a realistic test of the estimator because the parametric plugin estimator matches the AR(1) DGP. The two nonparametrically-adjusted critical values perform worse than the fixed-b critical values in all cases, and in some cases they perform so poorly that they are (literally) off this chart (also, compare the scales in Figure 5 v. Figure 2). These results do not suggest optimism about higher-order refinements being useful. Results in the Supplement are similar for an ARMA(2,1) DGP, except that the AR(1) plug-in adjustment now joins the other adjustments in being either matched or outperformed by the unadjusted fixed-b critical values.

Overall, we can draw four 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. In particular QS and EWP estimators outperform Newey-West, whose curves are shifted outwards in the figures, and this ranking is also found with stochastic regressors. Third, the data-dependent critical value refinements considered do not in general improve upon the performance of t critical values for EWP, in fact, outside the case that the plug-in adjustment is correctly specified, the adjusted critical values produce larger Monte Carlo size distortions than the fixed-b asymptotics. Fourth, 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.

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 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 bandwidths and fixed-*b* critical values provide meaningful size improvements over tests with small bandwidths and/or asymptotically normal critical values while sacrificing little power, and suggest that these qualitative results extend outside the Gaussian location model to tests involving stochastic regressors. These results lead us to recommend using the EWP test in empirical work.

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

Brillinger, D.R. (1975), *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

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. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: John Wiley and Sons.

Hillmer, C. E. and M. J. Hilmer (2014). *Practical Econometrics: Data Collection, Analayis, and Application, 1st Edition*. New York: McGraw-Hill.

Hwang, J. and Y. Sun (2015). "Simple, Robust and Accurate *F* and *t* Tests in Cointegrated Systems," manuscript, University of California, San Diego.

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. and T.J. Vogelsang (2002), "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093-2095.

Kiefer, N. and T.J. Vogelsang (2002), "A New Asymptotic Theory for Heteroskedasticity-

Autocorrelation Robust Tests," *Econometric Theory*, 21, 2093-2095.

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.

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

Phillips, P. C. B. (2005), "HAC Estimation by Automated Regression," *Econometric Theory* 21, 116–142.

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 Reivew* 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.

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 Econometric, 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.

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. (2014), “Let’s Fix It: Fixed-*b* Asymptotics versus Small-*b* Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference,” *Journal of Econometrics* 178, 659-677.

Sun, Y. and D.M. Kaplan (2014), “Fixed-smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator,” manuscript, UC-San Diego.

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

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 orthonormal series estimator. 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 must obtain results in which the relevant limits have been taken jointly according to the sequence in Assumption 4.

First, using the same steps as for equation (18), we can write,

$$\begin{aligned} E\hat{\Omega}_j^{OS} &= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)' \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\phi_j(t/T)\phi_j(s/T)\Gamma_{s-t} + O(1/T) \\ &= \sum_{u=-(T-1)}^{T-1}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\Gamma_u + O(1/T), \end{aligned} \tag{48}$$

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)). Thus,

$$E\hat{\Omega}^{OS} - \Omega = \sum_{u=-(T-1)}^{T-1}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u - \sum_{|u|\geq T}\Gamma_u + O\left(\frac{1}{T}\right). \tag{49}$$

For the last non-approximation-error term,

$$\left|\sum_{|u|\geq T}\Gamma_u\right| \leq \sum_{|u|\geq T}\left|\Gamma_u\right| \leq \frac{1}{T^q}\sum_{|u|\geq T}|u|^q\left|\Gamma_u\right| = o\left(T^{-q}\right) = o\left((B/T)^q\right), \tag{50}$$

by Assumptions 1(b) and 4, so that we may focus on the first summation. (Here, q = 1 or 2, and a more rigorous definition for this value will be provided below following (28).)

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

$$E\hat{\Omega}^{OS}-\Omega=\sum_{u=-L_T}^{L_T}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u$$

$$+\sum_{L_T<|u|<T}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u+o\left(\left(\frac{B}{T}\right)^q\right)+O\left(\frac{1}{T}\right), \quad (51)$$

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

$$\frac{T^q}{L_T^{q+\delta}B^q}+\frac{L_T B}{T}\to 0, \quad (52)$$

where δ is as in Assumption 1(b). Similar to the steps taken in (50), we have that

$$\left|\sum_{L_T<|u|<T}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u\right|$$

$$\le C\sum_{L_T<|u|<T}\left|\Gamma_u\right|\le\frac{C}{L_T^{q+\delta}}\sum_{L_T<|u|<T}|u|^{q+\delta}\left|\Gamma_u\right|=o(L_T^{-(q+\delta)})=o\left(\left(\frac{B}{T}\right)^q\right), \quad (53)$$

for some constant *C*, by Assumptions 1(b) and 3, and where the fact that $L_T^{-(q+\delta)}=o\left((B/T)^q\right)$ follows from the sequence (52).

We may accordingly focus attention on the terms in (51) for which $|u|\le L_T$, and can thus write that equation as

$$E\hat{\Omega}^{OS}-\Omega=\sum_{u=-L_T}^{L_T}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right)\right]-1\right\}\Gamma_u+o\left(\left(\frac{B}{T}\right)^q\right)+O\left(\frac{1}{T}\right). \quad (54)$$

We now consider the value in square brackets in the first term in this equation (or in

(49)). This value is similar to, but defined slightly differently than, the implied mean kernel in (20); accordingly, define

$$\tilde{k}_{B,T}^{OS}\left(\frac{u}{T}\right)=\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j\left(\frac{t}{T}-\frac{u}{T}\right). \tag{55}$$

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

$$\begin{aligned}
\tilde{k}_{B,T}^{OS}\left(\frac{u}{T}\right)-1&=\frac{1}{B}\sum_{j=1}^{B}\left\{\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\left[\phi_j\left(\frac{t}{T}\right)-\frac{u}{T}\phi_j'\left(\frac{t}{T}\right)+\frac{1}{2}\left(\frac{u}{T}\right)^2\phi_j''\left(\frac{t}{T}\right)\right.\right.\\
&\qquad\left.\left.-\frac{1}{6}\left(\frac{u}{T}\right)^3\phi_j'''\left(\frac{h_{t,u}}{T}\right)\right]\right\}-1\\
&=\frac{1}{B}\sum_{j=1}^{B}\left\{\frac{1}{T}\sum_{t=1}^{T}\phi_j\left(\frac{t}{T}\right)^2-1-\frac{1}{T}\sum_{1\le t\le u|T+u\le t\le T}\phi_j\left(\frac{t}{T}\right)^2-\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}\right.\\
&\qquad\left.+\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\frac{1}{2}\phi_j\left(\frac{t}{T}\right)\phi_j''\left(\frac{t}{T}\right)\left(\frac{u}{T}\right)^2-\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\frac{1}{6}\phi_j\left(\frac{t}{T}\right)\phi_j'''\left(\frac{h_{t,u}}{T}\right)\left(\frac{u}{T}\right)^3\right\}\\
&=-\frac{1}{B}\sum_{j=1}^{B}\left\{\frac{1}{T}\sum_{1\le t\le u|T+u\le t\le T}\phi_j\left(\frac{t}{T}\right)^2+\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}\right.\\
&\qquad\left.-\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\frac{1}{2}\phi_j\left(\frac{t}{T}\right)\phi_j''\left(\frac{t}{T}\right)\left(\frac{u}{T}\right)^2+\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\frac{1}{6}\phi_j\left(\frac{t}{T}\right)\phi_j'''\left(\frac{h_{t,u}}{T}\right)\left(\frac{u}{T}\right)^3\right\}\\
&\qquad+O(1/T),
\end{aligned}\tag{56}$$

where $\sum_{1\le t\le u|T+u\le t\le T}\phi_j\left(\frac{t}{T}\right)^2$ refers to the sum over either the indices $1\le t\le u$ (if $u>0$) or the indices $T+u\le t\le T$ (if $u<0$).

Note that for the last term in (56),

$$\left| \sum_{u=-L_T}^{L_T} \left\{ \frac{1}{B} \sum_{j=1}^{B} \frac{1}{T} \sum_{t=\max(1,u)}^{\min(T,T+u)} \frac{1}{6} \phi_j\left(\frac{t}{T}\right) \phi_j'''\left(\frac{h_{t,u}}{T}\right) \left(\frac{u}{T}\right)^3 \right\} \Gamma_u \right| \leq C L_T O\left(\left(\frac{B}{T}\right)^3 \right) \sum_{u=-L_T}^{L_T} |u|^2 \Gamma_u$$

$$= O\left(\frac{L_T B^{3-q}}{T^{3-q}} \right) O\left(\left(\frac{B}{T}\right)^q \right) = o\left(\left(\frac{B}{T}\right)^q \right), \quad (57)$$

for some constant C, where the first inequality uses that $\phi_j'''\left(\frac{h_{t,u}}{T}\right) = O(B^3)$ by Assumption 3 and the second line uses Assumption 1(b) and (52). We accordingly do not consider this term in (56) when evaluating (54), given that this introduces an error only of order $o\left((B/T)^q\right)$.

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 (56), for any value of u such that $|u| \leq L_T$, we have,

$$\begin{aligned}
\frac{1}{T} \sum_{1 \leq t \leq u | T+u \leq t \leq T} \phi_j\left(\frac{t}{T}\right)^2 &= \frac{1}{T} \sum_{t=1}^{u} \phi_j\left(\frac{t}{T}\right)^2 1\{u > 0\} + \frac{1}{T} \sum_{t=T+u}^{T} \phi_j\left(\frac{t}{T}\right)^2 1\{u < 0\} \\
&= \frac{1}{T} \sum_{t=1}^{u} \phi_j(0)^2 \left\{ 1 + O\left(\frac{u}{T}\right) \right\} 1\{u > 0\} + \frac{1}{T} \sum_{t=T+u}^{T} \phi_j(1)^2 \left\{ 1 + O\left(\frac{-u}{T}\right) \right\} 1\{u < 0\} \\
&= \left[\frac{u}{T} \phi_j(0)^2 + \frac{1}{T} O\left(\frac{u^2}{T}\right) \right] 1\{u > 0\} + \left[\frac{-u}{T} \phi_j(1)^2 + \frac{1}{T} O\left(\frac{u^2}{T}\right) \right] 1\{u < 0\} \\
&= \frac{u}{T} \phi_j(0)^2 1\{u > 0\} + \frac{-u}{T} \phi_j(1)^2 1\{u < 0\} + o\left(\frac{1}{T}\right), \qquad (58)
\end{aligned}$$

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 $|u| \leq L_T = o\left(T^{1/2}\right)$, by (52) and Assumption 4.

We now consider the second term in (56). Assume for now that $u > 0$. We first note that,

$$\frac{1}{T}\sum_{t=1}^{T}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}-\frac{1}{T}\sum_{t=u}^{T}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}=\frac{1}{T}\sum_{t=1}^{u-1}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}, \tag{59}$$

and further that,

$$\begin{aligned}\left|\sum_{u=1}^{L_T}\left\{\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=0}^{u-1}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}\right]-1\right\}\Gamma_u\right| &\le \sum_{u=1}^{L_T}\left|\left[\frac{1}{B}\sum_{j=1}^{B}\frac{1}{T}\sum_{t=0}^{u-1}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}\right]-1\right|\left|\Gamma_u\right| \\ &\le \frac{CO(B/T)}{T}\sum_{u=1}^{L_T}|u|^2\left|\Gamma_u\right| = o(1/T), \end{aligned} \tag{60}$$

where the second line uses that $\left|\frac{1}{T}\sum_{t=0}^{u-1}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\right| \le \left|\frac{1}{T}\sum_{t=0}^{u-1}C\times O(B)\right| \le u\times\left|C\times O(B)\right|$ for some constant C by Assumption 3, and further uses Assumptions 1(b) and 4. The same logic applies for $u < 0$. We need not worry about $u = 0$, since this value is fixed; these calculations are simply meant to show that approximating the sum $\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}$ by $\frac{1}{T}\sum_{t=1}^{T}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}$ introduces only a small approximation error in (54) despite having $\max u \to \infty$.

We thus deal directly with $\frac{1}{T}\sum_{t=1}^{T}\phi_j\left(\frac{t}{T}\right)\phi_j'\left(\frac{t}{T}\right)\frac{u}{T}$. It can be seen that uniformly in $|u| \le L_T$, this value can be approximated by Euler summation (as in Phillips (2005), Lemma A) as,

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

$$=\frac{u}{T}\int_0^1 \phi_j(s)\phi_j'(s)\,ds+O(1/T), \tag{61}$$

where [·] is the greatest lesser integer function, and the approximations in the second and third equalities hold by Assumption 3 and (52).

For the third term in (56), by similar steps as in (59)–(60), we can approximate the sum $\frac{1}{T}\sum_{t=\max(1,u)}^{\min(T,T+u)}\frac{1}{2}\phi_j\left(\frac{t}{T}\right)\phi_j''\left(\frac{t}{T}\right)\left(\frac{u}{T}\right)^2$ by $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\phi_j\left(\frac{t}{T}\right)\phi_j''\left(\frac{t}{T}\right)\left(\frac{u}{T}\right)^2$. Further, as in (61), we can approximate this latter value with the integral $\frac{1}{2}\left(\frac{u}{T}\right)^2\int_0^1 \phi_j(s)\phi_j''(s)\,ds$.

Combining all the results above, we can thus write (54) as,

$$E\hat{\Omega}^{OS}-\Omega=\sum_{u=-L_T}^{L_T}-\frac{1}{B}\sum_{j=1}^{B}\left\{\left[\phi_j(0)^2 1\{u>0\}+\phi_j(1)^2 1\{u<0\}\right]\left|\frac{u}{T}\right|\right.$$
$$\left.+\frac{u}{T}\int_0^1\phi_j(s)\phi_j'(s)\,ds-\frac{1}{2}\left(\frac{u}{T}\right)^2\int_0^1\phi_j(s)\phi_j''(s)\,ds\right\}\Gamma_u+o\left(\left(\frac{B}{T}\right)^q\right)+O\left(\frac{1}{T}\right). \tag{62}$$

Integrating the second term by parts,

$$\int_0^1\phi_j(s)\phi_j'(s)ds=\phi_j(1)^2-\phi_j(0)^2-\int_0^1\phi_j(s)\phi_j'(s)ds=\frac{1}{2}\left[\phi_j(1)^2-\phi_j(0)^2\right]. \tag{63}$$

We can use this to write the first two terms in (62) as follows, for each value j:

$$\sum_{u=-L_T}^{L_T}\left\{\left[\phi_j(0)^2 1\{u>0\}+\phi_j(1)^2 1\{u<0\}\right]\left|\frac{u}{T}\right|+\frac{u}{T}\int_0^1\phi_j(s)\phi_j'(s)\,ds\right\}\Gamma_u=$$
$$=\sum_{u=-L_T}^{L_T}\left\{\left[\phi_j(0)^2 1\{u>0\}+\phi_j(1)^2 1\{u<0\}\right]\left|\frac{u}{T}\right|+\frac{1}{2}\frac{u}{T}\left[\phi_j(1)^2-\phi_j(0)^2\right]\right\}\Gamma_u$$

$$= \sum_{u=1}^{L_T} \frac{1}{2}\frac{u}{T}\left[\phi_j(0)^2 + \phi_j(1)^2\right]\Gamma_u + \sum_{u=-L_T}^{-1} \frac{1}{2}\frac{-u}{T}\left[\phi_j(0)^2 + \phi_j(1)^2\right]\Gamma'_{-u}$$

$$= \sum_{u=1}^{L_T} \frac{1}{2}\frac{u}{T}\left[\phi_j(0)^2 + \phi_j(1)^2\right]\left(\Gamma_u + \Gamma'_u\right)$$

$$= \sum_{u=-L_T}^{L_T} \frac{1}{2}\left|\frac{u}{T}\right|\left[\phi_j(0)^2 + \phi_j(1)^2\right]\Gamma_u. \tag{64}$$

Thus (62) becomes,

$$E\hat{\Omega}^{OS} - \Omega = -\sum_{u=-L_T}^{L_T} \frac{1}{B}\sum_{j=1}^{B}\left\{\frac{1}{2}\left|\frac{u}{T}\right|\left[\phi_j(0)^2 + \phi_j(1)^2\right] - \frac{1}{2}\left(\frac{u}{T}\right)^2 \int_0^1 \phi_j(s)\phi_j''(s)ds\right\}\Gamma_u + o\left(\left(\frac{B}{T}\right)^q\right), \tag{65}$$

which further uses that $O(1/T) = o\left((B/T)^q\right)$ from (52). Thus,

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

$$= \begin{cases} -\left(\frac{B}{T}\right)\left(\frac{1}{B^2}\sum_{j=1}^{B}\frac{\phi_j(0)^2 + \phi_j(1)^2}{2}\right)\sum_{u=-\infty}^{\infty}|u|\Gamma_u + o\left(\left(\frac{B}{T}\right)\right), \text{ if } \lim_{B\to\infty}\frac{1}{B^2}\sum_{j=1}^{B}\frac{\phi_j(0)^2 + \phi_j(1)^2}{2} \neq 0 \\ -\left(\frac{B}{T}\right)^2\left(-\frac{1}{B^3}\sum_{j=1}^{B}\frac{\int_0^1 \phi_j(s)\phi_j''(s)ds}{2}\right)\sum_{u=-\infty}^{\infty}|u|^2\Gamma_u + o\left(\left(\frac{B}{T}\right)^2\right), \text{ otherwise.} \end{cases} \tag{66}$$

Equation (32) in the theorem follows from these expressions. These expressions and their relation to equations (54)–(55), along with the definition of the generalized derivative in (28), make clear that in the limit, we have $\tilde{k}^{OS(1)}(0) = \lim_{B\to\infty}\frac{1}{B^2}\sum_{j=1}^{B}\frac{\phi_j(0)^2 + \phi_j(1)^2}{2}$,

$\tilde{k}^{OS(2)}(0) = -\lim_{B\to\infty} \frac{1}{B^3} \sum_{j=1}^{B} \frac{\int_0^1 \phi_j(s)\phi_j''(s)ds}{2}$, where $\tilde{k}$ is as in (55). We note further that in a neighborhood around $\nu = 0$ and with $|u| \le L_T$, it must be the case in the limit that $\tilde{k}^{OS}(\nu) - k^{OS}(\nu) \to 0$ given the definitions in (20) and (55), 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 $0 < k^{OS(1)}(0) < \infty$, then $\frac{1-k^{OS}(\nu)}{\nu^2} \to \infty$ as $\nu \to 0$, so that q = 1. If not, note that as after equation (11), the frequency-domain weight function $K_T(\omega)$ for a kernel estimator must be nonnegative in order to guarantee positive semidefiniteness of the LRV estimate. Priestley (1981, p. 568) provides a proof that this implies that the Parzen characteristic exponent of the kernel (or, in the current case, implied mean kernel) must be no greater than q = 2, so that in our case q = 2 if $k^{OS(1)}(0) = 0$. Assumption 3 then guarantees the finiteness of $k^{OS(q)}(0)$.

Finally, as noted in Remark 6, 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 the condition that $k(0) = 1$ in Assumption 2 gives that $k^{(1)}(0) = -k'(0)$ for any kernel or implied mean kernel. Then using equation (21) and differentiating with respect to ν for $\nu > 0$,

$$\frac{dk_j^{OS}(B^{-1}\nu)}{d\nu} \equiv k_j^{OS\prime}(B^{-1}\nu) = B^{-1}\left[\frac{k_j(B^{-1}\nu)}{1-B^{-1}\nu} - \frac{\phi_j(B^{-1}\nu)\phi_j(0)}{1-B^{-1}\nu} - \frac{1}{1-B^{-1}\nu}\int_\nu^1 \phi_j(s)\phi_j{}'(s-B^{-1}\nu)\,ds\right] \quad (67)$$

so

$$k_j^{OS\prime}(0) = B^{-1}\left[k_j(0) - \phi_j(0)^2 - \int_0^1 \phi_j(s)\phi_j{}'(s)\,ds\right] = B^{-1}\left[1 - \frac{1}{2}\left(\phi_j(0)^2 + \phi_j(1)^2\right)\right], \quad (68)$$

where the final expression uses $k_j(0) = \int_0^1 \phi_j(s)^2 ds = 1$ and integrates by parts, as in (63). (The same expression for $k_j'(0)$ obtains starting from the expression for $k(v)$, $v \leq 0$.) Plugging this into (21) and taking the limit as $B \to \infty$ yields the same result as stated in the theorem for $q = 1$ estimators.

Priestley (1981, p. 460) also shows that for q even, $k^{(q)}(0) = -\frac{1}{q!}\left[\frac{d^q(k(v))}{dv^q}\right]_{v=0}$, yielding the relation $k^{(2)}(0) = -\frac{1}{2}k''(0)$. Then for $v > 0$, differentiating the expression in (67) again yields

$$k_j^{OS\prime\prime}(B^{-1}v) = B^{-2}\left[\frac{2k_j^{OS\prime}(B^{-1}v)}{1-B^{-1}v} + \frac{\phi_j(B^{-1}v)\phi_j'(0) - \phi_j'(B^{-1}v)\phi_j(0)}{1-B^{-1}v} + \frac{\int_v^1 \phi_j(s)\phi_j''(s - B^{-1}v)\,ds}{1-B^{-1}v}\right], \quad (69)$$

so that

$$k_j^{OS\prime\prime}(0) = B^{-2}\int_0^1 \phi_j(s)\phi_j''(s)\,ds\,, \quad (70)$$

and again the same expression obtains starting from $v \leq 0$. It follows from (21) that $k^{OS\prime\prime}(0)$ = $\frac{1}{B}\sum_{j=1}^{B} k_j^{OS\prime\prime}(0)$, and substituting (70) into this final expression and taking $B \to \infty$ 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 $E\hat{\Omega}^{OS}$ in equation (20), we can follow Priestley (1981, p. 459) and write

$$E\hat{\Omega}^{OS} - \Omega = \sum_{u=-(T-1)}^{T-1}\left\{k_{B,T}^{OS}(u/S)\left(1-|u/T|\right)-1\right\}\Gamma_u - \sum_{|u|\geq T}\Gamma_u$$

$$= \sum_{u=-(T-1)}^{T-1} \left\{ k_{B,T}^{OS}(u/S) - 1 \right\} \Gamma_u - \frac{1}{T} \sum_{u=-(T-1)}^{T-1} |u| k_{B,T}^{OS}(u/S) \Gamma_u - \sum_{|u| \geq T} \Gamma_u \, . \tag{71}$$

For the last term,

$$\left| \sum_{|u| \geq T} \Gamma_u \right| \leq \sum_{|u| \geq T} |\Gamma_u| \leq \frac{1}{T^q} \sum_{|u| \geq T} |u|^q |\Gamma_u| = o\left(T^{-q}\right) = o\left((B/T)^q\right), \tag{72}$$

by Assumptions 1(b) and 4. For the second term, given a bounded implied mean kernel (as in Assumption 3), similar steps to those taken for the last term give that

$$\frac{1}{T} \left| \sum_{u=-(T-1)}^{T-1} |u| k_{B,T}^{OS}(u/S) \Gamma_u \right| = O(1/T) = o\left((B/T)^q\right). \tag{73}$$

Finally, considering the first term in the bias expression, we can write

$$\begin{aligned} \sum_{u=-(T-1)}^{T-1} \left\{ k_{B,T}^{OS}(u/S) - 1 \right\} \Gamma_u &= -S^{-q} \sum_{u=-(T-1)}^{T-1} \left\{ \frac{1 - k_{B,T}^{OS}(u/S)}{\left(|u|/S\right)^q} \right\} |u|^q \, \Gamma_u \\ &= (B/T)^q k^{OS(q)}(0) \left[-2\pi s_z^{(q)}(0) \right] \left\{ 1 + o(1) \right\}, \end{aligned} \tag{74}$$

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 $s_z^{(q)}(0)$ 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 6).

(ii) For orthonormal series estimators with basis functions meeting Assumption 3, inspection of the proof of Theorem 2 in Sun (2014) or Theorem 4.1 in Sun (2013) shows that the higher-order size distortion is equal to $G'_m(\chi_m^\alpha)\chi_m^\alpha tr\left(A\Omega^{-1}\right)/m$, where $A = \lim_{T\to\infty} E\hat{\Omega} - \Omega$ as $B \to \infty$ such that $B/T \to 0$. Theorem 1(i) then gives that $A = (B/T)^q k^{OS(q)}(0)\left[-2\pi s_z^{(q)}(0)\right] + o\left((B/T)^q\right)$. (While this expression has not been shown formally for the SS series estimator, it is shown below in the derivation of Remark 6, which does not use this theorem in the case of the SS estimator.) Using $B = 1/b$ then yields equation (30). Note that in the scalar case of $p, m = 1$, we have that this higher-order size distortion (in both the kernel and orthonormal series case) is proportional to $\left|s_z''(0)/s_z(0)\right|$, so that all results below hold uniformly over all stochastic processes satisfying Assumption 1 with $\left|s_z''(0)/s_z(0)\right| \leq \kappa$ for finite κ.

For equation (31), the above steps carry through for the bias term, and the variance term follows from Theorem 5 and the proof of Theorem 2 in Sun (2011), with $\nu = B$ as in equation (25). The remainder terms follow from this theorem as well.

For kernel estimators, given Assumptions 1 and 2, equation (30) follows from Sun (2014) equation (16), along with the approximation $\dfrac{m\mathcal{F}_\infty^\alpha(m,b) - \chi_m^\alpha}{\chi_m^\alpha} = O(b)$ as $b \to 0$ given on page 665, where $\mathcal{F}_\infty^\alpha(m,b)$ is the fixed-*b* asymptotic distribution for kernel estimators. Equation (31) follows directly from the proof of Sun (2014) Theorem 5 for the case of the Gaussian location model. □

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 (30) and $\delta_{m,T}^\alpha(b)$ is defined implicitly by $\Pr_0\left[F_T^* > c_{m,T}^\alpha(b)\right] = \alpha + o(b) + o\left((bT)^{-q}\right)$. Taking a Taylor expansion of the null rejection rate around $c_m^\alpha(b)$,

$$\begin{aligned}\Pr_0\left[F_T^* > c_{m,T}^\alpha(b)\right] &= \alpha + G'_m(\chi_m^\alpha)\chi_m^\alpha \omega^{(q)} k^{(q)}(0)(bT)^{-q} - \delta_{m,T}^\alpha(b) G'_m(\chi_m^\alpha)\left[1 + O(b) + O\left((bT)^{-q}\right)\right] \\ &\quad + o(b) + o\left((bT)^{-q}\right) + o(\delta_{m,T}^\alpha(b)),\end{aligned} \tag{75}$$

where the fact that $\Pr_0{}'\left[F_T^* > c_m^\alpha(b)\right] = -G_m'(\chi_m^\alpha)\left[1+O(b)+O\left((bT)^{-q}\right)\right]+o(b)+o\left((bT)^{-q}\right)$, follows from Sun (2014) Theorem 2 and p. 665. Using this and $\Pr_0\left[F_T^* > c_{m,T}^\alpha(b)\right] =$ $\alpha + o(b) + o\left((bT)^{-q}\right)$ 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 (33) follows directly.

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

$$\Pr_\delta\left[F_T^* > c_{m,T}^\alpha(b)\right] = \left[1-G_{m,\delta^2}(\chi_m^\alpha)\right] + G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k^{(q)}(0)(bT)^{-q} - \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\nu^{-1}$$
$$-\delta_{m,T}^\alpha(b)G'_{m,\delta^2}(\chi_m^\alpha)\left[1+O(b)+O\left((bT)^{-q}\right)\right]+o(b)+o\left((bT)^{-q}\right)+o(\delta_{m,T}^\alpha(b)), \tag{76}$$

which analogously uses Sun (2014) Theorem 5. 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 (76) cancel, and using this along with $\delta_{m,T}^\alpha(b) = O\left((bT)^{-q}\right)$ yields the size-adjusted power relation given in equation (34). □

Proof of Theorem 3: This follows directly from equations (30) and (31). Fix a sequence b_1 for test F_1. Given equivalent values of q for tests F_1 and F_2, equation (30) gives that we must set $b_2 = \left(\frac{k_2^{(q)}(0)}{k_1^{(q)}(0)}\right)^{1/q} b_1$ in order to obtain equivalent higher-order size. We thus have that

$$G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k_2^{(q)}(0)(b_2T)^{-q} = G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha\omega^{(q)}k_1^{(q)}(0)(b_1T)^{-q}\ , \tag{77}$$

so that the corresponding second terms in the power expression (31) for F_1^* and F_2^* are equivalent. Using this along with equation (31) yields the desired relation. □

Proof of Corollary 1: Using equation (30) and the definition of the size distortion Δ_S, that size distortion can be written as follows:

$$\Delta_S = G'_m(\chi_m^\alpha)\chi_m^\alpha \omega^{(q)} k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right). \tag{78}$$

For kernel tests, using the definition of the size-adjusted power loss $\Delta_P(\delta)$ and the fact that $v = \left(b\int_{-\infty}^{\infty} k^2(x)dx\right)^{-1}$, we also have,

$$\Delta_P(\delta) = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha b\int_{-\infty}^{\infty} k^2(x)dx + o(b) + o\left((bT)^{-q}\right). \tag{79}$$

We can see that the leading terms in (78) and (79) 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 $O(b) = O\left((bT)^{-q}\right)$ under this sequence, the *o(b)* remainder term in (78) is also $o\left((bT)^{-q}\right)$, confirming that the leading term in (78) is in fact the first term. The same applies to the remainder terms in (79). These same steps apply for orthonormal series tests as well. □

Proof of Theorem 4: (i) Under the assumed sequence, we can rewrite the size distortion in (78) as

$$\begin{aligned} |\Delta_S|^{1/q} &= \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q} \left|\omega^{(q)}\right|^{1/q} \left(k^{(q)}(0)\right)^{1/q} (bT)^{-1}\left[1+o(1)\right]^{1/q} \\ &= \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q} \left|\omega^{(q)}\right|^{1/q} \left(k^{(q)}(0)\right)^{1/q} (bT)^{-1} + o\left((bT)^{-1}\right), \end{aligned} \tag{80}$$

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

$$T|\Delta_S|^{1/q} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/q} \left|\omega^{(q)}\right|^{1/q} \left(k^{(q)}(0)\right)^{1/q} b^{-1} + o(1/b). \tag{81}$$

Similarly, rewrite (79) (for kernel tests) as

$$\Delta_P(\delta) = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha b\int_{-\infty}^{\infty} k^2(x)dx + o(b). \tag{82}$$

Multiplying (81) and (82), 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

$$T\Delta_P(\delta)\left|\Delta_S\right|^{1/q} = a_{m,\alpha,q}(\delta)\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty} k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q} + o(1), \tag{83}$$

as stated. Identical steps for orthonormal series tests (using $v = 1/b$ in that case) then yield the tradeoff given in equation (40).

(ii) For kernel tests, we can express the maximum size-adjusted power loss $\Delta_P^{\max}$ using its definition in (38) and equation (79) as,

$$\Delta_P^{\max} = \sup_\delta \left\{\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\right\} b\int_{-\infty}^{\infty} k^2(x)dx + o(b), \tag{84}$$

since δ does not enter into the term $b\int_{-\infty}^{\infty} k^2(x)\,dx$. Thus following the same steps as in part (i) above, multiplying (81) and (84) yields equation (41). Identical steps for orthonormal series tests using the frontier in part (i) then yield equation (42).

(iii) Again using equation (30) and the same steps as in parts (i)-(ii), we can express $\sqrt{\left|\Delta_S\right|}$ for any test ($q = 1$ or $q = 2$) as,

$$\sqrt{\left|\Delta_S\right|} = \left(G'_m(\chi_m^\alpha)\chi_m^\alpha\right)^{1/w}\left|\omega^{(q)}\right|^{1/2}\left(k^{(q)}(0)\right)^{1/2}(bT)^{-q/2} + o\left((bT)^{-q/2}\right). \tag{85}$$

Multiplying this by (84), under the assumed sequence for b,

$$\Delta_P^{\max}\sqrt{|\Delta_S|} = \bar{a}_{m,\alpha,2}\left(\sqrt{k^{(q)}(0)}\int_{-\infty}^{\infty} k^2(x)\,dx\right)\left|\omega^{(q)}\right|^{1/2} T^{-1}\left(bT\right)^{1-q/2} + o(b^{3/2})\,. \tag{86}$$

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 $bT \to \infty$, and that this finding would hold as well in the orthonormal-series case (in which the expression would be identical but without the constant $\int_{-\infty}^{\infty} k^2(x)\,dx$). Thus comparing arbitrary kernel or orthonormal series tests with q = 1 and q = 2, for any two sets of values $k^{(q)}(0)$ and $\left|\omega^{(q)}\right|$, it must be that $\exists\, \bar{b}, \underline{T}$ 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. □

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 $q \le 2$, 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. 569-70), this is equivalent to minimizing $\left\{\int_{-\infty}^{\infty}\omega^2 K_j(\omega)\,d\omega\right\}^{1/2}\left\{\int_{-\infty}^{\infty} K_j^2(\omega)\,d\omega\right\}$, where K_j is the frequency-domain weight function corresponding to k. And for psd kernels, this minimum is in fact achieved exactly by the QS estimator, as proven in Priestley (1981, p. 571). Thus the QS estimator's size/size-adjusted power tradeoff defines the frontier for kernel tests.

Equation (45) (whose derivation does not use this theorem) then shows that the QS tradeoff dominates the tradeoff for EWP, which defines the frontier for orthonormal series tests, as shown in part (ii) of this theorem below. Thus we can conclude that the QS tradeoff defines the frontier for both kernel and orthonormal series HAR tests, as stated.

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 $k^{(2)}(0)=\pi^2/10$ (Table 7.1) and $\int_{-\infty}^{\infty}k^2(x)\,dx = 6/5$ (Table 6.1). Combining this with (41) yields (43) up to higher-order terms. Numerically calculating $\overline{a}_{m,\alpha,q} = \sup_\delta a_{m,\alpha,q}(\delta)$ for q = 2 and α = .05 yields $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.3368$ for m = 1, $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.6460$ for m = 2, and $\overline{a}_{m,\alpha,2}3\pi\sqrt{10}/25 \approx 0.9491$ for m = 3, as stated.

(ii) As in equations (22) and (25) (and discussed after (25)), only orthonormal series estimators yield fixed-b asymptotic distributions that are exact t (or exact F in the multivariate case). (As discussed after equation (23), the EWP estimator is unique among WP/SC estimators in having a fixed-b asymptotic distribution that allows for exact t- or F-based inference, but it too has an orthonormal series representation using the Fourier basis functions.) Thus we aim to achieve the size/size-adjusted power frontier, or equivalently to maximize higher-order power given equivalent higher-order size, among orthonormal series tests.

We can again confine the analysis to the q = 2 case. To fix higher-order size at a common value across all orthonormal series estimators, we can arbitrarily fix a value (or sequence) $B_{EWP} = 1/b_{EWP}$ for the EWP test. Theorem 3 then gives that the higher-order power gap between any alternative estimator *alt* (with q = 2) and EWP is given by $\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(B_{alt}^{-1} - B_{EWP}^{-1}\right)$, which follows from $\nu = B$ for orthonormal series estimators in equation (25). Thus having $B_{EWP} \geq B_{alt}$ for all alternatives with q = 2 such that EWP and *alt* have equivalent higher-order size is necessary and sufficient to prove the result that the exact t or F frontier is achieved by the EWP test.

From equation (30), having equivalent higher-order size requires setting

$$B_{alt} = \sqrt{\frac{k^{EWP(2)}(0)}{k^{alt(2)}(0)}}B_{EWP}. \tag{87}$$

Thus in order for $B_{EWP} \geq B_{alt}$ for all alternatives, it must be the case that $k^{EWP(2)}(0)$ is the minimum second generalized derivative value for the limiting implied mean kernel across all orthonormal series satisfying Assumption 3. (This is also apparent from the form of the tradeoff in (42).) From Theorem 1(i), a sufficient condition for this result can be obtained by showing that for any given value of B along the sequence, $k^{EWP(2)}(0)$ minimizes $\left|B^{-1}\sum_{j=1}^{B} k_j''(0)\right|$, where $k_j''(0) = B^{-2}\int_0^1 \phi_j(s)\phi_j''(s)\,ds$, across orthonormal series estimators.

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

$$\phi_j(s) = \sum_{l=1}^{\infty} a_{jl} e^{-i2\pi ls}, \tag{88}$$

where the a_{jl} values are as-yet undetermined projection coefficients. We know that for any orthonormal series,

$$\begin{aligned} 1 &= \int_0^1 \left|\phi_j(s)\right|^2 ds \\ &= \sum_l \sum_{l'} a_{jl} a_{jl'} \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} \\ &= \sum_l a_{jl}^2, \end{aligned} \tag{89}$$

and similarly

$$0 = \int_0^1 \phi_j(s)\overline{\phi_{j'\neq j}(s)}\,ds = \sum_l a_{jl} a_{j'\neq j,l}. \tag{90}$$

We then wish to minimize

$$\left|\frac{1}{B^3}\sum_{j=1}^{B}\int_0^1 \phi_j(s)\overline{\phi_j''(s)}\,ds\right| = \left|\frac{1}{B^3}\sum_{j=1}^{B}\sum_l \sum_{l'} a_{jl} a_{jl'} 4\pi^2 l^2 \int_0^1 e^{-i2\pi ls} e^{i2\pi ls} ds\right| = \frac{4\pi^2}{B^3}\sum_{j=1}^{B}\sum_l a_{jl}^2 l^2 \tag{91}$$

subject to the two constraints (89) and (90). But given the constraints, this is trivially solved by setting $a_{jj} = 1$, $a_{j,l\neq j} = 0$; that is, looking at the representation in (88), $\phi_j(s) = e^{-i2\pi js}$, so that we have in fact selected the Fourier basis functions themselves. Further, given that, as above, all orthonormal series in $L^2[0,1]$ are spanned by the Fourier basis functions, there is no such basis function for which $k^{(2)}(0) = 0$, which is a restatement of the Priestley (1981, p. 568) result discussed in the proof of Theorem 1(i) (i.e., that the Parzen characteristic exponent of a psd kernel or implied mean kernel must be no greater than $q = 2$). 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, which is equivalent to the EWP estimator; see also the derivation for the second part of Remark 6 below. (Note further that $\int_{-\infty}^{\infty} k^2(x)\,dx = 1$ for that kernel from Priestley, 1981, Table 6.1, as should be expected given that $\int_{-\infty}^{\infty} k^2(x)\,dx$ does not enter into the expression for the orthonormal series size/power tradeoff despite the fact that the EWP estimator can be expressed as a kernel estimator.) Thus we have that $\sqrt{k^{(2)}(0)} = \pi / \sqrt{6}$ for the EWP estimator. Combining this with (42) yields (44) 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. □

Derivation of Equation (45): As in the proof of Theorem 5, fix a sequence $B = 1/b_{EWP}$. To obtain equivalent higher-order size using the QS test, equation (30) gives that we must set

$$b_{QS} = \sqrt{\frac{k^{QS(2)}(0)}{k^{EWP(2)}(0)}}\, b_{EWP} = \sqrt{\frac{\pi^2/10}{\pi^2/6}}\, B^{-1} = \sqrt{\frac{3}{5}}\, B^{-1}, \tag{92}$$

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 $\nu_{EWP}^{-1} - \nu_{QS}^{-1} = B^{-1} - \frac{6}{5}\sqrt{\frac{3}{5}} B^{-1}$. Plugging this into the higher-order power difference in equation (35) (Theorem 3) yields the desired result.

Derivations for Remark 8: a. For the Legendre basis, 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 Stegum (1965)), so that the j^{th} normalized shifted Legendre polynomial is $\varphi_j(s) = p_j(s)\sqrt{2j+1}$. Because $p_j(0) = (-1)^j$ and $p_j(1) = 1$, we have $\varphi_j(0) = \sqrt{2j+1}(-1)^j$, $\varphi_j(1) = \sqrt{2j+1}$. Thus from Theorem 1(i), abusing notation slightly, we have $k_j^{Leg(1)}(0) = \frac{1}{B^2}\frac{1}{2}2(2j+1) = \frac{2j+1}{B^2}$ for each *j*, and thus $k_B^{Leg(1)}(0) = B^{-2}\sum_{j=1}^{B}(2j+1) = (B+2)/B \to 1$ as $B \to \infty$, as stated. As in Theorem 1(i), this also implies that $q = 1$ for the Legendre polynomials.

For SS, we can calculate $E\hat{\Omega}^{SS}$ 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 $\bar{x}_i - \bar{x} \equiv \frac{1}{T_i}\sum_{t \in T_i} x_t - \frac{1}{T}\sum_{t=1}^{T} x_t$ (where, abusing notation, T_i denotes both the number of observations in subsample *i* and the subsample that *t* indexes) can be written as $\bar{x}_i - \bar{x} = \frac{1}{T}\sum_{t=1}^{T}\left((B+1)1\{t \in T_i\} - 1\right)x_t = \frac{B+1}{T}\sum_{t=1}^{T}\left(1\{t \in T_i\} - \frac{1}{B+1}\right)x_t$. Thus summing over subsamples and squaring, we have,

$$\frac{1}{B}\sum_{i=1}^{B+1}(\bar{x}_i - \bar{x})^2 = \frac{1}{B}\sum_{i=1}^{B+1}\left(\frac{B+1}{T}\right)^2 \sum_{t=1}^{T}\sum_{s=1}^{T}\left(1\{t \in T_i\} - \frac{1}{B+1}\right)\left(1\{s \in T_i\} - \frac{1}{B+1}\right)x_t x_s. \qquad (93)$$

Taking the expectation of this value and performing the same change of variables as in (19),

$$E\frac{1}{B}\sum_{i=1}^{B+1}(\bar{x}_i-\bar{x})^2=\frac{B+1}{B}\frac{1}{T/(B+1)}\sum_{u=-(T-1)}^{T-1}\left[\left(1-\left|\frac{u}{T/(B+1)}\right|\right)1\left\{|u|\le\frac{T}{B+1}\right\}-\frac{1}{B+1}\left(1-\left|\frac{u}{T}\right|\right)\right]\Gamma_u$$

$$=\frac{B+1}{T}\sum_{u=-(T-1)}^{T-1}\left[\left(\frac{B+1}{B}-\frac{B+1}{B}\left|\frac{u}{T/(B+1)}\right|\right)1\left\{|u|\le\frac{T}{B+1}\right\}-\frac{1}{B}+\frac{1}{B}\left|\frac{u}{T}\right|\right]\Gamma_u. \tag{94}$$

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 (16) 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 $\nu = u/S$, we can write the SS implied mean kernel (i.e., the expression in brackets in (94)) as

$$k_B^{SS}(\nu)=\left(\frac{B+1}{B}-\frac{B+1}{B}|\nu|\right)1\{|\nu|\le 1\}-\frac{1}{B}+\frac{1}{B(B+1)}|\nu|. \tag{95}$$

Thus using the definition of the generalized first derivative in (28), we have $k_B^{SS(1)}(0)$ = $\frac{B+1}{B}-\frac{1}{B(B+1)}$ = $\frac{B+2}{B+1}\to 1$ as $B\to\infty$. Because $k^{SS(1)}(0)\neq 0$, $q = 1$ for the SS estimator. Further, comparing $E\hat{\Omega}^{SS}$ with Ω using (94) 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 for the Bartlett test, as stated.

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

$$k_B^{EWP(2)}(0) = -\frac{1}{2}\frac{1}{B}\sum_{j=1}^{B/2}\frac{1}{B^2}2\left(-4\pi^2 j^2\right) \;=\frac{\pi^2}{6}\frac{(B+1)(B+2)}{B^2} \rightarrow_{B\to\infty} \frac{\pi^2}{6}. \qquad (96)$$

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

$$k_B^{cos(2)} = -\frac{1}{2}\frac{1}{B}\sum_{j=1}^{B}\frac{1}{B^2}\left(-\pi^2 j^2\right) \;=\frac{\pi^2}{6}\frac{\left(B+1\right)\left(B+1/2\right)}{B^2} \rightarrow_{B\to\infty} \frac{\pi^2}{6}. \qquad (97)$$

Results (96) and (97) and Theorem 1(i) further imply that $q = 2$ for the Fourier and cosine estimators.

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

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

Note: b for QS is chosen so that its higher order size is the same as EWP with B terms.

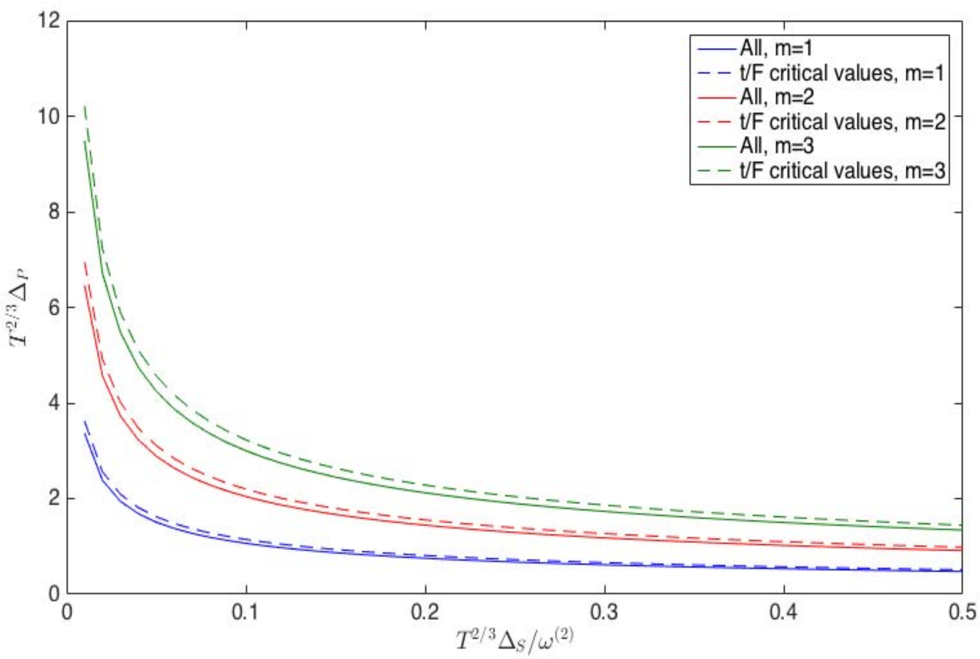

Figure 1. Higher-order frontier between the size distortion Δ_S and the maximum power loss $\Delta_P^{\max}$ of HAR tests in the Gaussian location model with dimension m, for stationary processes with average normalized spectral curvature $\omega^{(2)}$. Solid line: all kernel- and orthonormal series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

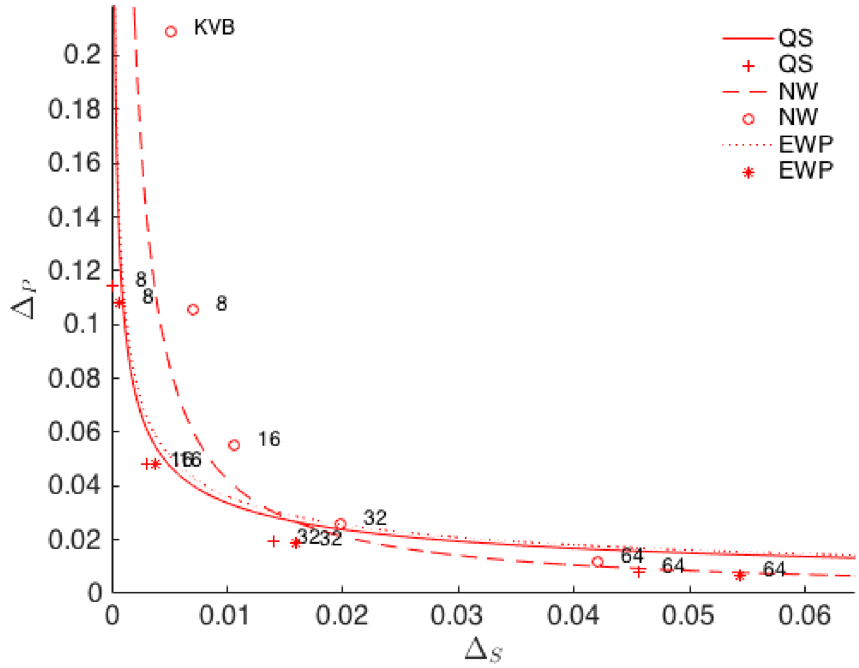

Figure 2. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Location model, $m = 1$, AR(1), $\rho = 0.5$, and $T = 200$.

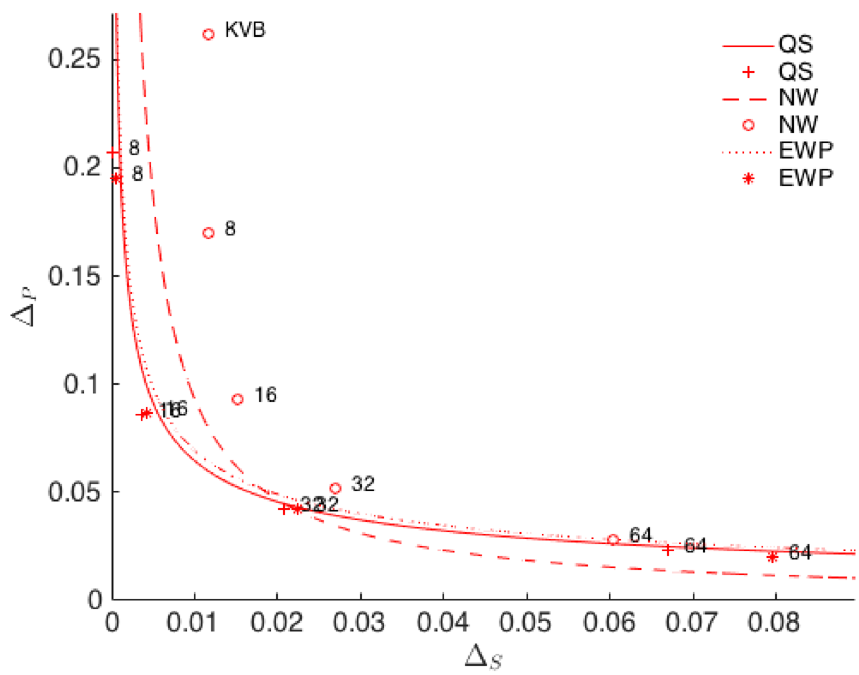

Figure 3. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Location model, $m = 2$, AR(1), $\rho = 0.5$, $T = 200$.

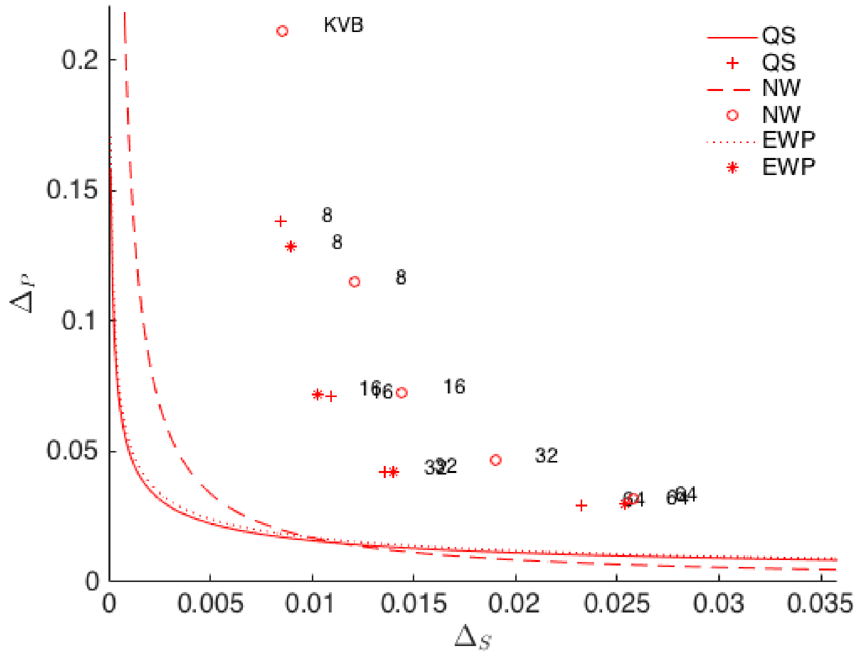

Figure 4. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for QS, Newey-West, and EWP estimators: Stochastic regressor, *m* = 1, AR(1), *ρ* = 0.5, T = 200. Theoretical curves are for the Gaussian location model.

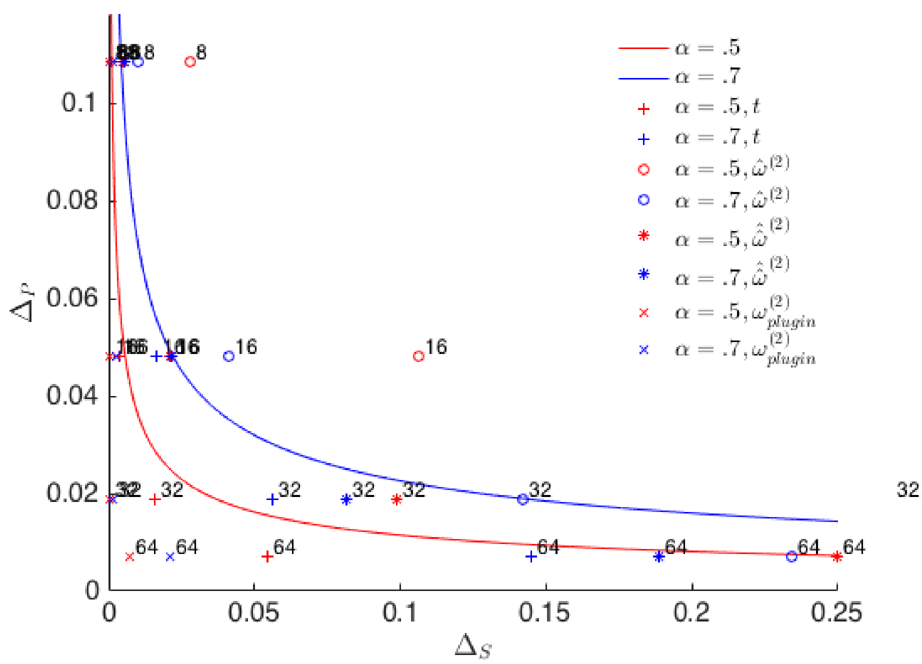

Figure 5. Theoretical (lines) and Monte Carlo (symbols) size distortion/power loss curves for the EWP estimator using feasible higher-order adjusted critical values:
Location model, *m* = 1, AR(1), *ρ* = 0.5 and *ρ* = 0.7, *T* = 200.