HAR Inference: Recommendations for Practice

April 19, 2016

Eben Lazarus

Department of Economics, Harvard University

Daniel J. Lewis

Department of Economics, Harvard University

James H. Stock

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

and

Mark W. Watson*

Department of Economics and the Woodrow Wilson School, Princeton University
and the National Bureau of Economic Research

*We thank Mikkel Plagborg-Møller and Yixiao Sun for helpful discussions. Replication files are available on Stock's Harvard Web site.

Abstract

The classic papers by Newey and West (1987) and Andrews (1991) spurred a large body of work investigating methods for improved heteroskedasticity- and autocorrelation-robust (HAR) inference in time series regression. Empirical practice, however, has not kept up with these developments. The goal of this paper is to draw on these developments to make a recommendation for practitioners about how to compute standard errors in regression settings with heteroskedasticity and serial correlation. A general conclusion from this literature is that the large size distortions of conventional HAR tests can be greatly reduced by using longer time-domain truncation parameters and modified "fixed-*b*" asymptotic critical values. In general, fixed-*b* distributions are nonstandard, however a class of tests (using orthonormal series estimators of the long-run variance) has standard *t* and *F* fixed-*b* asymptotic distributions. We provide new theoretical results on higher-order size and power that permit ranking tests within this class, and we compare tests in extensive Monte Carlo experiments that reflect both the benchmark designs used in the literature and data-based designs using U.S. macroeconomic time series. Theory and Monte Carlo results suggest that little is lost by restricting attention to tests with fixed-*b* *t*- and *F*-distributions. We conclude that estimating the long-run variance by the equal-weighted periodogram, averaged over the first $B/2$ periodogram ordinates, and using t_B and $F_{m,B-m+1}$ critical values (where m is the number of restrictions being tested) provides inference that is good in an absolute sense and represents a large improvement over currently-standard HAR methods. While the choice of the value B remains an open question, we discuss the tradeoffs involved in this choice and find that the value $B = 8$ performs well in leading cases with standard sample sizes.

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

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

1. Introduction

When the error term in time series regression is serially correlated, test statistics and confidence intervals must be computed using heteroskedasticity- and autocorrelation-robust (HAR) standard errors. Central to HAR inference is the estimation of the sum of the autocovariances of $z_t = x_t u_t$, where x_t is the vector of regressors and u_t is the regression error; this sum is the so-called long-run variance (LRV) of z_t. The two seminal papers in econometrics on HAR inference are Newey and West (1987) and Andrews (1991). Newey and West (1987) provided a consistent, positive semidefinite LRV estimator using the "tent" or Bartlett kernel to weight the sample autocovariances. Andrews (1991) asked, and answered, the questions of what kernel should be used and how its truncation parameter should be chosen. Drawing on the classical literature on spectral density estimation, Andrews (1991) framed the HAR problem as providing a consistent LRV estimator that minimizes the mean squared error (MSE), which entails trading off the variance against the square of the bias. The Newey-West/Andrews framework provides the recipe of using a kernel with a truncation parameter S_T, where S_T is chosen to be a function of the sample size T that ensures consistency and a low MSE. With this LRV in hand, inference proceeds using standard normal or chi-squared critical values.

There has been a large and active theoretical literature on HAR testing since the seminal Newey-West/Andrews papers; see Müller (2014) for a recent survey. This literature reaches three broad conclusions. First, the finite-sample performance of tests using minimum-MSE LRV estimators is not well approximated by the asymptotic theory, even with sample sizes that would usually be considered large (100 or 200 observations), leading to large size distortions when z_t has moderate or high persistence; see den Haan and Levin (1994, 1997) for an early simulation study.

Second, the source of these size distortions is that the relatively small truncation parameter needed to minimize MSE leads to considerable bias of the LRV estimator, which in turn distorts the size of the test. The theoretical basis for this conclusion is asymptotic expansions of null rejection probabilities, in which the bias (not bias^2) and variance of the LRV estimator appear as leading terms (Velasco and Robinson (2001), Sun, Phillips and Jin (2008)). Thus minimizing size distortions, rather than MSE, places more emphasis on minimizing bias and thus leads to larger truncation parameters than minimizing MSE.

Third, using larger truncation parameters has two drawbacks: the null distribution is no longer well-approximated by a standard normal/chi-squared distribution, and increasing the truncation parameter reduces size-adjusted power. Concerning the first of these drawbacks, Jansson (2004), and in more generality Sun, Phillips and Jin (2008) and Sun (2014), showed that inference using large-bandwidth estimators is improved (more precisely, provides a higher-order refinement as discussed below) by using critical values obtained using Kiefer and Vogelsang's (2005) "fixed-*b*" asymptotics. Under fixed-*b* asymptotics, the truncation parameter is treated as proportional to the sample size, so $S_T = bT$, where $b \in (0,1]$ is fixed. Fixed-*b* inference typically entails nonstandard critical values; the exception is the family of orthonormal series LRV estimators, for which the fixed-*b* asymptotic distributions are exact *t* and *F* (Brillinger (1975), Phillips (2005), Müller (2007), and Sun (2013)). We also refer to this class of estimators as "series LRV estimators" or "basis function LRV estimators" interchangeably. As pointed out by Phillips (2005), series LRV estimators can be computed from the explained sum of squares in a regression of $\hat{z}_t = x_t\hat{u}_t$ onto *B* orthonormal basis functions that span $L^2[0,1]$; the leading example of a series LRV estimator is the equal-weighted periodogram estimator, for which the *B* basis functions are $\sin(2\pi jt/T)$, $\cos(2\pi jt/T)$, $j = 1,\ldots, B/2$. As we show below, in the location model (mean-only), the family of series LRV estimators also includes the subsample estimator and test proposed by Ibragimov and Müller (2010). Concerning the second of these drawbacks, the reduction in power using large bandwidths, this ultimately entails a judgment between size and power, which we discuss in more detail below.

Despite all this work, no clear recommendation for practice has emerged, and the Newey-West/Andrews approach, typically implemented using the Bartlett window, remains the dominant method for HAR inference today.[1]

---

[1] The Newey-West estimator is implemented in standard econometric software, including Stata and Eviews. Among undergraduate textbooks, Stock and Watson (2011, eq (15.17)) recommend using the Newey-West estimator with the Andrews minimum-MSE bandwidth rule (with AR parameter .5), yielding $S_T = .75T^{1/3}$. Wooldridge (2006, sec. 12.5) recommends using the Newey-West estimator with either a rule of thumb for the bandwidth (he suggests S_T = 4 for quarterly data) or using Newey and West's (1987) rule, $S_T = 4(T/100)^{2/9}$ which, despite the slower rate, produces slightly larger bandwidths than the Stock-Watson rule for typical values of *T*. Dougherty (2011) and Hill, Griffiths, and Lim (2008) recommend Newey-West standard errors but do not discuss bandwidth choice. The reliance on Newey-West continues in the newest textbooks. Westhoff (2013, sec. 17.6) recommends Newey-West standard errors but does not mention bandwidth choice (leaving this to software defaults). Hilmer and Hilmer (2014) recommend Newey-West standard errors without discussing bandwidth choice, and their sole empirical example sets $S_T = 1$ using quarterly data.

The goal of this paper is to develop recommendations for empirical practice that are informed by the post–Newey-West/Andrews literature. In developing these recommendations, we adopt five criteria: (1) the procedure should provide the best possible control of size while sacrificing little power; (2) it should be easily implemented using standard statistical software; (3) it should be explainable to typical practitioners (not just a black box); (4) it should perform robustly in a wide range of econometric applications for data with short-run dependence; and (5) the LRV estimator should be positive semidefinite.

There is tension among these criteria. In light of the post–Newey-West/Andrews literature, criterion (1) suggests using tests with large bandwidths and fixed-b critical values, which are in general nonstandard. While in principle modern econometric packages could implement nonstandard distributions, outside of unit root, cointegration, and stability/break tests there is limited use of nonstandard distributions in practice. Thus criteria (2) and (3) point towards methods that have fixed-b asymptotic t- or F-distributions, as long as the size/power cost of this restriction is not too great. In criterion (4) we adopt the standard that size should be controlled over the family of applications in which there is short-range dependence but neither theory nor the data would point to unit root/long-range dependence. In these settings, LRV estimators have well-documented failures to control size (Müller (2007, 2014)), so different methods are needed. And while there exist methods to correct for non-positive-semidefiniteness (e.g., Politis (2011)), a desire for transparency and ease of communication leads us to consider only positive semidefinite LRV estimators, as in criterion (5).

This paper accordingly considers the class of tests based on kernel estimators or series LRV estimators.[2] Based on results in the literature, new theoretical results, and extensive Monte Carlo simulations, we reach six main conclusions:

---

[2] Our focus on asymptotic t- or F- inference under fixed-b asymptotics rules out an important class of LRV estimators based on parametric estimators of the LRV based on autoregressions (Berk (1974)) or vector autoregressions (den Haan and Levin (2000), Sun and Kaplan (2014)). This focus also rules out methods in which parametric models are used to prewhiten the spectrum, followed by nonparametric estimation (e.g. Andrews and Monahan (1992)), although for comparison we consider a prewhitening estimator in our Monte Carlo study. Finally, this focus, and the desire for transparency (criteria (2) and especially (3)), also leads us away from bootstrap methods (e.g. Gonçalves and Vogelsang (2011)).

1. The literature has established, and our numerical results confirm, that using fixed-b critical values substantially improves size control. We therefore focus on tests using fixed-b critical values.
2. As a matter of theory, among tests with the same (higher-order) size using fixed-b critical values, the greatest power is achieved using the quadratic spectral (QS) kernel estimator.
3. If one further restricts attention to tests having fixed-b distributions that are t or F, as opposed to nonstandard distributions, then as a matter of theory the test using the equal-weighted periodogram (EWP) LRV estimator achieves the maximum power. It turns out that the cost of this restriction to t- or F-based fixed-b inference is small: for example, for $b = 1/8$ and one linear restriction being tested, the maximal higher-order power gap between same-sized QS and EWP tests is 0.0074 over all alternatives.
4. These and additional theoretical results are borne out in our Monte Carlo simulations, which point to the near-equivalence of the EWP and QS tests, and the typically better performance of the EWP test than other kernel or orthonormal series tests.
5. We therefore recommend using the EWP estimator in practice, where critical values are obtained from the t_B or, when m restrictions are being tested, $F_{m,B-m+1}$ distributions. Computing the EWP estimator requires a choice of $B = 1/b$ for this orthonormal series estimator. Our numerical work suggests that choosing $B = 8$ provides a reasonable tradeoff between size and power given standard sample sizes, but the choice of B ultimately remains an open question, as discussed further below.
6. Although the Newey-West estimator is both theoretically dominated by the EWP estimator and has the disadvantage of nonstandard fixed-b critical values, the size and power differences between comparable Newey-West and EWP tests are numerically sufficiently small for relevant parameterizations that one could be justified in using the Newey-West estimator instead. If so, then we recommend setting the Newey-West bandwidth to $S_T = (3/(2B))T$ and evaluating the test statistic using critical values from the t_{B} or $F_{m,B-m+1}$ distributions.

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, essentially restating conclusions 2 and 3 above as theorems and providing related results. The Monte Carlo design is summarized in Section 5, results are given in Section 6, and Section 7 concludes. Proofs are given in the Appendix.

2. Notation and Class of LRV Estimators

2.1 The model and HAR testing problem

We consider the problem of inference about the coefficients β in the linear regression model with dependent variable y_t and p regressors x_t,

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

We assume that the process generating (x_t, u_t) is stationary, $E(u_t|x_t) = 0$, and that the OLS estimator has the asymptotic normal distribution,

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma_{XX}^{-1}\Omega\Sigma_{XX}^{-1}), \tag{2}$$

where $\Sigma_{XX} = E(x_t x_t')$ and Ω is the long-run variance of $z_t = x_t u_t$,

$$\Omega = \sum_{u=-\infty}^{\infty} \Gamma_u \text{, where } \Gamma_u = E(z_t z_{t-u}'). \tag{3}$$

The testing problem under consideration is to test the null hypothesis in (1) that $R\beta = r_0$, where R is the $m\times p$ matrix of m restrictions, against the two-sided alternative. Motivated by (2), HAR inference entails constructing the t-statistic testing the hypothesis that the j^{th} element of β, β_j, takes on the value β_{j0}, or, if $m > 1$, the F statistic, respectively given by

$$t(\hat{\Omega}) = \frac{\sqrt{T}\left(\hat{\beta}_j - \beta_{j0}\right)}{\sqrt{(\hat{\Sigma}_{XX}^{-1}\hat{\Omega}\hat{\Sigma}_{XX}^{-1})_{jj}}} \text{, and } F_T(\hat{\Omega}) = T(R\hat{\beta} - r_0)'(R\hat{\Sigma}_{XX}^{-1}\hat{\Omega}\hat{\Sigma}_{XX}^{-1}R')^{-1}(R\hat{\beta} - r_0)/m \tag{4}$$

where $\hat{\Sigma}_{XX} = T^{-1}\sum_{t=1}^{T} X_t X_t'$ and $\hat{\Omega}$ is an estimator of the long-run variance Ω.

The class of estimators $\hat{\Omega}$ considered here is comprised of two families of estimators, the family of positive semidefinite kernel estimators and the family of orthonormal series estimators.

2.2 Kernel estimators

Kernel estimators of Ω can be written as the weighted sum of sample autocovariances using the time-domain weight function, or kernel, $k(.)$:

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

where S is the time-domain truncation parameter, $\hat{z}_t = x_t \hat{u}_t$, where $\hat{u}_t$ are the OLS residuals, and the superscript "SC" denotes sum-of-covariances. Examples of kernels $k(v)$ include the Bartlett kernel used by Newey and West (1987), $k(v) = (1 - |v|)\mathbf{1}(|v|\leq 1)$, 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 Priestly (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:

$$\hat{\Omega}^{WP} = 2\pi\sum_{j=1}^{T-1} K_T(2\pi j/T) I_{\hat{z}\hat{z}}(2\pi j/T), \tag{6}$$

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,

$$I_{zz}(\omega) = d_z(\omega)\overline{d_z(\omega)}', \text{ where } d_z(\omega) = (2\pi T)^{-1/2}\sum_{t=1}^{T} z_t e^{-i\omega t}. \tag{7}$$

An important special case of kernel estimators is the equal-weighted periodogram (EWP) estimator, which is computed using the Daniell (flat) kernel that equally weights the first $B/2$ periodogram terms, where B is even:

$$\hat{\Omega}^{EWP} = \frac{2\pi}{B}\sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j/T) = \frac{2\pi}{B}\sum_{j=1}^{B/2} d_z(2\pi j/T)\overline{d_z(2\pi j/T)}'. \tag{8}$$

Kernel estimators are positive semidefinite with probability one if the frequency-domain weight function $K_T(\omega)$ is nonnegative. Under conditions given in Andrews (1991), as S, $T \to \infty$ and $S/T \to 0$, $\hat{\Omega}^{SC}$, $\hat{\Omega}^{WP} \xrightarrow{p} \Omega$ and asymptotic inference on β uses standard normal and chi-squared critical values.

2.3 Orthonormal series estimators

Orthonormal series estimators are obtained by projecting $\hat{z}_t$ onto the first B functions of a set of orthonormal basis functions for $L^2[0,1]$. The EWP estimator (8) is an example of an orthonormal series estimator using the Fourier basis, discussed further below.

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 let Φ denote the $T{\times}B$ matrix consisting of the $T{\times}1$ vectors of basis functions evaluated at t/T:

$$\Phi = [\Phi_1 \ldots \Phi_B], \text{ where } \Phi_j = [\varphi_j(1/T)\ \ \varphi_j(2/T) \ldots \varphi_j(1)]' \text{ and } \Phi'\Phi/T = \mathrm{I}_B. \tag{9}$$

The orthonormal series LRV estimator is,

$$\hat{\Omega}^{BF} = \frac{1}{B}\sum_{j=1}^{B}\hat{\Omega}_j, \text{ where } \hat{\Omega}_j = \hat{\Lambda}_j\hat{\Lambda}_j' \text{ and } \hat{\Lambda}_j = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t. \tag{10}$$

Note that Φ and $\hat{\Omega}^{BF}$ omit the $j = 0$ function, for which $\Phi_0 = \iota = (1\ 1 \ldots 1)'$ and $\Phi_0'\Phi = 0$. Because $\hat{\Omega}^{BF}$ is computed using an outer product, it is positive semidefinite with probability one.

The theory in this paper covers any basis functions that are twice differentiable, plus one nondifferentiable basis function based on splitting the sample. The four basis functions we examine numerically are Fourier, cosine, Legendre, and split-sample, which we shall refer to as the Ibragimov-Müller basis function, as it replicates the split-sample procedure of Ibragimov-Müller (2010) for $x_t = 1$. For convenience we use real-valued functions.

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

$$\{\varphi_{2j-1}(s), \varphi_{2j}(s)\} = \{\sqrt{2}\cos(2\pi js), \sqrt{2}\sin(2\pi js)\}, j = 1,\ldots, B. \tag{11}$$

The Fourier basis functions produce $\hat{\Omega}^{EWP}$ in (8). The first 4 Fourier basis functions (corresponding to periodogram ordinates 1 and 2) are shown in Figure 1a.

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:

$$\{\varphi_{Tj}(s)\} = \left\{\sqrt{2}\cos\left[2\pi j\left(s-\frac{1}{2T}\right)\right]\right\}, j = 1,\ldots, B. \tag{12}$$

Hwang and Sun (2015) refer to this as the shifted cosine function. A closely related alternative is Phillips' (2005) proposal of using $\{\sqrt{2}\sin[\pi j(s\text{-}1/(2T))]\}$, $j = 1,\ldots, B$, 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 second order in an expansion of B/T) to $\hat{\Omega}^{EWP}$. A related asymptotic equivalency result is shown in Section 4 for the type II cosine transform. The first 4 cosine basis functions are shown in Figure 1b.

Legendre polynomials can be constructed as the B functions of the Gram-Schmidt orthonormalization of $\varphi_j(s) = s^j$, $j = 1, \ldots, B$, on [0,1]. Abramowitz and Stegun (1965, ch. 8) give recursions for Legendre polynomials on [-1,1]. The first 4 shifted and normalized Legendre polynomial basis functions are shown in Figure 1c.

Ibragimov-Müller basis function. Ibragimov and Müller (2010) (IM) 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] We first consider the location model ($x_t = 1$), for which IM propose performing inference using

$$t^{IM} = \sqrt{B+1}\left(\bar{\hat{\beta}} - \beta_0\right) \Big/ \sqrt{S_{\hat{\beta}}^2} \text{ , where } S_{\hat{\beta}}^2 = \frac{1}{B}\sum_{i=1}^{B+1}\left(\hat{\beta}^{(i)} - \bar{\hat{\beta}}\right)^2, \tag{13}$$

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 IM estimator can be rewritten as an orthonormal series estimator. In this case, $\hat{\beta}^{(i)} = [T/(B+1)]^{-1}\sum_{t=[T/(B+1)](i-1)+1}^{[T/(B+1)]i} y_t$ and $\bar{\hat{\beta}} = \bar{y}$ (the full-sample mean of y_t), so $S_{\hat{\beta}}^2 = \xi' M_\iota \xi / B$, where $\xi = (\hat{\beta}_1 \ldots \hat{\beta}_{B+1})'$ and $M_\iota = \mathrm{I}_{B+1} - \iota\iota'/(B+1)$, where ι is the B+1 vector of ones. Let G_1 be the $(B+1)\times B$ matrix of eigenvectors corresponding to the B unit eigenvalues of M_ι, and let

$$\Phi^{IM} = \sqrt{B+1}\left[I_{B+1} \otimes \iota_{T/(B+1)}\right]G_1 \text{ and } \quad = \left(T^{-1/2}\Phi^{IM\prime}\hat{z}\right)'\left(T^{-1/2}\Phi^{IM\prime}\hat{z}\right)/B, \tag{14}$$

where $\otimes$ is the Kronecker product. Then Φ^{IM} is $T\times B$, $\Phi^{IM\prime}\Phi^{IM}/T = \mathrm{I}_B$ and the t-statistic in (13) can equivalently be written as, $t^{IM} = \sqrt{T}(\hat{\beta} - \beta_0) \Big/ \sqrt{\hat{\Omega}^{IM}}$; that is, the IM test statistic (13) can be written in the standard form (4) for orthonormal series estimators, computed using (14).

We will refer to Φ^{IM} in (14) as the IM-basis and to $\hat{\Omega}^{IM}$ as the IM-basis LRV estimator. The IM basis functions Φ^{IM} are plotted in Figure 1(d) for $B = 4$ (corresponding to 5 subsamples). Unlike the Fourier, cosine, and Legendre basis functions, the IM basis functions are discontinuous step-like functions.

For $B = 2^n - 1$, where n is an integer, the IM and Haar basis functions are equivalent, however for other B the Haar functions do not span Φ^{IM}.

---

[3] This subsample variance estimator is also referred to as the "batch mean estimator" in previous literature, see for example [XX]. We thank Yixiao Sun for pointing us to this literature.

The basis function representation (14) was developed for the location model. Although the IM-basis function estimator $\hat{\Omega}^{IM}$ can be computed for $p > 1$, the resulting test statistic equals the original split-sample Ibragimov-Müller test statistic (13) only in the location model. Mechanically, this distinction arises because, for general $p > 1$, the IM test statistic is based on the sample variance of the subsample estimators of $\hat{\beta}$, in which both the numerator and denominator are computed using the subsample, whereas $\hat{\Omega}^{IM}$ estimates Ω using the subsample estimates of the expression for $\hat{\beta} - \beta_0$ and the full-sample estimator of Σ_{XX}. This seemingly minor distinction prevents giving the IM statistic in (13) an orthonormal series interpretation for $p > 1$. More importantly, this distinction results in a lack of size control of the original Ibragimov-Müller (2010) test outside of the location model when there is nonzero long-run correlation between the regressor and the error, a situation common in time series applications (Lazarus, Lewis, and Stock (2015)). We therefore do not pursue the original Ibragimov-Müller test further, although we do examine the performance of the IM-basis test.

Implied mean kernels of orthonormal series estimators. Although the only series estimator with an exact kernel representation is the Fourier series 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. Using the device in Grenander and Rosenblatt (1957, p. 125), we can express the mean of the j^{th} contribution to an orthonormal series estimator approximately as the mean of a kernel estimator:

$$
\begin{aligned}
E\hat{\Omega}_j &= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t\right)' \\
&= E\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)z_t + o_p(1)\right)\left(\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)z_t + o_p(1)\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\phi_j(t/T)\phi_j(s/T)\Gamma_{s-t} + o(1)
\end{aligned}
$$

$$= \sum_{u=-(T-1)}^{T-1} k_{j,T}^{BF}(u/T)(1-|u/T|)\Gamma_u + o(1), \qquad (15)$$

where the second equality follows from the proof of Theorem 3.1 in Sun (2013), which gives that the limiting distribution of $\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)\hat{z}_t$ is identical to that of $\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_j(t/T)z_t$ as long as $\Phi_j'\iota = 0$ (as assumed above), and the third equality obtains by collecting terms and setting

$$k_{j,T}^{BF}(u/T) = \frac{1}{T-|u|}\sum_{t=1}^{T}\phi_j(t/T)\phi_j((t-u)/T)\mathbf{1}(1 \le t-u \le T). \qquad (16)$$

Thus

$$E\hat{\Omega}^{BF} = \frac{1}{B}\sum_{j=1}^{B}E\hat{\Omega}_j = \sum_{u=-(T-1)}^{T-1} k_T^{BF,B}(u/S)(1-|u/T|)\Gamma_u + o(1), \qquad (17)$$

where $k_T^{BF}(u/S) = \frac{1}{B}\sum_{j=1}^{B}k_{j,T}^{BF}\left(B^{-1}\frac{u}{S}\right)$, and $SB = T$, so that S corresponds to the usual time-domain truncation parameter. The change of variables from u/T to u/S aligns the implied mean kernel with the usual expression for kernels as a function of u/S, so that (17) matches (for example) Priestley (1981, eq. (6.2.120)).

For large T and fixed B, $\lim_{T\to\infty} k_T^{BF}(v) = k^{BF}(v)$, where

$$k^{BF}(v) = \frac{1}{B}\sum_{j=1}^{B}k_j^{BF}(v), \text{ where } k_j^{BF}(v) = \frac{1}{(1-|v|)}\int_{\max(0,v)}^{\min(1,1+v)}\phi_j(s)\phi_j(s-v)ds. \qquad (18)$$

Equations (17) and (18) 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 $T \to \infty$.

The approximate implied mean kernels $k_T^{BF}(u/S)$ for the Fourier, cosine, Legendre, and IM basis functions are plotted in Figure 2a for $B = 8$ and $T = 200$. The Fourier transform of these implied mean kernels, that is, the frequency-domain implied mean kernel, is plotted in Figure 2b at the frequencies $2\pi j/T$, $j = 1, \ldots, 32$. As in equation (8), the Fourier estimator is the only one of these four that has an exact kernel representation, and its frequency-domain implied kernel is the familiar flat (Daniell) kernel that gives equal weight to the first $B/2$ periodogram ordinates. The remaining three implied mean kernels in the frequency domain also concentrate their mass at low frequencies. The IM implied mean kernel in the time domain is the Bartlett kernel. The Legendre implied mean kernel is interesting in that, in the time domain, it is indistinguishable from the IM implied mean kernel near the origin, a result shown formally below, though the Legendre kernel places weight on more distant autocovariances. In the frequency domain both the Legendre and IM implied mean kernels have considerably more leakage than the Fourier and cosine kernels.

3. Fixed-*b* Asymptotics and Rejection Rate Expansions

This section summarizes results from the literature on fixed-*b* asymptotics. As shown by Jansson (2004) and Sun (2014), the use of fixed-*b* critical values provides higher-order corrections to the size of HAR tests.

At an intuitive level, fixed-*b* asymptotics are appealing because using large S_T to reduce bias means that it could be misleading to ignore the variance of $\hat{\Omega}$, as is done in the large-sample normal approximation, while in contrast fixed-*b* asymptotics treat $\hat{\Omega}$ as having a nondegenerate limiting distribution. This intuition has been formalized in a series of papers that provide asymptotic expansions of the rejection rate of HAR tests using normal and fixed-*b* critical values under the "small-*b*" sequence, $b \to 0$. The main finding is that using fixed-*b* critical values eliminates the leading term in *b*, which arises from the normal approximation ignoring the variance of $\hat{\Omega}$. This higher-order improvement was shown first by Jansson (2004) for the special case of the Kiefer-Vogelsang-Bunzel (2000) estimator in the Gaussian location model. This result was generalized to the family of kernel estimators by Sun (2014), and to the family of orthonormal series estimators by Sun (2013).

3.1 Fixed-*b* asymptotics

Fixed-*b* asymptotics provide a tractable approximation to the distribution of $\hat{\Omega}$ when the time-domain domain bandwidth S_T is large, so that bias is small but variance is not eliminated. In the frequency domain, this corresponds to holding the frequency-domain bandwidth fixed, for example, by choosing the kernel so that all or most of the weight concentrates on the first *B* periodogram ordinates, where *B* is fixed. Translated into orthonormal series estimators, this corresponds to considering projections onto the first *B* basis functions, where *B* is fixed.

***Kernel estimators*.** The first fixed-*b* asymptotic results for kernel estimators were provided in the classical spectral estimation literature for weighted periodogram (WP) estimators with truncated frequency-domain weights. For clarity, the discussion here focuses on the scalar case; the extension to the multivariate case is summarized at the end. Specifically, consider the WP estimator (6), let $K_{jT} = K_T(2\pi j/T)$, and suppose that $K_{jT} = 0$ for $j > B$ and that $K_{jT} \rightarrow K_j$ as $T \rightarrow$. 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 (8)). For example, for the QS kernel, $K_j = (3\pi B/2)^{-1}(1 - (j/B)^2)\mathbf{1}(j \leq B)$ and for the Daniell kernel $K_j = (2\pi \mathrm{B})^{-1}\mathbf{1}(j \leq B)$. Under these conditions,

$$\hat{\Omega}^{WP} \xrightarrow{d} \left(2\pi \sum_{j=1}^{B} K_j \xi_j\right)\Omega, \text{ where } \xi_j \text{ are i.i.d. } \chi_2^2/2, j = 1,\ldots, B, \tag{19}$$

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

Further, it is possible to obtain the general fixed-*b* asymptotic distribution for kernel estimators by extending (19) to kernels in which the frequency-domain kernel is not truncated, as is the case for the Bartlett kernel, under the assumption that $b = S_T/T$ is constant. Using Mercer's Theorem, the previous results from Brillinger (1981) and Priestley (1981), and an infinite-dimensional central limit theorem (as in Kandelaki & Sozanov 1964), it can be shown that the above representation still holds. Kiefer and Vogelsang (2005) obtained the fixed-*b* asymptotic distribution for kernel estimators by working directly with the time-domain representation. They showed that,

$$\hat{\Omega}^{SC} \xrightarrow{d} \left(\int_0^1 \int_0^1 k\left(\frac{r-s}{b}\right) dV(r)dV(s) \right) \Omega , \tag{20}$$

where V is a Brownian bridge. Sun (2014) provides additional expressions for the fixed-b asymptotic distribution.

Equation (19) represents the fixed-b asymptotic distribution of as a possibly infinite average of χ_2^2 random variables. 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. The EWP/Daniell estimator satisfies this condition. So too would an EWP estimator with equal weights on periodogram ordinates $j = 2, \ldots, B/2+1$, that is, shifting the Daniell kernel to drop the first ordinate and include the $B/2+1$ ordinate. However, this shifted-Daniell estimator would in general have greater bias, and the same variance, as the EWP estimator proximate to the origin, so we do not consider it further. With this caveat omitting trivially dominated estimators, 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 critical values in practice, Tukey (1949) proposed approximating the distribution (19) by a chi-squared, specifically,

$$\hat{\Omega}^{WP} \sim \left(\chi_\nu^2 / \nu\right)\Omega \text{, where } \nu = \left(b\int_{-\infty}^{\infty} k^2(x)\,dx \right)^{-1}, \tag{21}$$

where ν is the "equivalent degrees of freedom" of $\hat{\Omega}^{WP}$.

Orthonormal series estimators. If $|\varphi_j'|$ is bounded, then a standard central limit theorem for stationary processes shows that $T^{-1/2}\sum_{t=1}^{T} \phi_j(t/T)z_t \xrightarrow{d} N(0,\Omega)$. For B fixed, it follows that $\hat{\Omega}^{BF}$ in (10) has the fixed B asymptotic distribution,

$$\hat{\Omega}^{BF} \xrightarrow{d} \left(\frac{1}{B}\sum_{j=1}^{B} \zeta_j \right)\Omega \sim (\chi_\nu^2 / \nu)\Omega \text{, } \nu = B \text{ and } \zeta_j \text{ are i.i.d. } \chi_1^2, j = 1,\ldots, B. \tag{22}$$

Thus, for orthonormal series estimators, Tukey's approximation (21) to the asymptotic distribution holds exactly with $\nu = B$.

3.2. Fixed-*b* Inference

Fixed-*b* asymptotic inference proceeds by combining the fixed-*b* limiting distributions of $\hat{\Omega}$ above with the independent normal distribution of $\sqrt{T}\left(\hat{\beta}-\beta_0\right)$ to obtain a fixed-*b* limiting distribution of $t(\hat{\Omega})$ in (4). In general, this distribution is nonstandard and requires tabulated critical values, see Kiefer, Vogelsang, and Bunzel (2000). However, for those LRV estimators with fixed-*b* asymptotic distributions that are exactly chi-squared as in (22), the fixed-*b* asymptotic distribution of $t(\hat{\Omega})$ is t_v where $v = B$. This observation appears to date to Brillinger (1975, exercise 5.13.25). 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, so that among kernel and orthonormal series estimators, orthonormal series estimators uniquely have exact asymptotic t_v fixed-*b* distributions of $t(\hat{\Omega})$.

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

$$\hat{\Omega}^{BF} \xrightarrow{d} \Omega^{1/2}\Xi\,\Omega^{1/2\prime}, \text{ where } \Xi \sim W_p(\mathrm{I},B), \tag{23}$$

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 (4) will have an asymptotic Hotelling T^2 distribution. As in Sun (2013), it is convenient to rescale F_T so that it has the $F_{m,B-m+1}$ distribution. Specifically,

$$F_T^*(\hat{\Omega}^{BF}) \xrightarrow{d} F_{m,B-m+1}, \text{ where } F_T^*(\hat{\Omega}^{BF}) = \frac{B-m+1}{B}F_T(\hat{\Omega}^{BF}). \tag{24}$$

3.3 Higher-Order Expansions of Rejection Rates

The rejection rate asymptotic expansions use classical expressions for the variance and bias of spectral density estimators. The variance expressions depend on the kernel through the Tukey equivalent degrees of freedom v defined in (21) and (22). The bias expressions depend on the kernel through the so-called Parzen characteristic exponent of the kernel or implied mean kernel.

The Parzen characteristic exponent q of kernel (or implied mean kernel) k is the maximum integer such that

$$k^{(q)}(0) = \lim_{x \to 0} \frac{1 - k(x)}{|x|^q} < \infty \,. \tag{25}$$

The term $k^{(q)}(0)$ is called the q^{th} generalized derivative of k, evaluated at the origin.

Sun's higher-order expansions of fixed-b rejection rates of tests based on kernel estimators (Sun (2014)) and on orthonormal series estimators (Sun (2013)) play a central role in our analysis, so we summarize them here in unifying notation that covers both families of LRV estimators. Let F_T^* denote the modified F statistic in (24), and let $c_m^\alpha(b)$ denote the fixed-b critical value for the level α test with m degrees of freedom.

The asymptotic expansion of the null rejection rate is

$$\Pr_0\left[F_T^* > c_m^\alpha(b)\right] = \alpha + \mu_0(\alpha, m) k^{(q)}(0)(bT)^{-q} + o(b) + o\left((bT)^{-q}\right), \tag{26}$$

where $\mu_0(\alpha, m) = G_m'(\chi_m^\alpha)\chi_m^\alpha tr\left(\sum_{j=-\infty}^{\infty} |j|^q \, \Gamma_j \Omega^{-1}\right)/m$, G_m is the chi-squared cdf with m degrees of freedom, and χ_m^α is its 1-α quantile (the asymptotic chi-squared critical value as $b \to 0$). As discussed in Sun (2013, 2014) and in the Appendix, the second term in (26) depends only on the bias of the LRV estimator and features of the data and the test, so that minimizing the test's size distortion is equivalent to minimizing the bias of the 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,

$$\Pr_\delta\left[F_T^* > c_m^\alpha(b)\right] = [1 - G_{m,\delta^2}(\chi_m^\alpha)] + \mu_{\delta^2}(\alpha, m) k^{(q)}(0)(bT)^{-q} - \delta^2 \psi_{\delta^2}(\alpha, m)\nu^{-1}$$
$$+ o(b) + o\left((bT)^{-q}\right) + O(\log T / \sqrt{T}), \tag{27}$$

where G_{m,δ^2} is the noncentral chi-squared cdf with *m* degrees of freedom and noncentrality parameter δ^2, $\mu_{\delta^2}(\alpha,m) = G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha tr\left(\sum_{j=-\infty}^{\infty}|j|^q \Gamma_j \Omega^{-1}\right)/m$, $\psi_{\delta^2}(\alpha,m) = \frac{1}{2}G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha$, and the $O(\log T/\sqrt{T})$ term does not depend on *b*. This expression depends on both bias, as reflected in the third term, and variance, as reflected in ν^{-1}, of the LRV estimator.

4. Main Theoretical Results

This section summarizes the theoretical results comparing tests performed using fixed-*b* asymptotic critical values, with the goal of providing theoretical guidance on the choice of kernel or basis function. These results are based on the expansions (26) and (27), which use fixed-*b* critical values. Throughout, the term $\mu_0(\alpha,m)k^{(q)}(0)(bT)^{-q}$ in $(bT)^{-q}$ in (26) is referred to as the higher-order size distortion, and this term plus α is referred to as the higher-order size. The higher-order power using fixed-*b* critical values consists of all three explicit terms on the right side of (27).

In addition to the assumptions stated in equation (2), 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).

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 $r \in [0,2]$.

(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.

Assumption 2 (kernels). The kernels $k(\nu):\mathbb{R}\to[-1,1]$ are piecewise smooth, satisfy $k(\nu)=k(-\nu),\ k(0)=1,\ \int_0^\infty k(\nu)\nu d\nu<\infty,$ and have Parzen characteristic exponent q greater than or equal to 1 in equation (25).

Assumption 3 (basis functions). The orthonormal basis functions $\{\varphi_j\}$ on $L^2[0,1]$ have two bounded continuous derivatives, $j = 1, \ldots, B$, and they have implied mean kernels meeting Assumption 2.

Assumptions 1(a)-(c) provide general conditions under which the bias expressions and fixed-b distributions hold. Assumption 1(a) (or (b)) further implies that the Parzen (1957) generalized r^{th} derivative of the spectral density at frequency zero is finite for $r\le 2$:

$$\left|s_z^{(r)}(0)\right|\equiv\left|\frac{1}{2\pi}\sum_{u=-\infty}^{\infty}|u|^r\,\Gamma_u\right|<\infty,\quad r\in[0,2], \tag{28}$$

as will be useful below. The asymptotic expansions of the rejection rates in Velasco and Robinson (2001), Sun, Phillips and Jin (2008), and Sun (2013, 2014) further assume Gaussianity to simplify the calculations, and that additional condition is given in 1(c).[4]

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 two derivatives. The IM/step basis function does not satisfy this assumption, so it is 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 (18).

Theorems 1-5 provide our main theoretical results.

Theorem 1. Among kernel and orthonormal series tests computed using fixed-b critical values and having the same higher-order size:

[4] Velasco and Robinson (2001) provide an extension relaxing the Gaussianity assumption.

(i) The QS test has the greatest higher-order power among LRV estimators that are positive semidefinite with probability one, and

(ii) The higher-order power of tests with kernels (or, for orthonormal series, implied mean kernels) with Parzen characteristic exponent $q = 2$ asymptotically dominates that of tests with $q = 1$.

Theorem 2. Among kernel and orthonormal series tests computed using fixed-b critical values with the same higher-order size which in addition have fixed-b asymptotic distributions that are exact t (or, in the multivariate case, exact F), the EWP test achieves the maximum higher-order power.

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 $c_{1\alpha}(b_1)$ and $c_{2\alpha}(b_2)$. Fix b_1 and b_2 such that F_1^* and F_2^* have the same higher-order size. The difference between their higher-order power is then given by

$$\Pr_\delta[F_{1T}^* > c_{1\alpha}(b_1)] - \Pr_\delta[F_{2T}^* > c_{2\alpha}(b_2)] = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(\nu_2^{-1} - \nu_1^{-1}\right) + o(b) + o((bT)^{-q}). \quad (29)$$

Theorem 4. Let $c_{\alpha,T}(b)$ be the size-adjusted critical value of a test, that is, the critical value such that $\Pr_0[F_T^* > c_{\alpha,T}(b)] = \alpha$ to higher order. Then the size-adjusted power of the test is

$$\Pr_\delta[F_T^* > c_{\alpha,T}(b)] = \left[1 - G_{m,\delta^2}(\chi_m^\alpha)\right] - \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha \nu^{-1}$$
$$+ o(b) + o\left((bT)^{-q}\right) + O(\log T/\sqrt{T}). \quad (30)$$

Theorem 5. For orthonormal series estimators with basis functions satisfying Assumption 3,

(i) The asymptotic bias of $\hat{\Omega}^{BF}$ is of order $(B/T)^q$, where q is the Parzen characteristic exponent of the implied mean kernel;

(ii) The implied mean kernel is given by (18);

(iii) $k^{BF(1)}(0) = -B^{-1}\sum_{j=1}^{B} k_j^{BF\prime}(0)$, where $k_j^{BF\prime}(0) = 1-\frac{1}{2}\left(\phi_j(0)^2+\phi_j(1)^2\right)$. If $k^{BF(1)}(0) \neq 0$, then $q = 1$.

(iv) If $k^{BF(1)}(0) = 0$, then $q = 2$ and $k^{BF(2)}(0) = -\frac{1}{2}B^{-1}\sum_{j=1}^{B} k_j^{BF\prime\prime}(0)$, where $k_j^{BF\prime\prime}(0) = \int_0^1 \phi_j(s)\phi_j''(s)ds$.

Results (i)-(ii) also hold for the IM-basis function (which does not satisfy Assumption 3).

Remarks

1. Inspection of the expressions in the proof of the theorems show that these results hold uniformly over all stochastic processes satisfying Assumption 1 with $|s_z''(0)/s_z(0)| \leq \kappa$ for finite κ.
2. Theorem 1 says that the best test in the class of kernel and orthonormal series tests is the kernel test based on the QS kernel. The fixed-b distribution of the QS test is not an exact t or F, rather it is a weighted average of chi-squared random variables.
3. Theorem 2 says that, if one wishes to use exact t or F fixed-b critical values, then for any size distortion the EWP-based test achieves the greatest power.[5]
4. The price one must pay for using 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 (29), the power cost of using EWP relative to the higher-order best test with the same higher-order size is, neglecting the remainder terms,

[5] Any orthonormal series estimator for which $\int_0^1 \phi_j(s)\phi_k''(s) = 0$ for $j \neq k$ is also within this class of estimators that achieves maximum asymptotic power, as is evident in the form of the proof of Theorem 2 in the Appendix.

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

$$= \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi^\alpha_m)\chi^\alpha_m\left(1 - \frac{6\sqrt{3}}{5\sqrt{5}}\right)B^{-1}, \tag{31}$$

where $b_{EWP} = 1/B$ and the final expression is derived in the Appendix.

The function in the final expression in (31) is plotted in Figure 3 for $m = 1$ for various values of B. The maximum higher-order power loss from using EWP over all alternatives 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 power loss increases in m, for $B = 8$ it remains modest at roughly 0.02 even when testing $m = 4$ restrictions.

5. The expansions underlying these results can also be used to provide an expression for size-adjusted power, as given in Theorem 4. This expression shows that the size-adjusted power of a kernel or orthonormal series test depends solely on its equivalent degrees of freedom ν, and the difference between the size-adjusted power of two tests is again given by the expression on the right side of (29), this time for arbitrary ν_1 and ν_2. For orthonormal series tests with B bases, $\nu = B$, so all orthonormal series tests with the same value of B have the same size-adjusted power; however, given B, the EWP test has the smallest asymptotic size distortion (a result that follows from Theorem 2).
6. Like the kernel estimators, the bias of the orthonormal series estimators depends on the (generalized) derivative of the estimator at the origin, as in the proof of Theorem 5(i). Given a set of basis functions, these generalized derivatives can be calculated using the expressions in Theorem 5(iii)-(iv). It is shown in the appendix that: the Legendre basis function has $q = 1$; by a direct calculation not using Theorem 5, the IM-basis also has $q = 1$; and, surprisingly, the generalized first derivative at the origin of the two bases is the same, with $k^{Leg(1)}(0) = k^{IM(1)}(0) = (B+1)/B$. The Fourier and cosine bases both satisfy Assumption 2, and by calculations in the Appendix, are both $q = 2$ and are asymptotically equivalent to second order, with $k^{Fourier(2)}(0) = k^{cos(2)}(0) = \pi^2/6$. When the bias expansion is carried out to the next order, the next term is smaller for the cosine than the Fourier basis: $\frac{E\hat{\Omega}^{EWP} - \Omega}{\Omega}$ =

$\frac{s_z''(0)}{s_z(0)}\frac{\pi^2}{3}\left(\frac{B}{T}\right)^2\left(1+\frac{3}{B}\right)$, while $\frac{E\hat{\Omega}^{\cos}-\Omega}{\Omega} = \frac{s_z''(0)}{s_z(0)}\frac{\pi^2}{3}\left(\frac{B}{T}\right)^2\left(1+\frac{3}{2B}\right)$. This suggests that, in practice, the cosine basis might slightly outperform the Fourier basis, at least for B, T small.

7. The theorems consider the family of basis functions with two derivatives plus the IM basis function. We conjecture that other non-differentiable basis functions would have $q = 1$ and thus be dominated, but do not pursue this conjecture.
8. The discussion so far compares kernels and implied mean kernels for a given second-order size distortion. A different question is, given a kernel, what should B be? This is a difficult question because it involves a judgmental bias-variance tradeoff. Nevertheless, the asymptotic expansions in equation (26) and Theorem 4 permit quantifying the tradeoff between size and size-adjusted power for a given kernel. Figure 4 presents the higher-order size distortion and higher-order size-adjusted power for EWP with $m = 1$, $T = 200$, and $\alpha = 0.05$, for different data-generating processes across different values of B. Choosing B = 8 seems to provide a reasonable tradeoff between size and power even under moderate-to-high persistence (as encoded in $\left|s_z''(0)/s_z(0)\right|$), but this conclusion is preliminary, and we discuss the choice of B further in Section 7 below.
9. Equation (26) and Theorem 4 provide a pair of parametric equations in v which together constitute an envelope for the size/power tradeoff available using the family of kernel and orthonormal series tests. To obtain an explicit formula for this frontier, denote the normalized higher-order size distortion for an arbitrary test as

$$\widetilde{sd} = \frac{T^q}{tr\left(\sum_{j=-\infty}^{\infty}|j|^q\,\Gamma_j\Omega^{-1}\right)}\left\{\Pr_0\left[F_T^* > c_m^\alpha(b)\right]-\alpha\right\}. \tag{32}$$

Then using equation (26) and Theorem 4, along with the relation between b and ν, the higher-order size-adjusted power of the test (neglecting remainder terms) can be written immediately as

$$Pr_\delta[F_T^* > c_{\alpha,T}(b)] = \left[1-G_{m,\delta^2}(\chi_m^\alpha)\right]-\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(\frac{\widetilde{sd}\times m}{G'_m(\chi_m^\alpha)\chi_m^\alpha}\right)^{-1/q}\int_{-\infty}^{\infty}k^2(x)dx \tag{33}$$

in the case of a kernel estimator (and in the case of an orthonormal series estimator, we drop the $\int_{-\infty}^{\infty} k^2(x)dx$ term in this expression).

Given the result of Theorem 1, we can compute the frontier for the higher-order size-adjusted power vs. size distortion tradeoff by evaluating equation (33) for the QS test, which has $q = 2$ as above. We plot the resulting envelope in Figure 5 for $m = 1, 2, 3$ restrictions, and for nominal significance level of 0.05.[6] For reference in some leading cases, with $T = 200$, the value $T^2 / tr\left(\sum_{j=-\infty}^{\infty} j^2 \Gamma_j \Omega^{-1}\right)$ is equal to 10,000 in in a location model with error-term AR(1) coefficient of 0.5; 2,571 in the same model with AR(1) coefficient of 0.7; 8,182 in a 2-dimensional VAR(1) as in equation (34) below, with $A = \rho$I, $\rho = 0.5$, as in Monte Carlo design (a) below (and with MA parameter $\theta = 0$); and 2,070 in the same design with $\rho = 0.7$. The figure shows how the tradeoff between size-adjusted power and size can be quantified for any data-generating process using the theoretical results presented in this section.

5. Monte Carlo Design

The Monte Carlo analysis compares the size and power of hypothesis tests computed using various LRV estimators that differ both by kernel or basis function and by the bandwidth selection rule. The bandwidth selection rules correspond to minimum MSE, minimizing the coverage probability error (CPE), and fixed-*b* bandwidths (fixed *B*). The tests consider, alternatively, tests of the mean value, tests of a single stochastic regression coefficient, and tests of multiple regression coefficients. The tests are evaluated using standard normal and fixed-*b* asymptotic critical values.

The size of the tests is evaluated in six sets of Monte Carlo experiments: (a) contemporaneous regressions of y_t on x_t, where the data are generated according to vector autoregressions with varying degrees of persistence, in some cases with moving average errors (this design is the main benchmark in the HAC literature); (b) multi-step ahead direct forecasts with data generated by a VAR(1) with varying degrees of persistence and long-run correlation

[6] As in Figure 4, we choose the noncentrality parameter δ^2 for the alternative so that under first-order asymptotics, power is equal to 0.75 in the limit as $b \to 0$. In unreported results available on request, we also considered this tradeoff integrated over a range of noncentrality parameters delivering asymptotic power less than some threshold value. Results were very similar and were simply shifted vertically as compared to the frontiers shown in Figure 5.

between the errors and regressors; (c) multi-step direct regressions with long-run correlation and a persistent regressor as in the "predictive regression" literature (the Stambaugh (1999) bias literature); (d) multi-step ahead direct forecasts using data from a calibrated 3-variable VAR(1), where the VAR parameters are estimated using U.S. GDP growth, employment growth, and the commercial paper-Treasury bill spread; (e) a VAR with stochastic volatility; and (f) a data-sampling design in which the series are sampled from a 141-variable database of leading macroeconomic variables. We also compare the power of the tests in designs (a) and (f).

Estimators

1. Bartlett: Kernel estimator with time-domain Bartlett/Newey-West kernel, $k(v) = (1-|v|)\mathbf{1}(|v|\leq 1)$
    a. NW-MSE(.5): Minimum-MSE bandwidth selected using Andrew's (1991, eq. (5.3)) rule with AR(1) parameter fixed at $\rho = .5$. This yields the time-domain truncation parameter rule $S_T = .75T^{1/3}$ recommended in Stock and Watson (2013, eq. (15.17)).
    b. NW-CPE(.5): Bandwidth rule to minimize CPE at the 95% confidence level by minimizing the expression given in Sun, Phillips, and Jin (2008, eq. (34)), with AR(1) parameter fixed at $\rho = .5$, yielding the truncation parameter rule $S_T = 0.8747T^{1/2}$.
    c. NW-8: Truncation parameter $S_T = 3T/16$, corresponding to $b = 3/16$ and 8 degrees of freedom.
2. QS: Kernel estimator with QS/Bartlett-Priestley-Epanechnikov kernel.
    a. QS-MSE(.5): Minimum-MSE truncation parameter rule given by Andrews (1991), with AR(1) parameter fixed at $\rho = .5$, so $S_T = 2.3019T^{1/5}$.
    b. QS-CPE(.5): Minimum-CPE truncation parameter rule by minimizing the expression given in Sun, Phillips, and Jin (2008), with AR(1) parameter fixed at $\rho = .5$, so $S_T = 1.2909T^{1/3}$
    c. QS-plugin: Minimum-CPE truncation parameter rule by minimizing the expression given in Sun, Phillips, and Jin (2008), with AR(1) parameter estimated, so $S_T = 0.8132[2\hat{\rho}/(1-\hat{\rho})^2]T^{1/3}$.

d. QS-prewhiten: QS kernel applied to residuals from estimated VAR(1) for z_t, with AR(1) parameter estimated (Andrews-Monahan (1992), with minimum CPE bandwidth rule).

e. QS-8: Truncation parameter $S_T = T/8$, corresponding to $b = 1/8$ and 8 degrees of freedom.

3. KVB: The Kiefer-Vogelsang-Bunzel (2000) estimator using fixed-b asymptotic critical values.
4. Fourier : $\hat{\Omega}^{EWP}$ in (8), fixed B = 8,16, respectively denoted Fourier-8 and Fourier-16.
5. cos: Orthonormal series estimator $\hat{\Omega}^{\cos}$ computed using cosine basis functions (12), fixed B = 8, 16.
6. Legendre: Orthonormal series estimator $\hat{\Omega}^{Leg}$ computed using Legendre basis functions, fixed B = 8, 16.
7. IM-basis: Ibragimov-Muller orthonormal series estimator using (14), fixed B = 8, 16.

Monte Carlo Designs. This study uses six distinct Monte Carlo designs. The first three, a VARMA(1,1), an h-step ahead direct forecasting regression design, and a "predictive regression" are parametric with parameters chosen to match, or be similar to, designs commonly used in the literature. The fourth and fifth designs are parametric designs that are calibrated to U.S. data. The sixth design entails sampling series directly from a 140-variable U.S. macro/finance dataset, so that the correlation structure is the same as in the macro data, with a "flipped" data construction that separates the y and x data by 110 quarters thus arguably imposing the null of no linear predictability.

The VARMA model used to generate the data under the null in the first three experiments is,

$$\begin{pmatrix} y_t \\ \tilde{x}_t \end{pmatrix} = A\begin{pmatrix} y_{t-1} \\ \tilde{x}_{t-1} \end{pmatrix} + e_t,\ e_t = \eta_t + \theta\eta_{t-1},\ \text{where } \eta_t \text{ is i.i.d. } N\left(0,\Sigma_\eta\right), \tag{34}$$

where A and Σ_η are matrices and θ is a scalar.

(a) **Contemporaneous regression, size and power.** The Monte Carlo data (y_t, $\tilde{x}_t$) are generated according to (34) with $A = \rho I$ and $\Sigma_\eta = I$, so that y_t and $\tilde{x}_t$ are two independent Gaussian ARMA(1,1) processes and the null hypothesis is that $\beta = 0$. This is the leading benchmark design in the literature. The regression of interest is y_t on x_t, where $x_t = (1\ \ \tilde{x}_t)'$. In the case of testing the mean ($p = 1$), y_t is generated as an ARMA(1,1) with parameters ρ and θ. For $p = 2$, we consider the t test of the coefficient on $\tilde{x}_t$ and for $p > 2$, both the t test of the coefficient on the first stochastic regressor and the F-test of all the coefficients on $\tilde{x}_t$. An intercept is included in all regressions. Note that the persistence parameters ρ and θ are those of the dgp for y and x, not of $z_t = x_t u_t$. If x_t and u_t are AR(1) with parameter ρ, then z_t has an autocovariance function that declines as $\Gamma_u \propto \rho^{2|u|}$.

To compute power, data under a local alternative were generated as $\tilde{y}_t = \beta\tilde{x}_t + u_t$, where ($y_t$, $\tilde{x}_t$) are generated under the null hypothesis and β has the local parameterization, $\beta = b\left(\text{var}(\hat{\beta})\right)^{1/2}$, where var($\hat{\beta}$) is a function of the VAR design parameters and T.

Size is computed for $T = 200$ and power is computed for $T = 200$, 400, and 2000 for particular values of ρ.

(b) ***h*-step ahead forecasting, size.** The data (y_t, $\tilde{x}_t$) are generated by (34) with $A = \rho I$, $\Sigma_\eta = I$, and $\theta = 0$. The regression of interest is the cumulative h-step ahead direct forecasting regression,

$$y_{t+h}^h = \beta_0 + \beta_1 y_t + \beta_2 \tilde{x}_t + u_{t+h}^h \text{ , where } y_{t+h}^h = \sum_{s=1}^{h} y_{t+s} \tag{35}$$

and u_{t+h}^h is the error term in the h-step ahead regression. The coefficient of interest is β_2, so the test of $\beta_2 = 0$ is a test of whether $\tilde{x}_t$ is a predictor of y_{t+h}^h. Results are reported for $h = 4$ and 8. Size is computed for $T = 200$; no power results are computed for this design.

(c) ***h*-step ahead forecasts with long-run correlation, size.** The data (y_t, $\tilde{x}_t$) are generated by (34) with $A = \text{diag}(0, \rho)$, $\text{var}(\eta_{1t}) = \text{var}(\eta_{2t}) = 1$, and $\text{cov}(\eta_{1t},\eta_{2t}) = \chi$. The regression of interest is,

$$y_{t+h}^h = \beta_0 + \beta_1\tilde{x}_t + u_{t+h}^h \text{, where } y_{t+h}^h = \sum_{s=1}^{h} y_{t+s} \tag{36}$$

and u_{t+h}^h is the error term in the *h*-step ahead regression. This design is a benchmark in the "predictive regression" literature (e.g. Stambaugh, 1999) in which a persistent regressor is used to forecast a series that is a multi-period cumulation of martingale difference sequence under the null, where innovations to y_t and $\tilde{x}_t$ are correlated; for example, the dividend-yield might be used to predict future multi-month returns using monthly data. This design induces a long-run correlation between $\tilde{x}_t$ and u_t while maintaining $E(u_t|\tilde{x}_t) = 0$. The coefficient of interest is β_1, and results are reported for $h = 4$ and 8. Size is computed for $T = 200$; no power results are computed for this design.

(d) **Calibrated *h*-step ahead forecasts, size.** The VAR(1) parameters A and Σ_η , as well as a vector of intercepts, were estimated using GDP growth, manufacturing employment, and the 90-day corporate paper – 90-day Treasury bill spread. Random data were then generated using (34) with the addition of intercepts with these estimated parameters and $\theta = 0$. The random data were then used to estimate the direct forecasting regression (35), where the null values for the regression coefficients were computed from the VAR parameters. Results are reported for $h = 4$ and 8. Size is computed for $T = 200$; no power results are computed for this design.

(e) **Calibrated stochastic volatility, size.** The three time series in (c) were run through a particle filter and smoothing algorithm to fit the stochastic volatility model, $x_t = \rho_x x_{t-1} + \sigma_t\eta_t$, where $\ln(\sigma_t^2) = \mu + \varphi(\ln(\sigma_{t-1}^2)-\mu) + \tau\zeta_t$, where the innovations (η_t, ζ_t) are N(0,1). The estimated parameters were averaged to calibrate a stochastic volatility process ($\varphi = 0.8520$, $\tau = 0.1145$, $\mu = 0.4266$). Data were then generated using the

VAR(1) (34) with $A = \rho$I, Σ_η = I, and $\theta = 0$, with the modification that the errors have stochastic volatility with the estimated φ and μ parameters and with τ varying between 0 and 0.25 (just over twice the mean empirical value). For these simulations, $a = 0.7$ and $T = 200$. The stochastic volatility was implemented in two ways: (i) x_t has stochastic volatility, y_t does not, that is, $e_t = \text{diag}(1, \sigma_t)\eta_\tau$, and (ii) x_t and y_t have the same stochastic volatility path, that is, $e_t = \sigma_t\eta_\tau$,, where in both cases η_t are i.i.d. N(0,I). Because $\beta = 0$, (i) corresponds to x_t having stochastic volatility and the regression errors u_t being homoscedastic, and (ii) corresponds to x_t and u_t having the same stochastic volatility process. In both cases, the regression errors still satisfy $E(u_t|x_t) = 0$. Size is computed for $T = 200$; no power results are computed for this design.

(f) **"Macro data flipped-sampling" design, size and power.** The aim of this design is to draw from data that reflect those typically used in empirical macroeconomic applications. To that end, the design entails sampling (y_t, $\tilde{x}_t$) with replacement from the database of 141 macroeconomic quarterly time series in Stock and Watson (2015), with $T = 220$ quarterly observations. The time series are mainly obtained from the Federal Reserve Bank of St. Louis FRED database. These data were transformed to approximate stationarity by first differencing and by low-frequency detrending using the methods described in Stock and Watson (2012). Because the projection of randomly drawn y_t on randomly drawn $\tilde{x}_t$ will in general have a nonzero population coefficient, the following sampling scheme was adopted to impose the null that $\beta = 0$: (i) draw (y_t, $\tilde{x}_t$) at random from the data set; (ii) switch the first and second halves of the observations on $\tilde{x}_t$ to produce the regressor $\tilde{\tilde{x}}_t = (\tilde{x}_{T/2+1},...,\tilde{x}_T,\tilde{x}_1,...,\tilde{x}_{T/2})'$; (iii) randomly choose either the first or second half of the (y_t, $\tilde{\tilde{x}}_t$) observations, for 110 observations on (y_t, $\tilde{\tilde{x}}_t$). Under the assumption that the series have short-range dependence, we would expect that y_t cannot be linearly predicted given observations on $\tilde{x}_t$ separated by 110 quarters (27.5 years), so $E(y_t|\tilde{x}_{t-T/2}) = E(y_t|\tilde{x}_{t+T/2}) = 0$, that is, $E(y_t|\tilde{\tilde{x}}_t) = 0$. In the case of multiple regressors, the first three lags of $\tilde{x}_t$ were included in the flipping procedure and the random selection of one half of the data. Given the

110-observation sample on (y_t, $\tilde{\tilde{x}}_t$), three sets of regressions were run: (1) the regression of y_t on $\tilde{\tilde{x}}_t$; (2) the regression of y_t on , $\tilde{\tilde{x}}_{t-1}$, $\tilde{\tilde{x}}_{t-2}$, $\tilde{\tilde{x}}_{t-3}$; and (3) h-step ahead forecasting regressions in which $y_{t+1}+\ldots+y_{t+h}$ is regressed on (y_t, $\tilde{\tilde{x}}_t$) for $h = 4$ and 8. Finally, this entire exercise was repeated using HP-detrending instead of the Stock-Watson (2012) low frequency detrending and first differencing described above.

The data-based design was also used to compute size-adjusted power curves. Data under the alternative were generated by standardizing (y_t, $\tilde{\tilde{x}}_t$) (a step that is irrelevant for the size calculations) and constructing $\tilde{y}_t = \beta\tilde{\tilde{x}}_t + y_t$, where β is scaled to be in local parameterization units assuming $\Omega = 1$, so $\beta = T^{-1/2}b$ since (y_t, $\tilde{\tilde{x}}_t$) are standardized.

Inference. Inference is conducted using, alternatively, asymptotic normal/chi-squared critical values and fixed-b critical values. For the orthonormal series estimators and IM, fixed-b inference was conducted using t_B or $F_{q,B-q}$ critical values, using the t statistic and, for $m > 1$, the F_T^* statistic in (24). For tests using the QS and Bartlett kernel estimators, fixed-b critical values were computed using Sun's (2014, Theorem 4) F approximation to the fixed-b distribution. Tests using the KVB estimator are only evaluated using fixed-b critical values.

6. Monte Carlo Results

This section first describes the content of the tables and figures, then provides a series of summary findings that pool across the different designs and results.

Content of tables and figures. Tables 2-12 report size results. Tables 2-6 report results for design (a). Table 2 presents the null rejection rates for tests of the mean for VAR(1) data with AR coefficient $\rho = 0$, .5, .7, .9, and .95, at the nominal 5% and 10% significance level. Table 3 presents results for tests of a single stochastic regression coefficient ($p = 3$, $m = 1$), and Table 4 presents results for tests on a two stochastic regression coefficients ($p = 3$, $m = 2$), in both cases for VAR(1) data with various common AR parameters. The rejection rates in Tables 2a and 3a are computed using asymptotic normal critical values, and in Tables 2b and 3b are computed using fixed-*b* critical values; for all remaining tables, only fixed-*b* critical values are used. Tables 5 and 6 parallel Tables 3b and 4, respectively, where the data are generated by a VARMA(1,1) with AR parameter $\rho = .7$ and various MA parameters θ.

Size for the *h*-step ahead forecasts (design (b)) are presented in Table 7a for $h = 4$ and Table 7b for $h = 8$, where the data are generated by a bivariate VAR(1) with varying AR parameters. The reported rejection rates are for the *t*-test on the coefficient on the extra predictor ($\tilde{x}_t$ in (35)).

Table 8 presents results for the "predictive regression" design (c), in which y_t is serially uncorrelated, $\tilde{x}_t$ has a moderate autocorrelation parameter of .7, and regressions are 4- and 8-step ahead (left and right panel, respectively). The test is on the coefficient on $\tilde{x}_t$. The columns present null rejection rates (5% nominal level only) as the correlation between the innovations in y_t and $\tilde{x}_t$ is increased.

Tables 9 presents results for data generated by the VAR with coefficients estimated using macro data (design (d)). The first four columns pertain to 4-step ahead forecasts and the second four pertain to 8-step ahead forecasts. Within each set of forecasts, the first two columns test the coefficient on only the simulated manufacturing employment variable, and the second two jointly test the coefficients on both employment and the simulated spread variable.

Table 10 reports size for the calibrated stochastic volatility model (design (e)), in which $0 \leq \tau \leq 0.25$ (recall that the empirical calibrated value for the three macro series is $\tau = 0.1145$). The first block of columns are for the design in which x_t but not y_t has stochastic volatility, and in the second block x_t and y_t follow the same stochastic volatility process.

Tables 11 and 12 present results for the "flipped" macro data design (design (f)). In all cases, rejection rates are for the coefficient on the stochastic regressor(s), where the regression includes an intercept. The regressions in Table 11 are of y_t on flipped x_t and an intercept. The first four columns of Table 9 provide results for data that have been detrended by first differencing with subsequent low-frequency detrending; the final four columns are for data that have been detrended using a HP filter. Table 12 presents results for cumulative *h*-step ahead forecasting regressions ($h = 4, 8$), where the test is on the single coefficient on x_t in a regression of the cumulative value of y_t from t+1 to t+h on an intercept, y_t, and x_t.

Size-adjusted power results are presented graphically for a subset of the LRV estimators. Figure 6 presents results for the VAR(1) design (a), and Figure 7 presents results for the macro data flipped-sampling design (f). All the alternatives are local, with the variance of x_t normalized to 1, so the horizontal axis is comparable across the figures. Size adjustment is done by using the 5% quantile of the Monte Carlo distribution of test statistics under the null hypothesis for the given design.

Finally, Figure 8 compares size and size-adjusted power results from the Monte Carlo experiments to the QS-based theoretical frontier using a similar calculation to that used in Figure 5. The data are generated according to the VAR(1) design (a) and correspond to the tests in Table 3a (for NW-MSE(.5)) and Table 3b (for all other estimators) for the size results, and in Figure 6 for the size-adjusted power results, corresponding to the point at which the power of the test when the LRV is known (i.e., the asymptotic Gaussian curve) is equal to 0.75. We consider two parameterizations: one with moderate sample length and moderate VAR(1) persistence ($T = 200$, $\rho = 0.5$), and one with longer sample length and high VAR(1) persistence ($T = 400$, $\rho = 0.9$).

Discussion. Looking across the results for the different LRV estimators, we have nine main findings, the first four of which echo and confirm those in the large post Newey-West/Andrews literature.

First, in almost all cases, the worst size performance is when normal or chi-squared critical values are used with the NW-MSE(.5) kernel estimator, followed by the QS-MSE(.5) estimator.

Second, meaningful size improvements obtain from using fixed-*b* asymptotics instead of asymptotically normal critical values. The clearest example of this is comparing the results for the orthonormal series estimators in the case of estimating the mean in Table2a (normal critical values) and Table 2b (t_B critical values). In the case $\rho = 0$, these tests are exactly t_B distributed and using the t_B critical values improves size from roughly .08 for $B = 8$ to within Monte Carlo error of .05 (nominal 5% tests). For testing the mean in the AR case with $\rho = .9$, size is improved from .145 to .094 for Fourier-8. Comparable improvements from using fixed-*b* asymptotics are also seen in the case of testing a stochastic regressor (Table 2a v. Table 2b).

Third, in the parametric designs, the performance of NW-MSE and QS-MSE are improved by using the larger time-domain bandwidths produced by the minimum-CPE rule. The performance of the QS-plugin and QS-prewhiten tests varies across designs. Not surprisingly, in the AR(1) and VAR(1) designs, the QS-prewhiten (which supposes a global AR(1) model) performs quite well, comparably to or even better than the KVB and orthonormal series estimators. This advantage largely disappears, however, in the VARMA designs. In some cases with high non-AR persistence, the QS-plugin estimator performs quite poorly (e.g. Table 6). Even with these modifications, the sizes of the NW and QS tests generally exceed the size of the KVB and orthonormal series tests, sometimes by a substantial margin.

Fourth, for tests on stochastic regressors, the orthonormal series estimators with fixed B = 8 perform very well, matching or beating the size of the KVB estimator as well as the NW and QS MSE, CPE, plugin, and prewhiten variants.

Fifth, the theoretical predictions in Section 3 about comparing size distortions of the Fourier, cosine, Legendre and IM orthonormal series estimators are largely borne out in the parametric designs (Tables 2-10). The cosine basis function was predicted to have a slight edge in size distortion over the Fourier basis function in finite samples, however the two are in most cases within Monte Carlo error of each other. Both the Fourier and cosine basis functions were predicted to outperform the Legendre and IM basis functions, and that prediction is borne out (e.g. compare the B= 8 results for each). Moreover, the size distortions are modest, and arguably acceptable for empirical work, for moderate to large values of ρ. For example, for the test of a single coefficient with ρ = .7, the NW-MSE(.5) rejection rate is .116 while the Fourier-8 rejection rate is .070 at the 5% level. Even for ρ = .9, the Fourier-8 rejection rate remains .121 in the VAR(1) design (Table 3b). In the 4-step ahead regressions with VAR(1) data, the Fourier-8 rejections are .107 at 5% for h = 4 (Table 7a) and .125 for h = 8 (Table 7b), even for values of ρ up to 0.95.

Sixth, turning to size-adjusted power in the parametric settings (Figure 6), tests based on NW-MSE and QS-MSE have the greatest power, followed by the orthonormal series estimators (Fourier-8, cos-8, and Legendre-8, IMbasis-8) and QS-8 and NW-8, with no systematic ordering among these estimators, followed by KVB. The NW-MSE and QS-MSE tests have power approaching the asymptotic power envelope, for which Ω is known. These findings are consistent with the power based on asymptotic normality (large B) exceeding power based on the

t_B distribution, and both exceeding power based on the KVB distribution. It also bears out the prediction that, with equivalent effective degrees of freedom, QS-8 will perform as Fourier-8 and NW-8 as IM-8. This pattern is evident in all the configurations of T and ρ considered. These findings underscore the general tension between reducing size distortions (large S_T) and increasing size-adjusted power (small S_T) pointed out by Kiefer and Vogelsang (2005).

Seventh, the size and power performance of the QS-8 test is similar to Fourier-8, and the size and power of NW-8 is similar to IM-basis-8 (and to Leg-8). This also accords with the theory laid out above. The QS-8 and NW-8 tests set $b = 1/8$, which corresponds to $S_T = 25$ with $T = 200$, much larger than the $S_T = 7$ value used in NW-MSE(.5). The similarity of QS-8 to Fourier-8, and of NW-8 to IM-basis-8, also underscores that the most important driver of the size improvements is the use of a large time-domain truncation parameter and fixed-b asymptotics, not the choice of the kernel. This said, the results do confirm that the $q = 2$ kernels tend to perform somewhat better than the $q = 1$ kernels (actual kernels or implied mean kernels).

Eighth, the size and power results for the “flipped” macro data regressions have some similarities to, but some important differences from, the results for the parametric models. For the first-differenced data, the size distortions are all moderate for a single coefficient, perhaps not surprisingly because first differencing and low-frequency detrending has eliminated most of the persistence in these data. At least with these small size distortions, all the tests exhibit comparable rejection rates (Table 11, first two columns). The size distortions with the HP-filtered regressions are considerably larger, and here many of the patterns in the parametric results reemerge: the NW and QS estimators have substantially higher rejection rates than KVB or the orthonormal series estimators; the Fourier-8 rejection rate is comparable to KVB; the Legendre rejection rates are somewhat greater than the cos-8 and Fourier-8 rates; and the size-adjusted power of the orthonormal series tests falls between the QS and NW tests and the KVB test. Still, there are some noteworthy differences: the Fourier-8 test seems generally to have lower rejection rates than cos-8; the size improvements of the Fourier-8 and cos-8 tests are less than would be expected in the h-period ahead regressions, compared with the parametric design (note that one issue could be that these regressions include y_t but no lags so do not suitably account for dynamics in y_t); and the power of the orthonormal series tests in Figure 7 is substantially closer to the QS/NW power.

Ninth, the numerical results here confirm that the higher-order approximations used in Sections 3 and 4 perform reasonably well in Gaussian designs with moderate persistence or large sample sizes,[7] as illustrated by Figure 8. As expected, the theoretical QS frontier dominates the Monte Carlo results for all estimators, but the EWP estimator yields results close to the frontier and with small size distortions. Finally, while the EWP estimator theoretically dominates NW (a result that follows from Theorem 1(ii), as above) and exhibits greater size control in both sets of numerical results, the numerical results are sufficiently close (particularly for the parameterization with shorter sample length and lower persistence) that one could be justified in using NW instead and conducting inference using the F approximation to the fixed-b distribution.

7. Discussion and Conclusions

By combining new theoretical results with previous results from the associated literature, we characterize optimal HAR tests using fixed-b critical values. We find that tests based on a long-run variance estimator constructed using the QS kernel perform optimally within the class of kernel and orthonormal series estimators. However, restricting attention to tests admitting exact fixed-b t- and F-distributions entails only modest sacrifice with respect to higher-order power, controlling for higher-order size; further, within this class of tests, the test using the equal-weighted periodogram LRV estimator performs optimally. We confirm these results in extensive Monte Carlo experiments, which we also extend to include tests outside the classes considered in the theoretical results developed here. These experiments further 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, as in previous literature.

The question remains as to how to choose the bandwidth parameter B for the EWP test. The leading cases considered here use $B = 8$ (i.e., averaging over the first four periodogram ordinates) and t_8 and $F_{m,8-m+1}$ critical values, where m is the number of restrictions being tested,

[7] The approximations perform more poorly in the presence of highly-persistent series with low sample sizes. As discussed in the Introduction, however, this is not our main application of interest, given the documented failure of LRV estimators to control size with highly-persistent series (as in Müller (2014)). Results for other parameterizations are available upon request.

and this parameterization performs well both theoretically and in Monte Carlo experiments with reasonable sample lengths. A more data- or test-dependent rule, however, may ultimately be more desirable; for example, one may wish to select B as a function of the sample size T, the sample *span* (e.g., T normalized by the number of observations per year), and/or the number of restrictions m.[8] But our view is that calculating rejection rates obtained without controlling size makes the results of tests difficult to interpret, and we thus have not yet considered selection rules obtained by directly trading off size and unadjusted power. We hope that the work begun here will stimulate a focused discussion on issues of practical implementation, so that empirical practice can move beyond currently-standard methods of HAR inference and take advantage of advances in the previous 20 years.

[8] For example, setting $B = 2ceil((m+4)/2)$ ensures that the approximating F-distribution $F_{m,B-m+1}$ has finite variance. We thank Yixiao Sun for this suggestion.

Appendix

Proof of Equations (26) and (27): For kernel estimators, given Assumptions 1 and 2, equation (26) follows from Sun (2014) equation (16), along with the approximation $\frac{m\mathcal{F}_{\infty}^{\alpha}(m,b)-\chi_m^{\alpha}}{\chi_m^{\alpha}}=O(b)$ 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 (27) follows directly from Sun (2014) Theorem 5.

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 $B/T\to 0$. The proof of Theorem 5, which does not use this result, shows that this limiting bias is given by $A=(B/T)^q k^{BF(q)}(0)\left[-2\pi s_z^{(q)}(0)\right]+o\left((B/T)^q\right)$. Using $B=1/b$ yields equation (26). 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 $\left|s_z''(0)/s_z(0)\right|\leq$ κ for finite κ, as in Remark 2.

For equation (27), the above steps carry through for the bias term, and the variance term follows directly from Theorem 5 of Sun (2011), with $\nu=B$ as in equation (22). The remainder terms follow from this theorem as well. □

Proof of Theorem 1: (i) Note that as after equation (8), 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 implied mean kernel) must be no greater than q = 2. Thus we can confine this analysis to kernels (or implied mean kernels) with $q\leq 2$.

In order to fix higher-order size at a common value across all estimators in order to compare higher-order power, we can arbitrarily fix a value for b^{QS}. The QS estimator has q = 2;

see Priestley (1981), Table 7.1 (as well as for all other values of q and $k^{(q)}$ given below). Thus for any alternative estimator (denoted "*alt*") with q = 2, equation (26) gives that

$$b^{alt} = \sqrt{\frac{k^{(2),alt}}{k^{(2),QS}}} b^{QS}, \tag{37}$$

and $B^{alt} = 1/b^{alt}$ in the case that the alternative estimator is an orthonormal series estimator. Theorem 3, whose proof (see below) does not rely on this result, then allows us to compare the tests' higher-order power.

We first confine the analysis to kernel estimators. It is apparent from equation (29) that maximizing higher-order power, given equivalent higher-order size, is equivalent to maximizing ν, which is equal to $\nu = (b\int_{-\infty}^{\infty} k^2(x)dx)^{-1}$ as in (21). Given the form of b^{alt} above, we thus aim to minimize $\sqrt{k^{(2),alt}} \int_{-\infty}^{\infty} k^2(x)dx$. 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*. This value is minimized exactly by the QS estimator, as proven in Priestley (1981, p. 571).

This proves the optimality of the QS test among kernel estimators with q = 2. The optimality of QS as compared to orthonormal series estimators with *q=2* follows from Theorem 2 and equation (31) (whose proofs do not rely on Theorem 1), which tell us that (i) the EWP test achieves maximum higher-order power among orthonormal series tests, and (ii) the higher-order power of the QS test is greater than that of EWP. So the proof must be completed by showing that estimators with q = 2 asymptotically dominate those with q = 1, as in part (ii) of the theorem.

(ii) For an arbitrary estimator with characteristic exponent equal to 2, select a sequence for $b^{q=2}$ such that $b^{q=2} \to 0$, $b^{q=2}T \to \infty$, and $\left(b^{q=2}\right)^2 T \to 0$. We will show that q = 2 estimators dominate q = 1 estimators along this sequence.[9] For such an alternative estimator with q = 1, equation (26) now gives that in order to equalize the estimators' higher-order size, we must set

$$b^{alt} = T\left(b^{q=2}\right)^2 \frac{\mu_0^{alt}(\alpha,m)k^{(1)}(0)}{\mu_0^{q=2}(\alpha,m)k^{(2)}(0)}, \tag{38}$$

which implies from the sequence construction above that $b^{alt} \to 0$ while $S_T^{alt} = b^{alt}T \to \infty$. Using the value in (38) in the second term in equation (27), we obtain

$$\begin{aligned} \mu_{\delta^2}^{alt}(\alpha,m)k^{(1)}(0)(b^{alt}T)^{-1} &= \frac{\mu_{\delta^2}^{alt}(\alpha,m)\mu_0^{q=2}(\alpha,m)}{\mu_0^{alt}(\alpha,m)}k^{(2)}(0)(b^{q=2}\,T)^{-2} \\ &= \mu_{\delta^2}^{q=2}(\alpha,m)k^{(2)}(0)(b^{q=2}T)^{-2}, \end{aligned} \tag{39}$$

using the definition of μ_0 and μ_{δ^2} after (26) and (27), respectively, which gives that the bias terms in the higher-order-power comparison between *alt* and $q = 2$ again cancel. This leaves us with

$$\begin{aligned} &\Pr_\delta[F_T^{q=2} > c_\alpha^{q=2}(b^{q=2})] - \Pr_\delta[F_T^{alt} > c_\alpha^{alt}(b^{alt})] \\ &\quad = \frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left((\nu^{alt})^{-1} - (\nu^{q=2})^{-1}\right) + o((bT)^{-q}) + o(b), \end{aligned} \tag{40}$$

and using $\nu^{-1} \propto b$ and plugging the value for b^{alt} and $b^{q=2}$ from above into this expression yields that the higher-order power gap must be positive for large enough T, completing the proof. □

Proof of Theorem 2: As in equations (19) and (22), only orthonormal series estimators yield fixed-*b* asymptotic distributions that are exact *t* (or exact *F* in the multivariate case). (As discussed after (20), 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

---

[9] We must construct such a sequence so that the expansions in equations (26) and (27) still holds for the alternative q = 1 estimator. We thank Yixiao Sun for pointing this out and suggesting an argument similar to the one used here.

orthonormal series representation using the Fourier basis functions.) Thus we aim to maximize higher-order power, given equivalent higher-order size, among orthonormal series estimators.

Given Theorem 1(ii), we need only consider estimators for which the implied mean kernel has $q = 2$. Fix a value B_{EWP}. Theorem 3 (whose proof does not rely on Theorem 2) gives that the higher-order power gap between any alternative estimator *alt* (with $q = 2$) and EWP is given by $\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(\nu_{alt}^{-1}-\nu_{EWP}^{-1}\right)=\frac{1}{2}\delta^2 G'_{(m+2),\delta^2}(\chi_m^\alpha)\chi_m^\alpha\left(B_{alt}^{-1}-B_{EWP}^{-1}\right)$, where the equality follows from $\nu = B$ for orthonormal series estimators in equation (22). 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.

From equation (26), having equivalent higher-order size requires setting

$$B_{alt} = \sqrt{\frac{k_{EWP}^{(2)}(0)}{k_{alt}^{(2)}(0)}}\, B_{EWP}. \tag{41}$$

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 basis functions satisfying Assumption 3. From Theorem 5 (which does not rely on Theorem 2), this requires that for any given value of B, it must be the case that $k_{EWP}^{(2)}(0)$ minimizes $\left|B^{-1}\sum_{j=1}^{B} k_j''(0)\right|$, where $k_j''(0)=\int_0^1 \phi_j(s)\phi_j''(s)\,ds$, across orthonormal series estimators.

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

$$\phi_j(s) = \sum_{l=1}^{\infty} a_{jl} e^{-i2\pi ls}, \tag{42}$$

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

$$
\begin{aligned}
1 &= \int_0^1 \left|\phi_j(s)\right|^2 ds \\
&= \sum_l \sum_{l'} a_{jl} a_{jl'} \int_0^1 e^{-i2\pi ls} e^{i2\pi l's} \\
&= \sum_l a_{jl}^2,
\end{aligned} \tag{43}
$$

and similarly

$$
0 = \int_0^1 \phi_j(s)\overline{\phi_{j'\neq j}(s)}\, ds = \sum_l a_{jl} a_{j'\neq j,l}. \tag{44}
$$

We then wish to minimize

$$
\begin{aligned}
\left|\frac{1}{B}\sum_{j=1}^{B}\int_0^1 \phi_j(s)\overline{\phi_j''(s)}\, ds\right| &= \left|\frac{1}{B}\sum_{j=1}^{B}\sum_l \sum_{l'} a_{jl} a_{jl'} 4\pi^2 l^2 \int_0^1 e^{-i2\pi ls} e^{i2\pi l's}\, ds\right| \\
&= \frac{4\pi^2}{B}\sum_{j=1}^{B}\sum_l a_{jl}^2 l^2,
\end{aligned} \tag{45}
$$

subject to the two constraints (43) and (44). 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 (42), $\phi_j(s) = e^{-i2\pi js}$, so that we have in fact selected the Fourier basis functions themselves. Further, given that, as above, all basis functions 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.e., that the Parzen characteristic exponent of a kernel or implied mean kernel must be no greater than $q = 2$). We conclude that the EWP test achieves the minimum asymptotic bias among the class of orthonormal series estimators (see Theorem 5), and therefore given the argument above, the EWP test achieves the maximum higher-order power among kernel and orthonormal series tests with the same higher-order size which in addition have fixed-b asymptotic distributions that are exact t (or, in the multivariate case, exact F). □

Proof of Theorem 3: This follows directly from equations (26) and (27). Fix b_1 for test F_1. Given equivalent values of q for tests F_1 and F_2, equation (26) 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, similar to the expression given in the proof of Theorem 2. We thus have that

$$\mu_{\delta^2}(\alpha,m)k_2^{(q)}(0)(b_2T)^{-q} = \mu_{\delta^2}(\alpha,m)k_1^{(q)}(0)(b_1T)^{-q}, \tag{46}$$

so that the corresponding second terms in the power expression (27) for F_1 and F_2 are equivalent. Using this along with equation (27) yields the desired relation. □

Proof of Theorem 4: We can define the size-adjusted critical value as $c_{\alpha,T}(b) = c_m^\alpha(b) + \delta_{\alpha,T}$, where $c_m^\alpha(b)$ is the fixed-b critical value as in equation (26) and $\delta_{\alpha,T}$ is defined implicitly so that $\Pr_0[F_T^* > c_{\alpha,T}(b)] = \alpha$ to higher order. Taking a Taylor expansion around $c_m^\alpha(b)$, $\Pr_0\left[F_T^* > c_{\alpha,T}(b)\right] = \alpha + \mu_0(\alpha,m)k^{(q)}(0)(bT)^{-q} + \delta_{\alpha,T}G_m'(\chi_m^\alpha) + o(b) + o\left((bT)^{-q}\right)$, from which we can solve for $\delta_{\alpha,T}$ to higher order as $\delta_{\alpha,T} = k^{(q)}(0)(bT)^{-q}\chi_m^\alpha tr\left(\sum_{j=-\infty}^{\infty}|j|^q \Gamma_j \Omega^{-1}\right)/m$. Then taking a similar Taylor expansion, size-adjusted power is

$$\begin{aligned}\Pr_\delta\left[F_T^* > c_{\alpha,T}(b)\right] = \left[1 - G_{m,\delta^2}(\chi_m^\alpha)\right] + \mu_{\delta^2}(\alpha,m)k^{(q)}(0)(bT)^{-q} - \delta^2\psi_{\delta^2}(\alpha_m)\nu^{-1} + \delta_{\alpha,T}G'_{m,\delta^2}(\chi_m^\alpha) \\ + o(b) + o\left((bT)^{-q}\right) + O(\log T/\sqrt{T}).\end{aligned} \tag{47}$$

We have $\delta_{\alpha,T}G'_{m,\delta^2}(\chi_m^\alpha) = k^{(q)}(0)(bT)^{-q}G'_{m,\delta^2}(\chi_m^\alpha)\chi_m^\alpha tr\left(\sum_{j=-\infty}^{\infty}|j|^q \Gamma_j \Omega^{-1}\right)/m =$ $\mu_{\delta^2}(\alpha,m)k^{(q)}(0)(bT)^{-q}$ from the solution for $\delta_{\alpha,T}$ above and the definition of $\mu_{\delta^2}(\alpha,m)$. Thus the second and fourth terms in (47) cancel. Plugging in the definition of $\psi_{\delta^2}(\alpha_m)$ yields the relation given in equation (30). □

Proof of Theorem 5: (i) As in Phillips (2005), for these results we consider sequences under which $B \to \infty$ such that $\frac{T^{q-1}}{B^q} + \frac{B}{T} \to 0$. (This of course implies the condition that $B/T \to 0$ as in Sun, 2013, but this more stringent rate condition will insure that the dominant bias is $O\left((B/T)^q\right)$ even when q = 2.) Using the implied mean kernel representation for $E\hat{\Omega}^{BF}$ in equation (17), we can follow Priestley (1981, p. 459) and write

$$\begin{aligned} E\hat{\Omega}^{BF} - \Omega &= \sum_{u=-(T-1)}^{T-1} \{k_T^{BF,B}(u/S)(1-|u/T|)-1\}\Gamma_u - \sum_{|u|\geq T}\Gamma_u \\ &= \sum_{u=-(T-1)}^{T-1} \{k_T^{BF,B}(u/S)-1\}\Gamma_u - \frac{1}{T}\sum_{u=-(T-1)}^{T-1} |u|\, k_T^{BF,B}(u/S)\Gamma_u - \sum_{|u|\geq T}\Gamma_u . \end{aligned} \tag{48}$$

For the last term,

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

by Assumption 1(b) and under the assumed asymptotic sequence.

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

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

where the last equality holds by the rate condition introduced above. Note that in general, even for different sequences than the one considered here, this $O(1/T)$ term is $o(b)$ for $b = 1/B$, yielding the $o(b)$ term in (26) in the orthonormal series case.

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

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

$$= (B/T)^q k^{BF(q)}(0)\left[-2\pi s_z^{(q)}(0)\right]\left\{1+o(1)\right\}, \tag{51}$$

where the second equality holds by the definition $SB = T$ and by Parzen (1957) Theorem 5B or Priestley (1981, p. 459). This yields the desired result. Further, we did not require differentiability of the basis functions in order to obtain this result, so that this holds as well for the IM-basis function.

(ii) This follows directly from equations (16) and (17) as $T \to \infty$.

(iii) The condition that $k(0)=1$ in Assumption 2 gives that $k^{(1)}(0) = -k'(0)$ for any kernel or implied mean kernel. Using equation (18), for $v > 0$,

$$k_j^{BF\prime}(v) = \frac{k_j(v)}{1-v} - \frac{\phi_j(v)\phi_j(0)}{(1-v)} - \frac{1}{(1-v)}\int_v^1 \phi_j(s)\phi_j'(s-v)ds \tag{52}$$

so

$$k_j^{BF\prime}(0) = k_j(0) - \phi_j(0)^2 - \int_0^1 \phi_j(s)\phi_j'(s)ds = 1 - \frac{1}{2}\left(\phi_j(0)^2 + \phi_j(1)^2\right), \tag{53}$$

where the final expression uses $k_j(0) = \int_0^1 \phi_j(s)^2 ds = 1$ and integrates by parts. (The same expression for $k_j'(0)$ obtains starting from the expression for $k(v)$, $v \le 0$.) Then as in Priestley (1981, p. 460), it is apparent from the definition of the generalized derivative that if $0 < k^{(1)}(0) < \infty$, then $\frac{1-k(v)}{|v||v|} \to \infty$ as $v \to 0$, so that $q = 1$.

(iv) As in Theorem 1, we must have $q \le 2$, so following the result just above, if $k^{BF(1)}(0) = 0$, then $q = 2$. Priestley (1981, p. 460) 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 $\nu > 0$, differentiating the expression in part (iii) again yields

$$k_j^{BF''}(\nu) = \frac{2k_j^{BF'}(\nu)}{1-\nu} + \frac{\phi_j(\nu)\phi_j'(0) - \phi_j'(\nu)\phi_j(0)}{1-\nu} + \frac{\int_\nu^1 \phi_j(s)\phi_j''(s-\nu)\,ds}{1-\nu}, \tag{54}$$

giving that

$$k_j^{BF''}(0) = \int_0^1 \phi_j(s)\phi_j''(s)\,ds, \tag{55}$$

and again the same expression obtains starting from $\nu \le 0$. This completes the proof. □

Derivation of Equation (31): Fix a value $B^{-1} \equiv b^{EWP}$. From (37), using the values in Priestley (1981), Table 7.1, we have

$$b^{QS} = \sqrt{\frac{k^{(2),QS}(0)}{k^{(2),EWP}(0)}}\, b^{EWP} = \sqrt{\frac{\pi^2/10}{\pi^2/6}}\, B^{-1} = \sqrt{\frac{3}{5}}\, B^{-1}, \tag{56}$$

Priestley (1981) Table 6.1 also gives that $\int_{-\infty}^{\infty} k^{2,QS}(x)\,dx = \frac{6}{5}$, and $\int_{-\infty}^{\infty} k^{2,EWP}(x)\,dx = 1$, 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 (29) yields the desired result. □

Derivations for Remark 6: It follows from Theorem 5 that $k_j^{BF'}(0) = 0$ if $\varphi_j(1)^2 + \varphi_j(0)^2 = 2$. For bases comprised of even and odd functions, including the Fourier, cosine, and Legendre bases, $k^{BF'}(0) = 0$ if $\varphi_j(1) = |1|$ for $j = 1,\ldots, B$. This is true for the Fourier and cosine bases. For the Legendre basis, however, after translating the Legendre functions from [-1,1] to [0,1] and normalizing so that $\int_0^1 \phi_j(s)^2 ds = 1$, $\varphi_j(1) = \sqrt{2j+1}$, so that $k^{Leg(1)}(0) = (B+1)/B$. As in Theorem 5, this also implies that $q = 1$ for the Legendre polynomials, and $q = 2$ for the Fourier and cosine bases.

For IM-basis, a direct calculation gives that the IM implied mean kernel is the Bartlett kernel on a subsample of $T/(B+1)$ observations, so that $k^{IM}(v) = (1 - |v|)\mathbf{1}(|v|\leq 1)$. Because k^{BF} is defined with the degrees of freedom B in the denominator, not $B+1$, the generalized first derivative of $k^{IM}(v)$ is $k^{IM(1)}(v) = (B+1)/B$, not 1 which is the usual expression for the Bartlett kernel.

The fact that $k^{Fourier(2)}(0) = \pi^2/6$ follows directly from the Priestley (1981) Table 7.1 entry for the Daniell kernel. For $k^{cos(2)}(0)$, first note that Theorem 5(iv) gives that $k^{(2)}(0)$ depends on $\int_0^1 \phi_j(s)\phi_j''(s)ds$; second, that this value is asymptotically equivalent for the cosine basis functions and for the Phillips (2005) suggestion of $\{\sqrt{2}\sin[\pi j(s-1/(2T))]\}$, $j = 1,\ldots, B$; third, that Theorem 1(i) of Phillips (2005) gives that with these sine basis functions, the limiting bias $\lim_{T\to\infty} E\hat{\Omega}-\Omega$ is equal to $(B/T)^2\frac{\pi^2}{6}[2\pi s_z^{(2)}(0)]$; and fourth, equation (51) thus tells us that $k^{sin(2)}(0) = \frac{\pi^2}{6}$, and so we conclude that $k^{cos(2)}(0)$ is equal to this value as well.

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

Andrews, D.W.K. and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica* 60, 953-966.

Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502.

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.

den Haan, W.J. and A. Levin (2000), "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," NBER Technical Working Paper #255.

Dougherty, C. (2011), *Introduction to Econometrics, 4th Edition*. Oxford: Oxford University Press.

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.

Lazarus, E., D. Lewis, and J.H. Stock. 2015. "Size of Recentered Subsample HAR Tests with O(1/*T*) Estimator Bias," manuscript, Harvard University.

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.

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.

Stambaugh, R. (1999), "Predictive Regressions," *Journal of Financial Economics* 54, 375-421.

Stock, J.H. and M.W. Watson (2011), *Introduction to Econometrics*, *3rd Edition*. Boston: Addison-Wesley.

Stock, J.H. and M.W. Watson (2015), "Factor Models and Structural Vector Autoregressions in Macroeconomics," in preparation for *Handbook of Macroeconomics,* vol. 2.

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. (2013), “Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator,” *The Econometrics Journal*, 16, 1–26.

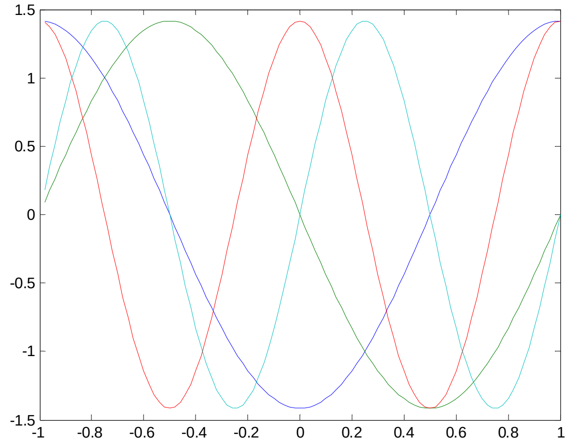
Sun, Y. (2014), “Let’s Fix It: Fixed-*b* Asymptotics versus Small-*b* Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference,” *Journal of Econometrics* 178, 659-677.

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

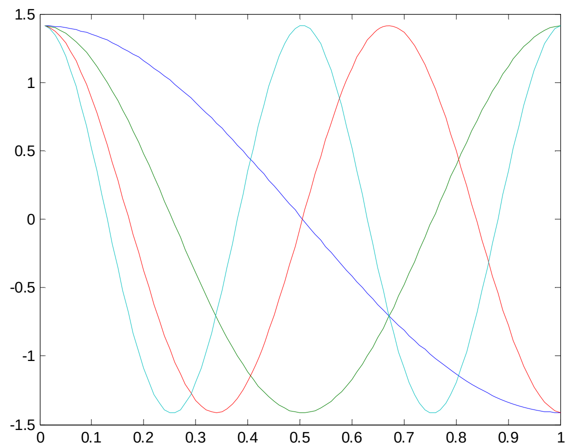
Velasco, C. and P.M. Robinson (2001), “Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean,” *Econometric Theory* 17, 497-539.

Westhoff, Frank (2013), *An Introduction to Econometrics: A Self-Contained Approach*. Cambridge: MIT Press.

Wooldridge, Jeffrey M. (2006). *Introductory Econometrics: A Modern Approach, 3rd Edition*. Thomson.



1.5
1
0.5
0
-0.5
-1
-1.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1



1.5
1
0.5
0
-0.5
-1
-1.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

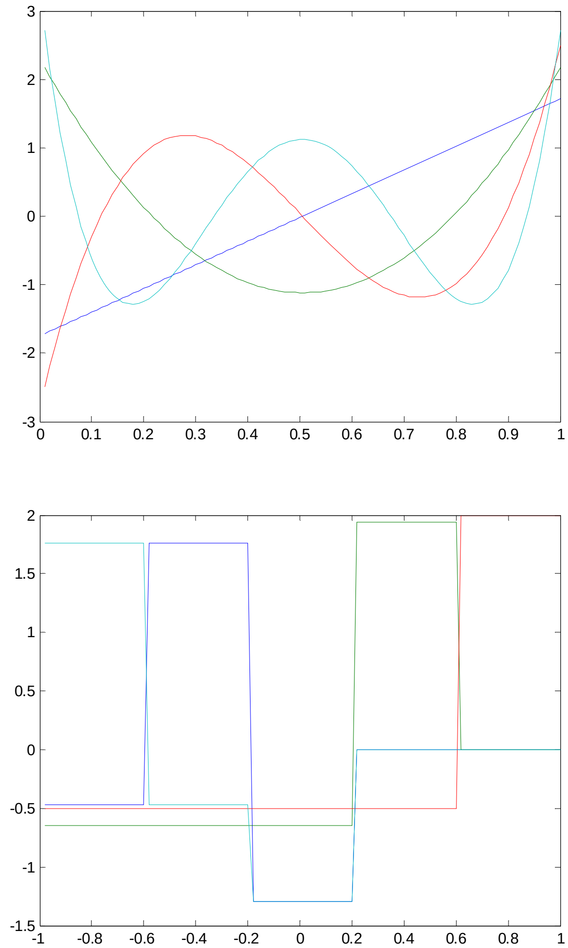


Figure 1. First 4 basis functions: (a) Fourier, (b) cosine, (c) Legendre, (d) Ibragimov-Müller.

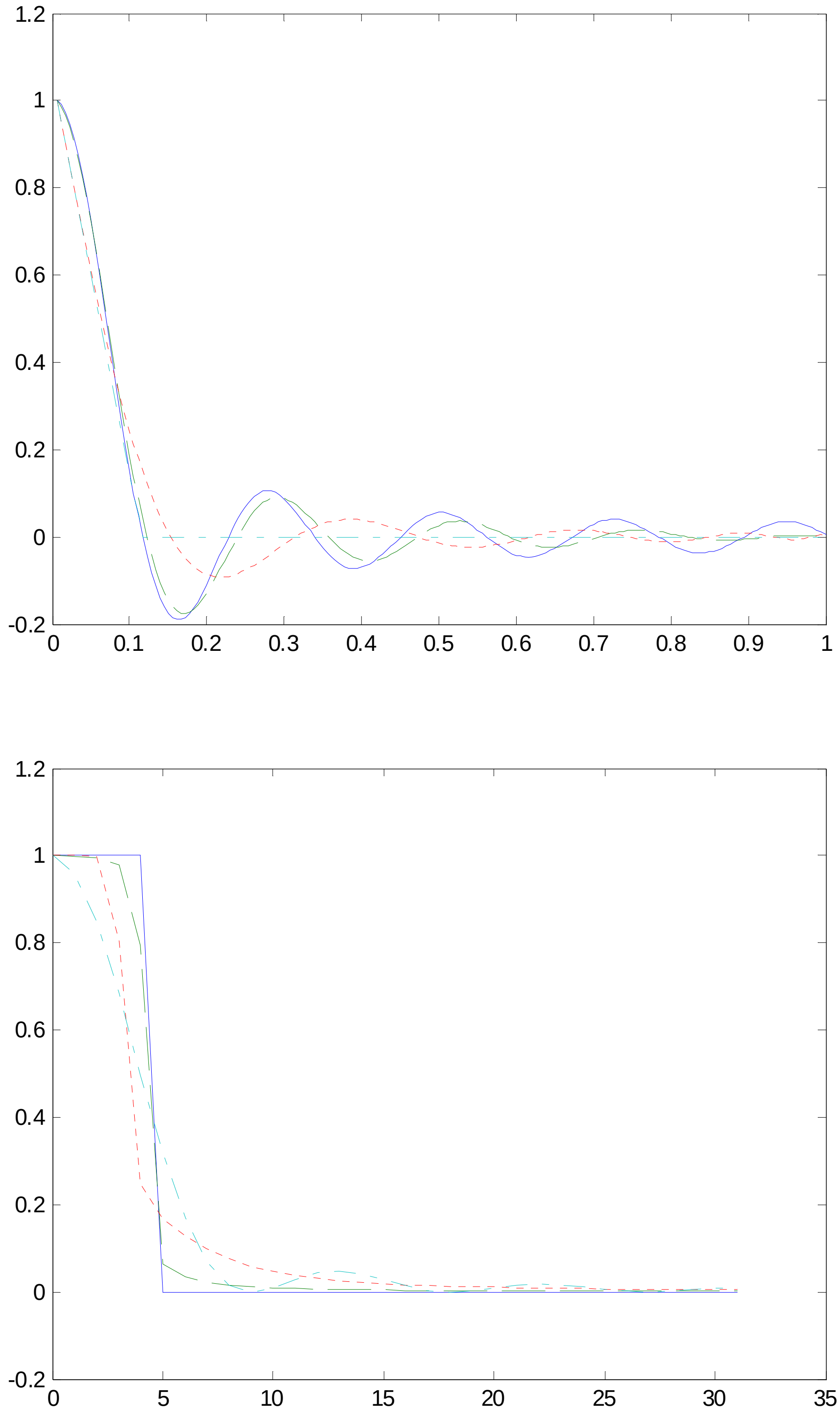


Figure 2. Implied mean kernel of basis function estimators with *B*=8: time domain (upper figure) and frequency domain (lower figure), for Fourier (dark blue, solid), cosine (light blue, dash), Legendre (red,

dot), and Ibragimov-Müller (teal, dash-dot). The frequency domain kernel is normalized to 1 at $\omega = 0$ and computed over the periodogram ordinates (so the horizontal axis value j corresponds to $2\pi j/T$, etc.)

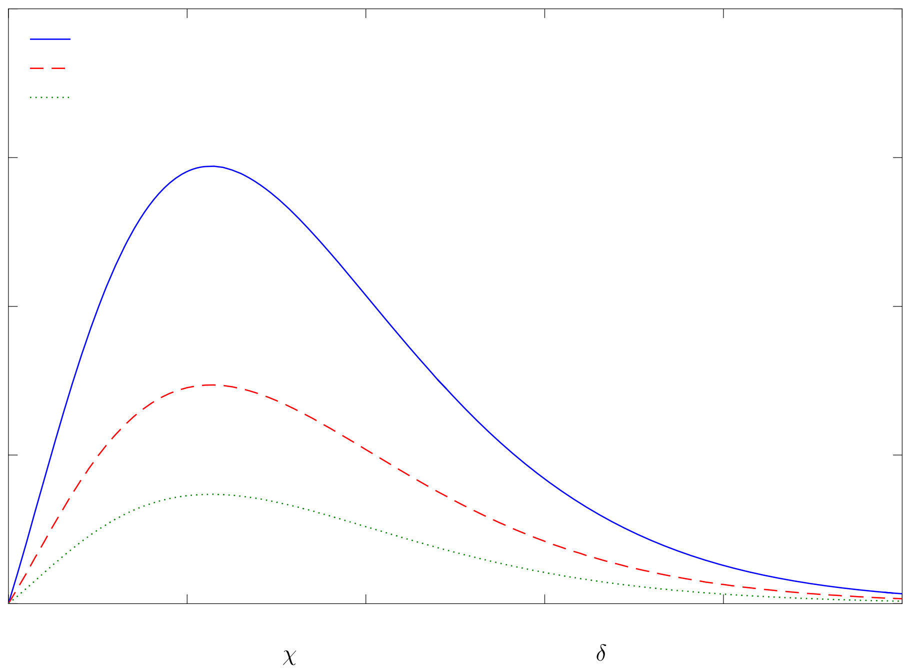


Figure 3: Power loss for EWP vs. QS given equivalent size (m=1, α=0.05)

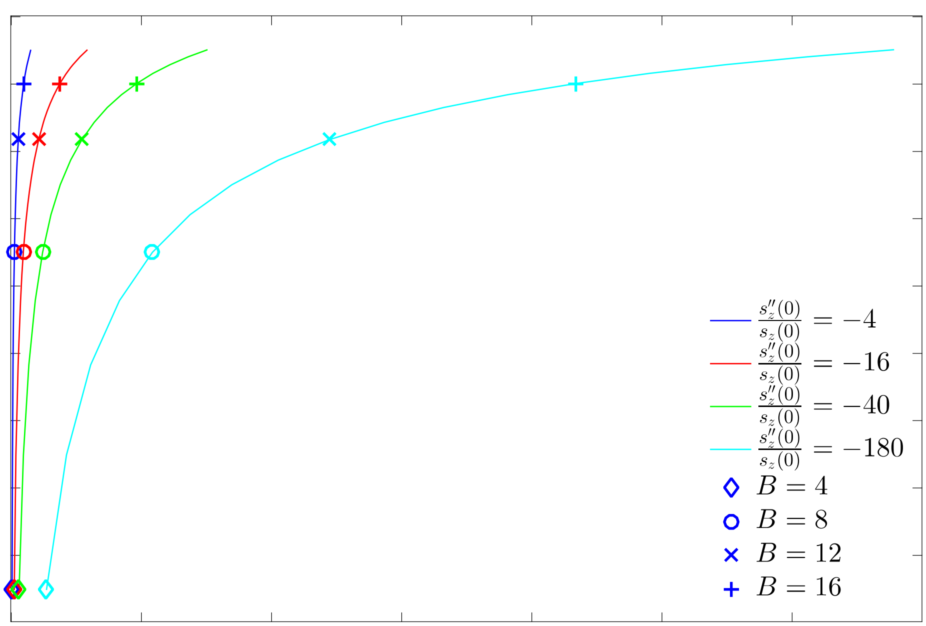


Figure 4: Size-distortion and size-adjusted power loss of EWP (T=200, m=1, α=0.05). In the context of the location model with AR(1) disturbances, the spectral density derivative ratios correspond to AR coefficients of 0.5, 0.7, 0.8, and 0.9, respectively. The δ^2 noncentrality parameter for the alternative is chosen so that power is asymptotically 0.75; this is also used in all subsequent figures requiring a choice of δ^2 for theoretical results.

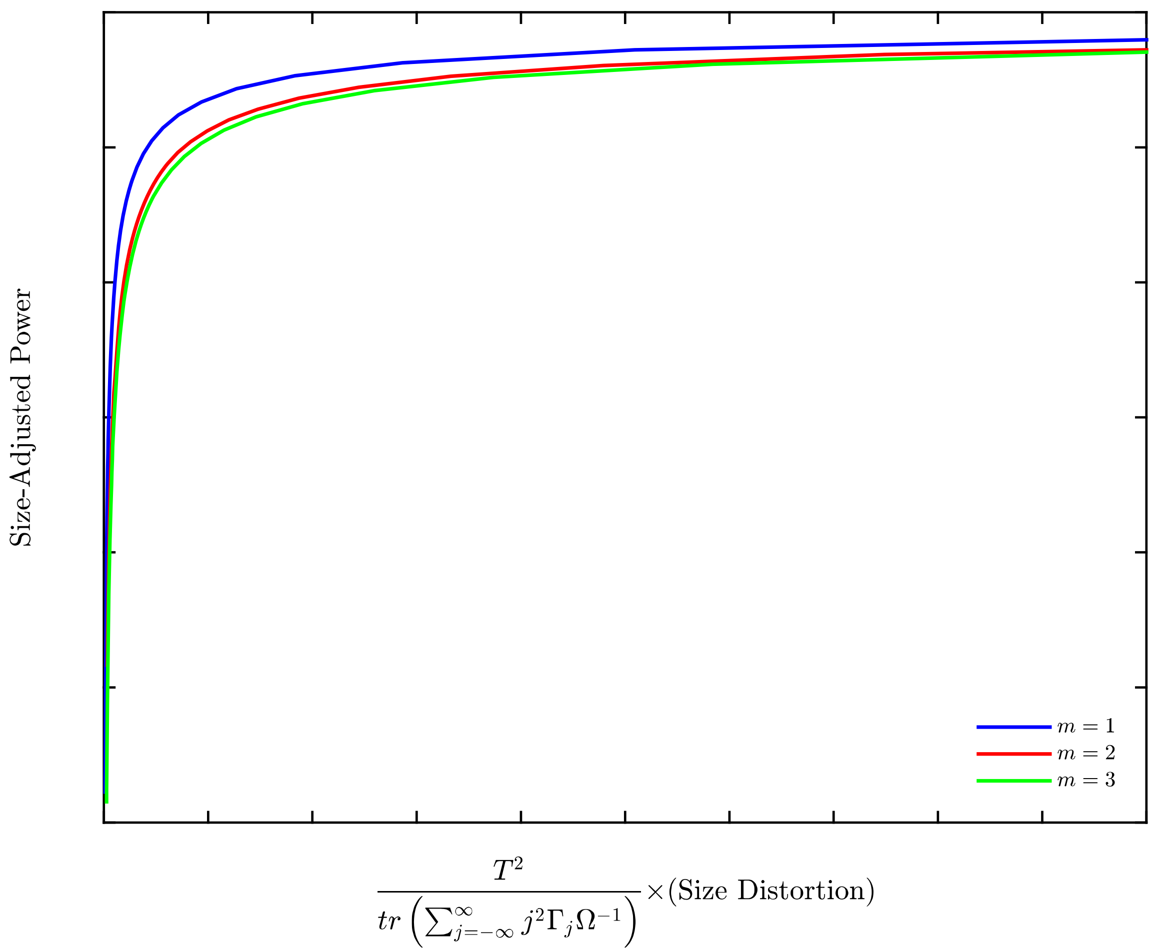


Figure 5: QS-based frontier for higher-order size-adjusted power and normalized size distortion

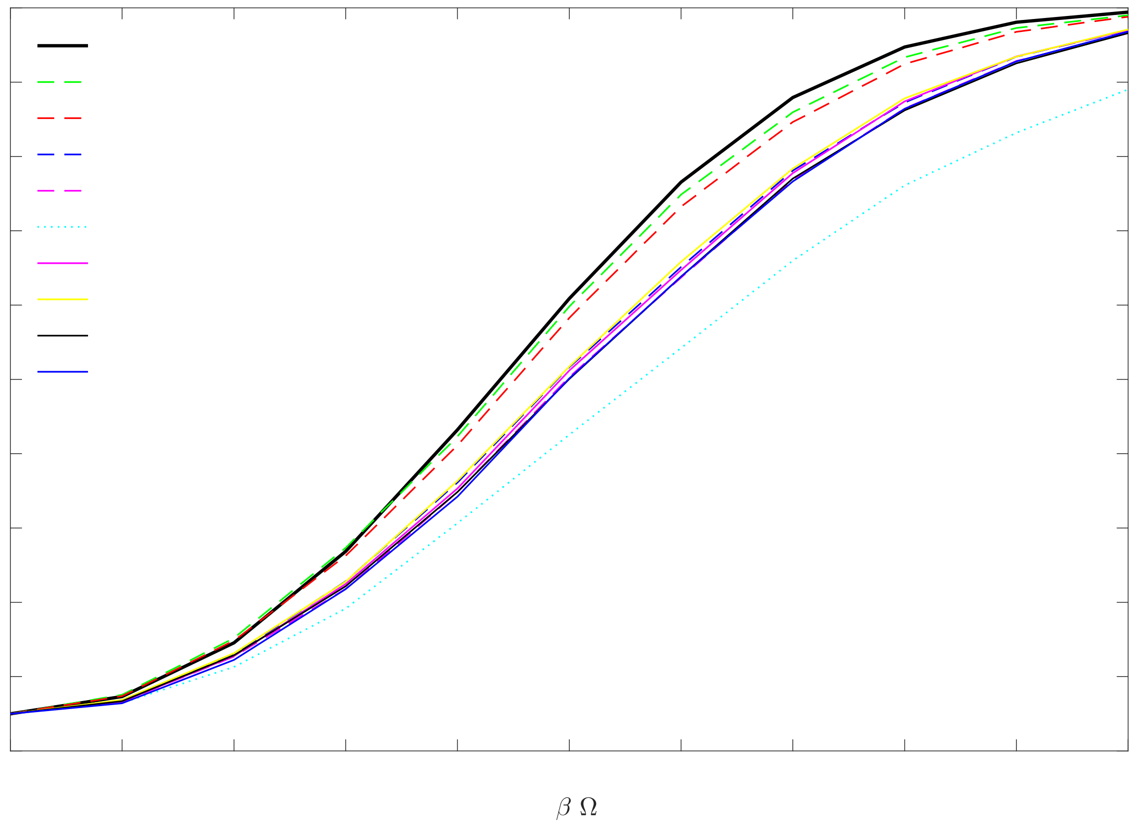


Figure 6(a): ρ = 0.5, T = 200

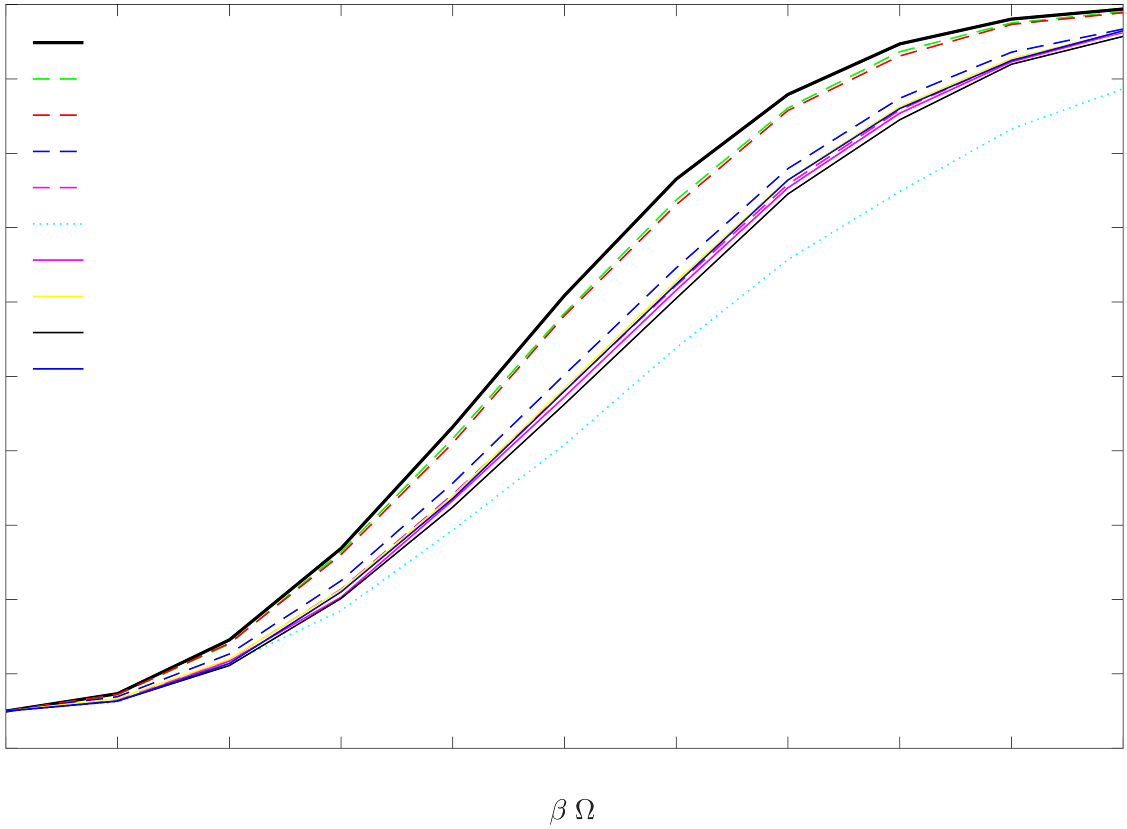


Figure 6(b): ρ = 0.5, T = 400

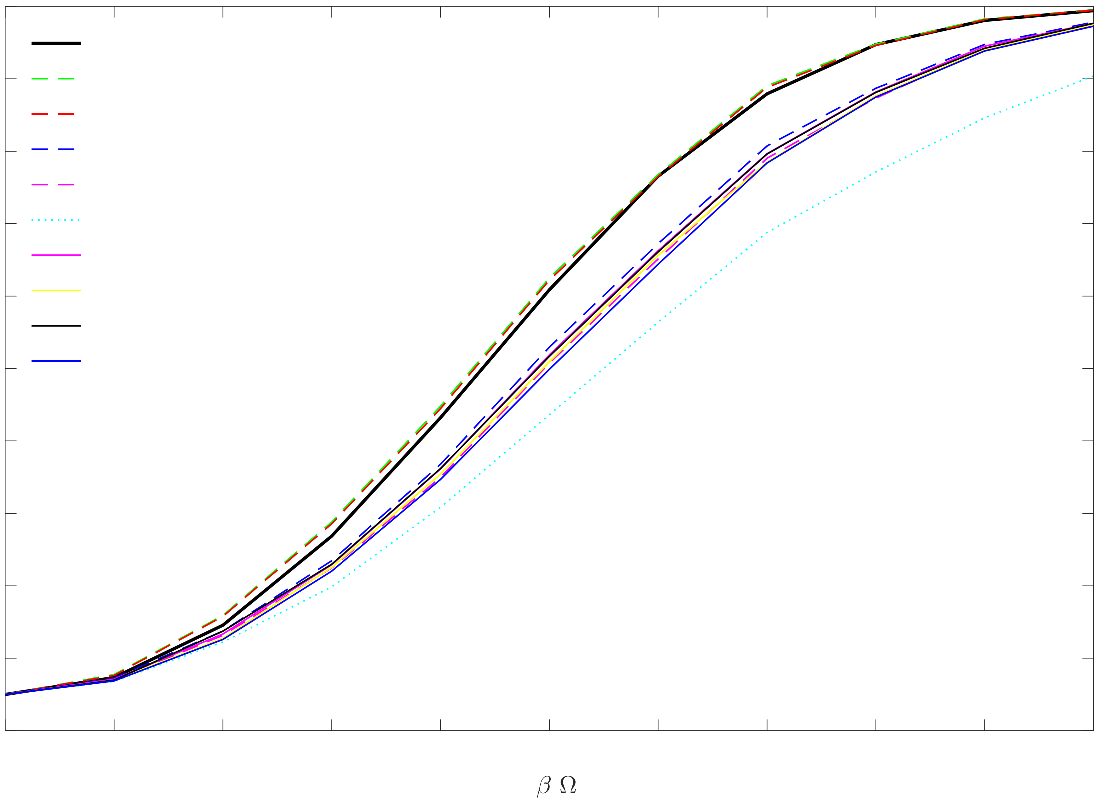


Figure 6(c): ρ = 0.5, T = 2000

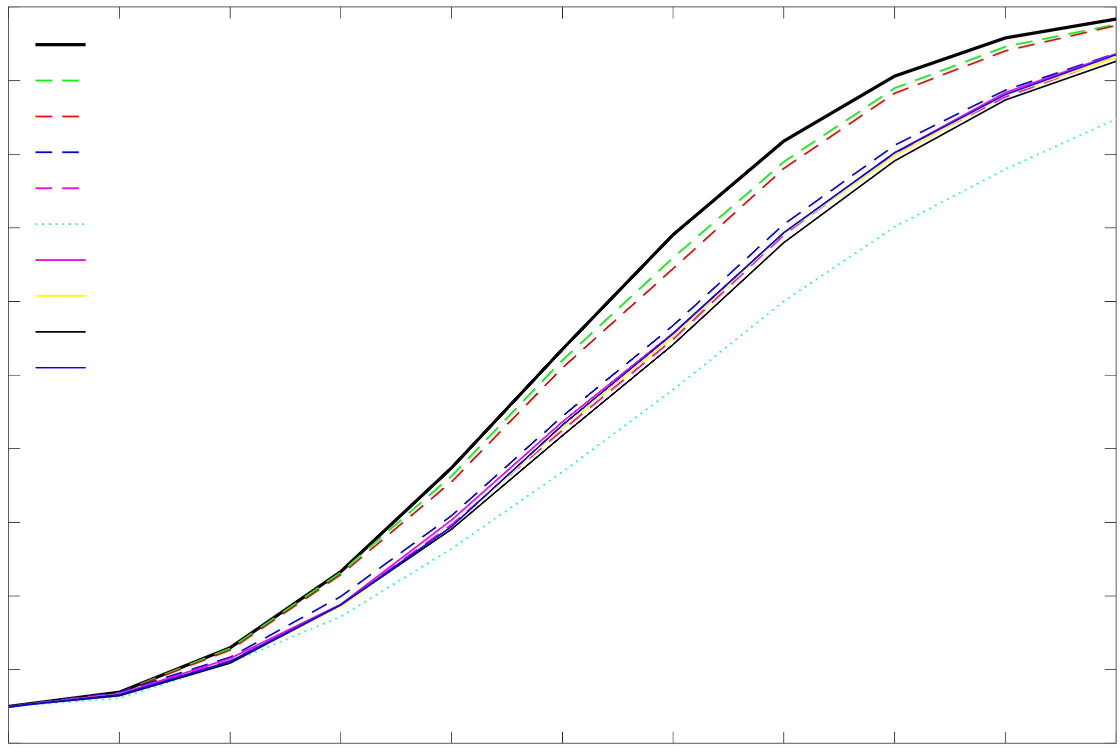


Figure 6(d): ρ = 0.7, T = 400

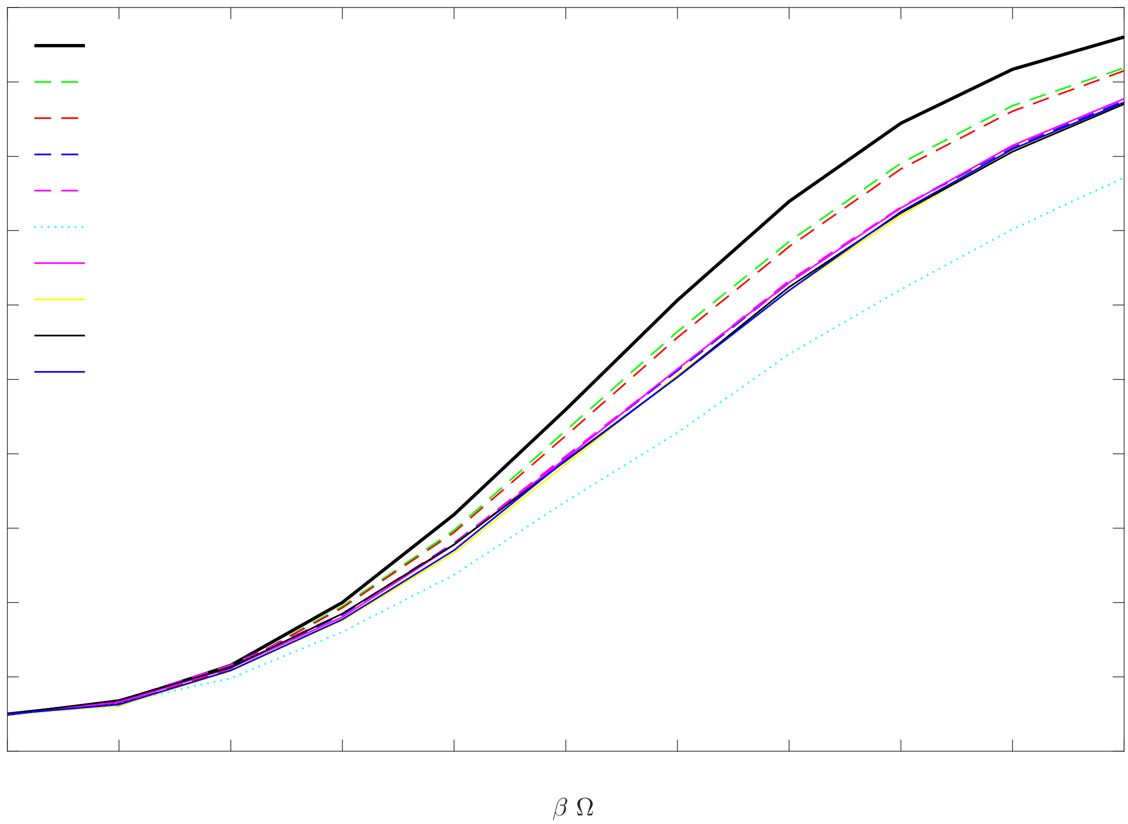


Figure 6(e): ρ = 0.9, T = 200

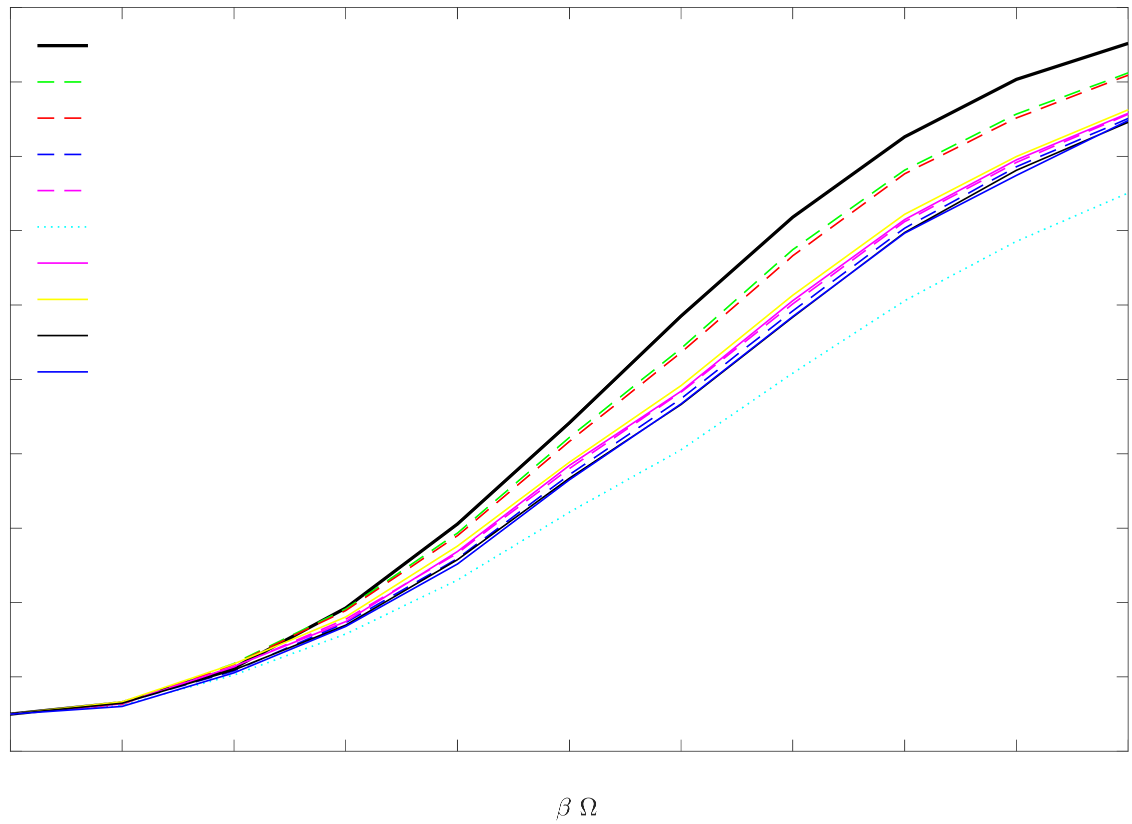


Figure 6(f): ρ = .95, T = 400

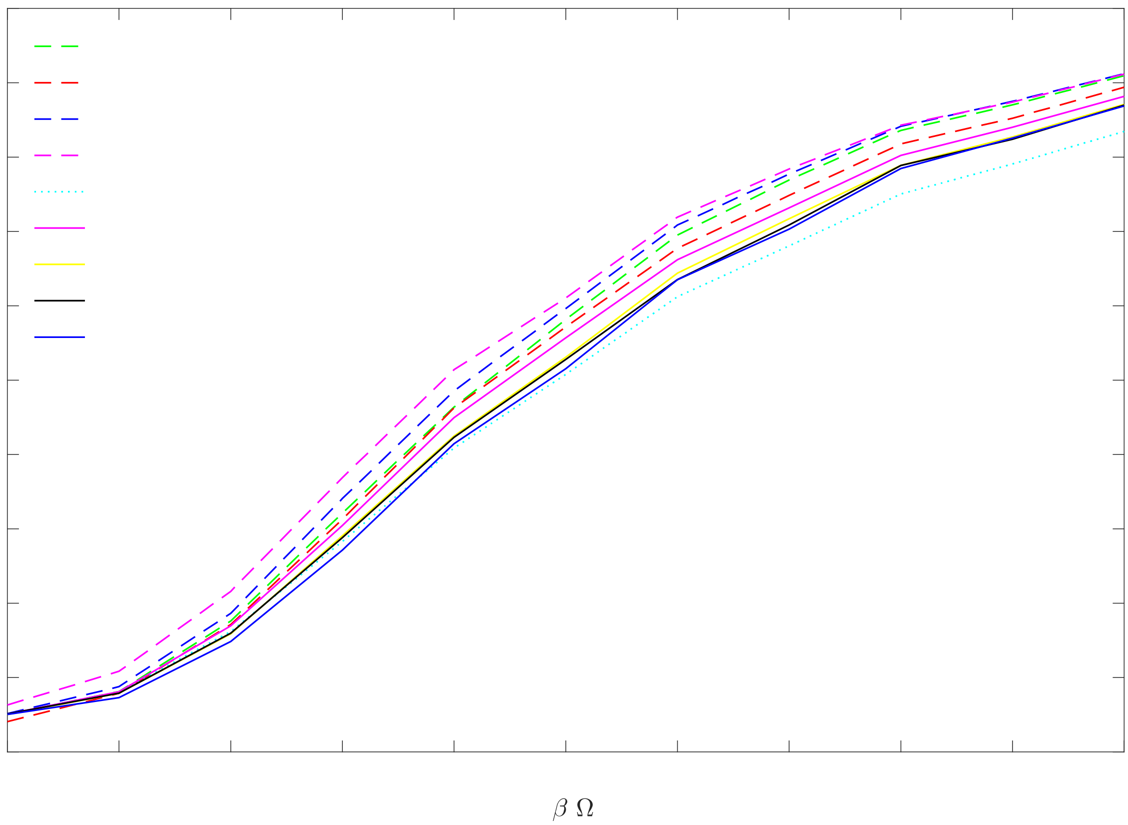


Figure 7: Size-adjusted power against standardized local alternative for selected HAR tests: "flipped" macro regressions, T = 110

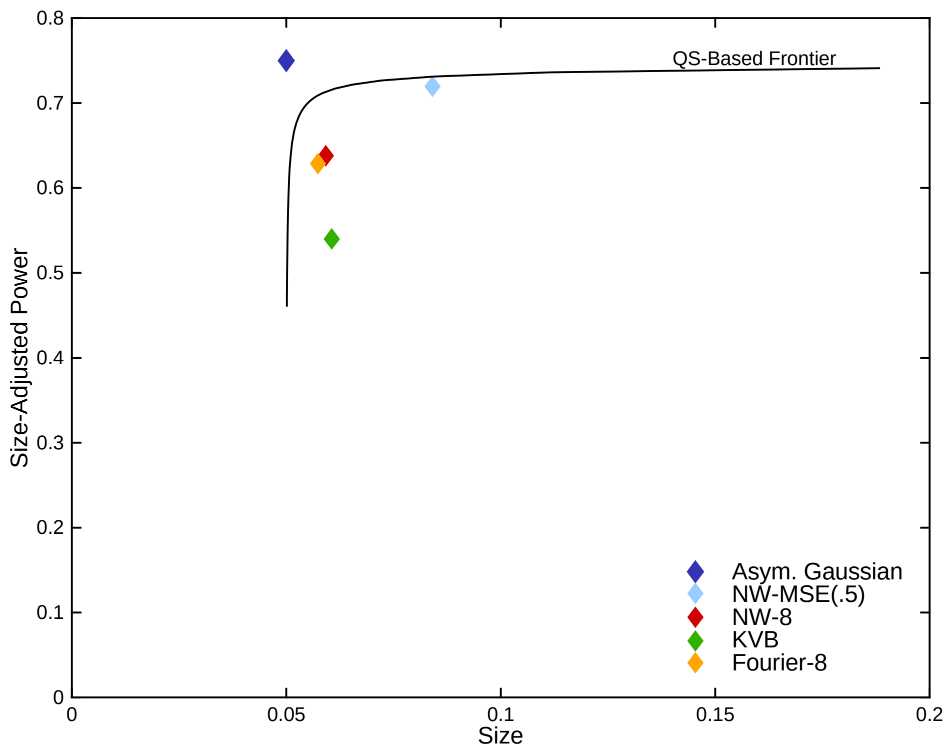


Figure 8(a): Theoretical frontier and Monte Carlo results for size-adjusted power and size (T = 200, m = 1, ρ = 0.5). Numerical values are from VAR(1) design with with AR coefficient matrix ρI, and test is for the coefficient on a single stochastic regressor. Asymptotic normal critical values are used for the NW-MSE(.5) test; fixed-*b* critical value are used for all other tests.

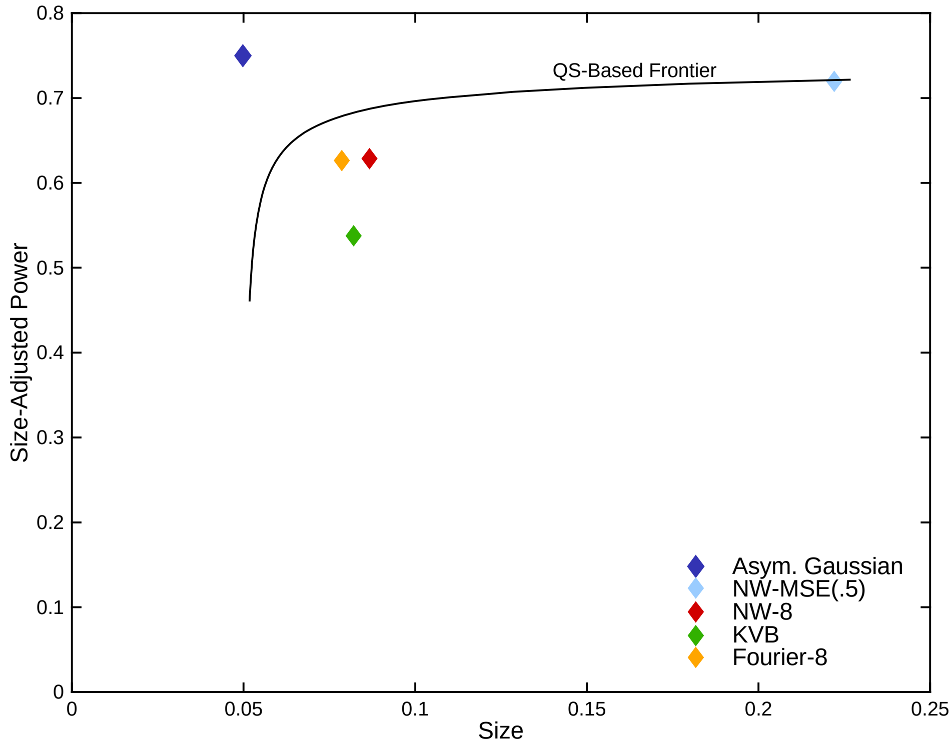


Figure 8(b): Theoretical frontier and Monte Carlo results for size-adjusted power and size (T = 400, m = 1, ρ = 0.9).
See the notes to Figure 8(a).

Table 1: Maximum power loss of EWP compared to QS over alternatives δ

	$\boldsymbol{B}$		
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

Table 2a. HAR test size, design (a): tests of mean (p = 1), normal/chi-squared critical values, AR(1)

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.058	0.104	0.168	0.376	0.539	0.111	0.168	0.246	0.459	0.607
NW-CPE(.5)	0.070	0.088	0.118	0.230	0.374	0.126	0.149	0.180	0.307	0.455
NW-8	0.112	0.122	0.138	0.184	0.273	0.175	0.184	0.202	0.253	0.350
QS-MSE(.5)	0.065	0.078	0.115	0.267	0.433	0.118	0.138	0.180	0.352	0.510
QS-CPE(.5)	0.068	0.077	0.111	0.247	0.407	0.121	0.138	0.173	0.330	0.487
QS-plugin	0.062	0.082	0.133	0.307	0.532	0.116	0.141	0.197	0.398	0.600
QS-prewhiten	0.068	0.067	0.073	0.099	0.141	0.121	0.122	0.128	0.153	0.197
QS-8	0.109	0.113	0.123	0.165	0.258	0.170	0.170	0.183	0.232	0.339
KVB	--	--	--	--	--	--	--	--	--	--
fourier-8	0.085	0.086	0.101	0.145	0.250	0.136	0.143	0.153	0.211	0.328
fourier-12	0.072	0.076	0.093	0.171	0.308	0.125	0.131	0.148	0.241	0.383
fourier-16	0.066	0.072	0.094	0.203	0.356	0.117	0.127	0.151	0.276	0.435
cos-8	0.083	0.092	0.104	0.168	0.298	0.138	0.146	0.158	0.234	0.378
cos-16	0.067	0.073	0.098	0.227	0.400	0.118	0.131	0.156	0.303	0.478
Leg-8	0.084	0.096	0.114	0.168	0.269	0.136	0.151	0.173	0.240	0.346
Leg-16	0.068	0.085	0.111	0.219	0.364	0.120	0.143	0.175	0.295	0.445
IM-basis-8	0.084	0.094	0.112	0.173	0.277	0.138	0.150	0.169	0.238	0.357
IM-basis-16	0.068	0.083	0.115	0.236	0.391	0.120	0.143	0.177	0.313	0.473

Notes: Data are generated by a Gaussian AR(1) described in the text, with the AR coefficient ρ and MA coefficient ϑ given in the first row. T = 200 with 10,000 Monte Carlo draws, so Monte Carlo standard error is .002 for 5% significance level and .003 for 10% significance level. The LRV estimators are described in Section 4.

Table 2b, design (a). HAR test size: tests of mean (p = 1), fixed-b critical values, AR(1)

	5% nominal significance level					**10% nominal significance level**				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.050	0.092	0.155	0.360	0.528	0.101	0.158	0.232	0.446	0.597
NW-CPE(.5)	0.049	0.063	0.091	0.195	0.340	0.101	0.123	0.151	0.274	0.422
NW-8	0.048	0.054	0.068	0.101	0.174	0.098	0.110	0.124	0.171	0.257
QS-MSE(.5)	0.051	0.061	0.094	0.238	0.403	0.100	0.118	0.159	0.324	0.487
QS-CPE(.5)	0.051	0.059	0.086	0.217	0.373	0.099	0.115	0.149	0.298	0.461
QS-plugin	0.052	0.068	0.072	0.282	0.515	0.102	0.125	0.130	0.372	0.588
QS-prewhiten	0.051	0.051	0.057	0.078	0.120	0.099	0.100	0.107	0.135	0.177
QS-8	0.051	0.052	0.061	0.089	0.168	0.103	0.108	0.115	0.157	0.249
KVB	0.047	0.055	0.063	0.090	0.132	0.100	0.106	0.118	0.148	0.203
fourier-8	0.048	0.051	0.061	0.094	0.184	0.101	0.102	0.112	0.163	0.272
fourier-12	0.048	0.052	0.066	0.136	0.256	0.099	0.103	0.122	0.209	0.350
fourier-16	0.050	0.054	0.071	0.170	0.321	0.099	0.108	0.131	0.251	0.407
cos-8	0.049	0.052	0.061	0.111	0.229	0.098	0.104	0.118	0.188	0.319
cos-16	0.051	0.054	0.077	0.194	0.366	0.099	0.110	0.136	0.275	0.451
Leg-8	0.049	0.059	0.072	0.112	0.203	0.095	0.109	0.131	0.187	0.291
Leg-16	0.051	0.062	0.088	0.185	0.331	0.100	0.123	0.151	0.268	0.417
IM-basis-8	0.049	0.055	0.067	0.114	0.213	0.099	0.107	0.126	0.191	0.300
IM-basis-16	0.050	0.063	0.088	0.206	0.357	0.103	0.121	0.154	0.287	0.445

Notes: Data generation and test statistics are the same as in Table 1a, except that fixed-b critical values are used here. The fixed-b critical values for NW and QS are computed using Sun's (2014) F-approximation, the KVB critical values are from Kiefer, Vogelsang, and Bunzel (2000), and the remaining tests use t_B critical values. See the notes to Table 2a.

Table 3a, design (a). HAR test size: tests of coefficient on a single stochastic regressor, normal/chi-squared critical values, VAR(1).

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.062	0.089	0.127	0.279	0.408	0.113	0.150	0.198	0.361	0.486
NW-CPE(.5)	0.075	0.095	0.118	0.208	0.300	0.128	0.155	0.183	0.284	0.382
NW-8	0.116	0.136	0.150	0.222	0.293	0.178	0.200	0.216	0.297	0.376
QS-MSE(.5)	0.070	0.085	0.107	0.207	0.315	0.123	0.144	0.170	0.287	0.400
QS-CPE(.5)	0.072	0.087	0.108	0.199	0.300	0.125	0.145	0.171	0.275	0.385
QS-plugin	0.065	0.084	0.133	0.277	0.474	0.118	0.141	0.197	0.358	0.545
QS-prewhiten	0.071	0.084	0.096	0.136	0.167	0.125	0.141	0.155	0.194	0.231
QS-8	0.111	0.131	0.139	0.203	0.267	0.174	0.195	0.205	0.274	0.348
KVB	--	--	--	--	--	--	--	--	--	--
fourier-8	0.085	0.103	0.115	0.171	0.235	0.143	0.161	0.173	0.238	0.312
fourier-12	0.076	0.092	0.107	0.170	0.242	0.126	0.149	0.166	0.240	0.323
fourier-16	0.072	0.085	0.105	0.178	0.264	0.121	0.141	0.161	0.247	0.343
cos-8	0.087	0.103	0.118	0.178	0.244	0.141	0.161	0.176	0.245	0.321
cos-16	0.071	0.087	0.105	0.183	0.274	0.120	0.142	0.165	0.257	0.354
Leg-8	0.087	0.116	0.127	0.199	0.263	0.144	0.174	0.188	0.269	0.342
Leg-16	0.074	0.093	0.118	0.200	0.287	0.121	0.155	0.179	0.277	0.368
IM-basis-8	0.091	0.105	0.122	0.191	0.259	0.145	0.163	0.182	0.260	0.342
IM-basis-16	0.072	0.089	0.116	0.211	0.307	0.121	0.148	0.178	0.284	0.390

Notes: The data were generated by a three-variable Gaussian VARMA(1,1) with AR coefficient matrix ρI and MA coefficient ϑI, where (ρ, ϑ) are given in the first row. All regressions include an intercept and two stochastic regressors. The reported *t* test is on the coefficient of the second regressor. See the notes to Table 2a.

Table 3b, design (a). HAR test size: tests of coefficient on a single stochastic regressor, fixed-*b* critical values, VAR(1).

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.055	0.079	0.116	0.264	0.394	0.104	0.139	0.186	0.347	0.475
NW-CPE(.5)	0.055	0.071	0.093	0.174	0.259	0.103	0.128	0.155	0.250	0.348
NW-8	0.051	0.062	0.077	0.134	0.189	0.102	0.124	0.136	0.206	0.275
QS-MSE(.5)	0.055	0.069	0.090	0.184	0.286	0.105	0.125	0.149	0.262	0.372
QS-CPE(.5)	0.054	0.068	0.087	0.173	0.267	0.104	0.124	0.147	0.248	0.354
QS-plugin	0.055	0.074	0.085	0.260	0.463	0.107	0.131	0.144	0.345	0.538
QS-prewhiten	0.055	0.065	0.076	0.114	0.142	0.104	0.119	0.132	0.172	0.208
QS-8	0.052	0.063	0.074	0.121	0.170	0.104	0.123	0.133	0.195	0.257
KVB	0.049	0.061	0.074	0.121	0.166	0.104	0.121	0.130	0.191	0.247
fourier-8	0.052	0.061	0.070	0.121	0.170	0.103	0.118	0.132	0.189	0.256
fourier-12	0.053	0.064	0.075	0.134	0.198	0.102	0.121	0.139	0.207	0.289
fourier-16	0.054	0.066	0.080	0.145	0.228	0.100	0.120	0.141	0.222	0.315
cos-8	0.048	0.062	0.075	0.125	0.175	0.100	0.123	0.134	0.198	0.268
cos-16	0.054	0.066	0.082	0.152	0.237	0.100	0.121	0.142	0.232	0.326
Leg-8	0.053	0.068	0.084	0.141	0.197	0.101	0.131	0.146	0.219	0.287
Leg-16	0.057	0.070	0.093	0.171	0.250	0.103	0.133	0.156	0.248	0.341
IM-basis-8	0.052	0.062	0.078	0.134	0.195	0.106	0.121	0.138	0.211	0.283
IM-basis-16	0.054	0.070	0.091	0.178	0.271	0.104	0.126	0.154	0.258	0.363

Notes: The data were generated as described in the notes to Table 2a, and the tests were evaluated using the fixed-*b* critical values described in the notes to Table 2b.

Table 4, design (a). HAR test size: tests of coefficient on two stochastic regressors (*m* = 2), fixed-*b* critical values, VAR(1).

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.056	0.089	0.147	0.387	0.580	0.111	0.154	0.226	0.481	0.655
NW-CPE(.5)	0.057	0.075	0.111	0.236	0.384	0.111	0.141	0.180	0.330	0.481
NW-8	0.054	0.066	0.088	0.170	0.267	0.112	0.129	0.158	0.259	0.373
QS-MSE(.5)	0.058	0.072	0.102	0.251	0.424	0.111	0.132	0.170	0.345	0.518
QS-CPE(.5)	0.057	0.070	0.099	0.230	0.396	0.111	0.133	0.165	0.326	0.490
QS-plugin	0.060	0.079	0.104	0.378	0.671	0.113	0.142	0.176	0.477	0.735
QS-prewhiten	0.057	0.066	0.086	0.152	0.231	0.112	0.127	0.148	0.223	0.305
QS-8	0.064	0.075	0.094	0.156	0.239	0.122	0.136	0.157	0.239	0.343
KVB	0.056	0.067	0.084	0.159	0.243	0.107	0.126	0.149	0.242	0.335
fourier-8	0.054	0.060	0.078	0.131	0.215	0.104	0.117	0.140	0.217	0.317
fourier-12	0.054	0.063	0.084	0.162	0.268	0.108	0.127	0.150	0.246	0.369
fourier-16	0.056	0.066	0.090	0.186	0.321	0.107	0.126	0.153	0.275	0.422
cos-8	0.054	0.059	0.081	0.140	0.221	0.102	0.118	0.146	0.228	0.328
cos-16	0.056	0.067	0.090	0.200	0.336	0.106	0.129	0.160	0.286	0.437
Leg-8	0.058	0.068	0.091	0.167	0.258	0.109	0.132	0.162	0.259	0.366
Leg-16	0.060	0.078	0.109	0.226	0.359	0.114	0.148	0.180	0.319	0.461
IM-basis-8	0.053	0.060	0.082	0.158	0.253	0.105	0.120	0.147	0.248	0.359
IM-basis-16	0.058	0.075	0.104	0.233	0.391	0.106	0.135	0.177	0.328	0.487

Notes: Tests are computed using the F_T^* statistic to test the coefficients on both stochastic regressors. See the notes to Tables 3a and 3b.

Table 5, design (a). HAR test size: tests of coefficient on a single stochastic regressor, fixed-*b* critical values, VARMA(1,1)

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(.7,-.5)**	**(.7,-.2)**	**(.7,0)**	**(.7,.2)**	**(.7,.5)**	**(.7,-.5)**	**(.7,-.2)**	**(.7,0)**	**(.7,.2)**	**(.7,.5)**
NW-MSE(.5)	0.066	0.102	0.116	0.119	0.124	0.119	0.170	0.186	0.184	0.193
NW-CPE(.5)	0.061	0.084	0.093	0.091	0.095	0.114	0.145	0.155	0.154	0.158
NW-8	0.054	0.071	0.077	0.077	0.078	0.110	0.137	0.136	0.136	0.140
QS-MSE(.5)	0.060	0.080	0.090	0.086	0.089	0.113	0.140	0.149	0.147	0.151
QS-CPE(.5)	0.058	0.079	0.087	0.083	0.086	0.112	0.136	0.147	0.145	0.148
QS-plugin	0.066	0.083	0.085	0.079	0.094	0.119	0.143	0.144	0.143	0.158
QS-prewhiten	0.058	0.074	0.076	0.067	0.056	0.112	0.130	0.132	0.122	0.104
QS-8	0.058	0.070	0.074	0.074	0.074	0.112	0.132	0.133	0.133	0.135
KVB	0.057	0.070	0.074	0.075	0.075	0.110	0.137	0.130	0.132	0.135
fourier-8	0.057	0.069	0.070	0.070	0.070	0.109	0.129	0.132	0.130	0.133
fourier-12	0.056	0.072	0.075	0.077	0.077	0.112	0.130	0.139	0.134	0.137
fourier-16	0.055	0.074	0.080	0.078	0.080	0.110	0.129	0.141	0.138	0.143
cos-8	0.055	0.070	0.075	0.071	0.073	0.107	0.130	0.134	0.132	0.137
cos-16	0.055	0.075	0.082	0.080	0.082	0.110	0.131	0.142	0.141	0.146
Leg-8	0.057	0.079	0.084	0.083	0.085	0.114	0.144	0.146	0.144	0.152
Leg-16	0.061	0.085	0.093	0.094	0.098	0.117	0.150	0.156	0.157	0.162
IM-basis-8	0.057	0.073	0.078	0.078	0.077	0.110	0.134	0.138	0.137	0.140
IM-basis-16	0.060	0.081	0.091	0.091	0.093	0.115	0.146	0.154	0.154	0.159

Notes: See the notes to Table 3a.

Table 6, design (a). HAR test size: tests of coefficient on 2 *x's* (*m* = 2), fixed-*b* critical values, VARMA(1,1)

	5% nominal significance level					**10% nominal significance level**				
(ρ, ϑ)	**(.7,-.5)**	**(.7,-.2)**	**(.7,0)**	**(.7,.2)**	**(.7,.5)**	**(.7,-.5)**	**(.7,-.2)**	**(.7,0)**	**(.7,.2)**	**(.7,.5)**
NW-MSE(.5)	0.072	0.125	0.147	0.151	0.159	0.134	0.202	0.226	0.228	0.244
NW-CPE(.5)	0.067	0.096	0.111	0.109	0.116	0.125	0.169	0.180	0.179	0.185
NW-8	0.061	0.081	0.088	0.089	0.091	0.120	0.150	0.158	0.157	0.164
QS-MSE(.5)	0.065	0.090	0.102	0.104	0.106	0.124	0.159	0.170	0.170	0.175
QS-CPE(.5)	0.066	0.087	0.099	0.100	0.103	0.122	0.157	0.165	0.165	0.169
QS-plugin	0.073	0.097	0.104	0.099	0.114	0.133	0.165	0.176	0.166	0.187
QS-prewhiten	0.066	0.082	0.086	0.079	0.068	0.122	0.149	0.148	0.135	0.118
QS-8	0.069	0.084	0.094	0.089	0.092	0.130	0.153	0.157	0.157	0.162
KVB	0.065	0.080	0.084	0.085	0.091	0.117	0.141	0.149	0.148	0.155
fourier-8	0.058	0.071	0.078	0.074	0.078	0.115	0.131	0.140	0.136	0.141
fourier-12	0.063	0.074	0.084	0.083	0.085	0.115	0.143	0.150	0.146	0.151
fourier-16	0.059	0.078	0.090	0.089	0.092	0.117	0.144	0.153	0.150	0.157
cos-8	0.058	0.071	0.081	0.079	0.080	0.113	0.136	0.146	0.139	0.143
cos-16	0.061	0.082	0.090	0.092	0.093	0.118	0.148	0.160	0.157	0.161
Leg-8	0.064	0.081	0.091	0.093	0.097	0.121	0.152	0.162	0.165	0.172
Leg-16	0.069	0.097	0.109	0.110	0.115	0.125	0.170	0.180	0.181	0.188
IM-basis-8	0.059	0.074	0.082	0.081	0.084	0.116	0.141	0.147	0.147	0.154
IM-basis-16	0.064	0.093	0.104	0.105	0.111	0.120	0.160	0.177	0.172	0.181

Notes: Tests are computed using the F_T^* statistic. See the notes to Table 3a.

Table 7a, design (b). HAR test size: Tests of coefficient on x_t in cumulative *h*-step ahead direct forecasting regression,
fixed-*b* critical values: VAR(1), *h* = 4.

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.052	0.085	0.098	0.120	0.151	0.108	0.147	0.163	0.192	0.227
NW-CPE(.5)	0.055	0.071	0.082	0.099	0.126	0.109	0.130	0.141	0.164	0.197
NW-8	0.052	0.063	0.071	0.086	0.112	0.107	0.123	0.127	0.152	0.182
QS-MSE(.5)	0.053	0.068	0.075	0.093	0.117	0.112	0.125	0.136	0.154	0.187
QS-CPE(.5)	0.053	0.068	0.075	0.091	0.117	0.111	0.123	0.134	0.152	0.185
QS-plugin	0.056	0.066	0.083	0.136	0.195	0.113	0.125	0.146	0.210	0.280
QS-prewhiten	0.051	0.061	0.062	0.069	0.085	0.108	0.114	0.115	0.121	0.141
QS-8	0.057	0.065	0.070	0.086	0.110	0.114	0.122	0.128	0.149	0.179
KVB	0.054	0.065	0.068	0.083	0.105	0.105	0.123	0.125	0.146	0.173
fourier-8	0.054	0.063	0.066	0.082	0.107	0.110	0.119	0.126	0.147	0.177
fourier-12	0.058	0.067	0.070	0.088	0.111	0.112	0.124	0.131	0.150	0.183
fourier-16	0.057	0.068	0.074	0.088	0.113	0.112	0.123	0.133	0.151	0.182
cos-8	0.056	0.065	0.068	0.088	0.111	0.113	0.123	0.133	0.151	0.180
cos-16	0.057	0.068	0.075	0.092	0.116	0.114	0.127	0.139	0.156	0.187
Leg-8	0.069	0.083	0.086	0.105	0.132	0.129	0.146	0.151	0.182	0.208
Leg-16	0.070	0.087	0.092	0.116	0.141	0.129	0.147	0.158	0.184	0.215
IM-basis-8	0.053	0.066	0.069	0.086	0.111	0.108	0.122	0.129	0.153	0.181
IM-basis-16	0.055	0.068	0.079	0.098	0.123	0.109	0.129	0.142	0.165	0.199

Notes: The regressions are of the form (27), where tests are of the hypothesis that the coefficient on $\tilde{x}_t$ is zero. See the notes to Table 3a.

Table 7b, design (b). HAR test size: Tests of coefficient on x_t in cumulative h-step ahead regression, fixed-b critical values: VAR(1), h = 8.

	5% nominal significance level					10% nominal significance level				
(ρ, ϑ)	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**	**(0, 0)**	**(0.5, 0)**	**(0.7, 0)**	**(0.9, 0)**	**(0.95, 0)**
NW-MSE(.5)	0.048	0.090	0.121	0.191	0.226	0.096	0.157	0.196	0.266	0.306
NW-CPE(.5)	0.050	0.074	0.087	0.134	0.159	0.099	0.134	0.154	0.206	0.233
NW-8	0.046	0.066	0.075	0.111	0.132	0.098	0.124	0.135	0.181	0.207
QS-MSE(.5)	0.048	0.070	0.081	0.130	0.153	0.099	0.127	0.148	0.202	0.231
QS-CPE(.5)	0.049	0.067	0.078	0.123	0.147	0.099	0.125	0.143	0.195	0.221
QS-plugin	0.048	0.084	0.157	0.346	0.461	0.100	0.150	0.236	0.427	0.535
QS-prewhiten	0.042	0.052	0.054	0.067	0.070	0.087	0.096	0.099	0.117	0.122
QS-8	0.050	0.063	0.073	0.104	0.124	0.104	0.120	0.130	0.172	0.196
KVB	0.048	0.067	0.073	0.102	0.125	0.104	0.126	0.130	0.175	0.195
fourier-8	0.051	0.064	0.072	0.105	0.125	0.105	0.121	0.131	0.172	0.198
fourier-12	0.054	0.066	0.077	0.112	0.132	0.104	0.123	0.136	0.183	0.209
fourier-16	0.054	0.069	0.080	0.119	0.141	0.104	0.128	0.141	0.192	0.215
cos-8	0.054	0.068	0.077	0.112	0.131	0.109	0.126	0.140	0.182	0.209
cos-16	0.055	0.071	0.082	0.123	0.146	0.107	0.134	0.147	0.196	0.220
Leg-8	0.069	0.086	0.097	0.136	0.157	0.129	0.152	0.164	0.212	0.237
Leg-16	0.071	0.092	0.106	0.150	0.179	0.131	0.160	0.175	0.227	0.256
IM-basis-8	0.049	0.067	0.076	0.111	0.136	0.103	0.126	0.133	0.184	0.211
IM-basis-16	0.052	0.073	0.088	0.132	0.157	0.103	0.132	0.153	0.204	0.238

Notes: See the notes to Table 7a.

Table 8. HAR test size, design (c): Tests of coefficient on x_t in "predictive regression" design with long-run correlation between x and u, h = 4 and 8, fixed-b critical values: VAR(1), α=0.7, T=200, 5% nominal significance level

h			**4**					**8**		
(ρ, χ)	**(0.7, 0)**	**(0.7, 0.5)**	**(0.7, 0.7)**	**(0.7, 0.9)**	**(0.7, 0.95)**	**(0.7, 0)**	**(0.7, 0.5)**	**(0.7, 0.7)**	**(0.7, 0.9)**	**(0.7, 0.95)**
NW-MSE(.5)	0.092	0.101	0.100	0.113	0.118	0.121	0.127	0.132	0.151	0.154
NW-CPE(.5)	0.077	0.087	0.090	0.105	0.107	0.087	0.099	0.107	0.132	0.141
NW-8	0.064	0.077	0.083	0.094	0.096	0.070	0.085	0.092	0.122	0.130
QS-MSE(.5)	0.071	0.082	0.083	0.096	0.099	0.084	0.093	0.097	0.119	0.122
QS-CPE(.5)	0.071	0.081	0.082	0.098	0.100	0.080	0.091	0.096	0.118	0.124
QS-plugin	0.067	0.079	0.081	0.095	0.096	0.101	0.106	0.111	0.128	0.128
QS-prewhiten	0.057	0.070	0.070	0.083	0.085	0.055	0.065	0.071	0.089	0.093
QS-8	0.064	0.076	0.084	0.096	0.099	0.066	0.083	0.093	0.121	0.130
KVB	0.062	0.075	0.078	0.090	0.095	0.072	0.085	0.088	0.116	0.118
fourier-8	0.064	0.073	0.084	0.094	0.098	0.067	0.081	0.096	0.121	0.129
fourier-12	0.068	0.079	0.085	0.099	0.100	0.072	0.084	0.098	0.126	0.132
fourier-16	0.072	0.082	0.084	0.100	0.102	0.078	0.091	0.098	0.126	0.134
cos-8	0.067	0.080	0.085	0.096	0.103	0.072	0.087	0.099	0.126	0.133
cos-16	0.077	0.085	0.086	0.103	0.104	0.081	0.092	0.101	0.126	0.134
Leg-8	0.083	0.098	0.105	0.117	0.118	0.090	0.108	0.122	0.149	0.158
Leg-16	0.089	0.101	0.107	0.123	0.125	0.101	0.118	0.126	0.156	0.163
IM-basis-8	0.065	0.076	0.082	0.093	0.097	0.073	0.086	0.098	0.125	0.133
IM-basis-16	0.076	0.082	0.089	0.102	0.107	0.085	0.100	0.107	0.135	0.140

Notes: The data were generated by a two-variable Gaussian VAR(1) with AR coefficient matrix diag(0, ρ) and innovation correlation corr(η_{1t}, η_{2t}) = χ. All regressions consist of the h-quarter ahead cumulated values of y_t on (1, x_t); the reported test is on the coefficient on x_t.

Table 9. HAR test size, design (d): Tests of coefficient on one or both additional regressors in cumulative *h*-step ahead regression, data generated from VAR estimated using macro data, fixed-*b* critical values: *h* = 4 and 8

Variables tested	**EMP**	**EMP**	**EMP&SPR**	**EMP&SPR**	**EMP**	**EMP**	**EMP&SPR**	**EMP&SPR**
h	**4**	**4**	**4**	**4**	**8**	**8**	**8**	**8**
Significance level	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
NW-MSE(.5)	0.081	0.143	0.098	0.165	0.087	0.157	0.121	0.194
NW-CPE(.5)	0.074	0.134	0.085	0.143	0.073	0.138	0.096	0.167
NW-8	0.069	0.126	0.080	0.146	0.070	0.135	0.096	0.163
QS-MSE(.5)	0.068	0.127	0.076	0.136	0.065	0.127	0.088	0.153
QS-CPE(.5)	0.067	0.126	0.076	0.133	0.065	0.126	0.087	0.152
QS-plugin	0.070	0.127	0.092	0.151	0.070	0.135	0.093	0.162
QS-prewhiten	0.061	0.114	0.068	0.123	0.058	0.112	0.079	0.135
QS-8	0.071	0.128	0.088	0.150	0.071	0.136	0.102	0.168
KVB	0.066	0.121	0.071	0.129	0.067	0.128	0.084	0.145
fourier-8	0.066	0.125	0.069	0.126	0.066	0.130	0.082	0.145
fourier-12	0.066	0.126	0.072	0.129	0.066	0.130	0.084	0.148
fourier-16	0.069	0.126	0.076	0.133	0.068	0.131	0.088	0.150
cos-8	0.069	0.126	0.070	0.128	0.069	0.130	0.085	0.144
cos-16	0.070	0.128	0.077	0.133	0.068	0.132	0.090	0.153
Leg-8	0.074	0.134	0.079	0.144	0.078	0.145	0.092	0.162
Leg-16	0.078	0.140	0.091	0.152	0.084	0.149	0.104	0.174
IM-basis-8	0.067	0.127	0.071	0.133	0.071	0.133	0.088	0.155
IM-basis-16	0.077	0.137	0.082	0.148	0.079	0.140	0.101	0.173

Notes: The data were generated by a three-variable Gaussian VAR(1) (T = 200), with autoregressive matrix and innovation covariance matrix estimated using quarterly U.S. GDP growth, employment growth, and the commercial paper-Treasury bill spread. Forecasts are of "GDP" using "employment" and the "spread" as predictors in *h*-step ahead forecasting regressions (27) including GDP and an intercept in the regression. Tests are of the coefficient on "employment" and/or the "spread", where the values of the coefficients under the null are computed from the (known) VAR coefficients.

Table 10. HAR test size, design (e): Tests of coefficient on x_t in stochastic volatility model, fixed-*b* critical values

	homoscedastic u_t						**stochastic volatility u_t**					
τ	**0.00**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.00**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**
NW-MSE(.5)	0.105	0.107	0.112	0.106	0.112	0.119	0.1047	0.1079	0.1106	0.1102	0.1171	0.1207
NW-CPE(.5)	0.080	0.082	0.086	0.085	0.086	0.094	0.0802	0.083	0.0859	0.0875	0.0931	0.097
NW-8	0.066	0.069	0.073	0.076	0.074	0.081	0.0654	0.07	0.0728	0.0771	0.0791	0.0854
QS-MSE(.5)	0.077	0.079	0.081	0.081	0.083	0.089	0.0762	0.0803	0.0816	0.083	0.0881	0.0907
QS-CPE(.5)	0.074	0.075	0.078	0.079	0.080	0.085	0.0743	0.0774	0.0799	0.0807	0.0858	0.0889
QS-plugin	0.070	0.073	0.076	0.077	0.079	0.087	0.0709	0.0745	0.0774	0.0786	0.0842	0.0905
QS-prewhiten	0.066	0.066	0.069	0.070	0.071	0.075	0.0647	0.067	0.0688	0.0722	0.075	0.0779
QS-8	0.063	0.070	0.070	0.071	0.072	0.078	0.0634	0.0688	0.0711	0.0743	0.075	0.0822
KVB	0.064	0.067	0.069	0.069	0.074	0.077	0.0649	0.0671	0.0703	0.0725	0.0758	0.0815
fourier-8	0.063	0.066	0.068	0.067	0.069	0.077	0.0632	0.0658	0.0696	0.0693	0.0748	0.0801
fourier-12	0.067	0.069	0.070	0.069	0.075	0.078	0.067	0.0681	0.0711	0.0735	0.0794	0.0819
fourier-16	0.068	0.069	0.074	0.075	0.076	0.081	0.0676	0.0697	0.0751	0.0773	0.0824	0.0844
cos-8	0.065	0.068	0.070	0.070	0.073	0.075	0.0654	0.0681	0.0712	0.0727	0.0776	0.0792
cos-16	0.069	0.075	0.074	0.076	0.078	0.085	0.0701	0.0733	0.0765	0.079	0.085	0.0849
Leg-8	0.075	0.076	0.082	0.082	0.084	0.086	0.0769	0.0754	0.0848	0.0865	0.0891	0.0936
Leg-16	0.083	0.087	0.091	0.090	0.091	0.097	0.0844	0.0865	0.092	0.0949	0.0957	0.102
IM-basis-8	0.069	0.071	0.074	0.072	0.079	0.083	0.0699	0.0696	0.0747	0.0746	0.0818	0.0863
IM-basis-16	0.082	0.083	0.087	0.087	0.088	0.095	0.0819	0.0823	0.0881	0.089	0.095	0.0994

Notes: Tests are on the coefficient on x_t in a regression of y_t on (1, x_t). The data were generated from a VAR(1) as in the notes to Table 2a (no MA component), modified for stochastic volatility. In the left panel, x_t has the stochastic volatility described in the text, which uses the average of parameters estimated for three macro time series, and the regression error u_t is homoscedastic. In the right panel, x_t and u_t have the same stochastic volatility realization. Only rejection rates at the 5% significance level are shown.

Table 11. HAR test size, design (f): Tests of the coefficient on stochastic regressor(s) with randomly drawn "flipped" macro time series: first-differenced and HP-detrended data, fixed-*b* critical values

Transformation	**Δ**	**Δ**	**Δ**	**Δ**	**HP**	**HP**	**HP**	**HP**
m	**1**	**1**	**4**	**4**	**1**	**1**	**4**	**4**
Significance level	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
NW-MSE(.5)	0.086	0.154	0.110	0.196	0.391	0.489	0.260	0.352
NW-CPE(.5)	0.084	0.154	0.111	0.199	0.331	0.444	0.242	0.344
NW-8	0.078	0.154	0.105	0.195	0.229	0.353	0.192	0.301
QS-MSE(.5)	0.089	0.159	0.114	0.195	0.306	0.422	0.254	0.350
QS-CPE(.5)	0.089	0.161	0.115	0.198	0.308	0.424	0.251	0.349
QS-plugin	0.088	0.154	0.115	0.198	0.305	0.415	0.254	0.345
QS-prewhiten	0.083	0.146	0.111	0.191	0.148	0.236	0.212	0.307
QS-8	0.086	0.168	0.116	0.201	0.246	0.368	0.203	0.314
KVB	0.070	0.140	0.099	0.170	0.159	0.259	0.142	0.239
fourier-8	0.080	0.156	0.102	0.184	0.270	0.398	0.213	0.320
fourier-12	0.086	0.160	0.111	0.195	0.322	0.435	0.245	0.345
fourier-16	0.090	0.159	0.112	0.192	0.298	0.418	0.254	0.354
cos-8	0.094	0.172	0.112	0.203	0.331	0.443	0.239	0.345
cos-16	0.097	0.167	0.113	0.192	0.306	0.421	0.267	0.358
Leg-8	0.108	0.195	0.142	0.234	0.296	0.417	0.240	0.346
Leg-16	0.106	0.189	0.145	0.231	0.384	0.486	0.282	0.377
IM-basis-8	0.082	0.155	0.114	0.194	0.264	0.386	0.212	0.326
IM-basis-16	0.086	0.153	0.119	0.198	0.323	0.436	0.236	0.335

Notes: For the m = 1 tests, two series are drawn from the macro database described in the text, the first and second halves of the x series are flipped, and either the first second half of the pair, y_t and "flipped" x_t is drawn, for 110 observations. The regression is of y_t on (1, x_t) and the coefficient on x_t is tested. For the m = 3 tests, the data are selected and manipulated in the same way, the regressors are (1, x_t, x_{t-1}, x_{t-2}), and the F_T^* statistic is used to test the hypothesis that all three coefficients on x are zero. For the first four columns, the data were preliminarily transformed by first differencing and low-frequency detrending. For the final four columns, the data were transformed by HP detrending (λ = 1600).

Table 12. HAR test size, design (f): Tests of coefficient on x_t in cumulative h-step ahead regression, with randomly drawn "flipped" macro time series: HP-detrended data, fixed-b critical values: h = 4 and 8

Transformation	**Δ**	**Δ**	**Δ**	**Δ**	**HP**	**HP**	**HP**	**HP**
h	**4**	**4**	**8**	**8**	**4**	**4**	**8**	**8**
Significance level	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
NW-MSE(.5)	0.119	0.197	0.139	0.220	0.263	0.350	0.304	0.387
NW-CPE(.5)	0.104	0.183	0.118	0.205	0.204	0.298	0.227	0.316
NW-8	0.094	0.170	0.098	0.183	0.150	0.236	0.150	0.235
QS-MSE(.5)	0.106	0.182	0.127	0.207	0.200	0.288	0.225	0.311
QS-CPE(.5)	0.105	0.178	0.125	0.208	0.193	0.282	0.216	0.303
QS-plugin	0.096	0.171	0.138	0.217	0.194	0.288	0.315	0.403
QS-prewhiten	0.093	0.157	0.093	0.156	0.127	0.196	0.104	0.161
QS-8	0.098	0.177	0.111	0.197	0.150	0.236	0.153	0.239
KVB	0.077	0.144	0.092	0.161	0.118	0.193	0.119	0.185
fourier-8	0.093	0.170	0.118	0.204	0.163	0.253	0.184	0.281
fourier-12	0.108	0.186	0.135	0.222	0.195	0.289	0.232	0.322
fourier-16	0.109	0.185	0.141	0.226	0.207	0.296	0.239	0.326
cos-8	0.108	0.186	0.133	0.218	0.187	0.279	0.217	0.310
cos-16	0.113	0.190	0.143	0.227	0.218	0.303	0.253	0.341
Leg-8	0.152	0.241	0.184	0.281	0.201	0.296	0.212	0.306
Leg-16	0.148	0.236	0.193	0.281	0.262	0.350	0.294	0.386
IM-basis-8	0.098	0.173	0.119	0.198	0.161	0.251	0.190	0.280
IM-basis-16	0.108	0.183	0.132	0.208	0.212	0.304	0.248	0.337

Notes: Entries are rejection rates for tests that the coefficient on x_t is zero in the regression of y_{t+h}^h = y_{t+1} + … + y_{t+h} on y_t and x_t, including an intercept, where y_t and x_t are randomly drawn randomly from the macro database and x is flipped as described in the notes to Table 10 (T = 110). The data were preliminarily transformed by first differencing and low-frequency detrending (left panel) and, alternatively, by HP detrending (right panel).