Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Check for updates

HAR Inference: Recommendations for Practice

Eben Lazarus
MIT Sloan School of Management, Cambridge, MA 02142 (elazarus@mit.edu)

Daniel J. Lewis
Federal Reserve Bank of New York, New York, NY 10045 (daniel.lewis@ny.frb.org)

James H. Stock
Department of Economics, Harvard University, Cambridge, MA 02138; and the National Bureau of Economic Research, Cambridge, MA 02138 (james_stock@harvard.edu)

Mark W. Watson
Department of Economics and the Woodrow Wilson School, Princeton University, Princeton, NJ 08540; and the National Bureau of Economic Research, Cambridge, MA 02138 (mwatson@princeton.edu)

The classic papers by Newey and West (1987) and Andrews (1991) spurred a large body of work on how to improve heteroscedasticity- and autocorrelation-robust (HAR) inference in time series regression. This literature finds that using a larger-than-usual truncation parameter to estimate the long-run variance, combined with Kiefer-Vogelsang (2002, 2005) fixed-b critical values, can substantially reduce size distortions, at only a modest cost in (size-adjusted) power. Empirical practice, however, has not kept up. This article therefore draws on the post-Newey West/Andrews literature to make concrete recommendations for HAR inference. We derive truncation parameter rules that choose a point on the size-power tradeoff to minimize a loss function. If Newey-West tests are used, we recommend the truncation parameter rule $S = 1.3T^{1/2}$ and (nonstandard) fixed-b critical values. For tests of a single restriction, we find advantages to using the equal-weighted cosine (EWC) test, where the long run variance is estimated by projections onto Type II cosines, using $\nu = 0.4T^{2/3}$ cosine terms; for this test, fixed-b critical values are, conveniently, t_ν or F. We assess these rules using first an ARMA/GARCH Monte Carlo design, then a dynamic factor model design estimated using a 207 quarterly U.S. macroeconomic time series.

KEY WORDS: Heteroscedasticity- and autocorrelation-robust estimation; HAC; Long-run variance.

1. INTRODUCTION

If the error term in a regression with time series data is serially correlated, ordinary least squares (OLS) is no longer efficient and the usual OLS standard errors are in general invalid. There are two solutions to this problem: efficient estimation by generalized least squares, and OLS estimation using heteroscedasticity- and autocorrelation robust (HAR) standard errors. In many econometric applications, the transformations required for generalized least squares result in inconsistent estimation, leaving as the only option OLS with HAR inference. Such applications include multi-period return regressions in finance, local projection estimators of impulse response functions in macroeconomics, distributed lag estimation of dynamic causal effects, and multi-step ahead direct forecasts.

The most common approach to HAR inference is to use Newey-West (1987) (NW) standard errors. To compute NW standard errors, the user must choose a tuning parameter, called the truncation parameter (S). Based on theoretical results in Andrews (1991), conventional practice is to choose a small value for S, then to evaluate tests using standard normal or chi-squared critical values. A large subsequent literature (surveyed in Müller (2014)) has shown, however, that this standard approach can lead to tests that incorrectly reject the null far too often. This literature has proposed many alternative procedures that do a better job of controlling the rejection rate under the null, but none has taken hold in practice.

The purpose of this article is to propose and to vet two tests that take on board the lessons of the vast post-Andrews (1991) literature and which improve upon the standard approach.

Before presenting the two procedures, we introduce some notation. Our interest is in the time series regression model,

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

where $X_t = (1 \quad x_t')'$, where x_t and u_t are second-order stationary stochastic processes and $E(u_t|X_t) = 0$. The HAR inference problem arises if $z_t = X_t u_t$ is heteroscedastic and/or serially correlated. The key step in computing HAR standard errors is estimating the long-run variance (LRV) matrix Ω of z_t,

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j = 2\pi s_z(0), \tag{2}$$

© 2018 American Statistical Association
Journal of Business & Economic Statistics
October 2018, Vol. 36, No. 4
DOI: 10.1080/07350015.2018.1506926
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

where $\Gamma_j = \text{cov}(z_t, z_{t-j}), j = 0, \pm 1, \ldots$ and $s_z(0)$ is the spectral density of z_t at frequency zero. The fundamental challenge of HAR inference is that Ω involves infinitely many autocovariances, which must be estimated using a finite number of observations.

The NW estimator of Ω, $\hat{\Omega}^{NW}$, is a linearly declining weighted average of the sample autocovariances of $\hat{z}_t = X_t\hat{u}_t$ through lag S, where $\hat{u}_t$ are the regression residuals. A typical rule for choosing S is $S = 0.75T^{1/3}$. This rule obtains from a formula in Andrews (1991), specialized to the case of a first-order autoregression with coefficient 0.25, and we shall refer to it, used in conjunction with standard normal or chi-squared critical values, as "textbook NW".[1]

The two proposed tests build on recent literature that combines the two workhorse tools of the theoretical literature, Edgeworth expansions (Velasco and Robinson (2001), Sun, Phillips, and Jin (2008)) and fixed-b asymptotics (Kiefer and Vogelsang (2002, 2005)), where b is the truncation parameter as a fraction of the sample size, that is, $b = S/T$. Both tests use fixed-b critical values, which deliver higher-order improvements to controlling the rejection rate under the null (Jansson (2004), Sun, Phillips, and Jin (2008), Sun (2013, 2014)). The choice of b entails a tradeoff, with small b (small S) resulting in biased estimates of Ω and large size distortions, and large b resulting in smaller size distortions but greater variance of the LRV estimator and thus lower power. The two tests are based on rules, derived in Section 3, that minimize a loss function that trades off size distortions and power loss. Using expressions for this tradeoff from Lazarus, Lewis, and Stock (2017) (LLS), we provide closed-form expressions for the rules. The rules are derived for the problem of testing the mean of a Gaussian time series (the Gaussian location model). We conduct extensive Monte Carlo simulations to assess their performance in the non-Gaussian location model and in time series regression.

Throughout, we restrict attention to tests that (i) use a positive semidefinite (psd) LRV estimator and (ii) have known fixed-b distributions. In addition, (iii) we consider persistence that is potentially high but not extreme, which we quantify as z_t having persistence no greater than an AR(1) with $\rho = 0.7$.

Criteria (i) and (ii) limit the class of estimators we consider[2], and (as explained in Section 3) results in LLS motivate us to further restrict attention to the quadratic spectral (QS), equal-weighted periodogram (EWP), and equal-weighted Type-II discrete cosine transform (EWC) estimators, along with the Newey-West estimator.

Criterion (iii) means that our tests are relevant for many, but not all, HAR problems, and in particular are not designed to cover the case of highly persistent regressors, such as the local-to-unity process examined by Müller (2014). In developing our procedures, we focus on the case of positive persistence, which (as shown in Section 3) corresponds to over-rejections under the null; in our Monte Carlo assessment, however, we include anti-persistent designs.

The first of our two proposed tests uses the NW estimator with the rule,

$$S = 1.3T^{1/2}\,(\textit{loss-minimizing rule},\ NW)\,. \tag{3}$$

This rule yields a much larger value of S than the textbook rule. For example, for $T = 200$, the rule (3) yields $S = 19$ (rounded up[3]), compared to $S = 5$ for the textbook rule. The test uses fixed-b critical values, which are nonstandard for the NW estimator.

The second test uses the Equal-Weighted Cosine (EWC) estimator of the long-run variance. The EWC test is a close cousin of the Equal-Weighted Periodogram (EWP) test, which, although less familiar to econometricians, has a long history. The EWP estimates the spectral density at frequency zero by averaging the squared discrete Fourier transform near frequency zero, that is, the squared projections of $\hat{z}_t$ onto sines and cosines at Fourier frequencies near zero. The EWC estimator (Müller (2004)) uses instead the Type II discrete cosine transform, so it is an equal-weighted average of projections on cosines only (see Equation (10) in Section 2). The free parameter for EWP and EWC is ν, the total number of sines and cosines for EWP or, for EWC, the total number of cosines. Our proposed rule is

$$\nu = 0.4T^{2/3}\,(\textit{loss-minimizing rule},\ EWP/EWC)\,. \tag{4}$$

As shown in Müller (2004), the EWC estimator produces t-ratios and Wald statistics with large-sample null distributions that are analogues of the finite-sample distributions from *i.i.d.* normal samples, where ν is the degrees of freedom of the Student t or Hotelling T^2 distribution. Specifically, the fixed-b distribution of the EWP/EWC t-ratio is t_ν, a result that, for EWP, appears to date to an exercise in Brillinger (1975, exercise 5.13.25). For multiple restrictions, the EWP/EWC F statistic

[1]The Newey-West estimator is implemented in standard econometric software, including Stata and EViews. Among undergraduate textbooks, Stock and Watson (2015, eq (15.17)) recommends using the Newey-West estimator with the Andrews rule $S = 0.75T^{1/3}$. Wooldridge (2012, sec. 12.5) recommends using the Newey-West estimator with either a rule of thumb for the bandwidth (he suggests $S = 4$ or 8 for quarterly data, not indexed to the sample size) or using either of the rules, $S = 4(T/100)^{2/9}$ or $S = T^{1/4}$ (Wooldridge also discusses fixed-b asymptotics and the Kiefer-Vogelsang-Bunzel (Kiefer, Vogelsang, and Bunzel 2000) statistic). Both the Stock-Watson (2015) and Wooldridge (2012) rules yield $S = 4$ for $T = 100$ and $S = 4$ (second Wooldridge rule) or 5 (Stock-Watson and first Wooldridge rule) for $T = 200$ (rounded up). Dougherty (2011) and Westhoff (2013) recommend Newey-West standard errors but do not mention bandwidth choice. Hill, Griffiths, and Lim (2018) and Hillmer and Hilmer (2014) recommend Newey-West standard errors without discussing bandwidth choice, and their empirical examples, which use quarterly data, respectively set $S = 4$ and $S = 1$.

[2]Criteria (i) and (ii) restrict our attention to kernel estimators and orthonormal series estimators of Ω, the class considered by LLS. Series estimators, also called orthogonal multitapers, are the sum of squared projections of $X_t\hat{u}_t$ on orthogonal series, see Brillinger (1975), Grenander and Rosenblatt (1957), Müller (2004), Stoica and Moses (2005), and Phillips (2005) (Sun (2013) discusses this class and provides additional references). This class includes subsample (batch-mean) estimators, the family considered in Ibragimov and Müller (2010), see LLS for additional discussion. Our requirement of a known fixed-b asymptotic distribution rules out several families of LRV estimators: (a) parametric LRV estimators based on autoregressions (Berk 1974) or vector autoregressions (VARHAC; den Haan and Levin (2000), Sun and Kaplan (2014)); (b) kernel estimators in which parametric models are used to prewhiten the spectrum, followed by nonparametric estimation (e.g., Andrews and Monahan 1992); (c) methods that require conditional adjustments to be psd with probability one, e.g., Politis (2011). Our focus on ease and speed of implementation in standard software leads us away from bootstrap methods (e.g. Gonçalves and Vogelsang 2011).

[3]All tests considered in this article use integer values of the truncation parameter S and the degrees of freedom ν. Our rounding convention is to round up S for NW. For EWP/EWC, $\nu = T/S$, so we round down ν.

Table 1. Monte Carlo rejection rates for 5% HAR t-tests under the null

	Estimator	Truncation rule	Critical values	Null imposed?	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$
1	NW	$S = 0.75T^{1/3}$	N(0,1)	No	0.088	0.114	0.180
2	NW	$S = 1.3T^{1/2}$	fixed-b (nonstandard)	No	0.067	0.079	0.108
3	EWC	$\nu = 0.4T^{2/3}$	fixed-b (t_ν)	No	0.062	0.071	0.097
4	NW	$S = 0.75T^{1/3}$	N(0,1)	Yes	0.077	0.095	0.149
5	NW	$S = 1.3T^{1/2}$	fixed-b (nonstandard)	Yes	0.056	0.060	0.074
6	EWC	$\nu = 0.4T^{2/3}$	fixed-b ($t\nu<\Sigma\psi\mu>$)	Yes	0.052	0.053	0.063
Theoretical bound based on Edgeworth expansions for the Gaussian location model							
7	NW	$S = 1.3T^{1/2}$	fixed-b (nonstandard)	No	0.054	0.058	0.067
8	EWC	$\nu = 0.4T^{2/3}$	fixed-b ($t\nu<\Sigma\psi\mu>$)	N/A	0.051	0.054	0.064

NOTES: Tests are on the single stochastic regressor in (1) with nominal level 5%, $T = 200$. The single stochastic regressor x_t and disturbance u_t are independent Gaussian AR(1)'s, with AR coefficients $\rho_X = \rho_u = \rho^{1/2}$, where ρ is given in the column heading. All regressions include an intercept. The tests differ in the kernel used for the standard errors (first column; NW is Newey-West/Bartlett kernel, EWC is equal-weighted cosine transform). The truncation rule for NW is expressed in terms of the truncation parameter S and for EWC as the degrees of freedom ν of the test, which is the number of cosine terms included. The third and fourth columns describe the critical values used and whether the estimator of Ω is computed under the null or is unrestricted. For the EWC test in the location model, the test statistic is the same whether the null is imposed or not.

(with a degrees-of-freedom correction) has a fixed-b asymptotic F distribution. The EWC and EWP tests are asymptotically equivalent for ν even, but EWC has the practical advantage of allowing ν to be any integer whereas ν must be even for EWP because the Fourier sines and cosines appear in pairs.

Sections 4 and 5 report results of extensive Monte Carlo simulations examining these two tests. Section 4 uses a tightly parameterized ARMA/GARCH design. To provide a more realistic test platform, the simulations in Section 5 use a design based on a dynamic factor model fit to 207 quarterly U.S. macroeconomic time series, both with Gaussian errors and with errors drawn from the empirical distribution in a way that preserves the time series structure of the second moments.

Although there are nuances, it turns out that most of our findings are captured by a simple Monte Carlo design with a single stochastic regressor, in which x_t and u_t are independent Gaussian AR(1) processes, for which results are shown in Table 1.[4] The table reports Monte Carlo rejection rates under the null for NW and EWC tests, and illustrates three of our five main findings.

First, as shown in the first row, the textbook NW method can result in large size distortions. The practitioner hopes to conduct a 5% test, but the test actually rejects 11.4% of the time under the null in the moderate-dependence case ($\rho_x = \rho_u = \sqrt{0.5}$) and 18.0% of the time in the high-dependence case ($\rho_x = \rho_u = \sqrt{0.7}$). This finding recapitulates simulation results dating at least to den Haan and Levin (1994, 1997). The large size distortions stem from bias in the NW estimator arising from using a small truncation parameter.

Second, as seen in rows 2 and 3, using the rules (3) or (4) with fixed-b critical values substantially reduces the size distortions observed in the textbook NW test.

Third, these improvements in size come with little cost in loss of power. This is shown in theory in Section 3 for the location model and is verified in the Monte Carlo simulations.

Fourth, although in theory the EWC test asymptotically dominates NW (both with the proposed rules and fixed-b critical values) for the location model, our simulation results suggest that, for sample sizes typically found in practice, neither test has a clear edge in regression applications. Broadly speaking, for inference about a single parameter, EWC has a slight edge, but in high dimensional cases with lower persistence, NW has the edge. These differences typically are small, and the choice between the two is a matter of convenience. In this regard, a convenient feature of EWC is its standard t and F fixed-b critical values.

Fifth, as seen in rows 4–6, for each of the three tests, the size distortions are substantially less using the restricted instead of unrestricted estimator of Ω. The difference between the two estimators is whether the null is imposed in constructing z: the unrestricted estimator uses $\hat{z}_t = X_t\hat{u}_t$ as described above, whereas the restricted estimator uses $\tilde{z}_t = X_t\tilde{u}_t$, where $\tilde{u}_t$ is the OLS residual from the restricted regression that imposes the null. However, this size improvement comes with a loss of power that, in some of our simulations, can be large; moreover, the econometric theory of the differences between the restricted and unrestricted estimators is incomplete. As a result, our recommendations focus on the use of the unrestricted estimator.

Figure 1 summarizes the derivation of the rules (3) and (4) for 5% two-sided tests of the mean in the scalar location case (that is, $X_t = 1$ in (1)) for a Gaussian AR(1) process with AR coefficient 0.7 and $T = 200$. Figure 1 plots the size-power tradeoff for a number of tests, all of which use fixed-b critical values. The horizontal axis is the size distortion Δ_S, which is the rejection rate of the test minus its nominal level $\alpha = 0.05$. The vertical axis is the maximum size-adjusted power loss, $\Delta_P^{\max}$, relative to the infeasible oracle test that uses the true value of Ω, where the maximum is computed over all alternatives.[5] The dashed line shows the theoretical tradeoff (derived in LLS) between the size distortion and the maximum power loss for the NW test; the points on this tradeoff are determined by different choices of b, with large b corresponding to points in the northwest and small b corresponding to points in the southeast. The solid line is the theoretical size-power tradeoff shared by the EWP and EWC estimators. LLS show that among psd estimators that produce

[4]If x_t and u_t are independent scalar AR(1)s with coefficients ρ_x and ρ_u, then z_t has the autocovariances of an AR(1) with coefficient $\rho_x\rho_u$. Throughout, we calibrate dependence through the dependence structure of z_t, which is what enters Ω.

[5]The theory of optimal testing aims to find the best test among those with the same size. To compare rejection rates under the alternative for two tests with different finite-sample sizes, an apples-to-apples comparison requires adjusting one or both of the tests so that they have the same rejection rate under the null. We elaborate on size adjustment in Section 3.

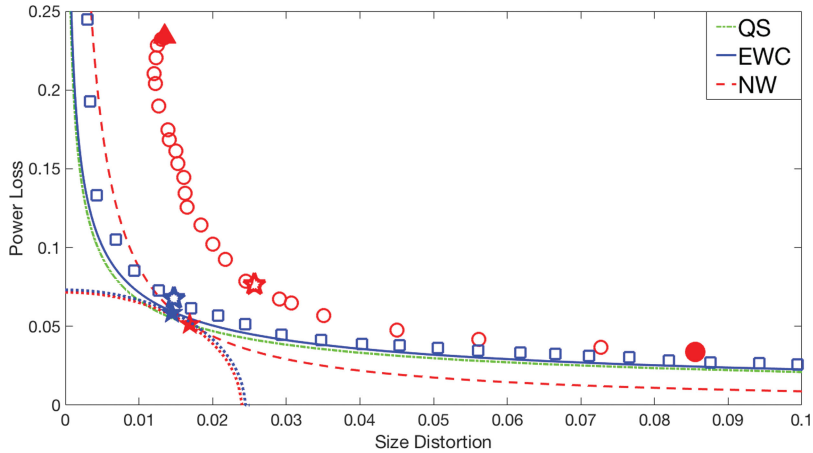

Figure 1. Tradeoff between size distortion, Δ_S, and maximum size-adjusted power loss relative to the oracle test, $\Delta_P^{\max}$, $T = 200$, in the Gaussian location problem, AR(1) disturbances with autoregression coefficient $\rho = 0.7$. Notes: Lines are the asymptotic frontier (green) and the theoretical tradeoffs for the NW test (dash) and EWP/EWC tests (solid); symbols are Monte Carlo values for NW (circle) and EWC (square); solid circle is NW using the Andrews rule $S = 0.75T^{1/3}$, solid triangle is the Kiefer-Vogelsang-Bunzel (2000) test, ellipses are iso-loss curves ((5) with $\kappa = 0.90$); stars are the loss-minimizing rule outcomes in theory (solid) and Monte Carlo (open).

tests with t or F fixed-b critical values, EWC and EWP share the most favorable asymptotic size-power tradeoff. In fact, the EWP/EWC theoretical tradeoff is very close to the asymptotic frontier for the size-power tradeoff for all kernel tests (the dotted line), which is achieved by the quadratic spectral (QS) kernel using nonstandard critical values. A point in Figure 1 has both better size control and less power loss than all points to its northeast. As the sample size grows, the curves shift in, but more slowly for NW so that eventually it is dominated by EWP/EWC, but for 200 observations with $\rho = 0.7$, the NW and EWP/EWC lines cross.

The squares and circles denote size distortions and power loss combinations computed by Monte Carlo simulation. For the EWC test (squares), the theoretical tradeoff evidently is a very good guide to the finite sample performance. In contrast, for NW, the finite-sample tradeoff (circles) are well to the northeast of the theoretical tradeoff, a finding we suspect relates to the remainder term in the Edgeworth expansion being of a lower order for NW than for EWC. The finite-sample EWC tradeoff dominates the NW tradeoff.

The solid circle indicates the textbook NW test: its null rejection rate is nearly 0.14, so its size distortion is nearly 0.09, and its maximum size-adjusted maximum power loss, relative to the oracle test, is 0.03. The solid triangle is the Kiefer-Vogelsang-Bunzel (2000) (KVB) test, which is NW with bandwidth equal to the sample size ($b = 1$). The textbook NW test is in the southeast corner of the plot: it prioritizes a small power loss at the cost of a large size distortion. In contrast, KVB prioritizes size control at the cost of a large power loss of 0.23.

These tradeoffs represent points that can be achieved by different choices of b; but which of these points should be chosen? We propose to decide by minimizing a loss function that is quadratic in Δ_S and $\Delta_P^{\max}$:

$$Loss = \kappa(\Delta_S)^2 + (1-\kappa)\left(\Delta_P^{\max}\right)^2. \tag{5}$$

This formulation departs from the classical approach to testing, which insists on perfect size control, and rather takes the practical compromise position that some size distortion is acceptable; that is, it is acceptable to choose a point on the size-power tradeoff in Figure 1. In the spirit of classical testing, we propose putting a large weight on controlling size, specifically, we use $\kappa = 0.9$. The loss function (5) maps out an ellipse, and the tangency point between the ellipse and the tradeoff is the loss-minimizing choice of b, which is shown as a solid star in Figure 1. A theoretical expression for the loss-minimizing bandwidth is given in Section 3; when evaluated for $\kappa = 0.9$ and an AR(1) with $\rho = 0.7$, the result is the rules (3) and (4). The open stars in Figure 1 show the Monte Carlo outcomes when the rules (3) and (4) are used. For EWC, the solid and open stars are very close to each other, but less so for NW.

Finally, we note that the asymptotically optimal test among the class we consider is based on the QS estimator (LLS), however theory also suggests that the EWP/EWC test performs nearly as well. We compared the finite-sample performance of QS and EWP/EWC in Monte Carlo simulations for the case of a single restriction and indeed found that the difference between the two was negligible. Because the fixed-b critical values are standard t and F for EWP/EWC but are nonstandard for QS, we focus on EWP/EWC and not QS.

The remainder of the article is organized as follows. Section 2 summarizes kernel LRV estimators and fixed-b distributions. Section 3 reprises key theoretical results for the Gaussian location model and uses those to derive loss-minimizing bandwidth formulas for general kernel estimators, which we specialize to NW and EWC. Simulation results are given in Section 4 for the location problem and in Section 5 for the regression problem. Section 6 concludes.

2. LRV ESTIMATORS AND FIXED-b CRITICAL VALUES

2.1 LRV Estimators for Time Series Regression

The NW, EWP, and QS estimators are kernel estimators of the long-run variance. The EWC estimator is not a kernel estimator, but rather falls in the class of orthogonal series estimators;

however, to the order of approximation used in this paper, the EWC estimator is asymptotically equivalent to the EWP estimator.

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

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{S}\right)\hat{\Gamma}_j \text{ where } \hat{\Gamma}_j = \frac{1}{T}\sum_{t=\max(1,j+1)}^{\min(T,T+j)} \hat{z}_t\hat{z}'_{t-j}, \quad (6)$$

The Newey-West estimator $\hat{\Omega}^{NW}$ uses the Bartlett kernel $k(v) = (1 - |v|)1(|v|\leq 1)$, and the QS estimator $\hat{\Omega}^{QS}$ uses the Bartlett-Priestley-Epanechnikov quadratic-spectral (QS) kernel, $k(v) = 3[\sin(\pi x)/\pi x - \cos\pi x]\ /(\pi x)^2$ for $x = 6v/5$; see Priestley (1981) and Andrews (1991) for other examples.

The estimator in (6) uses the sample autocovariances of the residual process, that is, $\hat{z}_t = y_t - \hat{\beta} = y_t - \bar{y}$ in the location model and $\hat{z}_t = X_t\hat{u}_t$ in the regression model, where $\hat{u}_t$ is the OLS residual from the unrestricted regression. We refer to estimators of Ω that use $\hat{z}_t$ as unrestricted estimators. Alternatively, $\hat{\Omega}$ can be computed imposing the null, that is, using $\tilde{z}_t(\beta_0) = X_t\tilde{u}_t(\beta_0)$, where $\tilde{u}_t(\beta_0)$ are the residuals from the restricted regression. We exposit the unrestricted estimator, which is the version typically used in practice, but return to the restricted estimator in the Monte Carlo analysis.

The estimator in (6) can alternatively be computed in the frequency domain as a weighted average of the periodogram:

$$\hat{\Omega} = 2\pi\sum_{j=1}^{T-1} K_T(2\pi j/T)I_{\hat{z}\hat{z}}(2\pi j/T), \quad (7)$$

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

$$I_{\hat{z}\hat{z}}(\omega) = (2\pi)^{-1}d_{\hat{z}}(\omega)\overline{d_{\hat{z}}(\omega)}', \text{ where}$$
$$d_{\hat{z}}(\omega) = T^{-1/2}\sum\nolimits_{t=1}^{T}\hat{z}_t\cos\omega t - iT^{-1/2}\sum\nolimits_{t=1}^{T}\hat{z}_t\sin\omega t. \quad (8)$$

The equal-weighted periodogram (EWP) estimator is computed using the Daniell (flat) frequency-domain kernel that equally weights the first $B/2$ values of the periodogram, where B is even:

$$\hat{\Omega}^{EWP} = \frac{2\pi}{B/2}\sum_{j=1}^{B/2} I_{\hat{z}\hat{z}}(2\pi j/T). \quad (9)$$

Kernel estimators are psd with probability one if the frequency-domain weight function $K_T(\omega)$ is nonnegative. The QS, NW, and EWP estimators satisfy this condition.

The EWC estimator. The expression for $d_{\hat{z}}(\omega)$ in (8) shows that the EWP estimator in (9) is the sum of squared normalized coefficients from a regression of $\hat{z}_t$ onto orthogonal sine and cosines, $\{\sin(2\pi j/T), \cos(2\pi j/T)\}, j = 1, \ldots, B/2$. Because the sines and cosines come in pairs, only even values of B are available for EWP. A closely related estimator, which allows for all integer values of B, replaces the sines and cosines of the periodogram with a set of just cosines; like the Fourier terms, these cosines constitute an orthonormal basis for L_2. Specifically, the EWC estimator is,

$$\hat{\Omega}^{\text{EWC}} = \frac{1}{B}\sum_{j=1}^{B}\hat{\Omega}_j, \text{ where } \hat{\Omega}_j = \hat{\Lambda}_j\hat{\Lambda}'_j \quad \text{and}$$
$$\hat{\Lambda}_j = \sqrt{\frac{2}{T}}\sum_{t=1}^{T}\hat{z}_t\cos\left[\pi j\left(\frac{t-1/2}{T}\right)\right]. \quad (10)$$

The projection $\hat{\Lambda}_j$ is the Type II discrete cosine transform of $\hat{z}_t$. The EWC estimator is psd with probability one.

The estimator $\hat{\Omega}^{\text{EWC}}$ cannot be written in the form (6) and thus is not a kernel estimator; rather $\hat{\Omega}^{\text{EWC}}$ is a member of the class of orthogonal series estimators (see Sun (2013) for a discussion and references). To the order of the expansions used here, $\hat{\Omega}^{\text{EWC}}$ is asymptotically equivalent to $\hat{\Omega}^{\text{EWP}}$ (LLS).[6]

2.2 Fixed-*b* Distributions

Fixed-b distributions of $\hat{\Omega}$ are commonly expressed as functionals of Brownian motions (e.g. Kiefer and Vogelsang (2002, 2005) and Sun 2014a). Here, we provide a simpler expression as a finite weighted average of chi-squareds for an important class of kernel estimators, those with frequency domain representations with nonzero weights on the first $B/2$ periodogram ordinates, a class that includes EWP and QS. This derivation shows that the fixed-b asymptotic distribution of EWP is a t distribution. This derivation draws on classical results in spectral analysis, and our exposition amounts to a sketch of the solution to, and minor generalization of, Brillinger's (1975) exercise 5.13.25. To keep things simple, we focus on the case $m = 1$ and a single stochastic regressor, where both y and x have mean zero, and $z_t = x_t u_t$. The arguments here go through for testing the mean of y (regression with an intercept only) and for regression with additional regressors, where (by Frisch-Waugh) they are projected out.

In large samples, $I_{\hat{z}\hat{z}}(2\pi j/T) \xrightarrow{d} s_z(0)\xi_j$ for fixed integer $j > 0$, where $s_z(0)$ is the spectral density of z_t at frequency 0 and $\xi_j \sim \chi^2_2/2$. This is true jointly for a fixed number of periodogram ordinates, where the orthogonality of sines and cosines at harmonic frequencies imply that their limiting distributions are independent. Because $\Omega = 2\pi s_z(0)$, it follows that, for a kernel estimator with frequency domain weights that are nonzero for only the first $B/2$ periodogram ordinates,

$$\hat{\Omega} \xrightarrow{d} \left(\sum_{j=1}^{B/2} K_{T,j}\xi_j\right)\Omega, \text{ where } \xi_j \text{ are iid } \chi^2_2/2,\ j = 1, \ldots, B/2, \quad (11)$$

where $K_{T,j} = K(2\pi j/T)$; see Brillinger (1981, p.145) and Priestley (1981, p. 466). In addition, let $\bar{z}_0 = T^{-1}\sum_{t=1}^{T} z_t(\beta_0)$, where $z_t(\beta_0) = x_t(y_t - \beta_0 x_t)$; then $\sqrt{T}\bar{z}_0 \xrightarrow{d} N(0, \Omega)$ under the null. Because the sine and cosine weights in the periodogram integrate to zero at harmonic frequencies, $\{I_{\hat{z}\hat{z}}(2\pi j/T)\}$,

[6]Phillips (2005) showed that the Type II sine transform has the same mean and variance expansion as EWP. We expect that his calculations could be modified to cover Type II cosines.

$j = 1, \ldots, B/2$ are asymptotically independent of $\bar{z}_0$. Thus, under the null for fixed B,

$$t = \frac{\sqrt{T}\bar{z}_0}{\sqrt{\hat{\Omega}}} \xrightarrow{d} \frac{z^*}{\sqrt{\sum_{j=1}^{B/2} K_j \xi_j}}, \tag{12}$$

where z^* is a standard normal random variable that is distributed independently of the $\chi_2^2/2$ random variables $\{\xi_j\}$. The limiting distribution in (12) is the fixed-b distribution of the HAR t statistic, because fixing $b = S_T/T$ is equivalent to fixing B in the frequency domain (for the EWP estimator, $B = b^{-1}$).

For the EWP estimator (Brillinger's exercise), $K_j = (B/2)^{-1}$ so $\sum_{j=1}^{B/2} K_j \xi_j = (B/2)^{-1} \sum_{j=1}^{B/2} \xi_j \sim \chi_B^2/B$, so the t-ratio in (12) is distributed t_B. With unequal weights, however, the limiting distribution of the t-ratio in (12) is nonstandard.

Tukey (1949) recognized that the distribution in (11) is not in general chi-squared, but can be approximated by a chi squared with what is called the "Tukey equivalent degrees of freedom,"

$$\nu = \left(b \int_{-\infty}^{\infty} k^2(x) dx \right)^{-1}. \tag{13}$$

For EWP, $\int_{-\infty}^{\infty} k^2(x)dx = 1$ so $\nu = b^{-1} = B$, and Tukey's approximation is exact. See Sun (2014a) for additional discussion. When discussing EWP (and EWC, which is asymptotically equivalent to EWP), we therefore refer to the number of included series as ν, which is also the degrees of freedom of the HAR fixed-b t-test.

In the case of $m > 1$ restrictions, the usual test statistic is $F_T = T\bar{z}_0\hat{\Omega}^{-1}\bar{z}_0/m$. For the EWC and EWP estimators, mF_T has a fixed-b asymptotic Hotelling $T^2(m,B)$ distribution. To turn this into a conventional F distribution, we consider the rescaled version of F_T,

$$F_T^* = \frac{B - m + 1}{B} F_T, \tag{14}$$

(see Rao (1973, Section 8.b.xii), Müller (2004), Sun (2013), and Hwang and Sun (2017)). When F_T is evaluated using $\hat{\Omega}^{\mathrm{EWC}}$, $F_T^* \xrightarrow{d} F_{m,B-m+1}$.

We denote the fixed-b critical value for a given kernel by $c_m^\alpha(b)$, for a test with nominal level α.

3. CHOICE OF b IN THE GAUSSIAN LOCATION MODEL

Although our interest is in HAR inference in the time series regression model (1) with stochastic regressors, the theoretical expressions we need are available only for the Gaussian location model. Under appropriately chosen sequences of b that tend to zero as T increases, kernel HAR tests are asymptotically equivalent to first order, yet as discussed in the introduction, simulation results show that they can have quite different performance in finite samples. The workhorse tool for comparing higher-order properties of tests that are first-order asymptotically equivalent is the Edgeworth expansion. Rothenberg (1984) surveys early work on Edgeworth expansions with i.i.d. data. Edgeworth expansions for HAR tests in the Gaussian location model date to Velasco and Robinson (2001), as extended by Sun, Phillips, and Jin (2008), Sun (2013, 2014a), and others. We first reprise these results, then use them to obtain loss-minimizing bandwidth rules.[7]

The Gaussian location model is,

$$y_t = b + u_t. \tag{15}$$

where y_t and u_t are $m \times 1$, β is the m-vector of means of y_t, and u_t is potentially serially correlated. When $m = 1$ and $X_t = 1$, the location and regression models coincide. For the location model, $z_t = u_t = y_t - \beta$; the t-ratio is $t = \sqrt{T}\bar{z}_0/\sqrt{\hat{\Omega}}$ as in (12); and the F-statistic is given in (14).

3.1 Small-b Approximate Size Distortions and Size-Power Frontier

The expansions for kernel HAR tests depend on the kernel through its so-called Parzen characteristic exponent (Parzen 1957), which is the maximum integer such that

$$k^{(q)}(0) = \lim_{x \to 0} \frac{1 - k(x)}{|x|^q} < \infty. \tag{16}$$

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

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

$$\omega^{(q)} = tr\left(m^{-1} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j \Omega^{-1} \right). \tag{17}$$

When $m = 1$ and $q = 2$, $\omega^{(2)}$ is the negative of the ratio of the second derivative of the spectral density of z_t at frequency zero to the value of the spectral density of z_t at frequency zero. If z_t follows a stationary AR(1) process with autoregressive coefficient ρ, then $\omega^{(1)} = 2\rho/(1-\rho^2)$ and $\omega^{(2)} = 2\rho/(1-\rho)^2$.

Higher-order size distortion. Let Δ_S denote the size distortion of the test, that is, the rejection rate of the test minus its nominal level α. Drawing on results from the Edgeworth expansion literature, Sun (2014 a) shows that, for kernel HAR tests using fixed-b critical values,

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

where G_m is the chi-squared c.d.f. with m degrees of freedom, G_m' is its derivative, and χ_m^α is its 100(1-α) percentile.

Higher order size-adjusted power loss. For a comparison of the rejection rates of two tests under the alternative to be meaningful, the two tests need to have the same rejection rate under the null. To illustrate this point, consider the following exact finite-sample example. Suppose y_t is iid $N(\beta,1)$, $t = 1, \ldots, T$. Consider two one-sided tests of $\beta = 0$ vs. $\beta > 0$: φ_1 is a 5% test that rejects if $T^{-1/2}\sum_{t=1}^{T} y_t > 1.645$, and φ_2 is the 10% test that rejects if $(.9T)^{-1/2}\sum_{t=1}^{.9T} y_t > 1.28$, so φ_1 uses all the data and φ_1 uses only 90% of the observations. For small values of β – specifically, for $\beta < (1.645 - 1.28)/(\sqrt{T} - \sqrt{.9T})$ – the test φ_2 rejects more frequently than φ_1 under the alternative. But

[7]Work on expansions in the GMM setting includes Inoue and Shintani (2006), Sun and Phillips (2009), and Sun (2014b).

this does not imply that the *statistic* used to compute φ_2, which discards 10% of the data, is the preferred test statistic, because the power comparisons is between two tests with different sizes. To determine which statistic should be used, the tests must have the same size, which is accomplished by adjusting the critical values. This size-adjustment would change the critical value for φ_1 to 1.28, or alternatively would change the critical value for φ_2 to 1.645. In either case, once the tests use size-adjusted critical values and have the same null rejection rate, the test using all the data has a uniformly higher power function than the test that discards 10% of the data, so the preferred statistic—the statistic with the greater size-adjusted power—is the sample mean using all the data.

Standard practice in the literature on higher-order comparison of tests that are equivalent to first order is to make two tests comparable, to higher order, by using size-adjusted critical values based on the leading higher-order term(s) in the null rejection probability expansion (e.g., Rothenberg (1984, Section 5)). In our case, the size-adjusted critical value, $c^{\alpha}_{m,T}(b)$, is an adjusted version of the critical value $c^{\alpha}_{m}(b)$ such that the leading term on the right hand side of (18) vanishes, that is, $\Pr_0[F^*_T > c^{\alpha}_{m,T}(b)] - \alpha = o(b) + o((bT)^{-q})$. LLS show that,

$$c^{\alpha}_{m,T}(b) = \left[1 + \omega^{(q)}k^{(q)}(0)(bT)^{-q}\right]c^{\alpha}_{m}(b). \tag{19}$$

This adjusted critical value depends on $\omega^{(q)}$ and thus is infeasible; however it allows the power of two tests with different second-order sizes to be compared.

Following LLS, we compare the power of two tests by comparing their worst-case power loss, relative to the infeasible (oracle) test using the unknown Ω, using size adjusted critical values.[8] Let $\delta = T^{-1/2}\Omega^{-1/2}\Sigma^{1/2}_{XX}\beta$ be the local alternative (in the location model, $\Sigma_{XX} = I$). Using results in Sun, Phillips, and Jin (2008) and Sun (2014), LLS show that the worst-case power loss $\Delta^{\max}_P$ against all local alternatives is,

$$\begin{aligned}\Delta^{\max}_P &= \max_\delta \left\{\left[1 - G_{m,\delta^2}(\chi^{\alpha}_m)\right] - \Pr_\delta\left[F^*_T > c^{\alpha}_{m,T}(b)\right]\right\} \\ &= \max_\delta \left[\tfrac{\delta^2}{2}G'_{m+2,\delta^2}(\chi^{\alpha}_m)\chi^{\alpha}_m\right]\nu^{-1} + o(b) + o\left((bT)^{-q}\right)\end{aligned} \tag{20}$$

where G_{m,δ^2} is the noncentral chi-squared cdf with m degrees of freedom and noncentrality parameter δ^2, G'_{m,δ^2} is its first derivative, and ν is the Tukey equivalent degrees of freedom (13).

The worst-case power loss occurs for the value of delta against which the oracle test has a rejection rate of 66% (a numerical finding that appears not to depend on m). Thus, the worst-case power can instead be thought of as power against this particular point alternative. This alternative seems reasonable to us and is in the range typically considered when constructing point-optimal tests. Two different strategies would be to consider power against an alternative much closer to the null, or instead a very different alternative. Doing so would produce different constants in the rules (3) and (4).

[8] In general, the higher-order terms in Edgeworth expansions can result in power curves that cross, so that the ranking of the test depends on the alternative, see Rothenberg (1984). Here, however, the higher order size-adjusted power depends on the kernel only through ν (see (20)) so the power ranking of size-adjusted tests is the same against all alternatives. Thus using, say, an average power loss criterion, would not change the ranking of tests, although it would change the constant in (20).

Size-power tradeoff. One definition of an optimal rate for b is that it makes Δ_S and $\Delta^{\max}_P$ vanish at the same rate; this condition is an implication of minimizing the loss function (5). This rate is $b \propto T^{-q/(1+q)}$ or equivalently $\nu \propto T^{q/(1+q)}$, for which Δ_S and $\Delta^{\max}_P$ are both $O(T^{-q/(1+q)})$. It follows that, asymptotically, size distortions and power losses for all $q = 2$ kernels dominate those for all $q = 1$ kernels.

For sequences with this optimal rate, the leading terms in (18) and (20) map out the tradeoff between absolute value of the size distortion and power loss, which LLS show is given by

$$\Delta_p(\delta)|\Delta_S|^{1/q} = a_{m,\alpha,q}(\delta)\left[\left(k^{(q)}(0)\right)^{1/q}\int_{-\infty}^{\infty}k^2(x)dx\right]\left|\omega^{(q)}\right|^{1/q}T^{-1}, \tag{21}$$

where $a_{m,\alpha,q}(\delta) = \frac{1}{2}\delta^2 G'_{m+2,\delta^2}(\chi^{\alpha}_m)\chi^{\alpha}_m(G'_m(\chi^{\alpha}_m)\chi^{\alpha}_m)^{1/q}$. Note that the absolute value of Δ_S can be replaced by Δ_S if z_t is persistent ($\omega^{(q)} > 0$). The asymptotic rate for this tradeoff is better for $q = 2$ (the largest value consistent with psd kernels) than for $q = 1$. Thus the best-possible tradeoff for all psd kernels minimizes $\sqrt{k^{(q)}(0)}\int_{-\infty}^{\infty}k^2(x)dx$, which is achieved by the QS kernel. Among tests with fixed-B distributions that are exact t_ν, LLS show that the best-possible tradeoff is achieved by the EWP and EWC tests.

3.2 Loss Function Approach to Choosing b

Choosing a bandwidth sequence amounts to choosing a point on the tradeoff curve (21). We make that choice by minimizing the quadratic loss function (5).

Specifically, consider sequences $b = b_0T^{-q/(1+q)}$, and minimize (5) over the constant b_0, where Δ_S and $\Delta^{\max}_P$ are respectively given by their leading terms in (18) and in the final line of (20). Because the spectral curvature $\omega^{(q)}$ enters (18), minimizing the loss requires either knowledge of $\omega^{(q)}$ or an assumption about $\omega^{(q)}$. In either case, let $\bar{\omega}^{(q)}$ denote the value of $\omega^{(q)}$ used in the minimization problem.

Minimizing the quadratic loss function (5) with $\omega^{(q)} = \bar{\omega}^{(q)}$ yields the sequence,

$$b^* = b_0T^{-q/(1+q)}, \text{ where } b_0 = \kappa_q d_{m,\alpha,q}\left(\frac{k^{(q)}(0)}{\int k^2}\right)^{1/(1+q)}\left(\bar{\omega}^{(q)}\right)^{1/(1+q)}, \tag{22}$$

where $\kappa_q = [q\kappa/(1-\kappa)]^{\frac{1}{2(1+q)}}$ and $d_{m,\alpha,q} = \{2G'_m(\chi^{\alpha}_m)/\max_\delta[\delta^2 G'_{m+2,\delta^2}(\chi^{\alpha}_m)]\}^{1/(1+q)}$ (see the online supplement for the derivation). The minimum-loss equivalent degrees of freedom (13) implied by (22) is,

$$\begin{aligned}&\nu^* = \nu_0T^{q/(1+q)}, \text{ where} \\ &\nu_0 = \kappa_q^{-1}d^{-1}_{m,\alpha,q}\left[\left(k^{(q)}(0)\right)^{1/q}\int k^2\right]^{-q/(1+q)}\left(\bar{\omega}^{(q)}\right)^{-1/(1+q)}.\end{aligned} \tag{23}$$

Table 2 evaluates the constants collectively multiplying $T^{-q/(1+q)}$ in (22) for various values of κ and $\bar{\omega}^{(q)}$ for NW and EWC. In the table, the assumed curvature is parameterized by the curvature for an AR(1) with autoregressive coefficient $\bar{\rho}$, a representation we find more directly interpretable. As can be seen from (22), (23), and the table, b^* is increasing and ν^* is decreasing in both $\bar{\rho}$ and κ.

Table 2. Constants in loss-minimizing rules for b for different values of assumed AR(1) parameter $\bar{\rho}$ and loss function parameter κ, for $m = 1$

(a) NW: $b = b_0 T^{-1/2}$, where b_0 is:

$\bar{\rho}\backslash\kappa$	0.5	0.75	0.8	0.85	0.9	0.95	0.99
0.1	0.20	0.27	0.29	0.31	0.35	0.43	0.64
0.2	0.29	0.39	0.41	0.45	0.51	0.61	0.92
0.3	0.37	0.48	0.52	0.57	0.64	0.77	1.16
0.4	0.44	0.58	0.63	0.68	0.77	0.92	1.40
0.5	0.52	0.69	0.74	0.81	0.91	1.09	1.65
0.6	0.62	0.82	0.88	0.96	1.08	1.30	1.96
0.7	0.75	0.99	1.06	1.16	**1.30**	1.57	2.37
0.8	0.96	1.26	1.35	1.48	1.66	2.00	3.02
0.9	1.40	1.84	1.97	2.15	2.42	2.92	4.40

(b) EWC: $\nu = \nu_0 T^{2/3}$, where ν_0 is:

$\bar{\rho}\backslash\kappa$	0.5	0.75	0.8	0.85	0.9	0.95	0.99
0.1	2.33	1.94	1.85	1.75	1.62	1.43	1.08
0.2	1.71	1.43	1.36	1.28	1.19	1.05	0.80
0.3	1.37	1.14	1.09	1.02	0.95	0.84	0.64
0.4	1.12	0.93	0.89	0.84	0.78	0.69	0.52
0.5	0.92	0.77	0.73	0.69	0.64	0.56	0.43
0.6	0.75	0.62	0.59	0.56	0.52	0.46	0.35
0.7	0.59	0.49	0.47	0.44	**0.41**	0.36	0.27
0.8	0.43	0.36	0.34	0.32	0.30	0.26	0.20
0.9	0.26	0.22	0.21	0.19	0.18	0.16	0.12

NOTES: For NW in part (a), entries evaluate b_0 in Equation (22), where $\bar{\omega}^{(1)} = 2\bar{\rho}/(1-\bar{\rho}^2)$. For EWC in part (b), entries evaluate ν_0 in Equation (23), where $\bar{\omega}^{(2)} = 2\bar{\rho}/(1-\bar{\rho})^2$. The values corresponding to $\bar{\rho} = 0.7$ and $\kappa = 0.9$ are bolded.

The size distortions and power losses along the optimal sequence, which are obtained by substitution of (22) and (23) into (18) and (20), are

$$\Delta_S^* = \kappa_q^{-q}\left(\bar{a}_{m,\alpha,q}\left(k^{(q)}(0)\right)^{1/q}\int k^2\right)^{q/(1+q)} \times\left(\frac{\omega^{(q)}}{\bar{\omega}^{(q)}}\right)^{q/(1+q)}\left(\omega^{(q)}\right)^{1/(1+q)}T^{-q/(1+q)}, \text{ and} \quad (24)$$

$$\Delta_P^{\max *} = \kappa_q\left(\bar{a}_{m,\alpha,q}\left(k^{(q)}(0)\right)^{1/q}\int k^2\right)^{q/(1+q)} \times\left(\bar{\omega}^{(q)}\right)^{1/(1+q)}T^{-q/(1+q)}. \quad (25)$$

The rule choice parameters enter the size distortion and power loss expressions in opposite ways. Larger values of κ and/or $\bar{\rho}$ lead to larger optimal values of b, less size distortion, and more power loss for a given true value ρ.

Figure 2 plots the theoretical size distortion and power loss in (24) and (25) for selected values of κ and $\bar{\rho}$. The preference parameters considered all place more weight on size than power, and the size distortion curves in these plots all lie below the power curves. The rates of the EWC curves are $T^{-2/3}$ and the rates of the NW curves are slower, $T^{-1/2}$, which is evident in Figure 2(b). A striking feature of these figures is that NW in Figure 2(b) has essentially the same size distortion but lower power loss than EWC (both using the proposed rules and fixed-b critical values) through $T \approx 300$. This is another manifestation of the crossing of the EWC and NW tradeoffs in Figure 1, and indicates that the asymptotic theoretical dominance of EWC and QS implied by the rates does not imply theoretical dominance at moderate, or even fairly large, values of T.

The choice of a specific constant from Table 2 requires choosing values of $\bar{\rho}$ and κ. Consistent with our focus on the problem of moderate persistence, we propose to use the spectral curvature corresponding to an AR(1) with $\bar{\rho} = 0.7$. Concerning κ, we adopt a value that weights size control much more heavily than power loss, and suggest $\kappa = 0.9$. The corresponding rule constants are bolded in Table 2 and produce the rules (3) and (4).

The theoretical performance of these rules is examined in Figures 2(c) and 2(d), for various true values of ρ. As the true ρ gets smaller, the size of each of the two tests improves. The power loss, however, depends on $\bar{\rho}$ but not on ρ so does not improve as ρ diminishes.

The constant in the optimal rule also depends on the number of restrictions being tested, m. Table 3 provides these constants for $1 \le m \le 10$ (the values for $m = 1$ are those in bold in Table 2). For both tests, the optimal value of b decreases with m (recall that for EWC, $\nu = b^{-1}$). For NW, the optimal S decreases by one-fourth from $m = 1$ to $m = 5$; for EWC, the optimal ν increases by one-fifth from $m = 1$ to $m = 5$. In practice, there are substantial advantages to producing a single estimate of Ω instead of using different tuning parameters for tests with different numbers of restrictions. Because the most common use of HAR inference is for a single restriction and confidence intervals for a single coefficient, we adopt b_0 and ν_0 for $m = 1$.

4. MONTE CARLO RESULTS I: ARMA/GARCH DGPs

The Monte Carlo results presented in this and the next section have two purposes. The first is to assess how well the theoretical tradeoffs characterize the finite-sample tradeoff between size and power. The second is to assess the finite-sample performance of the EWC and NW tests using proposed rules (3) and (4), including comparing their performance to each other and to the textbook NW test. In this section, the pseudo-data are generated from AR models, in some cases with GARCH errors and in some cases with non-Gaussian errors. The replication files for this article allow the reader to investigate finite sample tradeoffs for other data-generating processes and regression specifications. In the next section, the pseudo-data are generated according to a data-based design.

We first consider the location model, then turn to results for regressions. In all, we consider 11 DGPs in this section. Here, we provide illustrative results and high-level summaries; detailed results are provided in the online Supplement.

4.1 ARMA/GARCH DGPs

Location model. The disturbance u_t is generated according to the AR model $\rho(\mathrm{L})u_t = \eta_t$, with i.i.d. disturbances η_t with mean zero and variance 1, potentially with GARCH errors. The disturbance distribution includes Gaussian and non-Gaussian cases, where the non-Gaussian distributions are normal mixture distributions 2–5 from Marron and Wand (1992). The skewness and kurtosis of the innovations is summarized in the first two numeric columns of Table 4. The final two columns of Table 4

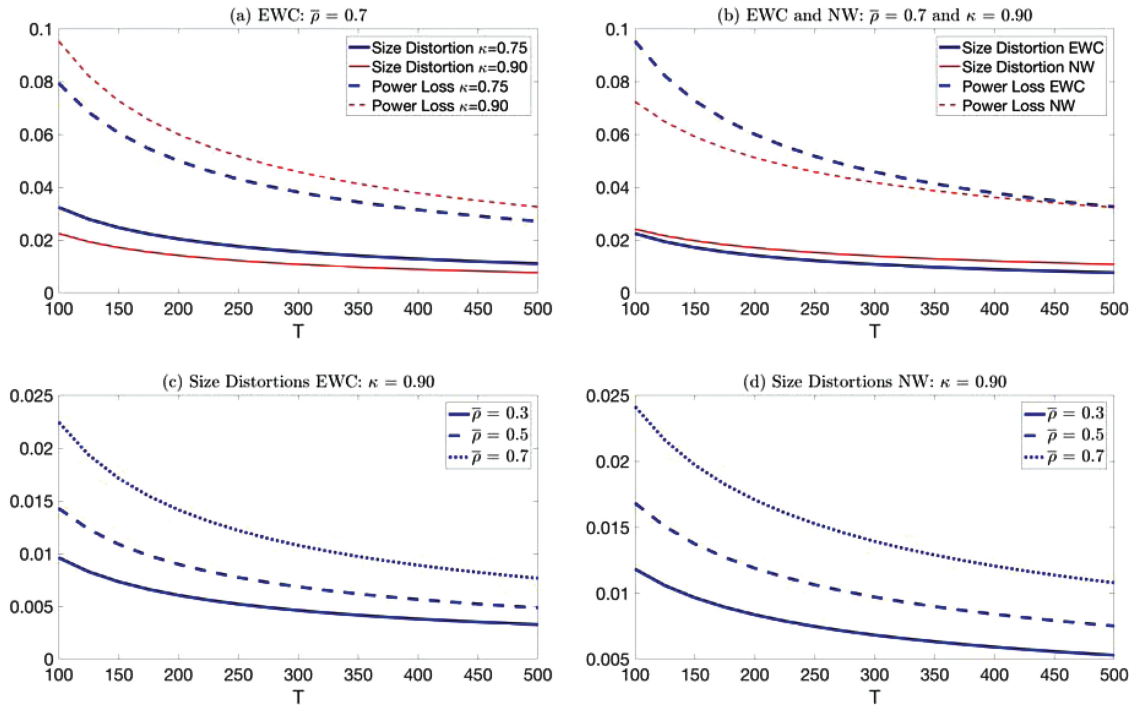

Figure 2. Theoretical size distortion and size-adjusted power loss along optimal bandwidth sequences for various values of preference parameter κ and autoregressive parameter $\bar{\rho}$. Notes: The plotted curves evaluate the size distortion Δ_S in (24) and power loss $\Delta_P^{\max}$ in (25), based on the loss-minimizing bandwidth sequence (22), for various values of loss parameter κ, rule parameter $\bar{\rho}$, and true value ρ.

Table 3. Constants in loss-minimizing rules for different values of m using parameter $\bar{\rho} = 0.7$, loss function parameter $\kappa = 0.9$, and $\alpha = 0.05$. Values of ν_0 for EWC and b_0 for NW

	m									
	1	2	3	4	5	6	7	8	9	10
NW: $S = b_0 T^{1/2}$, where $b_0 =$	1.30	1.15	1.07	1.01	0.97	0.93	0.90	0.88	0.86	0.84
EWC: $\nu = \nu_0 T^{2/3}$, where $\nu_0 =$	0.41	0.44	0.46	0.48	0.50	0.51	0.52	0.53	0.54	0.55

NOTES: Entries evaluate the constants in Equations (22) and (23); see the notes to Table 2.

provide the skewness and kurtosis of the AR(1) process for u_t with $\rho = 0.7$, which has less departure from normality than do the innovations. The null hypothesis is $\beta_0 = 0$. Data were generated under the local alternative δ by setting $y_t = \beta + u_t$, where $\beta = \sigma_u T^{-1/2}\delta$.

The size-adjusted critical values, which are used to compute $\Delta_P^{\max}$, are computed as the 100(1-α) percentile of the Monte Carlo distribution of the test statistic under the null for that design.

Table 4. Distributions of innovations and u_t: skewness and kurtosis

	skewness	kurtosis	skewness	kurtosis
Distribution	Innovations		u_t (AR(1), $\rho = 0.7$)	
$N(0,1)$	0	3	0	3
Marron-Wand 2	−0.7	4.0	−0.4	3.4
Marron-Wand 3	1.5	4.7	0.8	3.6
Marron-Wand 4	0	4.6	0	3.5
Marron-Wand 5	0	25.3	0	10.6

Regressions. Here, we consider regressions with a single stochastic regressor, that is, $y_t = \beta_0 + \beta_1 x_t + u_t$, where x_t and u_t follow independent Gaussian AR(1)'s. Results for additional ARMA/GARCH designs are reported in the online supplement; in general, the results for the more complicated designs give the same conclusions as the results for the independent AR(1) design here. Under the null, $\beta_1 = 0$, so y_t and x_t are two independent AR(1)'s.

We examine LRV estimators computed under the null (the restricted case) and under the alternative. In both cases the intercept is estimated. Thus the restricted LRV estimator uses $\tilde{z}_t = (x_t - \bar{x})(y_t - \beta_0 x_t - (\bar{y} - \beta_0\bar{x}))$, and the unrestricted LRV estimator uses $\hat{z}_t = (x_t - \bar{x})\hat{u}_t$, where $\hat{u}_t$ is the OLS residual. Sample autocovariances for the restricted NW estimator are computed as $T^{-1}\sum_t (\tilde{z}_t - \bar{\tilde{z}}_t)(\tilde{z}_t - \bar{\tilde{z}}_t)'$; the EWC estimator is invariant to the mean of $\tilde{z}_t$ so does not require demeaning.

To compute size-adjusted power, data on y under the alternative were generated as $y_t = \beta_1 x_t + u_t$, for some nonzero alternative value of β_1, given x and u. Thus changing β_1 under the alternative changes the value of y for a given draw of

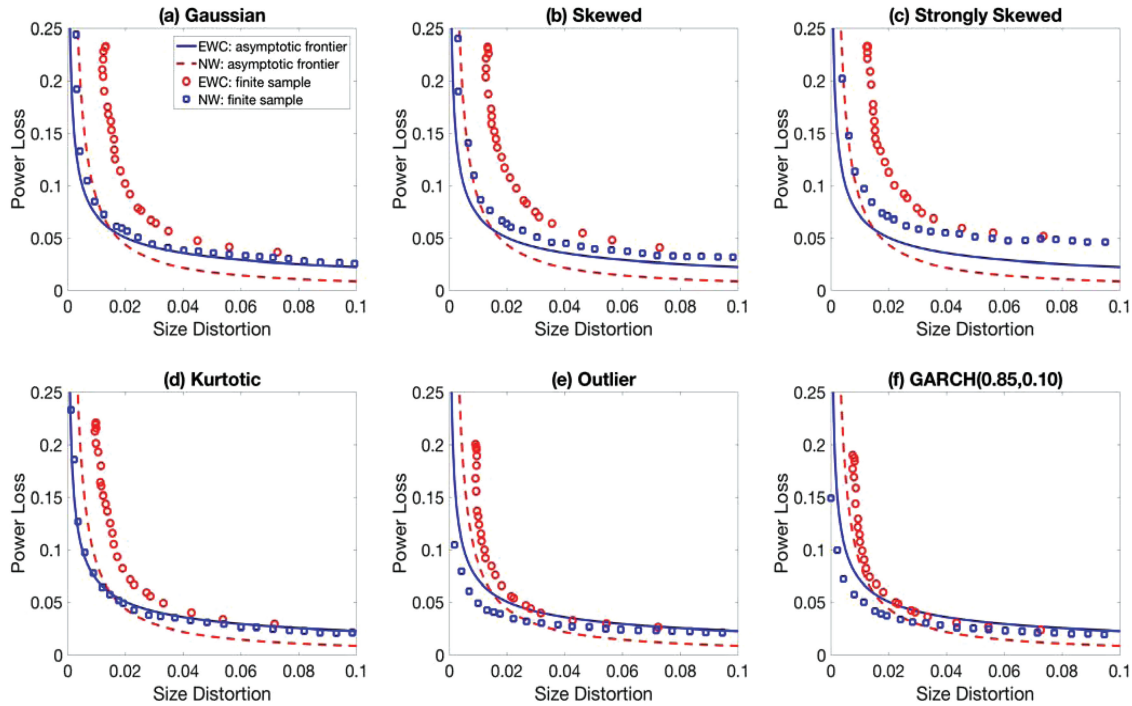

Figure 3. Size-power tradeoffs in the location model with Gaussian and non-Gaussian innovation distributions: Theoretical (lines) and Monte Carlo (symbols) for EWC (solid line & squares) and NW (dashed line & circles), $T = 200$, AR(1) disturbances with coefficient $\rho = 0.7$. Notes: The theoretical frontiers are from (21), the Monte Carlo frontiers (circles) are computed for different error distributions. In the first five figures, u_t follows an AR(1) with $\rho = 0.7$, with innovation distributions given in rows 1–5 of Table 4. For the final figure, the DGP is an AR(1) with coefficient 0.7, with errors that follow a highly persistent GARCH process.

x and u, but does not change z_t, the OLS residuals $\hat{u}_t$, the true long-run variance, or its unrestricted estimate based on $\hat{z}_t$. However, changing β_1 for a given draw of x and u does change $\tilde{z}_t$ and therefore changes the restricted estimate of the LRV (because the null is no longer true). The size-adjusted critical value is computed as the 95th percentile of the (two-sided) test statistic under the null, and the size-adjusted power is the rejection rate under the alternative using the size-adjusted critical value.

4.2 Results for the Location Model

Figure 3 presents the size-power tradeoff (Δ_S v. $\Delta_P^{\max}$) for six DGPs. The first figure presents results for Gaussian innovations to an AR(1) process with $\rho = 0.7$ and $T = 200$; these results also appear in Figure 1. As in Figure 1, the symbols are Monte Carlo results and the lines are the theoretical tradeoffs (21). The next four figures show results using Marron-Wand distributions 2–5 for the innovations, and the final figure shows results for GARCH errors with high GARCH persistence.

The results in Figure 3, combined with additional results in the supplement, suggest two conclusions. First, under Gaussianity, the theoretical approximations are numerically close to the Monte Carlo results for EWC but not for NW, where the finite-sample performance of NW is substantially worse than predicted by theory. The remainder terms in the expansion are $o(b^{-q/(1+q)})$, which is $o(T^{-1/2})$ for NW but $o(T^{-2/3})$ for EWC, so the remainder terms could be larger for NW than EWC – as appears to be the case in Figure 3. Nevertheless, the Monte Carlo points trace out a tradeoff with a similar shape to the theoretical tradeoff, just shifted to the northeast.

Second, although we would not expect the Gaussian tradeoffs to hold for non-Gaussian innovations (Velasco and Robinson 2001), the findings for Gaussianity seem to generalize to moderate departures from Gaussianity (Marron-Wand distributions 2 and 4), and even to heavily skewed distributions (Marron-Wand 3). Interestingly, in the cases of heavy-tailed distributions (Marron-Wand 5 and GARCH), the EWC Monte Carlo tradeoffs lie below the theoretical Gaussian tradeoffs.

Performance of proposed rules as a function of T. Figure 4 examines the performance of the proposed rules (3) and (4) as a function of T, for a Gaussian AR(1) DGP with $\rho = 0.7$. The theoretical tradeoffs are the same as in Figure 2; in finite samples, however, EWC performs better than NW. For EWC, the Monte Carlo size distortion and power loss are close to the theoretical predictions, but for NW they exceed the theoretical predictions, even for large values of T.

4.3 Results for Regressions

Figure 5 and Figure 6 are the same as Figure 3 and Figure 4, but now for the regression model. Figure 5 and Figure 6 present Monte Carlo tradeoffs for the regression model with a single stochastic regressor, in which u_t and x_t are independent Gaussian AR(1)'s with AR coefficients $\rho_u = \rho_x = \sqrt{0.7}$. Figure 7 summarizes size distortion and power loss results for the Gaussian AR(1) design with different values of $\rho_u = \rho_x$ (0.53, 0.64, 0.73, 0.80, which yield values of $\omega^{(2)}$ of 5, 10, 20, and 40); this figure

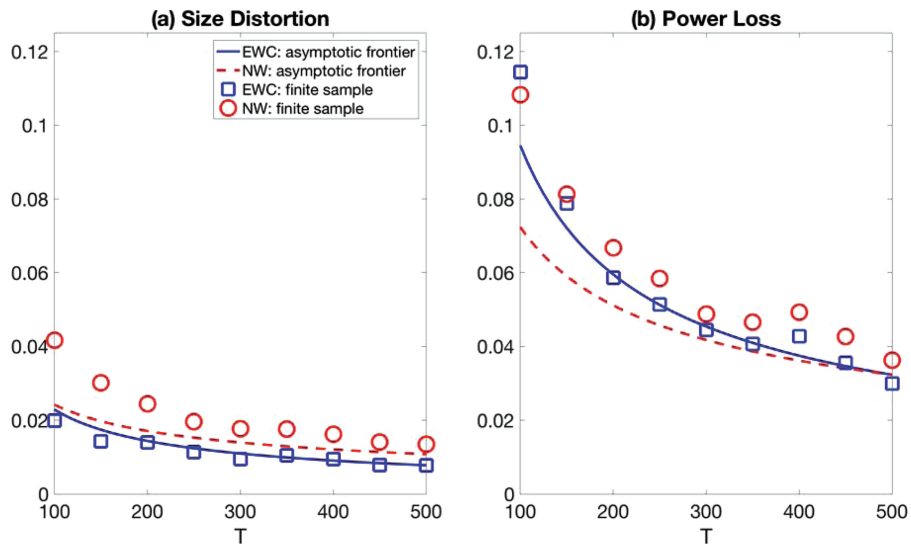

Figure 4. Monte Carlo (symbols) and theoretical values (lines) of size distortions and size-adjusted power losses for EWC (solid line & squares) and NW (dashed line & circles) tests in the Gaussian location model, Gaussian AR(1) disturbances with $\rho = 0.7$. Notes: The tests are evaluated for the loss-minimizing rules (3) (NW) and (4) (EWC).

shows results when S and ν are selected by rules (3) and (4) as well as values of b and ν that are half and twice as large as the proposed rule. The results for this AR(1) case are representative of those for other designs described in the online Supplement.

These results suggest four conclusions. First, because the marginal distribution of $x_t u_t$ is heavy tailed, one would expect the Gaussian location model tradeoff to be an imperfect approximation to the regression model tradeoff even if the null is imposed (i.e. the restricted LRV estimator is used). In fact, the Monte Carlo tradeoffs, particularly for the unrestricted LRV estimators, are further from the frontiers than many of those for the non-Gaussian disturbances in Figure 3.

Second, the performance of EWC is almost always better than, and only rarely worse than, NW for a given design.

Third, for a given b, the size is improved, with a slight power loss, by imposing the null for estimating the LRV. Often this size improvement is large. In Figure 5, if the restricted LRV estimator is used, the performance of the regression test is comparable to the location model, but if the unrestricted estimator is used, the tradeoff shifts far to the northeast, for both NW and EWC. The improvements from imposing the null persist with large sample sizes, and the Monte Carlo tradeoff for the test using the restricted estimator dominates that for the unrestricted estimator. This outward shift indicates that there are potentially large gains to testing by imposing the null, at least in some designs.

Fourth, the results in Figure 7 suggest that the tests based on the proposed rules provide a reasonable balance between size distortions and power loss, consistent with the intuition of Figure 1. Doubling S (halving ν) results in somewhat better size control but much worse power loss, while halving S (doubling ν) has the opposite effect. The textbook rule represents an extreme case of low power loss but high size distortion.

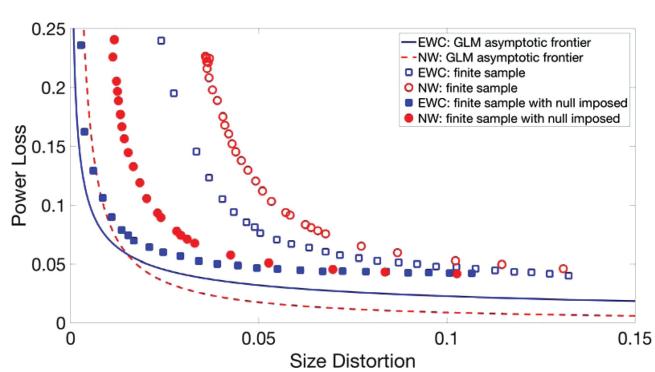

Figure 5. Monte Carlo size-power tradeoffs for the coefficient on a single stochastic regressor: unrestricted (open symbols) and restricted (solid symbols) LRV estimator, with theoretical tradeoff for Gaussian location model (line). Notes: Monte Carlo data on x_t and u_t are generated as independent Gaussian AR(1)'s with AR(1) parameters $\rho_x = \rho_u = \sqrt{.7}$ and $T = 200$. All regressions include an estimated intercept.

5. MONTE CARLO RESULTS II: DATA-BASED DGP

In this section, we examine the finite-sample performance of the proposed tests in a design that aims to reflect the time series properties of typical macroeconomic data. To this end, we fit a dynamic factor model (DFM) to a 207-series data set comprised of quarterly U.S. macroeconomic time series, taken from Stock and Watson (2016). We then simulate data from the estimated DFM and, with these data, test a total of 24 hypotheses, the location model plus 23 different time series regression tests that require HAR inference. Of the 23 regression tests, 9 test a single restriction ($m = 1$) and 14 test a joint hypothesis. Because the DFM parameters are known, we know the true value of the parameter(s) being tested and the true long-run variance, and can therefore compute the null rejection rate, size-adjusted critical values, power of the oracle test, and size-adjusted power loss for each of the tests.

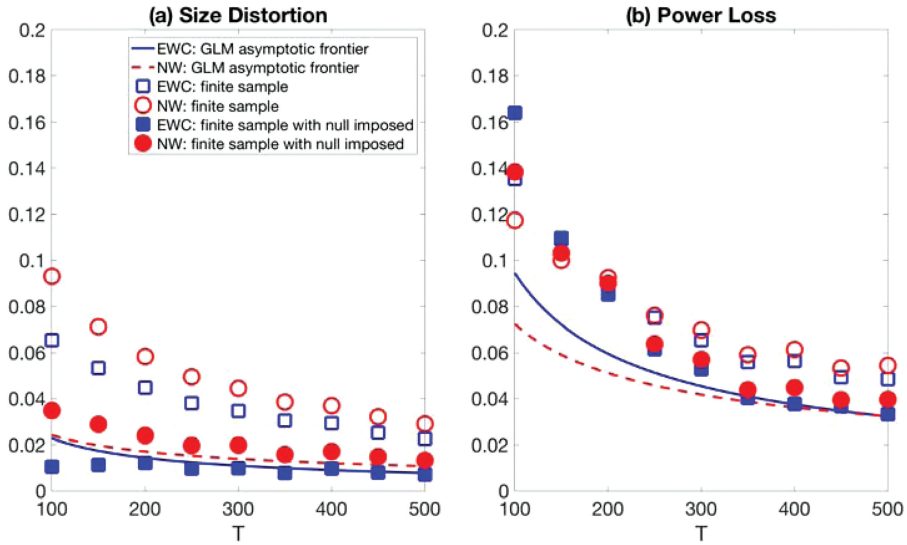

Figure 6. Monte Carlo size distortions and size-adjusted power losses for EWC and NW tests for the coefficient on a single stochastic regressor, with theoretical lines for the Gaussian location model. Notes: Monte Carlo data on x_t and u_t are generated as independent Gaussian AR(1)'s with $\rho_x = \rho_u = \sqrt{7}$. The tests are evaluated for the loss-minimizing rules (3) (NW) and (4) (EWC).

5.1 Data and Dynamic Factor Model DGP

The data set is taken from Stock and Watson (2016). The data are quarterly, 1959Q1-2014Q4, and include NIPA data, other real activity variables, housing market data, prices, productivity measures, and interest rates and other financial series. The series were transformed to be approximately integrated of order zero, for example real activity variables were typically transformed using growth rates. Some outliers were removed, and to eliminate low frequency trends, the series were then filtered using a low-frequency biweight filter (bandwidth

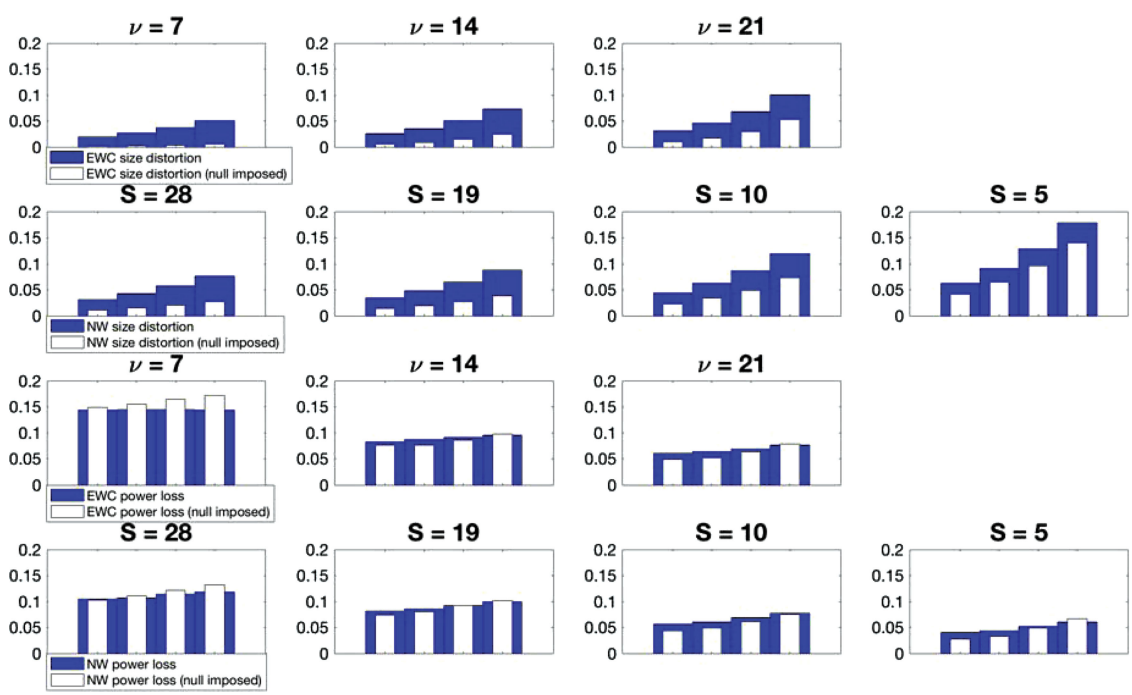

Figure 7. Size distortion and power loss for NW and EWC tests in the regression model with independent Gaussian AR(1) x_t and u_t, for the rules (3) and (4) (second column), larger b (first column), smaller b (right column), and textbook NW rule (furthest right column). Open bars are restricted, shaded bars are unrestricted LRV estimators. $T = 200$. Notes: First row is EWC size distortion, second row is NW size distortion, third row is EWC power loss, fourth row is NW power loss. Within each figure, the four bars are for AR(1)s with $\rho = (0.53, 0.64, 0.73, 0.80)$, which give rise to $\omega^{(2)} = (5, 10, 20, 40)$. The rule (4) selects $\nu = 14$ for EWC, and the rule (3) selects $S = 19$ for NW. The test based on NW textbook rule, which yields $S = 5$, is shown in the final column. All regressions include an estimated intercept.

100 quarters). For additional details, see Stock and Watson (2016).

These transformed series were then used to estimate a 6-factor DFM, where the factors follow a first-order vector autoregression and the idiosyncratic terms follow an AR(2). Let X_t denote the vector of 207 variables and let F_t denote the factors. The DFM is,

$$X_t = \Lambda F_t + e_t \tag{26}$$

$$F_t = \Phi F_{t-1} + \eta_t \tag{27}$$

$$\delta_i(L)\, e_{it} = \nu_{it}, \tag{28}$$

where Λ and Φ are coefficient matrices, δ_i(L) is an AR(2) lag polynomial, $E\eta_t\eta_t' = \Sigma_\eta$ and $E\nu_{it}\nu_{jt} = \sigma^2_{\nu_i}$ for $i = j$ and $= 0$ for $i \neq j$.

The estimated DFM was then used as the DGP for generating pseudo-data using serially uncorrelated draws of η_t (with covariance matrix Σ_η) and ν_{it}.

The disturbances (η_t, ν_t) were drawn in two ways: from an i.i.d. Gaussian distribution, and from the empirical distribution using a time reversal method that ensures that the disturbances are serially uncorrelated but preserve higher moment time series structure (such as GARCH). We describe this here for drawing from the joint distribution of two series, i and j, this extends to drawing from the joint distribution of additional series. Let T^{DFM} denote the number of observations used to estimate the DFM, let $\hat{\eta} = [\,\hat{\eta}_1\ \hat{\eta}_2\ \cdots\ \hat{\eta}_{T^{\mathrm{DFM}}}\,]$ denote the $6 \times T^{DFM}$ matrix of factor VAR residuals, let $\hat{\eta}^R = [\,\hat{\eta}_{T^{\mathrm{DFM}}}\ \hat{\eta}_{T^{\mathrm{DFM}}-1}\ \cdots\ \hat{\eta}_1\,]$ denote the $6 \times T^{\mathrm{DFM}}$ matrix of time-reversed factor VAR residuals, and similarly let $\hat{\nu}_i$ and $\hat{\nu}_i^R$ denote the $1 \times T^{\mathrm{DFM}}$ matrices of the ith idiosyncratic residual and its time-reversal. Also let $\Xi^{r,T}$ denote the $r \times T$ random matrix with elements that are ± 1, each with probability ½, where T is the number of observations used in the Monte Carlo regressions. The two methods of drawing the disturbances are,

(i) *Gaussian disturbances:* $\eta_t \sim$ iid $N(0,\Sigma_\eta)$, $\nu_{it} \sim$ iid $N(0,\sigma^2_{\nu_i})$, $\nu_{jt} \sim$ iid $N(0,\sigma^2_{\nu_j})$, where η_t, ν_{it}, and ν_{jt} are mutually independent.

(ii) *Empirical disturbances:* Randomly select a $6 \times (T + 500)$ segment $\tilde{\eta}$ from the matrix $[\,\ldots\hat{\eta}^R\ \hat{\eta}\ \hat{\eta}^R\ \hat{\eta}\ \ldots]$ and randomly draw a matrix $\Xi^{6,T+500}$; the factor VAR innovations for this draw are $\Xi^{6,T+500} \circ \tilde{\eta}$, where $\circ$ is the Hadamard product. Similarly, draw $\tilde{\nu}_i$ as a randomly chosen $1 \times (T + 500)$ segment of $[\,\ldots\ \hat{\nu}_i^R\ \hat{\nu}_i\ \hat{\nu}_i^R\ \hat{\nu}_i\ \ldots]$ and randomly draw $\Xi^{1,T+500}$; the idiosyncratic innovations for series i are $\Xi^{1,T+500} \circ \tilde{\nu}_i$. The idiosyncratic innovations for series j are constructed in the same way as for series i.[9]

In both cases, we generated time series of length $T + 500$ from the disturbances, which were used to generate the data, initializing the presample values of F and e with zeros. The first 500 observations were discarded to approximate a draw from the stationary distribution.

[9]The sampling scheme is described here as drawing from an infinite dimensional matrix. In practice it is implemented by randomly selecting a starting date, randomly selecting whether to start going forward or backwards in time, then obtaining the first T observations of the sequence $\hat{\eta}_\tau, \hat{\eta}_{\tau\pm1}, \hat{\eta}_{\tau\pm2}, \ldots$, with the direction of time reversed upon hitting $t = 1$ or T.

The scheme for sampling from the empirical disturbances has three desirable properties. First, because of multiplication by the random matrices Ξ, the disturbances are serially uncorrelated. Second, it preserves all even moments of the residuals (thus preserves outliers). Third, it preserves univariate time series structure in the second moments, that is, the fourth-order spectrum of the sampled series is the same as the fourth-order spectrum of the residuals. This approach draws on the concepts in MacKinnon (2006).

To compute power, data on y under the alternative were generated using values of β_1 that depart from the null, given x and u, where u is the (known) error term under the null. Specifically, consider the case $m = 1$, $y_t = \beta_1 x_t + \gamma w_t + u_t$, where w_t are additional regressors (possibly just the intercept), and let $\beta_{1,0}$ denote the known true values of β_1. For a given draw of (y_t, x_t, w_t), we use Frisch-Waugh and use the residuals of the regression on w_t, $y_t^\perp$ and $x_t^\perp$, to compute the error under the null, $u_t^\perp = y_t^\perp - \beta_{1,0}x_t^\perp$. Thus, given a draw of (x, w, u), a draw of $y_t^\perp$ under the alternative is computed as $y_t^\perp = (\beta_{1,0} + \delta)x_t^\perp + u_t^\perp$, where δ is a departure from the null. This mirrors the generation of (y, x) data under the alternative in the AR design in Section 4. Thus changing β_1 under the alternative does not change z_t ($= x_t^\perp u_t^\perp$), the OLS residuals $\hat{u}_t$, the true long-run variance, or its unrestricted estimate based on $\hat{z}_t$; however, changing β_1 for a given draw of x and u does change $\tilde{z}_t$ and therefore changes the restricted estimate of the LRV (because the null is no longer true).

For all the Monte Carlo results in this section we use $T = 200$, with 50,000 replications for each design.

5.2 Design and Computation

Because there are 207 series, there are 207 distinct DGPs for the location model, and we evaluate the performance of the tests in the location model for all 207 series.

For regressions with a single regressor (possibly including lags), there are a total of $207 \times 206 = 42{,}642$ bivariate DGPs. This is more than we could analyze computationally, so we sampled from all possible DGPs using a stratified sampling scheme. Specifically, each regression design involves a pair of variables (y, x) and possibly additional stochastic regressors w. For each design, each of the 207 variables was used as y along with a randomly selected variable for x (and w as needed); each of the 207 variables was also used as x with a randomly selected variable y (and w as needed). This yielded 414 DGPs for each regression design. Each design was implemented with Gaussian errors, then with empirical errors, for a total of 828 DGPs for each of the 23 designs.

Regressions. The regressions are summarized in Table 5. The 23 regression specifications include 3 distributed lag regressions (regressions 1–3), 11 single-predictor direct forecasting regressions (4–15), 4 local projection regressions (16–19), and 4 multiple-predictor direct forecasting regressions. In the table, a generic error term u_t is included so that the regressions appear in standard form, however the error term is not used to generate the data, rather, the dependent variable and the regressors are jointly drawn from the DFM directly as described in the previous section and knowledge of the DFM parameters is used to compute the true values of the coefficients being tested.

Table 5. Regressions Studied in the Data-Based Monte Carlo Simulation

	Description	m	Other	Test coefficients:
1		1		β_1
2	Distributed lag: $y_t = \beta_0 + \sum_{i=1}^{m} \beta_i x_{t-i+1} + u_t$	5		$\beta_1, \ldots, \beta_5$
3		10		$\beta_1, \ldots, \beta_{10}$
4		1	$h = 1, p = 0$	β_1
5		1	$h = 4, p = 0$	β_1
6		1	$h = 8, p = 0$	β_1
7	Predictor relevance in h-step ahead direct forecasting regression:	1	$h = 12, p = 0$	β_1
8	$\tilde{y}^h_{t+h} = \beta_0 + \sum_{i=1}^{m} \beta_i x_{t-i+1} + \sum_{i=1}^{p} \gamma_i y_{t-i+1} + u_{t+h},$	3	$h = 1, p = 0$	$\beta_1, \beta_2, \beta_3$
9		3	$h = 4, p = 0$	$\beta_1, \beta_2, \beta_3$
10		3	$h = 8, p = 0$	$\beta_1, \beta_2, \beta_3$
11	where $\tilde{y}^h_{t+h} = \begin{cases} y_{t+h} \text{ (}y\text{ is in levels)} \\ y_{t+h} - y_t \text{ (first differences)} \\ \text{see note a (second differences)} \end{cases}$	3	$h = 12, p = 0$	$\beta_1, \beta_2, \beta_3$
12		3	$h = 1, p = 3$	$\beta_1, \beta_2, \beta_3$
13		3	$h = 4, p = 3$	$\beta_1, \beta_2, \beta_3$
14		3	$h = 8, p = 3$	$\beta_1, \beta_2, \beta_3$
15		3	$h = 12, p = 3$	$\beta_1, \beta_2, \beta_3$
16	Local projections with 8 control variables:	1	$h = 1$	β_1
17	$\tilde{y}^h_{t+h} = \beta_0 + \beta_1 x_t + \sum_{i=1}^{2} \lambda_i x_{t-i} +$	1	$h = 4$	β_1
18	$+ \sum_{i=1}^{2} \phi_i y_{t-i} + \sum_{j=1}^{2} \sum_{i=1}^{2} \gamma_{ij} w_{jt-i} + u_{t+h}$	1	$h = 8$	β_1
19		1	$h = 12$	β_1
20		3	$h = 1$	$\beta_1, \beta_2, \beta_3$
21	Multiple predictor direct forecasting with 2 additional predictors:	3	$h = 4$	$\beta_1, \beta_2, \beta_3$
22	$\tilde{y}^h_{t+h} = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_{t+h}$	3	$h = 8$	$\beta_1, \beta_2, \beta_3$
23		3	$h = 12$	$\beta_1, \beta_2, \beta_3$

NOTES: u_t denotes a generic error term and is not used to generate the data (y and x are drawn jointly using the dynamic factor model as explained in Section 6.1).
[a]Some price indexes enter the DFM as second differences of logarithms. Let P_t denote the price series; then $y_t = \Delta \ln(P_t/P_{t-1})$ (the change in the rate of inflation), and $\tilde{y}^h_{t+h} = \ln(P_{t+h}/P_t) - h^{-1} \Delta \ln(P_t/P_{t-1})$ (the difference between the future h-period rate of inflation and the normalized 1-period rate of inflation from date t-1 to date t).

Eight of the specifications (regressions 12–19) include control variables, that is, regressors whose coefficients are not being tested. The control variables are lagged dependent variables (12-15) and, additionally, lags of the variable of interest and two lags each of two additional regressors (16-19).

The degree of persistence of z_t is expected to vary substantially among these specifications, in particular including lags of the variable of interest is expected to reduce the mean trace spectral curvature ω (by Frisch-Waugh, the m regressors of interest can be expressed as innovations to a projection on the other included lags). The regressions with the most persistent z_t would be expected to be 1–11 and 20–23, especially 7, 11, and 23, which have the additional persistence resulting from the overlap induced in the 12-step-ahead direct forecasting regression.

Guide to results. The results are summarized in Table 6, which has four parts, differentiated by whether the errors are Gaussian or empirical and by whether the variance estimator is restricted or unrestricted. The table aggregates over series and focuses on results for tests that choose b and ν using the rules (3) and (4). The online Supplement presents results for each y series individually and for other choices of b and ν. Each row of the tables corresponds to a different regression, using the numbering system of Table 5. In addition, Table 6(a) and (c) have a row 0, which summarizes results for the location model (for the location model, all 207 series were used sequentially for the DGP).

For many DGPs, $\omega^{(q)}$ is negative, indicating anti-persistence. For these series, the tests are undersized in theory and, in most cases, in the finite sample simulations as well. The size entries in Table 6 are therefore signed, not absolute values of size distortions. To be precise, for the location model results in row 0 in Table 6(a), the EWC size distortion is 0.01. Thus, for 95% of the 207 series, the rejection rate of the EWC test under the null hypothesis was at most 0.06; for some of the series (DGPs), the null rejection rate was less than 0.05. Whether the test is under- or over-sized, the size-adjusted power loss is computed the same way as the maximum power loss, compared to the oracle test, using finite-sample size-adjusted critical values (the 5% percentile of the Monte Carlo distribution of the test under the null).

5.3 Results for the Location Model

Before discussing the results in Table 6, we begin by plotting the size-power tradeoffs for six of the data-based DGPs. Specifically, Figure 8 is the counterpart of Figure 3, for

Table 6. Monte Carlo Results for Data-based DGP: 95% percentiles of signed size distortion and power loss using rule (3), rule (4), and the textbook rule

		EWC, $\nu = 0.4T^{2/3}$		NW, $S = 1.3T^{1/2}$		NW-TB, $S = 0.75T^{1/3}$	
Design	m	Size distortion	Power loss	Size distortion	Power loss	Size distortion	Power loss
A. Unrestricted variance estimator, Gaussian errors							
0 (mean)	1	0.01	0.08	0.03	0.08	0.11	0.03
1	1	0.02	0.10	0.03	0.09	0.04	0.05
2	5	0.02	0.30	0.03	0.20	0.03	0.09
3	10	0.02	0.56	0.05	0.28	0.07	0.15
4	1	0.02	0.10	0.02	0.08	0.04	0.05
5	1	0.03	0.11	0.03	0.10	0.05	0.07
6	1	0.03	0.13	0.04	0.12	0.07	0.08
7	1	0.03	0.12	0.04	0.11	0.09	0.07
8	3	0.02	0.20	0.02	0.14	0.03	0.07
9	3	0.02	0.20	0.02	0.12	0.02	0.04
10	3	0.03	0.19	0.02	0.10	0.03	0.02
11	3	0.03	0.17	0.01	0.09	0.03	0.01
12	3	0.02	0.20	0.02	0.15	0.02	0.07
13	3	0.02	0.17	0.02	0.11	0.03	0.02
14	3	0.02	0.16	0.02	0.09	0.03	0.01
15	3	0.02	0.16	0.02	0.09	0.03	0.01
16	1	0.01	0.09	0.01	0.07	0.02	0.03
17	1	0.02	0.09	0.02	0.08	0.02	0.04
18	1	0.03	0.09	0.03	0.08	0.03	0.04
19	1	0.03	0.10	0.03	0.08	0.04	0.05
20	3	0.03	0.21	0.04	0.16	0.06	0.11
21	3	0.04	0.22	0.05	0.18	0.09	0.13
22	3	0.05	0.24	0.07	0.19	0.11	0.16
23	3	0.05	0.22	0.07	0.18	0.13	0.16
B. Restricted variance estimator, Gaussian errors							
1	1	0.01	0.13	0.01	0.11	0.03	0.06
2	5	0.00	0.46	0.00	0.35	0.00	0.18
3	10	0.00	0.71	0.00	0.45	0.00	0.26
4	1	0.01	0.18	0.01	0.16	0.02	0.08
5	1	0.01	0.18	0.02	0.17	0.04	0.10
6	1	0.02	0.16	0.02	0.15	0.05	0.09
7	1	0.02	0.13	0.02	0.11	0.06	0.07
8	3	0.00	0.38	0.00	0.30	0.00	0.15
9	3	0.01	0.39	0.00	0.30	0.00	0.12
10	3	0.01	0.37	0.00	0.24	0.00	0.05
11	3	0.00	0.34	0.01	0.18	0.00	0.03
12	3	0.01	0.30	0.01	0.24	0.01	0.13
13	3	0.00	0.29	0.00	0.21	0.00	0.06
14	3	0.00	0.29	0.00	0.18	0.00	0.03
15	3	0.00	0.27	0.00	0.15	0.00	0.02
16	1	0.01	0.09	0.01	0.08	0.01	0.04
17	1	0.01	0.10	0.01	0.08	0.02	0.04
18	1	0.02	0.10	0.02	0.09	0.03	0.05
19	1	0.02	0.10	0.03	0.09	0.03	0.05
20	3	0.01	0.37	0.01	0.31	0.03	0.18
21	3	0.01	0.39	0.02	0.33	0.04	0.19
22	3	0.02	0.37	0.03	0.30	0.06	0.17
23	3	0.02	0.29	0.03	0.23	0.07	0.14

(*Continued on next page*)

Table 6. (Continued)

		EWC, $\nu = 0.4T^{2/3}$		NW, $S = 1.3T^{1/2}$		NW-TB, $S = 0.75T^{1/3}$	
Design	m	Size distortion	Power loss	Size distortion	Power loss	Size distortion	Power loss
C. Unrestricted variance estimator, empirical errors							
0 (mean)	1	0.01	0.08	0.02	0.07	0.10	0.03
1	1	0.08	0.14	0.08	0.13	0.10	0.08
2	5	0.09	0.29	0.09	0.18	0.10	0.07
3	10	0.07	0.54	0.10	0.23	0.13	0.09
4	1	0.05	0.13	0.05	0.12	0.07	0.08
5	1	0.05	0.15	0.06	0.13	0.08	0.09
6	1	0.05	0.16	0.06	0.14	0.09	0.10
7	1	0.05	0.14	0.06	0.13	0.11	0.09
8	3	0.07	0.19	0.06	0.14	0.06	0.06
9	3	0.08	0.22	0.05	0.11	0.04	0.03
10	3	0.07	0.19	0.03	0.08	0.04	0.01
11	3	0.06	0.16	0.02	0.07	0.04	0.01
12	3	0.05	0.19	0.05	0.14	0.06	0.06
13	3	0.05	0.17	0.04	0.09	0.04	0.01
14	3	0.05	0.15	0.03	0.07	0.04	0.01
15	3	0.05	0.14	0.02	0.06	0.04	0.01
16	1	0.03	0.10	0.03	0.08	0.04	0.04
17	1	0.04	0.09	0.04	0.08	0.05	0.04
18	1	0.04	0.10	0.04	0.09	0.05	0.05
19	1	0.04	0.10	0.04	0.08	0.05	0.05
20	3	0.07	0.23	0.09	0.19	0.12	0.13
21	3	0.08	0.25	0.09	0.19	0.12	0.15
22	3	0.08	0.23	0.10	0.18	0.14	0.14
23	3	0.07	0.20	0.09	0.16	0.14	0.13
D. Restricted variance estimator, empirical errors							
1	1	0.05	0.22	0.04	0.21	0.07	0.11
2	5	0.01	0.47	0.00	0.38	0.01	0.19
3	10	0.00	0.77	−0.01	0.50	0.00	0.28
4	1	0.02	0.23	0.03	0.21	0.04	0.11
5	1	0.03	0.25	0.03	0.23	0.05	0.12
6	1	0.03	0.21	0.03	0.20	0.06	0.12
7	1	0.02	0.21	0.03	0.20	0.07	0.11
8	3	0.01	0.39	0.00	0.33	0.01	0.18
9	3	0.02	0.41	0.00	0.32	0.00	0.12
10	3	0.02	0.39	0.01	0.28	0.00	0.08
11	3	0.01	0.35	0.01	0.22	0.00	0.06
12	3	0.01	0.37	0.01	0.31	0.01	0.15
13	3	0.00	0.33	0.00	0.25	0.00	0.07
14	3	0.00	0.31	0.01	0.20	0.00	0.04
15	3	0.00	0.30	0.01	0.19	0.00	0.04
16	1	0.02	0.20	0.02	0.18	0.02	0.07
17	1	0.02	0.18	0.02	0.17	0.03	0.08
18	1	0.02	0.18	0.03	0.18	0.03	0.08
19	1	0.03	0.20	0.03	0.17	0.04	0.09
20	3	0.02	0.42	0.03	0.36	0.05	0.23
21	3	0.03	0.43	0.03	0.37	0.05	0.23
22	3	0.03	0.37	0.04	0.32	0.07	0.19
23	3	0.02	0.35	0.03	0.30	0.06	0.18

Notes: Entries are the 95% percentiles of the distributions of size distortions, or of power losses, across the 207 DGPs in the location case and the 414 DGPs in the regression case, for the regression model listed in the first column (or, for row 0, the location model), and for the test indicated in the header row. The size distortions are signed, so a size distortion entry of 0.02 means that 95% of the DGPs for that row have a null rejection rate of less than 0.07, and some of those tests could be (in general, are) undersized, with a null rejection rate less than 0.05. The regression designs enumerated in the first column are described in Table 5.

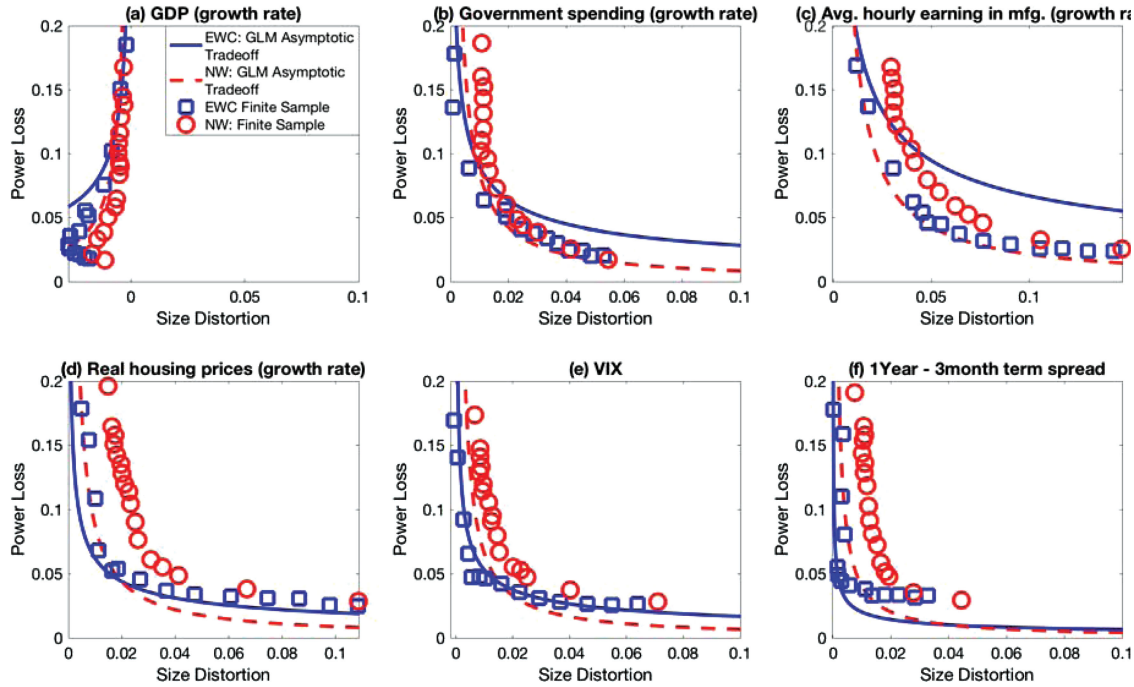

Figure 8. Size-power tradeoffs in the location model with data-based design and empirical errors: Theoretical tradeoff and Monte Carlo results for EWC (solid line & squares) and NW (dashed line & circles), $T = 200$. Notes: The DGPs are calibrated to the series in the figure captions, using the dynamic factor model design described in Section 5.1. The theoretical tradeoffs are calculated based on the values of $\omega^{(1)}$ and $\omega^{(2)}$ implied by the known DGP.

Monte Carlo simulations based on the data-based design. The theoretical tradeoffs are calculated based on the values of $\omega^{(1)}$ and $\omega^{(2)}$ implied by the known DGP. Figure 8 shows results for six series: GDP growth, government spending, average weekly hours in manufacturing, housing prices (OFHEO measure), the VIX, and the 1 year-90 day Treasury spread. The first two of these series are in growth rates and have low persistence, as measured by their AR(1) coefficients. The final four are in levels. Average weekly hours is dominated by business cycle movements, house prices exhibits volatility clustering, and the VIX has large outliers (volatility events). The figures show results for simulations using the empirical errors. Comparable plots for all 207 series, for both Gaussian and empirical errors, are given in the online Supplement; the six series used here were chosen because the results for these series are representative.

Four of the six plots look like those in Figure 1 and are unremarkable, but two merit discussion. The Monte Carlo tradeoff for average weekly hours in manufacturing is substantially better than the theoretical tradeoff for EWC, but is worse than the theory for NW. This pattern is consistent with the series having heavy tails, at least at low frequencies. More noticeable is the plot for GDP, for which the tradeoff slopes the opposite direction. This reversal of the tradeoff arises because, for this DGP, $\omega^{(1)}$ and $\omega^{(2)}$ are negative, that is, the series is antipersistent at low frequencies (recall that we preprocessed the series to remove very low frequency trends). Despite the reversal of slope, the Monte Carlo results approximately align with the theoretical line.

The main conclusion from Figure 8 is that the behavior of the tests in the data-based design is generally similar to that found in the AR designs in Figure 3. As in Figure 3, for most of the series the finite-sample EWC results are close to the theoretical tradeoff and the finite-sample NW results are above the tradeoff, with the exception of manufacturing hours just discussed. In all cases, the Monte Carlo EWC tradeoff is below the NW tradeoff.

The results for all 207 series (DGPs) are summarized in row 0 of Table 6(a) and (c) (as discussed above, the location tests are invariant to whether the null is imposed). For both EWC and NW (using the proposed rules), 95% of the DGPs have small size distortions, respectively less than 0.02 and 0.03 using either Gaussian or empirical errors. As discussed above, the size distortions in Table 6 are signed, and in some of these 95% of DGPs, the tests are undersized (for example, for GDP as seen in Figure 8(a)).

Looking across the results for the location model, for the vast majority of DGPs, the size distortions are small using the EWC and NW tests with the proposed rules, and are much smaller than the textbook test in the final column. The maximum power losses, compared to textbook NW are modest: 0.08 v. 0.03 for both Gaussian and empirical errors. These findings accord with those for the DGPs discussed in Section 4 and in the online Supplement.

5.4 Results for Regressions

The results in Table 6, along with those in the Supplement, suggest five main conclusions.

First, in many cases the unrestricted textbook NW test has large size distortions. In the 23 regressions with empirical errors and the unrestricted NW estimator using the textbook rule, the 95th percentile of size distortions ranges from 0.04 to 0.14, with a mean of 0.077.

Second, in the cases in which unrestricted textbook NW has large size distortions, those can be reduced substantially, often with little loss in power, by using the unrestricted EWC or NW tests with the proposed rules and fixed-b critical values. For example, in regression 7 (a regression with persistent z_t) with empirical errors, the textbook NW size distortion is 0.11, but using the proposed rules the size distortion is only 0.06 for NW and 0.05 for EWC. In this same case, the increase in the maximum power loss is only 0.05 for EWC and 0.04 for NW. In regressions with low persistence (e.g., regressions 16–19), the textbook NW size distortion is less (e.g., 0.05 for regression 19 with empirical errors), as is the improvement in size (to 0.04 in regression 19 for EWC and NW with the proposed rules), albeit still with a small power cost.

Third, the ranking of the EWC and NW estimator depends on the number of restrictions and, to a lesser extent, the persistence. When $m = 1$, the EWC estimator typically has slightly lower size distortion with nearly the same power. For $m = 3$, the EWC estimator tends to have slightly better size, but at a somewhat greater cost in power. For $m = 5$ or 10, the cost in power from using EWC is large. The reason for this cost is that the EWC F-statistic in (14) has denominator degrees of freedom that declines linearly with m, resulting in very large critical values, a feature not shared by the NW test. These results suggest that for large values of m, it could be worth using the larger values of ν for EWC, and possibly the smaller values of b for NW, given in Table 3.

Fourth, these broad patterns hold for the DGPs with Gaussian errors as for empirical errors. The main difference in the two sets of results is that the empirical error DGPs tend to have larger size distortions than the Gaussian error DGPs. After size-adjusting, however, the tests have similar maximum power losses for the Gaussian and empirical errors.

Fifth, consistent with Table 1, using the restricted LRV estimator results in substantial improvements in size; however, those improvements come at the cost of decreases in power that, depending on the design and regression, can be considerable. In the 23 regression designs with empirical errors, the largest size distortion for EWC with the proposed rule using the restricted estimator is 0.05, and in 8 regressions it is 0.01 or less. The size improvements for NW using the proposed rule are comparable. However, the maximal power loss increases are typically in the range of 0.05 to 0.40 when the restricted estimator is used, with the larger increases occurring when m is larger. This is consistent with the standard result that the LM test typically has worse power against moderate to distant alternatives than the Wald test, because the LM test estimates the variance under the null.

6. DISCUSSION AND CONCLUSIONS

Taken together, the theoretical and extensive Monte Carlo results in this paper indicate that using the NW test with the proposed rule (3) or the EWC test with the proposed rule (4), both with fixed-b critical values, can provide substantial improvements in size, with only modest costs in size-adjusted power, compared to the textbook NW test. Although the EWC test frequently dominates (in size-power space) the NW test in the location model, in the regression model there is no clear ranking between the two tests. When the number of restrictions being tested is low, EWC tends to outperform NW, but with multiple restrictions, NW tends to outperform EWC. For $m \leq 3$, these finite-sample performance differences between NW and EWC are small, especially compared with the frequently large improvements that both provide over the textbook NW test.

A surprising finding in our simulations is that large size improvements are possible by imposing the null when estimating Ω, that is, by using the restricted estimator. However, these large size improvements come at a substantial cost in power. The size-power loss plot in Figure 5 suggests that the restricted tradeoff dominates the unrestricted tradeoff, however for the proposed rules, neither dominates (Figure 6). This suggests that perhaps a different rule could retain at least some of the size improvements of the unrestricted estimator while substantially reducing the power cost. Obtaining such a rule is beyond the scope of this article; indeed, the proper starting point for such an investigation would be to develop reliable Edgeworth (or other) approximations for the regression case. To this end, we offer some thoughts. The expansions under the null of the mean and variance of the unrestricted LRV estimator have additional terms that are not present for the restricted estimator. These additional terms arise from the estimation of the regression coefficient and introduce a term involving the spectral density at frequency zero of x_t^2, $\Omega_{\tilde{x}^2}$. For example, the leading terms in the bias of the unrestricted estimator in the case of independent x and u is $[\omega^{(q)}k^{(q)}(0) + b^q T^{q-1}\Omega_{\tilde{x}^2}/\sigma_x^4](bT)^{-q}$, whereas the term in $\Omega_{\tilde{x}^2}$ is not present for the restricted estimator.[10] This term is of an order smaller than those appearing in Equation (18), and numerical investigation suggests that it can explain perhaps half the difference between the Monte Carlo tradeoffs for the restricted and unrestricted EWC estimator in Figure 5. Obtaining a better understanding of the merits, or not, of using the restricted estimator in HAR testing would be an interesting and potentially informative research project.

Based on these results, we recommend the use of either the NW test with the proposed rule (3) or the EWC test with the proposed rule (4), in both cases using the unrestricted estimator of Ω and fixed-b critical values. The rules (3) and (4) provide what we consider to be reasonable compromises between size distortions and power loss, both in theory and in the simulations. These sequences provide a compromise between the extreme emphasis on power loss of the textbook NW test, and the extreme emphasis on size of the KVB test. The relatively good size control of the NW and EWC tests with the proposed rules implies that confidence intervals constructed as $\pm$ the critical value, times the unrestricted standard error, will have better finite-sample coverage rates with little cost to accuracy than the textbook NW intervals.

Given their comparable performance in simulations, the decision about which test to use is a matter of convenience. The EWC test is not currently a standard software option, but it has

[10]As discussed by Hannan (1958), Ng and Perron (1996), Kiefer and Vogelsang (2005), and Sun (2014a), in the location model, estimating the mean produces downward bias in the NW LRV estimator. This bias is accounted for by fixed-b critical values in the case of the location model (and thus these terms do not appear in the higher-order expansions in LLS and this article), but it is not accounted for in the textbook NW case with standard critical values. This issue does not arise with EWP or EWC, or with QS if it is implemented in the frequency domain, because as discussed above these estimators are invariant to location shifts in z_t.

simple-to-use standard t and F critical values; the NW estimator is widely implemented, but our proposal requires nonstandard fixed-b critical values. The NW estimator also has an edge if one of the purposes for computing the standard errors is to test multiple restrictions, especially more than three restrictions. In both cases, the procedures could be conveniently integrated into conventional time series regression software without increasing computation time.

ACKNOWLEDGMENTS

We thank Ulrich Müller, Mikkel Plagborg-Møller, Benedikt Pötscher, Yixiao Sun, Tim Vogelsang, and Ken West for helpful comments and/or discussions. Replication files are available at *http://www.princeton.edu/∼mwatson/*.

[Received December 2017. Revised July 2018.]

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858 [541,542,545]

Andrews, D. W. K., and Monahan, J. C. (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966. [542]

Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502. [542]

Brillinger, D. R. (1975), *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston. [542]

den Haan, W. J., and Levin, A. (1994), "Vector Autoregressive Covariance Matrix Estimation," manuscript, Board of Governors of the Federal Reserve. [543]

——— (1997), "A Practitioners Guide to Robust Covariance Matrix Estimation," *Handbook of Statistics*, 15, 291–341. [543]

——— (2000), "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," NBER Technical Working Paper #255. [542]

Dougherty, C. (2011), *Introduction to Econometrics* (4th Ed.), Oxford: Oxford University Press. [542]

Grenander, U., and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*. New York: Wiley. [542]

Hannan, E. J. (1958), "The Estimation of the Spectral Density After Trend Removal," *Journal of the Royal Statistical Society,* Series B, 20, 323–333. [558]

Hill, R. C., Griffiths, W. F., and Lim, G. C. (2018), *Principles of Econometrics* (4th Ed.), New York: Wiley. [542]

Hillmer, C. E., and Hilmer, M. J. (2014), *Practical Econometrics: Data Collection, Analysis, and Application* (1st ed.), New York: McGraw-Hill. [542]

Hwang, J., and Sun, Y. (2017), "Asymptotic F and t Tests in an Efficient GMM Setting," *Journal of Econometrics*, 198, 277–295. [546]

Ibragimov, R., and Müller, U. K. (2010), "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468. [542]

Inoue, A., and Shintani, M. (2006), "Bootstrapping GMM Estimators for Time Series," *Journal of Econometrics*, 133, 531–555. [546]

Jansson, M. (2004), "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937–946. [542]

Kiefer, N., Vogelsang, T. J., and Bunzel, H. (2000), "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 69, 695–714. [542,544]

Kiefer, N., and Vogelsang, T. J. (2002), "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093–2095. [542,545]

——— (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 2093–2095. [542,545]

Lazarus, E., Lewis, D. J., and Stock, J. H. (2017), "The Size-Power Tradeoff in HAR Inference," manuscript, Harvard University. [542]

MacKinnon, J. G. (2006), "Bootstrap Methods in Econometrics," *Economic Record*, 82, s2–s18. [553]

Marron, J. S., and Wand, M. P. (1992), "Exact Mean Integrated Squared Error," *Annals of Statistics*, 20, 2, 712–736 [548]

Müller, U. (2004), "A Theory of Robust Long-Run Variance Estimation," Working paper, Princeton University. [546]

——— (2014), "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322. [541,542]

Newey, W. K., and West, K. D. (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [541]

Ng, S., and Perron, P. (1996), "The Exact Error in Estimating the Spectral Density at the Origin," *Journal of Time Series Analysis*, 17, 379–408. [558]

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348. [546]

Phillips, P. C. B. (2005), "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142. [542,545]

Politis, D. (2011), "Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices." *Econometric Theory*, 27, 703–744. [542]

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, London: Academic Press. [545]

Rao, C. R. (1973), *Linear Statistical Inference and its Application, second edition*, New York: Wiley. [546]

Rothenberg, T. G. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics." in *Handbook of Econometrics, Volume II*, eds. Z. Griliches and M. D. Intriligator, Amsterdam: Elsevier, pp. 881–935. [546,547]

Stock, J. H., and Watson, M. W. (2015), *Introduction to Econometrics*, (3rd ed., update), Boston: Addison-Wesley. [542]

——— (2016), "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." in *Handbook of Macroeconomics*, volume 2A, eds. J. Taylor and H. Uhlig, Amsterdam: Elsevier, pp. 415–525 [551,552]

Stoica, P., and Moses, R. (2005), *Spectral Analysis of Signals*, Englewood Cliffs, NJ: Pearson Prentice Hall. [542]

Sun, Y. (2013), "Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator," *The Econometrics Journal*, 16, 1–26. [542,545,546]

——— (2014a), "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroskedasticity and Autocorrelation Robust Inference," *Journal of Econometrics*, 178, 659–677. [542,545,546]

——— (2014b), "Fixed-smoothing Asymptotics in a Two-step GMM Framework," *Econometrica*, 82, 2327–2370. [546]

Sun, Y., and Kaplan, D. M. (2014), "Fixed-smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator," manuscript, UC-San\sDiego. [542,547]

Sun, Y., Phillips, P. C. B., and Jin, S. (2008), "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76(1): 175–194. [542,546,547]

Sun, Y., and Phillips, P. C. B. (2009), "Bandwidth Choice for Interval Estimation in GMM Regression," Working paper, Department of Economics, UC San Diego. [546]

Velasco, C., and Robinson, P. M. (2001), "Edgeworth Expansions for Spectral Density Estimates and Studentized Sample Mean," *Econometric Theory*, 17, 497–539. [542,546,550]

Westhoff, F. (2013), *An Introduction to Econometrics: A Self-Contained Approach*, Cambridge: MIT Press. [542]

Wooldridge, J. M. (2012), *Introductory Econometrics: A Modern Approach* (4th Ed.), Thomson. [542]

HAR Inference: Recommendations for Practice Rejoinder

Eben Lazarus
MIT Sloan School of Management, Cambridge, MA 02142 (*elazarus@mit.edu*)

Daniel J. Lewis
Federal Reserve Bank of New York, New York, NY 10045 (*daniel.lewis@ny.frb.org*)

James H. Stock
Department of Economics, Harvard University and the National Bureau of Economic Research, Cambridge, MA 02138 (*james_stock@harvard.edu*)

Mark W. Watson
Department of Economics and the Woodrow Wilson School, Princeton University, Princeton, NJ 08544-1013, and the National Bureau of Economic Research, Cambridge, MA 02138 (*mwatson@Princeton.edu*)

We are lucky to have such distinguished discussants and for them to provide such thoughtful comments and suggestions. We agree with much of what they say, and in this brief rejoinder we focus on a few common threads in their remarks.

First and most importantly, our aim has been to suggest a new convention for HAR inference that is informed by and builds on the massive body of research on HAR inference over the past twenty years. This research—which builds on fundamental contributions by our discussants—has been deep and has yielded important insights, yet it is largely ignored in empirical practice. It is time for that to change. The proposed rules embody those insights and turn them into concrete recommendations. It is true that the truncation parameters are unnecessarily large in some of our designs, but not in all of them, and if a single rule that is not data dependent is to be recommended, it should control size in situations with moderate dependence, not just low dependence.

Second, as West points out, in practice HAR inference might be used for testing individual coefficients, confidence intervals, and testing joint hypotheses, all in the same regression. Given the better power, and comparable size, of the NW test than EWC for testing multiple hypotheses, in our view this gives an edge to using NW with fixed-b asymptotics and our proposed rule.

Third, we neither pretend nor hope to have the last word on this topic. Indeed a number of the suggestions by the discussants are meritorious, and we have left some stones unturned. Müller's suggestion that the EWC estimator have degrees of freedom ν capped at 20 makes sense; at that point, the power loss is minimal compared to the oracle estimator, but capping ν controls the rejection rate under the null in a very large data set. We did not look at VARHAC estimators, but they should be on the table too, although deriving a lag length rule would require developing fixed-b theory for VARHAC. We have intentionally avoided data-dependent rules (an idea raised or implicit in the comments by Müller, Sun, and West), but for sophisticated practitioners data dependence might be warranted; in any event, data-dependent rules could be incorporated into standard software, making them easy to use. And there is clearly more to do concerning understanding whether it is desirable to impose the null for testing; here, Vogelsang's comments make good progress beyond what is in our article. Researchers should continue to investigate these and other methods. In this regard, we hope that our new data-based DFM design will provide a useful test bed for other researchers, and we encourage them to use it (the core software is compact and is available online). We think the design is sufficiently rich that it is unlikely that it can be "gamed" to produce tests that work well in this design but not in reality or in other settings with moderate dependence.

Fourth, we have followed LLS in approaching the truncation parameter choice through the lens of trading off the null rejection rate and size-adjusted power. As both West and Sun point out, a different approach is to start with a tradeoff between Type I and Type II errors, as done by Sun, Phillips, and Jin (2008). One can argue the intellectual merits of the two approaches but, in keeping with the practical motivation of this article, we instead point out that the two approaches have similar implications. In particular, by minimizing the linear loss function proposed in the comments by Sun—that is, $b^*_{SPJ} = \arg\min_b (w/(1+w))e_{\mathrm{I}}(b) + (1/(1+w))e_{\mathrm{II}}(b, \delta^2)$, where w is the ratio of the weight on the Type I error $e_{\mathrm{I}}(b)$ relative to the

© 2018 American Statistical Association
Journal of Business & Economic Statistics
October 2018, Vol. 36, No. 4
DOI: 10.1080/07350015.2018.1513251

weight on the Type II error $e_{\mathrm{II}}(b, \delta^2)$ and where the test uses fixed-b critical values – we obtain the optimal SPJ rule,

$$b^*_{\mathrm{SPJ}} = b_{0,\mathrm{SPJ}} T^{-q/(1+q)},$$

$$\text{where } b_{0,\mathrm{SPJ}} = \left[\frac{2q\left(wG'_m(\chi^\alpha_m) - G'_{m,\delta^2}(\chi^\alpha_m)\right)k^{(q)}(0)\omega^{(q)}}{\delta^2 G'_{m+2,\delta^2}(\chi^\alpha_m)\int k^2}\right]^{1/(1+q)}, \tag{1}$$

where terms are as defined in the text of our article.[1] This expression applies for positive low-frequency correlation, i.e., $\omega^{(1)}, \omega^{(2)} > 0$, which is the relevant case if the truncation parameter choice is not data-dependent.

The key point is that the rate in (1) is the same as in our Equation (22) for the size/size-adjusted power tradeoff; the only difference is the expression for the constant term. In particular, it can be seen that there exists some choice of w such that the sequence in (1) is equivalent to the sequence provided by the rule provided by our Equation (22). Assuming the local alternative δ is the worst-case alternative from a size-adjusted-power standpoint (as in our article), one obtains that

$$b_{0,\mathrm{SPJ}} = b_0 \kappa_q^{-1}\left(w - \frac{G'_{m,\delta^2}(\chi^\alpha_m)}{G'_m(\chi^\alpha_m)}\right)^{1/(1+q)} q^{1/(1+q)}, \tag{2}$$

where b_0 is the optimal constant for our rule given in our Equation (22) and κ_q is defined following (22). Solving (2) for $b_{0,\mathrm{SPJ}} = b_0$ yields the relation between κ and w for which the two approaches yield the same rules.

When (2) is evaluated for the NW and EWC/EWP kernels, 5% tests, an AR(1) with coefficient $\rho = 0.7$ (what we use for our rules), and $w = 9$ (which corresponds to our weight of $\kappa = 0.9$ on the squared size distortion and is within the range of values considered by Sun (2014a)), the Sun, Phillips, and Jin (2008) rule (1) implies,

$$S^*_{\mathrm{SPJ}} = 1.9T^{1/2} \text{ (NW Type I/II optimal)} \tag{3}$$

$$\nu^*_{\mathrm{SPJ}} = 0.3T^{2/3} \text{ (EWC/EWP Type I/II optimal)} \tag{4}$$

These are close to our proposed rules (respectively, $1.3T^{1/2}$ and $0.4T^{2/3}$). Thus, rules quite similar to ours, with identical asymptotic order, also emerge from the Type I/II approach.

We thank again the discussants for their efforts and insightful comments.

[1]We follow Sun's comment and refer to this as the SPJ rule, however this rule follows from the setup in Sun's comment but not strictly from Sun, Phillips, and Jin (2008): Sun's comment uses fixed-b critical values (as do we) whereas the SPJ optimal rule uses a polynomial correction to standard normal critical values.